INTERNATIONAL STANDARD

ISO 4259

Third edition
2006-08-01

Petroleum products — Determination and application of precision data in relation to methods of test

Produits pétroliers — Détermination et application des valeurs de fidélité relatives aux méthodes d'essai

Reference number
ISO 4259:2006(E)

PDF disclaimer

This PDF file may contain embedded typefaces. In accordance with Adobe's licensing policy, this file may be printed or viewed but shall not be edited unless the typefaces which are embedded are licensed to and installed on the computer performing the editing. In downloading this file, parties accept therein the responsibility of not infringing Adobe's licensing policy. The ISO Central Secretariat accepts no liability in this area.

Adobe is a trademark of Adobe Systems Incorporated.

Details of the software products used to create this PDF file can be found in the General Info relative to the file; the PDF-creation parameters were optimized for printing. Every care has been taken to ensure that the file is suitable for use by ISO member bodies. In the unlikely event that a problem relating to it is found, please inform the Central Secretariat at the address given below.

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 ... v

Introduction ... vi

1 Scope ... 1

2 Normative references ... 1

3 Terms and definitions ... 1

4 Stages in the planning of an inter-laboratory test programme for the determination of the precision of a test method ... 4
4.1 General ... 4
4.2 Preparing a draft method of test ... 5
4.3 Planning a pilot programme with at least two laboratories ... 5
4.4 Planning the inter-laboratory programme ... 5
4.5 Executing the inter-laboratory programme ... 5

5 Inspection of inter-laboratory results for uniformity and for outliers ... 6
5.1 General ... 6
5.2 Transformation of data ... 7
5.3 Tests for outliers ... 8
5.4 Rejection of complete data from a sample ... 11
5.5 Estimating missing or rejected values ... 12
5.6 Rejection test for outlying laboratories ... 13
5.7 Confirmation of selected transformation ... 14

6 Analysis of variance, calculation and expression of precision estimates ... 14
6.1 General ... 14
6.2 Analysis of variance ... 14
6.3 Expectation of mean squares and calculation of precision estimates ... 17
6.4 Expression of precision estimates of a method of test ... 20

7 Significance of repeatability (r) and reproducibility (R) ... 21
7.1 General ... 21
7.2 Repeatability, r ... 21
7.3 Reproducibility, R ... 22

8 Specifications ... 24
8.1 Aim of specifications ... 24
8.2 Construction of specifications limits in relation to precision ... 24

9 Quality control against specifications ... 25
9.1 General ... 25
9.2 Testing margin at the supplier ... 25
9.3 Testing margin at the recipient ... 25

10 Dispute procedure ... 25

Annex A (normative) **Determination of number of samples required** ... 28

Annex B (informative) **Derivation of equation for calculating the number of samples required** ... 30

Annex C (normative) **Notation and tests** ... 31

Annex D (normative) **Example results of test for determination of bromine number and statistical tables** ... 36

Annex E (normative) **Types of dependence and corresponding transformations** ... 45

Annex F (normative) **Weighted linear regression analysis** ... 49

Annex G (normative) **Rules for rounding off results** **56**
Annex H (informative) **Explanation of equations given in Clause 7** **57**
Annex I (informative) **Specifications that relate to a specified degree of criticality** **59**
Bibliography **62**

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.

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 4259 was prepared by Technical Committee ISO/TC 28, *Petroleum products and lubricants*.

This third edition cancels and replaces the second edition (ISO 4259:1992), Clauses 1, 5, 7 C.7, E.2 and F.3 and subclauses 4.2, 5.2, 6.3.2, 6.3.3.1, 6.3.3.3, 6.4, 8.2, 10.2, 10.4 and 10.5, which have been technically revised. It also incorporates the Technical Corrigendum ISO 4259:1992/Cor.1:1993.

Introduction

For purposes of quality control and to check compliance with specifications, the properties of commercial petroleum products are assessed by standard laboratory test methods. Two or more measurements of the same property of a specific sample by any given test method do not usually give exactly the same result. It is, therefore, necessary to take proper account of this fact, by arriving at statistically-based estimates of the precision for a method, i.e. an objective measure of the degree of agreement expected between two or more results obtained in specified circumstances.

ISO 4259 makes reference to ISO 3534-2 [11], which gives a different definition of true value (see 3.26). ISO 4259 also refers to ISO 5725-2. The latter is required in particular and unusual circumstances (see 5.2) for the purpose of estimating precision.

Petroleum products — Determination and application of precision data in relation to methods of test

1 Scope

This International Standard covers the calculation of precision estimates and their application to specifications. In particular, it contains definitions of relevant statistical terms (Clause 3), the procedures to be adopted in the planning of an inter-laboratory test programme to determine the precision of a test method (Clause 4), the method of calculating the precision from the results of such a programme (Clauses 5 and 6), and the procedure to be followed in the interpretation of laboratory results in relation both to precision of the test methods and to the limits laid down in specifications (Clauses 7 to 10).

The procedures in this International Standard have been designed specifically for petroleum and petroleum-related products, which are normally homogeneous. However, the procedures described in this International Standard can also be applied to other types of homogeneous products. Careful investigations are necessary before applying this International Standard to products for which the assumption of homogeneity can be questioned.

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 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*

3 Terms and definitions

For the purposes of this document, the following terms and definitions apply.

3.1
analysis of variance
technique that enables the total variance of a method to be broken down into its component factors

3.2
between-laboratory variance
element of the total variance attributable to the difference between the mean values of different laboratories

NOTE 1 When results obtained by more than one laboratory are compared, the scatter is usually wider than when the same number of tests are carried out by a single laboratory, and there is some variation between means obtained by different laboratories. These give rise to the between-laboratory variance which is that component of the overall variance due to the difference in the mean values obtained by different laboratories.

NOTE 2 There is a corresponding definition for between-operator variance.

NOTE 3 The term “between-laboratory” is often shortened to “laboratory” when used to qualify representative parameters of the dispersion of the population of results, for example as “laboratory variance”.

3.3
bias
difference between the true value (related to the method of test) and the known value, where this is available

NOTE For a definition of “true value” and “known value,” see 3.26 and 3.8, respectively.

3.4
blind coding
assignment of a different number to each sample so that no other identification or information on any sample is given to the operator

3.5
check sample
sample taken at the place where the product is exchanged, i.e. where the responsibility for the product quality passes from the supplier to the recipient

3.6
degrees of freedom
divisor used in the calculation of variance; one less than the number of independent results

NOTE The definition applies strictly only in the simplest cases. Complete definitions are beyond the scope of this International Standard.

3.7
determination
process of carrying out the series of operations specified in the test method, whereby a single value is obtained

3.8
known value
actual quantitative value implied by the preparation of the sample

NOTE The known value does not always exist, for example for empirical tests such as flash point.

3.9
mean
arithmetic mean
sum of the results divided by their number for a given set of results

3.10
mean square
sum of squares divided by the degrees of freedom

3.11
normal distribution
probability distribution of a continuous random variable, x, such that, if x is any real number, the probability density is

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma} \right)^2 \right], -\infty < x < \infty \qquad (1)$$

NOTE μ is the true value and σ is the standard deviation of the normal distribution ($\sigma > 0$).

3.12
operator
person who normally and regularly carries out a particular test

3.13
outlier
result far enough in magnitude from other results to be considered not a part of the set

3.14
precision
closeness of agreement between the results obtained by applying the experimental procedure several times on identical materials and under prescribed conditions

NOTE The smaller the random part of the experimental error, the more precise is the procedure.

3.15
random error
chance variation encountered in all test work despite the closest control of variables

3.16
recipient
any individual or organization who receives or accepts the product delivered by the supplier

3.17
repeatability
⟨qualitatively⟩ closeness of agreement between independent results obtained in the normal and correct operation of the same method on identical test material, in a short interval of time, and under the same test conditions (same operator, same apparatus, same laboratory)

NOTE The representative parameters of the dispersion of the population that can be associated with the results are qualified by the term “repeatability”, for example, repeatability standard deviation or repeatability variance. It is important that the term “repeatability” not be confused with the terms “between repeats” or “repeats” when used in this way (see 3.19). Repeatability refers to the state of minimum random variability of results. The period of time during which repeated results are to be obtained should therefore be short enough to exclude time-dependent errors, for example, environmental and calibration errors.

3.18
repeatability
⟨quantitatively⟩ value equal to or below which the absolute difference between two single test results obtained in the conditions specified that can be expected to lie with a probability of 95 %

NOTE For the details of the conditions specified, see 3.17.

3.19
replication
execution of a test method more than once so as to improve precision and to obtain a better estimation of testing error

NOTE Replication should be distinguished from repetition in that the former implies that repeated experiments are carried out at one place and, as far as possible, within one period of time. The representative parameters of the dispersion of the population that can be associated with repeated experiments are qualified by the term “between repeats”, or in shortened form “repeats”, for example, “repeats standard deviation”.

3.20
reproducibility
⟨qualitatively⟩ closeness of agreement between individual results obtained in the normal and correct operation of the same method on identical test material but under different test conditions (different operators, different apparatus and different laboratories)

NOTE The representative parameters of the dispersion of the population that can be associated with the results are qualified by the term “reproducibility”, for example, reproducibility standard deviation or reproducibility variance.

3.21
reproducibility
⟨quantitatively⟩ value equal to or below which the absolute difference between two single test results on identical material obtained by operators in different laboratories, using the standardized test method, may be expected to lie with a probability of 95 %

3.22
result
final value obtained by following the complete set of instructions in the test method; it may be obtained from a single determination or from several determinations depending on the instructions in the method

NOTE It is assumed that the result is rounded off according to the procedure specified in Annex G.

3.23
standard deviation
measure of the dispersion of a series of results around their mean, equal to the positive square root of the variance and estimated by the positive square root of the mean square

3.24
sum of squares
sum of squares of the differences between a series of results and their mean

3.25
supplier
any individual or organization responsible for the quality of a product just before it is taken over by the recipient

3.26
true value
for practical purposes, the value towards which the average of single results obtained by n laboratories tends, as n tends towards infinity

NOTE 1 Such a true value is associated with the particular method of test.

NOTE 2 A different and idealized definition is given in ISO 3534-2 [11].

3.27
variance
mean of the squares of the deviation of a random variable from its mean, estimated by the mean square

4 Stages in the planning of an inter-laboratory test programme for the determination of the precision of a test method

4.1 General

The stages in planning an inter-laboratory test programme are as follows:

a) preparing a draft method of test;

b) planning a pilot programme with at least two laboratories;

c) planning the inter-laboratory programme;

d) executing the inter-laboratory programme.

The four stages are described in turn in 4.2 to 4.5.

4.2 Preparing a draft method of test

This shall contain all the necessary details for carrying out the test and reporting the results. Any condition that could alter the results shall be specified.

A clause on precision is included in the draft method of the test at this stage only as a heading. It is recommended that the lower limit of the scope of the test method is not less than the region of the lowest value tested in the inter-laboratory programme, and is at least $2R$ greater than the lowest achievable result (see 8.2), where R is the reproducibility estimate. Similarly, it is recommended that the upper limit of the scope of a test method is not greater than the region of the highest value tested in the inter-laboratory programme, and is at least $2R$ less than the highest achievable result.

4.3 Planning a pilot programme with at least two laboratories

A pilot programme is necessary for the following reasons:

a) to verify the details in the operation of the test;

b) to find out how well operators can follow the instructions of the method;

c) to check the precautions regarding samples;

d) to estimate approximately the precision of the test.

At least two samples are required, covering the range of results to which the test method is intended to apply; however, at least twelve laboratory/sample combinations shall be included. Each sample is tested twice by each laboratory under repeatability conditions. If any omissions or inaccuracies in the draft test method are revealed, they shall now be corrected. The results shall be analysed for bias and precision; if either is considered to be too large, then alterations to the test method shall be considered.

4.4 Planning the inter-laboratory programme

There shall be at least five participating laboratories, but it is preferable that there are more in order to reduce the number of samples required.

The number of samples shall be sufficient to cover the range of the property measured at approximately equidistant intervals and to give reliability to the precision estimates. If precision is found to vary with the level of results in the pilot programme, then at least five samples shall be used in the inter-laboratory programme. In any case, it is necessary to obtain at least 30 degrees of freedom in both repeatability and reproducibility. For repeatability, this means obtaining a total of at least 30 pairs of results in the programme.

For reproducibility, Table A.1 gives the minimum number of samples required in terms of L, P and Q, where L is the number of participating laboratories, and P and Q are the ratios of variance component estimates obtained from the pilot programme. Specifically, P is the ratio of the interaction component to the repeats component and Q is the ratio of the laboratories component to the repeats component. Annex B gives the derivation of the equation used. If Q is much larger than P, then 30 degrees of freedom cannot be achieved; the blank entries in Table A.1 correspond to, or an approach to, this situation (i.e. when more than 20 samples are required). For these cases, there is likely to be a significant bias between laboratories.

4.5 Executing the inter-laboratory programme

One person shall be responsible for the entire programme, from the distribution of the texts of the test method and samples to the final appraisal of the results. He shall be familiar with the test method, but shall not personally take part in the tests.

The text of the test method shall be distributed to all the laboratories in time to allow any queries to be raised before the tests begin. If any laboratory wants to practice the method in advance, this shall be carried out with samples other than those used in the programme.

The samples shall be accumulated, subdivided and distributed by the organizer, who shall also keep a reserve of each sample for emergencies. It is most important that the individual laboratory portions be homogeneous. They shall be blind coded before distribution and the following information shall be sent with them:

a) agreed (draft) method of test;

b) handling and storage requirements for the samples;

c) order in which the samples are to be tested (a different random order for each laboratory);

d) statement that two results shall be obtained consecutively on each sample by the same operator with the same apparatus. For statistical reasons, it is imperative that the two results are obtained independently of each other, that is, that the second result is not biased by knowledge of the first. If this is regarded as impossible to achieve with the operator concerned, then the pairs of results shall be obtained in a blind fashion, but ensuring that they are carried out in a short period of time;

e) period of time during which repeated results are to be obtained and the period of time during which all the samples are to be tested;

f) blank form for reporting the results. For each sample, there shall be space for the date of testing, the two results, and any unusual occurrences. The unit of accuracy for reporting the results shall be specified;

g) statement that the test shall be carried out under normal conditions, using operators with good experience but not exceptional knowledge and that the duration of the test shall be the same as normal.

The pilot-programme operators may take part in the inter-laboratory programme. If their extra experience in testing a few more samples produces a noticeable effect, it serves as a warning that the test method is not satisfactory. They shall be identified in the report of the results so that any effect can be noted.

5 Inspection of inter-laboratory results for uniformity and for outliers

5.1 General

In 5.2 to 5.7, procedures are specified for examining the results reported in a statistically designed inter-laboratory programme (see Clause 4) in order to establish the following:

a) independence or dependence of precision and the level of results;

b) uniformity of precision from laboratory to laboratory;

c) and to detect the presence of outliers.

The procedures are described in mathematical terms based on the notation of Annex C and illustrated with reference to the example data (calculation of bromine number) set out in Annex D.

Throughout 5.2 to 5.7 (and Clause 6), the procedures used are first specified and then illustrated by a worked example using data given in Annex D.

It is assumed throughout this clause that all the results are either from a single normal distribution or capable of being transformed into such a distribution (see 5.2). Other cases (which are rare) require a different treatment that is beyond the scope of this International Standard. See Reference [8] for a statistical test on normality.

Although the procedures shown here are in a form suitable for hand calculation, it is strongly advised that an electronic computer with appropriately validated software be used to store and analyse inter-laboratory test results, based on the procedures of this International Standard (see, for example, Reference [9]).

5.2 Transformation of data

5.2.1 General

In many test methods, the precision depends on the level of the test result, and thus the variability of the reported results is different from sample to sample. The method of analysis outlined in this International Standard requires that this shall not be so and the position is rectified, if necessary, by a transformation.

The laboratories standard deviations, D_j, and the repeats standard deviations, d_j, for sample j (see Annex C) are calculated and plotted separately against the sample means, m_j. If the points so plotted can be considered as lying about a pair of lines parallel to the m-axis, then no transformation is necessary. If, however, the plotted points describe non-horizontal straight lines or curves of the form $D = f_1(m)$ and $d = f_2(m)$, then a transformation is necessary.

The relationships $D = f_1(m)$ and $d = f_2(m)$ are not, in general, identical. The statistical procedures of this International Standard require, however, that the same transformation be applicable both for repeatability and for reproducibility. For this reason, the two relationships are combined into a single dependency relationship $D = f(m)$ (where D now includes *d)* by including a dummy variable, T. This takes account of the difference between the relationships, if one exists, and provides a means of testing for this difference (see Clause F.1).

The single relationship $D = f(m)$ is best estimated by a weighted linear regression analysis, even though in most cases an unweighted regression gives a satisfactory approximation. The derivation of weights is described in Clause F.2, and the computational procedure for the regression analysis is described in Clause F.3. Typical forms of dependence $D = f(m)$ are given in Clause E.1. These are all expressed in terms of transformation parameters B and B_0.

The estimation of B and B_0, and the transformation procedure which follows, are summarized in Clause E.2. This includes statistical tests for the significance of the regression (i.e. is the relationship $D = f(m)$ parallel to the m-axis), and for the difference between the repeatability and reproducibility relationships, based at the 5 % significance level. If such a difference is found to exist, or if no suitable transformation exists, then the alternative sample-by-sample procedures of ISO 5725-2 shall be used. In such an event, it is not possible to test for laboratory bias over all samples (see 5.6) or separately estimate the interaction component of variance (see 6.2).

If it has been shown at the 5 % significance level that there is a significant regression of the form $D = f(m)$, then the appropriate transformation $y = F(x)$, where x is the reported result, is given by the equation:

$$F(x) = K \int \frac{dx}{f(x)} \tag{2}$$

where K is a constant. In that event, all results shall be transformed accordingly and the remainder of the analysis carried out in terms of the transformed results. Typical transformations are given in Clause E.1.

It is difficult to make the choice of transformation the subject of formalized rules. Qualified statistical assistance can be required in particular cases. The presence of outliers can affect judgement as to the type of transformation required, if any (see 5.7).

5.2.2 Worked example

Table 1 lists the values of m, D, and d for the eight samples in the example given in Annex D, correct to three significant digits. Corresponding degrees of freedom are in parentheses.

Table 1

Sample number	3	8	1	4	5	6	2	7
m	0,756	1,22	2,15	3,64	10,9	48,2	65,4	114
D	0,066 9 (14)	0,159 (9)	0,729 (8)	0,211 (11)	0,291 (9)	1,50 (9)	2,22 (9)	2,93 (9)
d	0,050 0 (9)	0,057 2 (9)	0,127 (9)	0,116 (9)	0,094 3 (9)	0,527 (9)	0,818 (9)	0,935 (9)

Inspection of the figures in Table 1 shows that both D and d increase with m, the rate of increase diminishing as m increases. A plot of these figures on log-log paper (i.e. a graph of log D and log d against log m) shows that the points may reasonably be considered as lying about two straight lines (see Figure F.1) From the example calculations given in Clause F.4, the gradients of these lines are shown to be the same, with an estimated value of 0,638. Bearing in mind the errors in this estimated value, the gradient may, for convenience, be taken as 2/3.

