INTERNATIONAL STANDARD

ISO 5725-2

Second edition
2019-12

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

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

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

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

Reference number
ISO 5725-2:2019(E)

Publishe

Foreword

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

The procedures used to develop this document and those intended for its further maintenance are described in the ISO/IEC Directives, Part 1. In particular the different approval criteria needed for the different types of ISO documents should be noted. This document was drafted in accordance with the editorial rules of the ISO/IEC Directives, Part 2 (see www.iso.org/directives).

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. ISO shall not be held responsible for identifying any or all such patent rights. Details of any patent rights identified during the development of the document will be in the Introduction and/or on the ISO list of patent declarations received (see www.iso.org/patents).

Any trade name used in this document is information given for the convenience of users and does not constitute an endorsement.

For an explanation of the voluntary nature of standards, the meaning of ISO specific terms and expressions related to conformity assessment, as well as information about ISO's adherence to the World Trade Organization (WTO) principles in the Technical Barriers to Trade (TBT), see www.iso.org/iso/foreword.html.

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

This second edition cancels and replaces the first edition (ISO 5725-2:1994), which has been technically revised. It also incorporates the Technical Corrigendum ISO 5725-2:1994/Cor 1:2002.

The main changes compared to the previous edition are as follows:

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

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

— guidance on the number of laboratories required for a precision study has been included;

— information on the computation of critical values has been included.

A list of all parts in the ISO 5725 series can be found on the ISO website.

Any feedback or questions on this document should be directed to the user's national standards body. A complete listing of these bodies can be found at www.iso.org/members.html.

Introduction

ISO 5725 uses two terms, "trueness" and "precision", t
"Trueness" refers to the closeness of agreement be
test results and the true or accepted reference valu
between test results.

General consideration of these quantities is given in I
ISO 5725-1 should be read in conjunction with all oth
gives the underlying definitions and general principl

This document is concerned solely with estima
reproducibility standard deviation based on an i
conducts a number of independent measurements o
There are other designs (such as nested, factorial or
estimation of precision: these are not dealt with in
parts of ISO 5725. Nor does this document consid
between the two principal measures; those are the s

In certain circumstances, the data obtained from ar
used also to estimate trueness and can be used to e
of trueness is not considered in this document; all as
of ISO 5725-4. The evaluation of measurement uncer
and precision, is the subject of ISO 21748.

Annex C provides practical examples of estimat
experiment. Worked examples are given to demonst
in one example a variable number of replicates pe
another some data were missing. This is because an
be unbalanced. Stragglers and outliers are also cons

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
$\boldsymbol{\theta}$	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 t
by an estimate based upon a sample, th
This is done in each of ISO 5725-1:1994

— s_L^2 is the estimate of the between-l

— s_W^2 is the estimate of the within-la

— s_r^2 is the arithmetic mean of s_W^2 a
mean is taken over all those labora
outliers have been excluded;

— s_R^2 is the estimate of the reproduci

$$s_R^2 = s_L^2 + s_r^2$$

6 Requirements for a precisio

6.1 Layout of the experiment

6.1.1 In the layout used in the basi
different levels of the test, are sent to
under repeatability conditions at eacl
uniform-level experiment.

6.1.2 The performance of these meas

a) Any preliminary checking of equip

b) Each group of n measurements b
conditions, i.e. within a short ir
intermediate recalibration of the
measurement.

c) It is essential that a group of n te
as if they were n tests on differen
is testing identical material, but t
purpose of the experiment is to de
If it is feared that, despite this war
thus the repeatability variance, it
of the q levels, coded in such a wa
given level. However, such a proced
apply between replicates. This is o
qn measurements can be performe

d) It is not essential that all the q grou
interval of time; different groups o

e) Measurements of all q levels shall
n measurements at a given level sh

f) If in the course of the measurem
may complete the measurements
n measurements at one level but or
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

ARPA TOSCANA - 2020716 - XPOD

ISO 5725-2:2019(E)

