भारतीय मानक
Indian Standard

IS 15393 (Part 2) : 2021
ISO 5725-2 : 2019

मापन पद्धतियों तथा परिणामों की यथार्थता (वास्तविक एंव परिशुद्ध)

भाग 2 मानक मापन पद्धति की पुनरावर्तनीयता तथा पुर्नउत्पादकता की मूल पद्धति ज्ञात करना

(पहला पुनरीक्षण)

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

(First Revision)

ICS 03.120.30; 17.020

भारतीय मानक ब्यूरो
BUREAU OF INDIAN STANDARDS
मानक भवन, 9 बहादुरशाह ज़फर मार्ग, नई दिल्ली – 110002
MANAK BHAVAN, 9 BAHADUR SHAH ZAFAR MARG
NEW DELHI-110002
www.bis.gov.in www.standardsbis.in

April 2021 **Price Group 12**

NATIONAL FOREWORD

This Indian Standard (Part 2) (First Revision) which is identical with ISO 5725-2 : 2019 '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' issued by the International Organization for Standardization (ISO) was adopted by the Bureau of Indian Standards on recommendation of Statistical Methods for Quality and Reliability Sectional Committee and approval of the Management and Systems Division Council.

This standard was originally published in 2003 which was identical with ISO 5724 : 1994. The first revision of this standard has been undertaken to align it with latest version of ISO 5725-2 : 2019.

The major changes in this revision are as follows:

a) permission is given to use alternative scrutiny and outlier detection tests provided that the performance is similar;

b) permission is given to apply modern statistical methods available for calculations of the relevant precision and trueness characteristics;

c) guidance on the number of laboratories required for a precision study has been included; and

d) information on the computation of critical values has been included.

This Indian Standard is published in various parts. The other parts in this series are:

Part 1	General principles and definitions
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

The text of ISO Standard has been approved as suitable for publication as an Indian Standard without deviations. Certain terminologies and conventions are, however, not identical to those used in Indian Standards. Attention is particularly drawn to the following:

a) Wherever the words 'International Standard' appear referring to this standard, they should be read as 'IndianStandard'.

b) Comma (,) has been used as a decimal marker, while in Indian Standards, the current practice is to use a point (.) as the decimalmarker.

In this adopted standard, reference appears to certain International Standards for which Indian Standards also exist. The corresponding Indian Standards, which are to be substituted in their respective place, are listed below along with their degree of equivalence for the editions indicated:

International Standard	*Corresponding Indian Standard*	*Degree of Equivalence*
ISO 3534-1 Statistics — Vocabulary and symbols — Part 1: General statistical terms and terms used in probability	IS 7920 (Part 1) : 2012 Statistics — Vocabulary and symbols: Part 1 General statistical terms and terms used in probability (*third revision*)	Modified

(*Continued on third cover*)

Contents

Page

Introduction ... **iii**

1 Scope ... **1**

2 Normative references ... **2**

3 Terms and definitions ... **2**

4 Symbols ... **2**

5 Estimates of the parameters in the basic model ... **4**

6 Requirements for a precision experiment ... **5**
6.1 Layout of the experiment ... 5
6.2 Recruitment of the laboratories ... 6
6.3 Preparation of the materials ... 6

7 Personnel involved in a precision experiment ... **7**
7.1 Panel ... 7
7.2 Statistical functions ... 8
7.3 Executive functions ... 8
7.4 Supervisors ... 9
7.5 Operators ... 10

8 Statistical analysis of a precision experiment ... **10**
8.1 Preliminary considerations ... 10
8.2 Tabulation of the results and notation used ... 11
8.2.1 Cells ... 11
8.2.2 Redundant data ... 11
8.2.3 Missing data ... 11
8.2.4 Outliers ... 11
8.2.5 Outlying laboratories ... 11
8.2.6 Erroneous data ... 11
8.2.7 Balanced uniform-level test results ... 11
8.2.8 Collation of data and intermediate values ... 12
8.2.9 Original test results ... 12
8.2.10 Cell means (Form B of Figure 2) ... 12
8.2.11 Measures of cell spread (Form C of Figure 2) ... 12
8.2.12 Corrected or rejected data ... 13
8.3 Scrutiny of results for consistency and outliers ... 13
8.3.1 Approaches for scrutiny of data ... 13
8.3.2 Graphical consistency technique ... 13
8.3.3 Numerical outlier technique ... 16
8.3.4 Cochran's test ... 16
8.3.5 Grubbs' tests ... 18
8.3.6 Repeated testing for outlying means or outlying data points ... 20
8.3.7 Alternative outlier inspection and test methods ... 20
8.4 Calculation of the general mean and variances ... 20
8.4.1 Method of analysis ... 20
8.4.2 Basic data ... 21
8.4.3 Non-empty cells ... 21
8.4.4 Calculation of the general mean, $\hat{m}$ ... 21
8.4.5 Calculation of variances ... 21
8.4.6 Alternative calculation methods for variances ... 22
8.4.7 Dependence of the variances upon m ... 23
8.5 Establishing a functional relationship between precision values, s, and the mean level, m ... 23
8.5.1 Choice of functional relationship ... 23

8.5.2 Fitting relationships I and II 24
8.5.3 Fitting relationship III 25
8.5.4 Fitting relationship IV 26
8.6 Statistical analysis as a step-by-step procedure 28
8.7 Report to the panel and decisions to be taken by the panel 30
8.7.1 Report by the statistical expert 30
8.7.2 Decisions to be taken by the panel 32
8.7.3 Full report 33
9 Statistical tables 33
Annex A (informative) **Number of laboratories required for an estimate of precision 38**
Annex B (informative) **Alternative calculations of variance components 41**
Annex C (informative) **Examples of the statistical analysis of precision experiments 44**
Annex D (informative) **Calculation of critical values and indicators 66**
Bibliography 69

Introduction

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.

General consideration of these quantities is given in ISO 5725-1 and so is not repeated in this document. 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.

This document is concerned solely with estimating the repeatability standard deviation and reproducibility standard deviation based on an interlaboratory design in which each laboratory conducts a number of independent measurements of the same sample under repeatability conditions. There are other designs (such as nested, factorial or split-level experiments) which can be used for the estimation of precision: these are not dealt with in this document but rather are the subject of other parts of ISO 5725. Nor does this document consider any other measures of precision intermediate between the two principal measures; those are the subject of ISO 5725-3.

In certain circumstances, the data obtained from an experiment carried out to estimate precision are used also to estimate trueness and can be used to evaluate measurement uncertainty. The estimation of trueness is not considered in this document; all aspects of the estimation of trueness are the subject of ISO 5725-4. The evaluation of measurement uncertainty, using inter-laboratory estimates of trueness and precision, is the subject of ISO 21748.

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

Indian Standard

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

(First Revision)

1 Scope

1.1 This document

— amplifies the general principles for 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 Modifications to this basic method for particular purposes are given in other parts of ISO 5725.

1.2 It 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 can 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 are 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.

1.4 The statistical model of ISO 5725-1:1994, Clause 5, 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 document, (usually) estimates 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 documents are referred to in the text in such a way that some or all of their content constitutes requirements of this document. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

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

ISO 3534-2, *Statistics — Vocabulary and symbols — Part 2: Applied statistics*

ISO 3534-3, *Statistics — Vocabulary and symbols — Part 3: Design of experiments*

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

3 Terms and definitions

For the purposes of this document, the definitions given in ISO 3534-1, ISO 3534-2, ISO 3534-3, and ISO 5725-1 apply.

ISO and IEC maintain terminological databases for use in standardization at the following addresses:

— ISO Online browsing platform: available at https://www.iso.org/obp

— IEC Electropedia: available at http://www.electropedia.org/

4 Symbols

α	Probability associated with a critical value of a test statistic, also referred to as a level of significance
a	Intercept in the relationship $s = a + bm$
a_v	Intercept parameter in the relationship $s_j^2 = a_v^2 + (b_v m)^2$
A	Factor used to calculate the uncertainty of an estimate
b	Slope in the relationship $s = a + bm$
b_v	Slope parameter in the relationship $s_j^2 = a_v^2 + (b_v m)^2$
B	Laboratory component of bias under repeatability 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
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
G	Grubbs' test statistic
h	Mandel's between-laboratory consistency test statistic
k	Mandel's within-laboratory consistency test statistic

$L(\theta)$	Log-likelihood for variance components θ
m	General mean of the test property; level
$\hat{m}$	Estimate of the general mean of the test property
$\mathbf{M}$	Transformation matrix used in REML estimation
N	Number of iterations
n	Number of test results obtained in one laboratory at one level (i.e. per cell)
n_j	Total number of test results obtained at level j of the interlaboratory experiment
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
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
$\mathbf{V}(\boldsymbol{\theta})$	Covariance matrix used in REML estimation
W	Weighting factor used in calculating a weighted regression
w	Weighting factor used in calculating a weighted mean
x	Datum used for Grubbs' test
$\mathbf{X}$	Design matrix for REML estimations
y	Test result
$\bar{\bar{y}}$	Grand mean of test results
$\mathbf{Y}$	Vector of all observations at a level j
θ	Vector of variance components used in REML estimation
μ	True value or accepted reference value of a test property
σ	True value of a standard deviation

Subscripts

i	Identifier for a particular laboratory Index for summation (Annex A)
j	Identifier for a particular level Index for summation (Annex A)
k	Identifier for a particular test result in a laboratory *i* at level *j*
L	Between-laboratory (interlaboratory)
P	Probability
r	Repeatability
R	Reproducibility
REML	Estimate arising from a restricted maximum likelihood calculation
v	Terms used in calculation of a relationship between mean and combined variance (see 8.5.1.3, relationship III)
W	Within-laboratory (intralaboratory)
1, 2, 3, ...	For test results, numbering in the order of obtaining them; for other cases (laboratories), as arbitrary identifiers
(1), (2), (3), ...	For test results, (1), (2) ... denote the 1st, 2nd ... etc. order statistic, that is, the 1st, 2nd... etc. value numbered in the order of increasing magnitude

5 Estimates of the parameters in the basic model

5.1 The procedures given in this document are based on the statistical model given in Clause 5 of ISO 5725-1:1994 and elaborated upon in ISO 5725-1:1994, 1.2. In particular, these procedures are based on Formulae (2) to (6) of ISO 5725-1:1994, Clause 5.

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.

NOTE The laboratory component of bias, B, represents the deviation of a laboratory mean from the general average m.

5.2 ISO 5725-1:1994, Formulae (2) to (6), 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.

5.3 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 is done in each of ISO 5725-1:1994, Formulae (2) to (6), 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:

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

6 Requirements for a precision experiment

6.1 Layout of the experiment

6.1.1 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.

6.1.2 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 knows 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 can 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 does not know which are the replicates for a given level. However, such a procedure can cause problems in ensuring that repeatability conditions apply between replicates. This is only possible if the measurements are of such a nature that all the qn measurements can 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 of time; 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 can 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.

6.1.3 In 6.1.2 and elsewhere in this document, reference is made to the operator. For some measurements, there can 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".

6.1.4 In commercial practice, the test results can 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 depends on the level *m*, different degrees of rounding can be necessary for different levels.

6.2 Recruitment of the laboratories

6.2.1 The general principles regarding recruitment of the laboratories to participate in an interlaboratory experiment are given in ISO 5725-1. Guidance on the number of laboratories is given in Annex A. 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.

6.2.2 For the purposes of this document, 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) can thus produce several "laboratories" if it can provide several operators each with independent sets of equipment and situations in which to perform the work.

6.3 Preparation of the materials

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

6.3.2 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 can necessitate using extra material. The amount of material prepared shall be sufficient to cover the experiment and allow an adequate stock in reserve.

6.3.3 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.

6.3.4 When a material is 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.

NOTE ISO Guide 35 gives information on evaluating homogeneity and stability for reference materials.

6.3.5 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 can be necessary 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 can result from shaking, for example during transport. When reaction with the atmosphere can 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 can be necessary to send them in a deep-frozen state to the participating laboratories with detailed instructions for the procedure for thawing.

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

7 Personnel involved in a precision experiment

NOTE 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.

7.1 Panel

7.1.1 The precision experiment should be overseen by a panel which should consist of experts familiar with the measurement method and its application.

7.1.2 The tasks of the panel are:

a) to plan and coordinate the precision experiment;

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

c) to appoint someone for the statistical functions (see 7.2);

d) to appoint someone for the executive functions (see 7.3);

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 can 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.

7.2 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 8.7.

7.3 Executive functions

7.3.1 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.

7.3.2 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 6.1.2 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 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 decimal places (or significant digits, as required). Such forms can 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.

NOTE The forms referred to in 7.3.2 d) can be electronic; for example a spreadsheet format suitably protected against unintended modification.

7.4 Supervisors

7.4.1 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.

7.4.2 The tasks of the supervisor are:

a) to ensure that the operators selected are those who 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 (or significant digits), including any anomalies and difficulties experienced, and comments made by the operators.

7.4.3 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 output 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 can have occurred during the measurements, including any change of operator, 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.

NOTE The output and forms referred to in 7.4.3 a) and b) can be electronic; for example a spreadsheet format suitably protected against unintended modification.

7.5 Operators

7.5.1 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.

7.5.2 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 is less temptation for them to discard or rework results that they feel are inconsistent.

7.5.3 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.

7.5.4 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 do 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 occasion(s) when they are unable to follow their instructions as this can also indicate a deficiency in the standard.

8 Statistical analysis of a precision experiment

8.1 Preliminary considerations

8.1.1 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 can exist.

8.1.2 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

— the mean, m.

NOTE The analysis includes a systematic application of statistical tests for outliers, a great variety of which are available from the literature and which can be used for the purposes of this document. For practical reasons, only a limited number of these tests, as explained in 8.3, have been incorporated in this document.

8.2 Tabulation of the results and notation used

8.2.1 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.

8.2.2 Redundant data

Sometimes a laboratory can 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 may be made from those available test results to choose the planned number of test results for analysis.

NOTE The calculations of precision terms that are permitted in this document can accommodate differing numbers of test results in each cell (see 8.2.7). However, substantial variation in the number of reported observations can adversely affect the interpretation of outlier tests and Mandel's statistics and can reduce the reliability of the grand mean.

8.2.3 Missing data

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

8.2.4 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.

8.2.5 Outlying laboratories

When several unexplained abnormal test results occur at different levels within the same laboratory, then that laboratory can 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 can then be reasonable to discard some or all of the data from such an outlying laboratory.

This document does not provide a statistical test by which suspected laboratories can 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.

8.2.6 Erroneous data

Obviously erroneous data should be investigated and corrected or discarded.

NOTE Erroneous data can be obvious when, for example, results are reported in incorrect units, the values reported cannot reasonably relate to the test material under consideration, or other reported information indicates substantial unintended deviation from the standard method under study.

