INTERNATIONAL STANDARD

ISO 5725-3

Second edition
2023-06

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

Part 3: Intermediate precision and alternative designs for collaborative studies

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

Partie 3: Fidélité intermédiaire et plans alternatifs pour les études collaboratives

Reference number
ISO 5725-3:2023(E)

ISO copyright office
CP 401 • Ch. de Blandonnet 8
CH-1214 Vernier, Geneva
Phone: +41 22 749 01 11
Email: copyright@iso.org
Website: www.iso.org

Published in Switzerland

Contents

Page

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 document 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).

ISO draws attention to the possibility that the implementation of this document may involve the use of (a) patent(s). ISO takes no position concerning the evidence, validity or applicability of any claimed patent rights in respect thereof. As of the date of publication of this document, ISO had not received notice of (a) patent(s) which may be required to implement this document. However, implementers are cautioned that this may not represent the latest information, which may be obtained from the patent database available at www.iso.org/patents. ISO shall not be held responsible for identifying any or all such patent rights.

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 5725-3:1994), which has been technically revised. It also incorporates the Technical Corrigendum ISO 5725-3:1994/Cor.1:2001.

The main changes are as follows:

— Several additional experimental designs have been added to this version compared to the previous version, some of them from ISO 5725-5. These are orthogonal array designs, split level designs, designs for heterogeneous sample material as well as designs across levels.

— Furthermore, the standard was supplemented by considerations on the selection of factors and modelling of the factorial effects, as well as by a section in which the reliability of the various interlaboratory test parameters (mean and precision parameters) are considered.

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

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

0.2 General consideration of these quantities is given in ISO 5725-1 and is not repeated here. It is stressed that ISO 5725-1 provides underlying definitions and general principles should be read in conjunction with all other parts of ISO 5725.

0.3 Many different factors (apart from test material heterogeneity) may contribute to the variability of results from a measurement method, including:

a) the laboratory;

b) the operator;

c) the equipment used;

d) the calibration of the equipment;

e) the batch of a reagent;

f) the time elapsed between measurements;

g) environment (temperature, humidity, air pollution, etc.);

h) other factors.

0.4 Two conditions of precision, termed repeatability and reproducibility conditions, have been found necessary and, for many practical cases, useful for describing the variability of a measurement method. Under repeatability conditions, none of the factors a) to h) in 0.3 are considered to vary, while under reproducibility conditions, all of the factors are considered to vary and contribute to the variability of the test results. Thus, repeatability and reproducibility conditions are the two extremes of precision, the first describing the minimum and the second the maximum variability in results. Intermediate conditions between these two extreme conditions of precision are also conceivable, when one or more of the factors listed in b) to g) are allowed to vary.

To illustrate the need for including a consideration of intermediate conditions in method validation, consider the operation of a present-day laboratory connected with a production plant involving, for example, a three-shift working system where measurements are made by different operators on different equipment. Operators and equipment are then some of the factors that contribute to the variability in the test results.

The standard deviation of test results obtained under repeatability conditions is generally less than that obtained under intermediate precision conditions. Generally, in chemical analysis, the standard deviation under intermediate precision conditions may be two or three times larger than that under repeatability conditions. It should not, of course, exceed the reproducibility standard deviation.

As an example, in the determination of copper in copper ore, a collaborative study among 35 laboratories revealed that the standard deviation under intermediate precision conditions (different times) was 1,5 times larger than that under repeatability conditions, both for the electrolytic gravimetry and $Na_2S_20_3$ titration methods.

0.5 This document focuses on intermediate precision and alternative designs for collaborative studies of a measurement method. Apart from the determination of intermediate precision measures, the aims of these alternative designs include reducing the number of required measurements, increasing the reliability of the estimates for precision and overall mean and taking into account test material heterogeneity.

Indeed, a t-factor fully-nested experiment with two levels per factor (inside each laboratory, there are $t-1$ factors) and two replicates per setting requires $2 \cdot 2^{t-1}$ test results from each laboratory, which can be an excessive requirement on the laboratories. For this reason, in the previous version of ISO 5725-3, the staggered nested design is also discussed. While the estimation of the precision parameters is more complex and subject to greater uncertainty in a staggered nested design, the workload is reduced. This document offers alternative strategies to reduce the workload without compromising the reliability of the precision estimates.

As far as the special designs for sample heterogeneity are concerned, they were discussed in the previous version of ISO 5725-5. However, it is convenient to have one part of this standard dedicated to the question of the design of experiments.

0.6 The repeatability precision as determined in accordance with ISO 5725-2 is computed as a mean across participating laboratories. Whether it can be used for quality control purposes depends on whether the repeatability standard deviation can be considered to remain constant across laboratories. For this reason, it is important to obtain information on how the repeatability standard deviation varies within and between the laboratories under different conditions.

0.7 In many collaborative studies, the between-laboratory variability is large in comparison to the repeatability, and it would be useful to a) decompose it into several different precision components, b) reduce, if possible, some sources of variability which are due to the intermediate precision conditions. This can be done by identifying factors (e.g. time, calibration, operator or equipment) which contribute to the variability under intermediate precision conditions of measurement, by quantifying the corresponding variability components and, wherever achievable, decreasing their contribution. In this manner, the intermediate precision component of the overall variance is enlarged while the between-laboratory component of the overall variance is reduced. Only random effects are considered: it is only reasonable to model a factor as a fixed effect after a method or calibration optimization study has been conducted. In this standard, different relationships between factors are taken into account, e.g. whether a particular factor is subsumed under another factor or not.

0.8 Estimates for precision and overall mean are subject to random variability. Accordingly, it is important to determine the uncertainty associated with each estimate, and to understand the relationships between this uncertainty, the number of participants and the design. Once these relationships are understood, it becomes possible to make much more informed decisions concerning the number of participants and the experimental design.

0.9 Provided different factorial effects do contribute to the variability, determining the respective precision components may make it possible to reduce the required number of participating laboratories, since the between-laboratory variability can be expected to be less dominant. However, it is highly recommended to have a reasonable number of participating laboratories in order to ensure a realistic assessment of the overall method variability obtained under routine conditions of operation.

0.10 In the uniform-level design according to part 2 of this standard, there is a risk that an operator will allow the result of a measurement on one sample to influence the result of a subsequent measurement on another sample of the same material, causing the estimates of the repeatability and reproducibility standard deviations to be biased. When this risk is considered to be serious, the split-level design described in this document may be preferred as it reduces this risk. Care should be taken that the two materials used at a particular level of the experiment are sufficiently similar to ensure that the same precision measures can be expected (in other words: the question arises whether the precision component associated with a particular factor remains unchanged across a range of similar matrices).

0.11 The experimental design presented in ISO 5725-2 requires the preparation of a number of identical samples of the material for use in the experiment. With heterogeneous materials this may not be possible, so that the use of the basic method then gives estimates of the reproducibility standard deviation that are inflated by the variation between the samples. The design for a heterogeneous material given in this document yields information about the variability between samples which is not obtainable from the basic method; it may be used to calculate an estimate of reproducibility from which the between-sample variation has been removed.

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

Part 3: Intermediate precision and alternative designs for collaborative studies

1 Scope

This document provides

a) a discussion of alternative experimental designs for the determination of trueness and precision measures including reproducibility, repeatability and selected measures of intermediate precision of a standard measurement method, including a review of the circumstances in which their use is necessary or beneficial, and guidance as to the interpretation and application of the resulting estimates, and

b) worked examples including specific designs and computations.

Each of the alternative designs discussed in this document is intended to address one (or several) of the following issues:

a) a discussion of the implications of the definitions of intermediate precision measures;

b) a guidance on the interpretation and application of the estimates of intermediate precision measures in practical situations;

c) determining reproducibility, repeatability and selected measures of intermediate precision;

d) improved[1)] determination of reproducibility and other measures of precision;

e) improving the estimate of the sample mean;

f) determining the range of in-house repeatability standard deviations;

g) determining other precision components such as operator variability;

h) determining the level of reliability of precision estimates;

i) reducing the minimum number of participating laboratories by optimizing the reliability of precision estimates;

j) avoiding distorted estimations of repeatability (split-level designs);

k) avoiding distorted estimations of reproducibility (taking the heterogeneity of the material into consideration).

Often, the performance of the method whose precision is being evaluated in a collaborative study will have previously been assessed in a single-laboratory validation study conducted by the laboratory which developed it. Relevant factors for the determination of intermediary precision will have been identified in this prior single-laboratory study.

1) Allowing a reduction in the number of laboratories.

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: General statistical terms and terms used in probability*

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

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

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

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

3 Terms and definitions

For the purposes of this document, the terms and definitions given in ISO 3534-1, ISO 3534-2 and ISO 5725-1 and the following 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 https://www.electropedia.org/

3.1
block
group of *settings* (3.7) conducted in parallel or within a short time interval, and with the same samples

EXAMPLE Two settings:

Operator 1 + Calibration 1 + Equipment 1 + Batch 1

and

Operator 1 + Calibration 2 + Equipment 2 + Batch 1

Note 1 to entry: This definition is more specific than the general definition given in ISO 3534-3:2013, 3.1.25, where block is defined as a collection of experimental units.

3.2
factor
feature under examination as a potential source of variation

EXAMPLE Operator, calibration, equipment, day, reagent batch, storage temperature, shaker orbit, shaker frequency.

Note 1 to entry: Strictly speaking, the factor laboratory is a factor just like any other. However, since the ISO 5725 standard focuses on method validation by means of interlaboratory studies, the factor laboratory can be considered to have a somewhat privileged role. The following characteristics distinguish it from other factors:

— The factor laboratory is indispensable: For each measurement, the name of the particular laboratory where it was performed will *always* be provided in a collaborative study.

— The factor laboratory will almost always have more levels than other factors.

It should also be noted that categories such as measurand, sample/matrix and level may also be considered to be factors. However, in collaborative studies, they are often not taken into account as such in the factorial design. The reason is that, for these factors, one is interested in a separate statistical analysis for each separate factor level. In other words, one is interested in obtaining separate precision measures for each particular measurand or concentration level, not across measurands or concentration levels. However, in cases where it is required to quantify precision across, say, matrices, then the factor sample/matrix should also be included in the design. Accordingly, in this document, designs are discussed to be applied for a particular measurand or concentration level by different laboratories all applying the same measurement procedure.

