SLOVENSKI STANDARD
SIST EN ISO 4259:1998

01-maj-1998

Naftni proizvodi - Določanje in uporaba stopenj natančnosti pri preskusnih metodah (ISO 4259:1992 in Popravek 1:1993

Petroleum products - Determination and application of precision data in relation to methods of test (ISO 4259:1992/Cor 1:1993)

Mineralölerzeugnisse - Bestimmung und Anwendung der Werte für die Präzision von Prüfverfahren (ISO 4259:1992/Cor 1:1993)

Produits pétroliers - Détermination et application des valeurs de fidélité relatives aux méthodes d'essai (ISO 4259:1992/Cor 1:1993)

Ta slovenski standard je istoveten z: **EN ISO 4259:1995**

ICS:

75.080	Naftni proizvodi na splošno	Petroleum products in general

SIST EN ISO 4259:1998 **en**

EUROPEAN STANDARD

NORME EUROPÉENNE

EUROPÄISCHE NORM

EN ISO 4259

August 1995

ICS 75.080

Descriptors: petroleum products, test results, accuracy, rules of calculation, reproducibility

English version

Petroleum products - Determination and application of precision data in relation to methods of test (ISO 4259:1992/Cor 1:1993)

Produits pétroliers - Détermination et application des valeurs de fidélité relatives aux méthodes d'essai (ISO 4259:1992/Cor 1:1993)

Mineralölerzeugnisse - Bestimmung und Anwendung der Werte für die Präzision von Prüfverfahren (ISO 4259:1992/Cor 1:1993)

This European Standard was approved by CEN on 1995-07-06. CEN members are bound to comply with the CEN/CENELEC Internal Regulations which stipulate the conditions for giving this European Standard the status of a national standard without any alteration.

Up-to-date lists and bibliographical references concerning such national standards may be obtained on application to the Central Secretariat or to any CEN member.

The European Standards exist in three official versions (English, French, German). A version in any other language made by translation under the responsibility of a CEN member into its own language and notified to the Central Secretariat has the same status as the official versions.

CEN members are the national standards bodies of Austria, Belgium, Denmark, Finland, France, Germany, Greece, Iceland, Ireland, Italy, Luxembourg, Netherlands, Norway, Portugal, Spain, Sweden, Switzerland and United Kingdom.

CEN

European Committee for Standardization
Comité Européen de Normalisation
Europäisches Komitee für Normung

Central Secretariat: rue de Stassart,36 B-1050 Brussels

Foreword

The text of the International Standard from ISO/TC 28 "Petroleum products and lubricants" of the International Organization for Standardization (ISO) has been taken over as a European Standard by the Technical Committee CEN/TC 19 "Petroleum products, lubricants and related products".

This European Standard shall be given the status of a National Standard, either by publication of an identical text or by endorsement, at the latest by February 1996, and conflicting national standards shall be withdrawn at the latest by February 1996.

According to CEN/CENELEC Internal Regulations, the following countries are bound to implement this European Standard: Austria, Belgium, Denmark, Finland, France, Germany, Greece, Iceland, Ireland, Italy, Luxembourg, Netherlands, Norway, Portugal, Spain, Sweden, Switzerland and the United Kingdom.

Endorsement notice

The text of the International Standard ISO 4259:1992/Cor 1:1993 has been approved by CEN as a European Standard without any modification.

INTERNATIONAL STANDARD

ISO 4259

Second edition
1992-12-15

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 4259:1992(E)

CONTENTS

Annexes

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

Foreword

ISO (the International Organization for Standardization) is a worldwide federation of national standards bodies (ISO member bodies). The work of preparing International Standards is normally carried out through ISO technical committees. Each member body interested in a subject for which a technical committee has been established has the right to be represented on that committee. International organizations, governmental and non-governmental, in liaison with ISO, also take part in the work. ISO collaborates closely with the International Electrotechnical Commission (IEC) on all matters of electrotechnical standardization.

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

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

This second edition cancels and replaces the first edition (ISO 4259:1979), of which it constitutes a technical revision.

