ERRATA

28

International Standard ISO 4259

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION•МЕЖДУНАРОДНАЯ ОРГАНИЗАЦИЯ ПО СТАНДАРТИЗАЦИИ•ORGANISATION INTERNATIONALE DE NORMALISATION

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

First edition — 1979-11-15

UDC 665.7 : 519.23 : 620.1

Ref. No. ISO 4259-1979 (E)

Descriptors : petroleum products, test results, reliability, statistical analysis, definitions

Price based on 27 pages

FOREWORD

ISO (the International Organization for Standardization) is a worldwide federation of national standards institutes (ISO member bodies). The work of developing International Standards is carried out through ISO technical committees. Every member body interested in a subject for which a technical committee has been set up 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.

Draft International Standards adopted by the technical committees are circulated to the member bodies for approval before their acceptance as International Standards by the ISO Council.

International Standard ISO 4259 was developed by Technical Committee ISO/TC 28, *Petroleum products and lubricants*, and was circulated to the member bodies in April 1977.

It has been approved by the member bodies of the following countries :

Australia	Hungary	Romania
Austria	India	South Africa, Rep. of
Brazil	Iran	Spain
Bulgaria	Israel	Turkey
Canada	Italy	United Kingdom
Finland	Japan	USA
France	Poland	USSR
Germany, F.R.	Portugal	

The member body of the following country expressed disapproval of the document on technical grounds :

Sweden

Printed in Switzerland

Published 1981-01-15

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION • МЕЖДУНАРОДНАЯ ОРГАНИЗАЦИЯ ПО СТАНДАРТИЗАЦИИ • ORGANISATION INTERNATIONALE DE NORMALISATION

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

ERRATUM

Page 1

In clause 1, paragraph 2, line 2, replace "tests" by "test".

Page 4

In sub-clause 4.1, paragraph 3, line 3, replace "weighed" by "weighted".

Page 6

In sub-clause 4.3, paragraph 2, line 3, replace "If" by "It".

Page 7

In sub-clause 5.1.2, formula (4), replace "SS_1" by "$S'S_1$".

In sub-clause 5.2, line 4, add "at" after "out".

Page 17

In clause C.3, in the expression for "W_j^2", replace "$\sum_{j=1}^{S}$" by "$\sum_{i=1}^{L}$".

CONTENTS

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

0 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 to be expected between two or more results obtained in specified circumstances.

1 SCOPE AND FIELD OF APPLICATION

This International Standard covers the calculation of precision estimates and their application to specifications. In particular, it contains definitions of relevant statistical terms (clause 2), the procedures to be adopted in the planning of an inter-laboratory test programme to determine the precision of a test method (clause 3), the method of calculating the precision from the results of such a programme (clauses 4 and 5), and the procedure to be followed in the interpretation of laboratory results in relation both to precision of the methods and to the limits laid down in specifications (clauses 6 to 9).

It must be emphasised that the procedures in this International Standard are designed to cover methods of tests for petroleum products only. The latter are, in general, homogeneous products with which serious sampling problems do not normally arise. It would not be appropriate, therefore, to consider the procedures to be necessarily of wider application, for example to heterogeneous solids.

2 DEFINITIONS

For the purposes of this International Standard, the following definitions apply :

2.1 analysis of variance : A technique which enables the total variance of a method to be broken down into its component factors.

2.2 between-laboratory variance : 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. (There is a corresponding definition for between-operator variance.)

2.3 bias : The difference between the true value (related to the method of test) (see 2.24) and the known value (see 2.8), where this is available.

2.4 blind coding : The assignment of a different number to each sample but not to repeats. No other identification or information on any sample is given to the operator.

2.5 check sample : A 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.

2.6 degrees of freedom : The 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.

2.7 determination : The process of carrying out the series of operations specified in the test method, whereby a single value is obtained.

2.8 known value : The 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.

2.9 mean; arithmetic mean; average : For a given set of results, the sum of the results divided by their number.

2.10 mean square : The sum of squares divided by the degrees of freedom.

2.11 normal distribution : The probability distribution of a continuous random variable X such that, if x is any real number, the probability density is

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right] \quad \ldots (1)$$

$$-\infty < x < +\infty$$

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

2.12 operator : A person who normally and regularly carries out a particular test.

2.13 outlier : A result far enough in magnitude from other results to be considered not a part of the set.