Hence, the same transformation is appropriate both for repeatability and reproducibility, and is given by the equation:

$$\int x^{-2/3} \mathrm{d}x = 3x^{1/3} \tag{3}$$

Since the constant multiplier may be ignored, the transformation thus reduces to that of taking the cube roots of the reported results (bromine numbers). This yields the transformed data shown in Table D.2, in which the cube roots are quoted correct to three decimal places.

5.3 Tests for outliers

5.3.1 General

The reported data, or if it has been decided that a transformation is necessary, the transformed results, shall be inspected for outliers. These are the values that are so different from the remaining data that it can only be concluded that they have arisen from some fault in the application of the test method or from testing a wrong sample. Many possible tests may be used and the associated significance levels varied, but those that are given below have been found to be appropriate for this International Standard. These outlier tests all assume a normal distribution of errors (see 5.1).

5.3.2 Uniformity of repeatability

5.3.2.1 General

The first outlier test is concerned with detecting a discordant result in a pair of repeat results. This test[1] involves calculating the e^2_{ij} over all the laboratory/sample combinations. Cochran's criterion at the 1 % significance level is then used to test the ratio of the largest of these e^2_{ij} values over their sum (see Clause C.5). If its value exceeds the value given in Table D.3, corresponding to one degree of freedom, n being the number of pairs available for comparison, then the member of the pair farthest from the sample mean shall be rejected and the process repeated, reducing n by 1, until no more rejections are called for. In certain cases, this test "snowballs" and leads to an unacceptably large proportion of rejections (say more than 10 %). If this is so, this rejection test shall be abandoned and some or all of the rejected results shall be retained. An arbitrary decision based on judgement is necessary in this case.

5.3.2.2 Worked example

In the case of the example given in Annex D, the absolute differences (ranges) between transformed repeat results, i.e. of the pairs of numbers in Table D.2, in units of the third decimal place, are shown in Table 2.

Table 2

Laboratory	Sample							
	1	2	3	4	5	6	7	8
A	42	21	7	13	7	10	8	0
B	23	12	12	0	7	9	3	0
C	0	6	0	0	7	8	4	0
D	14	6	0	13	0	8	9	32
E	65	4	0	0	14	5	7	28
F	23	20	34	29	20	30	43	0
G	62	4	78	0	0	16	18	56
H	44	20	29	44	0	27	4	32
J	0	59	0	40	0	30	26	0

The largest range is 0,078 for laboratory G on sample 3. The sum of squares of all the ranges is

$$0{,}042^2 + 0{,}021^2 + \ldots + 0{,}026^2 + 0^2 = 0{,}043\ 9$$

Thus, the ratio to be compared with Cochran's criterion is $\frac{0{,}078^2}{0{,}043\ 9} = 0{,}138$

There are 72 ranges and, as from Table D.3, the criterion for 80 ranges is 0,170 9, this ratio is not significant.

5.3.3 Uniformity of reproducibility

5.3.3.1 General

The following outlier tests are concerned with establishing uniformity in the reproducibility estimate and are designed to detect either a discordant pair of results from a laboratory on a particular sample or a discordant set of results from a laboratory on all samples. For both purposes, the Hawkins' test[2] is appropriate.

This involves forming for each sample, and finally for the overall laboratory averages (see 5.6), the ratio of the largest absolute deviation of laboratory mean from sample (or overall) mean to the square root of certain sums of squares (see Clause C.6).

The ratio corresponding to the largest absolute deviation shall be compared with the critical 1 % values given in Table D.4, where n is the number of laboratory/sample cells in the sample (or the number of overall laboratory means) concerned and where ν is the degrees of freedom for the sum of squares, which is additional to that corresponding to the sample in question. In the test for laboratory/sample cells, ν refers to other samples, but is zero in the test for overall laboratory averages.

If a significant value is encountered for individual samples, the corresponding extreme values shall be omitted and the process repeated. If any extreme values are found in the laboratory totals, then all the results from that laboratory shall be rejected.

If the test “snowballs”, leading to an unacceptably large proportion of rejections (say more than 10 %), then this rejection test shall be abandoned and some or all of the rejected results shall be retained. An arbitrary decision based on judgement is necessary in this case.

5.3.3.2 Worked example

The application of Hawkins' test to cell means within samples is shown below.

The first step is to calculate the deviations of cell means from respective sample means over the whole array. These are shown in Table 3, in units of the third decimal place.

The sum of squares of the deviations are then calculated for each sample. These are also shown in Table 3 in units of the third decimal place.

Table 3

Laboratory	Sample							
	1	2	3	4	5	6	7	8
A	20	8	14	15	10	48	6	3
B	75	7	20	9	10	47	6	3
C	64	35	3	20	30	4	22	25
D	314	33	18	42	7	39	80	50
E	32	32	30	9	7	18	18	39
F	75	97	31	20	30	8	74	53
G	10	34	32	20	20	61	9	62
H	42	13	4	42	13	21	8	50
J	1	28	22	29	14	8	10	53
Sum of squares	117	15	4	6	3	11	13	17

The cell tested is the one with the most extreme deviation. This was obtained by laboratory D from sample 1. The appropriate Hawkins' test ratio is therefore

$$B^* = \frac{0,314}{\sqrt{0,117 + 0,015 + ... + 0,017}} = 0,728\,1$$

The critical value, corresponding to $n = 9$ cells in sample 1 and $\nu = 56$ extra degrees of freedom from the other samples, is interpolated from Table D.4 as 0,372 9. The test value is greater than the critical value and so the results from laboratory D on sample 1 are rejected.

As there has been a rejection, the mean value, deviations and sum of squares are recalculated for sample 1, and the procedure is repeated. The next cell to be tested is that obtained by laboratory F from sample 2. The Hawkins' test ratio for this cell is:

$$B^* = \frac{0,097}{\sqrt{0,006 + 0,015 + ... + 0,017}} = 0,354\,2$$

The critical value corresponding to $n = 9$ cells in sample 2 and $\nu = 55$ extra degrees of freedom is interpolated from Table D.4 as 0,375 6. As the test ratio is less than the critical value, there are no further rejections.

5.4 Rejection of complete data from a sample

5.4.1 General

The laboratories standard deviation and repeats standard deviation shall be examined for any outlying samples. If a transformation has been carried out or any rejection made, new standard deviations shall be calculated.

If the standard deviation for any sample is excessively large, it shall be examined with a view to rejecting the results from that sample.

Cochran's criterion at the 1 % level can be used when the standard deviations are based on the same number of degrees of freedom. This involves calculating the ratio of the largest of the corresponding sums of squares (laboratories or repeats, as appropriate) to their total (see Clause C.5). If the ratio exceeds the critical value given in Table D.3, with n as the number of samples and ν the degrees of freedom, then all the results from the sample in question shall be rejected. In such an event, care should be taken that the extreme standard deviation is not due to the application of an inappropriate transformation (see 5.2), or undetected outliers.

There is no optimal test when standard deviations are based on different degrees of freedom. However, the ratio of the largest variance to that pooled from the remaining samples follows an F-distribution with ν_1 and ν_2 degrees of freedom (see Clause C.7). Here ν_1 is the degrees of freedom of the variance in question and ν_2 is the degrees of freedom for the remaining samples. If the ratio is greater than the critical value given in Tables D.6 to D.10, corresponding to a significance level of 0,01/S, where S is the number of samples, then results from the sample in question shall be rejected.

5.4.2 Worked example

The standard deviations of the transformed results, after the rejection of the pair of results by laboratory D on sample 1, are given in Table 4 in ascending order of sample mean, correct to three significant digits. Corresponding degrees of freedom are in parentheses.

Inspection shows that there is no outlying sample amongst these. It is noted that the standard deviations are now independent of the sample means, which was the purpose of transforming the results.

The figures in Table 5, taken from a test programme on bromine numbers over 100, illustrate the case of a sample rejection.

It is clear, by inspection, that the laboratories' standard deviation for sample 93 at 15,26 is far greater than the others. It is noted that the repeats standard deviation in this sample is correspondingly large.

Table 4

Sample number	3	8	1	4	5	6	2	7
Sample mean	0,910 0	1,066	1,240	1,538	2,217	3,639	4,028	4,851
Laboratories standard deviation	0,027 8 (14)	0,047 3 (9)	0,035 4 (13)	0,029 7 (11)	0,019 7 (9)	0,037 8 (9)	0,045 0 (9)	0,041 6 (9)
Repeats standard deviation	0,021 4 (9)	0,018 2 (9)	0,028 1 (8)	0,016 4 (9)	0,006 3 (9)	0,013 2 (9)	0,016 6 (9)	0,013 0 (9)

Table 5

Sample number	90	89	93	92	91	94	95	96
Sample mean	96,1	99,8	119,3	125,4	126,0	139,1	139,4	159,5
Laboratories standard deviation	5,10 (8)	4,20 (9)	15,26 (8)	4,40 (11)	4,09 (10)	4,87 (8)	4,74 (9)	3,85 (8)
Repeats standard deviation	1,13 (8)	0,99 (8)	2,97 (8)	0,91 (8)	0,73 (8)	1,32 (8)	1,12 (8)	1,36 (8)

Since laboratory degrees of freedom are not the same over all samples, the variance ratio test is used. The variance pooled from all samples excluding sample 93 is the sum of the sums of squares divided by the total degrees of freedom, that is:

$$\frac{\left[\left(8\times 5{,}10^2\right)+\left(9\times 4{,}20^2\right)+\ldots+\left(8\times 3{,}85^2\right)\right]}{(8+9+\ldots+8)}=19{,}96$$

The variance ratio is then calculated as $(15{,}26^2)/19{,}96 = 11{,}66$.

From Tables D.6 to D.10, the critical value corresponding to a significance level of $0{,}01/8 = 0{,}001\ 25$, for 8 and 63 degrees of freedom, is approximately 4. This is less than the test ratio and results from sample 93 shall. therefore, be rejected.

Turning to repeats standard deviations, it is noted that degrees of freedom are identical for each sample and that Cochran's test can therefore be applied. Cochran's criterion is the ratio of the largest sum of squares (sample 93) to the sum of all the sums of squares, that is:

$$2{,}97^2/(1{,}13^2 + 0{,}99^2 + \ldots + 1{,}36^2) = 0{,}510$$

This is greater than the critical value of 0,352 corresponding to $n = 8$ and $\nu = 8$ (see Table D.3), and confirms that results from sample 93 shall be rejected.

5.5 Estimating missing or rejected values

5.5.1 One of the two repeat values missing or rejected

If one of a pair of repeats (y_{ij1} or y_{ij2}) is missing or rejected, this shall be considered to have the same value as the other repeat in accordance with the least squares method.

5.5.2 Both repeat values missing or rejected

5.5.2.1 General

If both the repeat values are missing, estimates of a_{ij} $(= y_{ij1} + y_{ij2})$ shall be made by forming the laboratories × samples interaction sum of squares, including the missing values of the totals of the laboratories/samples pairs of results as unknown variables. Any laboratory or sample from which all the results were rejected shall be ignored and new values of L and S used. The estimates of the missing or rejected values shall then be found by forming the partial derivatives of this sum of squares with respect to each variable in turn and equating these to zero to solve as a set of simultaneous equations.

Equation (4) may be used where only one pair sum has to be estimated. If more estimates are to be made, the technique of successive approximation can be used. In this, each pair sum is estimated in turn from Equation (4), using L_1, S_1 and T_1 values which contain the latest estimates of the other missing pairs. Initial values for estimates can be based on the appropriate sample mean, and the process usually converges to the required level of accuracy within three complete iterations. See, for instance, Reference [5] in the bibliography for details.

If the value of one pair sum, a_{ij} , has to be estimated, the estimate is given by Equation (4):

$$a_{ij} = \frac{1}{(L-1)(S'-1)} = \left(LL_1 + S'S_1 - T_1\right) \qquad (4)$$

where

S' is S minus the number of samples rejected in 5.4;

L_1 is the total of remaining pairs in the ith laboratory;

S_1 is the total of remaining pairs in the jth sample;

T_1 is the total of all pairs except a_{ij}.

5.5.2.2 Worked example

The two results from laboratory D on sample 1 were rejected (see 5.3.3) and thus a_{ij} has to be estimated.

— total of remaining results in laboratory 4 = 36,354;

— total of remaining results in sample 1 = 19,845;

— total of all the results except a_{ij} = 348,358.

Also $S' = 8$ and $L = 9$.

Hence, the estimate of a_{ij}, is given by

$$a_{ij} = \frac{1}{(9-1)(8-1)}\left[\left(9 \times 36,354\right) + \left(8 \times 19,845\right) - 348,358\right]$$

Therefore $a_{ij} = \frac{137,588}{56} = 2,457$.

5.6 Rejection test for outlying laboratories

5.6.1 General

At this stage, one further rejection test remains to be carried out. This determines whether it is necessary to reject the complete set of results from any particular laboratory. It cannot be carried out at an earlier stage, except in the case where no individual results or pairs are missing or rejected. The procedure, again, consists of Hawkins' test (see 5.3.3), applied to the laboratory averages over all samples, with any estimated results included. If any laboratories are rejected on all samples, new estimates shall be calculated for any remaining missing values (see 5.5).

5.6.2 Worked example

The procedure on the laboratory averages shown in Table 6 below follows exactly that specified in 5.3.3.

The deviations of laboratory averages from the overall mean are given in Table 7 in units of the fourth decimal place, together with the sum of squares.

Hawkins' test ratio is, therefore,

$$B^* = 0,026\ 3 / \sqrt{0,002\ 221\ 9} = 0,558\ 0$$

Comparison with the value tabulated in Table D.4, for $n = 9$ and $\nu = 0$, shows that this ratio is not significant and, therefore, no complete laboratory rejections are necessary.

Table 6

Laboratory	A	B	C	D	E	F	G	H	J
Average	2,437	2,438	2,424	2,426[a]	2,444	2,458	2,410	2,427	2,462

[a] Including estimated value.

Table 7

Laboratory	A	B	C	D	E	F	G	H	J	SS[a]
Deviation	7	23	125	104	75	220	263	87	254	22,19

[a] Sum of squares.

5.7 Confirmation of selected transformation

5.7.1 General

At this stage, it is necessary to check that the rejections carried out have not invalidated the transformation used. If necessary, the procedure given in 5.2 shall be repeated with the outliers deleted, and if a new transformation is selected, outlier tests shall be re-applied.

5.7.2 Worked example

It is not considered necessary in this case to repeat the calculations from 5.2 with the outlying pair deleted.

6 Analysis of variance, calculation and expression of precision estimates

6.1 General

After the data have been inspected for uniformity, a transformation has been performed if necessary, and any outliers have been rejected (see Clause 5), an analysis shall be carried out. First an analysis-of-variance table shall be constructed, and finally the precision estimates derived.

6.2 Analysis of variance

6.2.1 Forming the sums of squares for the laboratories × samples interaction sum of squares

6.2.1.1 General

The estimated values, if any, shall be put in the array and an approximate analysis of variance performed.

Mean correction, $M_c = T^2/2L'S'$ (5)

where L' is L minus the number of laboratories rejected in 5.6 minus the number of laboratories with no remaining results after rejections in 5.3.3.

$$\text{Samples sum of squares} = \left[\sum_{j=1}^{S'}(g_j^2/2L')\right] - M_c \qquad (6)$$

$$\text{Laboratories sum of squares} = \left[\sum_{i=1}^{L'}(h_i^2 / 2S')\right] - M_c \quad (7)$$

$$\text{Pairs sum of squares} = (1/2)\left[\sum_{i=1}^{L'}\sum_{j=1}^{S'} a_{ij}^2\right] - M_c \quad (8)$$

The laboratories × samples interaction sum of squares, I, is given by:

I = (pairs sum of squares) − (laboratories sum of squares) − (sample sum of squares)

Ignoring any pairs in which there are estimated values,

$$E = \text{repeats sum of squares} = (1/2)\sum_{i=1}^{L'}\sum_{j=1}^{S'} e_{ij}^2 \quad (9)$$

The purpose of performing this approximate analysis of variance is to obtain the minimized laboratories × samples interaction sum of squares, I. This is then used as indicated in 6.2.2, to obtain the laboratories sum of squares.

If there were no estimated values, the above analysis of variance is exact and 6.2.2 shall be disregarded.

6.2.1.2 Worked example

$$\text{Mean correction} = \frac{350{,}815^2}{144} = 854{,}660\ 5$$

$$\text{Samples sum of squares} = \frac{22{,}302^2 + 72{,}512^2 + \ldots + 19{,}192^2}{18} - 854{,}660\,5 = 293{,}540\ 9$$

$$\text{Laboratories sum of squares} = \frac{38{,}992^2 + 39{,}020^2 + \ldots + 39{,}387^2}{16} - 854{,}660\,5 = 0{,}035\,6$$

Pairs sum of squares = $(1/2)(2{,}520^2 + 8{,}041^2 + \ldots + 2{,}238^2) - 854{,}660\ 5 = 293{,}690\ 8$

Repeats sum of squares = $(1/2)(0{,}042^2 + 0{,}021^2 + \ldots + 0^2) = 0{,}021\ 9$

Table 8 can then be derived.

Table 8

Source of variation	Sum of squares
Samples	293,540 9
Laboratories	0,035 6
Laboratories × samples	0,114 3
Pairs	293,690 8
Repeats	0,021 9

6.2.2 Forming the sum of squares for the exact analysis of variance

6.2.2.1 General

In 6.2.2, all the estimated pairs are disregarded and new values of g_j are calculated. The following sums of squares for the exact analysis of variance[3] are formed.

$$\text{Uncorrected sample sum of squares} = \sum_{j=1}^{S'} \frac{g_j^2}{S_j} \tag{10}$$

where $S_j = 2$ (L' – number of missing pairs in that sample).

$$\text{Uncorrected pairs sum of squares} = (1/2) \sum_{i=1}^{L'} \sum_{j=1}^{S'} a_{ij}^2 \tag{11}$$

The laboratories sum of squares is equal to (pairs sum of squares) – (samples sum of squares) – (the minimized laboratories × samples interaction sum of squares)

$$= (1/2) \left[\sum_{i=1}^{L'} \sum_{j=1}^{S'} a_{ij}^2 \right] - \left[\sum_{j=1}^{S'} \frac{g_j^2}{S_j} \right] - I \tag{12}$$

6.2.2.2 Worked example

$$\text{Uncorrected samples sum of squares} = \frac{19{,}845^2}{16} + \frac{75{,}512^2}{18} + \ldots + \frac{19{,}192^2}{18} = 1\,145{,}183\ 4$$

$$\text{Uncorrected pairs sum of squares} = \frac{2{,}520^2}{2} + \frac{8{,}041^2}{2} + \ldots + \frac{2{,}238^2}{2} = 1\ 145{,}332\ 9$$

Therefore, laboratories sum of squares = 1 145,332 9 – 1 145,183 4 – 0,114 3 = 0,035 2.