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 pr

b) to decide on the number of la
significant digits to be requir

c) to appoint someone for the st

d) to appoint someone for the ex

e) to consider the instructions t
measurement method;

f) to decide whether some oper
order to regain experience of
carried out on the official coll

g) to discuss the report of the st

h) to establish final values for t
deviation;

i) to decide if further actions ar
with regard to laboratories w

7.2 Statistical functions

At least one member of the pa
experiments. His/her tasks are:

a) to contribute his/her speciali

b) to analyse the data;

c) to write a report for submissi

7.3 Executive functions

7.3.1 The actual organization of
of the staff of that laboratory shou
appointed by the panel.

7.3.2 The tasks of the executive

a) to enlist the cooperation of th
appointed;

b) to organize and supervise th
samples; for each level, an ade

c) to draft instructions covering
early enough in advance for t
selected are those who norma

d) to design suitable forms for
report the test results to the r
Such forms can include the n
measured, the equipment use

e) to deal with any queries from

7.5 Operators

7.5.1 In each la
representative of t

7.5.2 Because tl
population of ope
should not be give
pointed out to the
vary in practice, s
inconsistent.

7.5.3 Although r
measurement met
whether the instru

7.5.4 The tasks

a) to perform th

b) to report any
the test resu
indicate a def

c) to comment
occasion(s) w
in the standa

8 Statistical

8.1 Prelimina

8.1.1 The analy
statistical expert,

a) critical exam
test the suita

b) computation

c) establishmen
between pre

8.1.2 The analy

— the repeatab

— the between-

— the reproduc

— the mean, *m*.

NOTE The ana
are available from
only a limited num

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 ($k = 1, 2, \ldots, n_{ij}$);

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

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

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

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):

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

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 – Recom	
Laboratory	1
1	
2	
⋮	
i	y_{i11} ⋮ y_{i1k} ⋮ y_{i1n}
⋮	
p	

Form B – Rec	
Laboratory	1
1	$\bar{y}_{11}$
2	
⋮	
i	
⋮	
p	$\bar{y}_{p1}$

Form C – Recommended	
Laboratory	1
1	s_{11}
2	
⋮	
i	
⋮	
p	s_{p1}

a Form A contains collated individ

b Form B contains cell means as ca

c Form C contains measures of spr

Figure 2 — Recomm

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 results, Cochran's test statistic, C, is given by Formula

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

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

a) If the test statistic is less than or equal to its correct.

b) If the test statistic is greater than its 5 % criti value, the item tested is called a straggler and is

c) If the test statistic is greater than its 1 % critica 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

8.3.4.3 Cochran's criterion applies strictly only w the same number (n) of test results obtained under t conditions, in this document). In actual cases, this nu This document assumes, however, that in a properly o of test results per cell are limited and can be ignored for n the number of test results occurring in the majo

8.3.4.4 Cochran's criterion tests only the highest va one-sided outlier test. Variance heterogeneity can als being comparatively too low. However, small values of by the degree of rounding of the original data and a seems unreasonable to reject the data from a laborat in its test results than the other laboratories. Hence C

8.3.4.5 A critical examination of Form C of Figure 2 for a particular laboratory are at all or at most lev can indicate that the laboratory works with a lowe laboratories, which in turn can be caused either by or incorrect application of the standard measureme the panel, which should then decide whether the po example of this is laboratory 2 in the experiment deta

8.3.4.6 If the highest standard deviation is classe and Cochran's test repeated on the remaining value excessive rejections when, as is sometimes the cas sufficiently upheld. The repeated application of Coch view of the lack of a statistical test designed for testir is not designed for this purpose and great caution s two or three laboratories give results having high st the levels, conclusions from Cochran's test should b stragglers and/or statistical outliers are found at di strong indication that the laboratory's within-laborat 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, $x_{(1)}, x_{(2)}, \ldots, x_{(p)}$, formed from $x_1, x_2, \ldots, x_p$ 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 Formula (14):

$$G = \frac{s^2_{p-1,p}}{s_0^2}$$

where

$$s_0^2 = \sum_{i=1}^{p}\left(x_{(i)} - \bar{x}\right)^2$$

and

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

and

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

Alternatively, to test the two smallest observations, com

$$G = \frac{s^2_{1,2}}{s_0^2}$$

where

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

and

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

8.3.5.3 Application of the Grubbs' test

When analysing a precision experiment, Grubbs' test can

a) The cell averages (Form B of Figure 2) for a given lev

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

and

$$p = p_j$$

where j is fixed.

Taking the data at one level, apply Grubbs' test fo described in 8.3.5.1. If a cell mean is shown to be an test at the other extreme cell mean (e.g. if the highes highest excluded), but do not apply Grubbs' test for t

If Grubbs' test does not sho
described in 8.3.5.2.

b) A single result within a cel
suspect.