8.2.7 Balanced uniform-level test results

The ideal case is p laboratories called i (i = 1, 2..., p), each testing q levels called j (j = 1, 2, ..., q) with n replicates at each level (each ij combination), giving a total of pqn test results. Because of missing (8.2.3)

or deviating (8.2.4) test results, or outlying laboratories (8.2.5) or erroneous data (8.2.6), this ideal situation is not always attained. Under these conditions the notations given in 8.2.9 to 8.2.11 and the procedures of 8.4 allow for differing numbers of test results.

8.2.8 Collation of data and intermediate values

Specimens of recommended forms for the statistical analysis are given in Figure 2. Form A includes collated individual results, Form B contains calculated cell means, and Form C contains calculated standard deviations. For convenience, they are referred to simply as forms A, B and C (of Figure 2).

NOTE The use of forms A, B and C for statistical analysis is not a requirement of this document. Forms A, B, and C are, however, convenient for collation of intermediate values in manual calculation and can easily be adapted for use in spreadsheet software. Alternative layouts are often more appropriate for use with statistical software.

8.2.9 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 ($k = 1, 2, ..., n_{ij}$);

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).

8.2.10 Cell means (Form B of Figure 2)

These are derived from the results in Form A using Formula (2) as follows.

$$\bar{y}_{ij} = \frac{1}{n_{ij}} \sum_{k=1}^{n_{ij}} y_{ijk} \tag{2}$$

The cell means should be recorded to at least one more significant digit than the reported test results (Form A).

8.2.11 Measures of cell spread (Form C of Figure 2)

These are derived from individual results (Form A, see 8.2.9) and means (Form B, see 8.2.10), as follows. For the general case, use the intracell standard deviation as in Formula (3),

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

or, equivalently

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

In using these formulae, 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 significant figures as required for the original reported data [see 7.3.2 b), 7.4.3 f)].

When using Form C, the standard deviation should be recorded to one more significant digit than the results in Form A. For values of n_{ij} less than 2, a dash should be inserted in Form C.

NOTE 1 If a cell ij contains two test results, the intracell standard deviation is

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

Therefore, for simplicity, absolute differences can be recorded in Form C and used for Cochran's test instead of standard deviations if all cells contain two test results.

NOTE 2 Formula (4) is prone to rounding error in computer calculation; for computer calculations, Formula (3) is more accurate.

8.2.12 Corrected or rejected data

As some of the data can be corrected or rejected on the basis of the tests mentioned in 8.3.3 and 8.3.4, the values of y_{ijk}, n_{ij} and p_j used for the final determinations of precision and mean can 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.

8.3 Scrutiny of results for consistency and outliers

8.3.1 Approaches for scrutiny of data

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 can 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.

NOTE Reference [3] provides further discussion of the treatment of inter-laboratory data.

8.3.2 Graphical consistency technique

8.3.2.1 Two measures called Mandel's h and k statistics are used. As well as describing the variability of the measurement method, these help in laboratory evaluation.

8.3.2.2 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):

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

in which, for $\bar{y}_{ij}$ see 8.2.10, and for $\bar{\bar{y}}_j$ see 8.4.4.

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 C.7).

Form A – Recommended form for the collation of the original data [a]									
Laboratory	Level								
	1	2	...	...	j	...	...	$q-1$	q
1									
2									
⋮									
i	y_{i11} ⋮ y_{i1k} ⋮ y_{i1n}				y_{ij1} ⋮ y_{ijk} ⋮ y_{ijn}				y_{iq1} ⋮ y_{iqk} ⋮ y_{iqn}
⋮									
p									

Form B – Recommended form for the collation of the means [b]									
Laboratory	Level								
	1	2	...	...	j	...	...	$q-1$	q
1	$\overline{y}_{11}$								$\overline{y}_{1q}$
2									
⋮									
i					$\overline{y}_{ij}$				
⋮									
p	$\overline{y}_{p1}$								$\overline{y}_{pq}$

Form C – Recommended form for the collation of the measures of spread within cells [c]									
Laboratory	Level								
	1	2	...	...	j	...	...	$q-1$	q
1	s_{11}								s_{1q}
2									
⋮									
i					s_{ij}				
⋮									
p	s_{p1}								s_{pq}

[a] Form A contains collated individual results. Cells each contain all n results for the laboratory.

[b] Form B contains cell means as calculated in 8.2.10.

[c] Form C contains measures of spread (standard deviations) as calculated in 8.2.11.

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

8.3.2.3 Calculate the within-laboratory consistency statistic, *k*, by first calculating the pooled within-cell standard deviation (Formula (7))

$$\sqrt{\frac{\sum s_{ij}^2}{p_j}} \tag{7}$$

for each level, and then calculate

$$k_{ij} = \frac{s_{ij}\sqrt{p_j}}{\sqrt{\sum s_{ij}^2}} \tag{8}$$

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 C.8).

8.3.2.4 Examination of the *h* and *k* plots can indicate that specific laboratories exhibit patterns of results that are markedly different from the others in the study. For example, laboratories can show 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 can:

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.

8.3.2.5 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 can 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 can 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 Table 7 and Table 8. These indicator lines serve as guides when examining patterns in the data.

8.3.2.6 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 can 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 (Table 7 and Table 8). These indicator lines serve as guides when examining patterns in the data.

8.3.2.7 When an *h* or *k* plot grouped by laboratory suggests that one laboratory has several *h* or *k* values near the indicator value line, the corresponding plot grouped by level should be studied. Often a value that appears large in a plot grouped by laboratory turns 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.

8.3.2.8 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 requires special treatment as the general underlying principle behind the methods described here assumes a single unimodal population.

8.3.3 Numerical outlier technique

8.3.3.1 The following practice is recommended for dealing with outliers.

a) The tests recommended in 8.3.4 and 8.3.5 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 mistake 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.

8.3.3.2 The tests given in 8.3.4 and 8.3.5 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.

8.3.4 Cochran's test

8.3.4.1 This document 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 to test the validity of this assumption. Several tests can be used for this purpose, but Cochran's test has been chosen.

8.3.4.2 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 given by Formula (9):

$$C = \frac{s_{\max}^2}{\sum_{i=1}^{p} s_i^2} \tag{9}$$

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 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.

Critical values for Cochran's test are given in Table 5.

Cochran's test shall be applied to Form C of Figure 2 at each level separately.

8.3.4.3 Cochran's criterion applies strictly only when all the standard deviations are derived from the same number (n) of test results obtained under the same conditions of measurement (repeatability conditions, in this document). In actual cases, this number can vary owing to missing or discarded data. This document assumes, however, that in a properly organized experiment such variations in the number of test results per cell are 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.

8.3.4.4 Cochran's criterion tests only the highest value in a set of standard deviations and is therefore a one-sided outlier test. Variance heterogeneity can also manifest itself in some of the standard deviations being comparatively too low. However, small values of standard deviation can 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.

8.3.4.5 A critical examination of Form C of Figure 2 can sometimes reveal that the standard deviations for a particular laboratory are at all or at most levels lower than those for other laboratories. This can indicate that the laboratory works with a lower repeatability standard deviation than the other laboratories, which in turn can 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.)

8.3.4.6 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 may be repeated but it can lead to excessive rejections when, as is sometimes the case, the underlying assumption of normality is not sufficiently upheld. 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 variance 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 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 can 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.

8.3.5 Grubbs' tests

8.3.5.1 One outlying observation

Given a set of data, $x_{(1)}$, $x_{(2)}$, ..., $x_{(p)}$, formed from x_1, x_2, ..., x_p arranged in ascending order, to determine whether the largest observation is an outlier using Grubbs' test, compute the test statistic, G_p, using Formula (10):

$$G_p = \frac{x_{(p)} - \bar{x}}{s} \tag{10}$$

where

$$\bar{x} = \frac{1}{p}\sum_{i=1}^{p} x_{(i)} \tag{11}$$

and

$$s = \sqrt{\frac{1}{p-1}\sum_{i=1}^{p}\left(x_{(i)} - \bar{x}\right)^2} \tag{12}$$

To test the significance of the smallest observation, compute the test statistic, G_1, using Formula (13):

$$G_1 = \frac{\left(\bar{x} - x_{(1)}\right)}{s} \tag{13}$$

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.

8.3.5.2 Two outlying observations

To test whether the two largest observations can be outliers, compute the test statistic, G, using Formula (14):

$$G = \frac{s^2_{p-1,p}}{s^2_0} \tag{14}$$

where

$$s^2_0 = \sum_{i=1}^{p} \left(x_{(i)} - \bar{x} \right)^2 \tag{15}$$

and

$$s^2_{p-1,p} = \sum_{i=1}^{p-2} \left(x_{(i)} - \bar{x}_{p-1,p} \right)^2 \tag{16}$$

and

$$x_{p-1,p} = \frac{1}{p-2} \sum_{i=1}^{p-2} x_{(i)} \tag{17}$$

Alternatively, to test the two smallest observations, compute the test statistic, G, using Formula (18):

$$G = \frac{s^2_{1,2}}{s^2_0} \tag{18}$$

where

$$s^2_{1,2} = \sum_{i=1}^{p-2} \left(x_{(i)} - \bar{x}_{1,2} \right)^2 \tag{19}$$

and

$$x_{1,2} = \frac{1}{p-2} \sum_{i=3}^{p} x_{(i)} \tag{20}$$

8.3.5.3 Application of the 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

$$x_i = \bar{y}_{ij} \tag{21}$$

and

$$p = p_j \tag{22}$$

where j is fixed.

Taking the data at one level, apply Grubbs' test for one outlying observation to cell means as described in 8.3.5.1. 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 Grubbs' test for two outlying observations described in 8.3.5.2.

If Grubbs' test does not show a cell mean to be an outlier, then apply Grubbs' test for two outliers described in 8.3.5.2.

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

NOTE According to 8.3.3.1, an item is called a statistical outlier if the test statistic is greater than its 1 % critical value. When Grubbs' test is first applied to a group of cell means, a critical value from Table 6 is used to test the highest cell mean using a test at the 0,5 % level, and to test the lowest cell mean at the 0,5 % level. This amounts to a test of the most extreme cell mean at the 1 % level in accordance with 8.3.3.1. If the most extreme cell mean is found to be a statistical outlier, Grubbs' test is then applied to the other extreme cell mean. It can be argued that a one-sided test should now be used. However, the procedure recommended in this document is to use only the critical values tabulated in Table 6 (the critical values for two-sided tests at the 1 % significance level) in order that all cell means be treated consistently. A similar argument can justify the use of the two-sided 5 % critical values in Table 6 for all tests for statistical stragglers.

8.3.6 Repeated testing for outlying means or outlying data points

8.3.6.1 It is often found that a single application of (for example) Grubbs' test, followed by removal of an outlier, leaves a smaller set of data which itself appears, on inspection, to contain further extreme values relative to the remainder of the set. In these circumstances the application of further tests for outlying data points or laboratory means is permitted.

8.3.6.2 Repeated outlier testing and rejection should be discontinued for a given level where it leads to rejection of an excessive proportion of data from that level. The panel shall determine whether the proportion of rejected data is excessive.

NOTE 1 It can be helpful to set an upper limit to the proportion of rejected data in advance of the statistical analysis.

NOTE 2 IUPAC[15] recommends that the proportion of data rejected from a given level of a study should not exceed the fraction 2/9, the fraction being considered after removal of any results deemed invalid on technical grounds prior to outlier inspection.

8.3.7 Alternative outlier inspection and test methods

Alternative inspection methods and statistical tests to those listed in 8.3.3 to 8.3.5 may be used, provided that:

— the same set of outlier inspection and statistical outlier tests is applied to all levels within a given study;

— the methods chosen are capable of identifying outlying individual data points within a laboratory, outlying laboratory means, and outlying laboratory dispersion;

— any bias in variance estimates for normally distributed data that arises from outlier rejection is not larger than that resulting from the application of subclauses 8.3.3 to 8.3.5.

NOTE All outlier rejection procedures lead to some downward bias in precision estimates unless an explicit correction for the resulting bias is made. Robust statistical procedures, which suffer much less bias in precision estimates for normally distributed data, are described in ISO 5725-5.

8.4 Calculation of the general mean and variances

8.4.1 Method of analysis

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

8.4.2 Basic data

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

— Form A containing the original test results;

— Form B containing the cell means;

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

8.4.3 Non-empty cells

As a consequence of the rule stated in 8.3.3.1 d), the number of non-empty cells to be used in the computation, for a specific level, are usually the same in Form B and Form C. An exception might occur if, owing to missing data, a cell in Form A contains only a single test result, which entails an empty cell in Form C but not in Form B. In that case it is permitted

a) to discard the solitary test result, which leads to empty cells in both Form B and Form C, or

b) if this is considered an undesirable loss of information, to insert a dash (missing value) in Form C.

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

8.4.4 Calculation of the general mean, $\hat{m}$

For level j, the general mean is

$$\hat{m}_j = \bar{\bar{y}}_j = \frac{\sum_{i=1}^{p} n_{ij}\bar{y}_{ij}}{\sum_{i=1}^{p} n_{ij}} \tag{23}$$

or, equivalently

$$\hat{m}_j = \bar{\bar{y}}_j = \frac{\sum_{i=1}^{p}\sum_{k=1}^{n_{ij}} y_{ijk}}{\sum_{i=1}^{p} n_{ij}} \tag{24}$$

8.4.5 Calculation of variances

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

8.4.5.1 The repeatability variance is given by Formula (25):

$$s_{rj}^2 = \frac{\sum_{i=1}^{p}\left(n_{ij}-1\right)s_{ij}^2}{\sum_{i=1}^{p}\left(n_{ij}-1\right)} \tag{25}$$

8.4.5.2 The between-laboratory variance is given by Formula (26):

$$s_{Lj}^2 = \frac{s_{dj}^2 - s_{rj}^2}{\bar{\bar{n}}_j} \tag{26}$$

where

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

$$= \frac{1}{p-1}\left[\sum_{i=1}^{p} n_{ij}\bar{y}_{ij}^2 - \bar{\bar{y}}^2\sum_{i=1}^{p} n_{ij}\right] \tag{27}$$

and

$$\bar{\bar{n}}_j = \frac{1}{p-1}\left[\sum_{i=1}^{p} n_{ij} - \frac{\sum_{i=1}^{p} n_{ij}^2}{\sum_{i=1}^{p} n_{ij}}\right] \tag{28}$$

These calculations are illustrated in the examples in C.1 and C.3.

NOTE The second form given for Formula (27), involving the difference between two summed terms, is mathematically identical to the first, but can result in severe round-off errors when used in computer calculations. The first form is much less subject to round-off errors.

8.4.5.3 For the particular case where all $n_{ij} = n = 2$, the simpler calculations of Formulae (29) and (30) may be used:

$$s_{rj}^2 = \frac{1}{2p}\sum_{i=1}^{p}\left(y_{ij1} - y_{ij2}\right)^2 \tag{29}$$

and

$$s_{Lj}^2 = \frac{1}{p-1}\sum_{i=1}^{p}\left(\bar{y}_{ij} - \bar{\bar{y}}_j\right)^2 - \frac{s_{rj}^2}{2} \tag{30}$$

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

8.4.5.4 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.