6.2.3 Degrees of freedom

6.2.3.1 General

The degrees of freedom for the laboratories are ($L' - 1$). The degrees of freedom for laboratories × samples interaction are ($L' - 1$) ($S' - 1$) for a complete array and are reduced by one for each pair which is estimated. The degrees of freedom for repeats are ($L'S'$) and are reduced by one for each pair in which one or both values are estimated.

6.2.3.2 Worked example

There are eight samples and nine laboratories in this example. As no complete laboratories or samples were rejected, then $S' = 8$ and $L' = 9$.

Laboratories degrees of freedom = $L' - 1 = 8$

Laboratories × samples interaction degrees of freedom, if there had been no estimates, would have been (9 – 1)(8 – 1) = 56. But one pair was estimated, hence laboratories × samples interaction degrees of freedom = 55. Repeats degrees of freedom would have been 72 if there had been no estimates. In this case, one pair was estimated, hence repeats degrees of freedom = 71.

6.2.4 Mean squares and analysis of variance

6.2.4.1 General

The mean square in each case is the sum of squares divided by the degrees of freedom. This leads to the analysis of variance shown in Table 9.

Table 9

Source of variation	Degrees of freedom	Sum of squares	Mean square
Laboratories	$L' - 1$	Laboratories sum of squares	M_L
Laboratories × samples	$(L' - 1)(S' - 1)$ – number of estimated pairs	I	M_{LS}
Repeats	$L'S'$ – number of pairs in which one or both values are estimated	E	M_r

The ratio M_L/M_{LS} is distributed as F with the corresponding laboratories and interaction degrees of freedom (see Clause C.7). If this ratio exceeds the 5 % critical value given in Table D.6, then bias between the laboratories is implied and the programme organizer shall be informed (see 4.5): further standardization of the test method can be necessary.

6.2.4.2 Worked example

The analysis of variance is shown in Table 10.

Table 10

Source of variation	Degrees of freedom	Sum of squares	Mean square
Laboratories	8	0,035 2	0,004 400
Laboratories × samples	55	0,114 3	0,002 078
Repeats	71	0,021 9	0,000 308

The ratio $M_L/M_{LS} = 0{,}004\ 4/0{,}002\ 078$ has a value 2,117. This is greater than the 5 % critical value obtained from Table D.6, indicating bias between laboratories.

6.3 Expectation of mean squares and calculation of precision estimates

6.3.1 Expectation of mean squares with no estimated values

For a complete array with no estimated values, the expectations of mean squares are:

Laboratories: $\sigma_0^2 + 2\sigma_1^2 + 2S'\sigma_2^2$

Laboratories × samples: $\sigma_0^2 + 2\sigma_1^2$

Repeats: σ_0^2

where

σ_1^2 is the component of variance due to interaction between laboratories and samples;

σ_2^2 is the component of variance due to differences between laboratories.

6.3.2 Expectation of mean squares with estimated values

6.3.2.1 General

The coefficients of σ_0^2 and σ_2^2 in the expectation of mean squares are altered in the cases where there are estimated values. The expectations of mean squares then become:

Laboratories: $\alpha\sigma_0^2 + 2\sigma_1^2 + \beta\sigma_2^2$

Laboratories × samples: $\gamma\sigma_0^2 + 2\sigma_1^2$

Repeats: σ_0^2

where

$$\beta = 2\frac{(K - S')}{(L' - 1)};$$

K is the number of laboratory × sample cells containing at least one result;

α and γ are computed as follows.

— If there are no cells with only a single estimated result, then $\alpha = \gamma = 1$.

— If there are no empty cells (i.e. every laboratory has tested every sample at least once, and $K = L' \times S'$), then α and γ are both 1 plus the proportion of cells with only a single result.

— If there are both empty cells and cells with only one result, then for each laboratory compute the proportion, p_i, of samples tested for which there is only one result, and the sum, P, of these proportions over all laboratories. For each sample, compute the proportion, q_j, of laboratories that have tested the sample for which there is only one result, and the sum, Q, of these proportions over all samples. Compute the total number of cells, W, with only one result and the proportion of these among all non-empty cells, W/K. Then:

$$\alpha = 1 + \frac{P - (W/K)}{L' - 1}$$

and

$$\gamma = 1 + \frac{W - P - Q + (W/K)}{K - L' - S' + 1}$$

NOTE The development in 6.3.2 is based upon the assumption that both samples and laboratories are “random effects”.

6.3.2.2 Worked example

For the example that has eight samples and nine laboratories, one cell is empty (laboratory D for sample 1), so $K = 71$ and

$$\beta = 2\frac{(71 - 8)}{(9 - 1)} = 15{,}75$$

None of the non-empty cells has a single result, so $\alpha = \gamma = 1$.

6.3.3 Calculation of precision estimates

6.3.3.1 Repeatability

The repeatability variance is twice the mean square for repeats. The repeatability estimate is the product of the repeatability standard deviation and the "t-value", t_v, with appropriate degrees of freedom, v (see Table D.5), corresponding to a two-sided probability of 95 %.

This calculated estimate shall be rounded to no fewer than three and no more than four significant digits.

Note that if a transformation $Y = F(x)$ has been used, then

$$r(x) = \left|\frac{\mathrm{d}x}{\mathrm{d}y}\right| r(y) \tag{13}$$

where $r(x)$, $r(y)$ are the corresponding repeatability functions (see Table E.1). A similar relationship applies to the reproducibility functions $R(x)$, $R(y)$.

6.3.3.2 Worked example

Repeatability variance $V_r = 2\sigma_0^2 = 0{,}000\ 616$

Repeatability of $y = t_{71}\sqrt{0{,}000\ 616} = 0{,}049\ 5$

Repeatability of $x = 3x^{2/3} \times\ 0{,}049\ 5 = 0{,}148\, x^{2/3}$

6.3.3.3 Reproducibility

The reproducibility variance, V_R, is expressed as

$$V_R = 2(\sigma_0^2 + \sigma_1^2 + \sigma_2^2)$$

and can be calculated using Equation (14):

$$V_R = \frac{2}{\beta} M_L + \left(1 - \frac{2}{\beta}\right) M_{LS} + \left(2 - \gamma + \frac{2}{\beta}(\gamma - \alpha)\right) M_r \tag{14}$$

where the symbols are as set out in 6.2.4 and 6.3.2.

The reproducibility estimate is the product of the reproducibility standard deviation and the "t-value", t_v, with appropriate degrees of freedom, v, (see Table D.5), corresponding to a two-sided probability of 95 %. An approximation[4] to the degrees of freedom of the reproducibility variance is given by Equation (15).

$$v = \frac{V_R^2}{\dfrac{r_1^2}{L' - 1} + \dfrac{r_2^2}{v_{LS}} + \dfrac{r_3^2}{v_r}} \tag{15}$$

where

r_1, r_2 and r_3 are the three successive terms in Equation (14);

v_{LS} is the degrees of freedom for laboratories × samples;

v_r is the degrees of freedom for repeats.

The calculated estimate of reproducibility shall also be rounded to no fewer than three and no more than four significant digits.

Substantial bias between laboratories results in a loss of degrees of freedom estimated by Equation (15). If reproducibility degrees of freedom are less than 30, then the programme organizer shall be informed (see 4.5); further standardization of the test method can be necessary.

6.3.3.4 Worked example

$$\text{Reproducibility variance} = \left(\frac{2}{15,75} \times 0,004\,40\right) + \left(\frac{13,75}{15,75} \times 0,002\,078\right) + 0,000\,308$$

$$= 0,000\,559 + 0,001\,814 + 0,000\,308$$

$$= 0,002\,681$$

$\nu = 718\ 8/(39 + 60 + 1) = 72$ (correct to nearest integer)

Reproducibility of $y = t_{72}\sqrt{0,002\,681} = 0,103\,4$

Reproducibility of $x = 0,310\,x^{2/3}$

6.4 Expression of precision estimates of a method of test

6.4.1 When the precision of a method of test has been determined in accordance with the procedures set out in this International Standard, it shall be included in the method as follows:

"**X Precision**
X.1 General
The precision, as determined by statistical examination in accordance with ISO 4259 of inter-laboratory test results on (type of products) with test results in the range (x to y), is given in X.2 and X.3.

X.2 Repeatability
The difference between two test results obtained by the same operator with the same apparatus under constant operating conditions on identical test material would, in the long run, in the normal and correct operation of the test method, exceed the following value (value given in Table M) (value shown in Figure N) in only one case in twenty:

$$r = f_r(x)$$

where x is the average of the test results being compared.

X.3 Reproducibility
The difference between two single and independent test results obtained by different operators working in different laboratories on identical test material would, in the long run, in the normal and correct operation of the test method, exceed the following value (value given in Table M) (value shown in Figure N) in only one case in twenty:

$$R = f_R(x)$$

where x is the average of the test results being compared."

6.4.2 Only in exceptional cases shall a precision estimate not based upon ISO 4259 be allowed. In those cases, the alternative introductory text below shall be used:

"The precision evaluation programme for the matrix of samples with (p) contents in the range (q to r) did not conform to the requirements of ISO 4259, and thus only an estimate of precision based upon inter-laboratory test results is given in X.2 and X.3."

6.4.3 The size of the matrix of samples used to generate the precision statement shall not be quoted unless it is for the reason given in 6.4.2 that has been exercised.

7 Significance of repeatability (r) and reproducibility (R)

7.1 General

The value of these quantities is estimated from analysis of variance (two-factor with replication) performed on the results obtained in a statistically designed inter-laboratory programme in which different laboratories each test a range of samples. Repeatability and reproducibility values shall be included in each published test method, and it is noted that the latter is usually greater than the former if the values are derived in accordance with this International Standard.

See in Annex H for an account of the statistical reasoning underlying the equations in this clause.

7.2 Repeatability, r

7.2.1 General

Most laboratories do not carry out more than one test on each sample for routine quality control purposes, except in abnormal circumstances, such as in cases of dispute or if the test operator wishes to confirm that his technique is satisfactory. In these abnormal circumstances, when multiple results are obtained, it is useful to check the consistency of repeated results against the repeatability of the method and the appropriate procedure is outlined in 7.2.2. It is also useful to know what degree of confidence can be placed on the average results, and the method of determining this is given in 7.2.3.

7.2.2 Acceptability of results

When only two results are obtained under repeatability conditions and their difference is less than or equal to r, the test operator may consider his work as being under control and may take the average of the two results as the estimated value of the property being tested.

If the two results differ by more than r, both shall be considered as suspect and at least three more results obtained. Including the first two, the difference between the most divergent result and the average of the remainder shall then be calculated and this difference compared with a new value, r_1, instead of r, given in Equation (16):

$$r_1 = r\sqrt{\frac{k}{2(k-1)}} \tag{16}$$

where k is the total number of results obtained.

If the difference is less than or equal to r_1, all the results shall be accepted. If the difference exceeds r_1, the most divergent result shall be rejected and the procedure specified in this subclause repeated until an acceptable set of results is obtained.

The average of the acceptable results shall be taken as the estimated value of the property. However, if two or more results from a total of not more than 20 have been rejected, the operating procedure and the apparatus shall be checked and a new series of tests made, if possible.

7.2.3 Confidence limits

When a single test operator, who is working within the precision limits of the method, obtains a series of k results under repeatability conditions, giving an average, $\bar{X}$, it can be assumed with 95 % confidence that the true value, μ, of the characteristic lies within the following limits:

$$\left(\bar{X} - \frac{R_1}{\sqrt{2}}\right) \leqslant \mu \leqslant \left(\bar{X} + \frac{R_1}{\sqrt{2}}\right) \tag{17}$$

where

$$R_1 = \sqrt{R^2 - r^2\left(1 - \frac{1}{k}\right)} \tag{18}$$

Similarly, for the single limit situation, when only one limit is fixed (upper or lower), it can be assumed with 95 % confidence that the true value, μ, of the characteristic is limited as follows:

$$\mu \leqslant \bar{X} + 0{,}59R_1 \quad \text{(upper limit)} \tag{19}$$

or

$$\mu \geqslant \bar{X} - 0{,}59R_1 \quad \text{(lower limit)} \tag{20}$$

The factor 0,59 is the ratio $0{,}84/\sqrt{2}$, where 0,84 is derived in Annex H.

However, since for most test methods r is much smaller than R, little improvement in the precision of the average is obtained by carrying out multiple testing under repeatability conditions.

If the reproducibility, R, of a test method has been found to be considerably greater than the repeatability, r, the reasons for the large value of the ratio R/r shall be analysed and the method, if possible, shall be improved.

7.3 Reproducibility, R

7.3.1 Acceptability of results

The procedure specified in this subclause is intended for judging the acceptability, with respect to the reproducibility of the test method, of results obtained by different laboratories in normal, day-to-day operations and transactions. In cases of dispute between a supplier and a recipient, the procedure specified in Clauses 8 to 10 shall be adopted.

When single results are obtained in two laboratories and their difference is less than or equal to R, the two results shall be considered as acceptable and their average, rather than either one separately, shall be considered as the estimated value of the tested property.

If the two results differ by more than R, both shall be considered as suspect. Each laboratory shall then obtain at least three other acceptable results (see 7.2.2).

In this case, the difference between the averages of all acceptable results of each laboratory shall be judged for conformity using a new value, R_2, instead of R, as given by Equation (21):

$$R_2 = \sqrt{R^2 - r^2\left(1 - \frac{1}{2k_1} - \frac{1}{2k_2}\right)} \tag{21}$$

where

R is the reproducibility of the method;

r is the repeatability of the method;

k_1 is the number of results of the first laboratory;

k_2 is the number of results of the second laboratory.

If the difference between the averages is less than or equal to R_2, then these averages are acceptable and their overall average shall be considered as the estimated value of the tested property. If the difference between the averages is greater than R_2, then the procedure specified in Clauses 8 to 10 shall be adopted.

If circumstances arise in which $(N+1)>2$ laboratories each supply one or more acceptable results, the difference between the most divergent laboratory average and the average of the remaining N laboratory averages shall be compared to R_3, where

$$R_3 = \sqrt{\frac{R_1^2}{2} + \frac{R_4^2}{2N}} \tag{22}$$

$$R_4 = \sqrt{R^2 - \frac{r^2}{N}\left(N - \frac{1}{k_1} - \frac{1}{k_2} - \ldots - \frac{1}{k_N}\right)} \tag{23}$$

R_1 is given in Equation (18), and corresponds to the most divergent laboratory average.

If this difference is equal to or less than R_3 in absolute value, all results shall be regarded as acceptable and their average taken as the estimated value of the property.

If the difference is greater than R_3, the most divergent laboratory average shall be rejected and the comparison using Equations (22) and (23) repeated until an acceptable set of laboratory averages is obtained. The average of these laboratory averages shall be taken as the estimated value of the property. However, if two or more laboratory averages from a total of not more than 20 have been rejected, the operating procedure and the apparatus shall be checked and a new series of tests made, if possible.

7.3.2 Confidence limits

When N laboratories obtain one or more results under conditions of repeatability and reproducibility, giving an average of laboratory averages $\bar{X}$, it may be assumed with 95 % confidence that the true value μ of the characteristic lies within the following limits:

$$\left(\bar{X} - \frac{R_4}{\sqrt{2N}}\right) \leqslant \mu \leqslant \left(\bar{X} + \frac{R_4}{\sqrt{2N}}\right) \tag{24}$$

Similarly for the single limit situation, when only one limit is fixed (upper or lower), it may be assumed with 95 % confidence that the true value μ of the characteristic is limited as follows:

$$\mu \leqslant \bar{X} + 0{,}59\frac{R_4}{\sqrt{N}} \quad \text{(upper limit)} \tag{25}$$

or

$$\mu \geqslant \bar{X} - 0{,}59\frac{R_4}{\sqrt{N}} \quad \text{(lower limit)} \tag{26}$$

These equations also allow a given laboratory ($N = 1$) to determine the confidence level that can be assigned to the average of results by comparison with the true value.

8 Specifications

8.1 Aim of specifications

The purpose of a specification is to fix a limit or limits to the true value of the property considered. In practice, however, this true value can never be established exactly. The property is measured in the laboratory by applying a standard test method, the results of which can show some scattering as defined by the repeatability and reproducibility. There is, therefore, some uncertainty as to the true value of the tested property.

Petroleum product specifications are controlled in accordance with Clauses 9 and 10. By prior agreement, a supplier and recipient can use the alternative procedures described in Annex I.

It is important that a test method is selected that is sufficiently precise to determine whether or not the product satisfies the specifications.

8.2 Construction of specifications limits in relation to precision

Usually specifications deal with limits for the values of the properties. To avoid uncertainty, such limits are normally expressed as “not less than” or “not greater than”. Limits are of two types:

— a double limit, upper and lower, for example viscosity not less than 5 mm^2/s and not greater than 16 mm^2/s; boiling point 100 °C ± 0,5 °C;

— a single limit, upper or lower, for example sulfur content not greater than 2 %; lead content not greater than 3,0 g/l; solubility of bitumen not less than 99 %.

The single limit situation becomes relevant when, as in most cases, there is an additional implied limit which effectively converts it into a double limit situation. This is illustrated by the examples above in which the additional implied limits are 0 %, 0 g/l, and 100 %, respectively. In cases of a true single limit situation, for example flash point not less than 60 °C, the following considerations do not apply. In Clauses 8 to 10, A_1 denotes the upper limit and A_2 denotes the lower limit.

The value chosen for a specification limit shall take into account the reproducibility of the test method adopted, as follows:

— for a double limit (A_1 and A_2), the specified range (stated or implied) shall be not less than four times the reproducibility R, i.e. $(A_1 - A_2) \geqslant 4R$;

— for a single limit (A_1 or A_2), the specified limit shall be a distance not less than twice the reproducibility, R, away from the implied limit, that is, if the upper implied limit is 100 %, then $(100 - A_1) \geqslant 2R$, or if the lower implied limit is zero, then $A_2 \geqslant 2R$.

The requirements of this International Standard apply to specifications drawn up in accordance with these principles.

In cases where, for practical reasons, the value of $(A_1 - A_2)$ is less than $4R$, the results obtained will be of doubtful significance in determining whether a sample does or does not satisfy the requirements of the specification. According to statistical reasoning, it is desirable for $(A_1 - A_2)$ to be considerably greater than $4R$. If not, one or both of the following courses shall be adopted:

a) the specification limits shall be examined to see whether they can be widened to fit in with the precision of the test method;

b) the test method shall be examined to see whether the precision can be improved, or an alternative test method adopted with an improved precision, to fit in with the desired specification limits.

As a consequence of the above restrictions on specification limits, it is recommended that the lower limit of the scope of petroleum test methods be a value not less than $2R$ greater than the lowest achievable result, and the upper limit of the scope be a value not greater than $2R$ less than the highest achievable result.