[SOURCE: ISO 3534-3:2013, 3.1.5, modified — Note 1 to entry was modified and Note 2 to entry was deleted.]

3.3
factor level
setting (3.7), value or assignment of a *factor* (3.2)

EXAMPLE Operator 1, Operator 2

Note 1 to entry: In many designs, the majority of factors will be varied across two levels.

3.4
fully-nested design
nested design, where there is a nesting hierarchy for every pair of *factors* (3.2)

EXAMPLE There are 2 operators in each laboratory, and each operator performs 2 calibrations, i.e., the study includes 2 operators and 4 calibrations for each laboratory.

3.5
partially-nested design
nested design where one *factor* (3.2) (the factor laboratory) is ranked higher than all other factors (i.e., all other factors are nested within the factor laboratory), and there is at least one factor pair without a nesting hierarchy

EXAMPLE There are 2 operators and 2 instruments in each laboratory, and each operator performs measurements on 2 instruments, i.e., the study includes 2 operators and 2 instruments for each laboratory.

3.6
run
actual measurement carried out for a particular *setting* (3.7) and for a particular laboratory

EXAMPLE Operator 1 + Equipment 1 + Batch 1 + Day 1 carried out in laboratory 1

Note 1 to entry: This definition is more specific than the general definition given in ISO 3534-3 (3.1.13), where run is defined as specific settings of every factor used on a particular experimental unit.

Note 2 to entry: "Identical" runs are called *replicates*, whereby "identical" means that the different time points are close enough to each other to allow for the results to be considered as obtained under repeatability conditions.

3.7
setting
combination of *factor levels* (3.3), for all *factors* (3.2) except the factor laboratory

EXAMPLE Operator 1 + Equipment 1 + Batch 1 + Day 1.

4 Symbols

B	Component in a test result representing the deviation of a laboratory from the general average (laboratory component of bias)

B_0	Component of B representing all factors that do not vary under intermediate precision conditions – laboratory bias *per se*
$B_{(1)}, B_{(2)}$, etc.	Components of B representing factors that vary under intermediate precision conditions
e	Component representing the random error occurring in every test result, corresponding to the analytical, repeatability, model or residual error
m	Overall mean of the measurand or test property for a particular matrix; level
$\hat{m}$	Estimate of the overall mean
n	Number of replicate test results obtained in one laboratory at one level for one setting
p	Number of laboratories participating in the collaborative study
q	Number of levels of the test property in the collaborative study
σ_w	Within-laboratory standard deviation of the residual term e
σ_r	Repeatability standard deviation
σ_R	Reproducibility standard deviation
σ_0	Standard deviation corresponding to factor B_0
$\sigma_{(1)}$	Standard deviation corresponding to factor $B_{(1)}$
$\sigma_{(2)}$	Standard deviation corresponding to factor $B_{(2)}$
σ_A	Standard deviation corresponding to factor A
$\sigma_{Interaction}$	Standard deviation corresponding to the interaction of two factors
σ_{AB}	Standard deviation corresponding to the interaction of the two factors A and B
s	Estimate of a standard deviation
se	Standard error
$Var(X)$	Variance of X
w	Range of a set of test results
y	Test result
$\bar{X}$	Mean of X
$\lvert X \rvert$	Absolute value of X

5 General requirements

In order to ensure that measurements are carried out in the same way, the measurement method shall have been standardized. All measurements obtained in the framework of an experiment within a specific laboratory or of a collaborative study shall be carried out according to that standard.

NOTE The terms collaborative experiment, collaborative trial and interlaboratory experiment are used interchangeably to denote a collaborative study conducted in order to characterize and/or assess the performance of a measurement method.

6 Intermediate measures of the precision of a standard measurement method

6.1 Factors and factor levels

6.1.1 Definitions and examples

In this document, the term factor denotes an identifiable and quantifiable source of variability such as time, calibration, operator or equipment (see 3.2). In order to investigate a factor's contribution to variability, it is necessary to conduct measurements under different conditions or states. For instance, measurements shall be carried out with different pieces of equipment, or with different operators. The different states associated with a particular factor are called factor levels (see 3.3). Table 1 provides typical examples of factors and their factor levels.

Table 1 — Examples of factors

Factor	Description/example of the different factor levels	Comments
Laboratory	The different participating laboratories, typically between 4 and 15 different laboratories.	Some of the special designs presented in this document allow reliable precision estimates with as few as 4 participating laboratories.
Point in time	Two different time points (e.g. different days, different weeks, etc.)	Differences between "measurements made at different times", i.e. separated by a relatively long time interval (as compared with the repeatability interval) will reflect effects which correspond to uncontrolled changes in environmental conditions as well as other "controlled" sources of variability such as the use of different reagent batches, etc.
Calibration	Before and after instrument is sent to the manufacturer for a recalibration	Calibration does not refer here to any calibration required as an integral part of obtaining a test result by the measurement method. It refers to the calibration process that takes place at regular intervals between groups of measurements within a laboratory.
Operator	The different technicians working in the laboratory	In some circumstances, the operator may be, in fact, a team of operators, each of whom performs some specific part of the procedure. In such a case, the team should be regarded as the operator, and any change in membership or in the allotment of duties within the team should be regarded as constituting a different operator.
Equipment	Two different pieces of equipment	Equipment is often a set of equipment, and any change in any significant component should be regarded as constituting different equipment. As to what constitutes a significant component, common sense must prevail (e.g. different burettes/pipettes, thermometers, pH meters, centrifuges, shaker orbits or frequencies).
Consumables (buffer solutions, reagents, calibrators, cartridges)	Different batches or producers	A change of a batch of a reagent should be considered a significant component. It can lead to different equipment or to a recalibration if such a change is followed by calibration.

NOTE 1 In practice, it may not be possible to consider factors in isolation from one another; this is due to a characteristic of experimental designs called *confounding*. In theory, it should always be possible to disentangle the effects of different factors by additional testing. For instance, if Operator 1 always carried out tests with Equipment 1 (e.g. HPLC system 1) and Operator 2 with Equipment 2, then it would be possible to tell the effects of the two factors *Operator* and *Equipment* apart by adding further runs for Operator 1 with Equipment 2 and for Operator 2 with Equipment 1.

NOTE 2 Further effects called interaction effects are not explicitly considered here. However, some interaction effects are implicitly taken into consideration. For instance, the effect of skill or fatigue of an operator may be considered to be the interaction of operator and time. Similarly, the performance of a piece of equipment may be different at the time it is first turned on and after many hours of use: this is an example of interaction between equipment and time.

NOTE 3 In ISO 5725-2, the factor laboratory is implicitly included in the analysis.

6.1.2 Selection of factors of interest

In the standard for a measurement method, the repeatability and reproducibility standard deviations should always be specified, but it is not necessary (or even feasible) to state all possible intermediate precision measures. The selection of relevant factors is informed by experience and an understanding of the relevant physical, chemical or microbiological processes.

Practical considerations in most laboratories, such as the desired precision of the final quoted result and the cost of performing the measurements, will govern the number and choice of factors taken into consideration in the standardization of the measurement method.

Finally, the choice of factors to include in the design should reflect concerns with uncontrollable variations between the laboratories.

It will often be sufficient to specify only one suitable intermediate precision measure, together with a detailed stipulation of the specific measurement conditions associated with it. The factors should be carefully defined; in particular, for the intermediate precision associated with the factor *Time*, a practical mean time interval between successive measurements should be specified.

It is assumed that, in the case of a standardized measurement method, the bias inherent in the method itself will have been corrected by technical means. For this reason, this document only addresses the bias arising in connection with different measurement conditions.

6.1.3 Random and fixed effects

This subclause provides a discussion of the question why, in this document, factors are modelled as random rather than as fixed effects.

The term fixed effect is used to describe a contribution to the deviation from the overall mean or true value whose direction and magnitude is predictable and can thus be determined. Say, for example, that measurements always lie below the true value with equipment 1 or reagent supplier 1 and above the true value with equipment 2 or reagent supplier 2. Then it would be appropriate to model the factor *Equipment* or *Reagent supplier* as a fixed effect.

On the other hand, the term *random effect* is used to describe a contribution to the deviation from the overall mean or true value whose direction varies – and thus cannot be determined. In such cases, the only quantity of interest is the magnitude of the contribution (independently of its direction) often described in terms of a standard deviation.

NOTE A factor is modelled as a fixed effect if the specific factor levels included in the experiment are of interest in and of themselves. On the other hand, if the aim is to characterize the variability associated with the underlying population from which the factor levels were selected, the factor is modelled as a random effect. In this document, it is usually the variability of the underlying population which is of interest, rather than the individual factor levels included in the experiment – this is the rationale for modelling factors as random.

The rationale for modelling factors as random rather than as fixed effects is now illustrated on the basis of several examples.

Table 2 — Rationale for modelling factors as random rather than as fixed effects

Factor	Discussion
Operator	Effects due to differences between operators include personal habits in operating measurement methods, e.g. in reading graduations on scales, etc. Thus, even though there is a bias in the test results obtained by an individual operator, this bias is not always constant. The magnitude of such a bias should be reduced by use of a clear operation manual and training. Under such circumstances, the effect of changing operators can be considered to be of a random nature.
Equipment	Effects due to different equipment include the effects due to different places of installation, particularly in fluctuations of the indicator, etc. Systematic differences should be corrected by calibration and such a procedure should be included in the standard method (e.g. a change in the batch of a reagent). An accepted reference value is needed for this, for which ISO Guide 33 and ISO Guide 35 shall be consulted. Remaining equipment effects are considered random.
Time	Effects due to time may be caused by environmental differences, such as changes in room temperature, humidity, etc. Standardization of environmental conditions should be attempted to minimize these effects. Clearly, achieving an ideal degree of standardization would make it appropriate to model the factor *Time* as a fixed effect. However, it is more realistic to model this factor in terms of random effects.

6.1.4 Statistical model

6.1.4.1 Basic model

For the reader's convenience and ease of reference, the basic model described in ISO 5725-1 is reproduced here. For estimating the accuracy (trueness and precision) of a measurement method, it is useful to assume that every test result y is the sum of three components given by Formula (1):

$$y = m + B + e \tag{1}$$

where, for the particular material tested

m is the overall mean (expectation);

B is the laboratory component of bias under repeatability conditions;

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

For a general discussion of these components, the reader is referred to ISO 5725-1, 5.1.