8.4.5.5 The reproducibility variance is given by Formula (31):

$$s_{Rj}^2 = s_{rj}^2 + s_{Lj}^2 \tag{31}$$

8.4.6 Alternative calculation methods for variances

8.4.6.1 For cases where all laboratories have the same number of replicates (after scrutiny for outliers and any subsequent removal of data points), the variance calculations above may be replaced by application of analysis of variance as described in Annex B.

8.4.6.2 The calculation process in 8.4.4 to 8.4.5 may be replaced with restricted likelihood estimation (REML)[13],[14] provided by appropriate computer software. Software that reproduces REML estimates of variance given in the examples in Annex B to three or more significant digits may be used.

NOTE 1 A procedure for restricted maximum likelihood estimation of variance components is provided in Annex B.

NOTE 2 There can be small differences between the results of REML estimation and the calculations provided elsewhere in this part. These differences are usually inconsequential.

8.4.6.3 Where an alternative method of calculation for variances is chosen, the same calculation method shall be applied to all levels of the precision experiment.

8.4.7 Dependence of the variances upon *m*

After calculating the precision for each level *m*, it should be investigated whether the precision depends upon *m* and, if so, the functional relationship should be determined. This is described in 8.5.

8.5 Establishing a functional relationship between precision values, *s*, and the mean level, *m*

8.5.1 Choice of functional relationship

8.5.1.1 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 is 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 considered, using knowledge of the materials and the measurement method, before the following procedure is applied.

8.5.1.2 Where no clear functional relationship can be established, separate values of precision may be reported for each material investigated. Where the different values of precision are similar across levels, an averaged precision value may be reported.

8.5.1.3 The reasoning and computation procedures presented in 8.5.2 to 8.5.4 apply both to repeatability and reproducibility standard deviations, but are presented here for repeatability only in the interests of brevity. Only four types of relationship are considered:

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

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

III. $s_r^2 = a_v^2 + (b_v m)^2$ (a combination of constant and relative contributions to a combined standard deviation)

IV. $\lg s_r = c + d \lg m$ (or $s_r = Cm^d$); $d \leq 1$ (an exponential relationship)

It is to be expected that in the majority of cases at least one of these formulae 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 formulae can 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.*

8.5.1.4 In general $d > 0$ so that relationships I and IV lead to $s = 0$ for $m = 0$, which can 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.

8.5.1.5 For $a = 0$ (or $a_v = 0$) and $d = 1$, all four relationships are algebraically identical (all can be reduced to relationship I), so when a lies near zero and/or d lies near unity, three or all four of these relationships 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.

8.5.1.6 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 can provide a satisfactory solution; but if for some reason a numerical method of fitting is preferred, the procedure of 8.5.2 is recommended for relationships I and II, and that of 8.5.4 for relationship IV.

8.5.2 Fitting relationships I and II

8.5.2.1 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.

8.5.2.2 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 $1/\left(\hat{s}_j\right)^2$, 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 can be complicated. The following iterative procedure, which has proved to be satisfactory in practice, is recommended.

8.5.2.3 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:

$$T_1 = \sum_j W_j \tag{32}$$

$$T_2 = \sum_j W_j \hat{m}_j \tag{33}$$

$$T_3 = \sum_j W_j \hat{m}_j^2 \tag{34}$$

$$T_4 = \sum_j W_j s_j \tag{35}$$

$$T_5 = \sum_j W_j \hat{m}_j s_j \tag{36}$$

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

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

$$a = \frac{T_3 T_4 - T_2 T_5}{T_1 T_3 - T_2^2} \tag{37}$$

and

$$b = \frac{T_1 T_5 - T_2 T_4}{T_1 T_3 - T_2^2} \tag{38}$$

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

$$b = \frac{\sum_j \left(s_j / \hat{m}_j \right)}{q} \tag{39}$$

and no iteration is necessary.

8.5.2.5 For relationship II, the initial values $\hat{s}_{0j}$ are the original values of s as obtained by the procedures given in 8.4. These are used to calculate

$$W_{0j} = 1 \Big/ \left(s_{0j} \right)^2 \quad (j = 1, 2, \ldots, q) \tag{40}$$

and to calculate a_1 and b_1 as in 8.5.6.2. This leads to

$$\hat{s}_{1j} = a_1 + b_1 \hat{m}_j \tag{41}$$

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

$$\hat{s}_{2j} = a_2 + b_2 \hat{m}_j \tag{42}$$

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

EXAMPLE 1 An example of fitting relationship I to repeatability data is given in Table 1. The data are taken from the case study of C.3 and have been used here only to illustrate the numerical procedure. The relationships are further discussed in C.3.

EXAMPLE 2 An example of fitting relationship II to repeatability data is given in Table 2. The data m_j, s_j are the same as for Table 1.

8.5.3 Fitting relationship III

8.5.3.1 For relationship III, the simplest approach is to fit the observed variances s_j^2 as a function of the squared means, giving a linear relationship in which the intercept a_v^2 is the variance at zero mean and the slope b_v^2 can be thought of as the square of the relative standard deviation as m becomes large. As for relationship II, however, while a simple linear regression can give a usable estimate, a good estimate requires weighted regression and, as for relationship II, the optimal weights depend on fitted values.

8.5.3.2 An approximate procedure for fitting relationship III, similar to that in 8.5.2, is as follows.

i) With weighting factor W_{vj} equal to $1/\hat{s}_{Nj}^4$, where N = 0, 1, 2, ... for successive iterations, then the calculated formulae are:

$$T_{v1} = \sum_j W_{vj} \tag{43}$$

$$T_{v2} = \sum_j W_{vj}\hat{m}_j^2 \tag{44}$$

$$T_{v3} = \sum_j W_{vj}\hat{m}_j^4 \tag{45}$$

$$T_{v4} = \sum_j W_{vj}s_j^2 \tag{46}$$

$$T_{v5} = \sum_j W_{vj}\hat{m}_j^2 s_j^2 \tag{47}$$

ii) Then, for relationship III, $s_r^2 = a_v^2 + \left(b_v m\right)^2$:

$$a_v^2 = \frac{T_{v3}T_{v4} - T_{v2}T_{v5}}{T_{v1}T_{v3} - T_{v2}^2} \tag{48}$$

$$b_v^2 = \frac{T_{v1}T_{v5} - T_{v2}T_{v4}}{T_{v1}T_{v3} - T_{v2}^2} \tag{49}$$

iii) Initial values a_{1v}^2 and b_{1v}^2 are then used to predict fitted values of s_j according to

$$\hat{s}_{1j}^2 = a_{1v}^2 + b_{1v}^2\hat{m}_j^2 \tag{50}$$

iv) Computations i) to iii) are then repeated once with $W_{1vj} = 1/\hat{s}_{1j}^4$ to produce the relationship

$$\hat{s}_{2j}^2 = a_{2v}^2 + b_{2v}^2\hat{m}_j^2 \tag{51}$$

The resulting formulae for $\hat{s}_{2j}$ should be considered as the final result.

NOTE The weights W_{vj} rely on the understanding that, for normally distributed data, the distribution of sample variance is a scaled chi-squared distribution whose variance is proportional to σ^4.

EXAMPLE An example of fitting relationship III to repeatability data is given in Table 3. The data m_j, s_j are the same as for Table 1, and are taken from Example C.3 of Annex C.

8.5.4 Fitting relationship IV

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

8.5.4.2 For relationship IV, the computational formulae are:

$$T_1 = \sum_j \lg \hat{m}_j \tag{52}$$

$$T_2 = \sum_j \left(\lg \hat{m}_j\right)^2 \tag{53}$$

$$T_3 = \sum_j \lg s_j \tag{54}$$

$$T_4 = \sum_j \left(\lg \hat{m}_j\right)\left(\lg s_j\right) \tag{55}$$

and thence

$$c = \frac{T_2 T_3 - T_1 T_4}{q T_2 - T_1^2} \tag{56}$$

and

$$d = \frac{q T_4 - T_1 T_3}{q T_2 - T_1^2} \tag{57}$$

EXAMPLE An example of fitting relationship IV to repeatability data is given in Table 4. The data m_j, s_j are the same as for Table 1, and are taken from Example C.3 of Annex C.

Table 1 — Relationship I, $s = bm$

$\hat{m}_j$	3,94	8,28	14,18	15,59	20,41
s_j	0,092	0,179	0,127	0,337	0,393
$s_j / \hat{m}_j$	0,023 4	0,021 6	0,008 9	0,021 6	0,019 3
$b = \frac{\sum_k \left(s_j / \hat{m}_j\right)}{q}$	$\frac{0,094\ 8}{5} = 0,019$				
$s = bm$	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
$s_1 = 0,058 + 0,009\ 0\ m$					
$\hat{s}_{1j}$	0,093	0,132	0,185	0,197	0,240
W_{1j}	116	57	29	26	17
$s_2 = 0,030 + 0,015\ 4\ m$					
$\hat{s}_{2j}$	0,092	0,159	0,251	0,273	0,348
W_{2j}	118	40	16	13	8
$s_3 = 0,032 + 0,015\ 4\ m$[a]					
$\hat{s}_{3j}$	0,093	0,160	0,251	0,273	0,348

NOTE The values of the weighting factors are not critical; two significant digits suffice.

[a] The difference from s_2 is negligible.

Table 3 — Relationship III, $s^2 = a_v^2 + (b_v m)^2$

W_{0vj}	13 959	974	3 844	77,5	41,9
$s_1^2 = 0,088^2 + (0,008\ 5\ m)^2$					
$\hat{s}_{1j}$	0,094	0,113	0,150	0,160	0,195
W_{1vj}	12 603	6 126	1 985	1 537	687
$s_2^2 = 0,061^2 + (0,017\ 8\ m)^2$					
$\hat{s}_{2j}$	0,093	0,159	0,260	0,284	0,368
W_{2vj}	13 356	1 546	220	154	54
$s_3^2 = 0,063^2 + (0,017\ 8\ m)^2$ [a]					
$\hat{s}_{3j}$	0,094	0,160	0,260	0,284	0,368

NOTE The values above were computed using full precision arithmetic and have been rounded for presentation.

[a] The difference from s_2 is negligible.

Table 4 — Relationship IV, $\lg s = c + d \lg m$

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

8.6 Statistical analysis as a step-by-step procedure

NOTE Figure 3 indicates in a stepwise fashion the procedure given in 8.6.

8.6.1 Collect all available test results in one form, Form A of Figure 2 (see 8.2). 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.

8.6.2 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 8.7.1).

8.6.3 From Form A, corrected according to 8.6.2 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 8.4.3 should be adopted.

8.6.4 Prepare the Mandel *h* and *k* plots as described in 8.3.1 and examine them for consistency of the data. These plots can 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 8.6.5 to 8.6.9.

8.6.5 Inspect forms B and C (see Figure 2) level by level for possible stragglers and/or statistical outliers [see 8.3.3.1 a)]. Apply the statistical tests given in 8.3 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 8.6.6 to 8.6.10 and proceed directly with 8.6.11.

8.6.6 Investigate whether there is or can 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 8.6.7 to 8.6.10 and proceed directly with 8.6.11.

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

8.6.7 If the distribution of the unexplained stragglers or statistical outliers in Form B or C does not suggest any outlying laboratories (see 8.2.5), ignore step 8.6.8 and proceed directly with 8.6.9.

8.6.8 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 8.7.1).

8.6.9 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.

8.6.10 If in the previous steps any entry in Form B has been rejected, then the corresponding entry in Form C shall be rejected also, and vice versa.

8.6.11 From the entries that have been retained as correct in forms B and C, compute, by the procedures given in 8.4, for each level separately, the mean level $\hat{m}_j$ and the repeatability and reproducibility standard deviations.

8.6.12 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 8.5.1) and not as functions of the level, ignore subclause 8.6.13 and proceed directly with 8.6.14.

8.6.13 Apply steps 1) to 6) below to s_r and to s_R separately.

NOTE For brevity, steps 1) to 5) are written out only in terms of s_r.

1) 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 2) and proceed with step 3). If s is judged to be independent of m, proceed with step 2). 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.

2) Use Formula (58)

$$\frac{1}{q}\sum_{j} s_j = s_r \qquad (58)$$

to calculate the final value of the repeatability standard deviation. Ignore steps 3) to 6) and proceed directly to 8.6.14.

3) Judge from the plot of s against m and of s^2 and m^2 whether the relationships between s and m, or between s^2 and m^2, can be represented by a straight line and, if so, whether relationship I ($s = bm$), relationship II ($s = a + bm$) or relationship III ($s_j^2 = a_v^2 + (b_v m)^2$) is appropriate (see 8.5.1). For relationships I and II, determine the parameter b, or the two parameters a and b, by the procedure of 8.5.2; for relationship III, use the procedure of 8.5.3. If one of these relationships is considered satisfactory, ignore step 4) and proceed directly with step 5). If not, proceed with step 4).

4) Plot lg s_j against lg $\hat{m}$ 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 IV (lg $s = c + d$ lg m) using the procedure given in 8.5.4.

5) If a satisfactory relation has been established in step 3) or 4), 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 6) and proceed with 8.6.14.

6) If no satisfactory relation has been established in step 3) or 4), 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. Where the data are too irregular to provide a functional relationship, precision values may be reported separately for each level.

8.6.14 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 8.3.2 can be useful in presenting the consistency or variability of the results.

8.7 Report to the panel and decisions to be taken by the panel

8.7.1 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 8.6.2 and 8.6.8, 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 step 8.6.13, 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.