Significant changes from the first edition include:

Independence of repeated results (subclause 3.17)
Blind coding in relation to repeated results (subclauses 3.4, 4.4)
Probability level for precision (subclauses 3.17, 3.19, 6.2.3)
Transformation procedure (subclause 5.1, annex E, annex F)
Hawkins' outlier test (subclauses 5.2.2, 5.5, clause C.5)
Rejection criteria for complete data from a sample (subclause 5.3)
Variance ratio test/bias between laboratories (subclauses 6.1.4, 6.2.3.2, clause C.6)
Formula for acceptability of results/confidence limits (clause 7, annex H)
Specifications which include a stated degree of criticality (subclause 8.1, annex J)

ISO 4259 makes reference to ISO 3534, *Statistics — Vocabulary and symbols,* which gives a different definition of "true value" (see subclause 3.24). ISO 4259 also refers to ISO 5725, *Precision of test methods — Determination of repeatability and reproducibility for a standard test method by inter-laboratory tests.* The latter will be required in particular and unusual circumstances (see subclause 5.1) for the purpose of estimating precision.

Annexes A, C, D, E, F and G form an integral part of this International Standard. Annexes B, H and J are for information only.

This page intentionally left blank

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

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

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 methods and to the limits laid down in specifications (clauses 7 to 10).

It is emphasised that the procedures in this International Standard are designed to cover methods of test 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 NORMATIVE REFERENCE

The following standard contains provisions which, through reference in this text, constitutes provisions of this International Standard. At the time of publication, the edition indicated was valid. All standards are subject to revision, and parties to agreements based on this International Standard are encouraged to investigate the possibility of applying the most recent edition of the standard listed below. Members of IEC and ISO maintain registers of currently valid International Standards.

ISO 5725: "Precision of Test Methods - Determination of Repeatability and Reproducibility for a Standard Test Method by Interlaboratory Tests".

3 DEFINITIONS

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

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

3.2 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.)

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 : The difference between the true value (related to the method of test) (see 3.24) and the known value (see 3.8), where this is available.

3.4 blind coding : The assignment of a different number to each sample. No other identification or information on any sample is given to the operator.

3.5 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.

3.6 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.

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

3.8 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.

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

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

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

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

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

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

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

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

3.14 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.

3.15 random error : The 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 :

a) *Qualitatively*

The 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).

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. The term "repeatability" shall not be confused with the terms "between repeats" or "repeats" when used in this way (see 3.18). Repeatability refers to the state of minimum random variability of results. The period of time during which repeated results are to be obtained shall therefore be short enough to exclude time - dependent errors, for example, environmental and calibration errors.

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 probability of 95 %.

3.18 replication : The execution of a test method more than once so as to improve precision and to obtain a better estimation of testing error. Replication shall be distinguished from repetition in that the former implies that repeated experiments are carried out at one place and, as far as possible, one period of time. The representative parameters of the dispersion of the population which may be associated with repeated experiments are qualified by the term "between repeats", or in shortened form "repeats", for example "repeats standard deviation".

3.19 reproducibility :

a) *Qualitatively*

The 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).

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 probability of 95 %.

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

3.21 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.

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

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

3.24 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.

3.25 variance : The mean of the squares of the deviation of a random variable from its mean, estimated by the mean square.

4 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 at least two laboratories.

c) Planning the inter-laboratory programme.

d) Executing the inter-laboratory programme.

The four stages are described in turn.

4.1 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.

4.2 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 shall 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.

4.3 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 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 11 (annex A) 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 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.

4.4 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 storage 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. 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) the 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) a 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) 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.

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

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

This clause specifies procedures for examining the results reported in a statistically designed inter-laboratory programme (see clause 4) to establish

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

b) 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 data (calculation of bromine number) set out in annex D.

Throughout this clause (and clause 6), 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 5.1). Other cases (which are rare) would require different treatment which 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 be used to store and analyse inter-laboratory test results, based on the procedures of this standard.

5.1 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 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 (see annex C) are calculated and plotted separately against the sample means m_j. If the points so plotted may 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 will be necessary.

The relationships $D = f_1(m)$ and $d = f_2(m)$ will not in general be 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 will take account of the difference between the relationships, if one exists, and will provide a means of testing for this difference (see annex F.1).

The single relationship $D = f(m)$ is best estimated by weighted linear regression analysis. Strictly speaking, an iteratively weighted regression should be used, but in most cases even an unweighted regression will give a satisfactory approximation. The derivation of weights is described in annex F.2, and the computational procedure for the regression analysis is described in annex F.3. Typical forms of dependence $D = f(m)$ are given in annex E.1. These are all expressed in terms of a single transformation parameter B.

