INTERNATIONAL STANDARD

**ISO
5725-6**

First edition
1994-12-15

Accuracy (trueness and precision) of measurement methods and results —

Part 6:
Use in practice of accuracy values

Exactitude (justesse et fidélité) des résultats et méthodes de mesure —

Partie 6: Utilisation dans la pratique des valeurs d'exactitude

ISO

Reference number
ISO 5725-6:1994(E)

ISO 5725-6:1994(E)

Contents

© ISO

ISO 5725-6:1994(E)

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 5725-6 was prepared by Technical Committee ISO/TC 69, *Applications of statistical methods*, Subcommittee SC 6, *Measurement methods and results*.

ISO 5725 consists of the following parts, under the general title *Accuracy (trueness and precision) of measurement methods and results*:

— *Part 1: General principles and definitions*

— *Part 2: Basic method for the determination of repeatability and reproducibility of a standard measurement method*

— *Part 3: Intermediate measures of the precision of a standard measurement method*

— *Part 4: Basic methods for the determination of the trueness of a standard measurement method*

— *Part 5: Alternative methods for the determination of the precision of a standard measurement method*

— *Part 6: Use in practice of accuracy values*

Parts 1 to 6 of ISO 5725 together cancel and replace ISO 5725:1986, which has been extended to cover trueness (in addition to precision) and intermediate precision conditions (in addition to repeatability and reproducibility conditions).

Annex A forms an integral part of this part of ISO 5725.

© ISO   **ISO 5725-6:1994(E)**

Introduction

0.1   ISO 5725 uses two terms "trueness" and "precision" to describe the accuracy of a measurement method. "Trueness" refers to the closeness of agreement between the arithmetic mean of a large number of test results and the true or accepted reference value. "Precision" refers to the closeness of agreement between test results.

0.2   The need to consider "precision" arises because tests performed on presumably identical materials in presumably identical circumstances do not, in general, yield identical results. This is attributed to unavoidable random errors inherent in every measurement procedure; the factors that influence the outcome of a measurement cannot all be completely controlled. In the practical interpretation of measurement data, this variability has to be taken into account. For instance, the difference between a test result and some specified value may be within the scope of unavoidable random errors, in which case a real deviation from such a specified value has not been established. Similarly, comparing test results from two batches of material will not indicate a fundamental quality difference if the difference between them can be attributed to the inherent variation in the measurement procedure.

0.3   Parts 1 to 5 of ISO 5725 dicuss the background to, and given methods for, the assessment of the precision (in terms of the repeatability standard deviation and the reproducibility standard deviation) and the trueness (in terms of the various components of bias) of measurements produced by a standard measurement method. Such assessment would, however, be pointless if there were no practical uses to which the results could be put.

0.4   Given that the accuracy of a measurement method has been established, this part of ISO 5725 applies that knowledge in practical situations in such a way as to facilitate commercial transactions and to monitor and improve the operational performance of laboratories.

Document provided by IHS Licensee=Shell Services International B.V./5924979112,
User=, 03/09/2003 22:24:02 MST Questions or comments about this message: please
call the Document Policy Management Group at 1-800-451-1584.

v

Accuracy (trueness and precision) of measurement methods and results —

Part 6:
Use in practice of accuracy values

1 Scope

1.1 The purpose of this part of ISO 5725 is to give some indications of the way in which accuracy data can be used in various practical situations by:

a) giving a standard method of calculating the repeatability limit, the reproducibility limit and other limits to be used in examining the test results obtained by a standard measurement method;

b) providing a way of checking the acceptability of test results obtained under repeatability or reproducibility conditions;

c) describing how to assess the stability of results within a laboratory over a period of time, and thus providing a method of "quality control" of the operations within that laboratory;

d) describing how to assess whether a given laboratory is able to use a given standard measurement method in a satisfactory way;

e) describing how to compare alternative measurement methods.

1.2 This part of ISO 5725 is concerned exclusively with measurement methods which yield measurements on a continuous scale and give a single numerical figure as the result, although this single figure may be the outcome of a calculation from a set of observations.

1.3 It is assumed that the estimates of trueness and precision for the method have been obtained in accordance with parts 1 to 5 of ISO 5725.

1.4 Any additional information regarding the field of application will be given at the beginning of each particular application.

2 Normative references

The following standards contain provisions which, through reference in this text, constitute provisions of this part of ISO 5725. At the time of publication, the editions indicated were valid. All standards are subject to revision, and parties to agreements based on this part of ISO 5725 are encouraged to investigate the possibility of applying the most recent editions of the standards indicated below. Members of IEC and ISO maintain registers of currently valid International Standards.

ISO 3534-1:1993, *Statistics — Vocabulary and symbols — Part 1: Probability and general statistical terms.*

ISO 5725-1:1994, *Accuracy (trueness and precision) of measurement methods and results — Part 1: General principles and definitions.*

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.*

ISO 5725-3:1994, *Accuracy (trueness and precision) of measurement methods and results — Part 3: Intermediate measures of the precision of a standard measurement method.*

ISO 5725-4:1994, *Accuracy (trueness and precision) of measurement methods and results — Part 4: Basic methods for the determination of the trueness of a standard measurement method.*

ISO 8258:1991, *Shewhart control charts.*

ISO Guide 33:1989, *Uses of certified reference materials.*

ISO Guide 35:1989, *Certification of reference materials — General and statistical principles.*

ISO/IEC Guide 25:1990, *General requirements for the competence of calibration and testing laboratories.*

3 Definitions

For the purposes of this part of ISO 5725, the definitions given in ISO 3534-1 and ISO 5725-1 apply.

The symbols used in ISO 5725 are given in annex A.

4 Determination of limits

4.1 Repeatability and reproducibility limits

4.1.1 In ISO 5725-2, attention has been focussed on estimating the standard deviations associated with operations under repeatability or reproducibility conditions. However, normal laboratory practice requires examination of the difference(s) observed between two (or more) test results, and for this purpose some measure akin to a critical difference is required, rather than a standard deviation.

4.1.2 When a quantity is based on sums or differences of n independent estimates each having a standard deviation σ, then that resultant quantity will have a standard deviation $\sigma\sqrt{n}$. The reproducibility limit (R) or repeatability limit (r) are for differences between two test results, so the associated standard deviation is $\sigma\sqrt{2}$. In normal statistical practice, for examining the difference between these two values the critical difference used is f times this standard deviation, i.e. $f\sigma\sqrt{2}$. The value of f (the critical range factor) depends on the probability level to be associated with the critical difference and on the shape of the underlying distribution. For the reproducibility and repeatability limits, the probability level is specified as 95 %, and throughout the analysis in ISO 5725 the assumption is made that the underlying distribution is approximately normal. For a normal distribution at 95 % probability level, f is 1,96 and $f\sqrt{2}$ then is 2,77. As the purpose of this part of ISO 5725 is to give some simple "rule of thumb" to be applied by non-statisticians when examining the results of tests, it seems reasonable to use a rounded value of 2,8 instead of $f\sqrt{2}$.

4.1.3 As has been stated, the process of estimating precision leads to estimates of the true standard deviations while the true standard deviations remain unknown. Therefore in statistical practice they should be denoted by s rather than σ. However, if the procedures given in ISO 5725-1 and ISO 5725-2 are followed, these estimates will be based on an appreciable number of test results, and will give the best information we are likely to have of the true values of the standard deviations. In other applications that follow, for estimates of these standard deviations based on more limited data, the symbol s (estimate of a standard deviation) is used. Therefore it seems best to use the symbol σ to denote the values obtained from a full precision experiment, and treat these as true standard deviations with which other estimates (s) will be compared.

4.1.4 In view of 4.1.1 to 4.1.3, when examining two single test results obtained under repeatability or reproducibility conditions, the comparison shall be made with the repeatability limit

$$r = 2,8\sigma_r$$

or the reproducibility limit

$$R = 2,8\sigma_R$$

4.2 Comparisons based on more than two values

4.2.1 Two groups of measurements in one laboratory

If, in one laboratory under repeatability conditions, two groups of measurements are performed with the first group of n_1 test results giving an arithmetic mean of $\bar{y}_1$ and the second group of n_2 test results giving an arithmetic mean of $\bar{y}_2$, then the standard deviation of $(\bar{y}_1 - \bar{y}_2)$ is

$$\sigma = \sqrt{\sigma_r^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}$$

and the critical difference for $|\bar{y}_1 - \bar{y}_2|$ is

$$CD = 2{,}8\sigma_r \sqrt{\frac{1}{2n_1} + \frac{1}{2n_2}}$$

at the 95 % probability level.

NOTE 1 If n_1 and n_2 are both unity, this reduces to $r = 2{,}8\sigma_r$, as expected.

4.2.2 Two groups of measurements in two laboratories

If the first laboratory obtains n_1 test results giving an arithmetic mean of $\bar{y}_1$ while the second laboratory obtains n_2 test results giving an arithmetic mean of $\bar{y}_2$, in each case under repeatability conditions, then the standard deviation of $(\bar{y}_1 - \bar{y}_2)$ is

$$\sigma = \sqrt{\sigma_L^2 + \frac{1}{n_1}\sigma_r^2 + \sigma_L^2 + \frac{1}{n_2}\sigma_r^2}$$

$$= \sqrt{2\sigma_L^2 + \sigma_r^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}$$

$$= \sqrt{2\left(\sigma_L^2 + \sigma_r^2\right) - 2\sigma_r^2 \left(1 - \frac{1}{2n_1} - \frac{1}{2n_2} \right)}$$

and the critical difference for $|\bar{y}_1 - \bar{y}_2|$ is

$$CD = \sqrt{(2{,}8\sigma_R)^2 - (2{,}8\sigma_r)^2 \left(1 - \frac{1}{2n_1} - \frac{1}{2n_2} \right)}$$

at the 95 % probability level.

NOTE 2 If n_1 and n_2 are both unity, this reduces to $R = 2{,}8\sigma_R$, as expected.

4.2.3 Comparison with a reference value for one laboratory

If n test results are obtained under repeatability conditions within one laboratory which give an arithmetic mean of $\bar{y}$, then the comparison with a given reference value μ_0 shall be made, in the absence of specific knowledge of the laboratory component of bias, using a standard deviation for $(\bar{y} - \mu_0)$ of

$$\sigma = \sqrt{\sigma_L^2 + \frac{1}{n}\sigma_r^2}$$

$$= \frac{1}{\sqrt{2}}\sqrt{2\left(\sigma_L^2 + \sigma_r^2\right) - 2\sigma_r^2 \left(1 - \frac{1}{n} \right)}$$

$$= \frac{1}{\sqrt{2}}\sqrt{2\left(\sigma_L^2 + \sigma_r^2\right) - 2\sigma_r^2 \left(\frac{n-1}{n} \right)}$$

and the critical difference for $|\bar{y} - \mu_0|$ is

$$CD = \frac{1}{\sqrt{2}}\sqrt{(2{,}8\sigma_R)^2 - (2{,}8\sigma_r)^2 \left(\frac{n-1}{n} \right)}$$

4.2.4 Comparison with a reference value for more than one laboratory

If p laboratories have obtained n_i test results giving arithmetic means of $\bar{y}_i$ (in each case under repeatability conditions) and the grand mean $\bar{\bar{y}}$ is computed by

$$\bar{\bar{y}} = \frac{1}{p}\sum \bar{y}_i$$

and this grand mean is to be compared with a reference value μ_0, then the standard deviation for $(\bar{\bar{y}} - \mu_0)$ is

$$\sigma = \sqrt{\frac{1}{p}\sigma_L^2 + \frac{1}{p^2}\sigma_r^2 \sum \frac{1}{n_i}}$$

$$= \frac{1}{\sqrt{2p}}\sqrt{2\left(\sigma_L^2 + \sigma_r^2\right) - 2\sigma_r^2 + \frac{2\sigma_r^2}{p}\sum \frac{1}{n_i}}$$

$$= \frac{1}{\sqrt{2p}} \sqrt{2\left(\sigma_L^2 + \sigma_r^2\right) - 2\sigma_r^2\left(1 - \frac{1}{p}\sum\frac{1}{n_i}\right)}$$

and the critical difference for $|\bar{\bar{y}} - \mu_0|$ is

$$CD = \frac{1}{\sqrt{2p}} \sqrt{(2{,}8\sigma_R)^2 - (2{,}8\sigma_r)^2\left(1 - \frac{1}{p}\sum\frac{1}{n_i}\right)}$$

at the 95 % probability level.

4.2.5 Quoting the results of a comparison

When the absolute difference exceeds the appropriate limit as given in the preceding clauses, then the difference shall be considered as suspect, and therefore all measurements that have given rise to this difference shall be considered as suspect and subject to further investigation.

5 Methods for checking the acceptability of test results and determining the final quoted result

5.1 General

5.1.1 The checking method described in this clause should be applied only to the case where the measurement was carried out according to a measurement method which has been standardized and whose standard deviations σ_r and σ_R are known. Therefore, when the range of N test results exceeds the appropriate limit as given in clause 4, it is considered that one, two or all of the N test results is or are aberrant. It is recommended that the cause of the aberrant result(s) should be investigated from the technical point of view. However, it may be necessary for commercial reasons to obtain some acceptable value, and in such cases the test results shall be treated according to the stipulations of this clause.

5.1.2 This clause has been prepared on the assumptions that the test results were obtained under repeatability and reproducibility conditions, and that the probability level to be used is 95 %. If intermediate conditions (see ISO 5725-3) were in force, then it is necessary to replace σ_r by the appropriate intermediate measure.

5.1.3 In some cases where the procedures described in 5.2 lead to the median being quoted as the final result, it might be better to abandon the data.

5.2 Methods for checking the acceptability of test results obtained under repeatability conditions

NOTE 3 In 5.2.2.1 and 5.2.2.2, reference made to measurements being expensive or inexpensive should be interpreted not only in financial terms but also whether the measurement is complex, troublesome or time-consuming.

5.2.1 Single test result

It is not common in commercial practice to obtain only one test result. When only one test result is obtained, it is not possible to make an immediate statistical test of the acceptability of that test result with respect to the given repeatability measure. If there is any suspicion that the test result may not be correct, a second test result should be obtained. Availability of two test results leads to the more common practice which is described below.

5.2.2 Two test results

The two test results should be obtained under repeatability conditions. The absolute difference between the two test results should then be compared with the repeatability limit $r = 2{,}8\sigma_r$.

5.2.2.1 Case where obtaining test results is inexpensive

If the absolute difference between the two test results does not exceed r, then both test results are considered acceptable, and the final quoted result should be quoted as the arithmetic mean of the two test results. If the absolute difference does exceed r, the laboratory should obtain two further test results.

If the range $(x_{max} - x_{min})$ of the four test results is equal to or less than the critical range at the 95 % probability level for $n = 4$, $CR_{0,95}(4)$, the arithmetic mean of the four test results should be reported as the final quoted result. Critical range factors, $f(n)$, for $n = 2$ to $n = 40$ and selected values from $n = 45$ to $n = 100$ are given in table 1 to be used to calculate the critical range according to the following equation:

$$CR_{0,95}(n) = f(n)\sigma_r$$

4

If the range of the four test results is greater than the critical range for $n = 4$, the median of the four test results should be reported as the final quoted result.

This procedure is summarized in the flowchart given in figure 1.

5.2.2.2 Case where obtaining test results is expensive

If the absolute difference between the two test results does not exceed r, then both test results are considered acceptable, and the final quoted result should be quoted as the arithmetic mean of the two test results. If the absolute difference does exceed r, the laboratory should obtain a further test result.

If the range $(x_{max} - x_{min})$ of the three test results is equal to or less than the critical range for $n = 3$, $CR_{0,95}(3)$, the arithmetic mean of the three test results should be reported as the final quoted result.

If the range of the three test results is greater than the critical range for $n = 3$, a decision on one of the following two cases shall be made.

a) **Case where it is impossible to obtain a fourth test result:**

The laboratory should use the median of the three test results as the final quoted result.

This procedure is summarized in the flowchart given in figure 2.

b) **Case where it is possible to obtain a fourth test result:**

The laboratory should obtain the fourth test result. If the range $(x_{max} - x_{min})$ of the four test results is equal to or less than the critical range for $n = 4$,

$CR_{0,95}(4)$, the arithmetic mean of the four test results should be reported as the final quoted result. If the range of the four test results is greater than the critical range for $n = 4$, the laboratory should use the median of the four test results as the final quoted result.