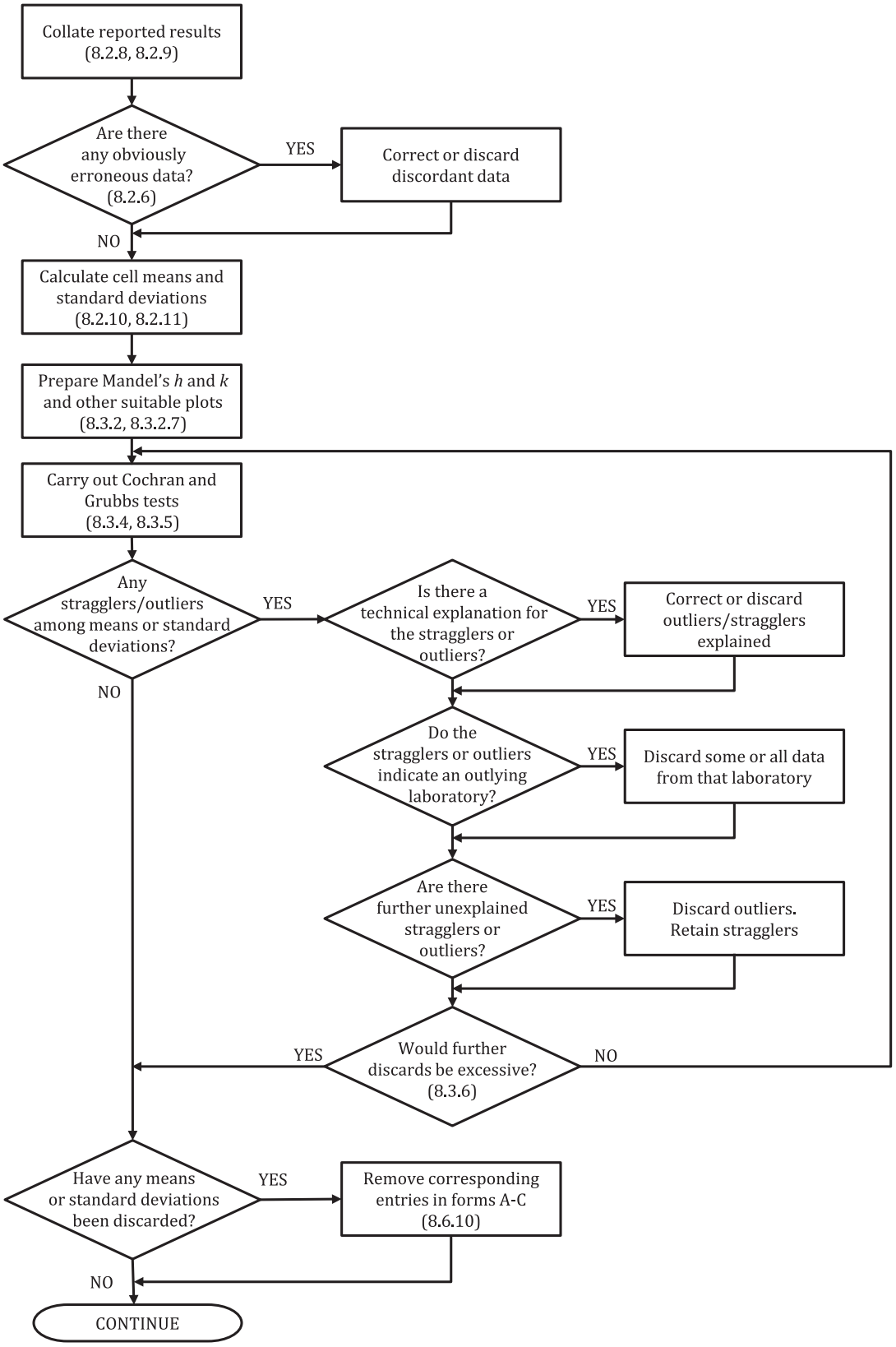


Figure 3 — Flow diagram of the principal steps in the statistical analysis (1 of 2)

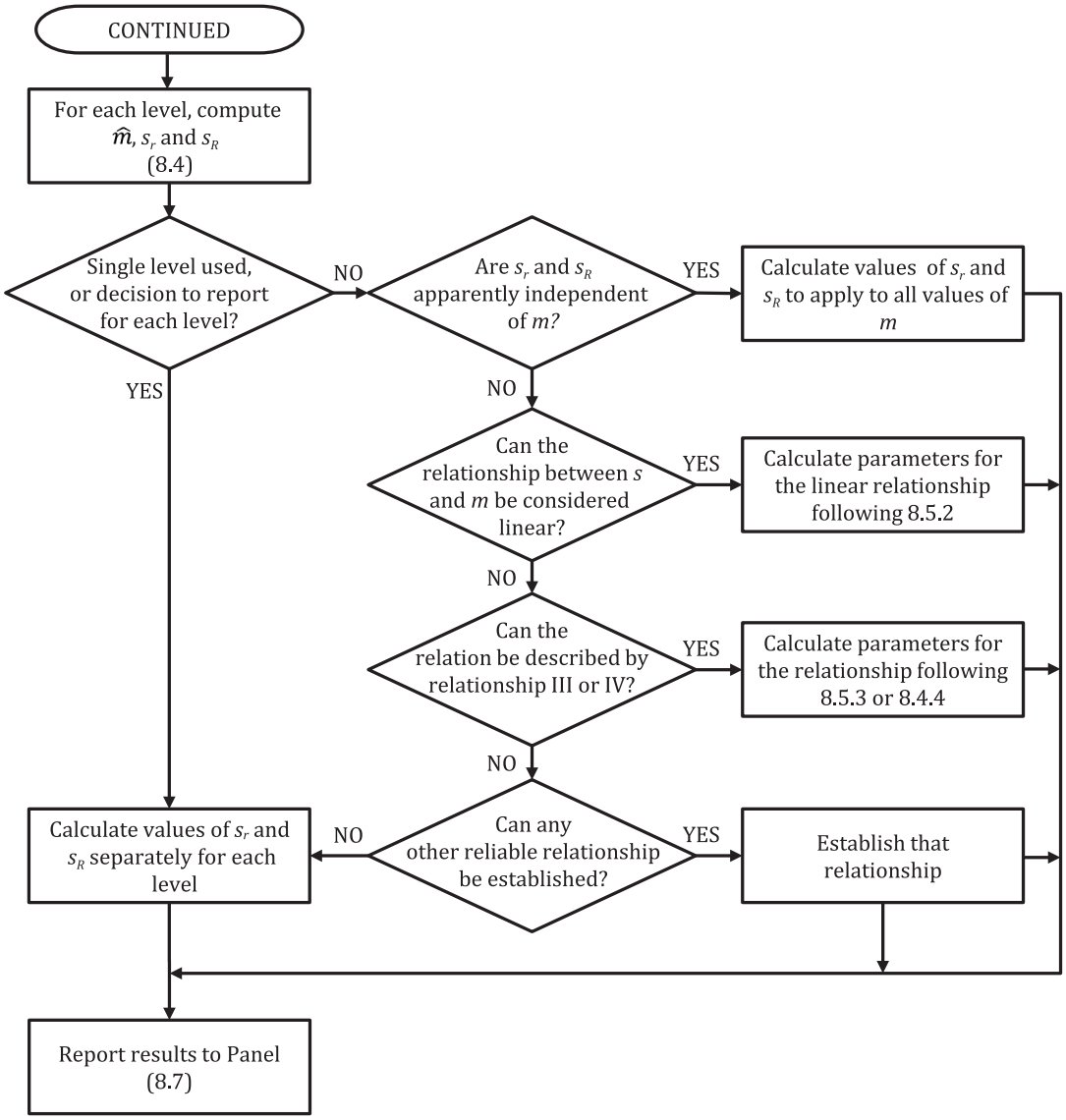


Figure 3 — Flow diagram of the principal steps in the statistical analysis (2 of 2)

8.7.2 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?

8.7.3 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.

9 Statistical tables

9.1 Critical values for Cochran's test (see 8.3.4) are given in Table 5.

9.2 Critical values for Grubbs' test (see 8.3.5) are given in Table 6.

For 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 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.

9.3 Indicators for Mandel's *h* and *k* statistics (see 8.3.2) are given in Table 7 and Table 8.

Table 5 — 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

p = number of laboratories at a given level

n = number of test results per cell (see 8.3.4.3)

NOTE Annex D, based on Reference [16], provides a method of calculating these critical values, in software.

Table 5 *(continued)*

p	$n = 2$		$n = 3$		$n = 4$		$n = 5$		$n = 6$	
	1 %	5 %	1 %	5 %	1 %	5 %	1 %	5 %	1 %	5 %
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 8.3.4.3)

NOTE Annex D, based on Reference [16], provides a method of calculating these critical values, in software.

Table 6 — 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

Reproduced with the permission of the American Statistical Association, from Reference [4].

p = number of laboratories at a given level

NOTE 1 The critical values given in this table are appropriate when two-sided tests are required. They are the critical values required by the procedure for applying Grubbs' outlier tests described in 8.3.5 of this document. They have been derived from the critical values for the corresponding one-sided tests as given in Reference [5].

NOTE 2 Annex D, based on Reference [16], provides a method of calculating these critical values, in software, from other well-known distributions.

Table 6 *(continued)*

p	One largest or one smallest		Two largest or two smallest	
	Upper 1 %	Upper 5 %	Lower 1 %	Lower 5 %
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].

p = number of laboratories at a given level

NOTE 1 The critical values given in this table are appropriate when two-sided tests are required. They are the critical values required by the procedure for applying Grubbs' outlier tests described in 8.3.5 of this document. They have been derived from the critical values for the corresponding one-sided tests as given in Reference [5].

NOTE 2 Annex D, based on Reference [16], provides a method of calculating these critical values, in software, from other well-known distributions.

Table 7 — 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 1 Supplied by Dr. J. Mandel and published with his permission.

NOTE 2 Annex D, based on Reference [16], provides a method of calculating these critical values, in software, from other well-known distributions.

Table 8 — 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	1,36
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 1 Supplied by Dr. J. Mandel and published with his permission.

NOTE 2 Annex D, based on Reference [16], provides a method of calculating these indicators, in software, from other well-known distributions.

Annex A
(informative)

Number of laboratories required for an estimate of precision

A.1 The various quantities represented by the symbol σ in different formulae in the main text of this document are true standard deviations whose values are not known, an object of a precision experiment being to estimate them. When an estimate (s) of a true standard deviation (σ) is to be made, conclusions can be drawn as to the range about σ within which the estimate (s) can be expected to lie. This is a well-understood statistical problem which is solved by the use of the chi-squared distribution and the number of results from which the estimate of s was based. One formula frequently used is:

$$P\left[-A<\frac{s-\sigma}{\sigma}<+A\right]=P \tag{A.1}$$

Often A is quoted in percentage terms, enabling a statement to be made that the estimated standard deviations (s) can be expected to be within A either side of the true standard deviation (σ) with a certain probability P.

A.2 For a single level of the test, the uncertainty in the repeatability standard deviation depends on the number of laboratories (p) and the number of test results within each laboratory (n) For the reproducibility standard deviation, the procedure is more complicated as this is determined from two standard deviations [see Formula (A.1)]. An extra factor γ is needed, representing the ratio of the reproducibility standard deviation to the repeatability standard deviation [see Formula (A.2)] that is:

$$\gamma=\sigma_R/\sigma_\mathrm{r} \tag{A.2}$$

A.3 Assuming a probability level P of 95 %, approximate formulae for the values of A have been prepared and are given below. The formulae are intended for the purposes of planning how many lab oratories to recruit and deciding how many test results are to be required from each laboratory at each level of the test. These formulae do not give confidence limits and so they should not be used during the analysis stage to calculate confidence limits. The formulae are as follows.

For repeatability,

$$A=A_r=1{,}96\sqrt{\frac{1}{2p(n-1)}} \tag{A.3}$$

For reproducibility,

$$A=A_R=1{,}96\sqrt{\frac{p\left[1+n(\gamma^2-1)\right]^2+(n-1)(p-1)}{2\gamma^4n^2(p-1)p}} \tag{A.4}$$

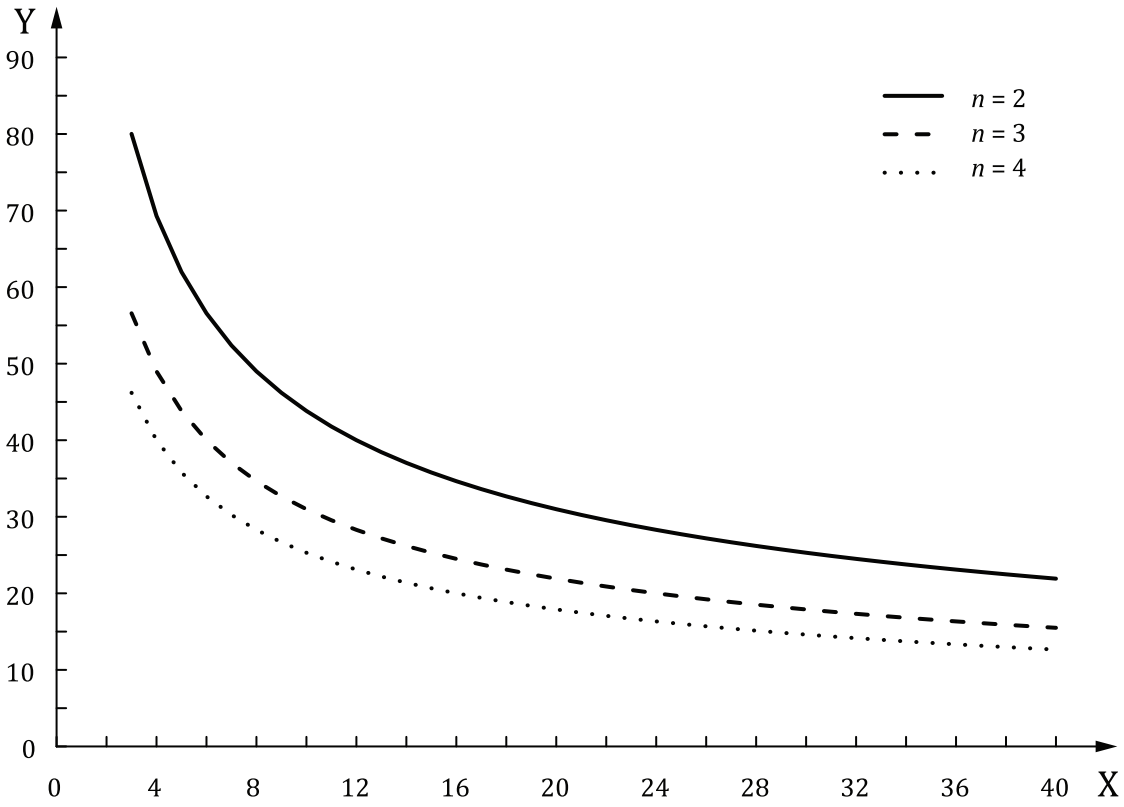
NOTE A sample variance which has v degrees of freedom and expectation σ^2 can be assumed to have, approximately, a normal distribution with variance $2\sigma^4/v$. Formulae (A.3) and (A.4) were derived by making this assumption about the variances involved in the estimation of σ_r and σ_R. The adequacy of the approximation was checked by an exact calculation.

A.4 The value of γ is not known, but often preliminary estimates are available of the within-laboratory standard deviations and the between-laboratory standard deviations obtained during the process of standardizing the measurement method. Values of the uncertainty (as a decimal fraction) for repeatability and reproducibility standard deviations with different numbers of laboratories (p) and

different numbers of results per laboratory (n) are given in Table A.1 and are also plotted in chart form as percentages in Figure A.1 for the repeatability standard deviation and Figure A.2, for the reproducibility standard deviation.