9 Quality control against specifications

9.1 General

Clause 9 provides general information to allow the supplier or the recipient to judge the quality of a product with regard to the specification when a single result is available. If it is necessary for the recipient to take action after examining this result, the procedure specified in Clause 10 shall be adopted.

9.2 Testing margin at the supplier

A supplier who has no other source of information on the true value of a characteristic than a single result shall consider that the product meets the specification limit, with 95 % confidence, only if the result, X, is such that

— in the case of a single upper limit, A_1:

$$X \leqslant A_1 - 0{,}59R \tag{27}$$

— in the case of a single lower limit, A_2:

$$X \geqslant A_2 + 0{,}59R \tag{28}$$

— in the case of a double limit (A_1 and A_2), both these conditions are satisfied (see 7.2.3).

The use of Equations (27) and (28) is for the guidance of the supplier and is not to be interpreted as an obligation. A reported value between the specification value and the limit from Equation (27) or (28) is not proof of non-compliance.

9.3 Testing margin at the recipient

A recipient who has no other source of information on the true value of a characteristic than a single result shall consider that the product fails the specification limit, with 95 % confidence, only if the result, X, is such that:

— in the case of a single upper limit A_1,

$$X > A_1 + 0{,}59R \tag{29}$$

— in the case of a single lower limit A_2,

$$X < A_2 - 0{,}59R \tag{30}$$

— in the case of a double limit (A_1 and A_2), either of these conditions applies.

10 Dispute procedure

10.1 If it is not possible for the supplier and the recipient to reach agreement about the quality of the product on the basis of their existing results, then the procedures given in 10.2 to 10.5 shall be adopted.

10.2 Each laboratory shall reject its original results and obtain at least three other acceptable results on their own check sample to ensure that the work has been carried out under repeatability conditions. The average of

the acceptable results in each laboratory shall then be computed, divergent results being discarded as indicated in 7.2.2. If the re-testing does not resolve the dispute, then continue as given below.

Let

$\bar{X}_S$ be the average of the supplier;

$\bar{X}_R$ be the average of recipient;

A_1 be the upper limit of the specification;

A_2 be the lower limit of the specification.

where

$$\bar{X}_S \leqslant A_1 < \bar{X}_R$$

$$\bar{X}_S \geqslant A_2 > \bar{X}_R$$

This means that $\bar{X}_S$ and $\bar{X}_R$ should be compared as follows with A_1 and A_2.

If $\dfrac{\bar{X}_S + \bar{X}_R}{2} \leqslant A_1$ or $\geqslant A_2$

— product meets specification if $\left|\bar{X}_S - \bar{X}_R\right| \leqslant 0{,}84R_2$ (for R_2, see 7.3.1);

— possible dispute if $\left|\bar{X}_S - \bar{X}_R\right| > 0{,}84R_2$.

In the latter case, it cannot be stated with confidence whether the product does or does not comply with the specification limit; hence resolution of the dispute may be by negotiation.

If $\dfrac{\bar{X}_S + \bar{X}_R}{2} > A_1$ or $< A_2$

dispute whatever the difference $\bar{X}_S - \bar{X}_R$.

10.3 In case of dispute, the two laboratories shall contact each other and compare their operating procedures and apparatus. Following these investigations, a correlation test between the two laboratories shall be carried out on their check samples. The average of at least three acceptable results shall be computed, in each laboratory, and these averages compared as indicated in 10.2.

10.4 If the disagreement remains, a third laboratory (neutral, expert and accepted by the two parties) shall be invited to carry out the test using a third sample. Suppose $\bar{X}_E$ is the average of the three or more acceptable results of the third laboratory. If the difference between the most divergent laboratory average and the average of the two other laboratory averages is less than or equal to R_3 (see 7.3.1) the following procedure shall be adopted:

If $\dfrac{\bar{X}_S + \bar{X}_R + \bar{X}_E}{3} \leqslant A_1$ or $\geqslant A_2$, product meets specification.

If $\dfrac{\bar{X}_S + \bar{X}_R + \bar{X}_E}{3} > A_1$ or $< A_2$, product fails specification.

10.5 If the difference between the most divergent laboratory average and the average, $\bar{X}$, of the two other laboratory averages is more than R_3, the following procedure shall be adopted:

If $\bar{X} \leqslant A_1$ or $\geqslant A_2$, product meets specification.

If $\bar{X} > A_1$ or $< A_2$, product fails specification.

Annex A
(normative)

Determination of number of samples required

NOTE See 4.4.

Table A.1 — Determination of number of samples required

$L = 5$										
	Q:	0	1	2	3	4	5	6	7	8	9
P:	0	4									
	1	5									
	2	6	11								
	3	6	9								
	4	7	8	16							
	5	7	8	12							
	6	7	8	11	19						
	7	7	8	10	15						
	8	7	8	9	13						
	9	7	8	9	11	17					

$L = 6$										
	Q:	0	1	2	3	4	5	6	7	8	9
P:	0	3									
	1	4	11								
	2	5	7								
	3	5	7	14							
	4	5	6	10							
	5	6	6	8	15						
	6	6	6	8	11						
	7	6	6	7	10	15					
	8	6	6	7	9	12					
	9	6	6	7	8	10	15				

$L = 7$										
	Q:	0	1	2	3	4	5	6	7	8	9
P:	0	3									
	1	4	7								
	2	4	6	17							
	3	4	5	9							
	4	5	5	7	13						
	5	5	5	6	9	19					
	6	5	5	6	8	12					
	7	5	5	6	7	10	15				
	8	5	5	6	7	8	12	20			
	9	5	5	6	6	8	10	14			

$L = 8$										
	Q:	0	1	2	3	4	5	6	7	8	9
P:	0	3									
	1	3	5								
	2	4	5	9							
	3	4	5	7	14						
	4	4	4	6	9	20					
	5	4	4	5	7	11					
	6	4	4	5	6	8	13				
	7	4	4	5	6	7	10	16			
	8	4	5	5	6	6	8	11	18		
	9	4	5	5	5	6	7	9	13		

$L = 9$										
	Q:	0	1	2	3	4	5	6	7	8	9
P:	0	2									
	1	3	4								
	2	3	4	7							
	3	3	4	5	9						
	4	4	4	5	6	11					
	5	4	4	5	6	7	12				
	6	4	4	4	5	6	9	14			
	7	4	4	4	5	6	7	10	15		
	8	4	4	4	5	5	6	8	10	16	
	9	4	4	4	5	5	6	7	8	11	18

$L = 10$										
	Q:	0	1	2	3	4	5	6	7	8	9
P:	0	2	8								
	1	3	4	11							
	2	3	4	5	12						
	3	3	3	4	6	13					
	4	3	4	4	5	7	14				
	5	3	4	4	5	6	8	14			
	6	3	4	4	4	5	6	9	14		
	7	3	4	4	4	5	6	7	9	14	
	8	3	4	4	4	5	5	6	7	10	14
	9	4	4	4	4	4	5	6	6	8	10

$L = 11$										
	Q:	0	1	2	3	4	5	6	7	8	9
P:	0	2	4								
	1	2	3	5							
	2	3	3	3	7						
	3	3	3	4	5	8					
	4	3	3	4	4	6	8	18			
	5	3	3	4	4	5	6	9	15		
	6	3	3	3	4	4	5	6	9	14	
	7	3	3	3	4	4	5	5	7	9	13
	8	3	3	3	4	4	4	5	6	7	9
	9	3	3	3	4	4	4	5	5	6	7

$L = 12$										
	Q:	0	1	2	3	4	5	6	7	8	9
P:	0	2	4								
	1	2	3	5							
	2	2	3	4	6	14					
	3	3	3	3	4	6	11				
	4	3	3	3	4	5	6	9			
	5	3	3	3	4	4	5	6	9	16	
	6	3	3	3	3	4	4	5	6	9	13
	7	3	3	3	3	4	4	5	5	6	8
	8	3	3	3	3	4	4	4	5	5	6
	9	3	3	3	3	3	4	4	4	5	6

$L = 13$										
	Q:	0	1	2	3	4	5	6	7	8	9
P:	0	2	3								
	1	2	3	4	12						
	2	2	3	3	4	8					
	3	2	3	3	4	5	7	14			
	4	3	3	3	3	4	5	7	10		
	5	3	3	3	3	4	4	5	6	9	15
	6	3	3	3	3	3	4	4	5	6	8
	7	3	3	3	3	3	4	4	4	5	6
	8	3	3	3	3	3	3	4	4	5	5
	9	3	3	3	3	3	3	4	4	4	5

Table A.1 (*continued*)

$L = 14$											
	Q:	0	1	2	3	4	5	6	7	8	9
P:	0	2	3								
	1	2	2	3	7						
	2	2	2	3	4	6	12				
	3	2	2	3	3	4	5	8	18		
	4	2	3	3	3	3	4	5	7	11	
	5	2	3	3	3	3	4	4	5	6	8
	6	3	3	3	3	3	3	4	4	5	6
	7	3	3	3	3	3	3	3	4	4	5
	8	3	3	3	3	3	3	3	4	4	4
	9	3	3	3	3	3	3	3	3	4	4

$L = 15$											
	Q:	0	1	2	3	4	5	6	7	8	9
P:	0	2	2	13							
	1	2	2	3	5	19					
	2	2	2	3	3	4	7				
	3	2	2	3	3	3	4	6	9		
	4	2	2	3	3	3	4	4	5	7	10
	5	2	2	3	3	3	3	4	4	5	6
	6	2	2	3	3	3	3	3	4	4	5
	7	2	2	3	3	3	3	3	3	4	4
	8	2	2	3	3	3	3	3	3	3	4
	9	2	2	3	3	3	3	3	3	3	3

$L = 16$											
	Q:	0	1	2	3	4	5	6	7	8	9
P:	0	1	2	5							
	1	2	2	3	4	8					
	2	2	2	2	3	4	5	9			
	3	2	2	2	3	3	4	4	6	9	
	4	2	2	2	3	3	3	4	4	5	6
	5	2	2	2	3	3	3	3	4	4	5
	6	2	2	2	3	3	3	3	3	4	4
	7	2	2	2	3	3	3	3	3	3	4
	8	2	2	2	3	3	3	3	3	3	3
	9	2	2	2	3	3	3	3	3	3	3

L = number of participating laboratories

$$P = \frac{\text{interaction variance component}}{\text{repeats variance component}}$$

$$Q = \frac{\text{laboratories variance component}}{\text{repeats variance component}}$$

Annex B
(informative)

Derivation of equation for calculating the number of samples required

NOTE See 4.4.

An analysis of variance is carried out on the results of the pilot programme. This yields rough estimates of the three components of variance, namely:

— σ_0^2 for repeats;

— σ_1^2 for laboratories × samples interaction;

— σ_2^2 for laboratories.

Substituting the above into Equation (15) (see 6.3.3.3) for calculating the reproducibility degrees of freedom, this becomes:

$$\frac{(1+P+Q)^2}{v} = \frac{\left[\left(\frac{1}{2}+P\right)/S+Q\right]^2}{(L-1)} + \frac{(S-1)\left(\frac{1}{2}+P\right)^2}{S^2(L-1)} + \frac{1}{4LS} \qquad \text{(B.1)}$$

where

P is the ratio σ_1^2/σ_0^2;

Q is the ratio σ_2^2/σ_0^2;

v is the reproducibility degrees of freedom;

L is the number of laboratories;

S is the number of samples.

The equation rearranges into the form:

$$aS + b = 0$$

where

$$a = vQ^2 - (1+P+Q)^2(L-1)$$

$$b = v\left[\left(2Q+\frac{1}{2}+P\right)\left(\frac{1}{2}+P\right)+0{,}25(L-1)/L\right]$$

Therefore

$$S = -\frac{b}{a} \qquad \text{(B.2)}$$

gives the values of S for given values of L, P, Q and v.

Table A.1 is based on $v = 30$ degrees of freedom. For non-integral values of P and Q, S can be estimated by second order interpolation from the table.

Annex C
(normative)

Notation and tests

C.1 Introduction

Throughout this International Standard the following notation is used:

— S is the number of samples;

— L is the number of laboratories;

— i is the subscript denoting laboratory number;

— j is the subscript denoting sample number;

— x is an individual test result;

— a is the sum of duplicate test results;

— e is the difference between duplicate test results;

— ν is the degrees of freedom.

C.2 Array of duplicate results from each of L laboratories on S samples and corresponding means m_j

Table C.1

Laboratory	**Sample**			
	1	2	j	S
1	x_{111} x_{112}	x_{121} x_{122}	x_{1j1} x_{1j2}	x_{1S1} x_{1S2}
2	x_{211} x_{212}	x_{221} x_{222}	x_{2j1} x_{2j2}	x_{2S1} x_{2S2}
i	x_{i11} x_{i12}	x_{i21} x_{i22}	x_{ij1} x_{ij2}	x_{iS1} x_{iS2}
L	x_{L11} x_{L12}	x_{L21} x_{L22}	x_{Lj1} x_{Lj2}	x_{LS1} x_{LS2}
Total	g_1	g_2	g_j	g_S
Mean	m_1	m_2	m_j	m_S
NOTE If a transformation $y = F(x)$ of the reported data is necessary (see 5.2), then corresponding symbols y_{ij1} and y_{ij2} are used in place of x_{ij1} and x_{ij2}.				

C.3 Array of sums of duplicate results, of laboratory totals h_i and sample totals g_j

Table C.2

Laboratory	Sample				
	1	2	j	S	**Total**
1	a_{11}	a_{12}	a_{1j}	a_{1S}	h_1
2	a_{21}	a_{22}	a_{2j}	a_{2S}	h_2
i	a_{i1}	a_{i2}	a_{ij}	a_{iS}	h_i
L	a_{L1}	a_{L2}	a_{Lj}	a_{LS}	h_L
Total	g_1	g_2	g_j	g_S	T

$a_{ij} = x_{ij1} + x_{ij2}$ (or $a_{ij} = {}_{yj}1 + y_{ij2}$, if a transformation has been used)

$e_{ij} = x_{ij1} - x_{ij2}$ (or $e_{ij} = y_{ij1} - y_{ij2}$, if a transformation has been used)

$$g_j = \sum_{i=1}^{L} a_{ij}$$

$$h_i = \sum_{j=1}^{S} a_{ij}$$

$$m_j = g_j \big/ 2L$$

$$T = \sum_{i=1}^{L} h_i = \sum_{j=1}^{S} g_j$$

If any results are missing from the complete array, then the divisor in the expression for m_j is correspondingly reduced.

C.4 Sums of squares and variances

NOTE See 5.2.

Repeats variance for sample j:

$$d_j^2 = \sum_{i=1}^{L} e_{ij}^2 \Big/ 2L \qquad \text{(C.1)}$$

where L is the repeats degrees of freedom for sample j. If either or both of a laboratory/sample pair of results is missing, the corresponding term in the numerator is omitted and the factor L is reduced by 1.

Between cells variance for sample j:

$$C_j^2 = \left(\sum_{i=1}^{L} \frac{a_{ij}^2}{n_{ij}} - \frac{g_j^2}{S_j} \right) \Big/ (L-1) \qquad \text{(C.2)}$$

Laboratories variance for sample j:

$$D_j^2 = \frac{1}{K_j}\left[C_j^2 + (K_j - 1)d_j^2\right] \tag{C.3}$$

where

$$K_j = \left(S_j^2 - \sum_{i=1}^{L} n_{ij}^2\right) \Big/ \left[S_j (L-1)\right]$$

n_{ij} is the number of results obtained by laboratory i from sample j;

S_j is the total number of results obtained from sample j;

L is the number of cells in sample j containing at least one result.

Laboratories degrees of freedom for sample j is given approximately[4] by

$$\nu_j = \frac{\left(K_j D_j^2\right)^2}{\dfrac{\left(C_j^2\right)^2}{(L-1)} + \dfrac{\left[\left(K_j - 1\right) d_j^2\right]^2}{L}} \quad \text{(rounded to the nearest integer)} \tag{C.4}$$

If either or both of a laboratory/sample pair of results is missing, the factor L is reduced by 1.

If both of a laboratory/sample pair of results is missing, the factor ($L - 1$) is reduced by 1.

C.5 Cochran's test

The largest sum of squares, SS_k, out of a set of n mutually independent sums of squares each based on ν degrees of freedom, can be tested for conformity in accordance with

$$\text{Cochran's criterion} = SS_k \Big/ \sum_{i=1}^{n} SS_i \tag{C.5}$$

The test ratio is identical if sum of squares values are replaced by mean squares (variance estimates). If the calculated ratio exceeds the critical value given in Table D.3, then the sum of squares in question, SS_k, is significantly greater than the others with a probability of 99 %. Examples of SS_i include e_{ij}^2 and d_j^2 (see Equation C.1).

C.6 Hawkins' test

An extreme value in a data set can be tested as an outlier by comparing its deviation from the mean value of the data set to the square root of the sum of squares of all such deviations. This is carried out in the form of a ratio. Extra information on variability can be provided by including independent sums of squares into the calculations. These are based on ν degrees of freedom and have the same population variance as the data set in question.

Table C.3 shows the values which are required to apply Hawkins' test to individual samples.

Table C.3

Characteristic	Sample			
	1	2	j	S
Number of cells	n_1	n_2	n_j	n_S
Mean of cell means	m'_1	m'_2	m'_j	m'_S
Sum of squares	SS_1	SS_2	SS_j	SS_S

where

n_j is the number of cells in sample j which contains at least one result;

m'_j is the mean of cell means in sample j;

SS_j is the sum of squares of deviations of cell means, a_{ij}/n_{ij}, from the mean of cell means, m'_j, and is given by:

$$SS_j = \sum_{i=1}^{L}\left(\frac{a_{ij}}{n_{ij}} - m'_j\right)^2 \quad \text{where} \quad m'_j = \frac{1}{n_j}\sum_{i=1}^{L}\left(\frac{a_{ij}}{n_{ij}}\right)$$

The test procedure is as follows.

a) Identify the sample, k, and cell mean, a_{ik}/n_{ik}, which has the most extreme absolute deviation $|m'_k - a_{ik}/n_{ik}|$. The cell identified is the candidate for the outlier test, be it high or low.

b) Calculate the total sum of squares of deviations:

$$SS = \sum_{j=1}^{S} SS_j \tag{C.6}$$

c) Calculate the test ratio:

$$B^* = \frac{\left|m'_k - a_{ik}/n_{ik}\right|}{\sqrt{SS}} \tag{C.7}$$

d) Compare the test ratio with the critical value from Table D.4, for $n = n_k$ and extra degrees of freedom ν where

$$\nu = \sum_{j=1}^{S}\left(n_j - 1\right), j \neq k \tag{C.8}$$

e) If B^* exceeds the critical value, reject results from the cell in question (sample k, laboratory i), modify the n_k, m'_k and SS_k values accordingly, and repeat from list item a).