NOTE 1 Depending on the context, m denotes either the theoretical (unknown) overall mean or its estimate. It is possible to use different symbols (e.g. m versus $\hat{m}$) in order to distinguish between a theoretical quantity and its estimate. However, this type of notational nuance seems unnecessary in this document. The same holds for the other symbols used to denote quantities which are to be estimated – though the symbol σ will be reserved for theoretical standard deviations and s for their estimates. The reader is referred to ISO 5725-1 for a discussion of this issue.

NOTE 2 In ISO 5725-4, the bias is further decomposed into two parts: method bias and laboratory bias. While laboratory bias is modelled as a random effect, method bias is modelled as a fixed effect.

6.1.4.2 Partitioning the laboratory bias term

The model described in Formula (1) is appropriate for the situation described in ISO 5725-2, where, within each laboratory, results are obtained under repeatability conditions (i.e. within a short period of time, by the same operator, etc.). Under these conditions, B can be considered constant and is called

the "laboratory component of bias". In practice, however, B arises from a combination of a number of effects. The statistical model as given in Formula (1) can be rewritten in the form given by Formula (2):

$$y = m + B_0 + B_{(1)} + B_{(2)} + \ldots + e \tag{2}$$

where B is partitioned into contributions from variates

B_0 the residual component of the laboratory bias;

$B_{(1)}, B_{(2)}, \ldots$ effects corresponding to intermediate precision factors (such as those in Table 1).

6.1.4.3 Terms B_0, $B_{(1)}$, $B_{(2)}$, etc.

Under repeatability conditions, these terms all remain constant and add to the bias of the test results. Under intermediate precision conditions, B_0 is the effect corresponding to the residual laboratory bias, i.e. it characterizes the background component of laboratory bias which remains invariant as the factors taken into consideration in the design are varied across their respective factor levels. B_0 represents the laboratory-specific bias which cannot be explained by $\boldsymbol{m}$, e, or any of the terms $B_{(1)}, B_{(2)}$, etc. The terms $B_{(1)}, B_{(2)}$, etc. are random effects corresponding to the factors included in the experimental design (the intermediate precision factors). These do not contribute to the residual laboratory bias; rather, they inflate the intermediate precision standard deviation so that it becomes larger than the repeatability standard deviation.

The variance of B is called the between-laboratory variance expressed as:

$$Var(B) = \sigma_L^2$$

In addition to the laboratory component of variance per se (i.e. $Var(B_0)$), this variance also reflects the effects of changes in the intermediate precision factors included in the experimental design. Thus, the quantity $Var(B)$ is composed of independent contributions from the factor laboratory and the intermediate precision factors:

$$Var(B) = Var(B_0) + Var\left(B_{(1)}\right) + Var\left(B_{(2)}\right) + \ldots$$

The variances are denoted:

$Var(B_0) = \sigma_0^2$,

$Var\left(B_{(1)}\right) = \sigma_{(1)}^2$,

$Var\left(B_{(2)}\right) = \sigma_{(2)}^2$, etc.

$Var(B)$ is estimated in practical terms as s_L^2 and corresponding intermediate precision estimates $s_0^2, s_{(1)}^2, s_{(2)}^2$, etc., may be obtained from suitably designed experiments.

6.1.4.4 Error term, e

This term represents a random error occurring in every test result.

Within a single laboratory, its variance is called the within-laboratory variance and is expressed as:

$$Var(e)=\sigma_w^2$$

It may be expected that σ_w^2 will have different values in different laboratories due to differences e.g. between the level of competence of the respective operators or between the sensitivities of different instruments. However, if such differences between laboratories are small, it is justifiable to establish a common value of within-laboratory variance for all the laboratories using the measurement method. This common value, which is estimated by the mean of the within-laboratory variances, is called the "repeatability variance" and is designated by:

$$\sigma_r^2=\overline{Var(e)}$$

This mean value is taken over all the laboratories taking part in the accuracy experiment after exclusion of outliers.

In certain circumstances, it may not be sufficient to establish a common value of within-laboratory variance for all the laboratories using the measurement method. In other words, the question may arise whether the value $\sigma_r^2=\overline{Var(e)}$ as a mean value across all laboratories is representative for each specific laboratory. In order to investigate this question, it is required to implement a design with a relatively large number of test results obtained under repeatability conditions (say, on the order of 10 replicates). This will then make it possible to determine whether the differences between the laboratory-specific estimates for σ_w^2 are systematic or random. Should systematic differences be identified, it may be appropriate to report the estimate for the range $\left[\min\sigma_w^2, \max\sigma_w^2\right]$ or for the ratio $\frac{\max\sigma_w^2}{\min\sigma_w^2}$ instead of a single repeatability value. Such expressions would then be understood as characterizing the range of laboratory-specific repeatability precision values for the method under consideration.

NOTE The error term corresponds to what is often called the residual term. There are different possible approaches to computing the residual term. One approach is based on repeated measurements (replicates). Depending on the required level of reliability and on available resources, replicates can be performed only for one setting, or for several or all settings, and the number of replicates per setting can be increased. However, it is important to note that repeated measurements are not required to obtain an estimate of the residual term. An alternative approach consists in simply computing the residuals themselves, i.e. the differences between the observed values and the values predicted by the model whose parameters have been estimated. Note that in the latter approach, it is not possible to distinguish the residuals per se from differences due to model inadequacies. On the other hand, the former approach is unsatisfactory if the replicates are not truly independent repeated measurements (see Clause 8).

6.2 Within-laboratory study and analysis of intermediate precision measures

6.2.1 Simplest approach

The simplest method of estimating an intermediate precision standard deviation within one laboratory consists of taking one sample (or, for destructive testing, one set of presumably identical samples) and performing a series of n measurements with a change of factor(s) between each measurement. It is recommended that n should be at least 15. This may not be satisfactory for the laboratory, and this method of estimating intermediate precision measures within a laboratory cannot be regarded as efficient when compared with other procedures. The analysis is simple, however, and it can be useful for studying time-different intermediate precision by making successive measurements on the same sample on successive days, or for studying the effects of calibration between measurements.

A graph of $(y_k-\bar{y})$ versus the measurement number k, where y_k is the k^{th} result of n replicated tests and $\bar{y}$ is the mean of the n replicate test results, is recommended to identify potential outliers. A more formal test of outliers consists of the application of Grubbs' test as given in ISO 5725-2:2019, 8.3.5.

The estimate of the intermediate precision standard deviation with M factor(s) different is given by Formula (3):

$$s_{I()} = \sqrt{\frac{1}{n-1}\sum_{k=1}^{n}(y_k - \bar{y})^2} \tag{3}$$

where symbols denoting the intermediate precision conditions should appear inside the empty parentheses.

NOTE Using this simplest design, and if more than one influencing factor has been changed, the intermediate precision standard deviation is estimated taking into account all factors which have been changed, in other words, factors are confounded.

6.2.2 Alternative method

6.2.2.1 Experimental design

An alternative method considers t groups of measurements, each comprising n replicate test results. For example, within one laboratory, a set of t materials could each be measured, then the intermediate precision factor(s) could be altered and the t materials remeasured, the procedure being repeated until there are n test results on each of the t materials. Each group of n test results shall be obtained on one identical sample (or set of presumed identical samples in the case of destructive testing), but it is not essential that the materials be identical. It is only required that the t materials all belong to the interval of test level within which one value of the intermediate precision standard deviation with M factor(s) different can be considered to apply. It is recommended that the value of $t(n-1)$ should be at least 15.

EXAMPLE One operator performs a single measurement on each of the t materials, then this is repeated by a second operator, and possibly by a third operator, and so on, allowing an estimate of $s_{I(0)}$ to be calculated.

6.2.2.2 Analysis

A graph of $(y_{jk} - \bar{y}_j)$ versus the material number, j, where y_{jk} is the k^{th} test result on the j^{th} material and $\bar{y}_j$ is the average of the n replicate test results on the j^{th} material, is recommended to identify potential outliers. A more formal test of outliers consists of the application of Grubbs' test as given in ISO 5725-2:2019, 8.3.5 either for each group separately or for all $t \cdot n$ test results combined.

The estimate of the intermediate precision standard deviation with M factor(s) different is then given by Formula (4):

$$s_{I()} = \sqrt{\frac{1}{t(n-1)}\sum_{j=1}^{t}\sum_{k=1}^{n}(y_{jk} - \bar{y}_j)^2} \tag{4}$$

For $n=2$ (i.e. two test results on each material), the formula simplifies to Formula (5):

$$s_{I()} = \sqrt{\frac{1}{2t}\sum_{j=1}^{t}(y_{j1} - y_{j2})^2} \tag{5}$$

An example of the statistical analysis of an intermediate precision experiment is given in K.1.

6.2.3 Effect of the measurement conditions on the final quoted result

The expectation of $\bar{y}$ is different between one combination and another of time, calibration, operator and equipment, even when only one of the four factors changes. This is a limitation on the usefulness of mean values. In chemical analysis or physical testing, $\bar{y}$ is reported as the final quoted result. In trading

raw materials, this final quoted result is often used for quality evaluation of the raw materials and affects the price of the product to a considerable extent.

EXAMPLE In the international trading of coal, the size of the consignment can often exceed 70 000 t, and the ash content is determined finally on a test portion of only 1 g. In a contract stipulating that each difference of 1 % in ash content corresponds to USD 1,5 per tonne of coal, a difference of 1 mg in the weighing of ash by a chemical balance corresponds to 0,1 % in ash content, or USD 0,15 per tonne, which for such a consignment amounts to a difference in proceeds of USD 10 500 (from 0,1 × 1,5 × 70 000). Consequently, the final quoted result of chemical analysis or physical testing should be sufficiently precise, highly reliable and, especially, universal and reproducible. A final quoted result which can be guaranteed only under conditions of a specific operator, equipment or time can be inadequate for commercial considerations.

7 Nested design

7.1 Balanced fully-nested design

A schematic layout of the balanced fully-nested experiment at a particular level of the test is given in Table 3 and Table 4.