Table A.1 — Values showing the uncertainty of estimates of the repeatability and reproducibility standard deviations

No. of laboratories p	A_r			A_R								
				$\gamma = 1$			$\gamma = 2$			$\gamma = 5$		
	$n = 2$	$n = 3$	$n = 4$	$n = 2$	$n = 3$	$n = 4$	$n = 2$	$n = 3$	$n = 4$	$n = 2$	$n = 3$	$n = 4$
5	0,62	0,44	0,36	0,46	0,37	0,32	0,61	0,58	0,57	0,68	0,67	0,67
10	0,44	0,31	0,25	0,32	0,26	0,22	0,41	0,39	0,38	0,45	0,45	0,45
15	0,36	0,25	0,21	0,26	0,21	0,18	0,33	0,31	0,30	0,36	0,36	0,36
20	0,31	0,22	0,18	0,22	0,18	0,16	0,28	0,27	0,26	0,31	0,31	0,31
25	0,28	0,20	0,16	0,20	0,16	0,14	0,25	0,24	0,23	0,28	0,28	0,27
30	0,25	0,18	0,15	0,18	0,15	0,13	0,23	0,22	0,21	0,25	0,25	0,25
35	0,23	0,17	0,14	0,17	0,14	0,12	0,21	0,20	0,19	0,23	0,23	0,23
40	0,22	0,16	0,13	0,16	0,13	0,11	0,20	0,19	0,18	0,22	0,22	0,22

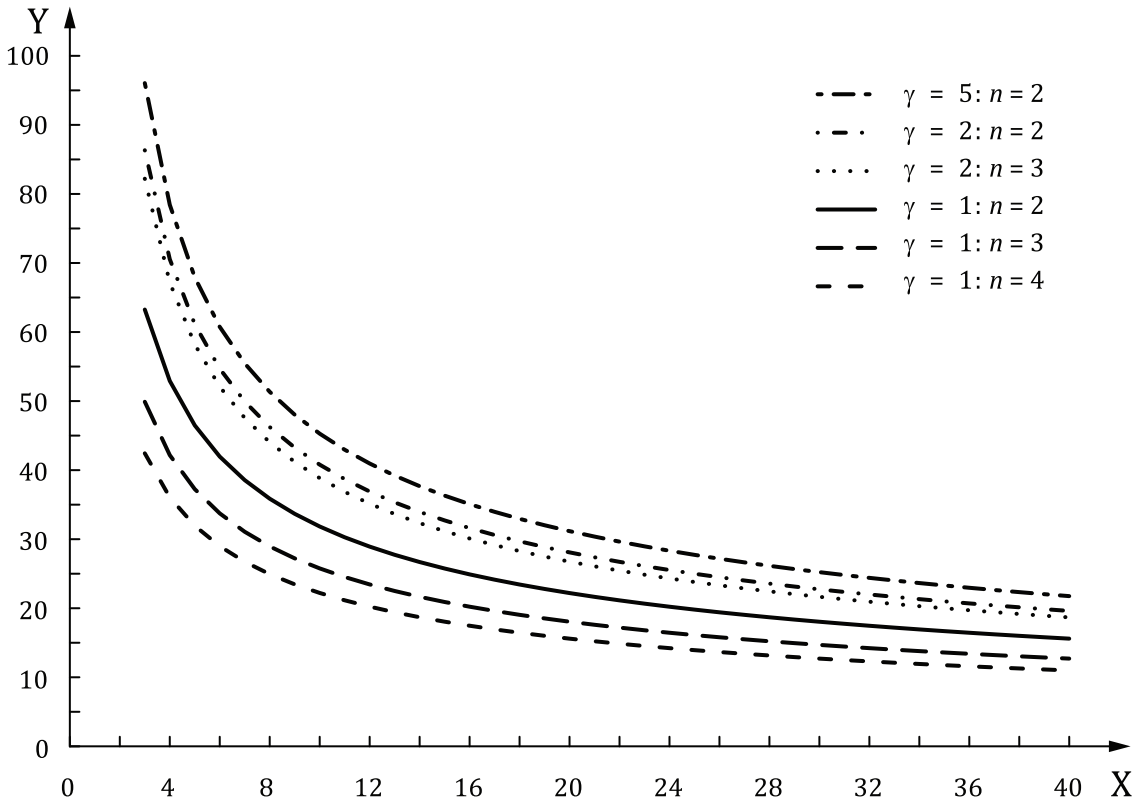


Key

X p

Y A_r (%)

Figure A.1 — The amount by which s_r can be expected to differ from the true value within a probability level of 95 %



Key

X p

Y A_R (%)

Figure A.2 — The amount by which s_R can be expected to differ from the true value within a probability of 95 %

Annex B
(informative)

Alternative calculations of variance components

B.1 Calculation from a one-way analysis of variance table

B.1.1 Analysis of variance (ANOVA) software is widely available for the 'one-way' case required by ISO 5725-2. Application of ANOVA to the data from a single level j of an interlaboratory study conventionally provides a table of the form shown in Table B.1. The relevant variance components s_{rj}^2 and s_{Lj}^2 can be determined from the mean squares in the table.

B.1.2 The repeatability variance is calculated as in Formula (B.1):

$$s_{rj}^2 = M_w \tag{B.1}$$

B.1.3 The between-laboratory variance is, for a balanced experiment in which all cells contain n reported values, calculated as in Formula (B.2):

$$s_{Lj}^2 = \frac{M_b - M_w}{n} \tag{B.2}$$

Table B.1 — Layout of a typical one-way ANOVA table

Source of variation	Sum of squares	Degrees of freedom	Mean square	F
Between-group	S_b	$p - 1$	$M_b = S_b/(p - 1)$	M_b/M_w
Within-group	S_w	$p(n - 1)$	$M_w = S_w/[p(n - 1)]$	
Total	$S_{tot} = S_b + S_w$	$np - 1$		

B.2 Restricted maximum likelihood (REML) calculation

B.2.1 The restricted maximum likelihood (REML) approach is a particular form of maximum likelihood estimation which does not base estimates on a maximum likelihood fit of all the information, but instead uses a likelihood function calculated from the transformed data that do not include the fixed effects parameters. A procedure suitable for this document is described below.

B.2.2 A transformation matrix

$$\mathbf{M} = \mathbf{I} - \mathbf{X}\left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathrm{T}} \tag{B.3}$$

where $\mathbf{X}^{\mathrm{T}}$ is the transpose of $\mathbf{X}$ and is created based on the design matrix $\mathbf{X}$. For the simple one-way layout used at each level j of the experiment in ISO 5725-2, $\mathbf{X}$ is a vector of length $n_j = \sum_{i=1}^{p}\sum_{k=1}^{n_i} n_{ijk}$, all elements being equal to 1. This corresponds to the design matrix for a fixed-effects model for a single

mean value. **M** is then a square n_j by n_j matrix such that all diagonal elements are equal to $1-1/n_j$ and all off-diagonal elements are equal to $-1/n_j$.

B.2.3 The last row of **M** is removed to give a modified matrix $\mathbf{M_m}$ After pre-multiplying by $\mathbf{M_m}$ on the vector of results **Y**, the single fixed effect is effectively removed from the model. Thus, only the parameters of random effects are estimated.

B.2.4 REML estimates the variance components $\boldsymbol{\theta} = (s_L, s_r)$ by maximising the log-likelihood of the redefined measurement vector $\mathbf{M_m Y}$, which is assumed to be normally distributed with mean 0 and covariance matrix $\mathbf{V}(\boldsymbol{\theta})$. $\mathbf{V}(\boldsymbol{\theta})$ is constructed such that:

— all diagonal elements are set to $s_L^2+s_r^2$ (using their current estimates in an iterative scheme);

— all off-diagonal elements for observations from the same laboratory are set to s_L^2;

— all other off-diagonal elements are set to zero.

The log-likelihood $L(\boldsymbol{\theta})$ of $\mathbf{M_m Y}$ given θ can then be written

$$L(\boldsymbol{\theta}) = -\frac{1}{2}\ln\det\left(\mathbf{M_m V}(\theta)\mathbf{M_m^T}\right) - \frac{1}{2}\left[\mathbf{Y}^{\mathrm{T}}\mathbf{M_m^T}\left(\mathbf{M_m V}(\theta)\mathbf{M_m^T}\right)^{-1}\mathbf{M_m Y}\right] \tag{B.4}$$

where the superscript $^{\mathrm{T}}$ denotes matrix transpose.

B.2.5 The variance components are found by maximising $L(\boldsymbol{\theta})$ iteratively as a function of $\boldsymbol{\theta}$. Starting values for $\boldsymbol{\theta}$ can be, for example, zero or (to give a simple upper limit for search algorithms) twice the standard deviation of the data. In the examples in this annex, implementations that reproduce the REML estimates of variance components to three or more significant digits are usually sufficient.

NOTE 1 It can be useful to iterate over values of $L(\boldsymbol{\theta})$ to ensure that estimates of the standard deviations remain strictly positive.

NOTE 2 REML estimation is readily available in most general-purpose statistical software packages, including some open source, free packages.

B.2.6 The REML estimate of reproducibility standard deviation is then calculated from the REML estimates of s_L and s_r using Formula (31).

B.2.7 The corresponding estimate of the mean $\bar{\bar{x}}_{\mathrm{REML}}$ can be estimated, using the REML estimates of s_L and s_r, from Formulae (B.5) and (B.6):

$$\bar{\bar{x}}_{\mathrm{REML}} = \frac{\sum_{i=1}^{p}\bar{x}_i w_i}{\sum_{i=1}^{p} w_i} \tag{B.5}$$

where

$$w_i = \frac{1}{s_L^2 + s_r^2/n_i} \tag{B.6}$$

B.2.8 Where necessary, a standard error $s(\bar{\bar{x}}_{\mathrm{REML}})$ for the mean $\bar{\bar{x}}_{\mathrm{REML}}$ can be obtained from Formula (B.7):

$$s(\bar{\bar{x}}_{\mathrm{REML}}) = 1 \Big/ \sqrt{\sum_{i=1}^{p} w_i} \tag{B.7}$$

Annex C
(informative)

Examples of the statistical analysis of precision experiments

C.1 Example 1: determination of the sulfur content of coal (several levels with no missing or outlying data)

C.1.1 Background

C.1.1.1 Measurement method

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

C.1.1.2 Source

See Reference [7].

C.1.1.3 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.

C.1.1.4 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 C.3.

C.1.2 Original data

These are given, as percentage by mass [mass fraction, %], in Table C.1 in the format of Form A of Figure 2 and do not invite any specific remarks.

Graphical presentations of these data are given in Figure C.1 to Figure C.4.

NOTE For the experiment quoted in Table C.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 document, 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 m_j, s_r and s_R.

Table C.1 — Original data — Sulfur content of coal (mass fraction, %)

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

C.1.3 Computation of cell means, $\bar{y}_{ij}$

The cell means, $\bar{y}_{ij}$, are given, as a percentage by mass, in Table C.2 in the format of Form B of Figure 2 (see 8.2.9).

C.1.4 Computation of standard deviations, s_{ij}

The standard deviations, s_{ij}, are given, as a percentage by mass, in Table C.3 in the format of Form C of Figure 2 (see 8.2.10).

C.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, the largest value of s is in laboratory 8, which has variance 0,000 625 (calculating from the rounded values in Table C.3):

$$\sum s^2 = 0{,}001\ 832;\ \text{test value } 0{,}000\ 625/0{,}001\ 832 = 0{,}341 \qquad \text{(C.1)}$$

Table C.2 — Cell means — Sulfur content of coal (mass fraction, %)

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 C.3 — Standard deviations — Sulfur content of coal (mass fraction, %)

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

For level 2, the largest value of s is in laboratory 5 (variance 0,001 849):

$$\sum s^2 = 0{,}006\ 36;\ \text{test value } 0{,}001\ 849/0{,}006\ 390 = 0{,}289 \quad \text{(C.2)}$$

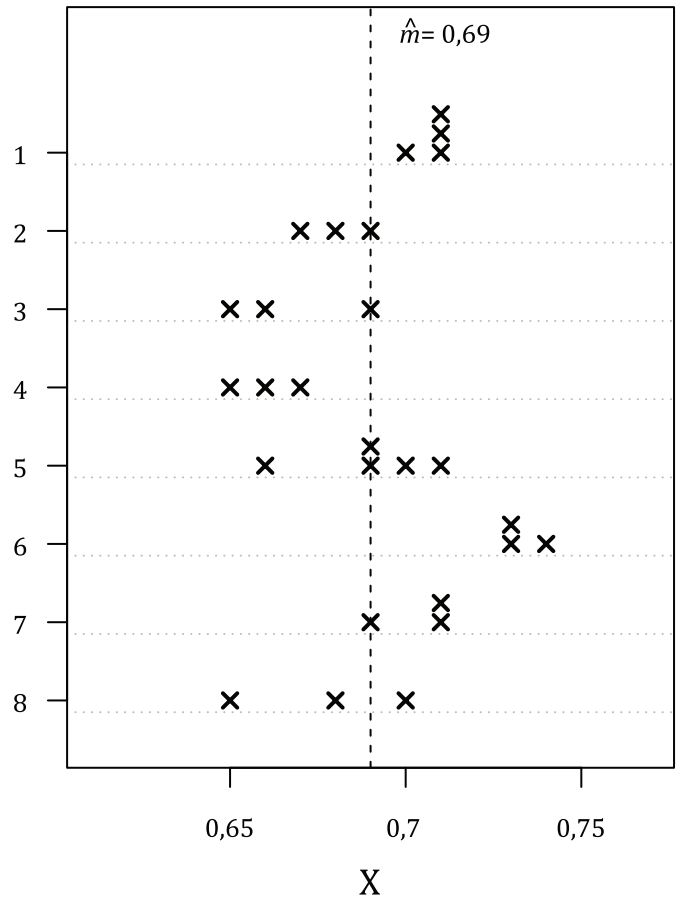
For level 3, the largest value of s is in laboratory 5 (variance 0,001 024):

$$\sum s^2 = 0{,}001\ 72;\ \text{test value} = 0{,}001\ 024/0{,}001\ 765 = 0{,}580 \quad \text{(C.3)}$$

For level 4, the largest value of s is in laboratory 4 (variance 0,001 444):

$$\sum s^2 = 0{,}004\ 63;\ \text{test value} = 0{,}001\ 444/0{,}004\ 649 = 0{,}311 \quad \text{(C.4)}$$

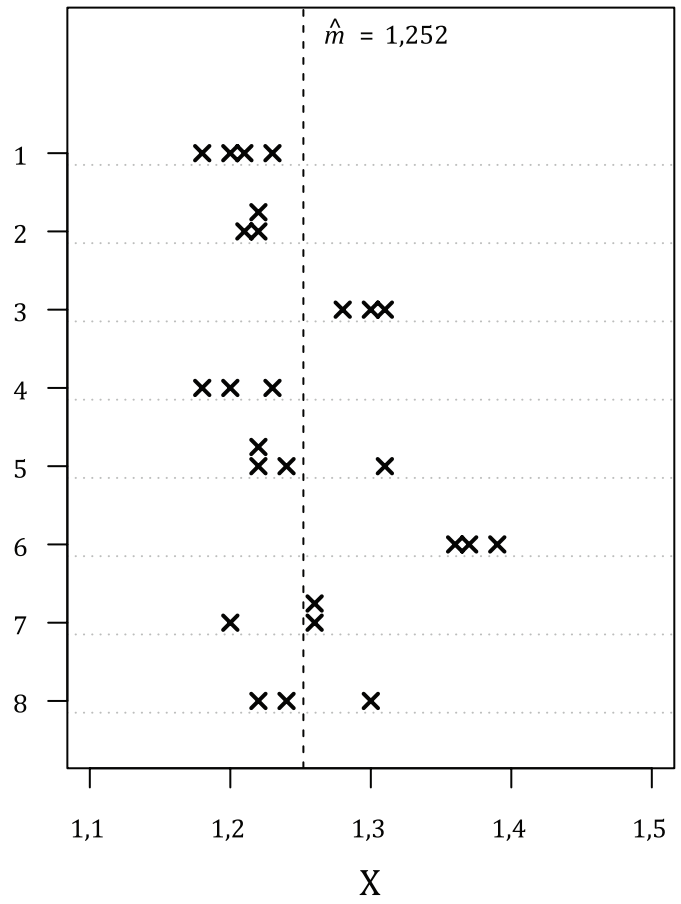
This indicates that one cell in level 3 may be regarded as a straggler, and there are no outliers. The straggler is retained in subsequent calculations.