This procedure is summarized in the flowchart given in figure 3.

Table 1 — Critical range factors, $f(n)$

n	$f(n)$	n	$f(n)$
2	2,8	25	5,2
3	3,3	26	5,2
4	3,6	27	5,2
5	3,9	28	5,3
6	4,0	29	5,3
7	4,2	30	5,3
8	4,3	31	5,3
9	4,4	32	5,3
10	4,5	33	5,4
11	4,6	34	5,4
12	4,6	35	5,4
13	4,7	36	5,4
14	4,7	37	5,4
15	4,8	38	5,5
16	4,8	39	5,5
17	4,9	40	5,5
18	4,9	45	5,6
19	5,0	50	5,6
20	5,0	60	5,8
21	5,0	70	5,9
22	5,1	80	5,9
23	5,1	90	6,0
24	5,1	100	6,1

NOTE — The critical range factor $f(n)$ is the 95 % quantile of the distribution of $(x_{max} - x_{min})/\sigma$ where x_{max} and x_{min} are the extreme values in a sample of size n from a normal distribution with standard deviation σ.

Start with two results

$$|x_1 - x_2| \leqslant r$$

Yes → $\dfrac{x_1 + x_2}{2}$ is the final quoted result

No

Obtain two further results

$$x_{max} - x_{min} \leqslant CR_{0.95}(4)$$

Yes → $\dfrac{x_1 + x_2 + x_3 + x_4}{4}$ is the final quoted result

No

$\dfrac{x_{(2)} + x_{(3)}}{2}$ is the final quoted result
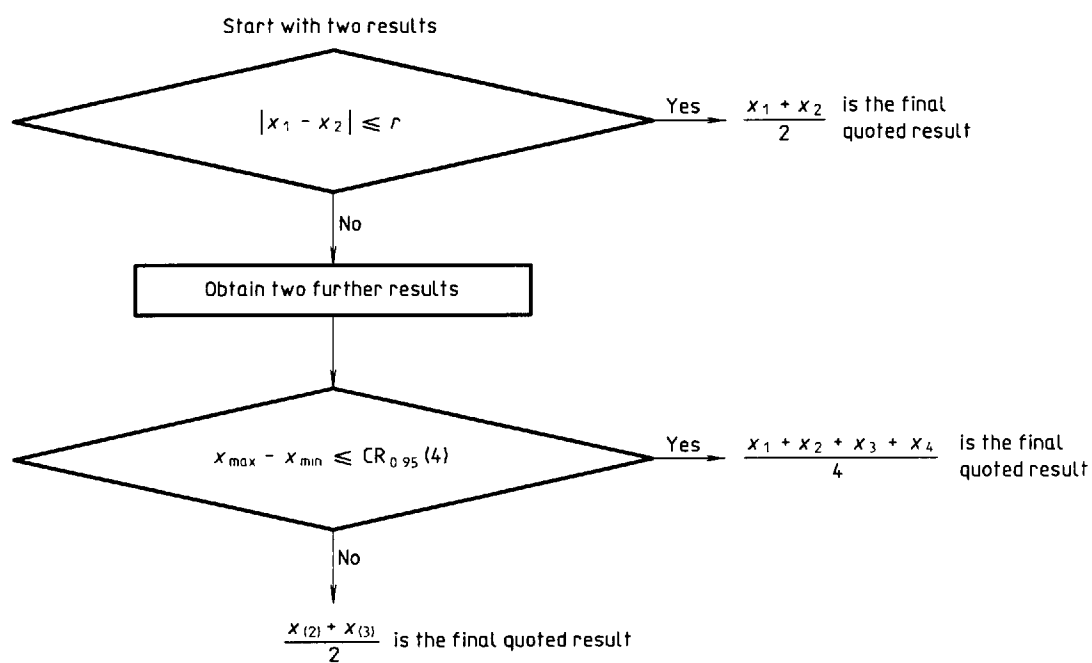
where

$x_{(2)}$ is the second smallest result

$x_{(3)}$ is the third smallest result

Figure 1 — Method for checking the acceptability of test results, obtained under repeatability conditions, when two test results are obtained to start with and obtaining test results is inexpensive: Case 5.2.2.1

Start with two results

where

$x_{(2)}$ is the second smallest result

Figure 2 — Method for checking the acceptability of test results, obtained under repeatability conditions, when two test results are obtained to start with and obtaining test results is expensive: Case 5.2.2.2 a)
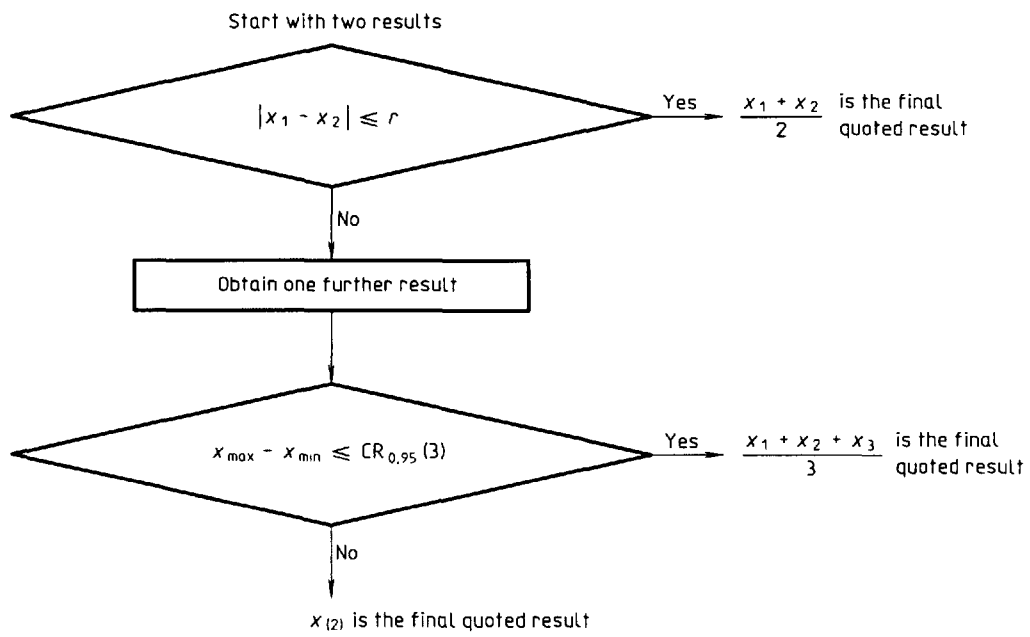
Start with two results

where

$x_{(2)}$ is the second smallest result

$x_{(3)}$ is the third smallest result

Figure 3 — Method for checking the acceptability of test results, obtained under repeatability conditions, when two test results are obtained to start with and obtaining test results is expensive: Case 5.2.2.2 b)

5.2.3 More than two test results to start with

It is sometimes practical to start with more than two test results. The method for obtaining the final quoted result under repeatability conditions for the cases where $n > 2$ is similar to the case for $n = 2$.

The range $(x_{max} - x_{min})$ of the test results is compared with the critical range $CR_{0,95}(n)$ calculated from table 1 for the appropriate value of n. If the range does not exceed the critical range, then the arithmetic mean of all the n test results is used as the final quoted result.

If the range does exceed the critical range $CR_{0,95}(n)$, then a decision on one of the cases A, B or C given in figures 4 to 6 shall be made to obtain the final quoted result.

Cases A and B correspond to the cases where obtaining test results is inexpensive and expensive, respectively. Case C is an alternative which is recommended when the starting number of test results is five or more and where obtaining each test result is inexpensive, or when the starting number of test results is four or more and where obtaining each test result is expensive.

For inexpensive measurements, the difference between case A and case C is that case A requires n further measurements, whereas case C requires less than half that number of further measurements. The decision will depend on the size of n and the ease of performing the measurements.

For expensive measurements, the difference between case B and case C is that case C requires further measurements, whereas in case B no further measurements are carried out. Case B shall only be considered where the performance of further measurements is so expensive as to be prohibitive.
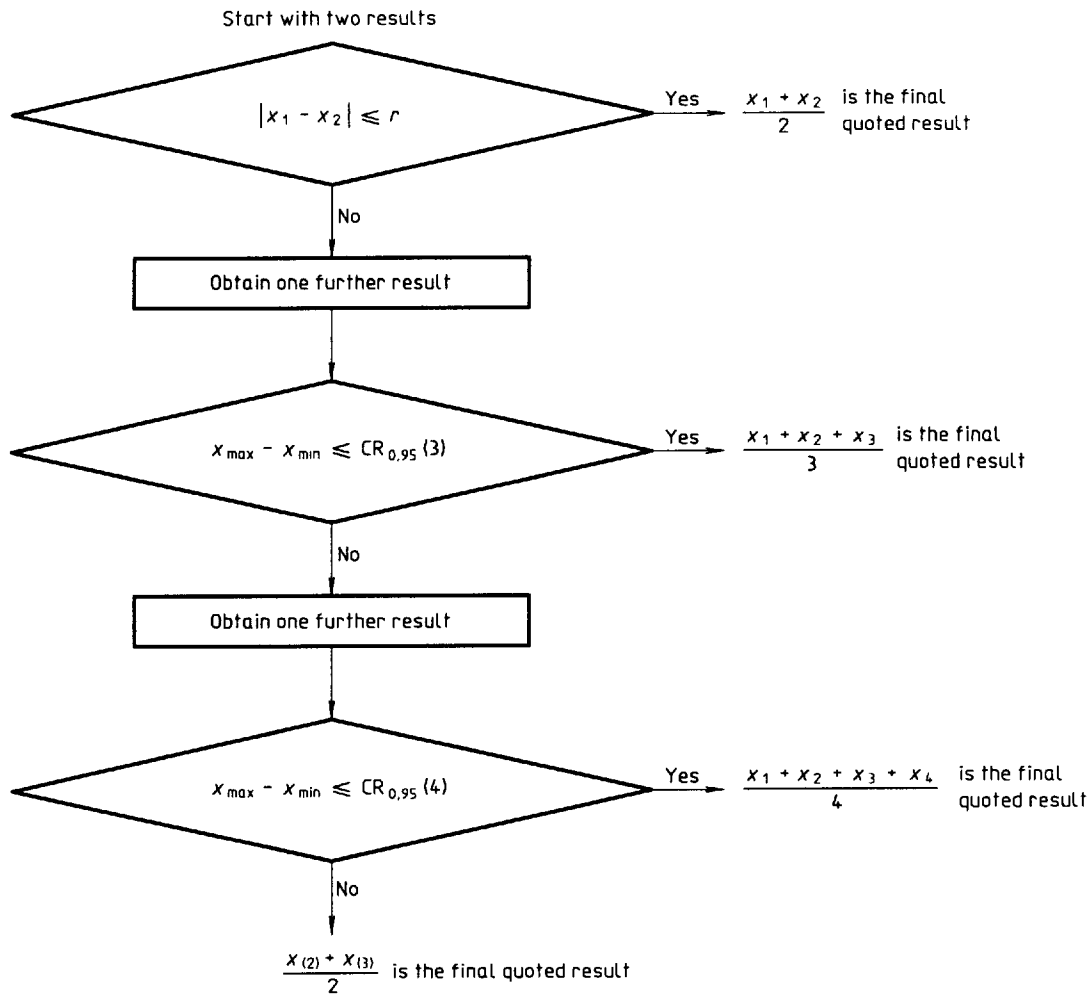

Figure 4 — Method for checking the acceptability of test results, obtained under repeatability conditions, when n test results are obtained to start with and obtaining test results is inexpensive: Case A

■■ 4851903 0594633 693 ■■

Start with n results

Range of n results $\leqslant CR_{0.95}(n)$ — Yes → Arithmetic mean of all n results is the final quoted result

No ↓

Median of all n results is the final quoted result

Figure 5 — Method for checking the acceptability of test results, obtained under repeatability conditions, when n test results are obtained to start with and obtaining test results is expensive: Case B

Start with n results

Range of n results $\leqslant CR_{0.95}(n)$ — Yes → Arithmetic mean of all n results is the final quoted result

No ↓

Obtain $m^{1)}$ further results

↓

Range of $(n + m)$ results $\leqslant CR_{0.95}(n + m)$ — Yes → Arithmetic mean of all $(n + m)$ results is the final quoted result

No ↓

Median of all $(n + m)$ results is the final quoted result

1) m has to be chosen as an integer satisfying the condition
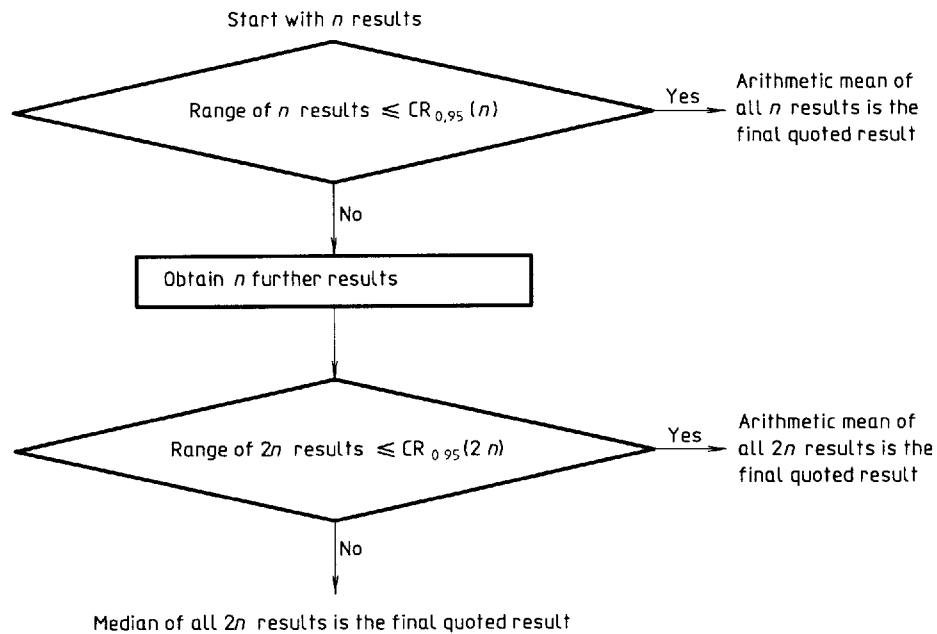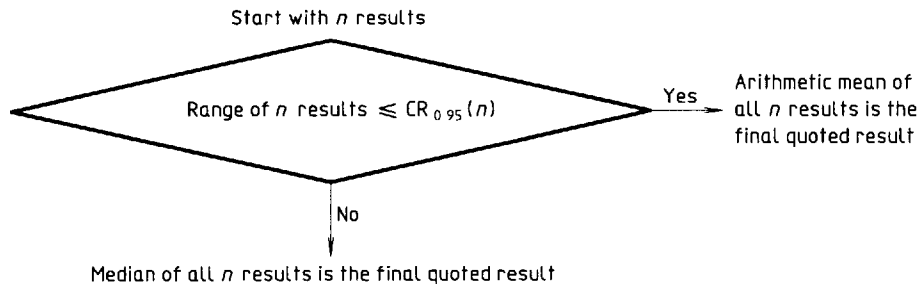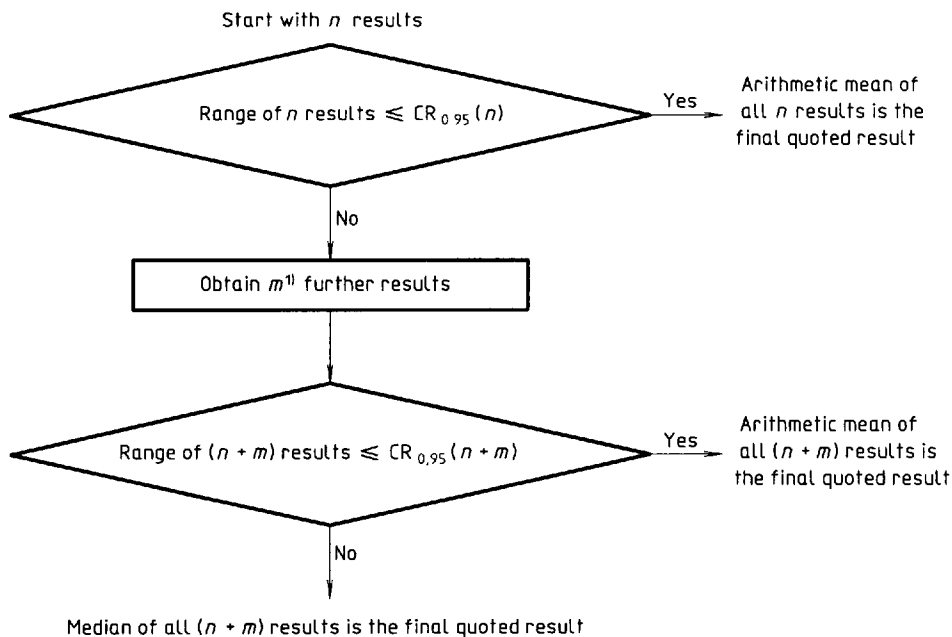$n/3 \leqslant m \leqslant n/2$

Figure 6 — Method for checking the acceptability of test results, obtained under repeatability conditions, when $n \geqslant 5$ and obtaining test results is inexpensive, or $n \geqslant 4$ and obtaining test results is expensive: Case C

5.2.4 Example of case B: An expensive chemical analysis

Expensive cases are often found in chemical analyses which consist of complicated and time-consuming procedures, requiring 2 or 3 days or more for one analysis. In such a case, it is troublesome and expensive to carry out a re-analysis when a technically questionable datum or an outlier is found in the first analysis, therefore usually three or four test results are obtained under repeatability conditions from the beginning, and test-processed according to case B. See figure 5.