The estimation of B, and the transformation procedure which follows, are summarised in annex 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 methods of ISO 5725 shall be used. In such an event it will not be possible to test for laboratory bias over all samples (clause 5.5) or separately estimate the interaction component of variance (clause 6.1).

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 formula

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

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 annex E.1.

The choice of transformation is difficult to make the subject of formalized rules. Qualified statistical assistance may be required in particular cases. The presence of outliers may affect judgement as to the type of transformation required, if any (see 5.6).

5.1.1 *Worked example*

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

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 in annex F). From the example calculations given in annex 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 formula

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

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.

5.2 Tests for outliers

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 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. These outlier tests all assume a normal distribution of errors (see 5.).

5.2.1 *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 combinations. Cochran's criterion at the 1 % significance 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 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 will be necessary in this case.

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	0,0669 (14)	0,159 (9)	0,729 (8)	0,211 (11)	0,291 (9)	1,50 (9)	2,22 (9)	2,93 (9)
d	0,0500(9)	0,0572(9)	0,127 (9)	0,116 (9)	0,0943(9)	0,527 (9)	0,818 (9)	0,935 (9)

5.2.1.1 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 16, in units of the third decimal place, are shown in table 2.

TABLE 2

	Sample							
Laboratory	**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

$$0{,}042^2 + 0{,}021^2 + \ldots + 0{,}026^2 + 0^2 = 0{,}0439.$$

Thus, the ratio to be compared with Cochran's criterion is

$$\frac{0{,}078^2}{0{,}0439} = 0{,}138.$$

There are 72 ranges and, as from table 17 (annex D), the criterion for 80 ranges is 0,1709, this ratio is not significant.

5.2.2 *Uniformity of reproducibility*

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 5.5), 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 annex C.5).

The ratio corresponding to the largest absolute deviation shall be compared with the critical 1% values given in table 18 (annex D), 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 will refer to other samples, but will be 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 will be necessary in this case.

5.2.2.1 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

	Sample							
Laboratory	**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 to be 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 :

$$B^* = \frac{0{,}314}{\sqrt{0{,}117 + 0{,}015 \ldots + 0{,}017}} = 0{,}7281.$$

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 18 (annex D) as 0,3729. 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 will be that obtained by laboratory F from sample 2. The Hawkins' test ratio for this cell is :

$$B^* = \frac{0{,}097}{\sqrt{0{,}006 + 0{,}015 + \ldots + 0{,}017}} = 0{,}3542.$$

The critical value corresponding to n = 9 cells in sample 2 and $\nu = 55$ extra degrees of freedom is interpolated from table 18 (annex D) as 0,3756. As the test ratio is less than the critical value there will be no further rejections.

5.3 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.

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 annex C, clause C.4). If the ratio exceeds the critical value given in table 17 (annex D), 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 5.1), 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 annex C, clause C.6). Here ν_1 is the degrees of freedom of the variance in question and ν_2 is the degrees of freedom from the remaining samples. If the ratio is greater than the critical value given in table 20 (annex D), 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.

5.3.1 *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 will be 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, 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. It is noted that the repeats standard deviation in this sample is correspondingly large.

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

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

The variance ratio is then calculated as

$$(15{,}26)^2 / 19{,}96 = 11{,}66.$$

From table 20 (annex D) the critical value corresponding to a significance level of 0,01/8 = 0,00125, on 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 will be the ratio of the largest sum of squares (sample 93) to the sum of all the sums of squares, that is

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

This is greater than the critical value of 0,352 corresponding to *n* = 8 and ν = 8 (see table 17, annex D), and confirms that results from sample 93 shall be rejected.

5.4 Estimating missing or rejected values

5.4.1 *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.

5.4.2 *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 x 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.

Formula (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 formula (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] for details.

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	0,0278(14)	0,0473(9)	0,0354(13)	0,0297(11)	0,0197(9)	0,0378(9)	0,0450(9)	0,0416(9)
Repeats standard deviation	0,0214(9)	0,0182(9)	0,0281(8)	0,0164(9)	0,0063(9)	0,0132(9)	0,0166(9)	0,0130(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)