Key

X mass fraction, %

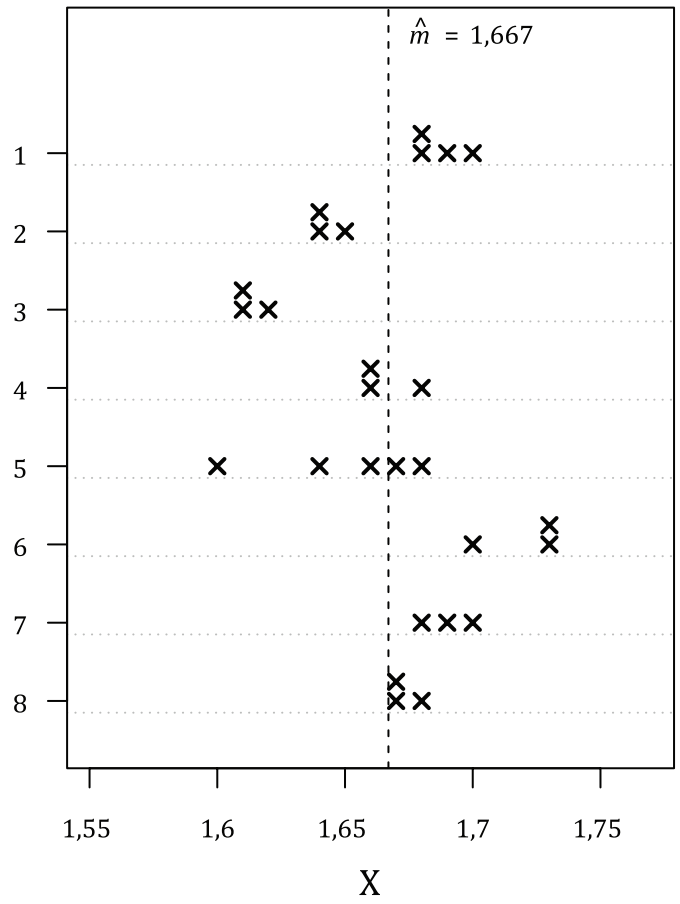
Figure C.1 — Sulfur content of coal, sample 1



Key

X mass fraction, %

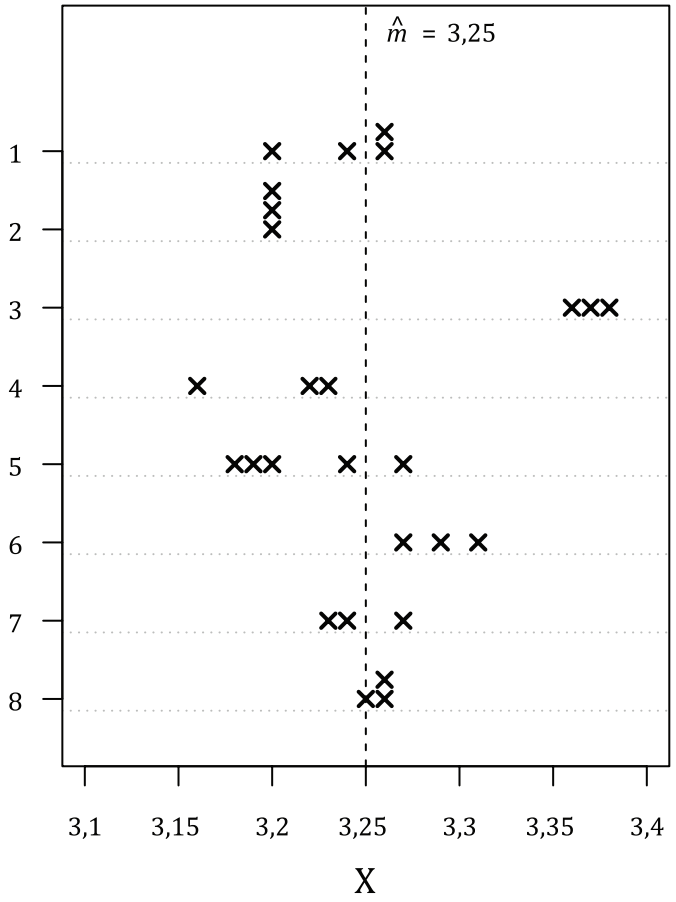
Figure C.2 — Sulfur content of coal, sample 2



Key

X mass fraction, %

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



Key

X mass fraction, %

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

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

Level	Single		Double		Type of test
	Low	High	Low	High	
1	1,24	1,80	0,539	0,298	Grubbs' test statistics
2	0,91	2,08	0,699	0,108	
3	1,67	1,58	0,378	0,460	
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	

NOTE Calculated test statistics are shown for both the single outlier test (8.3.5.1) and the double outlier test (8.3.5.2). The test statistics were calculated for the lowest and highest mean values at each level ("Low" and "High" respectively) and, for the double outlier test, for the two lowest and two highest means.

Grubbs' tests (for single and double outliers as described in 8.3.5.1 and 8.3.5.2 respectively) were applied to the cell means, giving the values shown in Table C.4. The table includes Grubbs' test statistics for the lowest and highest means or (for the double outlier test) lowest and highest pairs 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. Examples of the single and double outlier test calculations, for level 2 only and using the cell means in Table C.2, are as follows:

For the single Grubbs' test (see 8.3.5.1)

$$\bar{x} = \frac{1}{p}\sum_{i=1}^{p} x_i = \frac{1}{8} \times 10{,}036 = 1{,}255 \quad \text{(C.5)}$$

$$s = \sqrt{\frac{1}{p-1}\sum_{i=1}^{p}(x_i - \bar{x})^2} = 0{,}057 \quad \text{(C.6)}$$

For the high value (1,373):

$$G_1 = \frac{(x_1 - \bar{x})}{s} = \frac{(1{,}373 - 1{,}255)}{0{,}057} = 2{,}07 \quad \text{(C.7)}$$

For the low value (1,203):

$$G_1 = \frac{(x_1 - \bar{x})}{s} = \frac{(1{,}255 - 1{,}203)}{0{,}057} = 0{,}91 \quad \text{(C.8)}$$

For the double Grubbs' test (see 8.3.5.2)

$$s_0^2 = \sum_{i=1}^{p}(x_i - \bar{x})^2 = 0{,}022\ 61 \quad \text{(C.9)}$$

For the two highest values (1,297 and 1,373, which are omitted from the sum of squares below):

$$s_{p-1,p}^2 = \sum_{i=1}^{p-2}\left(x_i - \bar{x}_{p-1,p}\right)^2 = 0{,}002\,44 \quad \text{(C.10)}$$

$$G = \frac{s_{p-1,p}^2}{s_0^2} = \frac{0{,}002\ 44}{0{,}022\ 61} = 0{,}108 \quad \text{(C.11)}$$

For the two lowest values (1,203 and 1,205):

$$s_{1,2}^2 = \sum_{i=1}^{p-2}\left(x_i - \bar{x}_{1,2}\right)^2 = 0{,}015\ 81 \quad \text{(C.12)}$$

$$G = \frac{s_{1,2}^2}{s_0^2} = \frac{0{,}015\ 81}{0{,}022\ 61} = 0{,}699 \quad \text{(C.13)}$$

NOTE Some values in the example calculation differ from the tabulated values owing to slight changes in rounding for clarity in the example. These differences are inconsequential.

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

The variances defined in 8.4.4 and 8.4.5 are calculated as follows, using level 1 as an example and calculating from the rounded values in Table C.2 and Table C.3.

NOTE To implement Formulae (23) to (26) and (31), it is convenient to calculate some intermediate values, denoted T_1 to T_5 below, as they appear more than once in the calculation.

Number of laboratories, p = 8

$$T_1 = \sum n_i \bar{y}_i = 18{,}642 \quad \text{(C.14)}$$

$$T_2 = \sum n_i \left(\bar{y}_i\right)^2 = 12{,}883\ 7 \tag{C.15}$$

$$T_3 = \sum n_i = 27 \tag{C.16}$$

$$T_4 = \sum n_i^2 = 95 \tag{C.17}$$

$$T_5 = \sum \left(n_i - 1\right) s_i^2 = 0{,}004\ 411 \tag{C.18}$$

$$s_r^2 = \frac{T_5}{T_3 - p} = 0{,}000\ 232\ 2 \tag{C.19}$$

$$s_L^2 = \left[\frac{T_2 T_3 - T_1^2}{T_3 (p-1)} - s_r^2\right]\left[\frac{T_3 (p-1)}{T_3^2 - T_4}\right] \tag{C.20}$$

$$= 0{,}000\ 460\ 5$$

$$s_R^2 = s_L^2 + s_r^2 = 0{,}000\ 692\ 7 \tag{C.21}$$

$$\hat{m} = \frac{T_1}{T_3} = 0{,}690\ 44 \tag{C.22}$$

$$s_r = 0{,}015\ 24 \tag{C.23}$$

$$s_R = 0{,}026\ 32 \tag{C.24}$$

Table C.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}	RSD_r [a]	RSD_R [a]
1	8	0,690	0,015	0,026	0,022	0,038
2	8	1,252	0,029	0,061	0,023	0,049
3	8	1,667	0,017	0,035	0,010	0,021
4	8	3,250	0,026	0,058	0,008	0,018

[a] "RSD" denotes "relative standard deviation; that is, the standard deviation s divided by the mean m for the level.

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

C.1.7 Dependence of precision on m

An examination of the data in Table C.5 does not indicate any clear dependence and average values can be used.

C.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

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

C.1.9 Alternative calculation

For comparison, restricted maximum likelihood estimation (see B.2) of the values of $\hat{m}_j$, s_{rj} and s_{Rj}, using the same data as for Table C.5, yields the values in Table C.6. There are no material differences from the values in Table C.5.

Table C.6 — REML estimates of $\hat{m}_j$, s_{rj} and s_{Rj} for sulfur content of coal

Level	p	$\hat{m}$	s_r	s_R
1	8	0,690	0,015	0,027
2	8	1,254	0,029	0,062
3	8	1,668	0,017	0,036
4	8	3,253	0,026	0,060

C.2 Example 2: softening point of pitch (several levels with missing data)

C.2.1 Background

C.2.1.1 Measurement method

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

C.2.1.2 Source

Standard methods for testing tar and its products; Pitch section; Method Serial No. PT3 using neutral glycerine (see Reference [6]).

C.2.1.3 Material

This was selected from commercial batches of pitch collected and prepared as specified in the “Samples” chapter of the pitch section of Reference [6].

C.2.1.4 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).

C.2.1.5 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 C.3.

C.2.2 Original data

These are presented in Table C.7, in degrees Celsius, in the format of Form A of Figure 2 (see 8.2.9).

C.2.3 Cell means

These are given in Table C.8, in degrees Celsius, in the format of Form B of Figure 2 (see 8.2.10). A graphical presentation of these data is given in Figure C.5.

Table C.7 — Original data — Softening point of pitch (°C)

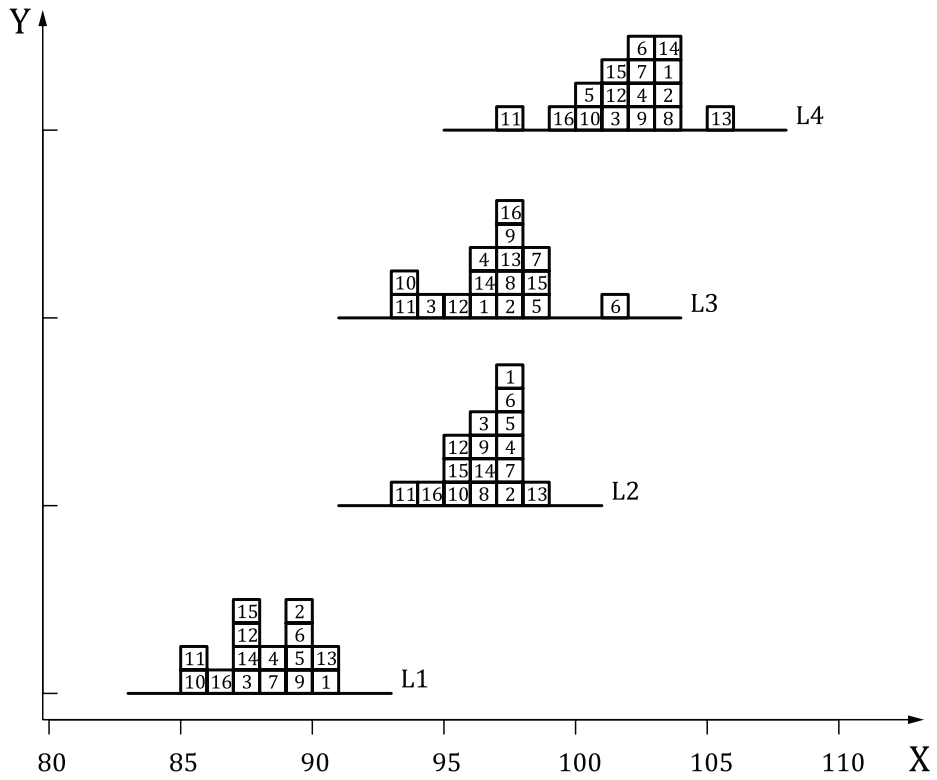
Laboratory *i*	Level *j*			
	1	2	3	4
1	91,0	97,0	96,5	104,0
	89,6	97,2	97,0	104,0
2	89,7	98,5	97,2	102,6
	89,8	97,2	97,0	103,6
3	88,0	97,8	94,2	103,0
	87,5	94,5	95,8	99,5
4	89,2	96,8	96,0	102,5
	88,5	97,5	98,0	103,5
5	89,0	97,2	98,2	101,0
	90,0	—	98,5	100,2
6	88,5	97,8	99,5	102,2
	90,5	97,2	103,2	102,0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

NOTE There are no obvious stragglers or statistical outliers.