2.14 precision : The closeness of agreement between the results obtained by applying the experimental procedure several times on identical materials and under prescribed conditions. The smaller the random part of the experimental error, the more precise is the procedure.

2.15 random error : The chance variation encountered in all test work despite the closest control of variables.

2.16 recipient : Any individual or organization who receives or accepts the product delivered by the supplier.

2.17 repeatability :

a) *Qualitatively*

The closeness of agreement between successive results obtained in the normal and correct operation of the same method on identical test material, under the same conditions (same operator, same apparatus, same laboratory, and short intervals of time).

NOTE – The representative parameters of the dispersion of the population which may be associated with the results are qualified by the term "repeatability", for example repeatability standard deviation, repeatability variance.

b) *Quantitatively*

The value equal to or below which the absolute difference between two single test results obtained in the above conditions may be expected to lie with a specified probability; in the absence of other indication, the probability level is 95 %.

2.18 replication : The execution of a test method more than once so as to improve precision and to obtain a closer estimation of sampling error. Replication should be distinguished from repetition in that the former implies that experiments are carried out at one place and, as far as possible, one period of time.

2.19 reproducibility :

a) *Qualitatively*

The closeness of agreement between individual results obtained in the normal and correct operations of the same method on identical test material but under different conditions (different operators, different apparatus and different laboratories).

NOTE – The representative parameters of the dispersion of the population which may be associated with the results are qualified by the term "reproducibility", for example reproducibility standard deviation, reproducibility variance.

b) *Quantitatively*

The 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 specified probability; in the absence of other indication, the probability level is 95 %.

2.20 result : The final value obtained by following the complete set of instructions in the test method; it may be obtained from a single determination of from several determinations depending on the instructions in the method. (It is assumed that all the results are rounded off according to the procedure specified in annex G.)

2.21 standard deviation : A 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.

2.22 sum of squares : The sum of squares of the differences between a series of results and their mean.

2.23 supplier : Any individual or organization responsible for the quality of a product just before it is taken over by the recipient.

2.24 true value : For practical purposes, the value towards which the average of single results obtained by n laboratories tends, as n tends towards infinity; consequently, such a true value is associated with the particular method of test.

NOTE – A different and idealized definition is given in ISO 3534, *Statistics – Vocabulary and symbols.*

2.25 variance : The mean of the squares of the deviation of a random variable from its mean.

3 STAGES IN PLANNING OF AN INTER-LABORATORY TEST PROGRAMME FOR THE DETERMINATION OF THE PRECISION OF A TEST METHOD

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 two laboratories.

c) Planning the inter-laboratory programme.

d) Executing the inter-laboratory programme.

The four stages are described in turn.

3.1 Preparing a draft method of test

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

The clause on precision will be included at this stage only as a heading.

3.2 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 roughly the precision of the test.

At least two samples are required, covering the range of results to which the test is intended to apply; however, at least 12 laboratory/sample combinations should be included. Each sample is tested twice by each laboratory under repeatability conditions. If any omissions or inaccuracies in the draft 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 method shall be considered.

3.3 Planning the inter-laboratory programme

There shall be at least five participating laboratories, but it is preferable to exceed this number in order to reduce the number of samples required.

The number of samples shall be sufficient to cover the range of the property measured, and to give reliability to the precision estimates. If any variation of precision with level was observed in the results of the pilot programme, then at least five samples shall be used in the inter-laboratory programme. In any case, it is advisable to aim for 30 degrees of freedom in both repeatability and reproducibility. For repeatability, this means obtaining a total of 30 pairs of results in the programme. For reproducibility, table 11 (annex A) gives the 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 formula used. If Q is much larger than P, then 30 degrees of freedom cannot be achieved; the blank entries in table 11 correspond to this situation or the approach of it (i.e. when more than 20 samples are required). For these cases, there is likely to be a significant bias between laboratories.

3.4 Executing the inter-laboratory programme

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

The text of the method shall be distributed to all the laboratories in time to raise any queries before the tests begin. If any laboratory wants to practice the method in advance, this shall be done 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 instructions shall be sent with them :

a) the agreed draft method of test;

b) the handling and storing requirements for the samples;

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

d) the statement that two results are to be obtained consecutively on each sample by the same operator with the same apparatus;

e) the period of time during which all the samples are to be tested;

f) a 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) a 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.