By carrying out the two-factor balanced fully-nested experiment collaboratively in several laboratories, one intermediate precision measure can be obtained at the same time as the repeatability and reproducibility standard deviations, i.e. σ_0, $\sigma_{(1)}$ and σ_r can be estimated. Likewise, the three-factor balanced fully-nested experiment can be used to obtain two intermediate precision measures, i.e. σ_0, $\sigma_{(1)}$, $\sigma_{(2)}$ and σ_r can be estimated.

The subscripts i, j and k suffixed to the data y in Table 3 for the two-factor balanced fully-nested experiment represent, for example, a laboratory and a day of experiment (i.e. the two factors) along with the replicate under repeatability conditions, respectively.

The subscripts i, j, k and l suffixed to the data y in Table 4 for the three-factor balanced fully-nested experiment represent, for example, a laboratory, a day of experiment and an operator (i.e. the three factors) along with the replicate under repeatability conditions, respectively.

Table 3 — Schematic layout for two-factor balanced fully-nested design

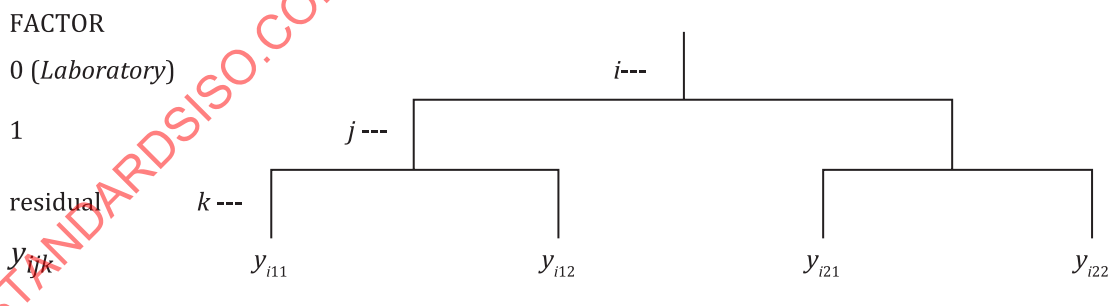
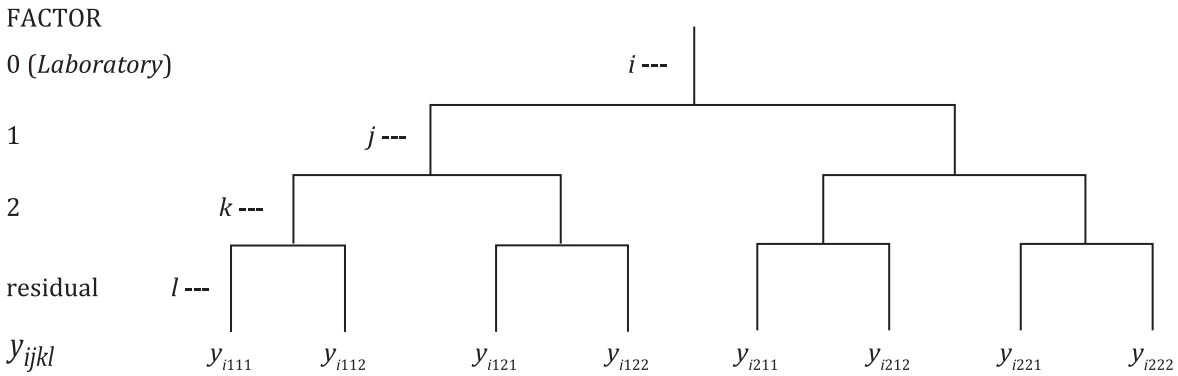


Table 4 — Schematic layout for three-factor balanced fully-nested design



NOTE 1 The factor index on the left corresponds to the subscript of the terms B_0, $B_{(1)}$, $B_{(2)}$ etc. from 6.1.4.2 and 6.1.4.3.

NOTE 2 In a design with t factors, there are $t+1$ hierarchical levels (i.e. $t+1$ ranks and subscripts). This is because the last level corresponds to replicate measurements – and multiple determination under repeatability conditions is not a factor. In this respect, there is a difference to the version of ISO 5725-3:1994, where multiple determination under repeatability conditions was counted as a factor. Accordingly, the two designs shown in the tables are referred to as three-factor and four-factor nested designs in the ISO 5725-3:1994 version.

NOTE 3 In this document, the first (highest) factor is always laboratory. Since this factor can be considered to be implicit in every experimental design in this standard, it is numbered 0, and the intermediate precision factors are numbered starting with $1,2,\ldots$

NOTE 4 The t -factor design is also known as a t -stage design.

As explained in Annex B, the balanced fully-nested design is appropriate when there is a natural *hierarchy* among all factors. The allocation of the factors in a nested design is arranged so that the factors lending themselves most to modelling in terms of fixed effects should be in the highest ranks (0,1,...), and those lending themselves most to modelling in terms of random effects should be in the lowest ranks. For example, in a three-factor design such as illustrated in Table 4 and Table 5, factor 0 could be the laboratory, factor 1 the operator and factor 2 the day on which the measurement is carried out.

Since the analysis is carried out separately for each level of the test (material), the procedure described in ISO 5725-2 is, in fact, a one-factor (laboratory) nested experimental design and produces two standard deviations, the repeatability and reproducibility standard deviations. If this design is increased by one factor, by having two operators in each laboratory, each obtaining two test results under repeatability conditions, then, in addition to the repeatability and reproducibility standard deviations, the intermediate precision standard deviation corresponding to the factor *Operator* can be determined. Alternatively, if there is only one operator in each laboratory, but the experiment is repeated on another day, it is possible to determine the intermediate precision standard deviation corresponding to the factor day. The addition of a further factor to the experiment (i.e. a three-factor design) makes it possible to obtain two intermediate precision standard deviations. For instance, if there are two operators in each laboratory, each performing two measurements, and the entire experiment is repeated the next day, it is possible to determine standard deviations corresponding to repeatability, reproducibility, the factor operator and the factor day.

In the nested design, the variability associated with intermediate factors is considered to remain constant across the levels of higher-ranked factors. For instance, it is assumed that the equipment variability is constant across laboratories and operators. If there are doubts as to the appropriateness of this assumption, then the design shall be modified.

The conditions under which each participating laboratory is required to perform measurements on the test items should be clearly described in the collaborative study announcement letter. Each selected factor used for the estimation of the intermediate precision should be specified. This will enable laboratories to determine whether they are eligible to participate.

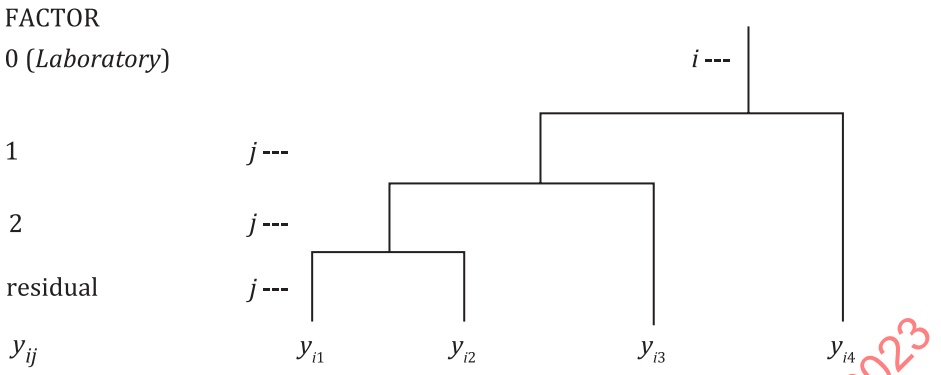
The analysis of the results of a balanced fully-nested experiment should be carried out by means of the statistical method Restricted Maximum Likelihood (REML). This method is described in general terms in Annex J. The calculation of the REML requires mathematical-statistical software. If this is not available, the statistical method Analysis of Variance (ANOVA) can be used instead. This is described in Annex C. For balanced data, REML and ANOVA estimation of variance components yield the same results (apart from the fact that the REML estimate can never be below zero).

7.2 Staggered-nested design

The staggered-nested design allows a considerable reduction in the workload. The drawback is increased complexity in the statistical analysis and larger uncertainties.

A schematic layout of the staggered-nested design at a particular level of the test is provided in Table 5.

Table 5 — Schematic layout of the three-factor staggered-nested design



The two-factor staggered-nested design requires each laboratory i to obtain three test results. Test results y_{i1} and y_{i2} are obtained under repeatability conditions, and y_{i3} under intermediate precision conditions, e.g. by obtaining y_{i3} on a different day from that on which y_{i1} and y_{i2} were obtained. An example of the statistical analysis of a two-factor staggered-nested design is given in J.2.

In a three-factor staggered-nested design, y_{i4} is obtained under intermediate precision conditions by varying two factors, e.g. by changing the day and the operator.

Just as in the case of the balanced fully-nested design, the allocation of the factors in a staggered-nested design is arranged in accordance with the natural hierarchy. For example, in a three-factor staggered-nested design such as illustrated in Table 5, factor 0 could be the laboratory, factor 1 the operator and factor 2 the day on which the measurement is carried out.

The t -factor balanced fully-nested design with two levels per factor (apart from the factor laboratory) and two replicates per setting requires $2 \cdot 2^{t-1}$ test results from each laboratory, which can be an excessive requirement on the laboratories. This is the main argument for the staggered-nested design. The latter design requires fewer test results to produce the same number of standard deviations, although the analysis is slightly more complex and there is a larger uncertainty in the estimates of the standard deviations due to the smaller number of test results.

NOTE 1 The staggered-nested design discussed here is fully-nested but not balanced.

The analysis of the results of a staggered-nested design should be carried out by means of the statistical method Restricted Maximum Likelihood (REML). This is described in Annex I. The calculation of the REML requires mathematical-statistical software. If this is not available, the statistical method Analysis of Variance (ANOVA) can be used instead. This is described in Annex C.

NOTE 2 Further details on the staggered-nested design can be found in References [5] and [6].

7.3 Balanced partially-nested design

A schematic layout of a balanced partially-nested experiment with three factors (laboratory, operator, reagent batch) at a particular level of the test and for laboratory, *i*, is provided in Table 6. Each of the two factors *Operator* and *Reagent batch* has two levels, denoted 1 and 2.