NOTE According to 8.3.3.1
1 % critical value. When Grubb
used to test the highest cell me
level. This amounts to a test of
most extreme cell mean is four
cell mean. It can be argued that
in this document is to use only
at the 1 % significance level) i
justify the use of the two-sided

8.3.6 Repeated testing for ou

8.3.6.1 It is often found that a
of an outlier, leaves a smaller set
values relative to the remainder
outlying data points or laboratory

8.3.6.2 Repeated outlier testin
to rejection of an excessive prop
proportion of rejected data is exc

NOTE 1 It can be helpful to set a
analysis.

NOTE 2 IUPAC[15] recommends t
exceed the fraction 2/9, the fractio
grounds prior to outlier inspection.

8.3.7 Alternative outlier insp

Alternative inspection methods
provided that:

— the same set of outlier insp
given study;

— the methods chosen are capa
outlying laboratory means, a

— any bias in variance estimate
larger than that resulting fro

NOTE All outlier rejection proc
correction for the resulting bias is n
estimates for normally distributed d

8.4 Calculation of the gene

8.4.1 Method of analysis

The method of analysis adopted
precision for each level j separat
value of j.

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 T
(REML)[13].
variance gi

NOTE 1 A
Annex B.

NOTE 2 T
elsewhere i

8.4.6.3 W
method sha

8.4.7 De

After calcu
upon *m* an

8.5 Esta
level, *m*

8.5.1 Ch

8.5.1.1 It
precision a
of the test
level *m*. Wi
regular fun
the materia

8.5.1.2 W
reported fo
an average

8.5.1.3 T
repeatabili
interests of

I. $s_r = bm$

II. $s_r = a +$

III. $s_r^2 = a_v^2$
deviati

IV. $\lg s_r =$

It is to be
If not, the
confusion,
a_r, b_r, ... for
this clause
the level *j*.

8.5.1.4 I
from an ex
that they a

8.5.1.5 For $a = 0$ (c
to relationship I), so
yield practically equ
the following simple

"Two test result

In statistical termin
all levels.

8.5.1.6 If, in a plo

reasonably close to
some reason a num
relationships I and I

8.5.2 Fitting rela

8.5.2.1 From a sta

and s_j are estimates
less), then errors in

8.5.2.2 A good es
because the standar

have to be proportio

However, $\hat{s}_j$ depend
for finding estimate
following iterative p

8.5.2.3 With weig

calculated formulae

$$T_1 = \sum_j W_j$$

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

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

$$T_4 = \sum_j W_j s_j$$

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

Then for relationshi

8.5.3.2 An

i) With we
calculat

T_{v1}

T_{v2}

T_{v3}

T_{v4}

T_{v5}

ii) Then, fo

$a_v^2 =$

$b_v^2 =$

iii) Initial v

$\hat{s}_{1j}^2$

iv) Comput

$\hat{s}_{2j}^2$

The resultin

NOTE Th
sample varia

EXAMPLE
the same as f

8.5.4 Fitti

8.5.4.1 Th
appropriate,

8.5.4.2 Fo

$T_1 = \sum_j$

W_{0vj}
$s_1^2 = 0,088$
$\hat{s}_{1j}$
W_{1vj}
$s_2^2 = 0,061$
$\hat{s}_{2j}$
W_{2vj}
$s_3^2 = 0,063$
$\hat{s}_{3j}$
NOTE Th
[a] The di

$\lg \hat{m}_j$
$\lg s_{0j}$
lg s = –1,5
or s = 0,0
s

8.6 Sta

NOTE

8.6.1 C
that this
contribut

In a unif
may be p

8.6.2 Ir
erroneou
impossibl
test resul
Such obvi
for furthe

8.6.3 Fi
and Form

When a c

2) Use Formula (58)

$$\frac{1}{q}\sum_j s_j = s_r$$

to calculate the final value of the repeatability standard d directly to 8.6.14.

3) Judge from the plot of s against m and of s^2 and m^2 whet between s^2 and m^2, can be represented by a straight line relationship II ($s = a + bm$) or relationship III ($s_j^2 = a_v^2$ relationships I and II, determine the parameter b, or the of 8.5.2; for relationship III, use the procedure of 8.5.3. satisfactory, ignore step 4) and proceed directly with ste

4) Plot lg s_j against lg $\hat{m}$ and judge from this whether th reasonably be represented by a straight line. If this is con ($\lg s = c + d \lg m$) using the procedure given in 8.5.4.