Table C.8 — 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 8.4.3).



Key

X temperature (°C)

Y frequency

L1 – L4 level 1 to level 4

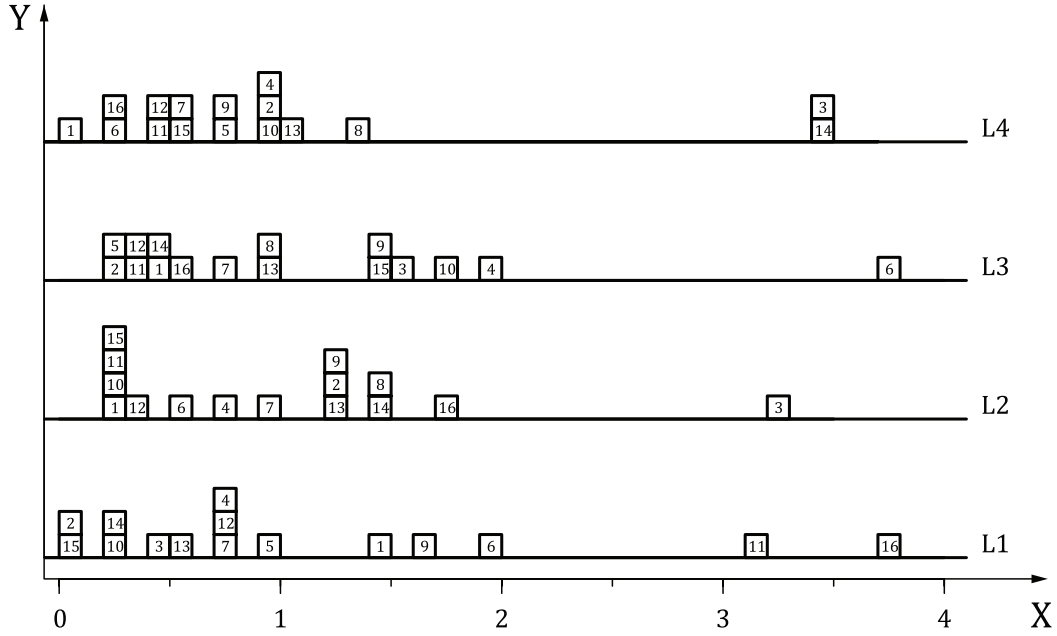
Figure C.5 — Softening point of pitch — Cell means

Table C.9 — 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

C.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 C.9, in the format of Form C of Figure 2. A graphical presentation of these data is given in Figure C.6.



Key

X temperature (°C)
Y frequency
L1 – L4 level 1 to level 4

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

Table C.10 — 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 C.11 — Application of Grubbs' test to cell means

Level *n*	Single low	Single high	Double low	Double high	Types 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	

C.2.5 Scrutiny for consistency and outliers

Application of Cochran's test leads to the values of the test statistic C given in Table C.10.

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 (Table C.11). No single or double stragglers or outliers were found.

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

These are calculated as in 8.4.4 and 8.4.5.

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

$$T_1 = \sum \bar{y}_i = 125,9500 \tag{C.25}$$

$$T_2 = \sum (\bar{y}_i)^2 = 1087,9775 \tag{C.26}$$

$$T_3 = \sum (y_{i1} - y_{i2})^2 = 36,910\ 0 \tag{C.27}$$

$$s_r^2 = \frac{T_3}{2p} = 1,2303 \tag{C.28}$$

$$s_L^2 = \left[\frac{pT_2 - T_1^2}{p(p-1)}\right] - \frac{s_r^2}{2} = 1,5575 \tag{C.29}$$

$$s_R^2 = s_L^2 + s_r^2 = 2,7878 \tag{C.30}$$

$$\hat{m} = \frac{T_1}{p}\ (\text{add } 80,00) = 88,396\ 7 \tag{C.31}$$

$$s_r = 1,109\ 2 \tag{C.32}$$

$$s_R = 1,6697 \tag{C.33}$$

The values for all four levels are given in Table C.12.

Table C.12 — Computed values of $\hat{m}_j$, s_{rj} and s_{Rj} for softening point of pitch

Level j	p_i	$\hat{m}_j$ (°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

C.2.7 Dependence of precision on *m*

A cursory examination of Table C.12 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.

C.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

C.2.9 Alternative calculation

For comparison, restricted maximum likelihood estimation (see B.2) of the values of $\hat{m}_j$, s_{rj} and s_{Rj}, using the same data as for Table C.12, yields the values in Table C.13. There are no material differences from the values in Table C.12.

Table C.13 — REML estimates of $\hat{m}_j$, s_{rj} and s_{Rj} for softening point of pitch

Level	p	$\hat{m}$	s_r	s_R
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	1,918

C.3 Example 3: thermometric titration of creosote oil (several levels with outlying data)

C.3.1 Background

C.3.1.1 Source

Standard methods for testing tar and its products; Creosote oil section; Method Serial No. Co. 18 (see Reference [6]).

C.3.1.2 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 [6].

C.3.1.3 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 %. The usual practice is to record test results to only one decimal place, but for this experiment operators were instructed to work to two decimal places.

C.3.2 Original data

These are presented in Table C.14, as a percentage by mass, in the format of Form A of Figure 2 (see 8.2.8).

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 fits much better at level 4. These points are discussed further in C.3.5.

C.3.3 Cell means

These are given in Table C.15, as a percentage by mass, in the format of Form B of Figure 2 (see 8.2.10).

Table C.14 — 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 C.15 — 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.

Table C.16 — 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.

C.3.4 Absolute differences within cells

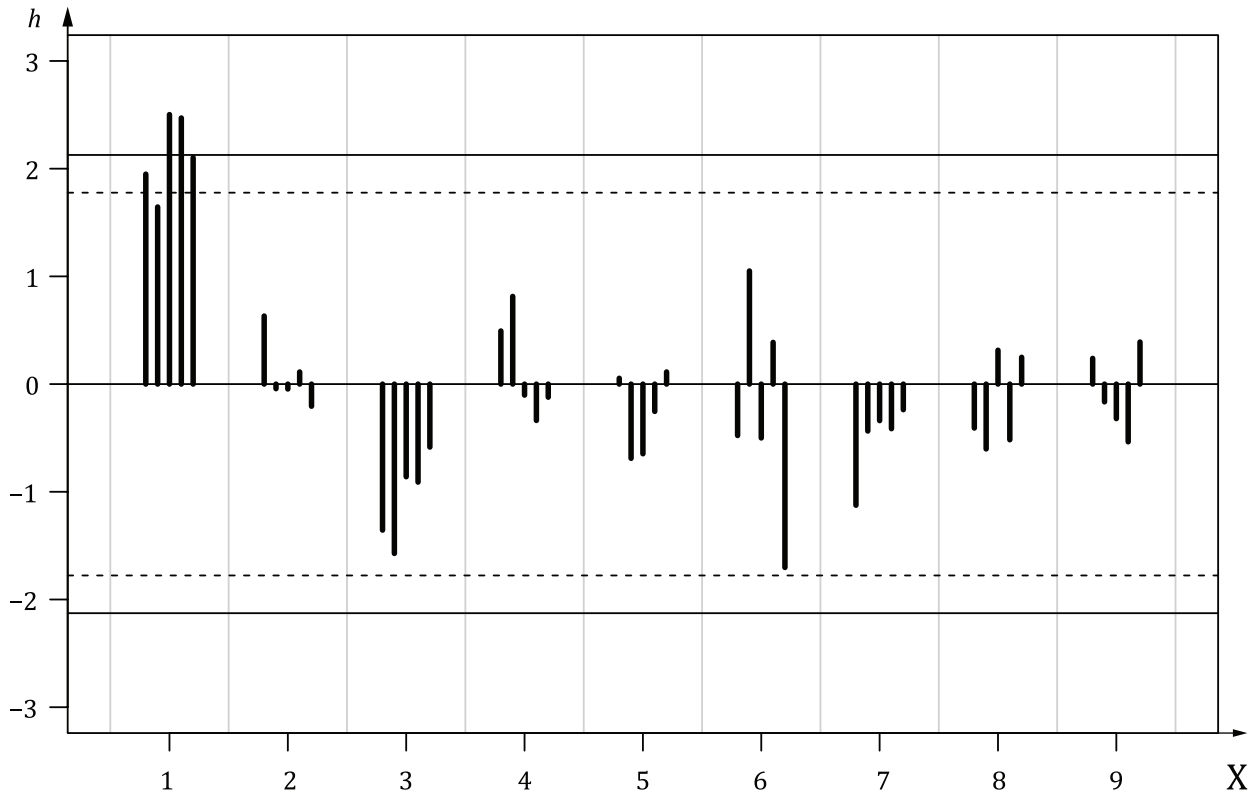
These are given in Table C.16, as w_{ij}, as a percentage by mass, in the format of Form C of Figure 2 (see 8.2.11).

C.3.5 Scrutiny for consistency and outliers

Calculation of Mandel's h and k consistency statistics (see 8.3.2) gave the values shown in Figure C.7 and Figure C.8. Horizontal lines are shown corresponding to the value of Mandel's indicators taken from 8.3.

The h graph (Figure C.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 C.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.

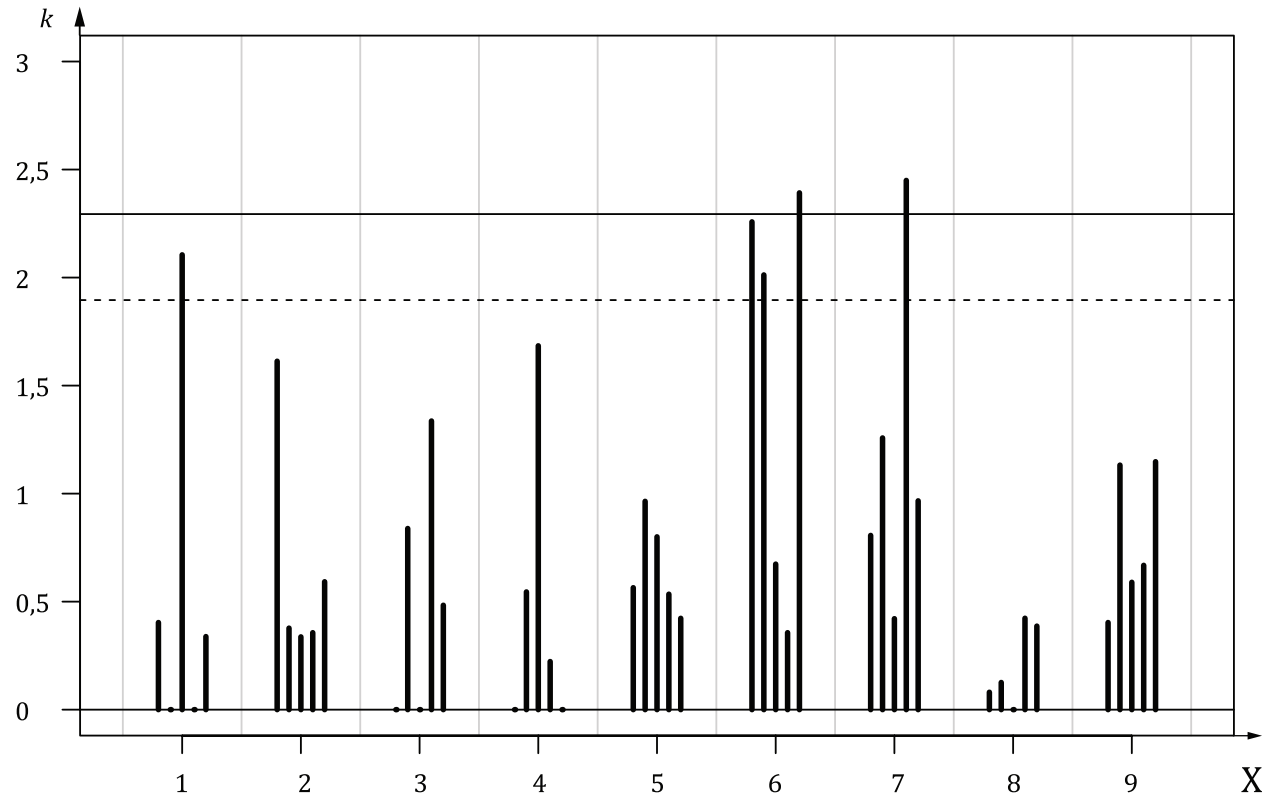


Key

X laboratory number

h Mandel's *h* statistic

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



Key

X laboratory number

k Mandel's *k* statistic

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

Table C.17 — 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	

Application of Cochran's test yields the following results.

— At level 4, the absolute difference 1,10 gave a test statistic value of 1,102/1,814 9 = 0,667.

— At level 5, the absolute difference 1,98 gave a test statistic value of 1,982/6,166 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 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 C.8) also gives rise to suspicion of these values.

Application of Grubbs' tests to the cell means gives the results shown in Table C.17.

For levels 3 and 4, because the single Grubbs' test indicates an outlier, the double Grubbs' test is not applied (see 8.3.5).

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 C.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.

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.

C.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 C.18, as a percentage by mass, calculated as in 8.4.4 and 8.4.5.

Table C.18 — 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

C.3.7 Dependence of precision on *m*

From Table C.18, it seems clear that the standard deviations tend to increase with higher values of $\hat{m}$, 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 8.5.2 to 8.5.4. The values of s_{rj} and s_{Rj} are plotted against $\hat{m}_j$ in Figure C.9. From Figure C.9 it is evident that the value for level 3 is strongly divergent and cannot be improved by any alternative procedures (see 8.5).