NOTE — 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 should serve as a warning that the method is not satisfactory. They should be identified in the report of the results so that any effect may be noted.

4 INSPECTION OF INTER-LABORATORY RESULTS FOR UNIFORMITY AND FOR OUTLIERS

4.0 Introduction

This clause specifies procedures for examining the results reported in a statistically designed inter-laboratory programme (see clause 3) to establish

a) the independence of precision,

b) the level of the results,

c) the uniformity of precision from laboratory to laboratory,

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 of calculation of the bromine number set out in annex D.

Throughout this clause (and clause 5), the procedures to be 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 4.1). Other cases (which are rare) would require different treatment which is beyond the scope of this International Standard.

4.1 Transformation of data

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 should not be so and the position is rectified, if necessary, by a transformation.

The laboratories standard deviations D_j, (see annex C, clause C.3) are calculated and plotted against the sample means m_j. If the points so plotted may be considered as lying about a line parallel to the m-axis, then no transformation is necessary. If, however, the plotted points lie about a curve of the form $D = f(m)$, then a transformation will be necessary.

The relationship $D = f(m)$ is best established by the technique of univariate regression analysis (strictly speaking, an iteratively weighed regression should be used, but in most cases an unweighted regression gives a satisfactory approximation).

An outline of the calculation necessary is given in annex F, but it is a standard programme for most computers. Normally, a 5 % significance level will be used to test whether a regression coefficient differs from zero.

If it has been shown that there is a significantly non-zero regression coefficient giving a dependence of the form $D = f(m)$, then the appropriate transformation $y = F(x)$, where x is the reported result, is given by the formula

$$F(x) = k \int \frac{dx}{f(x)} \qquad \ldots (2)$$

where k is constant.

The particular cases likely to be encountered, together with the required transformations, are listed in table 20 (annex E). A regression of log D_j on log m_j will show any dependence of the form $D = Am^B$.

The choice of transformation is difficult to make the subject of formalized rules and qualified statistical assistance may be required in particular cases.

After selecting a transformation on the basis of the dependence of D on m, it shall be verified that the same transformation is also relevant for the repeats standard deviation d (see annex C, clause C.3). If it is not, then either a separate transformation will be necessary or the results will not need transforming for the calculation of repeatability.

4.1.1 *Worked example*

Table 1 lists the values of m, D, and d for the eight samples in the example given in annex D.

Inspection of the figures in table 1 shows than 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 the figure in annex F). The gradients of these lines are 0,64 and 0,58 respectively and thus, bearing in mind the errors in these estimated gradients, they may for convenience be considered as parallel lines with gradient 2/3.

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

$$\int x^{-2/3}\, dx = 3x^{1/3} \qquad \ldots (3)$$

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 16 (annex D), in which the cube roots are quoted correct to three decimal places.

4.2 Tests for outliers

After application of the appropriate transformation (or transformations) to the reported data, or if it has been decided that this is not necessary, the transformed results shall be inspected for outliers. These are the values which are so different from the remainder that it can only be concluded that they have arisen from some fault in the application of the method or from testing a wrong sample. Many possible tests may be used and the associated significance levels varied, but those that are specified in the following sub-clauses have been found to be appropriate in this International Standard.

4.2.1 *Uniformity of repeatability*

The first outlier test is concerned with detecting a discordant result in a pair of repeat results. This test[1] involves calculating the e_{ij}^2 over all the laboratory/sample

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	0,067	0,159	0,729	0,211	0,291	1,50	2,22	2,93
d	0,050 0	0,057 2	0,127	0,115	0,094 3	0,527	0,817	0,935

combinations. Cochran's criterion at the 1 % level is then used to test the ratio of the largest of these values over their sum (see annex C, clause C.4). If its value exceeds the value given in table 17 (annex D), corresponding to the 1 % probability level, k 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 k 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 will be necessary in this case.

4.2.2 *Worked example*

In the case of the example given in annex D, the difference between transformed repeat results, i.e. of the pairs of numbers in table 16, 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	7	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 the Cochran's criterion is

$$\frac{0{,}078^2}{0{,}043\,9} = 0{,}138$$

There are 72 ranges and, as from table 17 (annex D), the criterion for 75 ranges is 0,180 9, this ratio is not significant.