NOTE Hawkins' test applies theoretically to the detection of only a single outlier laboratory in a sample. The technique of repeated tests for a single outlier, in the order of maximum deviation from sample mean, implies that the critical values in Table D.4 do not refer exactly to the 1 % significance level. It has been shown by Hawkins, however, that if $n > 5$ and the total degrees of freedom $(n + \nu)$ are greater than 20, then this effect is negligible, as are the effects of masking (one outlier hiding another) and swamping (the rejection of one outlier leading to the rejection of others).

When the test is applied to laboratories averaged over all samples, Table C.3 reduces to a single column where

n is the number of laboratories $= L$;

m is the overall mean equal to T/N, where N is the total number of results in the array;

SS is the sum of squares of deviations of laboratory means from the overall mean, and is given by

$$SS = \sum_{i=1}^{L} \left(\frac{h_i}{n_i} - m \right)^2 \text{, where } n_i \text{ is the number of results in laboratory } i. \qquad \text{(C.9)}$$

In the test procedure, therefore, identify the laboratory mean, h_i/n_i, that differs most from the overall mean, m. The corresponding test ratio then becomes:

$$B^* = \frac{\left| m - h_i / n_i \right|}{\sqrt{SS}} \qquad \text{(C.10)}$$

This shall be compared with the critical value from Table D.4 as before, but now with extra degrees of freedom $\nu = 0$. If a laboratory is rejected, adjust the values of n, m and SS accordingly and repeat the calculations.

C.7 Variance ratio test (F-test)

A variance estimate, V_1, based on ν_1 degrees of freedom, can be compared with a second estimate, V_2, based on ν_2 degrees of freedom, by calculating the ratio:

$$F = \frac{V_1}{V_2} \qquad \text{(C.11)}$$

If the ratio exceeds the appropriate critical value given in Tables D.6 to D.9, where ν_1 corresponds to the numerator (the larger variance estimate) and ν_2 corresponds to the denominator, then V_1 is greater than V_2 at the chosen level of significance.

Annex D
(normative)

Example results of test for determination of bromine number and statistical tables

Table D.1 — Bromine number for low boiling samples

Laboratory	Sample							
	1	2	3	4	5	6	7	8
A	1,9	64,5	0,80	3,7	11,0	46,1	114,8	1,2
	2,1	65,5	0,78	3,8	11,1	46,5	114,2	1,2
B	1,7	65,4	0,69	3,7	11,1	50,3	114,5	1,2
	1,8	66,0	0,72	3,7	11,0	49,9	114,3	1,2
C	1,8	63,5	0,76	3,5	10,4	48,5	112,4	1,3
	1,8	63,8	0,76	3,5	10,5	48,2	112,7	1,3
D	4,1	63,6	0,80	4,0	10,8	49,6	108,8	1,0
	4,0	63,9	0,80	3,9	10,8	49,9	108,2	1,1
E	2,1	63,9	0,83	3,7	10,9	47,4	115,6	1,3
	1,8	63,7	0,83	3,7	11,1	47,6	115,1	1,4
F	1,8	70,7	0,72	3,4	11,5	49,1	121,0	1,4
	1,7	69,7	0,64	3,6	11,2	47,9	117,9	1,4
G	1,9	63,8	0,77	3,5	10,6	46,1	114,1	1,1
	2,2	63,6	0,59	3,5	10,6	45,5	112,8	0,93
H	2,0	66,5	0,78	3,2	10,7	49,6	114,8	1,1
	1,8	65,5	0,71	3,5	10,7	48,5	114,5	1,0
J	2,1	68,2	0,81	4,0	11,1	49,1	115,7	1,4
	2,1	65,3	0,81	3,7	11,1	47,9	113,9	1,4

Table D.2 — Cube root of the bromine number for low boiling samples

Laboratory	Sample							
	1	2	3	4	5	6	7	8
A	1,239	4,010	0,928	1,547	2,224	3,586	4,860	1,063
	1,281	4,031	0,921	1,560	2,231	3,596	4,852	1,063
B	1,193	4,029	0,884	1,547	2,231	3,691	4,856	1,063
	1,216	4,041	0,896	1,547	2,224	3,682	4,853	1,063
C	1,216	3,990	0,913	1,518	2,183	3,647	4,826	1,091
	1,216	3,996	0,913	1,518	2,190	3,639	4,830	1,091
D	1,601	3,992	0,928	1,587	2,210	3,674	4,774	1,000
	1,587	3,998	0,928	1,574	2,210	3,682	4,765	1,032
E	1,281	3,998	0,940	1,547	2,217	3,619	4,871	1,091
	1,216	3,994	0,940	1,547	2,231	3,624	4,864	1,119
F	1,216	4,135	0,896	1,504	2,257	3,662	4,946	1,119
	1,193	4,115	0,862	1,533	2,237	3,632	4,903	1,119
G	1,239	3,996	0,917	1,518	2,197	3,586	4,850	1,032
	1,301	3,992	0,839	1,518	2,197	3,570	4,832	0,976
H	1,260	4,051	0,921	1,474	2,204	3,674	4,860	1,032
	1,216	4,031	0,892	1,518	2,204	3,647	4,856	1,000
J	1,281	4,086	0,932	1,587	2,231	3,662	4,873	1,119
	1,281	4,027	0,932	1,547	2,231	3,632	4,847	1,119

Table D.3 — Critical 1 % values of Cochran's criterion for n variance estimates and ν degrees of freedom

n	Degrees of freedom ν									
	1	2	3	4	5	10	15	20	30	50
3	0,993 3	0,942 3	0,883 1	0,833 5	0,793 3	0,674 3	0,614 5	0,577 5	0,532 7	0,487 2
4	0,967 6	0,864 3	0,781 4	0,721 2	0,676 1	0,553 6	0,496 4	0,462 0	0,421 3	0,380 8
5	0,927 9	0,788 5	0,695 7	0,632 9	0,587 5	0,469 7	0,416 8	0,385 5	0,348 9	0,313 1
6	0,882 8	0,721 8	0,625 8	0,563 5	0,519 5	0,408 4	0,359 7	0,331 2	0,298 2	0,266 1
7	0,837 6	0,664 4	0,568 5	0,508 0	0,465 9	0,361 6	0,316 7	0,290 7	0,260 6	0,231 6
8	0,794 5	0,615 2	0,520 9	0,462 7	0,422 7	0,324 8	0,283 2	0,259 2	0,231 6	0,205 2
9	0,754 4	0,572 7	0,481 0	0,425 1	0,387 0	0,295 0	0,256 3	0,234 0	0,208 6	0,184 2
10	0,717 5	0,535 8	0,446 9	0,393 4	0,357 2	0,270 4	0,234 2	0,213 5	0,189 8	0,167 3
11	0,683 7	0,503 6	0,417 5	0,366 3	0,331 8	0,249 7	0,215 7	0,196 3	0,174 2	0,153 2
12	0,652 8	0,475 1	0,391 9	0,342 8	0,309 9	0,232 1	0,200 0	0,181 8	0,161 1	0,141 4
13	0,624 5	0,449 8	0,369 5	0,322 3	0,290 9	0,216 9	0,186 5	0,169 3	0,149 8	0,131 3
14	0,598 5	0,427 2	0,349 5	0,304 3	0,274 1	0,203 6	0,174 8	0,158 5	0,140 0	0,122 6
15	0,574 7	0,406 9	0,331 8	0,288 2	0,259 3	0,191 9	0,164 5	0,149 0	0,131 5	0,115 0
20	0,479 9	0,329 7	0,265 4	0,228 8	0,204 8	0,149 6	0,127 4	0,115 0	0,101 0	0,087 9
25	0,413 0	0,278 2	0,222 0	0,190 4	0,169 9	0,123 0	0,104 3	0,093 9	0,082 2	0,071 3
30	0,363 2	0,241 2	0,191 4	0,163 5	0,145 5	0,104 6	0,088 5	0,079 4	0,069 4	0,060 0
35	0,324 7	0,213 4	0,168 5	0,143 5	0,127 4	0,091 2	0,076 9	0,069 0	0,060 1	0,051 9
40	0,294 0	0,191 6	0,150 7	0,128 1	0,113 6	0,080 9	0,068 1	0,061 0	0,053 1	0,045 7
45	0,269 0	0,174 0	0,136 4	0,115 8	0,102 5	0,072 7	0,061 1	0,054 7	0,047 5	0,040 9
50	0,248 1	0,159 6	0,124 8	0,105 7	0,093 5	0,066 1	0,055 5	0,049 6	0,043 1	0,037 0
60	0,215 1	0,137 1	0,106 8	0,090 2	0,079 6	0,056 1	0,046 9	0,041 9	0,036 3	0,031 1
70	0,190 3	0,120 4	0,093 5	0,078 8	0,069 5	0,048 7	0,040 7	0,036 3	0,031 4	0,026 9
80	0,170 9	0,107 5	0,083 2	0,070 1	0,061 7	0,043 1	0,036 0	0,032 0	0,027 7	0,023 6
90	0,155 3	0,097 2	0,075 1	0,063 1	0,055 5	0,038 7	0,032 2	0,028 7	0,024 8	0,021 1
100	0,142 4	0,088 8	0,068 5	0,057 5	0,050 5	0,035 1	0,029 2	0,026 0	0,022 4	0,019 1

NOTE These values are slightly conservative approximations calculated via Bonferroni's inequality[2] as the upper 0,01/n fractile of the beta distribution. If intermediate values are required along the n-axis, they may be obtained by linear interpolation of the reciprocals of the tabulated values. If intermediate values are required along the ν-axis, they may be obtained by second order interpolation of the reciprocals of the tabulated values.

Table D.4 — Critical values of Hawkins' 1 % outlier test for $n = 3$ to 50 and $\nu = 0$ to 200

n	Degrees of freedom ν											
	0	5	10	15	20	30	40	50	70	100	150	200
3	0,816 5	0,724 0	0,610 0	0,532 8	0,478 1	0,404 9	0,357 4	0,323 3	0,276 9	0,234 0	0,192 6	0,167 4
4	0,863 9	0,750 5	0,640 5	0,564 4	0,509 4	0,434 5	0,385 0	0,349 2	0,300 0	0,254 1	0,209 6	0,182 4
5	0,881 8	0,757 3	0,653 0	0,579 6	0,525 8	0,451 0	0,401 2	0,364 7	0,314 2	0,266 8	0,220 4	0,192 0
6	0,882 3	0,755 4	0,657 1	0,586 9	0,534 7	0,461 2	0,411 5	0,374 9	0,323 8	0,275 5	0,228 0	0,198 8
7	0,873 3	0,749 3	0,656 7	0,589 8	0,539 4	0,467 6	0,418 4	0,381 9	0,330 7	0,281 9	0,233 7	0,203 9
8	0,859 6	0,740 9	0,653 8	0,590 1	0,541 5	0,471 5	0,423 1	0,386 9	0,335 8	0,286 8	0,238 1	0,207 9
9	0,843 9	0,731 4	0,649 3	0,588 6	0,541 8	0,473 8	0,426 2	0,390 5	0,339 6	0,290 6	0,241 6	0,211 2
10	0,827 4	0,721 3	0,643 9	0,586 1	0,541 1	0,475 0	0,428 3	0,393 0	0,342 6	0,293 6	0,244 5	0,213 9
11	0,810 8	0,711 1	0,638 0	0,582 8	0,539 4	0,475 3	0,429 5	0,394 8	0,344 8	0,296 1	0,246 9	0,216 2
12	0,794 7	0,701 0	0,631 8	0,579 0	0,537 3	0,475 0	0,430 2	0,396 0	0,346 6	0,298 1	0,248 9	0,218 1
13	0,779 1	0,691 0	0,625 4	0,574 9	0,534 7	0,474 2	0,430 4	0,396 8	0,347 9	0,299 7	0,250 7	0,219 8
14	0,764 2	0,681 2	0,618 9	0,570 6	0,531 9	0,473 1	0,430 2	0,397 2	0,348 9	0,301 1	0,252 1	0,221 2
15	0,750 0	0,671 7	0,612 5	0,566 2	0,528 8	0,471 7	0,429 8	0,397 3	0,349 6	0,302 1	0,253 4	0,222 5
16	0,736 4	0,662 5	0,606 1	0,561 7	0,525 6	0,470 1	0,429 1	0,397 2	0,350 1	0,303 0	0,254 4	0,223 6
17	0,723 5	0,653 5	0,599 8	0,557 1	0,522 3	0,468 3	0,428 2	0,396 8	0,350 4	0,303 7	0,255 4	0,224 6
18	0,711 2	0,644 9	0,593 6	0,552 6	0,518 9	0,466 5	0,427 2	0,396 4	0,350 5	0,304 3	0,256 2	0,225 4
19	0,699 6	0,636 5	0,587 6	0,548 0	0,515 5	0,464 5	0,426 0	0,395 8	0,350 6	0,304 7	0,256 9	0,226 2
20	0,688 4	0,628 6	0,581 6	0,543 6	0,512 0	0,462 4	0,424 8	0,395 1	0,350 5	0,305 1	0,257 5	0,226 9
21	0,677 8	0,620 9	0,575 8	0,539 2	0,508 6	0,460 3	0,423 5	0,394 2	0,350 3	0,305 3	0,258 0	0,227 5
22	0,667 7	0,613 4	0,570 2	0,534 8	0,505 2	0,458 1	0,422 1	0,393 4	0,350 0	0,305 5	0,258 4	0,228 0
23	0,658 1	0,606 2	0,564 7	0,530 5	0,501 8	0,455 9	0,420 6	0,392 4	0,349 6	0,305 6	0,258 8	0,228 5
24	0,648 8	0,599 3	0,559 3	0,526 3	0,498 4	0,453 7	0,419 1	0,391 4	0,349 2	0,305 6	0,259 1	0,228 9
25	0,640 0	0,592 5	0,554 0	0,522 1	0,495 1	0,451 5	0,417 6	0,390 4	0,348 8	0,305 6	0,259 4	0,229 3
26	0,631 5	0,586 1	0,549 0	0,518 0	0,491 8	0,449 2	0,416 0	0,389 3	0,348 2	0,305 4	0,259 6	0,229 6
27	0,623 4	0,579 8	0,544 0	0,514 0	0,488 5	0,447 0	0,414 5	0,388 1	0,347 7	0,305 3	0,259 7	0,229 9
28	0,615 6	0,573 7	0,539 2	0,510 1	0,485 3	0,444 7	0,412 9	0,387 0	0,347 1	0,305 1	0,259 9	0,230 2
29	0,608 1	0,567 8	0,534 5	0,506 3	0,482 1	0,442 5	0,411 3	0,385 8	0,346 4	0,304 9	0,260 0	0,230 4
30	0,600 9	0,562 1	0,529 9	0,502 5	0,479 0	0,440 3	0,409 7	0,384 6	0,345 8	0,304 7	0,260 0	0,230 6
35	0,568 6	0,536 1	0,508 6	0,484 8	0,464 1	0,429 4	0,401 6	0,378 5	0,342 1	0,303 1	0,260 0	0,231 2
40	0,541 3	0,513 6	0,489 7	0,468 8	0,450 4	0,419 1	0,393 6	0,372 2	0,338 2	0,301 0	0,259 4	0,231 4
45	0,517 9	0,493 9	0,472 8	0,454 2	0,437 7	0,409 4	0,385 9	0,366 0	0,334 0	0,298 7	0,258 6	0,231 2
50	0,497 5	0,476 4	0,457 7	0,441 0	0,426 0	0,400 2	0,378 5	0,360 0	0,329 9	0,296 2	0,257 5	0,230 8

NOTE The critical values given in Table D.4 are correct to the 4th decimal place in the range $n = 3$ to 30 and $\nu = 0$, 5, 15 and 30[2]. Other values were derived from the Bonferroni inequality as:

$$B^* = t\left[\frac{(n-1)}{n\left(n+\nu-2+t^2\right)}\right]^{\frac{1}{2}} \qquad \text{(D.1)}$$

where t is the upper 0,005/n fractile of a t-variate with $n + \nu - 2$ degrees of freedom. The values so computed are only slightly conservative, and have a maximum error of approximately 0,000 2 above the true value. If critical values are required for intermediate values of n and ν, they may be estimated by second order interpolation using the square of the reciprocals of the tabulated values. Similarly, second order extrapolation can be used to estimate values beyond $n = 50$ and $\nu = 200$.

Table D.5 — Critical values of t

Degrees of freedom	Double-sided % significance level						
	50	40	30	20	10	5	1
1	1,000	1,376	1,963	3,078	6,314	12,706	63,657
2	0,816	1,061	1,386	1,886	2,920	4,303	9,925
3	0,765	0,978	1,250	1,638	2,353	3,182	5,841
4	0,741	0,941	1,190	1,533	2,132	2,776	4,604
5	0,727	0,920	1,156	1,476	2,015	2,571	4,032
6	0,718	0,906	1,134	1,440	1,943	2,447	3,707
7	0,711	0,896	1,119	1,415	1,895	2,365	3,499
8	0,706	0,889	1,108	1,397	1,860	2,306	3,355
9	0,703	0,883	1,100	1,383	1,833	2,262	3,250
10	0,700	0,879	1,093	1,372	1,812	2,228	3,165
11	0,697	0,876	1,088	1,363	1,796	2,201	3,106
12	0,695	0,873	1,083	1,356	1,782	2,179	3,055
13	0,694	0,870	1,079	1,350	1,771	2,160	3,012
14	0,692	0,868	1,076	1,345	1,761	2,145	2,977
15	0,691	0,866	1,074	1,341	1,753	2,131	2,947
16	0,690	0,865	1,071	1,337	1,746	2,120	2,921
17	0,689	0,863	1,069	1,333	1,740	2,110	2,898
18	0,688	0,862	1,067	1,330	1,734	2,101	2,878
19	0,688	0,861	1,066	1,328	1,729	2,093	2,861
20	0,687	0,860	1,064	1,325	1,725	2,086	2,845
21	0,686	0,859	1,063	1,323	1,721	2,080	2,831
22	0,686	0,858	1,061	1,321	1,717	2,074	2,819
23	0,685	0,858	1,060	1,319	1,714	2,069	2,807
24	0,685	0,857	1,059	1,318	1,711	2,064	2,797
25	0,684	0,856	1,058	1,316	1,708	2,060	2,787
26	0,684	0,856	1,058	1,315	1,706	2,056	2,779
27	0,684	0,855	1,057	1,314	1,703	2,052	2,771
28	0,683	0,855	1,056	1,313	1,701	2,048	2,763
29	0,683	0,854	1,055	1,311	1,699	2,045	2,756
30	0,683	0,854	1,055	1,310	1,697	2,042	2,750
40	0,681	0,851	1,050	1,303	1,684	2,021	2,704
50	0,680	0,849	1,048	1,299	1,676	2,008	2,678
60	0,679	0,848	1,046	1,296	1,671	2,000	2,660
120	0,677	0,845	1,041	1,289	1,658	1,980	2,617
∞	0,674	0,842	1,036	1,282	1,645	1,960	2,576

D.1 Critical values of F

D.1.1 General

See Reference [7] in the bibliography for the source of Tables D.6 to D.9.

