INTERNATIONAL STANDARD

ISO 5725-2

First edition
1994-12-15

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

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

Partie 2: Méthode de base pour la détermination de la répétabilité et de la reproductibilité d'une méthode de mesure normalisée

Reference number
ISO 5725-2:1994(E)

Contents

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

Printed in Switzerland

Annexes

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 5725-2 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. Annexes B and C are for information only.

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 General consideration of these quantities is given in ISO 5725-1 and so is not repeated in this part of ISO 5725. ISO 5725-1 should be read in conjunction with all other parts of ISO 5725, including this part, because it gives the underlying definitions and general principles.

0.3 This part of ISO 5725 is concerned solely with estimating by means of the repeatability standard deviation and reproducibility standard deviation. Although other types of experiment (such as the split-level experiment) are used in certain circumstances for the estimation of precision, they are not dealt with in this part of ISO 5725 but rather are the subject of ISO 5725-5. Nor does this part of ISO 5725 consider any other measures of precision intermediate between the two principal measures; those are the subject of ISO 5725-3.

0.4 In certain circumstances, the data obtained from an experiment carried out to estimate precision are used also to estimate trueness. The estimation of trueness is not considered in this part of ISO 5725; all aspects of the estimation of trueness are the subject of ISO 5725-4.

This page intentionally left blank

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

1 Scope

1.1 This part of ISO 5725

— amplifies the general principles to be observed in designing experiments for the numerical estimation of the precision of measurement methods by means of a collaborative interlaboratory experiment;

— provides a detailed practical description of the basic method for routine use in estimating the precision of measurement methods;

— provides guidance to all personnel concerned with designing, performing or analysing the results of the tests for estimating precision.

NOTE 1 Modifications to this basic method for particular purposes are given in other parts of ISO 5725.

Annex B provides practical examples of estimating the precision of measurement methods by experiment.

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

1.3 It assumes that in the design and performance of the precision experiment, all the principles as laid down in ISO 5725-1 have been observed. The basic method uses the same number of test results in each laboratory, with each laboratory analysing the same levels of test sample; i.e. a balanced uniform-level experiment. The basic method applies to procedures that have been standardized and are in regular use in a number of laboratories.

NOTE 2 Worked examples are given to demonstrate balanced uniform sets of test results, although in one example a variable number of replicates per cell were reported (unbalanced design) and in another some data were missing. This is because an experiment designed to be balanced can turn out to be unbalanced. Stragglers and outliers are also considered.

1.4 The statistical model of clause 5 of ISO 5725-1:1994 is accepted as a suitable basis for the interpretation and analysis of the test results, the distribution of which is approximately normal.

1.5 The basic method, as described in this part of ISO 5725, will (usually) estimate the precision of a measurement method:

a) when it is required to determine the repeatability and reproducibility standard deviations as defined in ISO 5725-1;

b) when the materials to be used are homogeneous, or when the effects of heterogeneity can be included in the precision values; and

c) when the use of a balanced uniform-level layout is acceptable.

1.6 The same approach can be used to make a preliminary estimate of precision for measurement methods which have not reached standardization or are not in routine use.

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.*

3 Definitions

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

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

4 Estimates of the parameters in the basic model

4.1 The procedures given in this part of ISO 5725 are based on the statistical model given in clause 5 of ISO 5725-1:1994 and elaborated upon in subclause 1.2 of ISO 5725-1:1994. In particular, these procedures are based on equations (2) to (6) of clause 5 of ISO 5725-1:1994.

The model is

$$y = m + B + e$$

where, for the particular material tested,

m is the general mean (expectation);

B is the laboratory component of bias under repeatability conditions;

e is the random error occurring in every measurement under repeatability conditions.

4.2 Equations (2) to (6) of ISO 5725-1:1994, clause 5 are expressed in terms of the true standard deviations of the populations considered. In practice, the exact values of these standard deviations are not known, and estimates of precision values must be made from a relatively small sample of all the possible laboratories, and within those laboratories from a small sample of all the possible test results.

4.3 In statistical practice, where the true value of a standard deviation, σ, is not known and is replaced by an estimate based upon a sample, then the symbol σ is replaced by s to denote that it is an estimate. This has to be done in each of the equations (2) to (6) of ISO 5725-1:1994, giving:

s_L^2 is the estimate of the between-laboratory variance;

s_W^2 is the estimate of the within-laboratory variance;

s_r^2 is the arithmetic mean of s_W^2 and is the estimate of the repeatability variance; this arithmetic mean is taken over all those laboratories taking part in the accuracy experiment which remain after outliers have been excluded;

s_R^2 is the estimate of the reproducibility variance:

$$s_R^2 = s_L^2 + s_r^2 \qquad \ldots (1)$$

5 Requirements for a precision experiment

5.1 Layout of the experiment

5.1.1 In the layout used in the basic method, samples from q batches of materials, representing q different levels of the test, are sent to p laboratories which each obtain exactly n replicate test results under repeatability conditions at each of the q levels. This type of experiment is called a balanced uniform-level experiment.

5.1.2 The performance of these measurements shall be organized and instructions issued as follows.

a) Any preliminary checking of equipment shall be as specified in the standard method.

b) Each group of n measurements belonging to one level shall be carried out under repeatability conditions, i.e. within a short interval of time and by the same operator, and without any intermediate recalibration of the apparatus unless this is an integral part of performing a measurement.

c) It is essential that a group of n tests under repeatability conditions be performed independently as if they were n tests on different materials. As a rule, however, the operator will know that he/she is testing identical material, but the point should be stressed in the instructions that the whole purpose of the experiment is to determine what differences in results can occur in actual testing. If it is feared that, despite this warning, previous results may influence subsequent test results and thus the repeatability variance, it should be considered whether to use n separate samples at each of the q levels, coded in such a way that the operator will not know which are the replicates for a given level. However, such a procedure could cause problems in ensuring that repeatability conditions will apply between replicates. This would only be possible if the measurements were of such a nature that all the qn measurements could be performed within a short interval of time.

d) It is not essential that all the q groups of n measurements each be performed strictly within a short interval; different groups of measurements may be carried out on different days.

e) Measurements of all q levels shall be performed by one and the same operator and, in addition, the n measurements at a given level shall be performed using the same equipment throughout.

f) If in the course of the measurements an operator should become unavailable, another operator may complete the measurements, provided that the change does not occur within a group of n measurements at one level but only occurs between two of the q groups. Any such change shall be reported with the results.

g) A time limit shall be given within which all measurements shall be completed. This may be necessary to limit the time allowed to elapse between the day the samples are received and the day the measurements are performed.

h) All samples shall be clearly labelled with the name of the experiment and a sample identification.

5.1.3 In 5.1.2 and elsewhere in this part of ISO 5725, reference is made to the operator. For some measurements, there may in fact be a team of operators, each of whom performs some specific part of the procedure. In such a case, the team shall be regarded as "the operator" and any change in the team shall be regarded as providing a different "operator".

5.1.4 In commercial practice, the test results may be rounded rather crudely, but in a precision experiment test results shall be reported to at least one more digit than specified in the standard method. If the method does not specify the number of digits, the rounding shall not be coarser than half the repeatability standard deviation estimate. When precision may depend on the level m, different degrees of rounding may be needed for different levels.

5.2 Recruitment of the laboratories

5.2.1 The general principles regarding recruitment of the laboratories to participate in an interlaboratory experiment are given in 6.3 of ISO 5725-1:1994. In enlisting the cooperation of the requisite number of laboratories, their responsibilities shall be clearly stated. An example of a suitable enlistment questionnaire is given in figure 1.

5.2.2 For the purposes of this part of ISO 5725, a "laboratory" is considered to be a combination of the operator, the equipment and the test site. One test site (or laboratory in the conventional sense) may thus produce several "laboratories" if it can provide several operators each with independent sets of equipment and situations in which to perform the work.

5.3 Preparation of the materials

5.3.1 A discussion of the points that need to be considered when selecting materials for use in a precision experiment is given in 6.4 of ISO 5725-1:1994.

5.3.2 When deciding on the quantities of material to be provided, allowance shall be made for accidental spillage or errors in obtaining some test results which may necessitate using extra material. The amount of material prepared shall be sufficient to cover the experiment and allow an adequate stock in reserve.

Questionnaire for interlaboratory study

Title of measurement method (copy attached) .....

1. Our laboratory is willing to participate in the precision experiment for this standard measurement method.

YES ☐ NO ☐ (tick appropriate box)

2. As a participant, we understand that:

a) all essential apparatus, chemicals and other requirements specified in the method must be available in our laboratory when the programme begins;

b) specified "timing" requirements such as starting date, order of testing specimens and finishing date of the programme must be rigidly met;

c) the method must be strictly adhered to;

d) samples must be handled in accordance with instructions;

e) a qualified operator must perform the measurements.

Having studied the method and having made a fair appraisal of our capabilities and facilities, we feel that we will be adequately prepared for cooperative testing of this method.

3. Comments

(Signed) .....

(Company or laboratory) .....

Figure 1 — Enlistment questionnaire for interlaboratory study

5.3.3 It should be considered whether it is desirable for some laboratories to obtain some preliminary test results for familiarization with the measurement method before obtaining the official test result and, if so, whether additional material (not precision experiment samples) should be provided for this purpose.

5.3.4 When a material has to be homogenized, this shall be done in the manner most appropriate for that material. When the material to be tested is not homogeneous, it is important to prepare the samples in the manner specified in the method, preferably starting with one batch of commercial material for each level. In the case of unstable materials, special instructions on storage and treatment shall be specified.

5.3.5 For the samples at each level, n separate containers shall be used for each laboratory if there is any danger of the materials deteriorating once the container has been opened (e.g. by oxidation, by losing volatile components, or with hygroscopic material). In the case of unstable materials, special instructions on storage and treatment shall be specified. Precautions may be needed to ensure that samples remain identical up to the time the measurements are made. If the material to be measured consists of a mixture of powders of different relative density or of different grain size, some care is needed because segregation may result from shaking, for example during transport. When reaction with the atmosphere may be expected, the specimens may be sealed into ampoules, either evacuated or filled with an inert gas. For perishable materials such as food or blood samples, it

may be necessary to send them in a deep-frozen state to the participating laboratories with detailed instructions for the procedure for thawing.

6 Personnel involved in a precision experiment

NOTE 3 The methods of operation within different laboratories are not expected to be identical. Therefore the contents of this clause are only intended as a guide to be modified as appropriate to cater for a particular situation.

6.1 Panel

6.1.1 The panel should consist of experts familiar with the measurement method and its application.

6.1.2 The tasks of the panel are:

a) to plan and coordinate the experiment;

b) to decide on the number of laboratories, levels and measurements to be made, and the number of significant figures to be required;

c) to appoint someone for the statistical functions (see 6.2);

d) to appoint someone for the executive functions (see 6.3);

e) to consider the instructions to be issued to the laboratory supervisors in addition to the standard measurement method;

f) to decide whether some operators may be allowed to carry out a few unofficial measurements in order to regain experience of the method after a long interval (such measurements shall never be carried out on the official collaborative samples);

g) to discuss the report of the statistical analysis on completion of the analysis of the test results;

h) to establish final values for the repeatability standard deviation and the reproducibility standard deviation;

i) to decide if further actions are required to improve the standard for the measurement method or with regard to laboratories whose test results have been rejected as outliers.

6.2 Statistical functions

At least one member of the panel should have experience in statistical design and analysis of experiments. His/her tasks are:

a) to contribute his/her specialized knowledge in designing the experiment;

b) to analyse the data;

c) to write a report for submission to the panel following the instructions contained in 7.7.

6.3 Executive functions

6.3.1 The actual organization of the experiment should be entrusted to a single laboratory. A member of the staff of that laboratory should take full responsibility; he/she is called the executive officer and is appointed by the panel.

6.3.2 The tasks of the executive officer are:

a) to enlist the cooperation of the requisite number of laboratories and to ensure that supervisors are appointed;

b) to organize and supervise the preparation of the materials and samples and the dispatch of the samples; for each level, an adequate quantity of material should be set aside as a reserve stock;

c) to draft instructions covering all the points in 5.1.2 a) to h), and circulate them to the supervisors early enough in advance for them to raise any comments or queries and to ensure that operators selected are those who would normally carry out such measurements in routine operations;

d) to design suitable forms for the operator to use as a working record and for the supervisor to report the test results to the requisite number of significant figures (such forms may include the name of the operator, the dates on which samples were received and measured, the equipment used and any other relevant information);

e) to deal with any queries from laboratories regarding the performance of the measurements;

f) to see that an overall time schedule is maintained;

g) to collect the data forms and present them to the statistical expert.

6.4 Supervisors

6.4.1 A staff member in each of the participating laboratories should be made responsible for organizing the actual performance of the measurements, in keeping with instructions received from the executive officer, and for reporting the test results.

6.4.2 The tasks of the supervisor are:

a) ensure that the operators selected are those who would normally carry out such measurements in routine operations;
b) to hand out the samples to the operator(s) in keeping with the instructions of the executive officer (and to provide material for familiarization experiments, if necessary);
c) to supervise the execution of the measurements (the supervisor shall not take part in performing the measurements);
d) to ensure that the operators carry out the required number of measurements;
e) to ensure adherence to the set timetable for performing the measurements;
f) to collect the test results recorded to the agreed number of decimal places, including any anomalies and difficulties experienced, and comments made by the operators.

6.4.3 The supervisor of each laboratory should write a full report which should contain the following information:

a) the test results, entered legibly by their originator on the forms provided, not transcribed or typed (computer or testing machine printout may be acceptable as an alternative);
b) the original observed values or readings (if any) from which the test results were derived, entered legibly by the operator on the forms provided, not transcribed or typed;
c) comments by the operators on the standard for the measurement method;
d) information about irregularities or disturbances that may have occurred during the measurements, including any change of operator that may have occurred, together with a statement as to which measurements were performed by which operator, and the reasons for any missing results;
e) the date(s) on which the samples were received;
f) the date(s) on which each sample was measured;
g) information about the equipment used, if relevant;
h) any other relevant information.

6.5 Operators

6.5.1 In each laboratory the measurements shall be carried out by one operator selected as being representative of those likely to perform the measurements in normal operations.

6.5.2 Because the object of the experiment is to determine the precision obtainable by the general population of operators working from the standard measurement method, in general the operators should not be given amplifications to the standard for the measurement method. However, it should be pointed out to the operators that the purpose of the exercise is to discover the extent to which results can vary in practice, so that there will be less temptation for them to discard or rework results that they feel are inconsistent.

6.5.3 Although normally the operators should receive no supplementary amplifications to the standard measurement method, they should be encouraged to comment on the standard and, in particular, to state whether the instructions contained in it are sufficiently unambiguous and clear.

6.5.4 The tasks of the operators are:

a) to perform the measurements according to the standard measurement method;
b) to report any anomalies or difficulties experienced; it is better to report a mistake than to adjust the test results because one or two missing test results will not spoil the experiment and many indicate a deficiency in the standard;
c) to comment on the adequacy of the instructions in the standard; operators should report any occasions when they are unable to follow their instructions as this may also indicate a deficiency in the standard.