For example, in the determination of gold and silver in ores by fire assay, although there are several methods, all of them require expensive specific equipment, highly skilled operators and a long time, usually about 2 days, for completing the entire processes and even more if the ore contains platinum-group metals or other specific co-existing elements.

The following four test results for gold content were obtained on a copper concentrate under repeatability conditions:

Au (in g/t):   11,0   11,0   10,8   10,5

These test results are processed according to method B.

The method for determination of gold and silver has not been established in an International Standard, however when a value of

$$\sigma_r = 0,12 \text{ g/t}$$

is given for the determination of gold,

$$CR_{0,95}(4) = 3,6 \times 0,12 = 0,43 \text{ g/t}$$

according to table 1, where $f(4) = 3,6$.

Since the range of the above four test results is

$$11,0 - 10,5 = 0,5 \text{ g/t}$$

which is greater than $CR_{0,95}(4)$, the final quoted result is the median of the four test results, i.e.

$$\frac{11,0 + 10,8}{2} = 10,9 \text{ g/t}$$

5.2.5 Note regarding precision experiment

If the procedures given in 5.2.2 or 5.2.3 result frequently in values exceeding the critical values, the precision of the measurement method for this lab-oratory and/or precision experiment should be investigated.

5.2.6 Reporting the final quoted result

If only the final quoted test result is presented, both of the following points should be specified:

— the number of test results used for the computation of the final quoted result; and

— whether the arithmetic mean or the median of the test results was taken.

5.3 Methods for checking the acceptability of test results obtained under both repeatability and reproducibility conditions

5.3.1 General

These methods cover the case where two laboratories obtain test results and there is some difference in the test results or in the arithmetic means of the test results. The reproducibility standard deviation becomes part of the statistical testing procedure as well as the repeatability standard deviation.

In all cases of obtaining test results on test samples, sufficient material should be provided to obtain the test results plus a reserve, which may be used if any re-testing becomes necessary. How large this reserve needs to be depends on the measurement method and its complexity. In any event, the surplus material should be carefully stored to protect against deterioration or adverse changes in the test material.

Test samples should be identical, that is, last-stage samples of the sample-preparing procedure should be used by both laboratories.

5.3.2 Statistical testing for agreement between test results from two laboratories

5.3.2.1 Case where only one test result is obtained in each laboratory

When each laboratory has obtained only one test result, the absolute difference between the two test results should be tested against the reproducibility limit $R = 2,8\sigma_R$. If the absolute difference between the two test results does not exceed R, the two test results are considered to be in agreement and the mean of the two test results may be used as the final quoted result.

If R is exceeded, then it is necessary to discover whether the difference is due to poor precision of the

measurement method and/or a difference in the test samples. To test the precision under repeatability conditions, each laboratory should follow the procedures described in 5.2.2.

5.3.2.2 Case where two laboratories obtain more than one single test result

It is assumed that each laboratory will have used the procedures of 5.2 and obtained its final quoted result. Thus, it is only necessary to consider the acceptability of the two final quoted results. To verify whether the quoted results of the laboratories are in agreement, the absolute difference between the two final quoted results should be tested against the critical difference, $CD_{0,95}$, as given below.

a) $CD_{0,95}$ for two arithmetic means of n_1 and n_2 test results, respectively:

$$CD_{0,95} = \sqrt{R^2 - r^2 \left(1 - \frac{1}{2n_1} - \frac{1}{2n_2}\right)}$$

Note that in the equation above if $n_1 = n_2 = 1$, the expression reduces to R as given in 5.3.2.1.

If $n_1 = n_2 = 2$, the expression reduces to

$$CD_{0,95} = \sqrt{R^2 - \frac{r^2}{2}}$$

b) $CD_{0,95}$ for an arithmetic mean of n_1 and a median of n_2 test results, respectively:

$$CD_{0,95} = \sqrt{R^2 - r^2 \left(1 - \frac{1}{2n_1} - \frac{\{c(n_2)\}^2}{2n_2}\right)}$$

where $c(n)$ is the ratio of the standard deviation of the median to the standard deviation of the arithmetic mean. Its value is given in table 2.

c) $CD_{0,95}$ for two medians of n_1 and n_2 test results, respectively:

$$CD_{0,95} = \sqrt{R^2 - r^2 \left(1 - \frac{\{c(n_1)\}^2}{2n_1} - \frac{\{c(n_2)\}^2}{2n_2}\right)}$$

See table 2 for values of $c(n)$.

If the critical difference is not exceeded, both final quoted results of the two laboratories are acceptable and the grand mean of these two final quoted results can be used. If the critical difference is exceeded,

then the procedures outlined in 5.3.3 should be followed.

Table 2 — Values of $c(n)$

Number of test results, n	$c(n)$
1	1,000
2	1,000
3	1,160
4	1,092
5	1,197
6	1,135
7	1,214
8	1,160
9	1,223
10	1,176
11	1,228
12	1,187
13	1,232
14	1,196
15	1,235
16	1,202
17	1,237
18	1,207
19	1,239
20	1,212

5.3.3 Resolving discrepancies between results from two laboratories

The cause of discrepancies between the test results or the final quoted results of the two laboratories could be due to

— systematic differences between the two laboratories,

— difference in test samples, or

— errors in the determination of σ_r and/or σ_R.

If it is possible to exchange the test samples and/or reference standard materials, each laboratory should obtain test results using the other's test sample to determine the existence and degree of systematic error. If exchange of test samples is not possible, each laboratory should obtain test results on a common sample (preferably a material of known value). The use of a material of known value has the advantage that systematic error can be ascribed to one or both laboratories. Where the use of a material of known value is not possible in order to ascribe systematic error to the laboratories, agreement should be reached between the two laboratories to refer to a third reference laboratory.

When the discrepancy appears to lie in differences between the test samples, both laboratories should

combine to make a joint sampling, or a third party should be invited to carry out the sampling.

5.3.4 Arbitration

The two parties to a contract may agree to an arbitration procedure at the time of concluding a contract or when a dispute arises.

6 Method for checking the stability of test results within a laboratory

6.1 Background

6.1.1 The first step in quality control is quantification by means of chemical analysis, physical test, sensory test, etc. The observed values obtained by these quantification methods are always accompanied by some errors, which can be divided into errors due to

— sampling,

— sample preparation,

— measurement, etc.

However, this clause will deal only with the error due to measurement; that is the measurement error including the inseparable variation between test portions of a test sample.

6.1.2 It is considered that the measurement error can be further divided into

— an error which is attributed to random cause (precision), and

— an error which is attributed to systematic cause (trueness).

6.1.3 In considering a measurement method, it is quite natural to expect that both the precision and trueness of the measurement method are satisfactory. However, there is no guarantee that the measurement method is satisfactory in trueness if it is satisfactory in precision. Accordingly, when the stability of test results is to be examined within a laboratory, it is necessary to check both the precision and trueness of the test results and maintain the two measures at desired levels, respectively, for a long period of time.

6.1.4 However, it can be that no true value exists for the measurement method or, even if a true value exists, there is no opportunity for checking the trueness of test results due to the unavailability of a reference material (RM). These examples are shown in table 3.

It is difficult to check the trueness of a test result if there is no RM. However, in practice, in many cases a test result obtained by a skilled operator in a well-equipped laboratory following a standard measurement method (or preferably a "definitive" method) strictly, thoroughly and carefully, can be used as a reference value in place of the certified value.

6.1.5 For checking the stability of test results within a laboratory, Shewhart control charts (see ISO 8258) and cumulative sum control charts are used in this part of ISO 5725.

In the situation where precision or trueness has a trend or shift, the cumulative sum control chart is more effective than the Shewhart control chart, whereas in the situation in which a sudden change might occur, no advantage is gained in applying the cumulative sum control chart instead of the Shewhart control chart.

Since a trend or shift is more likely to occur in trueness and sudden changes are more likely in precision, the cumulative sum control chart is recommended for checking trueness and the Shewhart control chart for checking precision.

However, it might be worthwhile to use both control charts in parallel for checking precision and trueness as well.

6.1.6 Because the checking procedures cover a longer period of time and probably involve changes of operator and equipment, true repeatability conditions do not apply. The checking, therefore, involves the use of intermediate precision measures which are described in ISO 5725-3.

6.2 Methods for checking stability

6.2.1 General

6.2.1.1 There are two cases to be considered when checking the stability of test results within a laboratory:

a) for routine test results to be used for process control, and

b) for test results to be used for price determination of raw materials and manufactured goods.

Table 3 — Classification for characteristics of test materials according to their true values and important parameters for checking accuracy (trueness and precision) of results

Classification[1]	Examples		
	Characteristics	Availability of RM[2]	Important parameters for checking accuracy[3]
A theoretical value based on scientific principles can be established practically as a *true value*.	Chemical component of benzoic acid	RM[4]	Δ and σ_W
Although a true value exists theoretically, a unique true value cannot be established in practice with the present technique; therefore the consensus value based on collaborative experimental work under the auspices of a scientific or engineering group is adopted as a *conventional true value*.	a) Percentage of Fe in an ore	RM	Δ and σ_W
	b) Percentage of S in pyrite	No[5]	σ_W and σ_L
An assigned value based on a reference test method established internationally, nationally or by a private organization is adopted as a *conventional true value*.	a) Octane value of gasoline	RM	Δ and σ_W
	b) Strength of coke	No[6]	σ_M/σ_W, σ_L and σ_W
	c) Melt flowrate of thermoplastics	No[7]	σ_W and σ_L

1) See ISO 3534-1.

2) See ISO Guide 35.

3) Δ is the laboratory bias; σ_W is the within-laboratory standard deviation; σ_L is the between-laboratory standard deviation; σ_M^2 is the between-test-sample standard deviation.

4) The test material itself may be used as a RM if it is pure and stable.

5) No RM can be established due to the material being unstable.

6) No RM can be established due to a large mass consisting of solid, fragile particles differing in particle size, shape and composition being needed for each test, which is destructive.

7) Reference value is defined by the measurement method itself.

6.2.1.2 In a), it is necessary to check the intermediate-precision standard deviations with one, two or three factors different to be obtained from the test results within the specific laboratory for a long period of time to see that the precision measure is maintained at a desired level (see example 2 in 6.2.3). In this case, the checking of the precision measure alone is sufficient for most cases, because even if the test results are biased, it is possible to check the process variation if the variation of the test results is sufficiently small compared to that of the production process. However, if the repeatability standard deviation is used for such a purpose, an over-reaction might result in the process control because of excessive sensitivity; therefore it is advisable to use an appropriate intermediate-precision standard deviation for this purpose.

6.2.1.3 In b), it is necessary to check the trueness (see example 3 in 6.2.4) as well as precision, to see that both measures are maintained at the desired level, respectively; therefore an accepted reference value is required in this case.

6.2.1.4 Four examples are presented as follows:

— examples 1 and 2 show how to check, by the Shewhart control chart method, the stability of a repeatability or of an intermediate precision measure;

— examples 3 and 4 show how to check trueness, using either the Shewhart control chart or the cusum method.

6.2.2 Example 1: Stability check of the repeatability standard deviation of a routine analysis

6.2.2.1 Background

a) **Measurement method:**

Determination of nickel content by the method given in ISO 6352:1985, *Ferronickel — Determination of nickel content — Dimethylglyoxime gravimetric method.*

b) **Source:**

Routine report in September 1985 of a laboratory of a ferronickel smelter.

c) **Description:**

In the works laboratory of the ferronickel smelter, chemical analysis is carried out every day to determine the chemical composition of the ferronickel products, together with a stability check of the nickel determination, using a private reference material prepared by the laboratory.

In order to check the stability of the above nickel determination, two test portions of the private reference material are analysed every day under repeatability conditions, i.e. by the same operator using the same equipment at the same time.

The chemical composition of the private reference material is:

Ni	47,21 %	Co	1,223 %	Si	3,50 %
Mn	0,015 %	P	0,003 %	S	0,001 %
Cr	0,03 %	Cu	0,038 %		

6.2.2.2 Original data

The routine analysis test results of the nickel content of the private reference material obtained under repeatability conditions are presented in table 5 as x_1 and x_2, expressed as a percentage by mass.

6.2.2.3 Stability check by the Shewhart control chart method

By applying the Shewhart control chart method (R-chart) (see ISO 8258) to the test results in table 5, the stability of the test results is checked, and the magnitude of the repeatability standard deviation is evaluated. In calculating the central line and control limits (UCL and LCL), the factors given in table 4 are used.

NOTE 4   To avoid confusion with the symbol R, used here for reproducibility, the R-chart of ISO 8258 will be referred to here as a range chart.

Table 4 — Factors for computing a range chart

Factors for computing the central line and action limits[1]			Factors for computing the warning limits[2]		
Number of observations in subgroup	**Factor for central line** d_2	**Factor for action limit** D_2	**Factors for warning limits** d_3	$D_1(2)$	$D_2(2)$
2	1,128	3,686	0,853	—	2,834
3	1,693	4,358	0,888	—	3,469
4	2,059	4,698	0,880	0,299	3,819
5	2,326	4,918	0,864	0,598	4,054

1) These data are extracted from table 2 of ISO 8258:1991.