The structure of the design is identical to the balanced fully-nested experiment and could also be described by Table 4. However, unlike the nested design, in the paially-nested design there is no complete hierarchy among the factors: all factors (apart from laboratory) are subsumed under the factor laboratory, and among these remaining factors there is no hierarchy. A detailed discussion of the differences between the different designs is found in Annex A.

Table 6 — Design for partially-nested design with three factors at a particular level of the test and for laboratory *i*

Operator	Reagent batch	Test result[a]
1	1	y_{i111} , y_{i112}
1	2	y_{i121} , y_{i122}
2	1	y_{i211} , y_{i212}
2	2	y_{i221} , y_{i222}

[a] The first three subscripts in $y_{i\ldots.}$, $y_{i\ldots.2}$ represent the levels of the three factors (laboratory, operator, reagent batch)

The reproducibility variance can be obtained by summing the precision variances corresponding to the factors and the residual variance. The latter may be interpreted as the repeatability variance.

The analysis of the results of a balanced partially-nested experiment is carried out by the statistical method Analysis of Variance (ANOVA) separately for each level of the test as described in Annex D. Alternatively, the statistical technique Restricted Maximum Likelihood (REML) can be used. This is described in Annex I.

7.4 Orthogonal array design

Consider the case that p laboratories take part in a collaborative study in which, in addition to the factor Laboratory, four factors (denoted A, B, C and D) are investigated with respect to contribution to overall variability. Within each laboratory, each of the factors has two levels (for instance, two operators, two instruments, two reagent batches, two batches of cartridges), there are two replicates per setting, and there is no complete hierarchy among the factors. In the balanced partially-nested design, each laboratory would have to carry out 32 measurements per level (see Table 7). More generally, in the balanced partially-nested design, if there are $(t-1$ factors within each laboratory, and if each factor (apart from laboratory) has 2 levels, each laboratory shall perform $n \cdot 2^{(t-1)}$ measurements per level and measurand, where n denotes the number of replicates.

Table 7 — Five-factor (laboratory, and factors A, B, C and D) balanced partially-nested design at level *i* of the factor laboratory (with 2 replicates)

Factor A	Factor B	Factor C	Factor D	Test result[a]
1	1	1	1	y_{i11111}, y_{i11112}
1	1	1	2	y_{i11121}, y_{i11122}
1	1	2	1	y_{i11211}, y_{i11212}
1	1	2	2	y_{i12211}, y_{i11222}
1	2	1	1	y_{i12111}, y_{i12112}
1	2	1	2	y_{i12121}, y_{i12122}
1	2	2	1	y_{i12211}, y_{i12212}
1	2	2	2	y_{i12221}, y_{i12222}
2	1	1	1	y_{i21111}, y_{i21112}
2	1	1	2	y_{i21121}, y_{i21122}
2	1	2	1	y_{i21211}, y_{i21212}

[a] The first five subscripts in $y_{i\ldots.}$, $y_{i\ldots.2}$ represent the levels of the five factors.

Table 7 *(continued)*

Factor A	Factor B	Factor C	Factor D	Test result[a]
2	1	2	2	y_{i21221}, y_{i21222}
2	2	1	1	y_{i22111}, y_{i22112}
2	2	1	2	y_{i22121}, y_{i22122}
2	2	2	1	y_{i22211}, y_{i22212}
2	2	2	2	y_{i22221}, y_{i22222}

[a] The first five subscripts in $y_{i....}$, $y_{i....2}$ represent the levels of the five factors.

Special designs called orthogonal array designs make it possible to considerably reduce the workload associated with a particular design without excessive loss of reliability. An orthogonal array design corresponding to the above example is provided in Table 8. As can be seen, the workload has been reduced by 50 %.

Table 8 — Orthogonal array design for the five-factor (laboratory, and factors A, B, C and D) experiment at level *i* of the factor laboratory (with 2 replicates)

Factor A	Factor B	Factor C	Factor D	Test result[a]
1	1	1	1	y_{i11111}, y_{i11112}
1	1	2	2	y_{i11221}, y_{i11222}
1	2	1	2	y_{i12121}, y_{i12122}
1	2	2	1	y_{i12211}, y_{i12212}
2	1	1	2	y_{i21121}, y_{i21122}
2	1	2	1	y_{i21211}, y_{i21212}
2	2	1	1	y_{i22111}, y_{i22112}
2	2	2	2	y_{i22221}, y_{i22222}

[a] The first five subscripts in $y_{i....}$, $y_{i....2}$ represent the levels of the five factors.

As can be seen, an orthogonal array design is obtained by omitting some of the settings (i.e. rows in Table 7). However, the reliability of the precision estimations obtained on the basis of the reduced design depends on correctly determining which rows to omit.

Orthogonal array designs are characterized by the following property: the number of occurrences of each factor level combination is the same for every pair of factors. In other words, for each factor pair, the number of measurements in a laboratory is the same for each factor level combination of the two factors.

Thus, in the example shown in Table 8, the factor level combinations corresponding to any factor pair are 1-1, 1-2, 2-1 and 2-2. For the factor pair Factor A-Factor B, it easily verified that each factor level combination occurs twice (indeed in the same order as in the previous sentence: 1-1, 1-1, 1-2, 1-2, 2-1, 2-1, 2-2, 2-2). For the factor pair Factor A-Factor C, the factor level combinations occur in a different order; however, it is easily verified (see Table 9) that each occurs twice.

Table 9 — Factor level combination occurrences for factor pair Factor A-Factor C

Factor A	Factor C	Factor level combination
1	1	1-1
1	2	1-2
1	1	1-1
1	2	1-2
2	1	2-1
2	2	2-2
2	1	2-1
2	2	2-2

Similarly, it can be verified that each factor level combination occurs twice for each of the four remaining factor pairs (A-D, B-C, B-D, C-D).

The orthogonal array design discussed here is only an example. Depending on the number of factors and the maximum number of runs one is willing to perform, different orthogonal array designs will be obtained. However, they all share the property just described.

Further reduction of workload can be achieved if 2 replicates are carried out only for one or two settings. In this case, only 9-10 measurements would have to be carried out per laboratory. Another approach consists in abstaining from replicates and determining the repeatability standard deviation by the residual standard deviation obtained from a restricted model. The analysis of the results of an orthogonal partially-nested experiment should be carried out by means of the statistical method Restricted Maximum Likelihood (REML). This is described in Annex I. The calculation of the REML requires mathematical-statistical software. If this is not available, the statistical method Analysis of Variance (ANOVA) may be used instead. This is described in Annex C.

NOTE 1 Orthogonal array designs can be derived e.g. from fractional factorial designs, latin squares and Hadamard matrices.

NOTE 2 The presented designs can be extended or adapted to specific requirements. A detailed discussion of efficient validation designs can be found in Reference [7]. An overview of factorial designs can be found e.g. in Reference [8].

8 Design for heterogeneous material

8.1 Applications of the design for a heterogeneous material

An example of a heterogeneous material is sand (that might be used, for example, for making concrete). This is laid down by the action of wind or water, in strata that always contain graduations in particle size, so when sand is used the particle size distribution is always of interest. In concrete technology the particle size distribution of sand is measured by sieve testing. In order to test a sand product, a bulk sample is taken from the product, then one or more test portions are produced from the bulk sample. Typically, the bulk sample will be about 10 kg in mass, and the test portions will be about 200 g. Because of the natural variability of the material, there will always be some variability between bulk samples of the same product. Hence, if a uniform level experiment is performed in which each laboratory is sent one bulk sample at each level, the variability between the bulk samples will increase the calculated reproducibility standard deviation of the test method. However, if laboratories are sent two bulk samples at each level, then values for the reproducibility standard deviation can be calculated that exclude this variation.

This example highlights another characteristic of heterogeneous materials: because of the variability of the material, the specimen or test portion preparation can be an important source of variation. Thus, with sieve tests on sand the process of preparing test portions from bulk samples is usually the major source of variability in the test method. If specimens or test portions are prepared for a precision

experiment in a way that does not correspond to normal practice (in an attempt to produce identical "samples") then the values of repeatability and reproducibility standard deviations produced by the experiment will not be representative of the variability experienced in practice. There are situations in which it can be desirable to produce identical "samples" by some special process designed to eliminate, as far as possible, the variability of the material (for example, for a proficiency test, or when a precision experiment is used as part of a program of work during the development of a measurement method). However, when the aim of the precision experiment is to discover the variability that will be experienced in practice (for example, when vendors and purchasers test samples of the same product) then it is necessary for the variability arising as a consequence of the heterogeneity of the material to be included in the measures of the precision of the measurement method.

Care should also be taken to ensure that each test result in an experiment is obtained by carrying out the test procedure independently of other tests. This will not be so if some stages of the specimen preparation are shared by several specimens, so that a bias or deviation introduced by the preparation will have a common influence on the test results derived from these specimens.

The design for heterogeneous materials proposed in this clause yields information about the variability between samples that is not obtainable from the uniform level design described in ISO 5725-2. There is, inevitably, a cost associated with obtaining extra information: the proposed design requires more samples to be tested. This extra information may be valuable. In the said example, information about the variability between bulk samples could be used to decide if the procedure for taking bulk samples is satisfactory or in need of improvement.

The design described in this clause is applicable to experiments involving two factors arranged in a hierarchy (laboratory and sample within laboratory) and replicate measurements (test results within sample).

8.2 Layout of the design for a heterogeneous material

The layout of the design for a heterogeneous material is shown in Table 10.

Table 10 — Design for heterogeneous material at a particular level of the test and for laboratory *i*

Sample	Test result
1	y_{i11} , y_{i12}
2	y_{i21} , y_{i22}

The p participating laboratories are each provided with two samples at q levels, and obtain two test results on each sample. Thus each laboratory reports four test results per level (two test results for each of two samples).

8.3 Statistical analysis

The statistical model for the experiment with a heterogeneous material is the same as for the balanced fully-nested design with two factors: laboratory, sample. The reader is referred to Annex B and Annex E. Let it be noted here that the variance corresponding to the between-sample variability is subsequently subtracted from the reproducibility variance in order to obtain a corrected reproducibility precision estimate.