5) If a satisfactory relation has been established in step 3 are the smoothed values obtained from this relationshi proceed with 8.6.14.

6) If no satisfactory relation has been established in step 3) whether some other relation between s and m can be est are so irregular that the establishment of a functional re Where the data are too irregular to provide a functio reported separately for each level.

8.6.14 Prepare a report showing the basic data and the re analysis, and present this to the panel. The graphical present the consistency or variability of the results.

8.7 Report to the panel and decisions to be taken by

8.7.1 Report by the statistical expert

Having completed the statistical analysis, the statistical expe the panel. In this report the following information should be g

a) a full account of the observations received from the ope standard for the measurement method;

b) a full account of the laboratories that have been rejected 8.6.8, together with the reasons for their rejection;

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

d) a form of the final results $\hat{m}_j$, s_r and s_R and an account illustrated by one of the plots recommended in these ste

e) Forms A, B and C (Figure 2) used in the statistical analysi

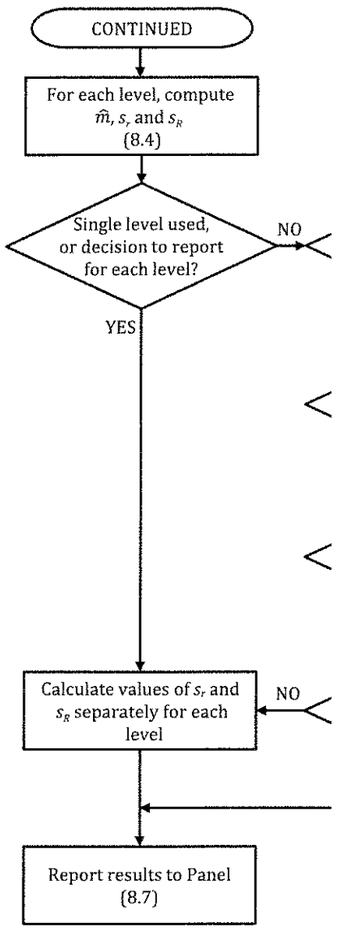


Figure 3 — Flow diagram of th

8.7.2 Decisions to be taken by the pa

The panel should then discuss this report

a) Are the discordant results, straggle standard for the measurement meth

b) What action should be taken with re

c) Do the results of the outlying labora supervisors indicate the need to im are the improvements required?

d) Do the results of the precision exper standard deviation and reproducibil form should they be published, and w

Table 5 *(continued)*

p	$n = 2$		$n = 3$		$n = 4$		$n = 5$		
	1 %	5 %	1 %	5 %	1 %	5 %	1 %	5 %	1 %
24	0,425	0,343	0,287	0,235	0,230	0,191	0,197	0,166	0,1
25	0,413	0,334	0,278	0,228	0,222	0,185	0,190	0,160	0,1
26	0,402	0,325	0,270	0,221	0,215	0,179	0,184	0,155	0,1
27	0,391	0,316	0,262	0,215	0,209	0,173	0,179	0,150	0,1
28	0,382	0,308	0,255	0,209	0,202	0,168	0,173	0,146	0,1
29	0,372	0,300	0,248	0,203	0,196	0,164	0,168	0,142	0,1
30	0,363	0,293	0,241	0,198	0,191	0,159	0,164	0,138	0,1
31	0,355	0,286	0,235	0,193	0,186	0,155	0,159	0,134	0,1
32	0,347	0,280	0,229	0,188	0,181	0,151	0,155	0,131	0,1
33	0,339	0,273	0,224	0,184	0,177	0,147	0,151	0,127	0,1
34	0,332	0,267	0,218	0,179	0,172	0,144	0,147	0,124	0,1
35	0,325	0,262	0,213	0,175	0,168	0,140	0,144	0,121	0,1
36	0,318	0,256	0,208	0,172	0,165	0,137	0,140	0,118	0,1
37	0,312	0,251	0,204	0,168	0,161	0,134	0,137	0,116	0,1
38	0,306	0,246	0,200	0,164	0,157	0,131	0,134	0,113	0,1
39	0,300	0,242	0,196	0,161	0,154	0,129	0,131	0,111	0,1
40	0,294	0,237	0,192	0,158	0,151	0,126	0,128	0,108	0,1