2) The factors applied for calculating the warning limits are as follows:

$$D_1(2) = d_2 - 2d_3$$

$$D_2(2) = d_2 + 2d_3$$

Table 5 — Control chart data sheet for example 1 (6.2.2)

1. Quality characteristic:	Nickel content of a private reference material
2. Unit of measurement:	% (m/m)
3. Analysis method:	ISO 6352
4. Period:	1985-09-01 to approx. 1985-09-30
5. Laboratory:	Works laboratory "A" of a ferronickel smelter

Date of analysis (subgroup number)	Observed values x_1	Observed values x_2	Range w	Description
1	47,379	47,333	0,046	
2	47,261	47,148	0,113	Above the warning limit
3	47,270	47,195	0,075	
4	47,370	47,287	0,083	
5	47,288	47,284	0,004	
6	47,254	47,247	0,007	
7	47,239	47,160	0,079	
8	47,239	47,193	0,046	
9	47,378	47,354	0,024	
10	47,331	47,267	0,064	
11	47,255	47,278	0,023	
12	47,313	47,255	0,058	
13	47,274	47,167	0,107	Above the warning limit
14	47,313	47,205	0,108	Above the warning limit
15	47,296	47,231	0,065	
16	47,264	47,247	0,017	
17	47,238	47,253	0,015	
18	47,181	47,255	0,074	
19	47,327	47,240	0,087	
20	47,358	47,308	0,050	
21	47,295	47,133	0,162	Above the action limit
22	47,310	47,244	0,066	
23	47,366	47,293	0,073	
24	47,209	47,185	0,024	
25	47,279	47,268	0,011	
26	47,178	47,200	0,030	
27	47,211	47,193	0,018	
28	47,195	47,216	0,021	
29	47,274	47,252	0,022	
30	47,300	47,212	0,088	
Total			1,660	
Average			0,055 3	$\bar{w}/d_2 = 0,049\ 0$

Remarks

$\sigma_r = 0,037\ 5$

a) Central line $= d_2\sigma_r = 1,128 \times 0,037\ 5 = 0,042\ 3$

b) Action limits

UCL $= D_2\sigma_r = 3,686 \times 0,037\ 5 = 0,138\ 2$
LCL $=$ none

c) Warning limits

UCL $= D_2(2)\sigma_r = 2,834 \times 0,037\ 5 = 0,106\ 2$
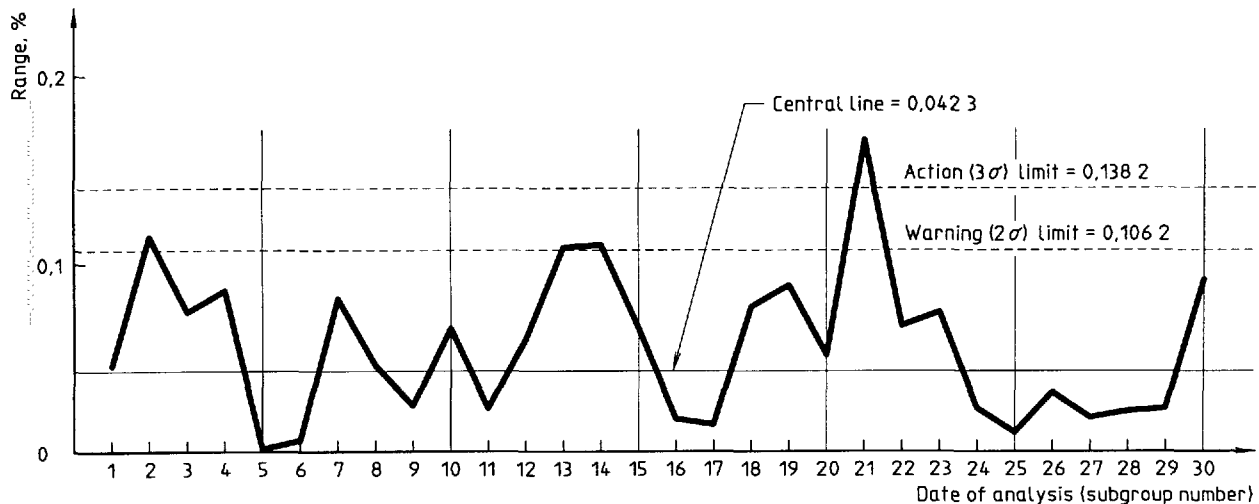LCL $=$ none

Figure 7 — Range chart for the nickel content (%) of a private reference material, obtained under repeatability conditions

Since the repeatability standard deviation obtained from the test results in the previous quarter of the year (σ_r) is given as the standard value for a range control chart for this example, the control chart is calculated as follows:

a) Central line $= d_2\sigma_r = 1{,}128 \times 0{,}037\ 5 = 0{,}042\ 3$

b) Action limits

$$UCL = D_2\sigma_r = 3{,}686 \times 0{,}037\ 5 = 0{,}138\ 2$$
$$LCL = \text{none}$$

c) Warning limits

$$UCL = D_2(2)\sigma_r = 2{,}834 \times 0{,}037\ 5 = 0{,}106\ 2$$
$$LCL = \text{none}$$

The estimate of the repeatability standard deviation (s_r) is derived from the following equations:

$$w = |x_1 - x_2|$$

$$s_r = \left(\sum_1^{30} w_i/30 \right) /d_2 = \overline{w}/d_2 = 0{,}055\ 3/1{,}128$$

$$= 0{,}049\ 0$$

The ranges are calculated for 30 subgroups, each containing 2 samples. Table 5 is an example of a work sheet to do this, and figure 7 is an example of the data plotted with the control limits shown.

The chart shown in figure 7 indicates that the test results are not stable because there is one point above the action limit and a pair of consecutive points above the warning limit.

6.2.3 Example 2: Stability check of the time-and-operator-different intermediate precision standard deviation of a routine analysis

6.2.3.1 Background

a) **Measurement method:**

Determination of the sulfur content in blast-furnace coke, with test results expressed as a percentage by mass, by the method given in ISO 351:1984, *Solid mineral fuels — Determination of total sulfur — High temperature combustion method*.

b) **Source:**

Routine report in August 1985 of a laboratory of a steel mill.

c) **Description:**

From a coke battery which produces blast-furnace coke, coke samples are taken routinely, from each production lot, every work-shift of the three-shift production scheme, every day. Then a test sample for chemical analysis is prepared in the laboratory for every production lot to determine the sulfur content [% (m/m)].

© ISO

6.2.3.2 Original data

The test results of a quality control analysis of sulfur content [% (m/m)] in coke test samples from the No. 1 coke battery in August 1985 are given in table 6. One coke test sample, which has been chosen at random and kept aside from the test samples which were analysed in a shift (x_1), is analysed again by another operator in another shift on the next day (x_2), and the test results are compared every day.

6.2.3.3 Stability check by Shewhart control chart method

By applying the Shewhart control chart method (range chart; see ISO 8258) to the data in table 6, the stability of the test results is checked and the magnitude of the time-and-operator-different intermediate precision standard deviation is evaluated.

Regarding the factors for calculating the central line, and the action and warning limits (UCL and LCL), see example 1 in 6.2.2. Since the time-and-operator-different intermediate precision standard deviation obtained from the test results in the previous quarter of the year, $\sigma_{I(TO)}$, is given as the standard value for the range chart for this example, the control chart is calculated as follows.

a) Central line = 1,128 × 0,013 3 = 0,015 0

b) Action limits

$$UCL = D_2\sigma_{I(TO)} = 3,686 \times 0,013\ 3 = 0,049\ 0$$
$$LCL = \text{none}$$

c) Warning limits

$$UCL = D_2(2)\sigma_{I(TO)} = 2,834 \times 0,013\ 3$$
$$= 0,037\ 8$$
$$LCL = \text{none}$$

The estimate of the time-and-operator-different intermediate precision standard deviation, $s_{I(TO)}$, is derived from the following equations:

$$w = |x_1 - x_2|$$

$$s_{I(TO)} = \left(\sum_1^{31} w_i/31\right) / d_2 = \bar{w}/d_2 = 0,014\ 2/1,128$$

$$= 0,012\ 6$$

The ranges are calculated for 31 subgroups, each containing 2 samples, as in table 6, and are plotted in figure 8 with the above calculated control limits.

The chart shown in figure 8 gives no evidence that the test results are not stable.
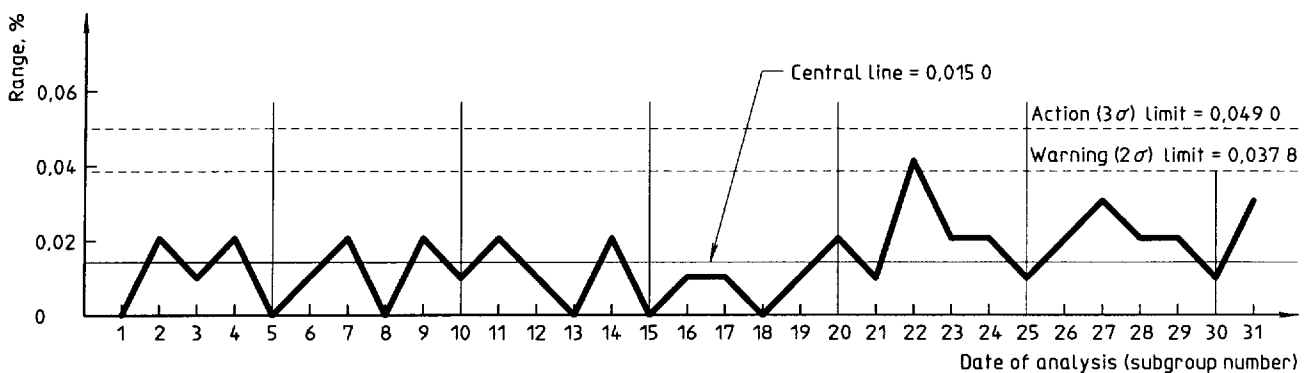

Figure 8 — Range chart for the sulfur content (%) in blast-furnace coke, obtained under time-and-operator-different intermediate precision conditions

Table 6 — Control chart data sheet for example 2 (6.2.3)

1. Quality characteristic.	Sulfur content of blast-furnace coke
2 Unit of measurement:	% (m/m)
3 Analysis method:	ISO 351
4 Period:	1985-08-01 to approx. 1985-08-31
5. Laboratory:	Works laboratory "B" of a steel mill

Date of analysis (subgroup number)	Observed values		Range	Description
	x_1	x_2	w	
1	0,56	0,56	0,00	
2	0,48	0,50	0,02	
3	0,57	0,58	0,01	
4	0,60	0,58	0,02	
5	0,58	0,58	0,00	
6	0,50	0,49	0,01	
7	0,56	0,58	0,02	
8	0,56	0,56	0,00	
9	0,48	0,46	0,02	
10	0,54	0,53	0,01	
11	0,55	0,57	0,02	
12	0,46	0,45	0,01	
13	0,58	0,58	0,00	
14	0,54	0,56	0,02	
15	0,56	0,56	0,00	
16	0,57	0,58	0,01	
17	0,46	0,45	0,01	
18	0,56	0,56	0,00	
19	0,56	0,57	0,01	
20	0,57	0,55	0,02	
21	0,44	0,45	0,01	
22	0,59	0,55	0,04	Above the warning limit[1]
23	0,55	0,57	0,02	
24	0,58	0,56	0,02	
25	0,46	0,45	0,01	
26	0,60	0,58	0,02	
27	0,59	0,56	0,03	
28	0,54	0,56	0,02	
29	0,47	0,49	0,02	
30	0,59	0,58	0,01	
31	0,49	0,52	0,03	
Total	16,84	16,72	0,44	
Average			0,014 2	$\overline{w}/d_2 = 0,012\ 6$

Remarks

$\sigma_{I(TO)} = 0,013\ 3$
x_1 : Routine analysis
x_2 · Second analysis on the next day by a different operator

a) Central line $= d_2\sigma_{I(TO)} = 1,128 \times 0,013\ 3 = 0,015\ 0$

b) Action limits

   $UCL = D_2\sigma_{I(TO)} = 3,686 \times 0,013\ 3 = 0,049\ 0$
   $LCL =$ none

c) Warning limits

   $UCL = D_2(2)\sigma_{I(TO)} = 2,834 \times 0,013\ 3 = 0,037\ 8$
   $LCL =$ none

1) The actual heating temperature for obtaining x_2 was lower than that specified.

6.2.4 Example 3: Stability check of the trueness of a routine analysis

6.2.4.1 Background

a) **Measurement method:**

Determination of the ash content in coal, expressed as a percentage by mass, by the method given in ISO 1171:1981, *Solid mineral fuels — Determination of ash*.

b) **Source:**

Routine report in June 1985 of a laboratory of a steel mill.

c) **Description:**

In the steel mill, coal blends are supplied to produce blast-furnace coke in a coke battery by a three-shift production scheme.

In order to control the quality of the coke products, the ash contents [% (m/m)] in coals are analysed every shift by the method given in ISO 1171. The stability check of the time-and-operator-different intermediate precision standard deviation of the routine analysis is carried out as in example 2 (6.2.3).

This example shows the method of checking the stability of the trueness of the routine analysis by using a private reference material (ash content = 10,29 %).

6.2.4.2 Original data

Every day, the private reference material is analysed by an operator who has been assigned at random from all the operators in the three shifts. The test results are presented as y in table 7.

6.2.4.3 Stability check by Shewhart control chart method

By applying the Shewhart control chart method to the data in table 7, the stability of the trueness of the routine analysis is checked and the magnitude of the bias is evaluated.

The repeatability standard deviation (s_r) cannot be used for checking the bias within this specific laboratory, where the routine analysis is carried out under time-and-operator-different intermediate precision conditions, hence s_r does not represent the actual precision of the test results obtained in the laboratory.

Rather than carry out an experiment to obtain the time-and-operator-different intermediate precision standard deviation, $s_{I(TO)}$, the moving range chart method is adopted as a simpler means.

The control chart is prepared using formulae as given in the remarks to table 7 and previously established values of μ and $\sigma_{I(TO)}$. The chart in figure 9 shows periods when both the bias and the ranges are very small, and other periods when the test results are much less stable, justifying an investigation of the reasons for these patterns.

6.2.4.4 Stability check by cumulative sum control chart method

Computation for $(H;K)$ in the cumulative sum control chart for $\hat{\delta}$ with $(h;k) = (4,79;0,5)$ is as follows (see figure 10).

Upper side:

$H = h\sigma_{I(TO)}$
$\quad = 4,79 \times 0,066\ 45$
$\quad = 0,318$

$K_1 = \mu + k\sigma_{I(TO)}$
$\quad = 10,29 + 0,5 \times$
$\qquad \times 0,066\ 45$
$\quad = 10,323$

Lower side:

$-H = -0,318$

$K_2 = \mu - k\sigma_{I(TO)}$
$\quad = 10,29 - 0,5 \times$
$\qquad \times 0,066\ 45$
$\quad = 10,257$

Table 7 — Control chart data sheet for example 3 (6.2.4)

1. Quality characteristic:	Ash content of a private reference material
2. Unit of measurement:	% (m/m)
3. Analysis method.	ISO 1171
4. Period:	1985-06-01 to approx. 1985-06-30
5. Laboratory:	Works laboratory "C" of a steel mill

Date of analysis (subgroup number)	Test result y	Estimate of bias $\hat{\delta}$	Moving range w	Description
1	10,30	0,01	0,01	
2	10,29	0,00	0,01	
3	10,28	− 0,01	0,02	
4	10,30	0,01	0,01	
5	10,29	0,00	0,00	
6	10,29	0,00	0,09	
7	10,20	− 0,09	0,08	
8	10,28	− 0,01	0,01	
9	10,29	0,00	0,00	
10	10,29	0,00	0,10	
11	10,19	− 0,10	0,10	
12	10,29	0,00	0,00	
13	10,29	0,00	0,00	
14	10,29	0,00	0,01	
15	10,28	− 0,01	0,02	
16	10,30	0,01	0,01	
17	10,29	0,00	0,00	
18	10,29	0,00	0,01	
19	10,28	− 0,01	0,00	
20	10,28	− 0,01	0,00	
21	10,28	− 0,01	0,03	
22	10,31	0,02	0,12	
23	10,19	− 0,10	0,10	
24	10,29	0,00	0,07	
25	10,36	0,07	0,00	
26	10,36	0,07	0,07	
27	10,29	0,00	0,01	
28	10,30	0,01	0,02	
29	10,28	− 0,01	0,09	
30	10,19	− 0,10		
Total	308,44	− 0,26	0,99	
Average		− 0,086 6	0,034 1	$\overline{w}/d_2 = 0,030\ 2$

Remarks

Ash content of the private reference material

$\mu = 10,29$

Standard deviation obtained from the test results of the previous quarter of the year

$\sigma_{I(TO)} = 0,066\ 45$

Estimate of bias

$\hat{\delta} = y - \mu$

Moving range

$w = |\hat{\delta}_{i+1} - \hat{\delta}_i|$

x-chart:

Central line	= 0
Action limits	$UCL = + 3\sigma_{I(TO)} = 0,199\ 4$
	$LCL = - 3\sigma_{I(TO)} = - 0,199\ 4$
Warning limits	$UCL = + 2\sigma_{I(TO)} = 0,132\ 9$
	$LCL = - 2\sigma_{I(TO)} = - 0,132\ 9$

Moving range chart:

Central line	$= d_2\sigma_{I(TO)} = 1,128 \times 0,066\ 45 = 0,074\ 96$
Action limits	$UCL = D_2\sigma_{I(TO)} = 3,396 \times 0,066\ 45 = 0,245$
	$LCL = $ none
Warning limits	$UCL = D_2(2)\sigma_{I(TO)} = 2,834 \times 0,066\ 45 = 0,188\ 3$
	$LCL = $ none
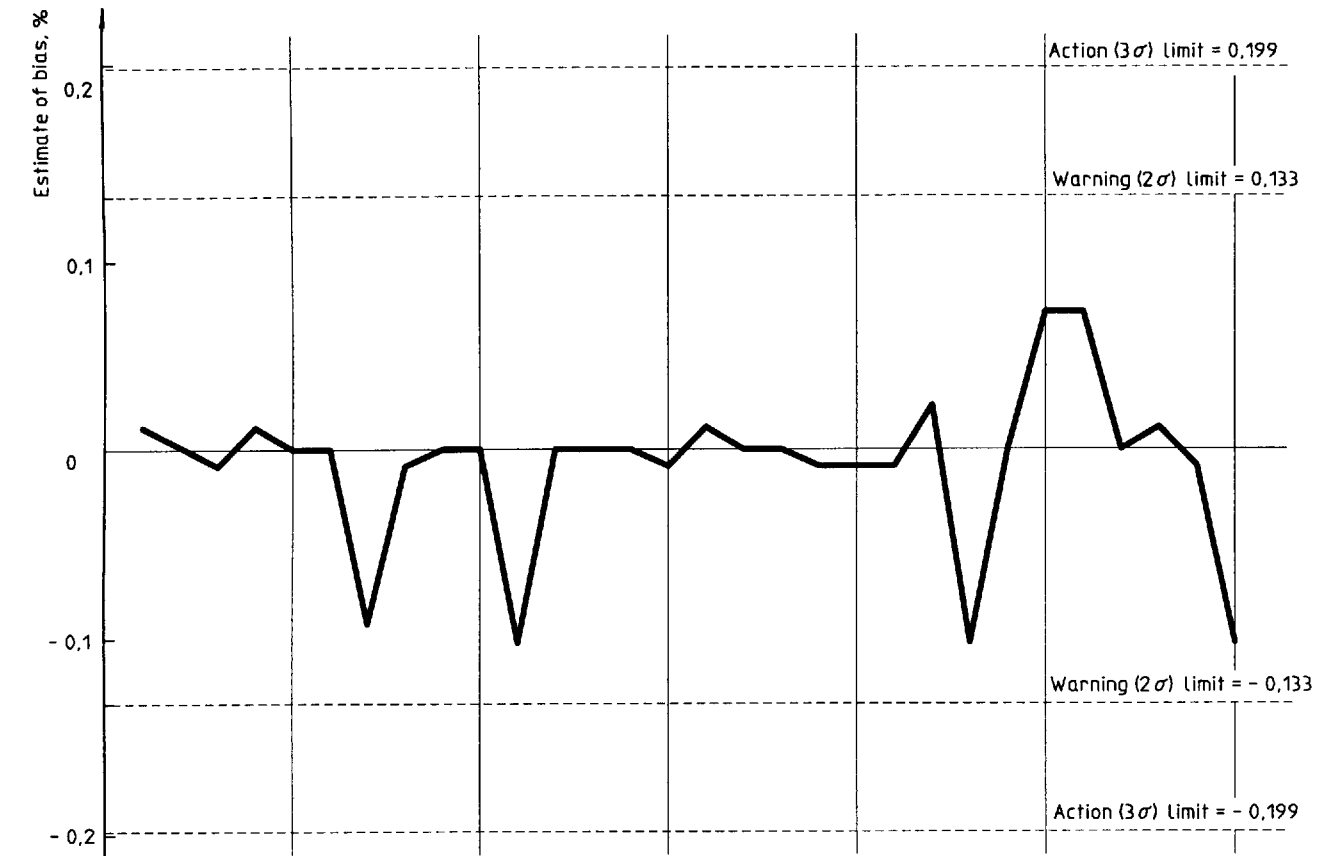
a)
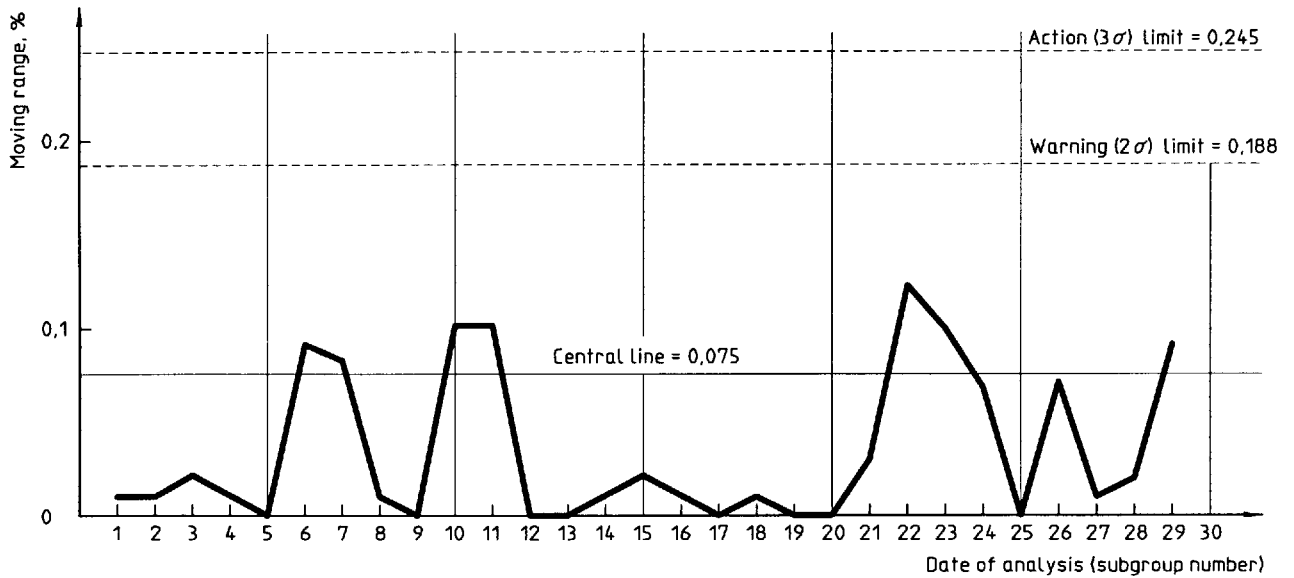
b)

Figure 9 — Shewhart control chart for $\hat{\delta}$ of the ash content [% (m/m)] of a private reference material
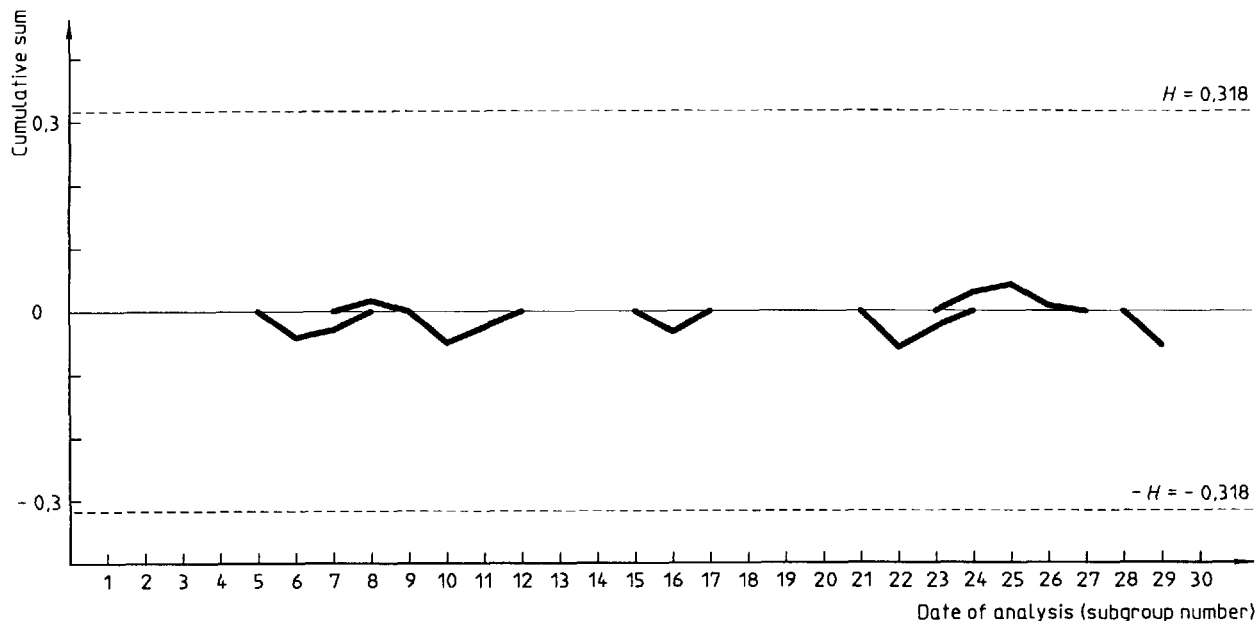
Figure 10 — Cumulative sum control chart for $\hat{\delta}$ of the ash content [% (m/m)] of a private reference material

6.2.5 Example 4: Another stability check of the trueness of a routine analysis

6.2.5.1 Background

a) **Measurement method:**

Determination of the arsenic content in zinc oxide by a silver diethyldithiocarbamate-arsine evolution colorimetric procedure.

b) **Source:**

Kanzelmeyer J.H. "Quality Control for Analytical Methods", *ASTM Standardization News*, October 1977, Figure 2, p. 27.

6.2.5.2 Original data

See table 8.

6.2.5.3 Stability check by Shewhart control chart method

The Shewhart control chart for $\bar{x}$ (see figure 11) is prepared using the formulae as given in the remarks to table 8, and previously established values of μ and σ_r.

The chart shows instability in the test results, as there is one point above the action limit, and two runs of seven or more test results below the central line.

6.2.5.4 Stability check by cumulative sum control chart method

Computation for $(H;K)$ in the cumulative sum control chart for $\bar{x}$ with $(h;k) = (4,79;0,5)$ is as follows (see figure 12).

Upper side:

$H = h\sigma_r/\sqrt{n}$

$= 4,79 \times 0,167$

$= 0,800$

$K_1 = \mu + k\sigma_r/\sqrt{n}$

$= 3,800 + 0,5 \times 0,167$

$= 3,88$

Lower side:

$-H = -0,800$

$K_2 = \mu - k\sigma_r/\sqrt{n}$

$= 3,800 - 0,5 \times 0,167$

$= 3,72$

23

ISO 5725-6:1994(E) © ISO

Table 8 — $\bar{x}$-chart data sheet for example 4 (6.2.5)

1. Quality characteristic:	As content of a private reference material
2. Unit of measurement:	ppm by mass
3. Analysis method:	Silver diethyldithiocarbamate-arsine evolution colorimetric procedure

Subgroup number	Observed values		$\bar{x}$	Description
	x_1	x_2		
1	3,70	3,80	3,75	
2	3,76	3,86	3,81	
3	3,64	3,38	3,51	
4	4,01	3,62	3,82	
5	3,40	3,52	3,46	
6	3,65	3,53	3,59	
7	3,20	3,58	3,39	
8	4,19	4,65	4,42	Above the action limit
9	3,97	3,77	3,87	
10	2,95	3,69	3,32	
11	3,43	3,55	3,49	
12	3,85	3,53	3,69	
13	3,77	3,17	3,47	
14	3,19	3,60	3,40	
15	3,75	3,45	3,60	
16	3,55	3,25	3,40	
17	3,98	3,76	3,87	
18	3,56	3,78	3,67	
19	3,54	4,02	3,78	
20	3,35	3,55	3,45	
21	3,37	3,25	3,31	
22	3,42	3,42	3,42	
23	3,71	3,87	3,79	
24	3,77	3,62	3,70	
25	3,82	3,58	3,70	
26	3,73	3,02	3,38	
27	3,48	3,28	3,38	
28	4,01	4,19	4,10	
29	3,63	3,11	3,37	
30	3,51	3,23	3,37	
Total			108,28	
Average			3,609	

Remarks

Arsenic content of the private reference material

$\mu = 3,80$

Standard deviation in the past

$\sigma_r = 0,236$

$\bar{x}$-chart

Central line $= 3,80$

Action limits

$UCL = \mu + 3\sigma_r/\sqrt{n} = 4,300$

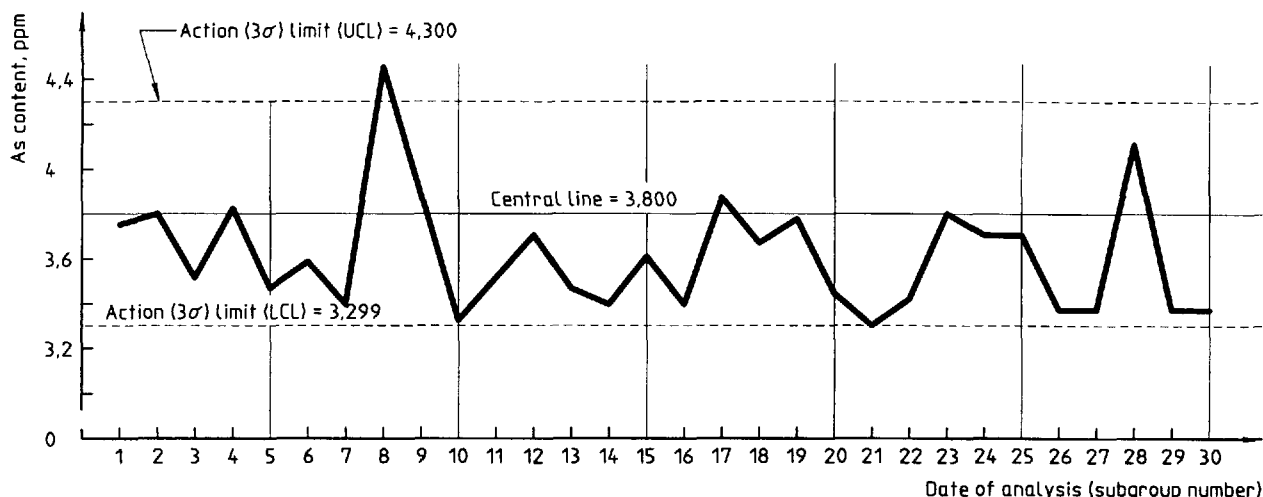$LCL = \mu - 3\sigma_r/\sqrt{n} = 3,299$

Figure 11 — Shewhart control chart for $\bar{x}$ of the As content for the silver diethyldithiocarbamate-arsine evolution colorimetric procedure for arsenic in zinc oxide
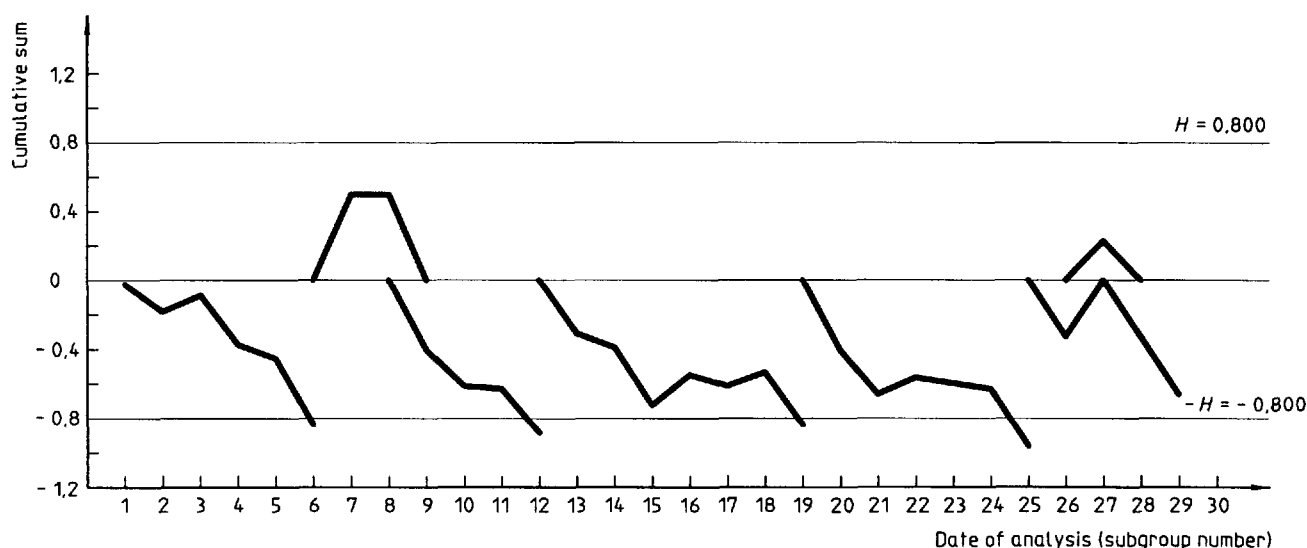
Figure 12 — Cumulative sum control chart for $\bar{x}$ of the As content for the silver diethyldithiocarbamate-arsine evolution colorimetric procedure for arsenic in zinc oxide

7 Use of repeatability and reproducibility standard deviations in assessing laboratories

7.1 Assessment method

7.1.1 General

This clause describes assessment of laboratories with regard to only a single measurement method which is standardized and which is in use in various laboratories. Consequently, it is possible to estimate the precision of the method in the form of the repeatability and the reproducibility standard deviations. It is assumed that these values have been determined in advance by a precision experiment.

