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)

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

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.