7 Statistical analysis of a precision experiment

7.1 Preliminary considerations

7.1.1 The analysis of the data, which should be considered as a statistical problem to be solved by a statistical expert, involves three successive stages:

a) critical examination of the data in order to identify and treat outliers or other irregularities and to test the suitability of the model;
b) computation of preliminary values of precision and means for each level separately;
c) establishment of final values of precision and means, including the establishment of a relationship between precision and the level m when the analysis indicates that such a relationship may exist.

7.1.2 The analysis first computes, for each level separately, estimates of

— the repeatability variance s_r^2

— the between-laboratory variance s_L^2

— the reproducibility variance $s_R^2 = s_r^2 + s_L^2$

— the mean m.

7.1.3 The analysis includes a systematic application of statistical tests for outliers, a great variety of which are available from the literature and which could be used for the purposes of this part of ISO 5725. For practical reasons, only a limited number of these tests, as explained in 7.3, have been incorporated.

7.2 Tabulation of the results and notation used

7.2.1 Cells

Each combination of a laboratory and a level is called a cell of the precision experiment. In the ideal case, the results of an experiment with p laboratories and q levels consist of a table with pq cells, each containing n replicate test results that can all be used for computing the repeatability standard deviation and the reproducibility standard deviation. This ideal situation is not, however, always attained in practice. Departures occur owing to redundant data, missing data and outliers.

7.2.2 Redundant data

Sometimes a laboratory may carry out and report more than the n test results officially specified. In that case, the supervisor shall report why this was done and which are the correct test results. If the answer is that they are all equally valid, then a random selection should be made from those available test results to choose the planned number of test results for analysis.

7.2.3 Missing data

In other cases, some of the test results may be missing, for example because of loss of a sample or a mistake in performing the measurement. The analysis recommended in 7.1 is such that completely empty cells can simply be ignored, while partly empty cells can be taken into account by the standard computational procedure.

7.2.4 Outliers

These are entries among the original test results, or in the tables derived from them, that deviate so much from the comparable entries in the same table that they are considered irreconcilable with the other data. Experience has taught that outliers cannot always be avoided and they have to be taken into consideration in a similar way to the treatment of missing data.

7.2.5 Outlying laboratories

When several unexplained abnormal test results occur at different levels within the same laboratory, then that laboratory may be considered to be an outlier, having too high a within-laboratory variance and/or too large a systematic error in the level of its test results. It may then be reasonable to discard some or all of the data from such an outlying laboratory.

This part of ISO 5725 does not provide a statistical test by which suspected laboratories may be judged. The primary decision should be the responsibility of the statistical expert, but all rejected laboratories shall be reported to the panel for further action.

7.2.6 Erroneous data

Obviously erroneous data should be investigated and corrected or discarded.

7.2.7 Balanced uniform-level test results

The ideal case is p laboratories called i ($i = 1, 2, \ldots, p$), each testing q levels called j ($j = 1, 2, \ldots, q$) with n replicates at each level (each ij combination), giving a total of pqn test results. Because of missing (7.2.3) or deviating (7.2.4) test results, or outlying laboratories (7.2.5) or erroneous data (7.2.6), this ideal situation is not always attained. Under these conditions the notations given in 7.2.8 to 7.2.10 and the procedures of 7.4 allow for differing numbers of test results. Specimens of recommended forms for the statistical analysis are given in figure 2. For convenience, they will be referred to simply as forms A, B and C (of figure 2).

7.2.8 Original test results

See form A of figure 2, where

n_{ij} is the number of test results in the cell for laboratory i at level j;

y_{ijk} is any one of these test results ($k = 1, 2, \ldots, n_{ij}$);

p_j is the number of laboratories reporting at least one test result for level j (after eliminating any test results designated as outliers or as erroneous).

Form A — Recommended form for the collation of the original data

Laboratory	Level								
	1	2	..	..	j	..	..	$q-1$	q
1									
2									
..									
.. .. i ..					.. .. y_{ijk} ..				
..									
p									

Form B — Recommended form for the collation of the means

Laboratory	Level								
	1	2	..	..	j	..	..	$q-1$	q
1									
2									
..									
i					$\bar{y}_{ij}$				
..									
p									

Form C — Recommended form for the collation of the measures of spread within cells

Laboratory	Level								
	1	2	..	..	j	..	..	$q-1$	q
1									
2									
..									
i					s_{ij}				
..									
p									

Figure 2 — Recommended forms for the collation of results for analysis

7.2.9 Cell means (form B of figure 2)

These are derived from form A as follows:

$$\bar{y}_{ij} = \frac{1}{n_{ij}} \sum_{k=1}^{n_{ij}} y_{ijk} \qquad \ldots (2)$$

The cell means should be recorded to one more significant figure than the test result in form A.

7.2.10 Measures of cell spread (form C of figure 2)

These are derived from form A (see 7.2.8) and form B (see 7.2.9) as follows.

For the general case, use the intracell standard deviation

$$s_{ij} = \sqrt{\frac{1}{n_{ij} - 1} \sum_{k=1}^{n_{ij}} (y_{ijk} - \bar{y}_{ij})^2} \quad \ldots (3)$$

or, equivalently

$$s_{ij} = \sqrt{\frac{1}{n_{ij} - 1} \left[\sum_{k=1}^{n_{ij}} (y_{ijk})^2 - \frac{1}{n_{ij}} \left[\sum_{k=1}^{n_{ij}} y_{ijk} \right]^2 \right]} \quad \ldots (4)$$

In using these equations, care shall be taken to retain a sufficient number of digits in the calculations; i.e. every intermediate value shall be calculated to at least twice as many digits as in the original data.

NOTE 4 If a cell ij contains two test results, the intracell standard deviation is

$$s_{ij} = | y_{ij1} - y_{ij2} | / \sqrt{2} \quad \ldots (5)$$

Therefore, for simplicity, absolute differences can be used instead of standard deviations if all cells contain two test results.

The standard deviation should be expressed to one more significant figure than the results in form A.

For values of n_{ij} less than 2, a dash should be inserted in form C.

7.2.11 Corrected or rejected data

As some of the data may be corrected or rejected on the basis of the tests mentioned in 7.1.3, 7.3.3 and 7.3.4, the values of y_{ijk}, n_{ij} and p_j used for the final determinations of precision and mean may be different from the values referring to the original test results as recorded in forms A, B and C of figure 2. Hence in reporting the final values for precision and trueness, it shall always be stated what data, if any, have been corrected or discarded.

7.3 Scrutiny of results for consistency and outliers

See reference [3].

From data collected on a number of specific levels, repeatability and reproducibility standard deviations are to be estimated. The presence of individual laboratories or values that appear to be inconsistent with all other laboratories or values may change the estimates, and decisions have to be made with respect to these values. Two approaches are introduced:

a) graphical consistency technique;

b) numerical outlier tests.

7.3.1 Graphical consistency technique

Two measures called Mandel's h and k statistics are used. It may be noted that, as well as describing the variability of the measurement method, these help in laboratory evaluation.

7.3.1.1 Calculate the between-laboratory consistency statistic, h, for each laboratory by dividing the cell deviation (cell mean minus the grand mean for that level) by the standard deviation among cell means (for that level):

$$h_{ij} = \frac{\bar{y}_{ij} - \bar{\bar{y}}_j}{\sqrt{\frac{1}{(p_j - 1)} \sum_{i=1}^{p_j} (\bar{y}_{ij} - \bar{\bar{y}}_j)^2}} \quad \ldots (6)$$

in which, for $\bar{y}_{ij}$ see 7.2.9, and for $\bar{\bar{y}}_j$ see 7.4.4.

Plot the h_{ij} values for each cell in order of laboratory, in groups for each level (and separately grouped for the several levels examined by each laboratory) (see figure B.7).

7.3.1.2 Calculate the within-laboratory consistency statistic, k, by first calculating the pooled within-cell standard deviation

$$\sqrt{\frac{\sum s_{ij}^2}{p_j}}$$

for each level, and then calculate

$$k_{ij} = \frac{s_{ij} \sqrt{p_j}}{\sqrt{\sum s_{ij}^2}} \quad \ldots (7)$$

for each laboratory within each level.

Plot the k_{ij} values for each cell in order of laboratory, in groups for each level (and separately grouped for the several levels examined by each laboratory) (see figure B.8).

7.3.1.3 Examination of the h and k plots may indicate that specific laboratories exhibit patterns of results that are markedly different from the others in the study. This is indicated by consistently high or low within-cell variation and/or extreme cell means across many levels. If this occurs, the specific laboratory should be contacted to try to ascertain the cause of the discrepant behaviour. On the basis of the findings, the statistical expert could:

a) retain the laboratory's data for the moment;

b) ask the laboratory to redo the measurement (if feasible);

c) remove the laboratory's data from the study.

7.3.1.4 Various patterns can appear in the h plots. All laboratories can have both positive and negative h values at different levels of the experiment. Individual laboratories may tend to give either all positive or all negative h values, and the number of laboratories giving negative values is approximately equal to those giving positive values. Neither of these patterns is unusual or requires investigation, although the second of these patterns may suggest that a common source of laboratory bias exists. On the other hand, if all the h values for one laboratory are of one sign and the h values for the other laboratories are all of the other sign, then the reason should be sought. Likewise, if the h values for a laboratory are extreme and appear to depend on the experimental level in some systematic way, then the reason should be sought. Lines are drawn on the h plots corresponding to the indicators given in 8.3 (tables 6 and 7). These indicator lines serve as guides when examining patterns in the data.

7.3.1.5 If one laboratory stands out on the k plot as having many large values, then the reason should be sought: this indicates that it has a poorer repeatability than the other laboratories. A laboratory could give rise to consistently small k values because of such factors as excessive rounding of its data or an insensitive measurement scale. Lines are drawn on the k plots corresponding to the indicators given in 8.3 (tables 6 and 7). These indicator lines serve as guides when examining patterns in the data.

7.3.1.6 When an h or k plot grouped by laboratory suggests that one laboratory has several h or k values near the critical value line, the corresponding plot grouped by level should be studied. Often a value that appears large in a plot grouped by laboratory will turn out to be reasonably consistent with other laboratories for the same level. If it is revealed as strongly different from values for the other laboratories, then the reason should be sought.

7.3.1.7 In addition to these h and k graphs, histograms of cell means and cell ranges can reveal the presence of, for example, two distinct populations. Such a case would require special treatment as the general underlying principle behind the methods described here assumes a single unimodal population.

7.3.2 Numerical outlier technique

7.3.2.1 The following practice is recommended for dealing with outliers.

a) The tests recommended in 7.3.3 and 7.3.4 are applied to identify stragglers or outliers:

   — if the test statistic is less than or equal to its 5 % critical value, the item tested is accepted as correct;

   — if the test statistic is greater than its 5 % critical value and less than or equal to its 1 % critical value, the item tested is called a straggler and is indicated by a single asterisk;

   — if the test statistic is greater than its 1 % critical value, the item is called a statistical outlier and is indicated by a double asterisk.

b) It is next investigated whether the stragglers and/or statistical outliers can be explained by some technical error, for example

   — a slip in performing the measurement,

   — an error in computation,

   — a simple clerical error in transcribing a test result, or

   — analysis of the wrong sample.

   Where the error was one of the computation or transcription type, the suspect result should be replaced by the correct value; where the error was from analysing a wrong sample, the result should be placed in its correct cell. After such correction has been made, the examination for stragglers or outliers should be repeated. If the explanation of the technical error is such that it proves impossible to replace the suspect test result, then it should be discarded as a "genuine" outlier that does not belong to the experiment proper.

c) When any stragglers and/or statistical outliers remain that have not been explained or rejected as

belonging to an outlying laboratory, the stragglers are retained as correct items and the statistical outliers are discarded unless the statistician for good reason decides to retain them.

d) When the data for a cell have been rejected for form B of figure 2 under the above procedure, then the corresponding data shall be rejected for form C of figure 2, and vice versa.

7.3.2.2 The tests given in 7.3.3 and 7.3.4 are of two types. Cochran's test is a test of the within-laboratory variabilities and should be applied first, then any necessary action should be taken, with repeated tests if necessary. The other test (Grubbs') is primarily a test of between-laboratory variability, and can also be used (if $n > 2$) where Cochran's test has raised suspicions as to whether the high within-laboratory variation is attributable to only one of the test results in the cell.

7.3.3 Cochran's test

7.3.3.1 This part of ISO 5725 assumes that between laboratories only small differences exist in the within-laboratory variances. Experience, however, shows that this is not always the case, so that a test has been included here to test the validity of this assumption. Several tests could be used for this purpose, but Cochran's test has been chosen.

7.3.3.2 Given a set of p standard deviations s_i, all computed from the same number (n) of replicate results, Cochran's test statistic, C, is

$$C = \frac{s_{\max}^2}{\sum_{i=1}^{p} s_i^2} \qquad \ldots (8)$$

where $s_{\max}$ is the highest standard deviation in the set.

a) If the test statistic is less than or equal to its 5 % critical value, the item tested is accepted as correct.

b) If the test statistic is greater than its 5 % critical value and less then or equal to its 1 % critical value, the item tested is called a straggler and is indicated by a single asterisk.

c) If the test statistic is greater than its 1 % critical value, the item is called a statistical outlier and is indicated by a double asterisk.

Critical values for Cochran's test are given in 8.1 (table 4).

Cochran's test has to be applied to form C of figure 2 at each level separately.

7.3.3.3 Cochran's criterion applies strictly only when all the standard deviations are derived from the same number (n) of test results obtained under repeatability conditions. In actual cases, this number may vary owing to missing or discarded data. This part of ISO 5725 assumes, however, that in a properly organized experiment such variations in the number of test results per cell will be limited and can be ignored, and therefore Cochran's criterion is applied using for n the number of test results occurring in the majority of cells.

7.3.3.4 Cochran's criterion tests only the highest value in a set of standard deviations and is therefore a one-sided outlier test. Variance heterogeneity may also, of course, manifest itself in some of the standard deviations being comparatively too low. However, small values of standard deviation may be very strongly influenced by the degree of rounding of the original data and are for that reason not very reliable. In addition, it seems unreasonable to reject the data from a laboratory because it has accomplished a higher precision in its test results than the other laboratories. Hence Cochran's criterion is considered adequate.

7.3.3.5 A critical examination of form C of figure 2 may sometimes reveal that the standard deviations for a particular laboratory are at all or at most levels lower than those for other laboratories. This may indicate that the laboratory works with a lower repeatability standard deviation than the other laboratories, which in turn may be caused either by better technique and equipment or by a modified or incorrect application of the standard measurement method. If this occurs it should be reported to the panel, which should then decide whether the point is worthy of a more detailed investigation. (An example of this is laboratory 2 in the experiment detailed in B.1.)

7.3.3.6 If the highest standard deviation is classed as an outlier, then the value should be omitted and Cochran's test repeated on the remaining values. This process can be repeated but it may lead to excessive rejections when, as is sometimes the case, the underlying assumption of normality is not sufficiently well approximated to. The repeated application of Cochran's test is here proposed only as a helpful tool in view of the lack of a statistical test designed for testing several outliers together. Cochran's test is not designed for this purpose and great caution should be exercised in drawing conclusions. When two or three

laboratories give results having high standard deviations, particularly if this is within only one of the levels, conclusions from Cochran's test should be examined carefully. On the other hand, if several stragglers and/or statistical outliers are found at different levels within one laboratory, this may be a strong indication that the laboratory's within-laboratory variance is exceptionally high, and the whole of the data from that laboratory should be rejected.