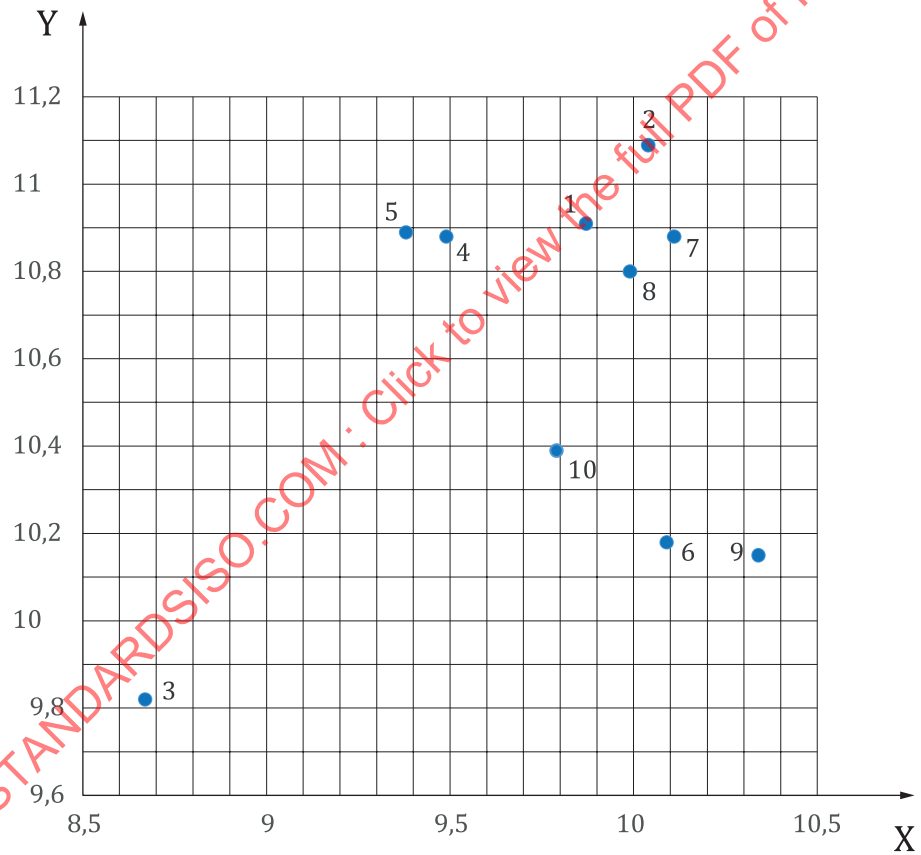
9 Split-level design

9.1 Applications of the split-level design

The uniform level design described in ISO 5725-2 requires two or more identical samples of a material to be tested in each participating laboratory and at each level of the experiment. With this design there is a risk that an operator may allow the result of a measurement on one sample to influence the result of

a subsequent measurement on another sample from the same material. If this happens, the results of the precision experiment will be distorted: the repeatability standard deviation σ_r will be underestimated and the between-laboratory standard deviation σ_L will be overestimated. In the split-level design, at each level of the experiment, each participating laboratory is provided with two samples corresponding to two similar materials, and the operators are told that the samples are not identical, but they are not told by how much the materials differ. The split-level design thus provides a method of determining the repeatability and reproducibility standard deviations of a standard measurement method in a way that reduces the risk that a test result obtained on one sample will influence a test result on another sample in the experiment.

For each level, the data obtained in a split-level experiment may be used to draw a graph (Youden plot) in which the data for one material are plotted against the data for the other, similar, material. An example is given in Figure 1. The data are taken from Table G.1, level 1. Each point in the Youden plot (displayed here as numbers) corresponds to a laboratory. If the two results lie close to each other, the point should lie on the main diagonal (i.e. the line joining the bottom-left an top-right corners of the plot). Points lying far from the main diagonal are associated with relatively large random error. Points lying far from the center of the plot along with the main diagonal show a *systematic* bias (i.e. a bias across both results). The Youden plot is thus useful in identifying the causes of laboratory errors with the aim of taking corrective action.



Key

X measurement result a

Y measurement result b

Figure 1 — Youden plot for the split-level design (see Annex G, level 1)

It is common for the repeatability and reproducibility standard deviations of a measurement method to depend on the level of the material. For example, when the test result is the proportion of an element obtained by chemical analysis, the repeatability and reproducibility standard deviations usually increase as the proportion of the element increases. It is necessary, for a split-level experiment, that the

two similar materials used at a level of the experiment are sufficiently similar in the sense that the same repeatability and reproducibility standard deviations can be expected. For the purposes of the split-level design, it is acceptable if the two materials used for a level of the experiment give almost the same level of measurement results, and nothing is to be gained by arranging that they differ substantially.

In many chemical analysis methods, the matrix containing the constituent of interest can influence the precision, so for a split-level experiment two materials with similar matrices are required at each level of the experiment. A sufficiently similar material can sometimes be prepared by spiking a material with a small addition of the constituent of interest. When the material is a natural or manufactured product, it can be difficult to find two products that are sufficiently similar for the purposes of a split-level experiment: a possible solution may be to use two batches of the same product. It should be remembered that the object of choosing the materials for the split-level design is to provide the operators with samples that they do not expect to be identical.

9.2 Layout of the split-level design

The layout of the split-level design is shown in Table 11.

Table 11 — Split-level design at a particular level of the test and for laboratory *i*

Material	Test result
a	y_{ia}
b	y_{ib}

For each level, the p participating laboratories each test two samples.

The two samples within a level are denoted a and b, where a represents a sample of one material, and b represents a sample of the other, similar, material.

9.3 Statistical analysis

The statistical analysis of the split-level experiment is described in Annex F. Let it be noted here that the model is a mixed model, i.e. a model with both random and fixed effects (more specifically: two random effects corresponding to the factors laboratory and test result, and one fixed effect for the factor material). Let it also be noted that the fixed effect (material) is a nuisance effect in the sense that it has no intrinsic interest. Rather, its inclusion is considered necessary in order to obtain undistorted estimates of repeatability and reproducibility.

The analysis of the results of a split-level experiment should be carried out by means of the statistical technique restricted maximum likelihood (REML). This is described in Annex I. The calculation of the REML requires mathematical-statistical software. If this is not available, the statistical technique analysis of variance (ANOVA) can be used instead (see Annex F). An example is provided in Annex G.

10 Design across levels

10.1 Applications of the design across levels

In the split-level design presented in Clause 9 and Annex F, two different (though similar) materials are tested. While the statistical model assigns a separate mean for each of the two materials, the laboratory bias is modelled as independent of the material. The rationale for this model is that, for a properly standardized method, two “sufficiently similar” materials used at a level of the experiment can be expected to give the same repeatability and reproducibility standard deviations. Indeed, this requirement constitutes the working criterion with respect to assessing whether the degree of similarity between the two materials is sufficient. If this requirement is not met, then the split-level design cannot be applied.

In such circumstances, questions will naturally arise regarding the relationship between the laboratory bias and the material tested. More generally, it is sometimes desirable to investigate this relationship within the scope of the validation. Indeed, it is recommended to include different levels in the validation study as long as an appropriate statistical model for the estimation of variance components is available. In particular, an appropriate model would reduce to the model in ISO 5725-2 if only one level is investigated.

This clause provides a design which allows the inclusion of different levels in the validation study.

10.2 Layout of the design across levels

The layout of the design across levels is shown in Table 12.

Table 12 — Design across levels for laboratory, *i*

Level	Test result
1	y_{i11} , y_{i12}
2	y_{i21} , y_{i22}
3	y_{i31} , y_{i32}
4	y_{i41} , y_{i42}
...	...

The layout provided here can be applied to investigate the relationship between laboratory bias and level. However, with minor modifications, the same design can be applied in order to investigate the relationship between laboratory bias and different materials within one particular level.

10.3 Statistical analysis

The statistical model for the experiment is described in Annex H. Let it be noted that the model is a mixed model, i.e. a model with both random and fixed effects (more specifically: three random effects corresponding to the factor laboratory, the interaction between the factors laboratory and level (i.e. the level-specific effect of the laboratory) and to the residual error, and one fixed effect for the factor level). Let it also be noted that the fixed effect (level) is a nuisance effect in the sense that it has no intrinsic interest.

Once the statistical analysis has been performed, it will then be possible to compare the between-laboratory variance corresponding to the laboratory biases across levels to the between-laboratory variance within a particular level. This makes it possible to determine whether the laboratory bias is dominated by level-specific effects. If this is the case, then the methods described in ISO 5725-2:2019, 8.5 can be applied to determine the relationship between level and precision.

NOTE The statistical model is described in Annex H and an example is described in Annex K.

11 Reliability of interlaboratory parameters

11.1 Reliability of precision estimates

Alternative designs allow a more efficient estimation of precision parameters. This means that the same level of reliability can be achieved with fewer participating laboratories. The reliability is assessed by the relative standard error of the precision estimate.

In order to assess the reliability of the reproducibility precision estimate, a simulation can be run on the basis of resampling (bootstrapping) from the available measured data (say, 1 000 resamples). Note that this procedure assumes that, for each resample, data are normally distributed. Alternatively, a Monte Carlo simulation can be run, whereby the parameters for the underlying distributions are estimated from the available measured data (again, on the order of 1 000 simulation runs). For each simulation

run or each resample, a precision estimate is then obtained. A confidence interval for the theoretical reproducibility can then be obtained by means of the 2,5 % and the 97,5 % percentiles of the resulting distribution of precision estimates. All necessary computations are easily implemented in standard statistical software.

NOTE 1 Software capable for this kind of calculation includes R[9] and JASP[10].

NOTE 2 Reliability is used here not in the sense of IEC 60050(192) but in the more general understanding of a method that consistently produces the same result.

11.2 Reliability of estimates of the overall mean

11.2.1 General

Alternative designs also allow an estimation of the overall mean, e.g. for the determination of trueness. In order to assess the reliability of the estimate of the overall mean, the formulae for the standard errors are provided in the following for the different models. These formulae are based on the assumption that two replicate measurements were performed per setting.

11.2.2 Balanced fully-nested design (2 factors)

The standard error of the overall mean is $se(\hat{m}) = \sqrt{\frac{\sigma_0^2}{p} + \frac{\sigma_{(1)}^2}{2p} + \frac{\sigma_r^2}{4p}}$ (see B.1).