There are three types of assessment depending on the existence of reference materials for the method or of a reference laboratory. When reference materials exist on an adequate number of levels, the assess-

ment may take place with the participation of the in-
dividual laboratory only. Concerning a measurement
method for which no reference materials exist, such
a simple assessment is not possible. The laboratory
has to be compared with a high-quality laboratory
which is widely recognized as providing an acceptable
benchmark for the assessment. For the continued
assessment of laboratories, a number of laboratories
often have to be assessed simultaneously. In this
situation a collaborative assessment experiment is
useful.

The purpose of carrying out a collaborative assess-
ment experiment is to compare the results of each
laboratory with those of the other laboratories with
the object of improving performance.

7.1.2 Implications of the definition of a collaborative assessment experiment

The repeatability standard deviation of a measure-
ment method measures the uncertainty of measure-
ments obtained under uniform conditions within a
laboratory. In this way it is an expression of the
within-laboratory precision of the laboratory under the
repeatability conditions defined in ISO 5725-1.

The bias of the laboratory can be determined im-
mediately when a true value of the property being
measured exists, and is known, as is the case with
reference materials. When a true value is not known,
the bias has to be determined indirectly. One way is
to compare the laboratory with another laboratory
with known bias. This solution, however, depends
strongly on the precision and bias of the "reference"
laboratory.

In the case of a collaborative assessment experiment,
the reproducibility indicates the accordance between
the results achieved in different laboratories. Conse-
quently, it can be used to evaluate the bias of each
laboratory. A laboratory which shows a large system-
atic deviation will appear as an outlier when the re-
producibility of an assessment experiment is
determined.

In this clause it is assumed that the precision of the
measurement method is determined in advance. This
means that the repeatability variance σ_r^2, the
between-laboratory variance σ_L^2, and the reproducibility
variance σ_R^2 are known.

The methods in clause 7 are principally intended to
check laboratory bias. The methods in clause 6 are
more effective in checking the repeatability of a lab-
oratory or its intermediate precision.

7.2 Evaluation of the use of a measurement method by a laboratory not previously assessed

7.2.1 Evaluation of laboratory practice

For general criteria for a laboratory evaluation, see
ISO/IEC Guide 25. The laboratory shall live up to good
laboratory practice, and have satisfactory internal
quality control. Methods for internal quality control
have already been described in clause 6.

This part of the control is only based on an inspection
of each laboratory in its usual working situation. This
can be carried out immediately without the use of
special test material and without involving other lab-
oratories.

It is necessary to carry out a control experiment in
order to evaluate quantitatively the laboratory's use
of the measurement method. This can be done either
internally in the laboratory by using reference ma-
terials (see 7.2.3) or by comparison with a good lab-
oratory (see 7.2.4).

7.2.2 General considerations concerning control experiments

The following questions should be considered when
a control experiment is planned.

a) On how many levels should the experiment be
   carried out (q)? This point is considered in
   ISO 5725-1:1994, 6.3.

b) How many replications should be carried out on
   each level (n)?

In the case of a collaborative assessment experiment:

c) How many laboratories will participate (p)?

When planning the experiment, subclause 6.1 in
ISO 5725-1:1994, as well as clauses 5 and 6 in
ISO 5725-2:1994 should be taken into consideration.

The test material shall be sent anonymously to the
laboratory, that is in such a way as to ensure that it is
treated in a manner consistent with the usual practice
within that laboratory and not given special treatment.

7.2.3 Measurement method for which reference materials exist

7.2.3.1 General

7.2.3.1.1 When reference materials exist, the assessment may take place in a single laboratory. As the precision of the method is known, the known value of the repeatability standard deviation is used when assessing the internal precision, while the bias is determined by comparing the test results with the reference value.

Sometimes it is relevant to introduce a detectable laboratory bias Δ_m as the minimum value of the laboratory bias that the experimenter wishes to detect with high probability from the results of the experiment.

7.2.3.1.2 It is necessary to carry out repeated measurements within the laboratory in order to assess the internal precision. After the considerations mentioned in 7.2.2, test material is sent out on q levels, and n replications of measurements are carried out on each level. When evaluating the results, use the method given in clause 7 of ISO 5725-2:1994. When assessing internal precision, the intracell standard deviation s_r is compared with the known repeatability standard deviation σ_r. The acceptance criterion is

$$s_r^2/\sigma_r^2 < \chi_{(1-\alpha)}^2(v)/v \qquad \ldots (1)$$

where $\chi_{(1-\alpha)}^2(v)$ is the $(1-\alpha)$-quantile of the χ^2 distribution with $v = n - 1$ degrees of freedom. Unless otherwise stated, the significance level α is assumed to be 0,05.

This inequality should be valid for about 95 % of the q levels. As normally q is rather small, this means that the criterion (1) shall be valid at all the q levels for the laboratory.

7.2.3.1.3 When assessing the bias, the average $\bar{y}$ for each level is compared with the corresponding reference value μ. Since

$$s_{(\bar{y})}^2 = s_L^2 + \frac{1}{n} s_r^2 = s_R^2 - s_r^2 \frac{(n-1)}{n} \qquad \ldots (2)$$

the acceptance criterion is

$$|\bar{y} - \mu| < 2\sqrt{\sigma_R^2 - \sigma_r^2 \frac{(n-1)}{n}} \qquad \ldots (3)$$

The acceptance criterion (3) shall be valid at each of the q levels.

When $n = 2$, criterion (3) is reduced to

$$|\bar{y} - \mu| < 2\sqrt{\sigma_R^2 - \frac{\sigma_r^2}{2}} \qquad \ldots (4)$$

In the case of a detectable bias, a further acceptance criterion is introduced as

$$|\bar{y} - \mu| < \Delta_m/2 \qquad \ldots (5)$$

7.2.3.2 Example: Determination of the cement content of concrete

7.2.3.2.1 Background

Cement content is important in that it affects the durability of concrete, and often a specification for concrete contains a minimum value for the cement content. The cement content can be determined from measurements of the calcium content of samples of the cement and aggregates and of the concrete specimens. For the assessment of a laboratory, it is possible to prepare concrete specimens of known cement content.

For the assessment of six laboratories, reference specimens with a cement content of 425 kg/m^3 were prepared. In each laboratory two determinations were performed.

7.2.3.2.2 Original data

See table 9. The values of the repeatability and reproducibility standard deviations are:

$$\sigma_r = 16$$

$$\sigma_R = 25$$

Table 9 — Cement content of concrete

Laboratory i	Observed values	
	y_{i1}	y_{i2}
1	406	431
2	443	455
3	387	431
4	502	486
5	434	456
6	352	399

7.2.3.2.3 Computation of cell means and ranges

See table 10.

Table 10 — Cell means and ranges

Laboratory	Cell mean	Range
1	418,5	25
2	449	12
3	409	44
4	494	16
5	445	22
6	375,5	47

7.2.3.2.4 Assessment of within-laboratory precision

The ranges in table 10 are compared with the repeatability standard deviation using the formula:

$$\frac{(y_{i1} - y_{i2})^2}{2\sigma_r^2} \leqslant \chi^2_{(1-\alpha)}(v)$$

When $\alpha = 0,05$ and $v = 1$, $\chi^2_{0,95}(1) = 3,841$.

Laboratory No. 6 was found to deviate:

$$(y_{6,1} - y_{6;2})^2 = 2\,209; \text{ test value} = 4,31.$$

7.2.3.2.5 Assessment of bias

Formula (4) for the acceptance criterion gives:

$$|\bar{y} - 425| < 44,59$$

For laboratory No. 4, the test value is

$$|\bar{y}_4 - 425| = 69$$

For laboratory No. 6, the test value is

$$|\bar{y}_6 - 425| = 50,5$$

Hence both laboratories have an unsatisfactory bias.

7.2.4 Measurement method for which no reference materials exist

7.2.4.1 When no reference materials are available, the assessment has to be performed through comparison with a high-quality laboratory. It is essential to find a laboratory that works with a satisfactory pre-

cision and bias in order to reach a reliable conclusion about the new laboratory.

As is the case with reference materials, it is sometimes relevant to introduce a detectable difference λ between the two laboratory biases. It is defined as the minimum value of the difference between the expected values of the results obtained by two laboratories that the experimenter wishes to detect with high probability.

7.2.4.2 Test materials are sent to both laboratories as described in 7.2.3.1.2 and the internal precision in each laboratory is assessed similarly. The two laboratories should preferably obtain the same number (n) of measurements at each level.

7.2.4.3 When assessing the bias of the measurement method, δ, the arithmetic means at each level from the two laboratories are compared. Generally, let n_1 be the number of test results from the first laboratory and n_2 the number of test results from the second laboratory. Since

$$s^2_{(\bar{y}(1) - \bar{y}(2))} = 2\sigma_L^2 + \sigma_r^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)$$

$$= 2 \left[\sigma_R^2 - \sigma_r^2 \left(1 - \frac{1}{2n_1} - \frac{1}{2n_2} \right) \right] \qquad \dots (6)$$

the acceptance criterion is

$$|\bar{y}_1 - \bar{y}_2| \leqslant 2\sqrt{2} \sqrt{\sigma_R^2 - \sigma_r^2 \left[1 - \frac{1}{2n_1} - \frac{1}{2n_2} \right]}$$

$$\dots (7)$$

The acceptance criterion (7) shall be valid at each of the q levels.

When $n_1 = n_2 = 2$, criterion (7) is reduced to

$$|\bar{y}_1 - \bar{y}_2| \leqslant 2\sqrt{2} \sqrt{\sigma_R^2 - \frac{\sigma_r^2}{2}} \qquad \dots (8)$$

7.3 Continued assessment of previously approved laboratories

7.3.1 General considerations on continued control experiments

To guarantee that an approved laboratory is still functioning in a satisfactory way, continued assessment is necessary and should be carried out either by inspection visits or by participation in assessment experiments. No hard and fast rule can be laid down to say how often the assessment should take place, as various factors contribute to the decision; i.e. technical, economical and security factors. The responsible authority should decide the frequency depending on the situation.

Continued assessment often causes a situation where many laboratories have to be assessed simultaneously. In this situation, comparison with a high-quality laboratory is not recommended, because even the best laboratory has to be checked itself. In this situation, it is necessary to conduct a collaborative assessment experiment.

7.3.2 Evaluation of laboratory practice

Laboratory practice is assessed by means of inspection visits as described in 7.2.1.

7.3.3 Measurement method for which reference materials exist

The method described in ISO 5725-4 can be applied correspondingly in the continued assessment of laboratories.

7.3.4 Measurement method for which no reference materials exist

7.3.4.1 General

7.3.4.1.1 In the case where no reference materials are available, the assessment of each laboratory is based on a collaborative assessment experiment with several laboratories participating.

Planning an assessment experiment is very similar to planning a precision experiment, so many of the considerations mentioned in parts 1 and 2 of ISO 5725 apply. The purpose is to assess each laboratory so the choice of number of replications at each level is similar to the situation with one laboratory described in 7.2.2.

As the purpose is an assessment, a smaller number of laboratories may participate than in a precision ex-

periment. An obvious procedure would, for instance, be to carry out the experiment exclusively with national participation. It is especially important that the reduction in the number does not reduce the systematic deviation between laboratories, in which case the risk of not being able to reveal an outlying laboratory would be increased.

7.3.4.1.2 After the considerations mentioned in 7.2.2, test material is sent out to p laboratories at q levels, and n measurements are carried out at each level. When evaluating the results, use the method given in clause 7 of ISO 5725-2:1994. Because of possible missing or additional test results, a varying number might be obtained in the cells.

The internal precision is assessed for each laboratory as described in clause 6.

7.3.4.1.3 For the overall assessment of the biases, the reproducibility variance is calculated at each level (see ISO 5725-2:1994, 7.5).

$$s_R^2 = s_L^2 + s_r^2 \qquad \ldots (9)$$

where

$$s_L^2 = \left\{ \frac{1}{p-1} \left[\sum_{i=1}^{p} n_i (\bar{y}_i - \bar{\bar{y}})^2 \right] - s_r^2 \right\} / \bar{n} \qquad \ldots (10)$$

and

$$\bar{n} = \frac{1}{p} \sum_{i=1}^{p} n_i \qquad \ldots (11)$$

The between-laboratory variance s_L^2 is compared with the known between-laboratory variance σ_L^2.

The acceptance criterion is

$$\frac{\bar{n} s_L^2 + s_r^2}{\bar{n} \sigma_L^2 + \sigma_r^2} \leqslant \frac{\chi_{(1-\alpha)}^2 (\nu)}{\nu} \qquad \ldots (12)$$

where $\chi_{(1-\alpha)}^2(\nu)$ is the $(1-\alpha)$-quantile of the χ^2 distribution with $\nu = p - 1$ degrees of freedom. Unless otherwise stated, α is assumed to be 0,05.

If the acceptance criterion (12) is valid, the between-laboratory variance s_L^2 is acceptable and it can be concluded that all laboratories have obtained sufficiently accurate results at the level in question.

When the criterion is not valid, the furthest outlying observation is found by calculation of Grubbs' test statistic, then the results from the laboratory in question are omitted and the variances are again estimated for the remaining $(p - 1)$ laboratories. If the corrected

variance fulfils the criterion (12), the $(p-1)$ laboratories are approved, otherwise Grubbs' test statistic is calculated again and the procedure is repeated several times, if necessary. As mentioned in ISO 5725-2, Grubbs' test is not suitable for repeated applications. Consequently, many outliers ought to lead to an inspection of all data at all levels. If the same laboratories deviate at several levels, it can be concluded that these laboratories work with a bias which is too high. If the deviations can be seen only at a single level, there is a good reason to examine the test material for irregularities. If the deviations occur at various levels for various laboratories, the deviations are possibly due to a defect in the assessment experiment. Then it is necessary to examine each individual part of the assessment experiment critically in order to be able to find explanations, if possible.

A laboratory which has appeared to be outlying (either as far as internal precision or bias is concerned) shall be informed of the results of the experiments and the methodology shall be examined in order to improve the laboratory practice.

7.3.4.1.4 Different test materials shall be used in consecutive assessment experiments so that the laboratories do not develop extraordinarily good precision when working on a specific test material. Furthermore, as mentioned in 7.2.2, the material shall be sent out anonymously to guarantee that the measurements are carried out with the usual care of the laboratory.

If an assessment experiment yields results which deviate considerably from earlier experiments, it is essential to analyse all available information in order to find possible explanations for these unexpected observations.

7.3.4.2 Example: Analysis of the alkalinity of water

7.3.4.2.1 Background

In controlling the quality of water, chemical water analyses are performed in many laboratories. To be approved, these laboratories have to be assessed repeatedly. The determination of total alkalinity is considered in this example. The method is potentiometric titration. No reference materials exist for this situation, so the assessment had to take place through an assessment experiment.

Eighteen laboratories participated in the experiment in which two levels were considered and two determinations were performed at each level in each laboratory.

7.3.4.2.2 Original data

See table 11.

Table 11 — Alkalinity of water

Labora-tory	Level 1	Level 2	Labora-tory	Level 1	Level 2
1	2,040 2,040	5,250 5,300	10	2,170 2,200	5,520 5,330
2	2,100 2,110	5,460 5,460	11	1,980 1,940	4,990 5,020
3	2,070 2,070	5,240 5,200	12	2,120 2,110	5,340 5,330
4	2,070 2,090	5,308 5,292	13	2,160 2,150	5,330 5,420
5	2,740 2,610	5,850 5,850	14	2,050 2,070	5,330 5,330
6	2,086 2,182	5,305 5,325	15	2,070 2,056	5,387 5,335
7	2,128 2,076	5,296 5,346	16	2,010 2,030	5,210 5,330
8	2,060 2,080	5,340 5,340	17	2,066 2,070	5,300 5,280
9	2,060 2,080	5,310 5,300	18	2,060 2,070	5,300 5,280

7.3.4.2.3 Computation of cell means and ranges

The cell means are given in table 12 and the ranges in table 13.