7.3.4 Grubbs' test

7.3.4.1 One outlying observation

Given a set of data x_i for $i = 1, 2, ..., p$, arranged in ascending order, then to determine whether the largest observation is an outlier using Grubbs' test, compute the Grubb's statistic, G_p.

$$G_p = (x_p - \bar{x})/s \quad \ldots (9)$$

where

$$\bar{x} = \frac{1}{p}\sum_{i=1}^{p} x_i \quad \ldots (10)$$

and

$$s = \sqrt{\frac{1}{p-1}\sum_{i=1}^{p}(x_i - \bar{x})^2} \quad \ldots (11)$$

To test the significance of the smallest observation, compute the test statistic

$$G_1 = (\bar{x} - x_1)/s$$

a) If the test statistic is less than or equal to its 5 % critical value, the item tested is accepted as correct.

b) If the test statistic is greater than its 5 % critical value and less than or equal to its 1 % critical value, the item tested is called a straggler and is indicated by a single asterisk.

c) If the test statistic is greater than its 1 % critical value, the item is called a statistical outlier and is indicated by a double asterisk.

7.3.4.2 Two outlying observations

To test whether the two largest observations may be outliers, compute the Grubbs' test statistic G:

$$G = s^2_{p-1,p}/s^2_0 \quad \ldots (12)$$

where

$$s^2_0 = \sum_{i=1}^{p}(x_i - \bar{x})^2 \quad \ldots (13)$$

and

$$s^2_{p-1,p} = \sum_{i=1}^{p-2}(x_i - \bar{x}_{p-1,p})^2 \quad \ldots (14)$$

and

$$\bar{x}_{p-1,p} = \frac{1}{p-2}\sum_{i=1}^{p-2} x_i \quad \ldots (15)$$

Alternatively, to test the two smallest observations, compute the Grubbs' test statistic G:

$$G = s^2_{1,2}/s^2_0 \quad \ldots (16)$$

where

$$s^2_{1,2} = \sum_{i=3}^{p}(x_i - \bar{x}_{1,2})^2 \quad \ldots (17)$$

and

$$\bar{x}_{1,2} = \frac{1}{p-2}\sum_{i=3}^{p} x_i \quad \ldots (18)$$

Critical values for Grubbs' test are given in 8.2 (table 5).

7.3.4.3 Application of Grubbs' test

When analysing a precision experiment, Grubbs' test can be applied to the following.

a) The cell averages (form B of figure 2) for a given level j, in which case

$$x_i = \bar{y}_{ij}$$

and

$$p = p_j$$

where j is fixed.

Taking the data at one level, apply the Grubbs' test for one outlying observation to cell means as described in 7.3.4.1. If a cell mean is shown to be an outlier by this test, exclude it, and repeat the test at the other extreme cell mean (e.g. if the highest is an outlier then look at the lowest with the highest excluded), but do not apply the

Grubbs' test for two outlying observations described in 7.3.4.2. If the Grubbs' test does not show a cell mean to be an outlier, then apply the double-Grubbs' test described in 7.3.4.2.

b) A single result within a cell, where Cochran's test has shown the cell standard deviation to be suspect.

7.4 Calculation of the general mean and variances

7.4.1 Method of analysis

The method of analysis adopted in this part of ISO 5725 involves carrying out the estimation of m and the precision for each level separately. The results of the computation are expressed in a table for each value of j.

7.4.2 Basic data

The basic data needed for the computations are presented in the three tables given in figure 2:

— table A containing the original test results;

— table B containing the cell means;

— table C containing the measures of within-cell spread.

7.4.3 Non-empty cells

As a consequence of the rule stated in 7.3.2.1 d), the number of non-empty cells to be used in the computation will, for a specific level, always be the same in tables B and C. An exception might occur if, owing to missing data, a cell in table A contains only a single test result, which will entail an empty cell in table C but not in table B. In that case it is possible

a) to discard the solitary test result, which will lead to empty cells in both tables B and C, or

b) if this is considered an undesirable loss of information, to insert a dash in table C.

The number of non-empty cells may be different for different levels, hence the index j in p_j.

7.4.4 Calculation of the general mean $\hat{m}$

For level j, the general mean is

$$\hat{m}_j = \bar{\bar{y}}_j = \frac{\sum_{i=1}^{p} n_{ij}\,\bar{y}_{ij}}{\sum_{i=1}^{p} n_{ij}} \qquad \ldots (19)$$

7.4.5 Calculation of variances

Three variances are calculated for each level. They are the repeatability variance, the between-laboratory variance and the reproducibility variance.

7.4.5.1 The repeatability variance is

$$s_{rj}^2 = \frac{\sum_{i=1}^{p} (n_{ij} - 1) s_{ij}^2}{\sum_{i=1}^{p} (n_{ij} - 1)} \qquad \ldots (20)$$

7.4.5.2 The between-laboratory variance is

$$s_{Lj}^2 = \frac{s_{dj}^2 - s_{rj}^2}{\bar{\bar{n}}_j} \qquad \ldots (21)$$

where

$$s_{dj}^2 = \frac{1}{p-1} \sum_{i=1}^{p} n_{ij} (\bar{y}_{ij} - \bar{\bar{y}}_j)^2$$

$$= \frac{1}{p-1} \left[\sum_{i=1}^{p} n_{ij} (\bar{y}_{ij})^2 - (\bar{\bar{y}}_j)^2 \sum_{i=1}^{p} n_{ij} \right] \qquad \ldots (22)$$

and

$$\bar{\bar{n}}_j = \frac{1}{p-1} \left[\sum_{i=1}^{p} n_{ij} - \frac{\sum_{i=1}^{p} n_{ij}^2}{\sum_{i=1}^{p} n_{ij}} \right] \qquad \ldots (23)$$

These calculations are illustrated in the examples in B.1 and B.3 in annex B.

7.4.5.3 For the particular case where all $n_{ij} = n = 2$, the simpler formulae may be used, giving

$$s_{rj}^2 = \frac{1}{2p} \sum_{i=1}^{p} (y_{ij1} - y_{ij2})^2$$

and

$$s_{Lj}^2 = \frac{1}{p-1} \sum_{i=1}^{p} (\bar{y}_{ij} - \bar{\bar{y}}_j)^2 - \frac{s_{rj}^2}{2}$$

These are illustrated by the example given in B.2.

7.4.5.4 Where, owing to random effects, a negative value for s_{Lj}^2 is obtained from these calculations, the value should be assumed to be zero.

7.4.5.5 The reproducibility variance is

$$s_{Rj}^2 = s_{rj}^2 + s_{Lj}^2 \qquad \ldots (24)$$

7.4.6 Dependence of the variances upon m

Subsequently, it should be investigated whether the precision depends upon m and, if so, the functional relationship should be determined.

7.5 Establishing a functional relationship between precision values and the mean level m

7.5.1 It cannot always be taken for granted that there exists a regular functional relationship between precision and m. In particular, where material heterogeneity forms an inseparable part of the variability of the test results, there will be a functional relationship only if this heterogeneity is a regular function of the level m. With solid materials of different composition and coming from different production processes, a regular functional relationship is in no way certain. This point should be decided before the following procedure is applied. Alternatively, separate values of precision would have to be established for each material investigated.

7.5.2 The reasoning and computation procedures presented in 7.5.3 to 7.5.9 apply both to repeatability and reproducibility standard deviations, but are presented here for repeatability only in the interests of brevity. Only three types of relationship will be considered:

I: $s_r = bm$ (a straight line through the origin)

II: $s_r = a + bm$ (a straight line with a positive intercept)

III: $\lg s_r = c + d \lg m$ (or $s_r = Cm^d$); $d \leqslant 1$ (an exponential relationship)

It is to be expected that in the majority of cases at least one of these formulae will give a satisfactory fit. If not, the statistical expert carrying out the analysis should seek an alternative solution. To avoid confusion, the constants a, b, c, C and d occurring in these equations may be distinguished by subscripts, a_r, b_r, ... for repeatability and a_R, b_R, ... when considering reproducibility, but these have been omitted in this clause again to simplify the notations. Also s_r has been abbreviated simply to s to allow a suffix for the level j.

7.5.3 In general $d > 0$ so that relationships I and III will lead to $s = 0$ for $m = 0$, which may seem unacceptable from an experimental point of view. However, when reporting the precision data, it should be made clear that they apply only within the levels covered by the interlaboratory precision experiment.

7.5.4 For $a = 0$ and $d = 1$, all three relationships are identical, so when a lies near zero and/or d lies near unity, two or all three of these relationships will yield practically equivalent fits, and in such a case relationship I should be preferred because it permits the following simple statement.

> "Two test results are considered as suspect when they differ by more than (100 b) %."

In statistical terminology, this is a statement that the coefficient of variation (100 s/m) is a constant for all levels.

7.5.5 If in a plot of s_j against $\hat{m}_j$, or a plot of $\lg s_j$ against $\lg \hat{m}_j$, the set of points are found to lie reasonably close to a straight line, a line drawn by hand may provide a satisfactory solution; but if for some reason a numerical method of fitting is preferred, the procedure of 7.5.6 is recommended for relationships I and II, and that of 7.5.8 for relationship III.

7.5.6 From a statistical viewpoint, the fitting of a straight line is complicated by the fact that both $\hat{m}_j$ and s_j are estimates and thus subject to error. But as the slope b is usually small (of the order of 0,1 or less), then errors in $\hat{m}$ have little influence and the errors in estimating s predominate.

7.5.6.1 A good estimate of the parameters of the regression line requires a weighted regression because the standard error of s is proportional to the predicted value of s_j ($\hat{s}_j$).

The weighting factors have to be proportional to $1/(\hat{s}_j)^2$, where $\hat{s}_j$ is the predicted repeatability standard deviation for level j. However $\hat{s}_j$ depends on parameters that have yet to be calculated.

A mathematically correct procedure for finding estimates corresponding to the weighted least-squares of residuals may be complicated. The following procedure, which has proved to be satisfactory in practice, is recommended.

7.5.6.2 With weighting factor W_j equal to $1/(\hat{s}_{Nj})^2$, where $N = 0, 1, 2 \ldots$ for successive iterations, then the calculated formulae are:

$$T_1 = \sum_j W_j$$

$$T_2 = \sum_j W_j \hat{m}_j$$

$$T_3 = \sum_j W_j \hat{m}_j^2$$

$$T_4 = \sum_j W_j s_j$$

$$T_5 = \sum_j W_j \hat{m}_j s_j$$

Then for relationship I ($s = bm$), the value of b is given by T_5/T_3.

For relationship II ($s = a + bm$):

$$a = \left[\frac{T_3 T_4 - T_2 T_5}{T_1 T_3 - T_2^2} \right] \quad \ldots (25)$$

and

$$b = \left[\frac{T_1 T_5 - T_2 T_4}{T_1 T_3 - T_2^2} \right] \quad \ldots (26)$$

7.5.6.3 For relationship I, algebraic substitution for the weighting factors $W_j = 1/(\hat{s}_j)^2$ with $\hat{s}_j = b\hat{m}_j$ leads to the simplified expression:

$$b = \frac{\sum_j \left(s_j / \hat{m}_j \right)}{q} \quad \ldots (27)$$

and no iteration is necessary.

7.5.6.4 For relationship II, the initial values $\hat{s}_{0j}$ are the original values of s as obtained by the procedures given in 7.4. These are used to calculate

$$W_{0j} = 1/(\hat{s}_{0j})^2 \quad (j = 1, 2, \ldots, q)$$

and to calculate a_1 and b_1 as in 7.5.6.2.

This leads to

$$\hat{s}_{1j} = a_1 + b_1 \hat{m}_j$$

The computations are then repeated with $W_{1j} = 1/(\hat{s}_{1j})^2$ to produce

$$\hat{s}_{2j} = a_2 + b_2 \hat{m}_j$$

The same procedure could now be repeated once again with weighting factors $W_{2j} = 1/(\hat{s}_{2j})^2$ derived from these equations, but this will only lead to unimportant changes. The step from W_{0j} to W_{1j} is effective in eliminating gross errors in the weights, and the equations for $\hat{s}_{2j}$ should be considered as the final result.

7.5.7 The standard error of lg s is independent of s and so an unweighted regression of lg s on lg $\hat{m}$ is appropriate.

7.5.8 For relationship III, the computational formulae are:

$$T_1 = \sum_j \lg \hat{m}_j$$

$$T_2 = \sum_j \left(\lg \hat{m}_j \right)^2$$

$$T_3 = \sum_j \lg s_j$$

$$T_4 = \sum_j \left(\lg \hat{m}_j \right) \left(\lg s_j \right)$$

and thence

$$c = \frac{T_2 T_3 - T_1 T_4}{qT_2 - T_1^2} \quad \ldots (28)$$

and

$$d = \frac{qT_4 - T_1 T_3}{qT_2 - T_1^2} \quad \ldots (29)$$

7.5.9 Examples of fitting relationships I, II and III of 7.5.2 to the same set of data are now given in 7.5.9.1 to 7.5.9.3. The data are taken from the case study of B.3 and have been used here only to illustrate the numerical procedure. They will be further discussed in B.3.

7.5.9.1 An example of fitting relationship I is given in table 1.

7.5.9.2 An example of fitting relationship II is given in table 2 ($\hat{m}_j$, s_j are as in 7.5.9.1).

7.5.9.3 An example of fitting relationship III is given in table 3.

Table 1 — Relationship I: $s = bm$

$\hat{m}_j$ s_j	3,94 0,092	8,28 0,179	14,18 0,127	15,59 0,337	20,41 0,393
$s_j/\hat{m}_j$	0,023 4	0,021 6	0,008 9	0,021 6	0,019 3
$b = \frac{\sum_j (s_j/\hat{m}_j)}{q}$	$\frac{0,094\ 8}{5} = 0,019$				
$s = bm$	0,075	0,157	0,269	0,296	0,388

Table 2 — Relationship II: $s = a + bm$

W_{0j}	118	31	62	8,8	6,5
$s_1 = 0,058 + 0,009\ 0\ m$					
$\hat{s}_{1j}$ W_{1j}	0,093 116	0,132 57	0,185 29	0,197 26	0,240 17
$s_2 = 0,030 + 0,015\ 6\ m$					
$\hat{s}_{2j}$ W_{2j}	0,092 118	0,159 40	0,251 16	0,273 13	0,348 8
$s_3 = 0,032 + 0,015\ 4\ m$					
$\hat{s}_{3j}$ 1)	0,093	0,160	0,251	0,273	0,348
NOTE — The values of the weighting factors are not critical; two significant figures suffice.					
1) The difference from s_2 is negligible.					

Table 3 — Relationship III: $\lg s = c + d \lg m$

$\lg \hat{m}_j$ $\lg s_{0j}$	+ 0,595 − 1,036	+ 0,918 − 0,747	+ 1,152 − 0,896	+ 1,193 − 0,472	+ 1,310 − 0,406
$\lg s = -1,506\ 5 + 0,772 \lg m$ or $s = 0,031\ m^{0,77}$					
s	0,089	0,158	0,239	0,257	0,316

7.6 Statistical analysis as a step-by-step procedure

NOTE 5 Figure 3 indicates in a stepwise fashion the procedure given in 7.6.