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 sm	
	Upper 1 %	Upper 5 %	Lower 1 %	Lo
3	1,155	1,155	—	
4	1,496	1,481	0,000 0	0,
5	1,764	1,715	0,001 8	0,
6	1,973	1,887	0,011 6	0,
7	2,139	2,020	0,030 8	0,
8	2,274	2,126	0,056 3	0,
9	2,387	2,215	0,085 1	0,
10	2,482	2,290	0,115 0	0,
11	2,564	2,355	0,144 8	0,
12	2,636	2,412	0,173 8	0,
13	2,699	2,462	0,201 6	0,
14	2,755	2,507	0,228 0	0,

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 a values required by the procedure for applying Grubbs' outlier tests described in 8.3.5 of this document. Th 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 softwa 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 Mand

p	h			
		2	3	4
3	1,15	1,65	1,53	1,4
4	1,42	1,76	1,59	1,5
5	1,57	1,81	1,62	1,5
6	1,66	1,85	1,64	1,5
7	1,71	1,87	1,66	1,5
8	1,75	1,88	1,67	1,5
9	1,78	1,90	1,68	1,5
10	1,80	1,90	1,68	1,5
11	1,82	1,91	1,69	1,5
12	1,83	1,92	1,69	1,5
13	1,84	1,92	1,69	1,5
14	1,85	1,92	1,70	1,5
15	1,86	1,93	1,70	1,5
16	1,86	1,93	1,70	1,5
17	1,87	1,93	1,70	1,5
18	1,88	1,93	1,71	1,5
19	1,88	1,93	1,71	1,5
20	1,89	1,94	1,71	1,5
21	1,89	1,94	1,71	1,6
22	1,89	1,94	1,71	1,6
23	1,90	1,94	1,71	1,6
24	1,90	1,94	1,71	1,6
25	1,90	1,94	1,71	1,6
26	1,90	1,94	1,71	1,6
27	1,91	1,94	1,71	1,6
28	1,91	1,94	1,71	1,6
29	1,91	1,94	1,72	1,6
30	1,91	1,94	1,72	1,6

p = number of laboratories at a given level

n = number of replicates within each laboratory

NOTE 1 Supplied by Dr. J. Mandel and published

NOTE 2 Annex D, based on Reference [16], prov 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:

$$\gamma=\sigma_R/\sigma_r \tag{A.2}$$

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.

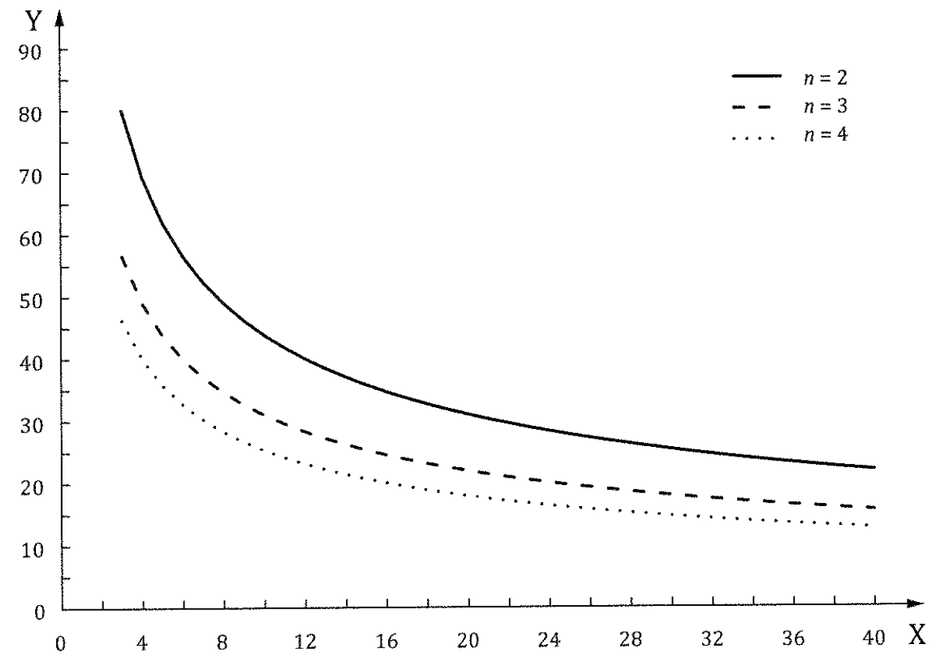
A.4 The value of γ is not known, but often preliminary estimates are available of the within-laboratory standard deviations and the between-laboratory standard deviations obtained during the process of standardizing the measurement method. Values of the uncertainty (as a decimal fraction) for repeatability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

ARPA TOSCANA - 2020716 - XPOD

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.