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

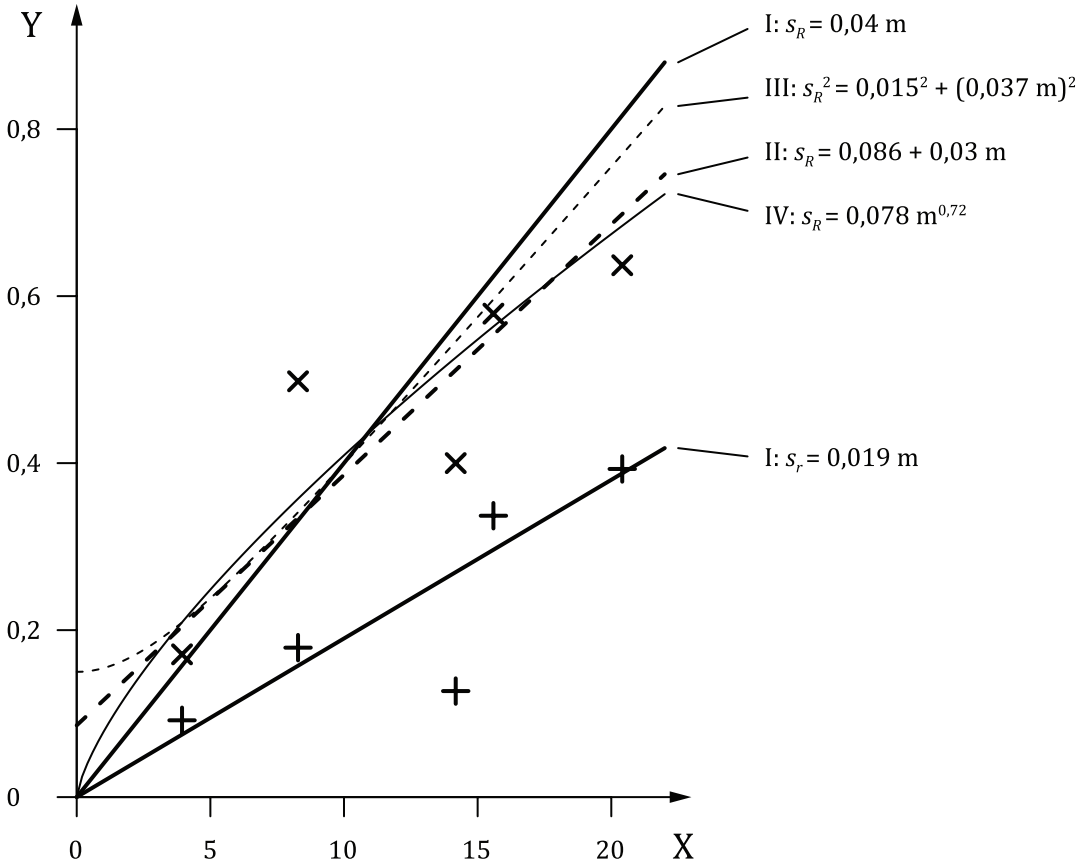
For reproducibility, all three lines show adequate fit with the data, relationship IV perhaps showing the best fit. Someone familiar with the requirements for a standard measurement method for creosote oil can select the most suitable relationship.

C.3.8 Alternative calculation

For comparison, restricted maximum likelihood estimation (see Clause B.2) of the values of $\hat{m}_j$, s_{rj} and s_{Rj}, using the same data as for Table C.18, yields the values in Table C.19. There are no material differences from the values in Table C.18.

Table C.19 — REML estimates of $\hat{m}_j$, s_{rj} and s_{Rj} for thermometric titration of creosote oil

Level	p	$\hat{m}$	s_r	s_R
Level 1	8	3,94	0,092	0,171
Level 2	8	8,28	0,179	0,498
Level 3	8	14,18	0,127	0,400
Level 4	8	15,59	0,337	0,579
Level 5	7	20,41	0,393	0,637



Key

X m (mass fraction, %)
Y s_r, s_R (mass fraction, %)
× reproducibility standard deviation, s_{Rj}
\+ repeatability standard deviation, s_{rj}

Figure C.9 — Plot of s_{rj} and s_{Rj} against $\hat{m}_j$ of the data from Table C.18, showing the functional relationships I-IV from 8.5 fitted to these data

Annex D
(informative)

Calculation of critical values and indicators

D.1 Calculation of critical values for the Cochran's test

The Cochran's test for unusually high laboratory dispersion within one level of a precision study is described in 8.3.4. Critical values at the 1 % and 5 % level of significance (99 % confidence and 95 % confidence respectively) are provided in Table 5. For use in software and for other numbers of laboratories p or levels of significance α, the indicator values can be calculated from Formula (D.1):

$$C_{p,n;1-\alpha} = \frac{1}{\left(1+(p-1)F_{\nu_1,\nu_2;\alpha/p}\right)} \tag{D.1}$$

where

$C_{p,n;1-\alpha}$ denotes the one-tailed upper critical value for Cochran's test at a significance level α for p laboratories each reporting n replicate observations;

p is the number of laboratories;

n is the number of observations in each laboratory;

$F_{\nu_1,\nu_2;\alpha/p}$ is the α/p quantile of the F distribution with $\nu_1 = (p - 1)(n - 1)$ and $\nu_2 = (n - 1)$ degrees of freedom.

NOTE The use of the probability α/p in the F distribution ensures that the probability of one or more values of the test statistic C exceeding $C_{p,n;1-\alpha}$ by chance in a data set of size p is α.

D.2 Calculation of critical values for Grubbs' tests

D.2.1 One outlying observation

Grubbs' test statistic G_1 can be calculated for a single data point or laboratory mean value as described in 8.3.5.1. Critical values are given in Table 6. For use in software and for other numbers of observations n or levels of significance α, the critical values can be calculated using Formula (D.2):

$$G_{1,n;1-\alpha} = \frac{(n-1)t_{p-2;1-\alpha/(2n)}}{\sqrt{n\left(n-2+t^2_{p-2;1-\alpha/(2n)}\right)}} \tag{D.2}$$

where

$G_{1,n;1-\alpha}$ denotes the two-tailed critical value for Grubbs' single outlier test at a significance level α for n results or mean values;

n is the number of values in the data set tested;

$t_{n-2;1-\alpha/(2n)}$ is the $[1 - \alpha/(2n)]$ quantile of Student's t distribution with $n - 2$ degrees of freedom.

NOTE Formula (D.2) has been adapted from Reference [16] to return a two-tailed critical value as used in Table 6; the formula in Reference [16] is for a one-tailed test.

D.2.2 Two outlying observations

Grubbs' test statistic G for two outlying high (or low) observations is calculated as described in 8.3.5.2. Critical values are given in Table 6. For use in software and for other numbers of observations n, the critical values can be calculated using Formula (D.3):

$$G_{n;\alpha} = \frac{1}{1+\frac{2}{n-3}F_{\nu_1,\nu_2;(1-\alpha)^{1/f(n)}}} \quad \text{(D.3)}$$

where

$G_{n;\alpha}$ denotes the one-tailed critical value for Grubbs' outlier test at a significance level α for n results or mean values;

n is the number of values in the data set tested;

$F_{\nu_1,\nu_2;(1-\alpha)^{1/f(n)}}$ is the $[(1-\alpha)^{1/f(n)}]$ quantile of the F distribution with $\nu_1 = 2$ and $\nu_2 = n - 3$ degrees of freedom;

$f(n)$ is the function given in Formula (D.4), with coefficients in Table D.1.

$$f(n) = g_{0,\alpha} + g_{1,\alpha}n + g_{2,\alpha}n^2 \quad \text{(D.4)}$$

NOTE 1 Formula (D.3) provides one-tailed critical values. To reproduce the values in Table 6, the significance level α in Formula (D.3) is replaced by $\alpha/2$; for example, the tabulated values for significance level of 5 %, can be reproduced by using $\alpha = 0{,}025$ in Formula (D.3), with the corresponding coefficients from Table D.1.

NOTE 2 Formula (D.3) and the coefficients in Table D.1 are not exact, but reproduce exact values to within ±0,003 for the significance levels listed in Table D.1[16].

Table D.1 — Coefficients for Grubbs' test critical values

α	$g_{0,\alpha}$	$g_{1,\alpha}$	$g_{2,\alpha}$
0,001 0	−4,249 3	1,001 2	0,044 3
0,005 0	−3,661 3	0,955 8	0,038 8
0,010 0	−3,310 1	0,925 0	0,036 2
0,025 0	−2,858 0	0,883 3	0,032 2
0,050 0	−2,507 5	0,850 1	0,028 9
0,100 0	−2,161 5	0,816 9	0,025 1
SOURCE: Reference [16], reproduced with permission of Springer-Verlag.			

D.3 Calculation of indicators for Mandel's h and k statistics

D.3.1 Mandel's h statistic

Mandel's h statistic for a given data point can be calculated as described in 8.3.2.2, from all of the p_j laboratory means at a given level j of a precision experiment. Indicators at the 1 % and 5 % level of significance are provided in Table 7 and Table 8 respectively. For use in software and for other numbers of laboratories p or levels of significance α, the indicator values can be calculated from Formula (D.5)

$$h_{p;1-\alpha/2} = \frac{(p-1)t_{p-2;1-\alpha/2}}{\sqrt{p\left(p-2+t^2_{p-2;1-\alpha/2}\right)}} \quad \text{(D.5)}$$

where

$h_{p;1-\alpha/2}$ denotes the two-tailed indicator value for Mandel's h at a significance level α for a total of p laboratories;

p is the number of laboratories;

$t_{p-2;1-\alpha/2}$ is the $(1 - \alpha/2)$ quantile of Student's t distribution with p-2 degrees of freedom.

NOTE 1 Plots for Mandel's h are constructed with indicator lines at $h_{p;1-\alpha/2}$ and at $-h_{p;1-\alpha/2}$.

NOTE 2 Individual values outside $\pm h_{p;1-\alpha/2}$ are expected to occur with frequency α by chance. The probability of one or more such values occurring by chance in a data set of size p is accordingly $1-(1-\alpha)^p$, which is typically much larger than α.

D.3.2 Mandel's k statistic

Mandel's k statistic for a given laboratory is calculated as described in 8.3.2.3, from p_j variances of n observations at a given level j of a precision experiment. Indicators at the 1 % and 5 % level of significance are provided in Table 7 and Table 8 respectively. For use in software and for other numbers of laboratories p, number of observations n or levels of significance α, the indicator values for k can be calculated from Formula (D.6)

$$k_{p,n;1-\alpha} = \sqrt{\frac{p}{1+(p-1)F_{\nu_1,\nu_2;\alpha}}} \tag{D6}$$

where

$k_{p,n;1-\alpha}$ denotes the one-tailed upper indicator value for Mandel's k at a significance level α for p laboratories each reporting n replicate observations;

p is the number of laboratories;

n is the number of observations in each laboratory;

$F_{\nu_1,\nu_2;\alpha}$ is the α quantile of the F distribution with $\nu_1 = (p - 1)(n - 1)$ and $\nu_2 = (n - 1)$ degrees of freedom.

NOTE 1 Plots for Mandel's k are constructed with one indicator line at $k_{p,n;1-\alpha}$.

NOTE 2 Individual values above $k_{p,n;1-\alpha}$ are expected to occur with frequency α by chance. The probability of one or more such values occurring by chance in a data set of size p is accordingly $1-(1-\alpha)^p$, which is typically much larger than α.

Bibliography

[1] ISO Guide 33, *Reference materials — Good practice in using reference materials*

[2] ISO Guide 35, *Reference materials — Guidance for characterization and assessment of homogeneity and stability*

[3] ASTM E691-18, *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., Procedures for detecting outlying observations in samples. *Technometrics*, **11**, 1969, pp. 1–21

[5] Grubbs F.E., Beck G., Extension of sample sizes and percentage points for significance tests of outlying observations. *Technometrics*, **14**, 1972, pp. 847–854

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

[7] Tomkins S.S. *Industrial and Engineering Chemistry (Analytical edition)*, **14**, 1942, pp. 141–145

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

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

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

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

[12] Corbeil R.R., Searle S.R., Restricted Maximum Likelihood (REML) Estimation of Variance Components in the Mixed Model, *Technometrics*, **18**, 1976, 31-38

[13] Harville D.A., Maximum Likelihood Approaches to Variance Component Estimation and to Related Problems. *Journal of the American Statistical Association*, **72** (358), 1977, 320–338

[14] Searle S.R., Casella G., McCulloch C.E., (1992). Variance components. Wiley, New York

[15] International Union of Pure and Applied Chemistry, Protocol for the design, conduct and interpretation of method-performance studies. *Pure and Appl. Chem.*, Vol. **67**, No. 2, 1995, pp. 331-343

[16] Wilrich P, Critical values of Mandel's h and k, the Grubbs and the Cochran test statistic. *Advances in Statistical Analysis* **97**, 2013, 1-10, DOI 10.1007/s10182-011-0185-y

[17] ISO 21748, *Guidance for the use of repeatability, reproducibility and trueness estimates in measurement uncertainty evaluation*

(*Continued from second cover*)

International Standard	*Corresponding Indian Standard*	*Degree of Equivalence*
ISO 3534-2 Statistics — Vocabulary and symbols — Part 2: Applied statistics	IS 7920 (Part 2) : 2012 Statistics — Vocabulary and symbols: Part 2 Applied statistics (*third revision*)	Identical with ISO 3534-2 : 2006
ISO3534-3 Statistics — Vocabulary and symbols — Part 3: Design of experiments	IS 7920 (Part 3) : 2018 Statistics — Vocabulary and symbols: Part 3 Design of experiments	Identical with ISO 3534-3 : 2013
ISO 5725-1 Accuracy (trueness and precision) of measurement methods and results — Part 1: General principles and definitions	IS 15393 (Part 1) : 2003 Accuracy (trueness and precision) of measurement methods and results: Part 1 General principles and definitions	Identical with ISO 5725-1 : 1994

Annexes A to D of this standard are for information only.

Bureau of Indian Standards

BIS is a statutory institution established under the *Bureau of Indian Standards Act*, 2016 to promote harmonious development of the activities of standardization, marking and quality certification of goods and attending to connected matters in the country.

Review of Indian Standards

Amendments are issued to standards as the need arises on the basis of comments. Standards are also reviewed periodically; a standard along with amendments is reaffirmed when such review indicates that no changes are needed; if the review indicates that changes are needed, it is taken up for revision. Users of Indian Standards should ascertain that they are in possession of the latest amendments or edition by referring to the latest issue of 'BIS Catalogue' and 'Standards: Monthly Additions'.

This Indian Standard has been developed from Doc No.: MSD 03 (16392).

Amendments Issued Since Publication

Amend No.	Date of Issue	Text Affected

BUREAU OF INDIAN STANDARDS

Headquarters:

Manak Bhavan, 9 Bahadur Shah Zafar Marg, New Delhi 110002
Telephones: 2323 0131, 2323 3375, 2323 9402 *Website*: www.bis.gov.in

Regional Offices:		*Telephones*
Central	: Manak Bhavan, 9 Bahadur Shah Zafar Marg, NEW DELHI 110002	2323 7617, 2323 3841
Eastern	: 1/14 C.I.T. Scheme VII M, V.I.P. Road, Kankurgachi, KOLKATA 700054	2337 8499, 2337 8561, 2337 8626, 2337 9120
Northern	: Plot No. 4-A, Sector 27-B, Madhya Marg, CHANDIGARH 160019	265 0206, 265 0290
Southern	: C.I.T. Campus, IV Cross Road, CHENNAI 600113	2254 1216, 2254 1442, 2254 2519, 2254 2315
Western	: Manakalaya, E9 MIDC, Marol, Andheri (East), MUMBAI 400093	2832 9295, 2832 7858, 2832 7891, 2832 7892

Branches : AHMEDABAD. BENGALURU. BHOPAL. BHUBANESHWAR. COIMBATORE. DEHRADUN. DURGAPUR. FARIDABAD. GHAZIABAD. GUWAHATI. HYDERABAD. JAIPUR. JAMMU. JAMSHEDPUR. KOCHI. LUCKNOW. NAGPUR. PARWANOO. PATNA. PUNE. RAIPUR. RAJKOT. VISAKHAPATNAM.

Published by BIS, New Delhi