Table D.6 — Critical 5 % values of F

ν_2	ν_1															
	3	4	5	6	7	8	9	10	15	20	30	50	100	200	500	∞
3	9,28	9,12	9,01	8,94	8,89	8,85	8,81	8,79	8,70	8,66	8,62	8,58	8,55	8,54	8,53	8,53
4	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96	5,86	5,80	5,75	5,70	5,66	5,65	5,64	5,63
5	5,41	5,19	5,05	4,95	4,88	4,82	4,77	4,74	4,62	4,56	4,50	4,44	4,41	4,39	4,37	4,37
6	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06	3,94	3,87	3,81	3,75	3,71	3,69	3,68	3,67
7	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,64	3,51	3,44	3,38	3,32	3,27	3,25	3,24	3,23
8	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,35	3,22	3,15	3,08	3,02	2,97	2,95	2,94	2,93
9	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,14	3,01	2,94	2,86	2,80	2,76	2,73	2,72	2,71
10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,98	2,85	2,77	2,70	2,64	2,59	2,56	2,55	2,54
15	3,29	3,06	2,90	2,79	2,71	2,64	2,59	2,54	2,40	2,33	2,25	2,18	2,12	2,10	2,08	2,07
20	3,10	2,87	2,71	2,60	2,51	2,45	2,39	2,35	2,20	2,12	2,04	1,97	1,91	1,88	1,86	1,84
30	2,92	2,69	2,53	2,42	2,33	2,27	2,21	2,16	2,01	1,93	1,84	1,76	1,70	1,66	1,64	1,62
50	2,79	2,56	2,40	2,29	2,20	2,13	2,07	2,03	1,87	1,78	1,69	1,60	1,52	1,48	1,46	1,44
100	2,70	2,46	2,31	2,19	2,10	2,03	1,97	1,93	1,77	1,68	1,57	1,48	1,39	1,34	1,31	1,28
200	2,65	2,42	2,26	2,14	2,06	1,98	1,93	1,88	1,72	1,62	1,52	1,41	1,32	1,26	1,22	1,19
500	2,62	2,39	2,23	2,12	2,03	1,96	1,90	1,85	1,69	1,59	1,48	1,38	1,28	1,21	1,16	1,11
∞	2,60	2,37	2,21	2,10	2,01	1,94	1,88	1,83	1,67	1,57	1,46	1,35	1,24	1,17	1,11	1,00

Table D.7 — Critical 1 % values of F

ν_2	ν_1															
	3	4	5	6	7	8	9	10	15	20	30	50	100	200	500	∞
3	29,5	28,7	28,2	27,9	27,7	27,5	27,3	27,2	26,9	26,7	26,5	26,4	26,2	26,2	26,1	26,1
4	16,7	16,0	15,5	15,2	15,0	14,8	14,7	14,5	14,2	14,0	13,8	13,7	13,6	13,5	13,5	13,5
5	12,1	11,4	11,0	10,7	10,5	10,3	10,2	10,1	9,72	9,55	9,38	9,24	9,13	9,08	9,04	9,02
6	9,78	9,15	8,75	8,47	8,26	8,10	7,98	7,87	7,56	7,40	7,23	7,09	6,99	6,93	6,90	6,88
7	8,45	7,85	7,46	7,19	6,99	6,84	6,72	6,62	6,31	6,16	5,99	5,86	5,75	5,70	5,67	5,65
8	7,59	7,01	6,63	6,37	6,18	6,03	5,91	5,81	5,52	5,36	5,20	5,07	4,96	4,91	4,88	4,86
9	6,99	6,42	6,06	5,80	5,61	5,47	5,35	5,26	4,96	4,81	4,65	4,52	4,42	4,36	4,33	4,31
10	6,55	5,99	5,64	5,39	5,20	5,06	4,94	4,85	4,56	4,41	4,25	4,12	4,01	3,96	3,93	3,91
15	5,42	4,89	4,56	4,32	4,14	4,00	3,89	3,80	3,52	3,37	3,21	3,08	2,98	2,92	2,89	2,87
20	4,94	4,43	4,10	3,87	3,70	3,56	3,46	3,37	3,09	2,94	2,78	2,64	2,54	2,48	2,44	2,42
30	4,51	4,02	3,70	3,47	3,30	3,17	3,07	2,98	2,70	2,55	2,39	2,25	2,13	2,07	2,03	2,01
50	4,20	3,72	3,41	3,19	3,02	2,89	2,79	2,70	2,42	2,27	2,10	1,95	1,82	1,76	1,71	1,68
100	3,98	3,51	3,21	2,99	2,82	2,69	2,59	2,50	2,22	2,07	1,89	1,73	1,60	1,52	1,47	1,43
200	3,88	3,41	3,11	2,89	2,73	2,60	2,50	2,41	2,13	1,97	1,79	1,63	1,48	1,39	1,33	1,28
500	3,82	3,36	3,05	2,84	2,68	2,55	2,44	2,36	2,07	1,92	1,74	1,56	1,41	1,31	1,23	1,16
∞	3,78	3,32	3,02	2,80	2,64	2,51	2,41	2,32	2,04	1,88	1,70	1,52	1,36	1,25	1,15	1,00

Table D.8 — Critical 0,1 % values of F

ν_2	ν_1															
	3	4	5	6	7	8	9	10	15	20	30	50	100	200	500	∞
3	141	137	135	133	132	131	130	129	127	126	125	125	124	124	124	124
4	56,2	53,4	51,7	50,5	49,7	49,0	48,5	48,0	46,8	46,1	45,4	44,9	44,5	44,3	44,1	44,0
5	33,2	31,1	29,8	28,8	28,2	27,6	27,2	26,9	25,9	25,4	24,9	24,4	24,1	23,9	23,8	23,8
6	23,7	21,9	20,8	20,0	19,5	19,0	18,7	18,4	17,6	17,1	16,7	16,3	16,0	15,9	15,8	15,8
7	18,8	17,2	16,2	15,5	15,0	14,6	14,3	14,1	13,3	12,9	12,5	12,2	11,9	11,8	11,7	11,7
8	15,8	14,4	13,5	12,9	12,4	12,0	11,8	11,5	10,8	10,5	10,1	9,80	9,57	9,46	9,39	9,34
9	13,9	12,6	11,7	11,1	10,7	10,4	10,1	9,89	9,24	8,90	8,55	8,26	8,04	7,93	7,86	7,81
10	12,6	11,3	10,5	9,92	9,52	9,20	8,96	8,75	8,13	7,80	7,47	7,19	6,98	6,87	6,81	6,76
15	9,34	8,25	7,57	7,09	6,74	6,47	6,26	6,08	5,53	5,25	4,95	4,70	4,51	4,41	4,35	4,31
20	8,10	7,10	6,46	6,02	5,69	5,44	5,24	5,08	4,56	4,29	4,01	3,77	3,58	3,48	3,42	3,38
30	7,05	6,12	5,53	5,12	4,82	4,58	4,39	4,24	3,75	3,49	3,22	2,98	2,79	2,69	2,63	2,59
50	6,34	5,46	4,90	4,51	4,22	4,00	3,82	3,67	3,20	2,95	2,68	2,44	2,24	2,14	2,07	2,03
100	5,85	5,01	4,48	4,11	3,83	3,61	3,44	3,30	2,84	2,59	2,32	2,07	1,87	1,75	1,68	1,62
200	5,64	4,81	4,29	3,92	3,65	3,43	3,26	3,12	2,67	2,42	2,15	1,90	1,68	1,55	1,46	1,39
500	5,51	4,69	4,18	3,82	3,54	3,33	3,16	3,02	2,58	2,33	2,05	1,80	1,57	1,43	1,32	1,23
∞	5,42	4,62	4,10	3,74	3,47	3,27	3,10	2,96	2,51	2,27	1,99	1,73	1,49	1,34	1,21	1,00

Table D.9 — Critical 0,05 % **values of** F

ν_2	ν_1															
	3	4	5	6	7	8	9	10	15	20	30	50	100	200	500	∞
3	225	218	214	211	209	208	207	206	203	201	199	198	197	197	196	196
4	80,1	76,1	73,6	71,9	70,6	69,7	68,9	68,3	66,5	65,5	64,6	63,8	63,2	62,9	62,7	62,6
5	44,4	41,5	39,7	38,5	37,6	36,9	36,4	35,9	34,6	33,9	33,1	32,5	32,1	31,8	31,7	31,6
6	30,4	28,1	26,6	25,6	24,9	24,3	23,9	23,5	22,4	21,9	21,4	20,9	20,5	20,3	20,2	20,1
7	23,5	21,4	20,2	19,3	18,7	18,2	17,8	17,5	16,5	16,0	15,5	15,1	14,7	14,6	14,5	14,4
8	19,4	17,6	16,4	15,7	15,1	14,6	14,3	14,0	13,1	12,7	12,2	11,8	11,6	11,4	11,4	11,3
9	16,8	15,1	14,1	13,3	12,8	12,4	12,1	11,8	11,0	10,6	10,2	9,80	9,53	9,40	9,32	9,26
10	15,0	13,4	12,4	11,8	11,3	10,9	10,6	10,3	9,56	9,16	8,75	8,42	8,16	8,04	7,96	7,90
15	10,8	9,48	8,66	8,10	7,68	7,36	7,11	6,91	6,27	5,93	5,58	5,29	5,06	4,94	4,87	4,83
20	9,20	8,02	7,28	6,76	6,38	6,08	5,85	5,66	5,07	4,75	4,42	4,15	3,93	3,82	3,75	3,70
30	7,90	6,82	6,14	5,66	5,31	5,04	4,82	4,65	4,10	3,80	3,48	3,22	3,00	2,89	2,82	2,78
50	7,01	6,01	5,37	4,93	4,60	4,34	4,14	3,98	3,45	3,16	2,86	2,59	2,37	2,25	2,17	2,13
100	6,43	5,47	4,87	4,44	4,13	3,89	3,70	3,54	3,03	2,75	2,44	2,18	1,95	1,82	1,74	1,67
200	6,16	5,23	4,64	4,23	3,92	3,68	3,49	3,34	2,83	2,56	2,25	1,98	1,74	1,60	1,50	1,42
500	6,01	5,09	4,51	4,10	3,80	3,56	3,36	3,21	2,72	2,45	2,14	1,87	1,61	1,46	1,34	1,24
∞	5,91	5,00	4,42	4,02	3,72	3,48	3,30	3,14	2,65	2,37	2,07	1,79	1,53	1,36	1,22	1,00

D.1.2 Approximate equation for critical values of F

Critical values of F for untabulated values of ν_1 and ν_2 may be approximated by second order interpolation from Tables D.6 to D.9.

Critical values of F corresponding to $\nu_1 > 30$ and $\nu_2 > 30$ degrees of freedom and a significance level 100 $(1 - p)$ %, where p is the probability, can also be approximated from the equation:

$$\log_{10}(F) = \frac{A(p)}{\sqrt{b - B(p)}} - C(p)\left(\frac{1}{\nu_1} - \frac{1}{\nu_2}\right) \tag{D.2}$$

where

$$\frac{1}{b} = \frac{1}{2}\left(\frac{1}{\nu_1} + \frac{1}{\nu_2}\right)$$

Values of $A(p)$, $B(p)$ and $C(p)$ are given in Table D.10 for typical values of significance level 100 $(1 - p)$.

Table D.10 — Typical values of equation parameters

100 $(1 - p)$%	$A(p)$	$B(p)$	$C(p)$
10	1,113 1	0,77	0,527
5	1,428 7	0,95	0,681
2,5	1,702 3	1,14	0,846
1	2,020 6	1,40	1,073
0,5	2,237 3	1,61	1,250
0,1	2,684 1	2,09	1,672
0,05	2,858 0	2,30	1,857

For values of p not given in Table D.10, critical values of F may be obtained by second order interpolation/extrapolation of $\log_{10}(F)$ (either tabulated or estimated from the equation) against $\log_{10}(1 - p)$.

D.2 Critical values of the normal distribution

Critical values, Z, corresponding to a single-sided probability, p, or to a double-sided significance level $2(1 - p)$, are given in Table D.11 in terms of the "standard normal deviate", where

$$Z = \frac{x - \mu}{\sigma} \tag{D.3}$$

and where μ and σ are the mean and standard deviation respectively of the normal distribution.

Table D.11 — Critical values of the normal distribution

p	0,70	0,80	0,90	0,95	0,975	0,99	0,995
Z	0,524	0,842	1,282	1,645	1,960	2,326	2,576
$2(1 - p)$	0,60	0,40	0,20	0,10	0,05	0,02	0,01

When p is less than 0,5, the appropriate critical value is the negative of the value corresponding to a probability $(1 - p)$.

Annex E
(normative)

Types of dependence and corresponding transformations

NOTE See 5.2.

E.1 Types of dependence

The forms of dependence given in Table E.1 are shown graphically in Figures E.1 to E.8. In all cases, K can be any positive constant, and "ln" refers to Naperian logarithms. The form of the line to be fitted includes a dummy variable, T (see Clause F.1), by which it is possible to test for a difference in the transformation as applied to repeatability and reproducibility.

Table E.1

Form of dependence	Transformation	Form of line to be fitted	$\frac{\mathrm{d}x}{\mathrm{d}y}$	Remarks
1. $D = K(m + B)$ $0 < (m + B)$	$Y = \ln(x + B)$ "log model"	$\ln(D) = b_0 + b_1 \ln(m + B) + ...$ $... + b_2 T + b_3 T \ln(m + B)$ Test: $b_1 = 1$	$(x + B)$	Care shall be taken if $(x + B)$ is small, since rounding becomes critical.
2. $D = Km^B$ $B \neq 1$	$y = x^{1-B}$ "power model"	$\ln(D) = b_0 + b_1 \ln(m) + b_2 T + ...$ $... + b_3 T \ln(m)$ Test: $b_1 \neq 0$	$\frac{X^B}{(1-B)}$	The fitted line will pass through the origin. B = 1/2 or 2 are common cases.
3. $D = K(m + B_0)^B$ $B \neq 1$ $B_0 \neq 0$ $0 < (m + B_0)$	$y = (x + B_0)^{1-B}$ "power-with-intercept model"	$\ln(D) = b_0 + b_1 \ln(m + B_0) + ...$ $... + b_2 T + b_3 T \ln(m + B_0)$ Test: $b_1 \neq 0$	$\frac{X^B}{(1-B)}$	The fitted line will not pass through the origin.
4. $d = K\sqrt{\frac{m}{B}\left(1-\frac{m}{B}\right)}$ $0 \leqslant m \leqslant B$	$y = \arcsin\sqrt{\frac{x}{B}}$ "arcsin model"	$\ln(D) = b_0 + b_1 \ln[m(B-m)] + ...$ $... + b_2 T + b_3 T \ln[m(B-m)]$ Test: $b_1 = 1/2$	$2\sqrt{x(B-x)}$	This case often arises when results are reported as percentages or qualitatively as "scores". If x is always small, the transformation reduces to $y = \sqrt{x}$, a special case of 2 above.
5. $D = K\frac{m}{B}\left(1-\frac{m}{B}\right)$ $0 \leqslant m \leqslant B$	$y = \ln\left(\frac{x}{B-x}\right)$ "logistic model"	$\ln(D) = b_0 + b_1 \ln[m(B-m)] + ...$ $... + b_2 T + b_3 T \ln[m(B-m)]$ Test: $b_1 = 1$	$\frac{x(B-x)}{B}$	This case arises when results are reported on a rating scale of 0 to B. If x is always small, then the transformation reduces to $y = \ln(x)$, a special case of 1 above.
6. $D = K\left(\frac{m^2 + B^2}{B}\right)$ $B > 0$	$y = \arctan\left(\frac{x}{B}\right)$ "arctan model"	$\ln(D) = b_0 + b_1 \ln(m^2 + B^2) + ...$ $... + b_2 T + b_3 T \ln(m^2 + B^2)$ Test: $b_1 = 1$	$\frac{(x^2 + B^2)}{B}$	The fitted line does not pass through the origin. If B is small, the transformation reduces to $y = 1/x$, a special case of 2 above.

E.2 Transformation procedure

The following steps shall be taken in identifying the correct type of transformation and its parameters B and B_0.

a) Plot laboratories standard deviations, D, and repeats standard deviations, d, against sample means in the form of scatter diagrams. Refer to Figures E.1 to E.8 and identify the type of transformation to be applied (if any).

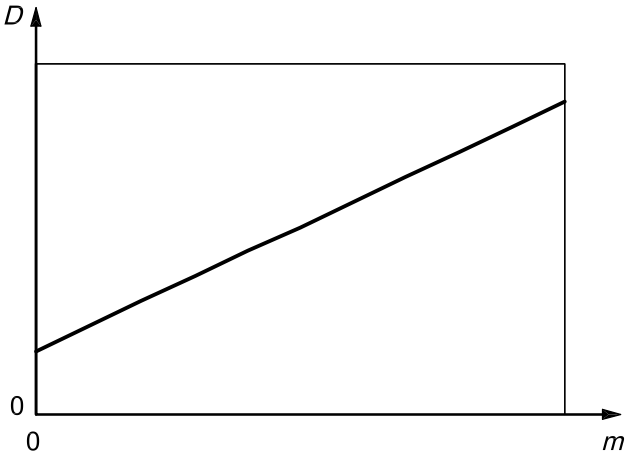
b) With the exception of the power transformations (types 2 and 3 in Table E.1), estimate the transformation parameter, B, from the scatter diagrams. These are known for the arcsin and the logistic transformation (types 4 and 5, respectively, in Table E.1), since B, in both cases, is the upper limit of the rating scale or "score" which defines results. For the log transformation (type 1 in Table E.1), calculate B from the intercept and slope estimated from the scatter diagrams. Similarly, estimate B from the intercept in the case of the arctan transformation (type 6 in Table E.1). In every case, B shall be rounded to give a meaningful value that satisfies the plots for both the laboratories and repeats standard deviations. For the power-with-intercept transformation (type 3 in Table E.1), also estimate the parameter B_0 from the scatter diagrams (but see the Note below).

c) Fit the line specified in Table E.1, corresponding to the transformation in question, according to the computational procedure in Annex F.3 (but see the Note below). For both types of power transformations, coefficient b_1 shall differ significantly from zero and provide an estimate of B, which shall be rounded to a meaningful value. For the arcsin transformation, b_1 shall have a value not significantly different from 0,5. Similarly, b_1 shall not significantly differ from a value of 1 for the logistic, log and arctan transformations. In every case, the test specified in Table E.1 shall be applied at the 5 % significance level.

   Failure of this test implies either that the type of transformation or its parameters B and/or B_0 is/are incorrect. Similarly, the coefficient b_3 shall in every case be tested as zero. Failure in this case implies that the transformation is different for repeatability and reproducibility. In some cases, the presence of outliers (see 5.3) can give rise to this difference.

d) If the tests applied in a) to c) above are satisfactory, transform all the results accordingly, recalculate means and standard deviations using transformed results and create new scatter diagrams as in paragraph a) above. These now show a uniform level for laboratories standard deviation, and a uniform (but not necessarily the same) level for repeats standard deviation. A statistical test for uniformity is given in 5.4.