11.2.3 Staggered nested design (2 factors)

The standard error of the overall mean is $se(\hat{m}) = \sqrt{\frac{\sigma_0^2}{p} + \frac{5\sigma_{(1)}^2}{9p} + \frac{\sigma_r^2}{3p}}$ (see C.1).

11.2.4 Balanced partially-nested design

The standard error of the overall mean is $se(\hat{m}) = \sqrt{\frac{\sigma_0^2}{p} + \frac{\sigma_{(1)}^2}{2p} + \frac{\sigma_{(2)}^2}{2p} + \frac{\sigma_r^2}{8p}}$ (see Annex D, the interaction term is omitted here).

11.2.5 Orthogonal array design

In the case of the orthogonal array design with five factors (laboratory and factors A, B, C and D subsumed under laboratory; see 74), the standard error of the overall mean (omitting interaction terms) is:

$$se(\hat{m}) = \sqrt{\frac{\sigma_0^2}{p} + \frac{\sigma_A^2}{2p} + \frac{\sigma_B^2}{2p} + \frac{\sigma_C^2}{2p} + \frac{\sigma_D^2}{2p} + \frac{\sigma_r^2}{16p}}.$$

It should be noted that the formula provided here is only appropriate for the orthogonal array design discussed in 7.4. Different formulae must be derived for the standard errors of different orthogonal array designs.

NOTE In order to compare the reliability of the orthogonal array design and the basic design from ISO 5725-2, consider the case that 6 laboratories implement the orthogonal array design presented in 7.4 with one measurement per run instead of two (i.e. eight measurements per laboratory). This case will be compared to the basic design from ISO 5725-2 with nine laboratories and four replicates. The corresponding standard errors will now be computed and compared on the following basis:

Orthogonal array design:		Basic design from ISO 5725-2:
$\sigma_0^2 + \sigma_A^2 + \sigma_B^2 + \sigma_C^2 + \sigma_D^2$	→	σ_L^2
σ_r^2	→	σ_r^2

For the variance components

$$\sigma_0^2 = 2$$

$$\sigma_A^2 = \sigma_B^2 = \sigma_C^2 = \sigma_D^2 = 1$$

$$\sigma_r^2 = 2$$

the following standard error values are obtained for the two designs:

Orthogonal array design ($p = 6$):

$$se(\hat{m}) = \sqrt{\frac{\sigma_0^2}{p} + \frac{\sigma_A^2}{2p} + \frac{\sigma_B^2}{2p} + \frac{\sigma_C^2}{2p} + \frac{\sigma_D^2}{2p} + \frac{\sigma_r^2}{8p}} = 0,84$$

Basic design from ISO 5725-2 ($p = 9$):

$$se(\hat{m}) = \sqrt{\frac{\sigma_L^2}{p} + \frac{\sigma_r^2}{4 \cdot p}} = 0,85$$

As can be seen, the reliability of the orthogonal array design is slightly better than that of the basic design from ISO 5725-2 with only two thirds the number of laboratories.

11.2.6 Split-level design

The standard error of the overall mean is $se(\hat{m}) = \sqrt{\frac{\sigma_0^2}{p} + \frac{\sigma_r^2}{2p}}$ (see Annex F).

Annex A
(informative)

Fully- and partially-nested designs

Factors can be related to one another in different ways. One fundamental consideration in this regard is the distinction between the fully and partially nested designs.

Both the balanced fully-nested design and the staggered-nested design are fully-nested designs. Fully-nested designs are appropriate when there is a nesting hierarchy among all factors. For instance, consider the two factors operator and calibration. If each operator performs his or her own calibration, then the factor operator can be considered to be ranked higher than the factor calibration. With the same argumentation it follows that in collaborative studies generally the factor laboratory has the highest rank.

A crucial observation is that, in the presence of a hierarchy among the factors, lower-ranked factors do not behave in the same way across factor levels of higher-ranked factors. This implies that it is not appropriate to average across factor levels of higher-ranked factors. For instance, take the case that there are two operators and two calibrations per operator in each laboratory. For each operator, the two calibrations are denoted `C1` and `C2`. Clearly, it is not permissible to average the results corresponding to `C1` across the two operators. The reason is very simple: `C1` is actually a different calibration for each of the two operators. In other words, it would be more accurate to denote the calibrations for the second operator `C3` and `C4`; in this manner, one would obtain four different factor levels for the factor calibration, and it would be clear that averaging across operators for a particular calibration is meaningless.

In a fully-nested design, there is also a hierarchy among the variabilities associated with the different factors. In the case discussed above the variabilities are ranked as follows: variability between different laboratories – variability between different operators within the laboratory – variability between different calibrations for one operator.

In the presence of a complete hierarchy among the factors, it is necessary to apply a fully-nested design. By contrast, if all factors (apart from laboratory) are subsumed under the factor laboratory, and in the absence of a hierarchy among these factors subsumed under laboratory, then a fully-nested design is not appropriate and a different design shall be applied: the partially-nested design. For instance, take the case that the number of factor levels is insufficient for the implementation of a fully-nested design, e.g. there are only two instruments available in each laboratory, and each operator conducts analyses on both instruments. It would then be appropriate to compute, for each laboratory, the average for each operator across instruments. Conversely, it would also be possible to compute, for each laboratory, the average for each instrument across operators.

In a partially-nested experimental design, all factors are nested within the factor laboratory, but there is at least one pair of factors where there is no nesting hierarchy: such a situation exists, for example, for the two factors operator and instrument: averaging the results over the two operators for each of the two instruments provides information about systematic differences between the two instruments, and similarly, averaging the results over the two instruments for each of the two operators provides information about systematic differences between the two operators. Thus, there is generally no hierarchy between operator and instrument.

All designs described in this standard that are not fully-nested are considered partially-nested designs.

NOTE If statistically significant systematic effects are observed between instruments in different laboratories, correction for measurement method bias is usually possible. For this reason, in collaborative studies, all factors can be subsumed to and nested within the laboratory factor laboratory.

Finally, it should be noted here that for both the fully-nested and the partially-nested design, there exists a possibility of reducing the workload involved. For the fully-nested design, such a workload reduction is achieved by means of the staggered design. For the partially-nested design, the workload can be reduced by means of orthogonal array designs.

Annex B
(informative)

Analysis of variance for balanced fully-nested design

B.1 Two-factor balanced fully-nested design

The analysis of variance described here shall be carried out separately for each level of the test included in the collaborative study. The formulae provided are only applicable for the case that two replicates are available per setting and laboratory. For the sake of readability, the subscript indicating the level of the test has been suppressed.

NOTE The subscript *j* is used in this document for factor 1 (factor 0 being the laboratory), while in the other parts of ISO 5725 it is used for the level of the test.

The exact analysis of the data can be very complicated when some of the test results from a laboratory are missing. If it is decided that some of the test results from a laboratory are stragglers or outliers and should be excluded from the analysis, then it is recommended that all the data from that laboratory (at the level affected) should be excluded from the analysis. Alternatively, the statistical technique REML (Annex I) can be applied.

The outlier tests and the computation of the precision estimates for the two-factor balanced fully-nested experiment will now be described.

For each level, the data obtained in the experiment are denoted y_{ijk} (with i representing factor 0, i.e. laboratory, $i=1,\ldots,p$; j representing factor 1, $j=1,2$; and k representing the replicate, $k=1,2$).

The basic model used in this document is described in 6.1.4 as Formula (1). For a balanced nested design with two factors, this model is expanded to

$$y_{ijk} = m + B_i + B_{ij} + e_{ijk} \tag{B.1}$$

The term B_i represents the effect corresponding to factor 0 (laboratory) with variance σ_0^2.

The term B_{ij} represents the effect corresponding to factor 1 with variance $\sigma_{(1)}^2$.

The term e_{ijk} represents the residual or repeatability effect with variance σ_r^2.

In a first step, the mean values are computed as follows:

$$\bar{y}_{ij} = \frac{1}{2}\left(y_{ij1} + y_{ij2}\right)$$

$$\bar{y}_i = \frac{1}{2}\left(\bar{y}_{i1} + \bar{y}_{i2}\right)$$

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

where p denotes the number of laboratories which have participated in the collaborative study.

The total sum of squares, SST, can be partitioned as

$$SST = \sum_i \sum_j \sum_k \left(y_{ijk} - \bar{\bar{y}}\right)^2 = SS0 + SS1 + SSe$$

where

$$SS0 = \sum_i \sum_j \sum_k \left(\bar{y}_i - \bar{\bar{y}}\right)^2 = 4\sum_i \left(\bar{y}_i - \bar{\bar{y}}\right)^2$$

$$SS1 = \sum_i \sum_j \sum_k \left(\bar{y}_{ij} - \bar{y}_i\right)^2 = 2\sum_i \sum_j \left(\bar{y}_{ij} - \bar{y}_i\right)^2 = 2\sum_i s_{iA}^2$$

$$SSe = \sum_i \sum_j \sum_k \left(y_{ijk} - \bar{y}_{ij}\right)^2 = \sum_i \sum_j s_{ij}^2 = 2\sum_i s_{iB}^2$$

The methods described in ISO 5725-2:2019, 8.3 should be applied to check the data for consistency and outliers.

The Cochran test should be applied to check

the variances s_{ij}^2, $i = 1,\ldots,p$, $j = 1,2$, for outliers within the replicates

the variances s_{iB}^2, $i = 1,\ldots,p$, for repeatability outliers

the variances s_{iA}^2, $i = 1,\ldots,p$, for intermediate precision outliers

The Grubbs test should be applied to check the laboratory mean values $\bar{y}_i$, $i = 1,\ldots,p$ for outliers.

Since the degrees of freedom for the sums of squares $SS0$, $SS1$ and SSe are $p-1$, p and $2p$, respectively, the ANOVA table is composed as shown in Table B.1.