7.6.1 Collect all available test results in one form, form A of figure 2 (see 7.2). It is recommended that this form be arranged into p rows, indexed $i = 1, 2, ..., p$ (representing the p laboratories that have contributed data) and q columns, indexed $j = 1, 2, ..., q$ (representing the q levels in increasing order).

In a uniform-level experiment the test results within a cell of form A need not be distinguished and may be put in any desired order.

7.6.2 Inspect form A for any obvious irregularities, investigate and, if necessary, discard any obviously erroneous data (for example, data outside the range of the measuring instrument or data which are impossible for technical reasons) and report to the panel. It is sometimes immediately evident that the test results of a particular laboratory or in a particular cell lie at a level inconsistent with the other data. Such obviously discordant data shall be discarded immediately, but the fact shall be reported to the panel for further consideration (see 7.7.1).

7.6.3 From form A, corrected according to 7.6.2 when needed, compute form B containing cell means and form C containing measures of within-cell spread.

When a cell in form A contains only a single test result, one of the options of 7.4.3 should be adopted.

7.6.4 Prepare the Mandel h and k plots as described in 7.3.1 and examine them for consistency of the data. These plots may indicate the suitability of the data for further analysis, the presence of any possible outlying values or outlying laboratories. However, no definite decisions are taken at this stage, but are delayed until completion of 7.6.5 to 7.6.9.

7.6.5 Inspect forms B and C (see figure 2) level by level for possible stragglers and/or statistical outliers [see 7.3.2.1 a)]. Apply the statistical tests given in 7.3 to all suspect items, marking the stragglers with a single asterisk and the statistical outliers with a double asterisk. If there are no stragglers or statistical outliers, ignore steps 7.6.6 to 7.6.10 and proceed directly with 7.6.11.

7.6.6 Investigate whether there is or may be some technical explanation for the stragglers and/or statistical outliers and, if possible, verify such an explanation. Correct or discard, as required, those stragglers and/or statistical outliers that have been satisfactorily explained, and apply corresponding corrections to the forms. If there are no stragglers or statistical outliers left that have not been explained, ignore steps 7.6.7 to 7.6.10 and proceed directly with 7.6.11.

NOTE 6 A large number of stragglers and/or statistical outliers may indicate a pronounced variance inhomogeneity or pronounced differences between laboratories and may thereby cast doubt on the suitability of the measurement method. This should be reported to the panel.

7.6.7 If the distribution of the unexplained stragglers or statistical outliers in form B or C does not suggest any outlying laboratories (see 7.2.5), ignore step 7.6.8 and proceed directly with 7.6.9.

7.6.8 If the evidence against some suspected outlying laboratories is considered strong enough to justify the rejection of some or all the data from those laboratories, then discard the requisite data and report to the panel.

The decision to reject some or all data from a particular laboratory is the responsibility of the statistical expert carrying out the analysis, but shall be reported to the panel for further consideration (see 7.7.1).

7.6.9 If any stragglers and/or statistical outliers remain that have not been explained or attributed to an outlying laboratory, discard the statistical outliers but retain the stragglers.

7.6.10 If in the previous steps any entry in form B has been rejected, then the corresponding entry in form C has to be rejected also, and vice versa.

7.6.11 From the entries that have been retained as correct in forms B and C, compute, by the procedures given in 7.4, for each level separately, the mean level $\hat{m}_j$ and the repeatability and reproducibility standard deviations.

7.6.12 If the experiment only used a single level, or if it has been decided that the repeatability and reproducibility standard deviations should be given separately for each level (see 7.5.1) and not as functions of the level, ignore steps 7.6.13 to 7.6.18 and proceed directly with 7.6.19.

NOTE 7 The following steps 7.6.13 to 7.6.17 are applied to s_r and s_R separately, but for brevity they are written out only in terms of s_r.

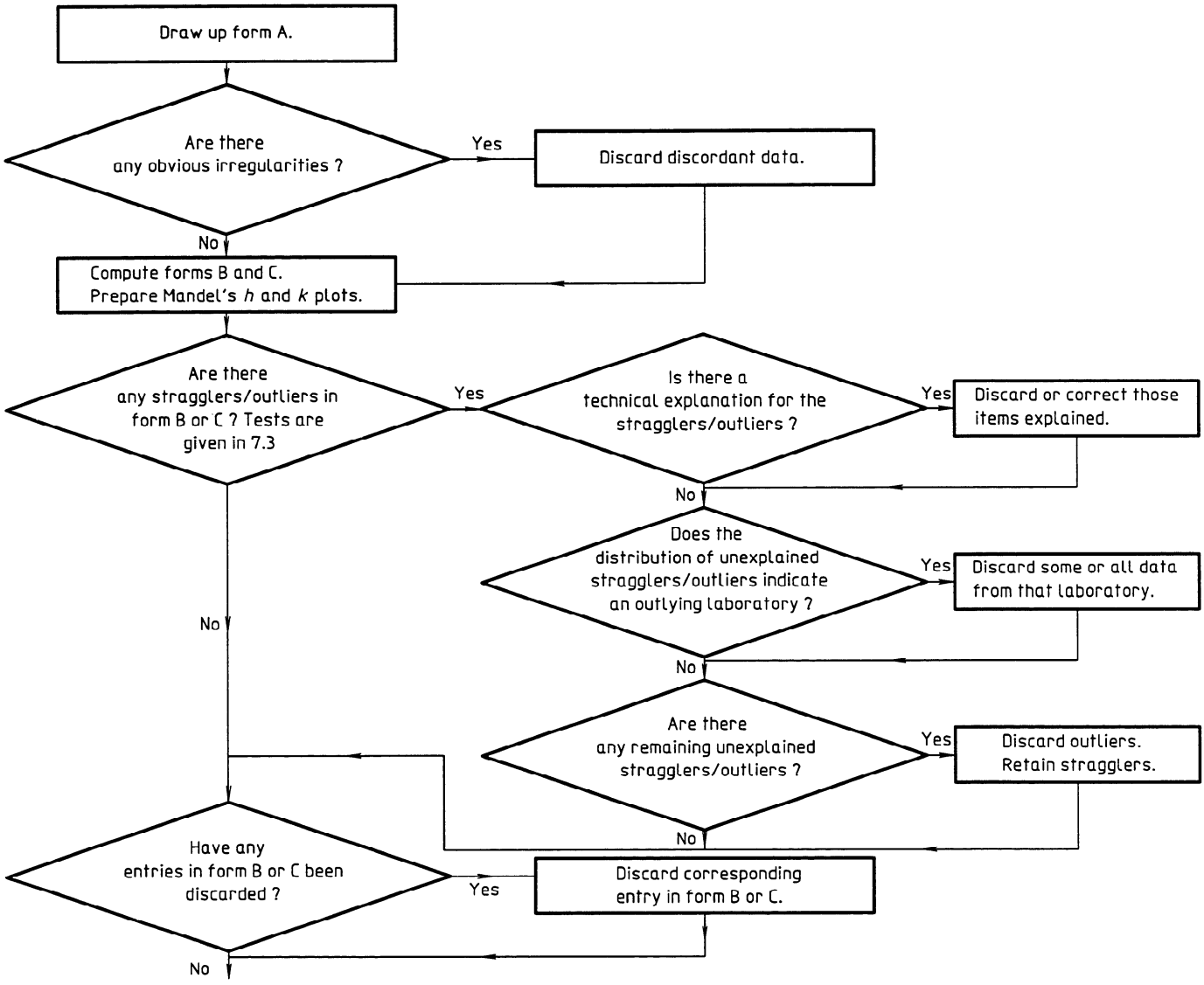

Figure 3 — Flow diagram of the principal steps in the statistical analysis *(continued on page 19)*

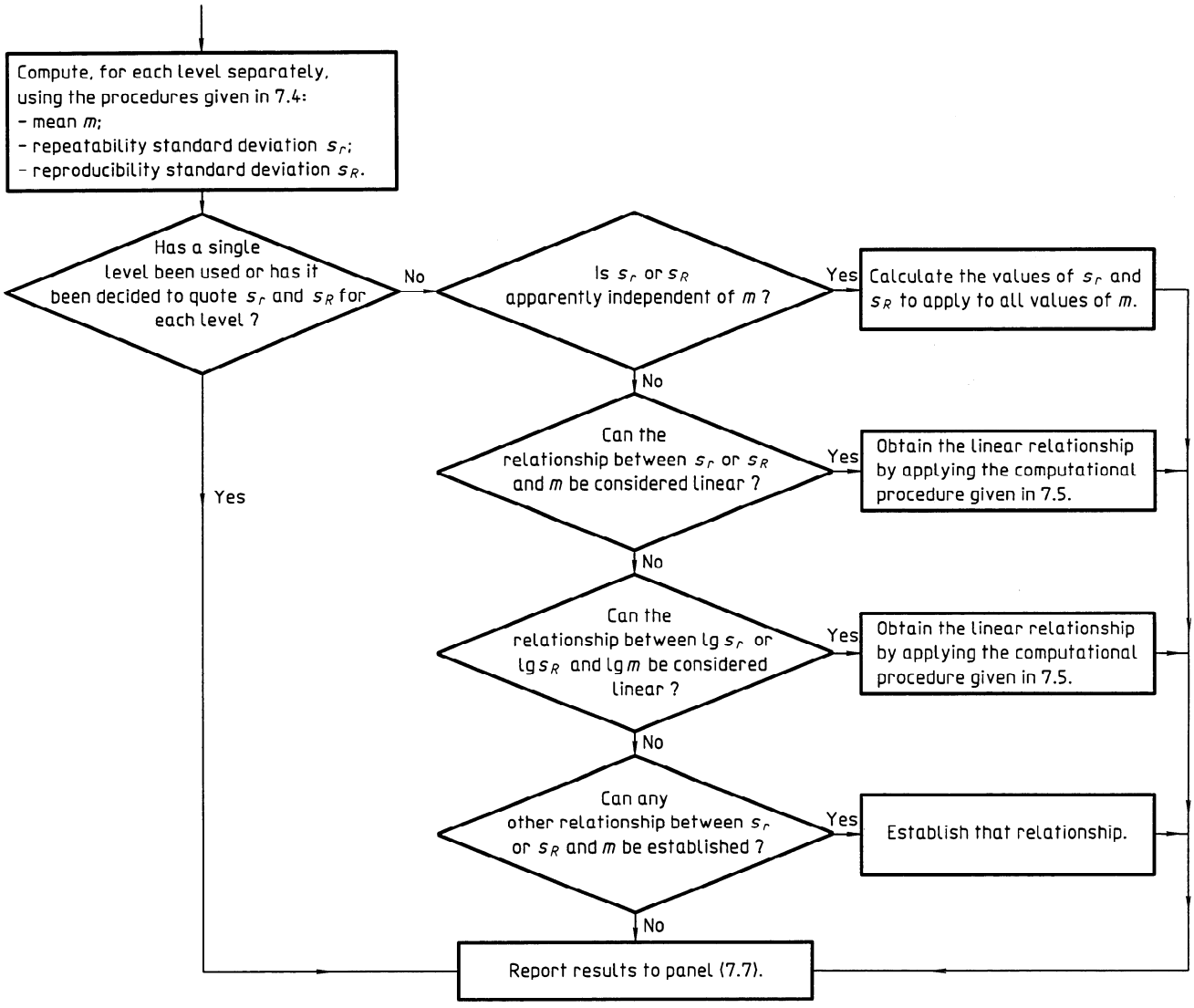

Figure 3 — Flow diagram of the principal steps in the statistical analysis

7.6.13 Plot s_j against $\hat{m}_j$ and judge from this plot whether s depends on m or not. If s is considered to depend on m, ignore step 7.6.14 and proceed with 7.6.15. If s is judged to be independent of m, proceed with step 7.6.14. If there should be doubt, it is best to work out both cases and let the panel decide. There exists no useful statistical test appropriate for this problem, but the technical experts familiar with the measurement method should have sufficient experience to take a decision.

7.6.14 Use $\frac{1}{q}\Sigma s_j = s_r$ as the final value of the repeatability standard deviation. Ignore steps 7.6.15 to 7.6.18 and proceed directly with 7.6.19.

7.6.15 Judge from the plot of 7.6.13 whether the relationship between s and m can be represented by a straight line and, if so, whether relationship I ($s = bm$) or relationship II ($s = a + bm$) is appropriate (see 7.5.2). Determine the parameter b, or the two parameters a and b, by the procedure of 7.5.6. If the linear relationship is considered satisfactory, ignore step 7.6.16 and proceed directly with 7.6.17. If not, proceed with step 7.6.16.

7.6.16 Plot lg s_j against lg $\hat{m}_j$ and judge from this whether the relationship between lg s and lg m can reasonably be represented by a straight line. If this is considered satisfactory, fit the relationship III (lg $s = c + d$ lg m) using the procedure given in 7.5.8.

7.6.17 If a satisfactory relation has been established in step 7.6.15 or 7.6.16, then the final values of s_r (or s_R) are the smoothed values obtained from this relationship for given values of m. Ignore step 7.6.18 and proceed with 7.6.19.

7.6.18 If no satisfactory relation has been established in step 7.6.15 or 7.6.16, the statistical expert should decide whether some other relation between s and m can be established, or alternatively whether the data are so irregular that the establishment of a functional relationship is considered to be impossible.

7.6.19 Prepare a report showing the basic data and the results and conclusions from the statistical analysis, and present this to the panel. The graphical presentations of 7.3.1 may be useful in presenting the consistency or variability of the results.

7.7 The report to, and the decisions to be taken by, the panel

7.7.1 Report by the statistical expert

Having completed the statistical analysis, the statistical expert should write a report to be submitted to the panel. In this report the following information should be given:

a) a full account of the observations received from the operators and/or supervisors concerning the standard for the measurement method;

b) a full account of the laboratories that have been rejected as outlying laboratories in steps 7.6.2 and 7.6.8, together with the reasons for their rejection;

c) a full account of any stragglers and/or statistical outliers that were discovered, and whether these were explained and corrected, or discarded;

d) a form of the final results $\hat{m}_j$, s_r and s_R and an account of the conclusions reached in steps 7.6.13, 7.6.15 or 7.6.16, illustrated by one of the plots recommended in these steps;

e) forms A, B and C (figure 2) used in the statistical analysis, possibly as an annex.

7.7.2 Decisions to be taken by the panel

The panel should then discuss this report and take decisions concerning the following questions.

a) Are the discordant results, stragglers or outliers, if any, due to defects in the description of the standard for the measurement method?

b) What action should be taken with respect to rejected outlying laboratories?

c) Do the results of the outlying laboratories and/or the comments received from the operators and supervisors indicate the need to improve the standard for the measurement method? If so, what are the improvements required?

d) Do the results of the precision experiment justify the establishment of values of the repeatability standard deviation and reproducibility standard deviation? If so, what are those values, in what form should they be published, and what is the region in which the precision data apply?

7.7.3 Full report

A report setting out the reasons for the work and how it was organized, including the report by the statistician and setting out agreed conclusions, should be prepared by the executive officer for approval by the panel. Some graphical presentation of consistency or variability is often useful. The report should be circulated to those responsible for authorizing the work and to other interested parties.