Table 12 — Cell means of table 11

Laboratory	Level 1	Level 2
1	2,040	5,275
2	2,105	5,460
3	2,070	5,220
4	2,080	5,300
5	2,675	5,850
6	2,134	5,315
7	2,102	5,321
8	2,070	5,340
9	2,070	5,305
10	2,185	5,425
11	1,960	5,005
12	2,115	5,335
13	2,155	5,375
14	2,060	5,330
15	2,063	5,361
16	2,020	5,270
17	2,068	5,290
18	2,065	5,290

Table 13 — Cell ranges of table 11

Laboratory	Level	
	1	2
1	0,000	0,050
2	0,010	0,000
3	0,000	0,040
4	0,020	0,016
5	0,130	0,000
6	0,096	0,020
7	0,052	0,050
8	0,020	0,000
9	0,020	0,010
10	0,030	0,190
11	0,040	0,030
12	0,010	0,010
13	0,010	0,090
14	0,020	0,000
15	0,014	0,052
16	0,020	0,120
17	0,004	0,020
18	0,010	0,020

The previously established values of the repeatability and reproducibility standard deviations at the two levels are:

$$\sigma_{r1} = 0,023 \qquad \sigma_{r2} = 0,027$$

$$\sigma_{R1} = 0,045 \qquad \sigma_{R2} = 0,052$$

7.3.4.2.4 Assessment of internal precision

The ranges in table 13 are compared with the repeatability standard deviation using the formula:

$$w_{ij}^2 / 2\sigma_{rj}^2 < \chi_{(1-\alpha)}^2(v)/v$$

With $\alpha = 0,05$ and $v = 1$, $\chi_{0,95}^2(v)/v = 3,841$.

For level 1, the following laboratories are found to deviate:

laboratory 5: $w^2 = 0,016\ 9$ test value = 15,974

laboratory 6: $w^2 = 0,009\ 216$ test value = 8,711

For level 2, the following laboratories are found to deviate:

laboratory 10: $w^2 = 0,036\ 1$ test value = 24,76

laboratory 13: $w^2 = 0,008\ 1$ test value = 5,55

laboratory 16: $w^2 = 0,014\ 4$ test value = 9,88

7.3.4.2.5 Assessment of bias

From table 12, the between-laboratory variance is computed using the formula:

$$s^2 = \frac{1}{p-1} \sum_{i=1}^{p} n_i(\bar{y}_i - \bar{\bar{y}})^2 = \bar{\bar{n}} s_L^2 + s_r^2$$

For level 1, the following values are found:

$$n\sigma_L^2 + \sigma_r^2 = n\sigma_R^2 - (n-1)\sigma_r^2 = 0,003\ 521$$

$$s^2 = 0,044\ 36$$

test value = 12,60

With $\alpha = 0,05$ and $v = 17$, $\chi_{(1-\alpha)}^2(v)/v = 1,623$.

The furthest outlying value is found for laboratory No. 5.

Grubbs' test value for laboratory No. 5 is

$$G = (2,675 - 2,113\ 2)/0,148\ 9 = 3,77$$

This is compared with the critical 5 % value in clause 9 of ISO 5725-2:1994. For $p = 18$, this value is 2,651.

Computations with the results from laboratory No. 5 omitted give:

$$s^2 = 0,005\ 357$$

test value = 1,521

With $\alpha = 0,05$ and $v = 16$, $\chi_{(1-\alpha)}^2(v)/v = 1,644$. The conclusion is that all laboratories except laboratory No. 5 have obtained sufficiently accurate results at level 1.

For level 2, the following values are found:

$$n\sigma_L^2 + \sigma_r^2 = 0,004\ 679$$

$$s^2 = 0,050\ 34$$

test value = 10,758

With $\alpha = 0,05$ and $v = 17$, $\chi_{(1-\alpha)}^2(v)/v = 1,623$.

The furthest outlying value is found for laboratory No. 5.

Grubbs' test value for laboratory No. 5 is

$$G = (5,85 - 5,337\ 0)/0,158\ 6 = 3,235$$

The critical 5 % value is 2,651 for $p = 18$.

Computations with the results from laboratory No. 5 omitted give:

$$s^2 = 0,018\ 67$$

test value $= 3,990$

With $\alpha = 0,05$ and $v = 16$, $\chi^2_{(1-\alpha)}(v)/v = 1,644$.

The furthest outlying value is now found for laboratory No. 11.

Grubbs' test value for laboratory No. 11 is

$$G = (5,005 - 5,306\ 9)/0,096\ 61 = -3,125$$

The critical 5 % value is 2,620 for $p = 17$.

Computations with the results from laboratory No. 11 omitted give:

$$s^2 = 0,007\ 00$$

test value $= 1,496$

With $\alpha = 0,05$ and $v = 15$, $\chi^2_{(1-\alpha)}(v)/v = 1,666$.

The conclusion is that all laboratories except laboratories No. 5 and No. 11 have obtained sufficiently accurate results at level 2.

7.3.4.2.6 Conclusions

The assessment experiment has revealed that several laboratories are working with an unsatisfactory internal precision. These laboratories are Nos. 5, 6, 10, 13 and 16. A further two laboratories show a significant bias at one or both levels. These are Nos. 5 and 11. All the deviating laboratories should be informed about the result.

8 Comparison of alternative measurement methods

8.1 Origin of alternative measurement methods

An international standard method is a measurement method that has been subjected to a standardization process in order to satisfy various requirements. Among these requirements are the following.

a) It shall be applicable to a wide range of levels of characteristics to cover most materials that are internationally traded. For example, a method for the determination of total iron content in iron ores shall be applicable to as many internationally traded iron ores as possible.

b) Equipment, reagents and personnel shall be available on an international basis.

c) The cost of performing the measurement shall be acceptable.

d) The precision and trueness of the measurement method shall be acceptable for the users of the results.

These methods are usually compromises that may be too tedious to apply to routine work. A particular laboratory may find that a simpler method is sufficient for its own needs. For example, in the case where most of the materials to be measured come from the same source and the variations in their characteristics are relatively small, a simpler less expensive method may be sufficient.

Some measurement methods may be preferred in certain regions for historical reasons. In this case, an alternative international standard method may be desirable.

The comparison described in this clause is based on results from one test sample. It is strongly recommended that more than one test sample should be used for comparing precision and trueness of two measurement methods. The number of test samples required depends on various factors, such as the range of level of characteristics of interest, the sensitivity of the measurement methods to changes in the composition of the samples, etc.

8.2 Purpose of comparing measurement methods

8.2.1 Subclause 8.2 describes the procedure for comparing precision and trueness of two measurement methods where one of them (method A) is either an international standard method or a prime candidate for an international standard method. It provides evidence as to whether the two methods have different precision and/or trueness. It does not recommend which one is more suitable than the other for a particular application. This decision should be made in conjunction with other factors; i.e. cost, availability of equipment, etc.

8.2.2 Subclause 8.2 is primarily designed for the following applications.

a) In the development of an international standard method, sometimes the technical committee is faced with the problem of choosing which of the candidate methods is suitable for adoption as an international standard. Precision and trueness are

among the criteria used as the basis for this choice.

b) Sometimes it is found necessary to develop an alternative standard method. The candidate for this method should be as accurate as the first method. This comparison procedure will help to determine if the candidate method meets the requirements.

c) For some laboratories, most of the samples to be measured come from the same source. These samples have generally very much the same composition. In this situation, application of an international standard method as a routine method may be unnecessarily costly. It may be desirable for this laboratory to adopt a simpler method for routine applications. This method should produce test results with trueness and precision equal to the existing international standard method.

8.3 Method B is a candidate for an alternative standard method ("Standardization experiment" not defined)

The comparison between methods A and B shall be made on the results of precision experiments. If method A is a well-established standard method, the precision of method A can be used as the basis for comparison. If method A is itself still under development as a standard method, it shall also be subjected to a precision experiment. Both precision experiments shall be conducted in accordance with ISO 5725-2.

The objectives of the experiment are the following.

a) To determine whether method B is as precise as method A. The experimental results should be able to detect if the ratio between the precision measures of method B and method A is greater than a specified value.

b) To determine whether the trueness of method B is equal to that of method A, by showing that the difference between the grand means of the results of precision experiments involving identical samples for both methods is statistically insignificant, or showing that the difference between the certified value of a reference material and the grand mean of the test results obtained with method B in a precision experiment, using the certified reference material as test sample, is statistically insignificant.

In addition, it should be possible to detect whether the difference either between the expected values of the results of the two methods, or between the expected values of the results of each method and the certified value, is greater than a specified value.

8.4 Accuracy experiment

8.4.1 General requirements

The accuracy experiment shall be conducted in accordance with the general rules described in ISO 5725-1.

The procedures for both methods shall be documented in sufficient detail so as to avoid misinterpretation by the participating laboratories. No modification to the procedure is permitted during the experiment.

The participating laboratories shall be a representative sample of potential users of the method.

8.4.2 Test samples

The precision of many measurement methods is affected by the matrix of the test sample as well as the level of the characteristic. For these methods, comparison of the precision is best done on identical test samples. Furthermore, comparison of the trueness of the methods can only be made when identical test samples are used. For this reason, communication between the working groups who conduct the accuracy experiments on each method should be achieved by appointment of a common executive officer.

The main requirement for a test sample is that it shall be homogeneous; i.e. each laboratory shall use identical test samples. If within-unit inhomogeneity is suspected, clear instructions on the method of taking test portions shall be included in the document. The use of reference materials (RMs) for some of the test samples has some advantages. The homogeneity of the RM has been assured and the results of the method can be examined for bias relative to the certified value of the RM. The drawback is usually the high cost of the RM. In many cases, this can be overcome by redividing the RM units. For the procedure for using a RM as a test sample, see ISO Guide 33.

8.4.3 Number of test samples

The number of test samples used varies depending on the range of the characteristic levels of interest,

and on the dependency of the accuracy on the level. In many cases, the number of test samples is limited by the amount of work involved and the availability of a test sample at the desired level.

8.4.4 Number of laboratories and number of measurements

8.4.4.1 General

The number of laboratories and the number of measurements per laboratory required for the inter-laboratory test programme for both methods depend on:

a) precisions of the two methods;

b) detectable ratio, ϱ or ϕ, between the precision measures of the two methods; this is the minimum ratio of precision measures that the experimenter wishes to detect with high probability from the results of experiments using two methods; the precision may be expressed either as the repeatability standard deviation, in which case the ratio is termed ϱ, or as the square root of the between-laboratory mean squares, in which case the ratio is termed ϕ;

c) detectable difference between the biases of the two methods, λ; this is the minimum value of the difference between the expected values of the results obtained by the two methods.

It is recommended that a significance level of $\alpha = 0{,}05$ is used to compare precision estimates and that the risk of failing to detect the chosen minimum ratio of standard deviations, or the minimum difference between the biases, is set at $\beta = 0{,}05$.

With those values of α and β, the following equation can be used for the detectable difference:

$$\lambda = 4\sqrt{\left(\sigma_{LA}^2 + \sigma_{rA}^2/n_A\right)/p_A + \left(\sigma_{LB}^2 + \sigma_{rB}^2/n_B\right)/p_B}$$

$$\ldots (13)$$

where the subscripts A and B refer to method A and method B, respectively.

In most cases, the precision of method B is unknown. In this case, use the precision of method A as a substitute to give

$$\lambda = 4\sqrt{\left(\sigma_{LA}^2 + \sigma_{rA}^2/n_A\right)/p_A + \left(\sigma_{LA}^2 + \sigma_{rA}^2/n_B\right)/p_B}$$

$$\ldots (14)$$

The experimenter should try substituting values of n_A, n_B, p_A and p_B in equation (13) or (14) until values are found which are large enough to satisfy the equation. The values of these parameters which are needed to give an adequate experiment to compare precision estimates should then be considered.

Table 14 shows the minimum ratios of standard deviation for given values of α and β as a function of the degrees of freedom v_A and v_B.

For repeatability standard deviations

$$v_A = p_A(n_A - 1) \text{ and } v_B = p_B(n_B - 1)$$

For between-laboratory mean squares

$$v_A = p_A - 1 \text{ and } v_B = p_B - 1$$

If the precision of one of the methods is well established, use degrees of freedom equal to 200 from table 14.

8.4.4.2 Example: Determination of iron in iron ores

8.4.4.2.1 Background

Two analytical methods for the determination of the total iron in iron ores are investigated. They are presumed to have equal precision:

$$\sigma_{rA} = \sigma_{rB} = 0{,}1 \text{ \% Fe}$$

$$\sigma_{LA} = \sigma_{LB} = 0{,}2 \text{ \% Fe}$$

8.4.4.2.2 Requirements

$$\lambda = 0{,}4 \text{ \% Fe}$$

$$\varrho = \phi = 4$$

The minimum number of laboratories required for each interlaboratory test programme are computed assuming equal numbers of laboratories and duplicate analyses:

$$p_A = p_B \text{ and } n_A = n_B = 2$$

a) For the trueness requirement:

$$0{,}4 = 4\sqrt{\left(0{,}2^2 + 0{,}1^2/2\right)/p_A + \left(0{,}2^2 + 0{,}1^2/2\right)/p_B}$$

hence

$$p_A = p_B = 9$$

b) For the precision requirement:

From table 14 it can be seen that $\varrho = 4$ or $\phi = 4$ is given by $v_A = v_B = 9$.

To compare repeatability standard deviations, $v_A = p_A$ and $v_B = p_B$, so $p_A = p_B = 9$.

To compare between-laboratory mean squares, $v_A = p_A - 1$ and $v_B = p_B - 1$, so $p_A = p_B = 10$.

8.4.4.2.3 Conclusions

The minimum number of participating laboratories required for each interlaboratory test programme is 10.

8.4.5 Test sample distribution

The executive officer of the interlaboratory test programme shall take the final responsibility for obtaining, preparing and distributing the test samples. Precautions shall be taken to ensure that the samples are received by the participating laboratories in good condition and are clearly identified. The participating laboratories shall be instructed to analyse the samples on the same basis, for example, on dry basis; i.e. the sample is to be dried at 105 °C for x h before weighing.

8.4.6 Participating laboratory

The participating laboratory shall assign a staff member to be responsible for organizing the execution of the instructions of the coordinator. The staff member shall be a qualified analyst. Unusually skilled staff (such as a research personnel or the "best" operator) should be avoided in order to prevent obtaining an unrealistically low estimate of the standard deviation of the method. The assigned staff member shall perform the required number of measurements under repeatability conditions. The laboratory is responsible for reporting the test results to the coordinator within the time specified.