4.2.3 *Uniformity of reproducibility*

The remaining 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, one of the range of Dixon r tests[2] is appropriate. This involves forming for each sample, and finally for the laboratory totals (see 5.2), ratios of various differences between the pair sums a_{ij} (see annex C, clause C.5).

The appropriate ratio shall be compared with the critical 1 % values given in table 18 (annex D), with the value of n determined by the number of laboratories concerned. If 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 this laboratory shall be rejected.

4.2.3.1 WORKED EXAMPLE

The application of Dixon's test to sample 1 is shown in detail below. (See note.)

The first step is to place the pair sums for each laboratory which tested sample 1 in ascending order of magnitude, as shown in table 3.

The appropriate Dixon ratio for nine laboratories is r_{11}.

For testing the highest value,

$$r_{11} = \frac{3{,}188 - 2{,}562}{3{,}188 - 2{,}409}$$

$$= 0{,}804$$

This value is greater than the tabulated value and so the results from laboratory D on this sample are rejected.

As there has been a rejection, the procedure is repeated for high values without the results for laboratory D being taken into account. This gives

$$r_{11} = \frac{2{,}562 - 2{,}540}{2{,}562 - 2{,}409}$$

$$= 0{,}144$$

Comparison of this value with the corresponding value in table 18 (annex D), for eight laboratories shows that it is not significant and so there are no further outliers.

For testing the lowest value,

$$r_{11} = \frac{2{,}409 - 2{,}409}{2{,}562 - 2{,}409}$$

$$= 0$$

This value is compared with the corresponding value in table 18 (annex D), namely 0,677.

TABLE 3

Laboratory	B	F	C	H	E	A	G	J	D
Pair sum	2,409	2,409	2,432	2,476	2,497	2,520	2,540	2,562	3,188

As the calculated value is less than the one in table 18, there are no outliers at the low end.

This procedure is repeated for each sample. In this example there were no further significant ratios, and so the only rejections made were those for sample 1 obtained by laboratory D.

NOTE – If the two lowest values or the two highest values are equal, there can be no corresponding outlier.

4.3 Rejection of complete data from a sample

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. If is not possible to give an exact criterion for defining "excessively large" in this context, but it is felt that this action should be taken only in extreme cases[1].

NOTE – At this stage it is desirable to check that the rejections carried out have not invalidated the transformation used. If necessary, the procedure from 4.1 should be repeated with the outliers deleted.

4.3.1 *Worked example*

The laboratories 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.

Inspections shows that there is no outlying sample amongst these. It will be noted that the laboratories standard deviations are now independent of the sample means, which was the purpose of transforming the results. It was not considered necessary in this case to repeat the calculations with the outlier deleted.

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

It is clear, by inspection, that the laboratories standard deviation of sample 93 at 15,26 is far greater than the others, which are close together lying between 3,85 and 5,10, and so should be rejected. It is noted that the size of the repeats standard deviation in this sample also tends to confirm it as an outlier.

5 ANALYSIS OF VARIANCE AND CALCULATION OF PRECISION ESTIMATES

5.0 Introduction

After the data have been inspected for uniformity, a transformation has been performed if necessary, and any outliers have been rejected (see clause 4), an analysis shall be carried out. First the missing values shall be estimated by the least squares method, then an analysis of variance table constructed, and finally the precision estimates derived.

5.1 Estimating missing or rejected values

5.1.1 *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.

5.1.2 *Both repeat values missing or rejected*

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 from which all the results were rejected by Dixon's test shall be ignored and the new value of L 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.

Formula (4) may be used where only one pair sum has to be estimated. If more estimates are to be made, see, for instance, reference [5] for details.

TABLE 4

Sample number	3	8	1	4	5	6	2	7
Sample mean	0,910 1	1,066	1,240	1,538	2,217	3,639	4,028	4,851
Laboratories standard deviation	0,027 8	0,047 4	0,035 7	0,029 7	0,019 6	0,037 8	0,044 8	0,041 6

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	4,20	15,26	4,40	4,09	4,87	4,74	3,85
Repeats standard deviation	1,13	0,99	2,97	0,91	0,73	1,32	1,12	1,36

1) A test which may prove to be appropriate, but of which no experience is available in this context, is that which involves the ratio of the maximum to the minimum of a set of variances (at the 1 % level), as described in *Biometrika tables for statisticians*, volume 1, table 31.