8 Statistical tables

8.1 Critical values for Cochran's test (see 7.3.3) are given in table 4.

Table 4 — Critical values for Cochran's test

p	$n = 2$		$n = 3$		$n = 4$		$n = 5$		$n = 6$	
	1 %	5 %	1 %	5 %	1 %	5 %	1 %	5 %	1 %	5 %
2	—	—	0,995	0,975	0,979	0,939	0,959	0,906	0,937	0,877
3	0,993	0,967	0,942	0,871	0,883	0,798	0,834	0,746	0,793	0,707
4	0,968	0,906	0,864	0,768	0,781	0,684	0,721	0,629	0,676	0,590
5	0,928	0,841	0,788	0,684	0,696	0,598	0,633	0,544	0,588	0,506
6	0,883	0,781	0,722	0,616	0,626	0,532	0,564	0,480	0,520	0,445
7	0,838	0,727	0,664	0,561	0,568	0,480	0,508	0,431	0,466	0,397
8	0,794	0,680	0,615	0,516	0,521	0,438	0,463	0,391	0,423	0,360
9	0,754	0,638	0,573	0,478	0,481	0,403	0,425	0,358	0,387	0,329
10	0,718	0,602	0,536	0,445	0,447	0,373	0,393	0,331	0,357	0,303
11	0,684	0,570	0,504	0,417	0,418	0,348	0,366	0,308	0,332	0,281
12	0,653	0,541	0,475	0,392	0,392	0,326	0,343	0,288	0,310	0,262
13	0,624	0,515	0,450	0,371	0,369	0,307	0,322	0,271	0,291	0,243
14	0,599	0,492	0,427	0,352	0,349	0,291	0,304	0,255	0,274	0,232
15	0,575	0,471	0,407	0,335	0,332	0,276	0,288	0,242	0,259	0,220
16	0,553	0,452	0,388	0,319	0,316	0,262	0,274	0,230	0,246	0,208
17	0,532	0,434	0,372	0,305	0,301	0,250	0,261	0,219	0,234	0,198
18	0,514	0,418	0,356	0,293	0,288	0,240	0,249	0,209	0,223	0,189
19	0,496	0,403	0,343	0,281	0,276	0,230	0,238	0,200	0,214	0,181
20	0,480	0,389	0,330	0,270	0,265	0,220	0,229	0,192	0,205	0,174
21	0,465	0,377	0,318	0,261	0,255	0,212	0,220	0,185	0,197	0,167
22	0,450	0,365	0,307	0,252	0,246	0,204	0,212	0,178	0,189	0,160
23	0,437	0,354	0,297	0,243	0,238	0,197	0,204	0,172	0,182	0,155
24	0,425	0,343	0,287	0,235	0,230	0,191	0,197	0,166	0,176	0,149
25	0,413	0,334	0,278	0,228	0,222	0,185	0,190	0,160	0,170	0,144
26	0,402	0,325	0,270	0,221	0,215	0,179	0,184	0,155	0,164	0,140
27	0,391	0,316	0,262	0,215	0,209	0,173	0,179	0,150	0,159	0,135
28	0,382	0,308	0,255	0,209	0,202	0,168	0,173	0,146	0,154	0,131
29	0,372	0,300	0,248	0,203	0,196	0,164	0,168	0,142	0,150	0,127
30	0,363	0,293	0,241	0,198	0,191	0,159	0,164	0,138	0,145	0,124
31	0,355	0,286	0,235	0,193	0,186	0,155	0,159	0,134	0,141	0,120
32	0,347	0,280	0,229	0,188	0,181	0,151	0,155	0,131	0,138	0,117
33	0,339	0,273	0,224	0,184	0,177	0,147	0,151	0,127	0,134	0,114
34	0,332	0,267	0,218	0,179	0,172	0,144	0,147	0,124	0,131	0,111
35	0,325	0,262	0,213	0,175	0,168	0,140	0,144	0,121	0,127	0,108
36	0,318	0,256	0,208	0,172	0,165	0,137	0,140	0,118	0,124	0,106
37	0,312	0,251	0,204	0,168	0,161	0,134	0,137	0,116	0,121	0,103
38	0,306	0,246	0,200	0,164	0,157	0,131	0,134	0,113	0,119	0,101
39	0,300	0,242	0,196	0,161	0,154	0,129	0,131	0,111	0,116	0,099
40	0,294	0,237	0,192	0,158	0,151	0,126	0,128	0,108	0,114	0,097

p = number of laboratories at a given level

n = number of test results per cell (see 7.3.3.3)

8.2 Critical values for Grubbs' test (see 7.3.4) are given in table 5.

For the Grubbs' test for one outlying observation, outliers and stragglers give rise to values which are larger than the tabulated 1 % and 5 % critical values respectively.

For the Grubbs' test for two outlying observations, outliers and stragglers give rise to values which are smaller than the tabulated 1 % and 5 % critical values respectively.

8.3 Indicators for Mandel's h and k statistics (see 7.3.1) are given in tables 6 and 7.

Table 5 — Critical values for Grubbs' test

p	One largest or one smallest		Two largest or two smallest	
	Upper 1 %	Upper 5 %	Lower 1 %	Lower 5 %
3	1,155	1,155	—	—
4	1,496	1,481	0,000 0	0,000 2
5	1,764	1,715	0,001 8	0,009 0
6	1,973	1,887	0,011 6	0,034 9
7	2,139	2,020	0,030 8	0,070 8
8	2,274	2,126	0,056 3	0,110 1
9	2,387	2,215	0,085 1	0,149 2
10	2,482	2,290	0,115 0	0,186 4
11	2,564	2,355	0,144 8	0,221 3
12	2,636	2,412	0,173 8	0,253 7
13	2,699	2,462	0,201 6	0,283 6
14	2,755	2,507	0,228 0	0,311 2
15	2,806	2,549	0,253 0	0,336 7
16	2,852	2,585	0,276 7	0,360 3
17	2,894	2,620	0,299 0	0,382 2
18	2,932	2,651	0,320 0	0,402 5
19	2,968	2,681	0,339 8	0,421 4
20	3,001	2,709	0,358 5	0,439 1
21	3,031	2,733	0,376 1	0,455 6
22	3,060	2,758	0,392 7	0,471 1
23	3,087	2,781	0,408 5	0,485 7
24	3,112	2,802	0,423 4	0,499 4
25	3,135	2,822	0,437 6	0,512 3
26	3,157	2,841	0,451 0	0,524 5
27	3,178	2,859	0,463 8	0,536 0
28	3,199	2,876	0,475 9	0,547 0
29	3,218	2,893	0,487 5	0,557 4
30	3,236	2,908	0,498 5	0,567 2
31	3,253	2,924	0,509 1	0,576 6
32	3,270	2,938	0,519 2	0,585 6
33	3,286	2,952	0,528 8	0,594 1
34	3,301	2,965	0,538 1	0,602 3
35	3,316	2,979	0,546 9	0,610 1
36	3,330	2,991	0,555 4	0,617 5
37	3,343	3,003	0,563 6	0,624 7
38	3,356	3,014	0,571 4	0,631 6
39	3,369	3,025	0,578 9	0,638 2
40	3,381	3,036	0,586 2	0,644 5

Reproduced, with the permission of the American Statistical Association, from reference [4] in annex C.

p = number of laboratories at a given level

Table 6 — Indicators for Mandel's h and k statistics at the 1 % significance level

p	h	k								
		n								
		2	3	4	5	6	7	8	9	10
3	1,15	1,71	1,64	1,58	1,53	1,49	1,46	1,43	1,41	1,39
4	1,49	1,91	1,77	1,67	1,60	1,55	1,51	1,48	1,45	1,43
5	1,72	2,05	1,85	1,73	1,65	1,59	1,55	1,51	1,48	1,46
6	1,87	2,14	1,90	1,77	1,68	1,62	1,57	1,53	1,50	1,47
7	1,98	2,20	1,94	1,79	1,70	1,63	1,58	1,54	1,51	1,48
8	2,06	2,25	1,97	1,81	1,71	1,65	1,59	1,55	1,52	1,49
9	2,13	2,29	1,99	1,82	1,73	1,66	1,60	1,56	1,53	1,50
10	2,18	2,32	2,00	1,84	1,74	1,66	1,61	1,57	1,53	1,50
11	2,22	2,34	2,01	1,85	1,74	1,67	1,62	1,57	1,54	1,51
12	2,25	2,36	2,02	1,85	1,75	1,68	1,62	1,58	1,54	1,51
13	2,27	2,38	2,03	1,86	1,76	1,68	1,63	1,58	1,55	1,52
14	2,30	2,39	2,04	1,87	1,76	1,69	1,63	1,58	1,55	1,52
15	2,32	2,41	2,05	1,87	1,76	1,69	1,63	1,59	1,55	1,52
16	2,33	2,42	2,05	1,88	1,77	1,69	1,63	1,59	1,55	1,52
17	2,35	2,44	2,06	1,88	1,77	1,69	1,64	1,59	1,55	1,52
18	2,36	2,44	2,06	1,88	1,77	1,70	1,64	1,59	1,56	1,52
19	2,37	2,44	2,07	1,89	1,78	1,70	1,64	1,59	1,56	1,53
20	2,39	2,45	2,07	1,89	1,78	1,70	1,64	1,60	1,56	1,53
21	2,39	2,46	2,07	1,89	1,78	1,70	1,64	1,60	1,56	1,53
22	2,40	2,46	2,08	1,90	1,78	1,70	1,65	1,60	1,56	1,53
23	2,41	2,47	2,08	1,90	1,78	1,71	1,65	1,60	1,56	1,53
24	2,42	2,47	2,08	1,90	1,79	1,71	1,65	1,60	1,56	1,53
25	2,42	2,47	2,08	1,90	1,79	1,71	1,65	1,60	1,56	1,53
26	2,43	2,48	2,09	1,90	1,79	1,71	1,65	1,60	1,56	1,53
27	2,44	2,48	2,09	1,90	1,79	1,71	1,65	1,60	1,56	1,53
28	2,44	2,49	2,09	1,91	1,79	1,71	1,65	1,60	1,57	1,53
29	2,45	2,49	2,09	1,91	1,79	1,71	1,65	1,60	1,57	1,53
30	2,45	2,49	2,10	1,91	1,79	1,71	1,65	1,61	1,57	1,53

p = number of laboratories at a given level

n = number of replicates within each laboratory at that level

NOTE — Supplied by Dr. J. Mandel and published with his permission.

Table 7 — Indicators for Mandel's h and k statistics at the 5 % significance level

p	h	k								
		n								
		2	3	4	5	6	7	8	9	10
3	1,15	1,65	1,53	1,45	1,40	1,37	1,34	1,32	1,30	1,29
4	1,42	1,76	1,59	1,50	1,44	1,40	1,37	1,35	1,33	1,31
5	1,57	1,81	1,62	1,53	1,46	1,42	1,39	1,36	1,34	1,32
6	1,66	1,85	1,64	1,54	1,48	1,43	1,40	1,37	1,35	1,33
7	1,71	1,87	1,66	1,55	1,49	1,44	1,41	1,38	1,36	1,34
8	1,75	1,88	1,67	1,56	1,50	1,45	1,41	1,38	1,36	1,34
9	1,78	1,90	1,68	1,57	1,50	1,45	1,42	1,39	1,36	1,35
10	1,80	1,90	1,68	1,57	1,50	1,46	1,42	1,39	1,37	1,35
11	1,82	1,91	1,69	1,58	1,51	1,46	1,42	1,39	1,37	1,35
12	1,83	1,92	1,69	1,58	1,51	1,46	1,42	1,40	1,37	1,35
13	1,84	1,92	1,69	1,58	1,51	1,46	1,43	1,40	1,37	1,35
14	1,85	1,92	1,70	1,59	1,52	1,47	1,43	1,40	1,37	1,35
15	1,86	1,93	1,70	1,59	1,52	1,47	1,43	1,40	1,38	1,36
16	1,86	1,93	1,70	1,59	1,52	1,47	1,43	1,40	1,38	1,36
17	1,87	1,93	1,70	1,59	1,52	1,47	1,43	1,40	1,38	1,36
18	1,88	1,93	1,71	1,59	1,52	1,47	1,43	1,40	1,38	1,36
19	1,88	1,93	1,71	1,59	1,52	1,47	1,43	1,40	1,38	1,36
20	1,89	1,94	1,71	1,59	1,52	1,47	1,43	1,40	1,38	1,36
21	1,89	1,94	1,71	1,60	1,52	1,47	1,44	1,41	1,38	1,36
22	1,89	1,94	1,71	1,60	1,52	1,47	1,44	1,41	1,38	1,36
23	1,90	1,94	1,71	1,60	1,53	1,47	1,44	1,41	1,38	1,36
24	1,90	1,94	1,71	1,60	1,53	1,48	1,44	1,41	1,38	1,38
25	1,90	1,94	1,71	1,60	1,53	1,48	1,44	1,41	1,38	1,36
26	1,90	1,94	1,71	1,60	1,53	1,48	1,44	1,41	1,38	1,36
27	1,91	1,94	1,71	1,60	1,53	1,48	1,44	1,41	1,38	1,36
28	1,91	1,94	1,71	1,60	1,53	1,48	1,44	1,41	1,38	1,36
29	1,91	1,94	1,72	1,60	1,53	1,48	1,44	1,41	1,38	1,36
30	1,91	1,94	1,72	1,60	1,53	1,48	1,44	1,41	1,38	1,36

p = number of laboratories at a given level

n = number of replicates within each laboratory at that level

NOTE — Supplied by Dr. J. Mandel and published with his permission.

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
$F_p(\nu_1, \nu_2)$	p-quantile of the F-distribution with ν_1 and ν_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

$\bar{y}$	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

Annex B
(informative)

Examples of the statistical analysis of precision experiments

B.1 Example 1: Determination of the sulfur content of coal (Several levels with no missing or outlying data)

B.1.1 Background

a) **Measurement method**

Determination of the sulfur content in coal with test results expressed as a percentage by mass.

b) **Source**

Tomkins, S. S. *Industrial and Engineering Chemistry.* (See reference [6] in annex C.)

c) **Description**

Eight laboratories participated in the experiment, carrying out the analysis according to a standardized measurement method described in the source cited. Laboratory 1 reported four test results and laboratory 5 reported four or five; the other laboratories all carried out three measurements.

d) **Graphical presentation**

Mandel's h and k statistics should be plotted, but because in this example they showed little of note they have been omitted in order to allow space for a different example of the graphical presentation of data. Mandel's plots are fully illustrated and discussed in the example given in B.3.

B.1.2 Original data

These are given, as percentage by mass [% (m/m)], in table B.1 in the format of form A of figure 2 (see 7.2.8) and do not invite any specific remarks.

Graphical presentations of these data are given in figures B.1 to B.4.