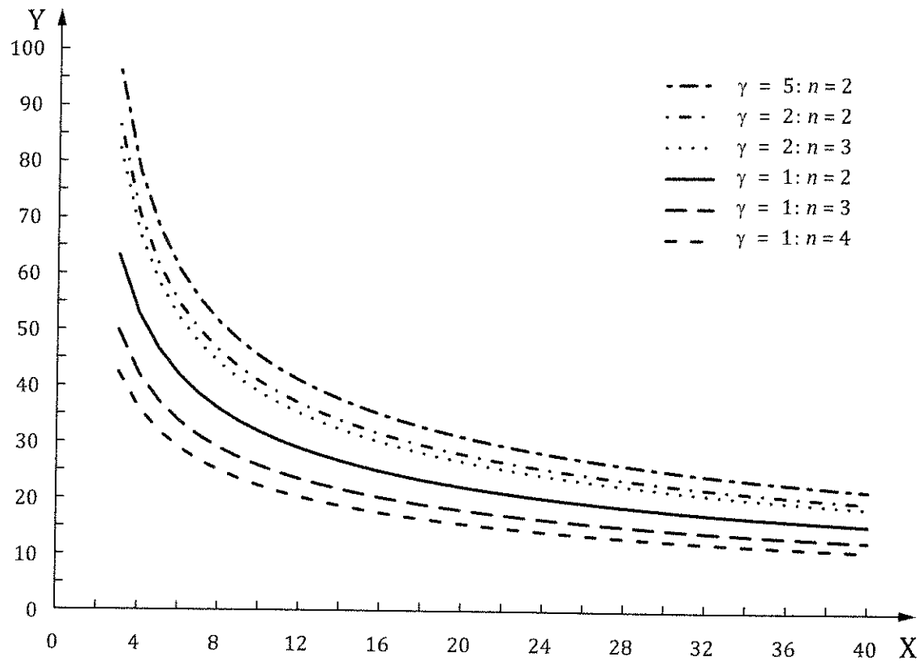
Table A.1 — Values showing the uncertainty of estimates of the repeatability and reproducibility standard deviations

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

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

where the superscript ^{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):

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

C.1

C.1

Det

C.1

See

C.1

Eig
me
lab

C.1

Ma
hav
Ma

C.1

The
Eig

Gra

NOT
mea
doc
all d
num
sele
has

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):

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

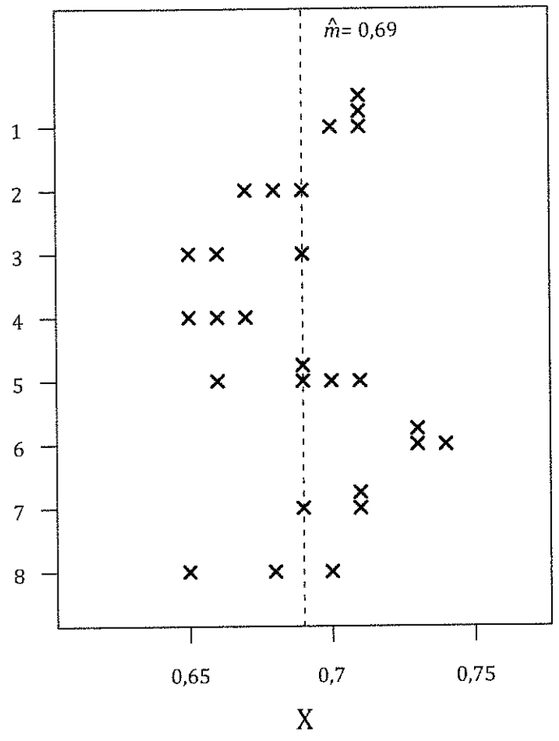
For level 3, the largest value of s is in laboratory 5 (variance 0,001 024):