NOTE For the power-with-intercept transformation, B and B_0 cannot be estimated together according to the linear least squares technique described in Clause F.3. A non-linear and iterative technique is required instead, such as the simplex procedure of Nelder and Mead[10], necessitating the use of appropriate computer software[9].

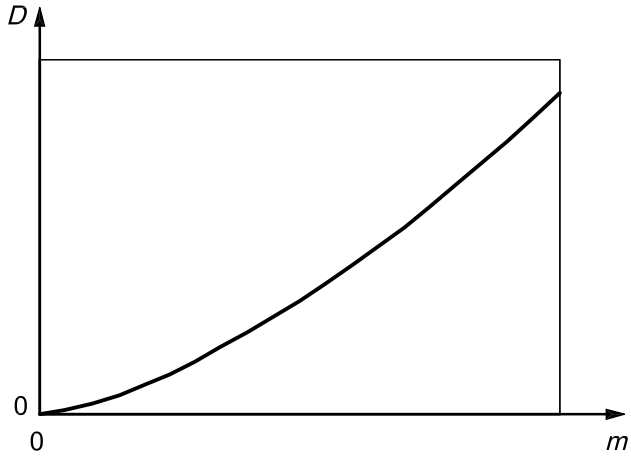


Key

m mean

D standard deviation, $D = K(m + B)$, $(m + B) > 0$

Figure E.1

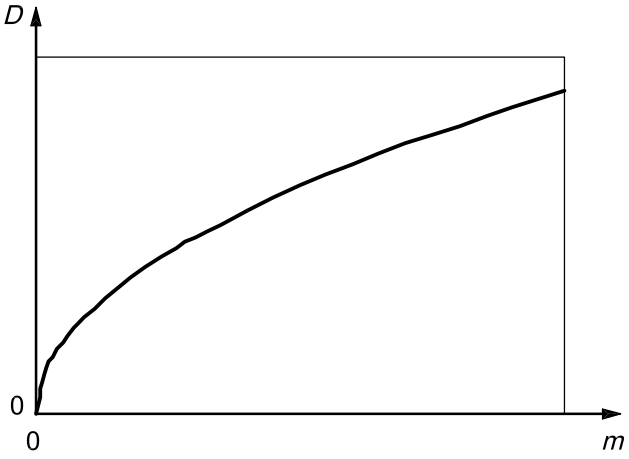


Key

m mean

D standard deviation, $D = Km^B$, $B > 1$

Figure E.2

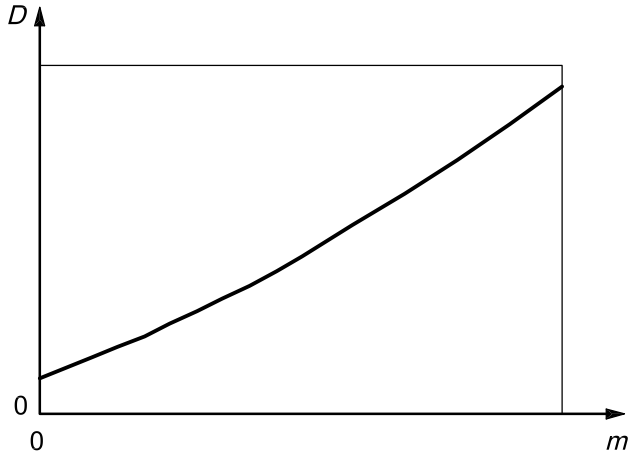


Key

m mean

D standard deviation, $D = Km^B$, $0 < B < 1$

Figure E.3

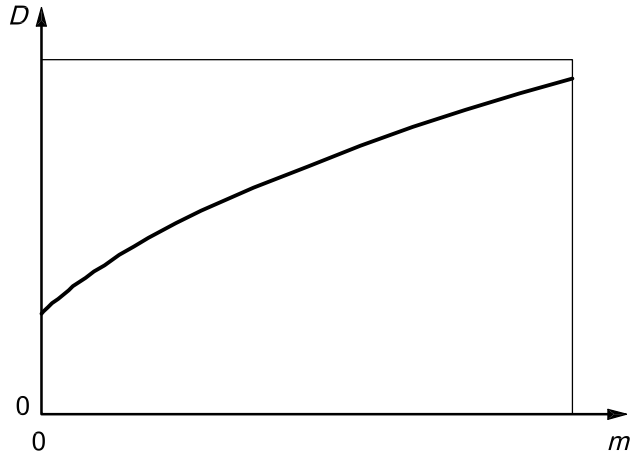


Key

m mean

D standard deviation, $D = K(m + B_0)^B$, $B > 1$, $B_0 \neq 0$

Figure E.4

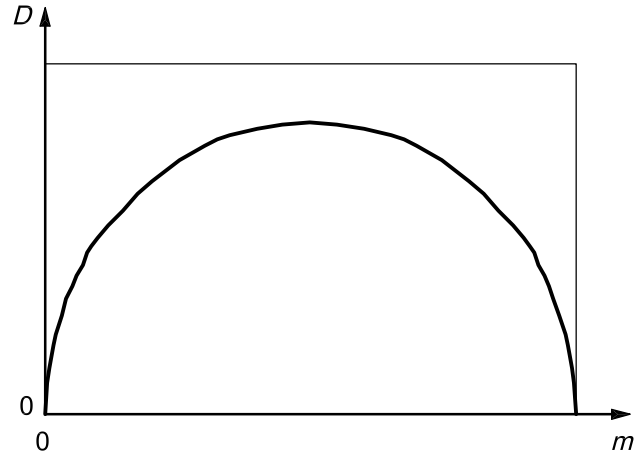


Key

m mean

D standard deviation, $D = K(m + B_0)^B$, $0 < B < 1$, $B_0 \neq 0$

Figure E.5

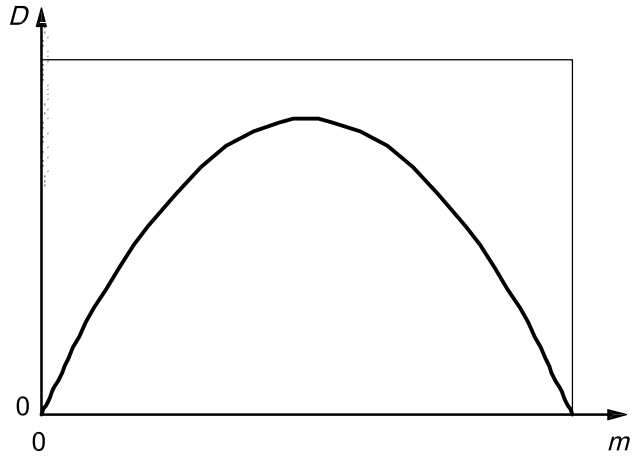


Key

m mean

D standard deviation, $D = K\sqrt{\frac{m}{B}\left(1-\frac{m}{B}\right)}$, $0 \leqslant m \leqslant B$

Figure E.6

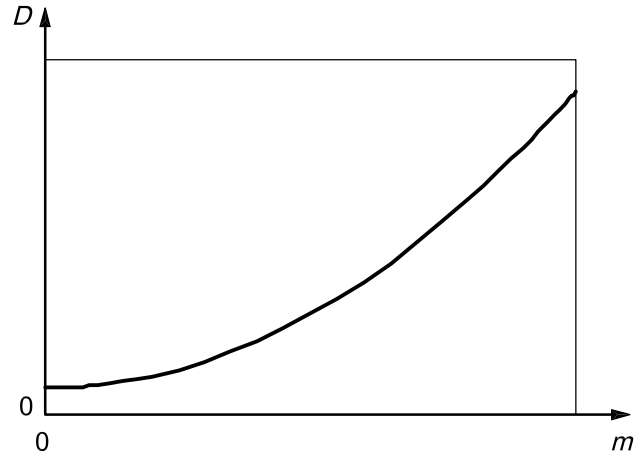


Key

m mean

D standard deviation, $D = K\frac{m}{B}\left(1-\frac{m}{B}\right)$, $0 \leqslant m \leqslant B$

Figure E.7



Key

m mean

D standard deviation, $D = K\left(\frac{m^2 + B^2}{B}\right)$, $B > 0$

Figure E.8

Annex F
(normative)

Weighted linear regression analysis

NOTE See 5.2.

F.1 Explanation for the use of a dummy variable

Two different variables, Y_1 and Y_2, when plotted against the same independent variable, X, in general give different linear relationships of the form:

$$Y_1 = b_{10} + b_{11}X$$

$$Y_2 = b_{20} + b_{21}X \quad \text{(F.1)}$$

where the coefficients b_{ij} are estimated by regression analysis. In order to compare the two relationships, a dummy variable, T, can be defined such that:

$T = T_1$, a constant value for every observation of Y_1

$T = T_2$, a constant value for every observation of Y_2

$T_1 \neq T_2$

Letting Y represent the combination of Y_1 and Y_2, plot a single relationship,

$$Y = b_0 + b_1X + b_2T + b_3TX \quad \text{(F.2)}$$

where, as before, the coefficients b_i are estimated by regression analysis. By comparing Equations (F.1) and (F.2), it is evident that

$$b_{10} = b_0 + b_2T_1$$

$$b_{20} = b_0 + b_2T_2 \quad \text{(F.3)}$$

and that therefore

$$b_{10} - b_{20} = b_2(T_1 - T_2) \quad \text{(F.4)}$$

Similarly,

$$b_{11} - b_{21} = b_3(T_1 - T_2) \quad \text{(F.5)}$$

In order to test for a difference between b_{10} and b_{20}, therefore, it is only necessary to test for a non-zero coefficient b_2. Similarly, to test for a difference between b_{11} and b_{21}, test for a non-zero coefficient b_3.

Any non-zero values can be chosen for T_1 and T_2. However, since reproducibility is the basis of tests for quality control against specifications (see Clauses 9 and 10), weighting shall reflect this in the estimation of precision relationships. An “importance ratio” of 2:1 in the favour of reproducibility shall be applied by setting $T_1 = 1$ and $T_2 = -2$, where T_1 refers to the plot of laboratories standard deviation and T_2 refers to the repeats standard deviation.

F.2 Derivation of weights used in regression analysis

In order to account for the relative precision of fitted variables in a regression analysis, weights shall be used that are inversely proportional to the variance of the fitted variables.

For a variable, D, which is an estimate of population standard deviation, σ, based on $\nu(D)$ degrees of freedom, the variance of D is given by

$$Var(D) = \sigma^2/2\,\nu(D) \tag{F.6}$$

Replacing σ^2 by its estimate D^2, the weight for this variable is approximated by

$$w(D) = 2\,\nu(D)/D^2 \tag{F.7}$$

It is clear that as standard deviation, D, increases, so does the weight decrease. For this reason, the fitted variable in the weighted regression shall instead be a function of standard deviation which yields weights independent of the fitted variable.

In cases where a function $g(D)$, rather than D itself, is fitted, the variance equation becomes

$$Var[g(D)] = \left(\frac{\delta g}{\delta D}\right)^2 Var(D) \tag{F.8}$$

Hence, for the Naperian logarithm function:

$$Var[\ln(D)] = \frac{1}{D^2} Var(D) = \frac{1}{D^2}\frac{\sigma^2}{2\nu(D)} \tag{F.9}$$

Once again, replacing σ^2 by its estimate D^2, the weight for $\ln(D)$ will be approximated by:

$$w[\ln(D)] = 2\,\nu(D) \tag{F.10}$$

In relation to laboratories standard deviation D and repeats standard deviation d, therefore, it is necessary to perform regression analysis in terms of $\ln(D)$ and $\ln(d)$, since weighting then takes account only of the amount of data on which the standard deviation was based. A relationship estimated in this way is less dependent on samples that have a high proportion of missing results.

Denoting degrees of freedom as $\nu(D)$ for laboratory standard deviations, D, and $\nu(d)$ for repeats standard deviations, d, equations for calculating weights then become

$$w[\ln(D)] = 2\,\nu(D) \tag{F.11}$$

$$w[\ln(d)] = 2\,\nu(d) \tag{F.12}$$

NOTE Unweighted regression corresponds to weighted regression in which all the weights have a constant value of 1.

F.3 Computational procedure in regression analysis

The following technique gives the best fitting straight line of the form of Equation (F.2) (but see the Note to E.2).

First draw up a table (see Table F.1) giving values of the variables to be plotted in the regression, together with corresponding weights. Functions g_1 and g_2 are always Naperian logarithms corresponding to the transformation in question, as specified in Clause E.2.

Table F.1

Sample	Standard deviation function g_1	Sample mean function g_2	Dummy T	Tg_2	Weight
1	$g_1(D_1)$	$g_2(m_1)$	1	$g_2(m_1)$	$2\nu(D_1)$
2	$g_1(D_2)$	$g_2(m_2)$	1	$g_2(m_2)$	$2\nu(D_2)$
3	$g_1(D_3)$	$g_2(m_3)$	1	.	.
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.
S	$g_1(D_S)$	$g_2(m_S)$	1	$g_2(m_S)$	$2\nu(D_S)$
1	$g_1(d_1)$	$g_2(m_1)$	–2	$-2g_2(m_1)$	$2\nu(d_1)$
2	$g_1(d_2)$	$g_2(m_2)$	–2	.	$2\nu(d_2)$
3	$g_1(d_3)$	$g_2(m_3)$	–2	.	.
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.
S	$g_1(d_S)$	$g_2(m_S)$	–2	$-2g_2(m_S)$	$2\nu(d_S)$
Symbol	y_i	x_{1i}	x_{2i}	x_{3i}	w_i

Using the symbols defined in Table F.1, the line to be fitted (Equation F.2) becomes:

$$y = b_0 + b_1x_1 + b_2x_2 + b_3x_3 \tag{F.13}$$

The intercept, b_0, can be eliminated by rewriting this as

$$\left(y - \bar{y}\right) = b_1\left(x_1 - \bar{x}_1\right) + b_2\left(x_2 - \bar{x}_2\right) + b_3\left(x_3 - \bar{x}_3\right) \tag{F.14}$$

where $\bar{y}$, $\bar{x}_1$, $\bar{x}_2$ and $\bar{x}_3$ are weighted mean values, for example

$$\bar{x}_2 = \frac{\sum_{i=1}^{n} w_i x_{2i}}{\sum_{i=1}^{n} w_i} \tag{F.15}$$

and where n is the number of points (twice the number of samples) to be plotted.

The least squares solution of Equation (F.14) requires the solution of the set of simultaneous equations of the form:

$$\begin{aligned} a_{y1} &= a_{11}b_1 + a_{12}b_2 + a_{13}b_3 \\ a_{y2} &= a_{21}b_1 + a_{22}b_2 + a_{23}b_3 \\ a_{y3} &= a_{31}b_1 + a_{32}b_2 + a_{33}b_3 \end{aligned} \tag{F.16}$$

Examples of a_{ij} and a_{yi} elements, in terms of weighted means $\bar{x}_i$, are as follows:

$$a_{22} = \sum w_i (x_{2i} - \bar{x}_2)^2 \qquad a_{23} = \sum w_i (x_{2i} - \bar{x}_2)(x_{3i} - \bar{x}_3)$$

$$a_{y2} = \sum w_i \left(y_i - \bar{y}\right)\left(x_{2i} - \bar{x}_2\right) \qquad a_{yy} = \sum w_i \left(y_i - \bar{y}\right)^2$$

Having solved the equations for b_1, b_2 and b_3, calculate the intercept from the weighted means of the variables as

$$b_0 = \bar{y} - b_1\bar{x}_1 - b_2\bar{x}_2 - b_3\bar{x}_3 \tag{F.17}$$

The coefficient estimates, b_i, can be summarized in tabular form, together with test statistics, as given in Table F.2:

Table F.2

Fitted variable	Coefficient estimate	Standard error of estimate	t-Ratio
Intercept	b_0	e_0	t_0
Sample mean	b_1	e_1	t_1
Dummy	b_2	e_2	t_2
Dummy × mean	b_3	e_3	t_3

In order to complete the table, it is necessary to calculate the standard deviation of the observed y values about the estimated line. This is called the residual standard deviation, s, and is given by

$$s = \sqrt{\frac{1}{n-4}(a_{yy} - b_1 a_{y1} - b_2 a_{y2} - b_3 a_{y3})} \tag{F.18}$$

Standard errors of the estimates then become

$e_i = s\sqrt{c_{ii}}$, for $i = 1$ to 3

and

$$e_0 = S\sqrt{\frac{1}{n} + c_{11}\bar{x}_1^2 + c_{22}\bar{x}_2^2 + c_{33}\bar{x}_3^2 + 2c_{12}\bar{x}_1\bar{x}_2 + 2c_{13}\bar{x}_1\bar{x}_3 + 2c_{23}\bar{x}_2\bar{x}_3} \tag{F.19}$$

where the elements c_{ij} correspond to the inverse of the matrix containing elements a_{ij}.

The t-ratios are the ratios $(b_i - K)/e_i$, where K is a constant, and by comparing these to the critical values of t in Table D.5, it is possible to test if coefficient b_i differs from K. If t_i is greater than the critical value corresponding to 5 % significance and $(n - 4)$ degrees of freedom, then the coefficient can be regarded as differing from K. In particular, t_1 identifies an inappropriate slope b_1 and t_3 indicates whether the slope is different for laboratories and repeats standard deviations. Since laboratories standard deviation is generally larger than repeats standard deviation at the same level of sample mean, t_2 in general indicates a non-zero coefficient, b_2.

F.4 Worked example

This subclause describes the fitting of a power function (form of dependence 2 of Annex E) using weighted linear regression according to the procedure of Clause E.2. Rounded sample means and standard deviations are given in Table 1, based on the bromine data given in D.1.

a) Scatter diagrams identify the power transformation as appropriate, as indicated by the log-log plot shown in Figure F.1.

b) It is not necessary to estimate the transformation parameter, B, from Figure F.1, since it is given in the regression analysis that follows.

c) The form of the line that is being fitted (see Table E.1) is

$$\ln(D) = b_0 + b_1 \ln(m) + b_2 T + b_3 T \ln(m)$$

The table of values that are being fitted (see Table F.1) is as given in Table F.3.