Table B.1 — Original data: Sulfur content of coal

Laboratory i	Level j			
	1	2	3	4
1	0,71 0,71 0,70 0,71	1,20 1,18 1,23 1,21	1,68 1,70 1,68 1,69	3,26 3,26 3,20 3,24
2	0,69 0,67 0,68	1,22 1,21 1,22	1,64 1,64 1,65	3,20 3,20 3,20
3	0,66 0,65 0,69	1,28 1,31 1,30	1,61 1,61 1,62	3,37 3,36 3,38
4	0,67 0,65 0,66	1,23 1,18 1,20	1,68 1,66 1,66	3,16 3,22 3,23
5	0,70 0,69 0,66 0,71 0,69	1,31 1,22 1,22 1,24 —	1,64 1,67 1,60 1,66 1,68	3,20 3,19 3,18 3,27 3,24
6	0,73 0,74 0,73	1,39 1,36 1,37	1,70 1,73 1,73	3,27 3,31 3,29
7	0,71 0,71 0,69	1,20 1,26 1,26	1,69 1,70 1,68	3,27 3,24 3,23
8	0,70 0,65 0,68	1,24 1,22 1,30	1,67 1,68 1,67	3,25 3,26 3,26

NOTE 8 For the experiment quoted in table B.1, the laboratories were not instructed as to how many measurements were to be made, only a minimum number. By the recommended procedures given in this part of ISO 5725, for laboratories 1 and 5 a random selection should be made from the values given in order to reduce all cells to exactly three test results. However, in order to illustrate the computational procedures for variable numbers of test results, all test results have been retained in this example. The reader may make random selections to reduce the number of test results to three in each cell if he/she wishes to verify that such a procedure has relatively little effect on the values of $\hat{m}_j$, s_r and s_R.

B.1.3 Computation of cell means ($\bar{y}_{ij}$)

The cell means are given, as a percentage by mass [% (*m/m*)], in table B.2 in the format of form B of figure 2 (see 7.2.9).

B.1.4 Computation of standard deviations (s_{ij})

The standard deviations are given, as a percentage by mass [% (*m/m*)], in table B.3 in the format of form C of figure 2 (see 7.2.10).

B.1.5 Scrutiny for consistency and outliers

Cochran's test with $n = 3$ for $p = 8$ laboratories gives critical values of 0,516 for 5 % and 0,615 for 1 %.

For level 1, largest value of s is in laboratory 8:

$\Sigma s^2 = 0{,}001\ 82$; test value = 0,347

For level 2, largest value of s is in laboratory 5:

$\Sigma s^2 = 0{,}006\ 36$; test value = 0,287

For level 3, largest value of s is in laboratory 5:

$\Sigma s^2 = 0{,}001\ 72$; test value = 0,598

For level 4, largest value of s is in laboratory 4:

$\Sigma s^2 = 0{,}004\ 63$; test value = 0,310

Table B.2 — Cell means: Sulfur content of coal

Laboratory i	Level j							
	1		2		3		4	
	$\bar{y}_{ij}$	n_{ij}	$\bar{y}_{ij}$	n_{ij}	$\bar{y}_{ij}$	n_{ij}	$\bar{y}_{ij}$	n_{ij}
1	0,708	4	1,205	4	1,688	4	3,240	4
2	0,680	3	1,217	3	1,643	3	3,200	3
3	0,667	3	1,297	3	1,613	3	3,370	3
4	0,660	3	1,203	3	1,667	3	3,203	3
5	0,690	5	1,248	4	1,650	5	3,216	5
6	0,733	3	1,373	3	1,720	3	3,290	3
7	0,703	3	1,240	3	1,690	3	3,247	3
8	0,677	3	1,253	3	1,673	3	3,257	3

Table B.3 — Standard deviations: Sulfur content of coal

Laboratory i	Level j							
	1		2		3		4	
	s_{ij}	n_{ij}	s_{ij}	n_{ij}	s_{ij}	n_{ij}	s_{ij}	n_{ij}
1	0,005	4	0,021	4	0,010	4	0,028	4
2	0,010	3	0,006	3	0,006	3	0,000	3
3	0,021	3	0,015	3	0,006	3	0,010	3
4	0,010	3	0,025	3	0,012	3	0,038	3
5	0,019	5	0,043	4	0,032	5	0,038	5
6	0,006	3	0,015	3	0,017	3	0,020	3
7	0,012	3	0,035	3	0,010	3	0,021	3
8	0,025	3	0,042	3	0,006	3	0,006	3

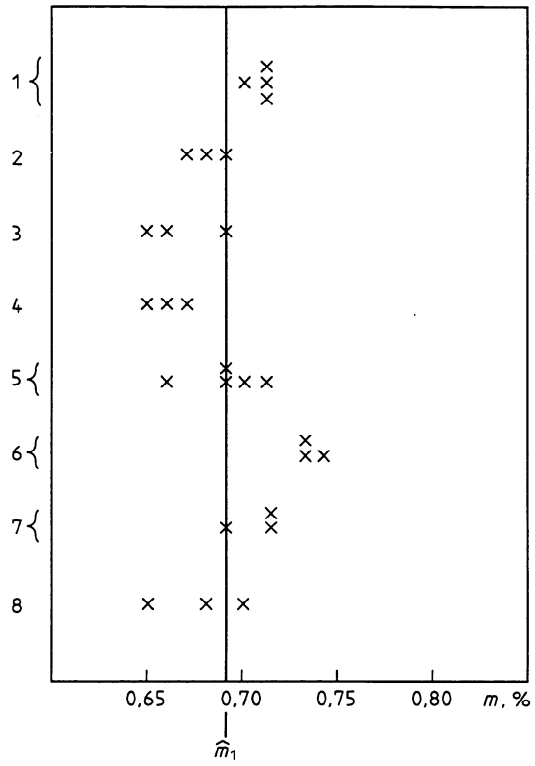

Figure B.1 — Sulfur content of coal, sample 1

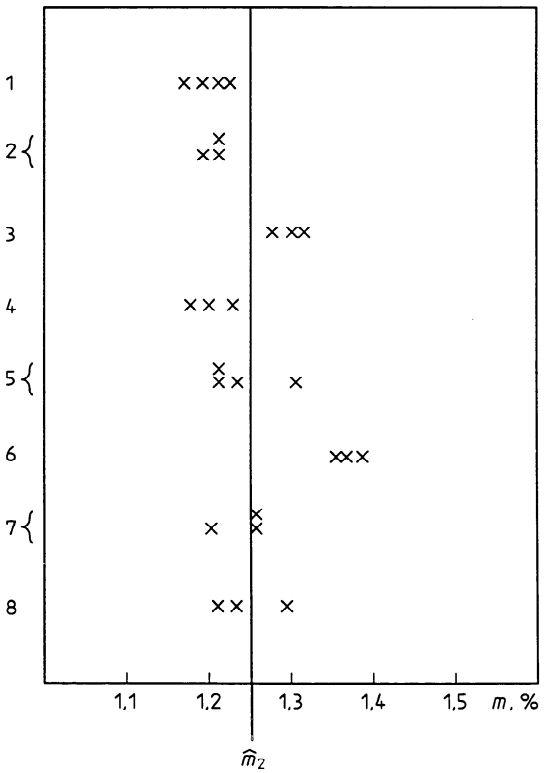

Figure B.2 — Sulfur content of coal, sample 2

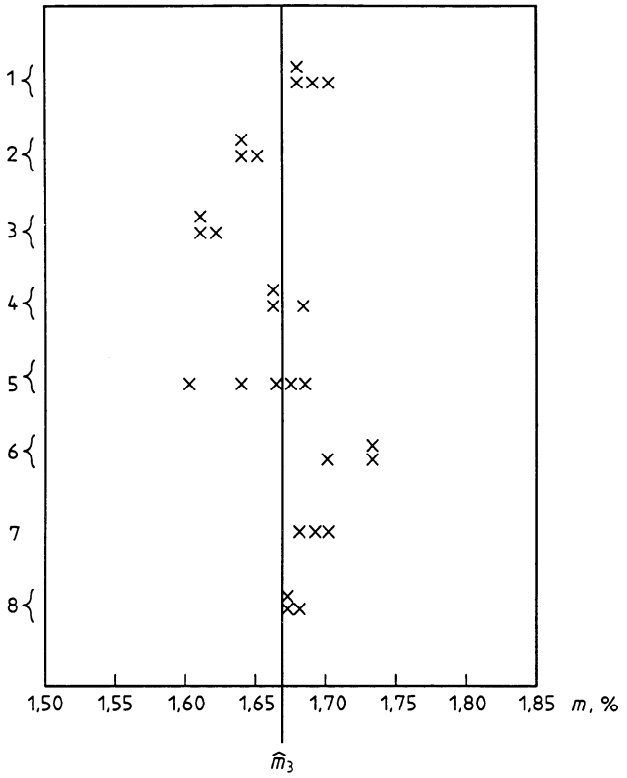

Figure B.3 — Sulfur content of coal, sample 3

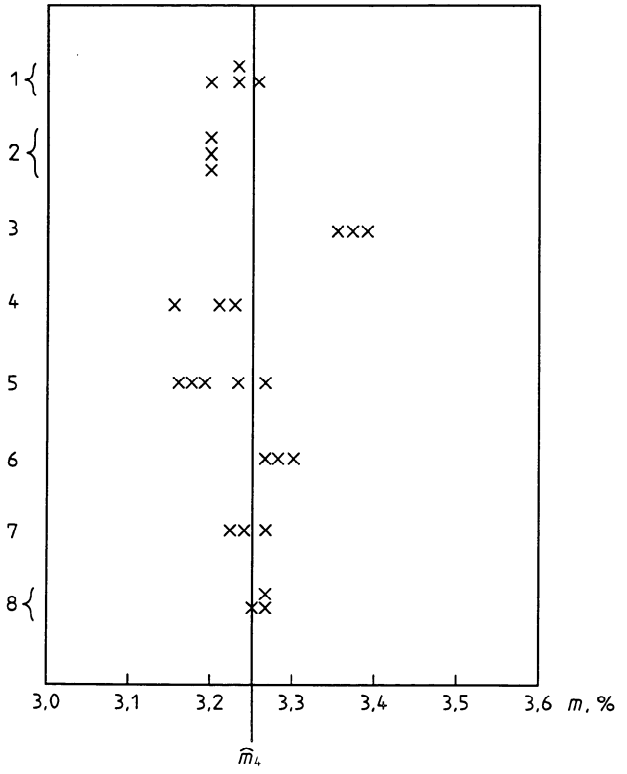

Figure B.4 — Sulfur content of coal, sample 4

This indicates that one cell in level 3 may be regarded as a straggler, and there are no outliers. The straggler is retained in subsequent calculations.

Grubbs' tests were applied to the cell means, giving the values shown in table B.4. There are no single stragglers or outliers. At levels 2 and 4, the high results for laboratories 3 and 6 are stragglers according to the double-high test; these were retained in the analysis.

B.1.6 Computation of $\hat{m}_j$, s_{rj} and s_{Rj}

The variances defined in 7.4.4 and 7.4.5 are calculated as follows, using level 1 as an example.

Number of laboratories, $p = 8$

$T_1 = \Sigma n_i \bar{y}_i = 18{,}642$

$T_2 = \Sigma n_i (\bar{y}_i)^2 = 12{,}883\ 7$

$T_3 = \Sigma n_i = 27$

$T_4 = \Sigma n_i^2 = 95$

$T_5 = \Sigma (n_i - 1) s_i^2 = 0{,}004\ 411$

$$s_r^2 = \frac{T_5}{T_3 - p} = 0{,}000\ 232\ 2$$

$$s_L^2 = \left[\frac{T_2\, T_3 - T_1^2}{T_3(p-1)} - s_r^2 \right] \left[\frac{T_3(p-1)}{T_3^2 - T_4} \right] =$$

$$= 0{,}000\ 460\ 3$$

$$s_R^2 = s_L^2 + s_r^2 = 0{,}000\ 692\ 5$$

$$\hat{m} = \frac{T_1}{T_3} = 0{,}690\ 44$$

$$s_r = 0{,}015\ 24$$

$$s_R = 0{,}026\ 32$$

The calculations for levels 2, 3 and 4 may be carried out similarly to give the results shown in table B.5.

B.1.7 Dependence of precision on m

An examination of the data in table B.5 does not indicate any dependence and average values can be used.

B.1.8 Conclusions

The precision of the measurement method should be quoted, as a percentage by mass, as

repeatability standard deviation, $s_r = 0{,}022$

reproducibility standard deviation, $s_R = 0{,}045$

Table B.4 — Application of Grubbs' test to cell means

Level	Single low	Single high	Double low	Double high	Type of test
1	1,24	1,80	0,539	0,298	
2	0,91	2,09	0,699	0,108	Grubbs' test statistics
3	1,67	1,58	0,378	0,459	
4	0,94	2,09	0,679	0,132	
Stragglers	2,126	2,126	0,110 1	0,110 1	Grubbs' critical values
Outliers	2,274	2,274	0,056 3	0,056 3	

Table B.5 — Computed values of $\hat{m}_j$, s_{rj} and s_{Rj} for sulfur content of coal

Level j	p_j	$\hat{m}_j$	s_{rj}	s_{Rj}
1	8	0,690	0,015	0,026
2	8	1,252	0,029	0,061
3	8	1,667	0,017	0,035
4	8	3,250	0,026	0,058

These values may be applied within a range 0,69 % (m/m) to 3,25 % (m/m). They were determined from a uniform-level experiment involving 8 laboratories covering that range of values, in which four stragglers were detected and retained.

B.2 Example 2: Softening point of pitch (Several levels with missing data)

B.2.1 Background

a) **Measurement method**

The determination of the softening point of pitch by ring and ball.

b) **Source**

Standard methods for testing tar and its products; Pitch section; Method Serial No. PT3 using neutral glycerine (reference [5] in annex C).

c) **Material**

This was selected from commercial batches of pitch collected and prepared as specified in the "Samples" chapter of the pitch section of reference [5].

d) **Description**

This was the determination of a property involving temperature measurement in degrees Celsius. Sixteen laboratories cooperated. It was intended to measure four specimens at about 87,5 °C, 92,5 °C, 97,5 °C and 102,5 °C to cover the normal commercial range of products, but wrong material was chosen for level 2 with a mean temperature of about 96 °C which was similar to level 3. Laboratory 5 applied the method incorrectly at first on the sample for level 2 (the first one they measured) and there was then insufficient material remaining for more than one determination. Laboratory 8 found that they did not have a sample for level 1 (they had two specimens for level 4).

e) **Graphical presentations**

Mandel's h and k statistics should be plotted, but again in this example they have been omitted in order to provide for another type of graphical presentation of data. Mandel's plots are fully illustrated and discussed in the example given in B.3.

B.2.2 Original data

These are presented in table B.6, in degrees Celsius, in the format of form A of figure 2 (see 7.2.8).

Table B.6 — Original data: Softening point of pitch (°C)

Laboratory i	Level j			
	1	2	3	4
1	91,0 89,6	97,0 97,2	96,5 97,0	104,0 104,0
2	89,7 89,8	98,5 97,2	97,2 97,0	102,6 103,6
3	88,0 87,5	97,8 94,5	94,2 95,8	103,0 99,5
4	89,2 88,5	96,8 97,5	96,0 98,0	102,5 103,5
5	89,0 90,0	97,2 —	98,2 98,5	101,0 100,2
6	88,5 90,5	97,8 97,2	99,5 103,2	102,2 102,0
7	88,9 88,2	96,6 97,5	98,2 99,0	102,8 102,2
8	— —	96,0 97,5	98,4 97,4	102,6 103,9
9	90,1 88,4	95,5 96,8	98,2 96,7	102,8 102,0
10	86,0 85,8	95,2 95,0	94,8 93,0	99,8 100,8
11	87,6 84,4	93,2 93,4	93,6 93,9	98,2 97,8
12	88,2 87,4	95,8 95,4	95,8 95,4	101,7 101,2
13	91,0 90,4	98,2 99,5	98,0 97,0	104,5 105,6
14	87,5 87,8	97,0 95,5	97,1 96,6	105,2 101,8
15	87,5 87,6	95,0 95,2	97,8 99,2	101,5 100,9
16	88,8 85,0	95,0 93,2	97,2 97,8	99,5 99,8

NOTE — There are no obvious stragglers or statistical outliers.