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

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

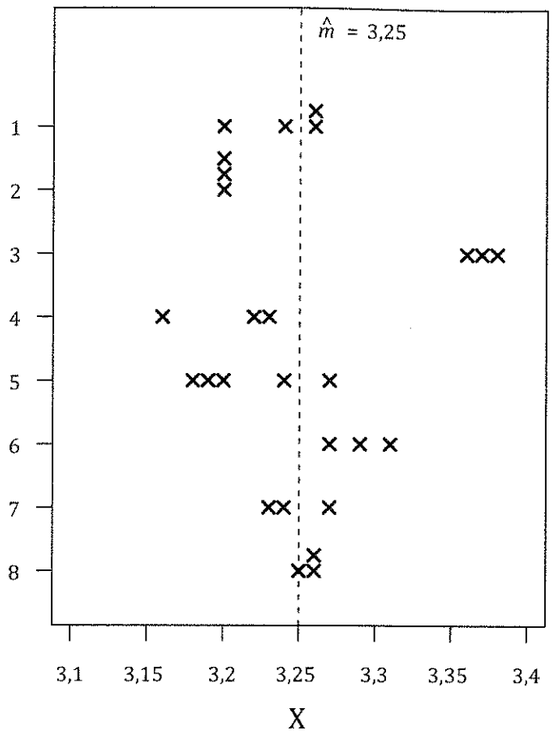
$$\sum s^2 = 0{,}004\ 63;\ \text{test value} = 0{,}001\ 444/0{,}004\ 649 = 0{,}311 \qquad \text{(C.4)}$$

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.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, $p = 8$

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

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

$$T_3 = \sum n_i = 27 \quad \text{(C.16)}$$

$$T_4 = \sum n_i^2 = 95 \quad \text{(C.17)}$$

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

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

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

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

$$s_R^2 = s_L^2 + s_r^2 = 0{,}000\ 692\ 7 \quad \text{(C.21)}$$

$$\hat{m} = \frac{T_1}{T_3} = 0{,}690\ 44 \quad \text{(C.22)}$$

$$s_r = 0{,}015\ 24 \quad \text{(C.23)}$$

$$s_R = 0{,}026\ 32 \quad \text{(C.24)}$$

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 applie
determined from a uniforı
which four stragglers were

C.1.9 Alternative calcı

For comparison, restricte

using the same data as fo
from the values in Table C.

Table C.6 –

C.2 Example 2: softe

C.2.1 Background

C.2.1.1 Measurement n

The determination of the s

C.2.1.2 Source

Standard methods for tes
glycerine (see Reference [

C.2.1.3 Material

This was selected from co
chapter of the pitch sectio

C.2.1.4 Description

This was the determinat
Sixteen laboratories coop
97,5 °C and 102,5 °C to co
for level 2 with a mean te
the method incorrectly a
then insufficient material
not have a sample for leve