Table 14 — Values of $\varrho(v_A, v_B, \alpha, \beta)$ or $\phi(v_A, v_B, \alpha, \beta)$ for $\alpha = 0,05$ and $\beta = 0,05$

v_B	v_A																	
	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	25	50	200
6	5,82	5,40	5,10	4,88	4,72	4,58	4,47	4,38	4,31	4,24	4,19	4,14	4,09	4,06	4,02	3,89	3,65	3,47
7	5,40	4,99	4,71	4,50	4,34	4,21	4,10	4,01	3,94	3,88	3,82	3,78	3,74	3,70	3,67	3,54	3,30	3,13
8	5,10	4,71	4,43	4,23	4,07	3,94	3,84	3,76	3,68	3,62	3,57	3,52	3,48	3,45	3,41	3,29	3,06	2,89
9	4,88	4,50	4,23	4,03	3,87	3,75	3,65	3,56	3,49	3,43	3,38	3,33	3,29	3,26	3,23	3,11	2,88	2,71
10	4,72	4,34	4,07	3,87	3,72	3,59	3,50	3,41	3,34	3,28	3,23	3,19	3,15	3,11	3,08	2,96	2,73	2,57
11	4,58	4,21	3,94	3,75	3,59	3,47	3,38	3,29	3,22	3,16	3,11	3,07	3,03	2,99	2,96	2,85	2,62	2,45
12	4,47	4,10	3,84	3,65	3,50	3,38	3,28	3,20	3,13	3,07	3,02	2,97	2,93	2,90	2,87	2,75	2,52	2,36
13	4,38	4,01	3,76	3,56	3,41	3,29	3,20	3,12	3,05	2,99	2,94	2,89	2,85	2,82	2,79	2,67	2,44	2,28
14	4,31	3,94	3,68	3,49	3,34	3,22	3,13	3,05	2,98	2,92	2,87	2,83	2,79	2,75	2,72	2,60	2,38	2,21
15	4,24	3,88	3,62	3,43	3,28	3,16	3,07	2,99	2,92	2,86	2,81	2,77	2,73	2,69	2,66	2,55	2,32	2,15
16	4,19	3,82	3,57	3,38	3,23	3,11	3,02	2,94	2,87	2,81	2,76	2,72	2,68	2,64	2,61	2,50	2,27	2,10
17	4,14	3,78	3,52	3,33	3,19	3,07	2,97	2,89	2,83	2,77	2,72	2,67	2,63	2,60	2,57	2,45	2,22	2,05
18	4,09	3,74	3,48	3,29	3,15	3,03	2,93	2,85	2,79	2,73	2,68	2,63	2,60	2,56	2,53	2,41	2,18	2,01
19	4,06	3,70	3,45	3,26	3,11	2,99	2,90	2,82	2,75	2,69	2,64	2,60	2,56	2,53	2,50	2,38	2,15	1,98
20	4,02	3,67	3,41	3,23	3,08	2,96	2,87	2,79	2,72	2,66	2,61	2,57	2,53	2,50	2,46	2,35	2,12	1,95
25	3,89	3,54	3,29	3,11	2,96	2,85	2,75	2,67	2,60	2,55	2,50	2,45	2,41	2,38	2,35	2,23	2,00	1,82
50	3,65	3,30	3,06	2,88	2,73	2,62	2,52	2,44	2,38	2,32	2,27	2,22	2,18	2,15	2,12	2,00	1,75	1,56
200	3,47	3,13	2,89	2,71	2,57	2,45	2,36	2,28	2,21	2,15	2,10	2,05	2,01	1,98	1,95	1,82	1,56	1,32

NOTES

1 $\varrho = \dfrac{\sigma_{rB}}{\sigma_{rA}}$; $v_A = p_A(n_A - 1)$; $v_B = p_B(n_B - 1)$

2 $\phi = \sqrt{\left[\dfrac{n_B \sigma_{LB}^2 + \sigma_{rB}^2}{n_A \sigma_{LA}^2 + \sigma_{rA}^2} \right]}$; $v_A = p_A - 1$; $v_B = p_B - 1$

8.4.7 Collection of test results

The coordinator of the test programme for each method is responsible for collecting all the test results within a reasonable time.

It is his/her responsibility to scrutinize the test results for physical aberrants. These are test results that due to explainable physical causes do not belong to the same distribution as the other test results.

8.4.8 Evaluation of test results

The test results shall be evaluated by a qualified statistician using the procedure described in ISO 5725-2. For each test sample, the following quantities are to be computed:

s_{rA} estimate of the repeatability standard deviation for method A

s_{rB} estimate of the repeatability standard deviation for method B

s_{RA} estimate of the reproducibility standard deviation for method A

s_{RB} estimate of the reproducibility standard deviation for method B

$\bar{\bar{y}}_A$ grand mean for method A

$\bar{\bar{y}}_B$ grand mean for method B

8.4.9 Comparison between results of method A and method B

The results of the interlaboratory test programmes shall be compared for each level. It is possible that method B is more precise and/or biased at lower levels of the characteristic but less precise and/or biased at higher levels of the characteristic values or vice versa.

8.4.9.1 Graphical presentation

Graphical presentation of the raw data for each level is desirable. Sometimes the difference between the results of the two methods in terms of precision and/or bias is so obvious that further statistical evaluation is unnecessary.

Graphical presentation of the precision and grand means of all levels is also desirable.

8.4.9.2 Comparison of precision

8.4.9.2.1 Method A is an established standard method

The precision of method A is well established.

a) **Within-laboratory precision**

If

$$\frac{s_{rB}^2}{\sigma_{rA}^2} \leqslant \frac{\chi_{(1-\alpha)}^2(v_{rB})}{v_{rB}}$$

there is no evidence that the within-laboratory precision of method B is not as good as that of method A;

if

$$\frac{s_{rB}^2}{\sigma_{rA}^2} > \frac{\chi_{(1-\alpha)}^2(v_{rB})}{v_{rB}}$$

there is evidence that the within-laboratory precision of method B is poorer than that of method A.

$\chi_{(1-\alpha)}^2(v_{rB})$ is the $(1-\alpha)$-quantile of the χ^2 distribution with v_{rB} degrees of freedom, and

$$v_{rB} = p_B(n_B - 1)$$

b) **Overall precision**

If

$$\frac{s_{RB}^2 - (1 - 1/n_B)s_{rB}^2}{\sigma_{RA}^2 - (1 - 1/n_B)\sigma_{rA}^2} \leqslant \frac{\chi_{(1-\alpha)}^2(v_{LB})}{v_{LB}}$$

there is no evidence that the mean square of method B is not as good as that of method A;

if

$$\frac{s_{RB}^2 - (1 - 1/n_B)s_{rB}^2}{\sigma_{RA}^2 - (1 - 1/n_B)\sigma_{rA}^2} > \frac{\chi_{(1-\alpha)}^2(v_{LB})}{v_{LB}}$$

there is evidence that the mean square of method B is not as good as that of method A.

$\chi_{(1-\alpha)}^2(v_{LB})$ is the $(1-\alpha)$-quantile of the χ^2 distribution with v_{LB} degrees of freedom, and

$$v_{LB} = p_B - 1$$

8.4.9.2.2 Both methods are new candidate standard methods

a) **Within-laboratory precision**

$$F_r = \frac{s_{rB}^2}{s_{rA}^2}$$

If

$$F_{\alpha/2}(v_{rA}, v_{rB}) \leqslant F_r \leqslant F_{(1 - \alpha/2)}(v_{rA}, v_{rB})$$

there is no evidence that the methods have different within-laboratory precisions;

if

$$F_r < F_{\alpha/2}(v_{rA}, v_{rB})$$

there is evidence that method B has better within-laboratory precision than method A;

if

$$F_r > F_{(1 - \alpha/2)}(v_{rB}, v_{rA})$$

there is evidence that method B has poorer within-laboratory precision than method A.

$F_{\alpha/2}(v_{rA}, v_{rB})$ and $F_{(1 - \alpha/2)}(v_{rA}, v_{rB})$ are the $\alpha/2$- and $(1 - \alpha/2)$-quantiles of the F distribution with degrees of freedom of numerator v_{rA} and denominator v_{rB}

$$v_{rA} = p_A(n_A - 1)$$

$$v_{rB} = p_B(n_B - 1)$$

b) **Overall precision**

$$F_R = \frac{s_{RB}^2 - (1 - 1/n_B)s_{rB}^2}{s_{RA}^2 - (1 - 1/n_A)s_{rA}^2}$$

If

$$F_{\alpha/2}(v_{RB}, v_{RA}) \leqslant F_R \leqslant F_{(1 - \alpha/2)}(v_{RB}, v_{RA})$$

there is no evidence that the methods have different between-laboratories precisions;

if

$$F_R < F_{\alpha/2}(v_{RB}, v_{RA})$$

there is evidence that method B has better overall precision than method A;

if

$$F_R > F_{(1 - \alpha/2)}(v_{RB}, v_{RA})$$

there is evidence that method B has poorer overall precision than method A.

$F_{\alpha/2}(v_{RB}, v_{RA})$ and $F_{(1 - \alpha/2)}(v_{RB}, v_{RA})$ are the $\alpha/2$- and $(1 - \alpha/2)$-quantiles of the F distribution with degrees of freedom of numerator v_{RB} and denominator v_{RA}, and

$$v_{LA} = p_A - 1$$

$$v_{LB} = p_B - 1$$

NOTE 5   Many tables list only the $(1 - \alpha/2)$-quantiles of the F distribution. In this case, the following relationships can be used to find the $\alpha/2$-quantiles:

$$F_{\alpha/2}(v_{rB}, v_{rA}) = 1/F_{(1 - \alpha/2)}(v_{rA}, v_{rB})$$

$$F_{\alpha/2}(v_{RB}, v_{RA}) = 1/F_{(1 - \alpha/2)}(v_{RA}, v_{RB})$$

8.4.9.3 Comparison of trueness

8.4.9.3.1 Comparison of the mean with the certified value of an RM

The grand mean of each method can be compared with the certified value of the RM used as one of the test samples. The following test may be used:

a)   if

$$|\mu - \bar{\bar{y}}| \leqslant 2\sqrt{\left[s_{RB}^2 - (1 - 1/n_B)s_{rB}^2\right]/p_B}$$

the difference between the grand mean of the results of the method and the certified value is statistically insignificant;

b)   if

$$|\mu - \bar{\bar{y}}| > 2\sqrt{\left[s_{RB}^2 - (1 - 1/n_B)s_{rB}^2\right]/p_B}$$

the difference between the grand mean of the results of the method and the certified value is statistically significant.

There are two possibilities:

1)   if

$$|\mu - \bar{\bar{y}}| \leqslant \delta_m/2$$

there is no evidence that the measurement method is unacceptably biased; or

2)   if

$$|\mu - \bar{\bar{y}}| > \delta_m/2$$

there is evidence that the measurement method is unacceptably biased;

where δ_m is the minimum difference between the expected value of the results of the method and the certified value of the reference material that the experimenter wishes to detect with high probability from the results of an experiment.

8.4.9.3.2 Comparison between the means of method A and method B

a) If

$$\left| \frac{\bar{\bar{y}}_A - \bar{\bar{y}}_B}{s} \right| \leqslant 2,0$$

the difference between the means of method A and method B is statistically insignificant;

b) if

$$\left| \frac{\bar{\bar{y}}_A - \bar{\bar{y}}_B}{s} \right| > 2,0$$

the difference between the means of method A and method B is statistically significant;

where

$$s = \sqrt{s_A^2 + s_B^2}$$

$$s_A^2 = \left[s_{RA}^2 - (1 - 1/n_A) s_{rA}^2 \right] / p_A$$

$$s_B^2 = \left[s_{RB}^2 - (1 - 1/n_B) s_{rB}^2 \right] / p_B$$

There are two possibilities:

1) if

$$|\bar{\bar{y}}_A - \bar{\bar{y}}_B| \leqslant \lambda/2$$

there is no evidence that the difference between the biases of the two methods is unacceptable;

2) if

$$|\bar{\bar{y}}_A - \bar{\bar{y}}_B| > \lambda/2$$

there is evidence that the difference between the biases of the two methods is unacceptable;

where λ is the detectable difference between the biases.

8.5 Method B is a candidate for a routine method

8.5.1 Parameters

The parameters of interest for a routine laboratory method are the long-term mean μ_t, the precision under repeatability conditions (expressed as the repeatability standard deviation σ_r) and the intermediate precision (expressed as the time-different intermediate precision standard deviation $\sigma_{I(T)}$).

To estimate these parameters, the laboratory shall conduct a quasi-interlaboratory test programme, replacing the participating laboratories by "time" (see ISO 5725-3). The mathematical model used to represent this quasi-interlaboratory test programme is the same as that used for an interlaboratory programme, replacing the subscript L by T (laboratory by time). In this case, the time-different variation includes variation due to various changes that normally occur in the laboratory, such as calibration of equipment, different reagents, different analysts, ambient conditions, etc. The quasi-interlaboratory programme should therefore cover the duration that normally covers these changes. The procedures for comparing the precision are the same as those described in 8.4.9.3.

The bias can be determined by applying each method to a certified reference material, where μ is the accepted value of the reference material.

8.5.2 Long-term bias test

Compute the long-term arithmetic mean

$$\bar{y} = \sum_{i=1}^{p_{tB}} \sum_{j=1}^{n_B} \frac{y_{ij}}{n_B p_{tB}}$$

where i and j are indices associated with long-term (intermediate precision) and short-term (repeatability condition) measurements respectively.

a) If

$$|\bar{y} - \mu_t| \leqslant 2\sqrt{\left(s_{tB}^2 + \frac{s_{rB}^2}{n_B} \right) / p_{tB}}$$

the difference between the long-term mean and the accepted value is statistically insignificant;

b) if

$$|\bar{y} - \mu_t| > 2\sqrt{\left(s_{tB}^2 + \frac{s_{rB}^2}{n_B} \right) / p_{tB}}$$

the difference between the long-term mean and the accepted value is statistically significant.

There are two possibilities:

1) If

$$|\bar{\bar{y}} - \mu_t| \leqslant \delta_m/2$$

there is no evidence that the long-term bias of the method is unacceptable;

2) if

$$|\bar{\bar{y}} - \mu_t| > \delta_m/2$$

there is evidence that the long-term bias of the method is unacceptable;

where δ_m is the long-term detectable difference preset by the experimenter.

Annex A

(normative)

Symbols and abbreviations used in ISO 5725

a	Intercept in the relationship $s = a + bm$
A	Factor used to calculate the uncertainty of an estimate
b	Slope in the relationship $s = a + bm$
B	Component in a test result representing the deviation of a laboratory from the general average (laboratory component of bias)
B_0	Component of B representing all factors that do not change in intermediate precision conditions
$B_{(1)}$, $B_{(2)}$, etc.	Components of B representing factors that vary in intermediate precision conditions
c	Intercept in the relationship $\lg s = c + d \lg m$
C, C', C''	Test statistics
C_{crit}, C'_{crit}, C''_{crit}	Critical values for statistical tests
CD_P	Critical difference for probability P
CR_P	Critical range for probability P
d	Slope in the relationship $\lg s = c + d \lg m$
e	Component in a test result representing the random error occurring in every test result
f	Critical range factor
$F_p(v_1, v_2)$	p-quantile of the F-distribution with v_1 and v_2 degrees of freedom
G	Grubbs' test statistic
h	Mandel's between-laboratory consistency test statistic

k	Mandel's within-laboratory consistency test statistic
LCL	Lower control limit (either action limit or warning limit)
m	General mean of the test property; level
M	Number of factors considered in intermediate precision conditions
N	Number of iterations
n	Number of test results obtained in one laboratory at one level (i.e. per cell)
p	Number of laboratories participating in the interlaboratory experiment
P	Probability
q	Number of levels of the test property in the interlaboratory experiment
r	Repeatability limit
R	Reproducibility limit
RM	Reference material
s	Estimate of a standard deviation
$\hat{s}$	Predicted standard deviation
T	Total or sum of some expression
t	Number of test objects or groups
UCL	Upper control limit (either action limit or warning limit)
W	Weighting factor used in calculating a weighted regression
w	Range of a set of test results
x	Datum used for Grubbs' test
y	Test result

© ISO

$\bar{y}$	Arithmetic mean of test results	
$\bar{\bar{y}}$	Grand mean of test results	
α	Significance level	
β	Type II error probability	
γ	Ratio of the reproducibility standard deviation to the repeatability standard deviation (σ_R/σ_r)	
Δ	Laboratory bias	
$\hat{\Delta}$	Estimate of Δ	
δ	Bias of the measurement method	
$\hat{\delta}$	Estimate of δ	
λ	Detectable difference between two laboratory biases or the biases of two measurement methods	
μ	True value or accepted reference value of a test property	
ν	Number of degrees of freedom	
ϱ	Detectable ratio between the repeatability standard deviations of method B and method A	
σ	True value of a standard deviation	
τ	Component in a test result representing the variation due to time since last calibration	
ϕ	Detectable ratio between the square roots of the between-laboratory mean squares of method B and method A	
$\chi_p^2(\nu)$	p-quantile of the χ^2-distribution with ν degrees of freedom	

Symbols used as subscripts

C	Calibration-different
E	Equipment-different
i	Identifier for a particular laboratory
$I(\ )$	Identifier for intermediate measures of precision; in brackets, identification of the type of intermediate situation
j	Identifier for a particular level (ISO 5725-2). Identifier for a group of tests or for a factor (ISO 5725-3)
k	Identifier for a particular test result in a laboratory i at level j
L	Between-laboratory (interlaboratory)
m	Identifier for detectable bias
M	Between-test-sample
O	Operator-different
P	Probability
r	Repeatability
R	Reproducibility
T	Time-different
W	Within-laboratory (intralaboratory)
1, 2, 3...	For test results, numbering in the order of obtaining them
(1), (2), (3)...	For test results, numbering in the order of increasing magnitude

ISO 5725-6:1994(E)

ICS 03.120.30

Descriptors: measurement, tests, test results, accuracy, reproducibility, statistical analysis.

Price based on 41 pages