B.2.3 Cell means

These are given in table B.7, in degrees Celsius, in the format of form B of figure 2 (see 7.2.9).

A graphical presentation of these data is given in figure B.5.

B.2.4 Absolute differences within cells

In this example there are two test results per cell and the absolute difference can be used to represent the variability. The absolute differences within cells, in degrees Celsius, are given in table B.8, in the format of form C of figure 2 (see 7.2.10).

A graphical presentation of these data is given in figure B.6.

Table B.7 — Cell means: Softening point of pitch (°C)

Laboratory i	Level j			
	1	2	3	4
1	90,30	97,10	96,75	104,00
2	89,75	97,85	97,10	103,10
3	87,75	96,15	95,00	101,25
4	88,85	97,15	97,00	103,00
5	89,50	—	98,35	100,60
6	89,50	97,50	101,35	102,10
7	88,55	97,05	98,60	102,50
8	—	96,75	97,90	103,25
9	89,25	96,15	97,45	102,40
10	85,90	95,10	93,90	100,30
11	86,00	93,30	93,75	98,00
12	87,80	95,60	95,60	101,45
13	90,70	98,85	97,50	105,05
14	87,65	96,25	96,85	103,50
15	87,55	95,10	98,50	101,20
16	86,90	94,10	97,50	99,65

NOTE — The entry for $i = 5$, $j = 2$ has been dropped (see 7.4.3).

Table B.8 — Absolute differences within cells: Softening point of pitch (°C)

Laboratory i	Level j			
	1	2	3	4
1	1,4	0,2	0,5	0,0
2	0,1	1,3	0,2	1,0
3	0,5	3,3	1,6	3,5
4	0,7	0,7	2,0	1,0
5	1,0	—	0,3	0,8
6	2,0	0,6	3,7	0,2
7	0,7	0,9	0,8	0,6
8	—	1,5	1,0	1,3
9	1,7	1,3	1,5	0,8
10	0,2	0,2	1,8	1,0
11	3,2	0,2	0,3	0,4
12	0,8	0,4	0,4	0,5
13	0,6	1,3	1,0	1,1
14	0,3	1,5	0,5	3,4
15	0,1	0,2	1,4	0,6
16	3,8	1,8	0,6	0,3

B.2.5 Scrutiny for consistency and outliers

Application of Cochran's test leads to the values of the test statistic C given in table B.9.

The critical values (see 8.1) at the 5 % significance level are 0,471 for $p = 15$ and 0,452 for $p = 16$ where $n = 2$. No stragglers are indicated.

Grubbs' tests were applied to the cell means. No single or double stragglers or outliers were found.

B.2.6 Computation of $\hat{m}_j$, s_{rj} and s_{Rj}

These are calculated as in 7.4.4 and 7.4.5.

Using level 1 for example, the calculations are as follows. To ease the arithmetic, 80,00 has been subtracted from all the data. The method for $n = 2$ replicates per cell is used.

Number of laboratories, $p = 15$

Number of replicates, $n = 2$

$T_1 = \Sigma\bar{y}_i = 125{,}950\ 0$

$T_2 = \Sigma(\bar{y}_i)^2 = 1\ 087{,}977\ 5$

$T_3 = (y_{i1} - y_{i2})^2 = 36{,}910\ 0$

$s_r^2 = \dfrac{T_3}{2p} = 1{,}230\ 3$

$s_L^2 = \left[\dfrac{pT_2 - T_1^2}{p(p-1)}\right] - \dfrac{s_r^2}{2} = 1{,}557\ 5$

$s_R^2 = s_L^2 + s_r^2 = 2{,}787\ 8$

$\hat{m} = \dfrac{T_1}{p}$ (add 80,00) $= 88{,}396\ 6$

$s_r = 1{,}109\ 2$

$s_R = 1{,}669\ 7$

The values for all four levels are given in table B.11.

Table B.9 — Values of Cochran's test statistic, C

Level j	1	2	3	4
C	0,391 (15)	0,424 (15)	0,434 (16)	0,380 (16)
NOTE — Number of laboratories is given in parentheses.				

Table B.10 — Application of Grubbs' test to cell means

Level; n	Single low	Single high	Double low	Double high	Type of test
1; 15	1,69	1,56	0,546	0,662	Grubbs' test statistics
2; 15	2,04	1,77	0,478	0,646	
3; 16	1,76	2,27	0,548	0,566	
4; 16	2,22	1,74	0,500	0,672	
Stragglers					Grubbs' critical values
$n = 15$	2,549	2,549	0,336 7	0,336 7	
$n = 16$	2,585	2,585	0,360 3	0,360 3	
Outliers					
$n = 15$	2,806	2,806	0,253 0	0,253 0	
$n = 16$	2,852	2,852	0,276 7	0,276 7	

Table B.11 — Computed values of $\hat{m}_j$, s_{rj} and s_{Rj} for softening point of pitch

Level j	p_j	$\hat{m}_j$ (°C)	s_{rj}	s_{Rj}
1	15	88,40	1,109	1,670
2	15	96,27	0,925	1,597
3	16	97,07	0,993	2,010
4	16	101,96	1,004	1,915

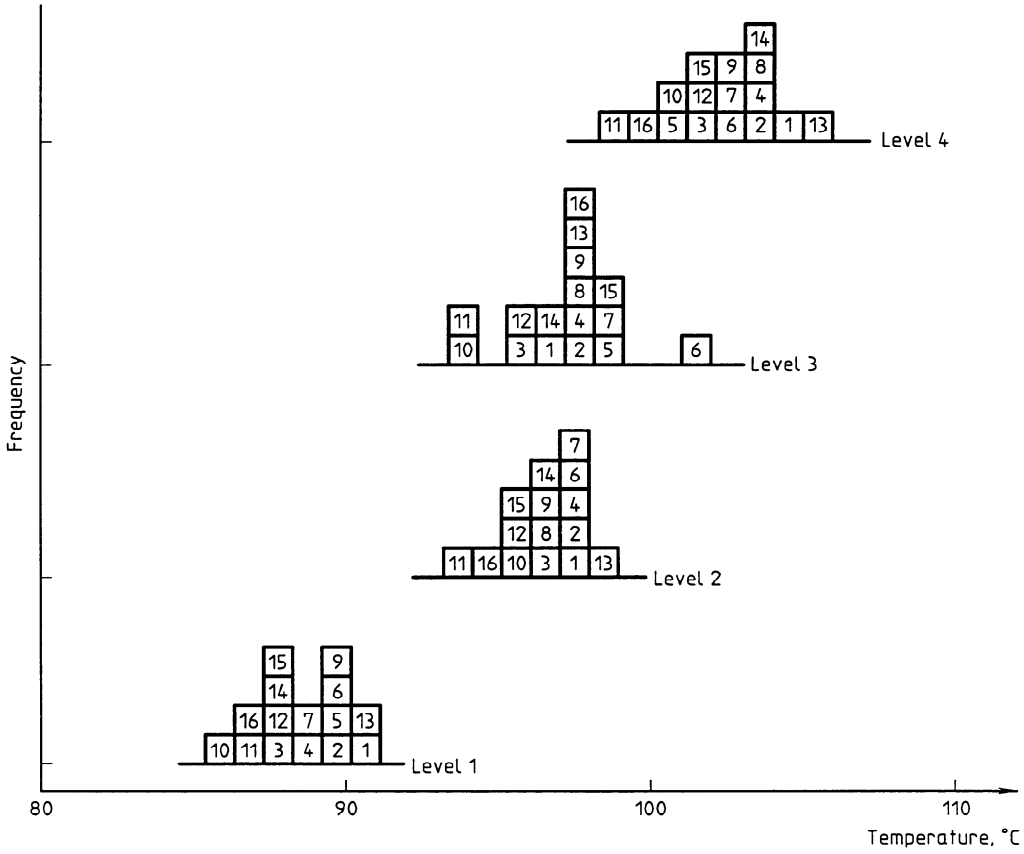

Figure B.5 — Softening point of pitch: Cell means

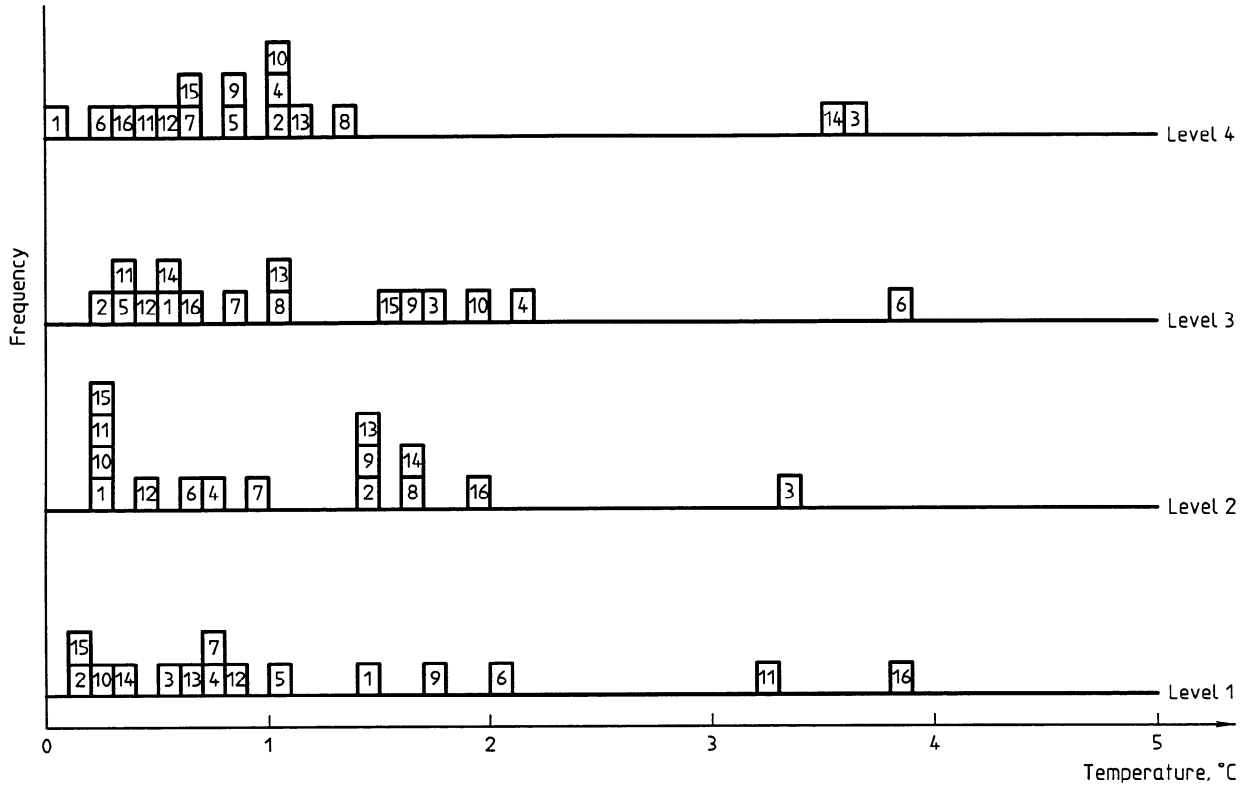

Figure B.6 — Softening point of pitch: Absolute differences within cells

B.2.7 Dependence of precision on m

A cursory examination of table B.11 does not reveal any marked dependence, except perhaps in reproducibility. The changes over the range of values of m, if any at all, are too small to be considered significant. Moreover, in view of the small range of values of m and the nature of the measurement, a dependence on m is hardly to be expected. It seems safe to conclude that precision does not depend on m in this range, which was stated as covering normal commercial material, so that the means may be taken as the final values for repeatability and reproducibility standard deviations.

B.2.8 Conclusions

For practical applications, the precision values for the measurement method can be considered as independent of the level of material, and are

repeatability standard deviation, $s_r = 1{,}0$ °C

reproducibility standard deviation, $s_R = 1{,}8$ °C

B.3 Example 3: Thermometric titration of creosote oil (Several levels with outlying data)

B.3.1 Background

a) **Source**

Standard methods for testing tar and its products; Creosote oil section; Method Serial No. Co. 18 (reference [5] in annex C).

b) **Material**

This was selected from commercial batches of creosote oil collected and prepared as specified in the "Samples" chapter of the creosote oil section of reference [5].

c) **Description**

This was a standard measurement method for chemical analysis involving a thermometric titration, with results expressed as a percentage by mass. Nine laboratories participated by measuring five specimens in duplicate, the specimens

measured having been selected so as to cover the normal range expected to be encountered in general commercial application. These were chosen to lie at the approximate levels of 4, 8, 12, 16 and 20 [% (*m/m*)]. The usual practice would be to record test results to only one decimal place, but for this experiment operators were instructed to work to two decimal places.

B.3.2 Original data

These are presented in table B.12, as a percentage by mass, in the format of form A of figure 2 (see 7.2.8).

The test results for laboratory 1 were always higher, and at some levels considerably higher, than those of the other laboratories.

The second test result for laboratory 6 at level 5 is suspect; the value recorded would fit much better at level 4.

These points are discussed further in B.3.5.

B.3.3 Cell means

These are given in table B.13, as a percentage by mass, in the format of form B of figure 2 (see 7.2.9).

Table B.12 — Original data: Thermometric titration of creosote oil

Laboratory i	Level j									
	1		2		3		4		5	
1	4,44	4,39	9,34	9,34	17,40	16,90	19,23	19,23	24,28	24,00
2	4,03	4,23	8,42	8,33	14,42	14,50	16,06	16,22	20,40	19,91
3	3,70	3,70	7,60	7,40	13,60	13,60	14,50	15,10	19,30	19,70
4	4,10	4,10	8,93	8,80	14,60	14,20	15,60	15,50	20,30	20,30
5	3,97	4,04	7,89	8,12	13,73	13,92	15,54	15,78	20,53	20,88
6	3,75	4,03	8,76	9,24	13,90	14,06	16,42	16,58	18,56	16,58
7	3,70	3,80	8,00	8,30	14,10	14,20	14,90	16,00	19,70	20,50
8	3,91	3,90	8,04	8,07	14,84	14,84	15,41	15,22	21,10	20,78
9	4,02	4,07	8,44	8,17	14,24	14,10	15,14	15,44	20,71	21,66

Table B.13 — Cell means: Thermometric titration of creosote oil

Laboratory i	Level j				
	1	2	3	4	5
1	4,415	9,340	17,150**	19,230**	24,140*
2	4,130	8,375	14,460	16,140	20,155
3	3,700	7,500	13,600	14,800	19,500
4	4,100	8,865	14,400	15,550	20,300
5	4,005	8,005	13,825	15,660	20,705
6	3,890	9,000	13,980	16,500	17,570
7	3,750	8,150	14,150	15,450	20,100
8	3,905	8,055	14,840	15,315	20,940
9	4,045	8,305	14,170	15,290	21,185

* Regarded as a straggler.