C.2.1.5 Graphical prese

Mandel's h and k statistics
to provide for another ty
discussed in the example

C.2.2 Original data

These are presented in Ta

C.2.3 Cell means

These are given in Table
graphical presentation of

Tab

Labo													
NOTE													

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

$$T_1 = \sum \bar{y}_i = 125,9500 \quad \text{(C.25)}$$

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

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

$$s_r^2 = \frac{T_3}{2p} = 1,2303 \quad \text{(C.28)}$$

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

$$s_R^2 = s_L^2 + s_r^2 = 2,7878 \quad \text{(C.30)}$$

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

$$s_r = 1,109\ 2 \quad \text{(C.32)}$$

$$s_R = 1,6697 \quad \text{(C.33)}$$

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

Thi
"Sa

C.3

Thi
wit
spe
ran
app
dec

C.3

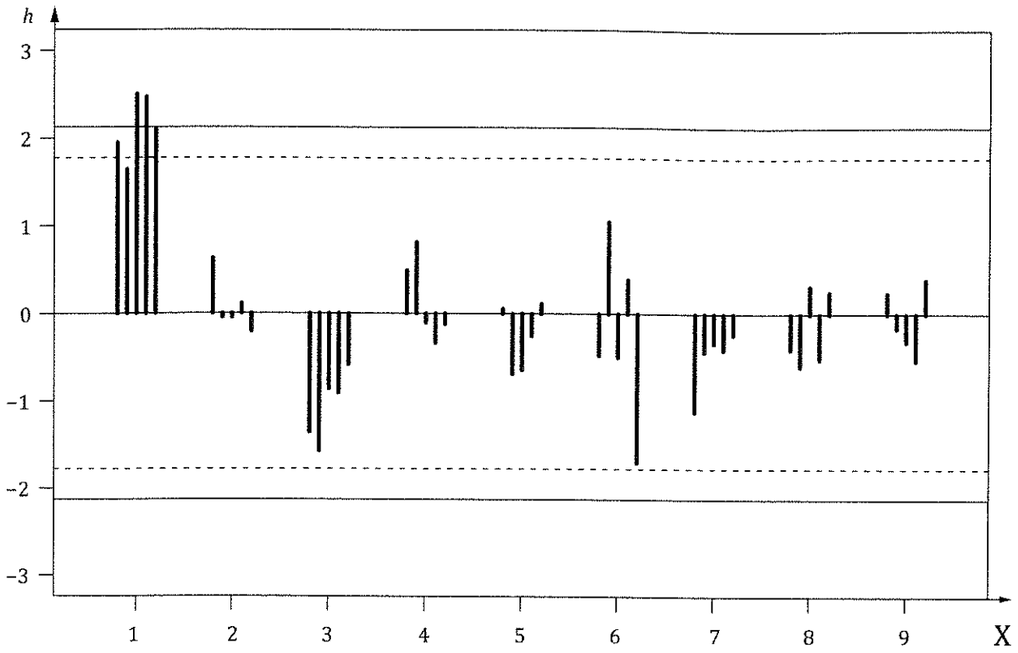
The

The
of t
fits

C.3

The

L									

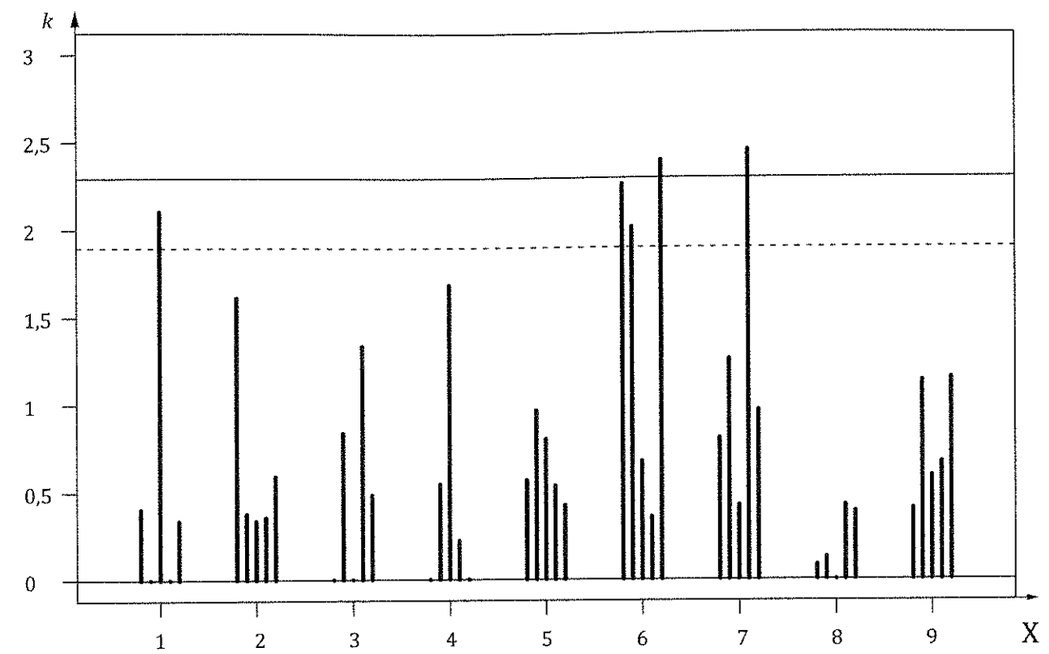


Key

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

$h_{p;1-\alpha/2}$ denotes the two-tailed ir
of p laboratories;

p is the number of laborat

$t_{p-2;1-\alpha/2}$ is the $(1 - \alpha/2)$ quantile

NOTE 1 Plots for Mandel's h are construct

NOTE 2 Individual values outside $\pm h_{p;1-\alpha}$
of one or more such values occurring by chan
much larger than α.

D.3.2 Mandel's k statistic

Mandel's k statistic for a given laborat
of n observations at a given level j of a
significance are provided in Table 7 and T
of laboratories p, number of observations
calculated from Formula (D.6)

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

where

$k_{p,n;1-\alpha}$ denotes the one-tailed upp
p laboratories each reporti

p is the number of laborator

n is the number of observati

$F_{\nu_1,\nu_2;\alpha}$ is the α quantile of the F di
freedom.

NOTE 1 Plots for Mandel's k are constructe

NOTE 2 Individual values above $k_{p,n;1-\alpha}$
of one or more such values occurring by chanc
much larger than α.