Table F.3

Sample	Logarithm of standard deviation	Logarithm of sample mean	Dummy T	Dummy × ln(m)	Weight
1	– 0,315 8	0,765 5	1	0,765 5	16
2	0,796 9	4,180 4	1	4,180 4	18
3	– 2,704 6	– 0,280 2	1	– 0,280 2	28
4	– 1,556 8	1,293 2	1	1,293 2	22
5	– 1,235 8	2,388 8	1	2,388 8	18
6	0,402 9	3,875 5	1	3,875 5	18
7	1,076 2	4,737 8	1	4,737 8	18
8	– 1,840 1	0,197 5	1	0,197 5	18
1	– 2,064 4	0,765 5	–2	– 1,530 9	18
2	– 0,201 5	4,180 4	–2	– 8,360 9	18
3	– 2,995 7	– 0,280 2	–2	0,560 5	18
4	– 2,158 5	1,293 2	–2	– 2,586 4	18
5	– 2,361 3	2,388 8	–2	– 4,777 5	18
6	– 0,641 5	3,875 5	–2	– 7,751 0	18
7	– 0,067 4	4,737 8	–2	– 9,475 6	18
8	– 2,861 2	0,197 5	–2	– 0,394 9	18
Symbol	y_i	x_{1i}	x_{2i}	x_{3i}	w_i

Least squares regression requires the solution of the simultaneous equations given below:

$$614{,}671 = 999{,}894 b_1 - 35{,}852\ 4 b_2 - 493{,}045 b_3$$

$$188{,}526 = -\ 35{,}852\ 4 b_1 + 673{,}920 b_2 + 1\ 409{,}58 b_3$$

$$195{,}477 = -\ 493{,}045 b_1 + 1\ 409{,}58 b_2 + 5\ 362{,}27 b_3$$

Also required are

$a_{yy} = 505{,}668$

$s = 2{,}238\ 68$ [see Equation (F.18)]

The solution is summarized in Table F.4 (see Table F.2):

Table F.4

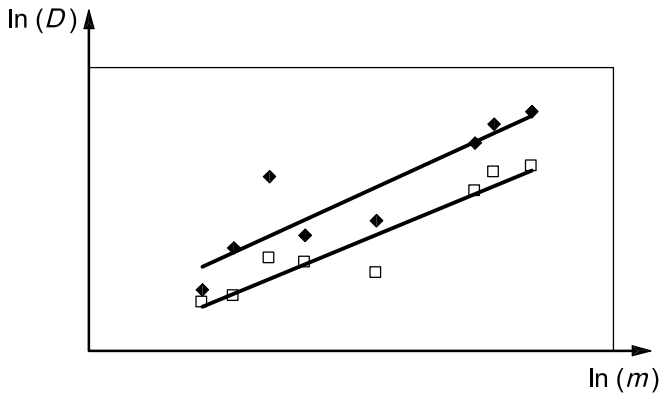
Fitted variable	Coefficient estimate b_i	Standard error of estimate	t-Ratio
Intercept	–2,406 4	—	—
ln(m)	0,637 73	0,073 59	8,67
Dummy	0,254 96	0,130 52	1,95
Dummy × ln (m)	0,028 08	0,047 31	0,59

Comparing the t-ratios with the critical 5 % values for 12 degrees of freedom (namely 2,179) given in Table D.5, it can be seen that the slope is significantly non-zero ($b_1 = 0{,}638$), confirming that a transformation was required. Furthermore, since coefficient b_3 does not significantly differ from zero, the slope (and resulting transformation) are the same for both laboratories and repeats standard deviations.

d) As the slope $b_1 = 0{,}638$ has a standard error of 0,074, the approximate 66 % confidence region of $0{,}638 \pm 0{,}074$ contains the value 2/3. Rounding to this value is, therefore, reasonable, and leads to the convenient transformation

$$y = x^{1/3}$$

Having applied this transformation and recalculated sample means and standard deviations, corresponding scatter diagrams are shown in Figure F.2. These show uniform levels for both laboratories and repeats standard deviations for all samples except sample 1. In the case of the latter sample, the extreme point is due to outliers (see 5.3.3.2).



Key

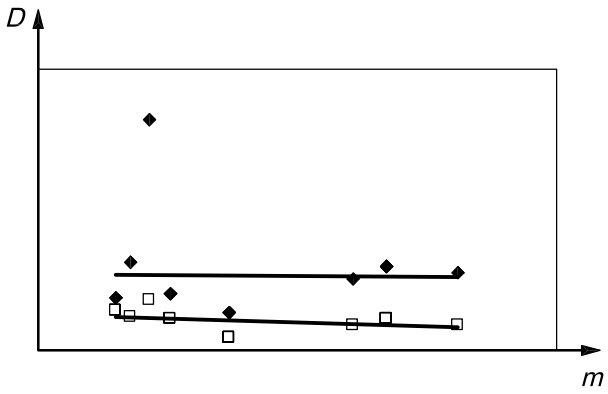
ln(m) logarithm of the sample mean

ln(D) logarithm of the standard deviation

◆ laboratories standard deviation

□ repeats standard deviation

Figure F.1



Key

m sample mean

D standard deviation

◆ laboratories standard deviation

□ repeats standard deviation

Figure F.2

Annex G
(normative)

Rules for rounding off results

NOTE See 3.22.

G.1 The general rule is that results shall not be rounded off coarser than one-tenth of the reproducibility of the test method. In practice, this means that for the decimal system, this choice is made from the series of 1 — 0,5 — 0,2 — 0,1 — 0,05 — 0,02, etc., (since 1, 2 and 5 are the only integral factors of 10). If the figure for one-tenth of the reproducibility is not among this series, the nearest lower number shall be used. Thus, if the reproducibility is 5, the result shall be rounded off to 0,5, but if the reproducibility is 4, the result shall be rounded off to 0,2.

G.2 When a number is rounded off to a rounding unit, the nearest unit shall be chosen. If two choices are available, as when the number to be rounded is exactly halfway between the two nearest rounded numbers, the rounded number that is an even multiple of the rounding unit shall be chosen. For example, if the rounding unit is 0,1, then 23,55 shall be rounded to 23,6, while 23,45 shall be rounded to 23,4. Again, if the rounding unit is 0,02, then 5,03 shall be rounded to 5,04, while 5,01 shall be rounded to 5,00.

G.3 There are cases, however, in which the rounding-off procedure is either specified or otherwise implied by the test method itself. In these cases, the above rules based on statistical reasoning might not always be applicable. For example, in a standard test method for the determination of viscosity of cutback bitumen and road oil, the reproducibility is given as 2 s, or 10 % of the mean for efflux times of 20 s and above, but a rounding-off interval of 1 s is specified.

Annex H
(informative)

Explanation of equations given in Clause 7

Let σ_0^2 be the variance of results obtained under repeatability conditions.

Let σ_1^2 be the variance of all other errors which contribute to reproducibility errors. Then ($\sigma_0^2 + \sigma_1^2$) is the variance of results obtained under reproducibility conditions.

r is calculated as $Z\sqrt{2}\sigma_0$ (H.1)

and

R is calculated as $Z\sqrt{2}\sqrt{\sigma_0^2+\sigma_1^2}$ (H.2)

where Z is the factor[6] for converting a standard deviation to a confidence limit (see Table D.11), and which corresponds in this case to a double-sided 95 % probability level, having a value of 1,96.

Furthermore, the variance of means of k results obtained under repeatability conditions is σ_0^2/k.

In a set of k such results, therefore, the variance of the difference between a single result and the mean of the remainder is

$$\sigma_0^2+\frac{\sigma_0^2}{(k-1)}=\frac{r^2}{1{,}96^2}\times\frac{k}{2(k-1)} \tag{H.3}$$

and the 95 % confidence limit for the absolute difference is

$$r_1=1{,}96\sqrt{\sigma_0^2+\frac{\sigma_0^2}{k-1}}=r\sqrt{\frac{k}{2(k-1)}} \tag{H.4}$$

If the mean of k results is obtained in each of several laboratories, these laboratory means have a variance

$$\frac{\sigma_0^2}{k}+\sigma_1^2=\frac{1}{2}\times\frac{1}{1{,}96^2}\left[R^2-r^2\left(1-\frac{1}{k}\right)\right] \tag{H.5}$$

Let

$$R_1=\sqrt{R^2-r^2\left(1-\frac{1}{k}\right)} \tag{H.6}$$

The double-sided 95 % confidence limits for such means are

$$\text{mean}\pm 1{,}96\sqrt{\frac{\sigma_0^2}{k}+\sigma_1^2}=\text{mean}\pm\frac{R_1}{\sqrt{2}} \tag{H.7}$$

Confidence limits for probability levels other than 95 % may be calculated by selecting the appropriate Z value from Table D.11 (with single- or double-sided probability as required) and multiplying by the conversion factor Z/1,96. For a single-sided probability of 95 %, Z = 1,64 and the conversion factor is 0,84.

In general, N laboratories obtain average results from k_1, k_2, ...k_N results, respectively. The variance of the average of N such laboratory averages is

$$\frac{1}{N^2}\left[\left(\frac{\sigma_0^2}{k_1}+\sigma_1^2\right)+\ldots+\left(\frac{\sigma_0^2}{k_N}+\sigma_1^2\right)\right]=\frac{1}{2N1{,}96^2}\left[R^2-\frac{r^2}{N}\left(N-\frac{1}{k_1}-\ldots-\frac{1}{k_N}\right)\right] \tag{H.8}$$

$$\text{Let } R_4=\sqrt{R^2-\frac{r^2}{N}\left(N-\frac{1}{k_1}-\ldots-\frac{1}{k_N}\right)} \tag{H.9}$$

The double-sided 95 % confidence limits for such means then becomes

$$\text{mean } \pm\frac{R_4}{\sqrt{2N}} \tag{H.10}$$

Confidence limits for probability levels other than 95 %, single- or double-sided as required, can be calculated by selecting the appropriate value Z from Table D.11, and multiplying by the conversion factor Z/1,96.

In a set of $N+1$ such averages, therefore, the variance of the difference between a single average of k results and the mean of the remaining N averages is

$$\frac{R_1^2}{2\times 1{,}96^2}+\frac{R_4^2}{2N1{,}96^2} \tag{H.11}$$

The 95 % confidence limit, R_3, for the absolute difference is therefore

$$R_3=\sqrt{\frac{R_1^2}{2}+\frac{R_4^2}{2N}} \tag{H.12}$$

In the case of only two laboratory averages (when $N=1$) this equation reduces to

$$R_2=\sqrt{R^2-r^2\left(1-\frac{1}{2k_1}-\frac{1}{2k_2}\right)} \tag{H.13}$$

Annex I
(informative)

Specifications that relate to a specified degree of criticality

I.1 Criticality of specifications

NOTE See 8.1.

Some specifications, because of the product characteristic or the end use of the product, or both, require that the recipient has a high degree of assurance that the product meets or exceeds the quality level indicated by the specification level(s). For the purpose of this International Standard, such specifications are called *critical* specifications.

Specifications that require assurance only that the quality is not substantially poorer than is indicated by the specification are called *non-critical* specifications for the purposes of this International Standard.

In any exchange of product, the *degree of criticality* of a specification is defined to be the maximum probability (risk) that the recipient can tolerate of accepting a shipment which fails specifications. The degree of criticality is denoted by p_c for the purposes of this International Standard. The risk borne by the supplier, that a shipment which *marginally* meets the specification be rejected by the recipient, will thus be $1 - p_c$. In some cases p_c is subject to prior agreement among the parties, and, like the specification limits and the test method, shall in that case be considered as an integral part of the specification.

I.2 Construction of specifications

This is described in 8.2.

I.3 Quality control against specifications

I.3.1 General

Clause I.3 provides general information to allow the supplier or the recipient to judge the quality of a product with regard to the specification when one or more results are available from one or more laboratories. If it is necessary for the recipient to take action after examining these results, the procedure specified in Clause I.4 shall be adopted.

The procedures of this subclause assume a test method which follows the normal probability distribution, with no bias, and with repeatability, r, and reproducibility, R. They also assume that a degree of criticality, p_c, has been agreed to in advance by supplier and recipient.

I.3.2 Testing margin for a single result

A supplier or a recipient who has no other source of information on the true value of a characteristic than a single result shall consider that the product meets the specification limit, with confidence 100 $(1 - p_c)$ %, if the result X is such that

in the case of a single upper limit A_1:

$$X \leqslant A_1 + 0{,}361(Z \times R) \qquad \text{(I.1)}$$

in the case of a single lower limit A_2,

$$X \geqslant A_2 - 0{,}361(Z \times R) \qquad \text{(I.2)}$$

and, in the case of a double limit (A_1 and A_2), both these conditions are satisfied.

The factor Z in these equations is the value of the standard normal distribution corresponding to a probability, p (see Table D.11). Note that for critical specifications ($p_c < 0{,}5$), Z has a negative value, and that the confidence $100\,(1 - p_c)\%$ that the product meets the specification limit is greater than for non-critical specifications. The factor 0,361 is the reciprocal of $1{,}96\sqrt{2}$ (see Equation H.2) used to convert reproducibility to a standard deviation.

If the reproducibility, R, is a function of the true value of the property in question, as it is in the example of Clauses 5 and 6, the value of R applied in Equation (I.1) is that which is appropriate for a true value of A_1, whereas for Equation (I.2), R shall be computed assuming the true value is A_2.

A recipient shall not take action, as specified in Clause I.4, on the basis of a single result. Instead, he shall obtain a minimum of at least three acceptable results (see 7.2.2) and compute the average. The test margin of I.3.3 shall then apply.

I.3.3 Test margin for several results from the same laboratory

A supplier or recipient who has obtained k acceptable test results from the product shall consider that the product meets specifications if the average of the test results, $\overline{X}$, satisfies Equation (I.1) or (I.2) or both, as appropriate, with R replaced by R_1, in accordance with Equation (18).

A recipient who has obtained a minimum of three acceptable test results is justified in taking action under Clause I.4 if the average of his results fails the test of this section.

(As in I.3.2, if R or r depend on the true value of the property considered, then they shall be computed as though the true value were A_1 in Equation (I.1) or A_2 in Equation (I.2).)

I.3.4 Test margin for results from several laboratories

In the event that two or more laboratories have independently obtained one or more acceptable results each from the same product, the mean, $\overline{X}$, of the individual laboratories' average test results shall be used to test for satisfaction of the specifications, provided that these averages pass the acceptability (reproducibility) test of 7.3.1. The supplier or recipient shall consider that the product meets specifications when either or both of Equation (I.1) and/or Equation (I.2) are satisfied, as appropriate, with R replaced by $R_4/\sqrt{N}$, where R_4 is as defined in Equation (23) of 7.3.1 and N is the number of laboratories involved.

[As above, if R or r depend on the true value of the property considered, then they shall be computed as though the true value were A_1 in Equation (I.1) or A_2 in Equation (I.2).]

I.4 Dispute procedure

I.4.1 If it is not possible for the supplier and the recipient to reach agreement about the quality of the product on the basis of their existing results, then the procedures given in I.4.2 to I.4.10 shall be adopted.

I.4.2 Each laboratory shall reject its original results and obtain at least three other acceptable results on the check sample to ensure that the work has been carried out under repeatability conditions. The average of the acceptable results in each laboratory shall then be computed, divergent results being discarded as indicated in 7.2.2. The supplier's and recipient's averages will be denoted $\overline{X}_S$ and $\overline{X}_R$, respectively.

I.4.3 If $\bar{X}_S$ and $\bar{X}_R$ are acceptable in terms of reproducibility (see 7.3.1), and if $\bar{X} = \left(\bar{X}_S + \bar{X}_R\right)/2$ satisfies Equation (I.1) or Equation (I.2) or both, as appropriate, with R replaced by R_2 of 7.3.1, then the product meets specification.

I.4.4 If $\bar{X}_S$ and $\bar{X}_R$ are acceptable in terms of reproducibility (see 7.3.1) and if $\bar{X} = \left(\bar{X}_S + \bar{X}_R\right)/2$ fails either Equation (I.1) or Equation (I.2), as appropriate, with R replaced by R_2 of 7.3.1, then the product fails specification.

I.4.5 In the event that the difference in laboratory means, $|\bar{X}_S - \bar{X}_R|$, exceeds R_2 of 7.3.1, and if the dispute cannot otherwise be settled at this point, the procedures of I.4.6 shall be applied.

I.4.6 In the case of unacceptable laboratory averages, the two laboratories shall contact each other and compare their operating procedures and apparatus. Following these investigations, a correlation test between the two laboratories shall be carried out on the two check samples. The average of at least three acceptable results shall be computed, in each laboratory, and these averages compared as given in I.4.2 to I.4.5.

I.4.7 If the disagreement remains, a third laboratory (neutral, expert and accepted by the two parties) shall be invited to carry out the test using a third sample. Suppose $\bar{X}_E$ is the average of the three or more acceptable results of the third laboratory. If the averages $\bar{X}_S$, $\bar{X}_R$ and $\bar{X}_E$ are acceptable in terms of reproducibility (see 7.3.1) then:

a) if $\bar{X} = \left(\bar{X}_S + \bar{X}_R + \bar{X}_E\right)/3$ satisfies Equation (I.1) or Equation (I.2) or both, as appropriate, with R replaced by $R_4/\sqrt{3}$, then the product meets specification (see Equation 23 with $N = 3$).

b) if $\bar{X} = \left(\bar{X}_S + \bar{X}_R + \bar{X}_E\right)/3$ fails either Equation (I.1) or Equation (I.2), as appropriate, with R replaced by $R_4/\sqrt{3}$, then the product fails specification.

I.4.8 If the averages $\bar{X}_S$, $\bar{X}_R$ and $\bar{X}_E$ are not acceptable in terms of reproducibility (see 7.3.1), then the most divergent laboratory average shall be rejected, and the average of the two remaining averages are denoted as $\bar{X}$. R_4 shall be recomputed based on the number of test results obtained by the remaining two laboratories, and becomes identical with R_2 (Equation 21).

I.4.9 If $\bar{X}$ satisfies Equation (I.1) or Equation (I.2) or both, as appropriate, with R replaced by $R_4/\sqrt{2}$, then the product meets specification.

I.4.10 If $\bar{X}$ fails either Equation (I.1) or Equation (I.2), as appropriate, with R replaced by $R_4/\sqrt{2}$, then the product fails specification.

Bibliography

[1] COCHRAN, W.G., *Ann. Eugen.,* vol. 11, 1941, pp. 47-52

[2] HAWKINS, D.M., *Identification of Outliers,* 1980, pp. 136-138

[3] KOLODZIEJCZYK, S., *Biometrika,* vol. 27, 1935, pp. 161-190

[4] WELCH, B.L., *Biometrika,* vol. 29, 1938, pp. 350-362

[5] DAVIES, O.L. *et al., Design and Analysis of Industrial Experiments,* second edition 1963, example 6B.1, pp. 236-238

[6] SMITH, I.J., *J. Inst. Pet.* 1963, 49, pp. 155-162

[7] MERRINGTON, M. and THOMPSON, C.M., *Biometrika,* vol. 33, 1943, pp. 73-88

[8] SHAPIRO, S.S. and WILK, M.B., *Biometrika,* vol. 52, 1965, pp. 591-611

[9] LAWREY, D.M.G., Software Program D2PP *Determination of Precision and Bias Data for use in Test Methods for Petroleum Products*[1)]

[10] NELDER, J.A. and MEAD, R., *A simplex method for function minimization*, Computer J. vol. 7, pp. 308-313

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

1) Available from ASTM International, 100 Barr Harbor Drive, West Conshohocken, PA 19428-2959, USA.

ICS 75.080

Price based on 62 pages