** Regarded as a statistical outlier.

B.3.4 Absolute differences within cells

These are given in table B.14, as w_{ij}, as a percentage by mass, in the format of form C of figure 2 (see 7.2.10).

B.3.5 Scrutiny for consistency and outliers

Calculation of Mandel's h and k consistency statistics (see 7.3.1) gave the values shown in figures B.7 and B.8. Horizontal lines are shown corresponding to the value of Mandel's indicators taken from 8.3.

The h graph (figure B.7) shows clearly that laboratory 1 obtained much higher test results than all other laboratories at all levels. Such results require attention on the part of the committee running the interlaboratory study. If no explanations can be found for these test results, the members of the committee should use their judgement, based on additional and perhaps non-statistical considerations, in deciding whether to include or exclude this laboratory in the calculation of the precision values.

The k graph (figure B.8) exhibits rather large variability between replicate test results for laboratories 6 and 7. However, these test results do not seem so severe as to require any special action beyond a search for possible explanations and, if necessary, remedial action for these test results.

Application of Cochran's test yields the following results.

At level 4, the absolute difference 1,10 gave a test statistic value of $1{,}10^2/1{,}814\ 9 = 0{,}667$.

At level 5, the absolute difference 1,98 gave a test statistic value of $1{,}98^2/6{,}166\ 3 = 0{,}636$.

For $p = 9$, the critical values for Cochran's test are 0,638 for 5 %, and 0,754 for 1 %.

The value 1,10 at level 4 is clearly a straggler, and the value 1,98 at level 5 is so near the 5 % level as to be also a possible straggler. As these two values are so different from all the others, and as their presence has inflated the divisor used in Cochran's test statistic, they have both been regarded as stragglers and marked with an asterisk. The evidence against them so far, however, cannot be regarded as sufficient for rejection, although Mandel's k plot (figure B.8) also gives rise to suspicion of these values.

Application of Grubbs' tests to the cell means gives the results shown in table B.15.

For levels 3 and 4, because the single Grubbs test indicates an outlier, the double Grubbs test is not applied (see 7.3.4).

The cell means for laboratory 1 in levels 3 and 4 are found to be outliers. The cell mean for this laboratory for level 5 is also high. This is also clearly indicated on Mandel's h plot (figure B.7).

On further enquiry, it was learned that at least one of the samples for laboratory 6, level 5, might by mistake have come from level 4. As the absolute difference for this cell was also suspect, it was decided that this pair of test results may also have to be rejected. Without the "help" of this pair of values, the test result for laboratory 1 at level 5 is now definitely suspicious.

Table B.14 — Cell ranges: Thermometric titration of creosote oil

Laboratory i	Level j				
	1	2	3	4	5
1	0,05	0,00	0,50	0,00	0,28
2	0,20	0,09	0,08	0,16	0,49
3	0,00	0,20	0,00	0,60	0,40
4	0,00	0,13	0,40	0,10	0,00
5	0,07	0,23	0,19	0,24	0,35
6	0,28	0,48	0,16	0,16	1,98*
7	0,10	0,30	0,10	1,10*	0,80
8	0,01	0,03	0,00	0,19	0,32
9	0,05	0,27	0,14	0,30	0,95

* Regarded as a straggler.

Because of these test results, it was decided to reject the pair of test results from laboratory 6 for level 5 because it was uncertain what material had been measured and to reject all the test results from laboratory 1 as coming from an outlying laboratory.

Without these test results, the Cochran's test statistic at level 4 was then compared with the critical value for 8 laboratories (0,680 at 5 %) and this no longer appeared as a straggler and was retained.

Table B.15 — Application of Grubbs' test to cell means

Level	Single low	Single high	Double low	Double high	Type of test
1	1,36	1,95	0,502	0,356	Grubbs' test statistics
2	1,57	1,64	0,540	0,395	
3	0,86	2,50	—	—	
4	0,91	2,47	—	—	
5	1,70	2,10	0,501	0,318	
Stragglers	2,215	2,215	0,149 2	0,149 2	Grubbs' critical values
Outliers	2,387	2,387	0,085 1	0,085 1	

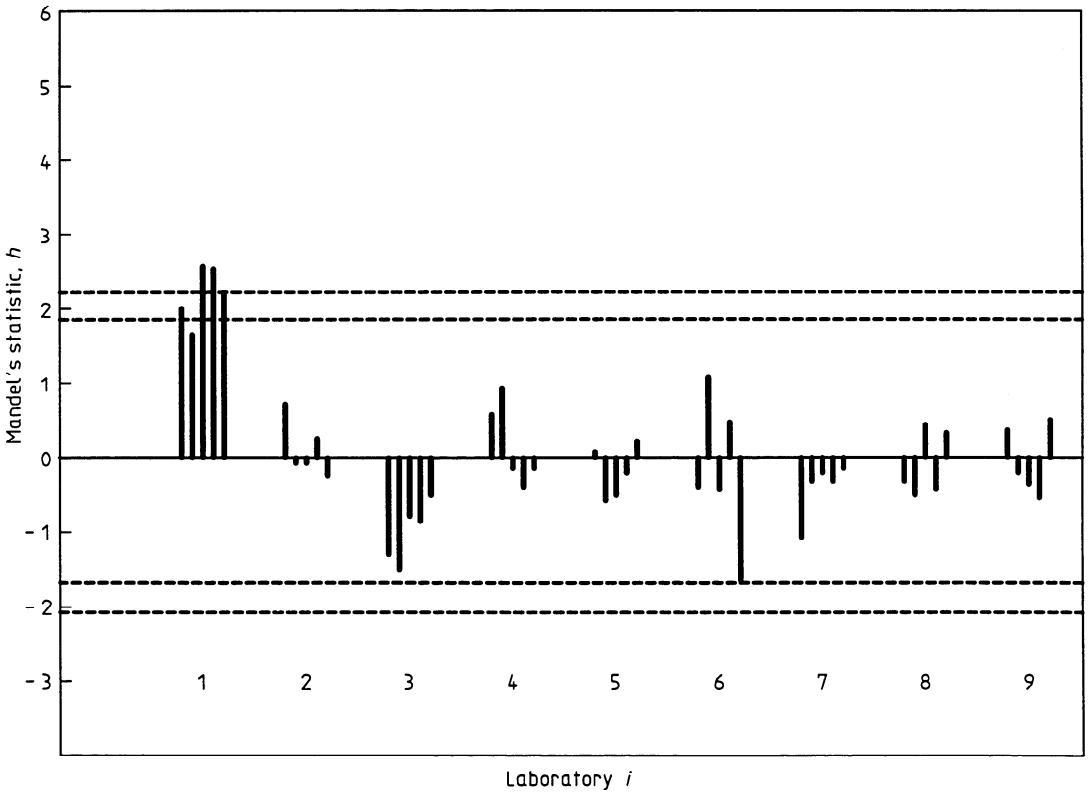

Figure B.7 — Titration of creosote oil: Mandel's between-laboratory consistency statistic, *h*, grouped by laboratories

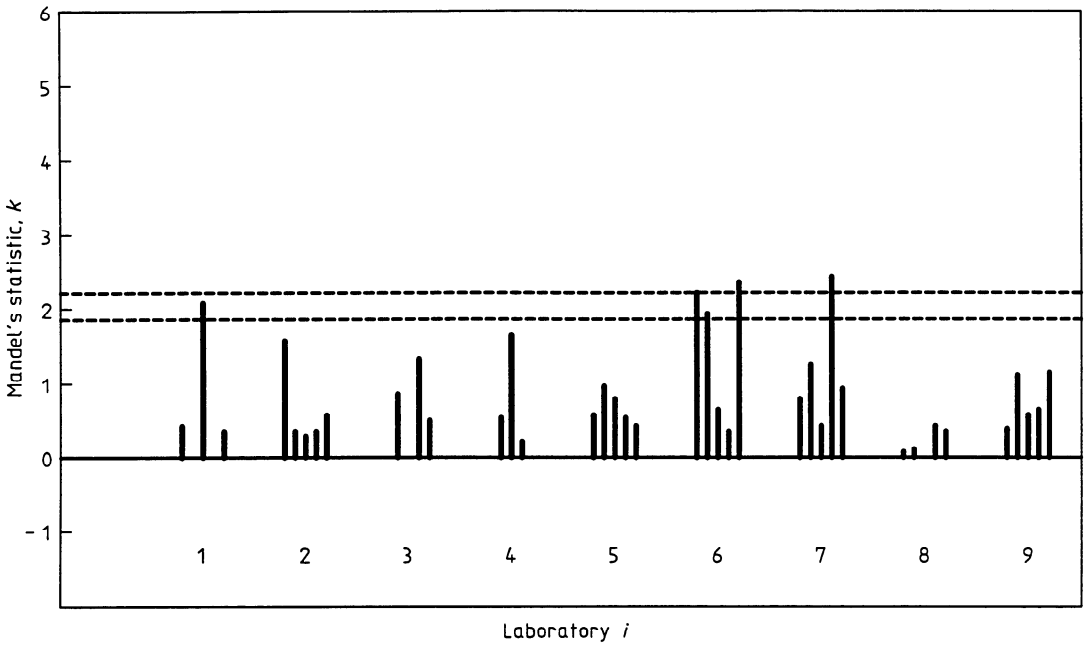

Figure B.8 — Titration of creosote oil: Mandel's within-laboratory consistency statistic, k, grouped by laboratories

B.3.6 Computation of $\hat{m}_j$, s_{rj} and s_{Rj}

The values of $\hat{m}_j$, s_{rj} and s_{Rj} computed without the test results of laboratory 1 and the pair of test results from laboratory 6, level 5, are given in table B.16, as a percentage by mass, calculated as in 7.4.4 and 7.4.5.

B.3.7 Dependence of precision on m

From table B.16, it seems clear that the standard deviations tend to increase with higher values of m, so it is likely that it might be permissible to establish some form of functional relationship. This view was supported by a chemist familiar with the measurement method, who was of the view that the precision was likely to be dependent on the level.

The actual calculations for fitting a functional relationship are not given here as they have already been set out in detail for s_r in 7.5.9. The values of s_{rj} and s_{Rj} are plotted against $\hat{m}_j$ in figure B.9.

From figure B.9 it is evident that the value for level 3 is strongly divergent and could not be improved by any alternative procedures (see 7.5.2).

For repeatability, a straight line through the origin seems adequate.

For reproducibility, all three lines show adequate fit with the data, relationship III showing the best fit. Someone familiar with the requirements for a standard measurement method for creosote oil may be able to select the most suitable relationship.

B.3.8 Final values of precision

The final values, duly rounded, should be

repeatability standard deviation, $s_r = 0{,}019m$

reproducibility standard deviation,
$s_R = 0{,}086 + 0{,}030m$ or

$s_R = 0{,}078m^{0{,}72}$

B.3.9 Conclusions

There are no statistical reasons for preferring either one of the two equations for s_R in B.3.8. The panel should decide which one to use.

The reason for the outlying test results of laboratory 1 should be investigated.

This seems to have been a rather unsatisfactory precision experiment. One of the 9 laboratories had to be rejected as an outlier, and another laboratory had tested a wrong specimen. The material for level 3 seems to have been wrongly selected, having almost the same value as level 4 instead of lying midway

between levels 2 and 4. Moreover, the material for level 3 seems to have been somewhat different in nature, perhaps being more homogeneous than the other material. It might be worthwhile to repeat this experiment, taking more care over the selection of the materials for the different levels.

Table B.16 — Computed values of $\hat{m}_j$, s_{rj} and s_{Rj} for thermometric titration of creosote oil

Level j	p_j	$\hat{m}_j$	s_{rj}	s_{Rj}
1	8	3,94	0,092	0,171
2	8	8,28	0,179	0,498
3	8	14,18	0,127	0,400
4	8	15,59	0,337	0,579
5	7	20,41	0,393	0,637

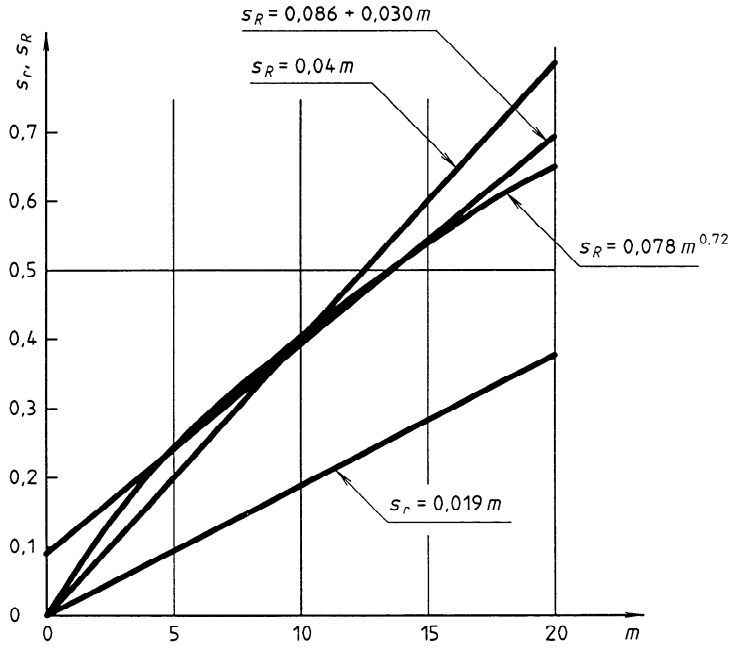

Figure B.9 — Plot of s_{rj} and s_{Rj} against $\hat{m}_j$ of the data from table B.16, showing the functional relationships fitted in 7.5 from these data

Annex C
(informative)

Bibliography

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

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

[3] ASTM E691-87, *Standard Practice for Conducting an Interlaboratory Study to Determine the Precision of a Test Method.* American Society for Testing and Materials, Philadelphia, PA, USA.

[4] GRUBBS, F.E. and BECK, G. Extension of sample sizes and percentage points for significance tests of outlying observations. *Technometrics,* **14**, 1972, pp. 847-854.

[5] "Standard Methods for Testing Tar and its Products". 7th Ed. Standardisation of Tar Products Tests Committee, 1979.

[6] TOMKINS, S.S. *Industrial and Engineering Chemistry* (Analytical edition), **14**, 1942, pp. 141-145.

[7] GRUBBS, F.E. Procedures for detecting outlying observations in samples. *Technometrics,* **11**, 1969, pp. 1-21.

[8] ISO 3534-2:1993, *Statistics — Vocabulary and symbols — Part 2: Statistical quality control.*

[9] ISO 3534-3:1985, *Statistics — Vocabulary and symbols — Part 3: Design of experiments.*

[10] 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.*

[11] 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.*

[12] ISO 5725-5:—[1], *Accuracy (trueness and precision) of measurement methods and results — Part 5: Alternative methods for the determination of the precision of a standard measurement method.*

[13] ISO 5725-6:1994, *Accuracy (trueness and precision) of measurement methods and results — Part 6: Use in practice of accuracy values.*

1) To be published.

This page intentionally left blank

This page intentionally left blank

This page intentionally left blank

ICS 03.120.30

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

Price based on 42 pages