Table B.1 — ANOVA table for a two-factor balanced fully-nested experiment

Source	Sum of Squares	Degrees of freedom	Mean square	Expected mean square
0	$SS0$	$p-1$	$MS0 = SS0/(p-1)$	$\sigma_r^2 + 2\sigma_{(1)}^2 + 4\sigma_0^2$
1	$SS1$	p	$MS1 = SS1/p$	$\sigma_r^2 + 2\sigma_{(1)}^2$
Residual	SSe	$2p$	$MSe = SSe/(2p)$	σ_r^2
Total	SST	$4p-1$		

The unbiased estimates $s_0^2, s_{(1)}^2$ and s_r^2 for $\sigma_0^2, \sigma_{(1)}^2$ and σ_r^2, respectively, can be obtained from the mean squares $MS0$, $MS1$ and MSe as

$$s_0^2 = \frac{1}{4}(MS0 - MS1)$$

$$s_{(1)}^2 = \frac{1}{2}(MS1 - MSe)$$

$$s_r^2 = MSe\,.$$

The estimates of the repeatability, intermediate and reproducibility variances are, respectively, as follows:

$$s_r^2$$

$$s_{I(1)}^2 = s_r^2 + s_{(1)}^2$$

$$s_R^2 = s_r^2 + s_{(1)}^2 + s_0^2\,.$$

NOTE Strictly speaking, the outlier tests given here and in ISO 5725-2 are valid under the assumption that the underlying distributions are approximately normal; in practice, they work for most distributions provided they are unimodal.

B.2 Three-factor balanced fully-nested design

The analysis of variance described here shall be carried out separately for each level of the test included in the collaborative study. The formulae provided are only applicable for the case that two replicates are available per setting and laboratory. For the sake of readability, the subscript indicating the level of the test has been suppressed.

NOTE The subscript *j* is used in this document for factor 1 and factor 2 (factor 0 being the laboratory), while in the other parts of ISO 5725 it is used for the level of the test.

The exact analysis of the data can be very complicated when some of the test results from a laboratory are missing. If it is decided that some of the test results from a laboratory are stragglers or outliers and should be excluded from the analysis, then it is recommended that all the data from that laboratory (at the level affected) should be excluded from the analysis. Alternatively, the statistical technique REML (Annex J) can be applied.

The outlier tests and the computation of the precision estimates for the three-factor balanced fully-nested experiment will now be described.

For each level, the data obtained in the experiment are denoted y_{ijkl} (with i representing factor 0, i.e. laboratory, $i = 1,\ldots,p$; j representing factor 1, $j = 1,2$; k representing factor 2, $k = 1,2$; and l representing the replicate, $l = 1,2$).

The basic model used in this document is described in 6.1.4 as Formula (1). For a balanced nested design with two factors, this model is expanded to

$$y_{ijkl} = m + B_i + B_{ij} + B_{ijk} + e_{ijkl} \qquad \text{(B.2)}$$

The term B_i represents the effect corresponding to factor 0 (laboratory) with variance σ_0^2.

The term B_{ij} represents the effect corresponding to factor 1 with variance $\sigma_{(1)}^2$.

The term B_{ijk} represents the effect corresponding to factor 2 with variance $\sigma_{(2)}^2$.

The term e_{ijkl} represents the residual or repeatability effect with variance σ_r^2.

In a first step, the mean values are computed as follows:

$$\bar{y}_{ijk} = \frac{1}{2}\left(y_{ijk1} + y_{ijk2}\right)$$

$$\bar{y}_{ij} = \frac{1}{2}\left(\bar{y}_{ij1} + \bar{y}_{ij2}\right)$$

$$\bar{y}_i = \frac{1}{2}\left(\bar{y}_{i1} + \bar{y}_{i2}\right)$$

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

where p denotes the number of laboratories which have participated in the collaborative study.

The total sum of squares, SST, can be partitioned as

$$SST = \sum_i \sum_j \sum_k \sum_l \left(y_{ijkl} - \bar{\bar{y}}\right)^2 = SS0 + SS1 + SS2 + SSe$$

where

$$SS0 = \sum_i \sum_j \sum_k \sum_l \left(\bar{y}_i - \bar{\bar{y}}\right)^2 = 8\sum_i \left(\bar{y}_i - \bar{\bar{y}}\right)^2$$

$$SS1 = \sum_i \sum_j \sum_k \sum_l \left(\bar{y}_{ij} - \bar{y}_i\right)^2 = 4\sum_i \sum_j \left(\bar{y}_{ij} - \bar{y}_i\right)^2 = 4\sum_i s^2_{iA(1)}$$

$$SS2 = \sum_i \sum_j \sum_k \sum_l \left(\bar{y}_{ijk} - \bar{y}_{ij}\right)^2 = 2\sum_i \sum_j \sum_k \left(\bar{y}_{ijk} - \bar{y}_{ij}\right)^2 = 4\sum_i \sum_j s^2_{iA(2)}$$

$$SSe = \sum_i \sum_j \sum_k \sum_l \left(y_{ijkl} - \bar{y}_{ijk}\right)^2 = 2\sum_i \sum_j \sum_k s^2_{ij} = 4\sum_i s^2_{iB}$$

The methods described in ISO 5725-2:2019, 8.3 should be applied to check the data for consistency and outliers.

The Cochran test should be applied to check

the variances s^2_{ij}, $i = 1, \ldots, p$, $j = 1, 2$, for outliers within the replicates

the variances s^2_{iB}, $i = 1, \ldots, p$, for repeatability outliers

the variances $s^2_{iA(1)}$, $i = 1, \ldots, p$, for intermediate precision outliers regarding factor 1

the variances $s^2_{iA(2)}$, $i = 1, \ldots, p$, for intermediate precision outliers regarding factor 2

The Grubbs test should be applied to check the laboratory mean values $\bar{y}_i$, $i = 1, \ldots, p$ for outliers.

Since the degrees of freedom for the sums of squares $SS0$, $SS1$, $SS2$ and SSe are $p-1$, p, $2p$ and $4p$, respectively, the ANOVA table is composed as shown in Table B.2.

The unbiased estimates $s^2_0, s^2_{(1)}, s^2_{(2)}$ and s^2_r for $\sigma^2_0, \sigma^2_{(1)}, \sigma^2_{(2)}$ and σ^2_r, respectively, can be obtained from the mean squares $MS0$, $MS1$, $MS2$ and MSe as

$$s^2_0 = \frac{1}{8}(MS0 - MS1)$$

$$s^2_{(1)} = \frac{1}{4}(MS1 - MS2)$$

$$s^2_{(2)} = \frac{1}{2}(MS2 - MSe)$$

$s_r^2 = MSe$.

Table B.2 — ANOVA table for a three-factor balanced fully-nested experiment

Source	Sum of Squares	Degrees of freedom	Mean square	Expected mean square
0	$SS0$	$p-1$	$MS0 = SS0/(p-1)$	$\sigma_r^2 + 2\sigma_{(2)}^2 + 4\sigma_{(1)}^2 + 8\sigma_0^2$
1	$SS1$	p	$MS1 = SS1/p$	$\sigma_r^2 + 2\sigma_{(2)}^2 + 4\sigma_{(1)}^2$
2	$SS2$	$2p$	$MS2 = SS2/(2p)$	$\sigma_r^2 + 2\sigma_{(2)}^2$
Residual	SSe	$4p$	$MSe = SSe/(4p)$	σ_r^2
Total	SST	$8p-1$		

The estimates of the repeatability variance, one-factor-different intermediate precision variance, two-factor-different intermediate precision variance and reproducibility variances are, respectively, as follows:

s_r^2

$s_{I(1)}^2 = s_r^2 + s_{(2)}^2$

$s_{I(2)}^2 = s_r^2 + s_{(2)}^2 + s_{(1)}^2$

$s_R^2 = s_r^2 + s_{(2)}^2 + s_{(1)}^2 + s_0^2$.

NOTE Strictly speaking, the outlier tests given here and in ISO 5725-2 are valid under the assumption that the underlying distributions are approximately normal; in practice, they work for most distributions provided they are unimodal.

Annex C
(informative)

Analysis of variance for staggered design

C.1 Two-factor staggered design

The analysis of variance described here shall be carried out separately for each level of the test included in the collaborative study. For the sake of readability, the subscript indicating the level of the test has been suppressed.

NOTE The subscript j is used in this annex for factor 1 (factor 0 being the laboratory), while in the other parts of ISO 5725 it is used for the level of the test.

The exact analysis of the data can be very complicated when some of the test results from a laboratory are missing. If it is decided that some of the test results from a laboratory are stragglers or outliers and should be excluded from the analysis, then it is recommended that all the data from that laboratory (at the level affected) should be excluded from the analysis. Alternatively, the statistical technique REML (Annex I) can be applied.

The outlier tests and the computation of the precision estimates for the two-factor staggered experiment will now be described.

For each level, the data obtained in the experiment are denoted y_{ijk} (with i representing factor 0, i.e. laboratory, $i = 1, \ldots, p$; j representing factor 1, $j = 1, 2$; and k representing the replicate, with k ranging from 1 to $k(j)$, with $k(1) = 2$ and $k(2) = 1$).

The basic model used in this document is described in 6.1.4 as Formula (1). For a staggered design with two factors, this model is expanded to Formula (C.1):

$$y_{ijk} = m + B_i + B_{ij} + e_{ijk} \tag{C.1}$$

The term B_i represents the effect corresponding to factor 0 (laboratory) with variance σ_0^2.

The term B_{ij} represents the effect corresponding to factor 1 with variance $\sigma_{(1)}^2$.

The term e_{ijk} represents the residual or repeatability effect with variance σ_r^2.

For the sake of simplicity, for each level, the data obtained in the experiment within laboratory i can also be denoted y_{ij}, $j = 1, 2, 3$ (see Table 5 in 7.2). This simplified notation will be used in the following.

In the first step, mean values and ranges are computed as follows:

$$\bar{y}_{i(1)} = \frac{1}{2}(y_{i1} + y_{i2})$$

$$\bar{y}_{i(2)} = \frac{1}{3}(y_{i1} + y_{i2} + y_{i3})$$

$$\bar{\bar{y}} = \frac{1}{p}\sum_i \bar{y}_{i(2)}$$

$$w_{i(1)} = |y_{i1} - y_{i2}|$$

$$w_{i(2)} = \left| \bar{y}_{i(1)} - y_{i3} \right|$$

where p is the number of laboratories which have participated in the collaborative study.

The total sum of squares, SST, can be partitioned as follows:

$$SST = \sum_i \sum_j \left(y_{ij} - \bar{\bar{y}} \right)^2 = SS0 + SS1 + SSe$$

where

$$SS0 = 3 \sum_i \left(\bar{y}_{i(2)} \right)^2 - 3p \left(\bar{\bar{y}} \right)^2$$

$$SS1 = \frac{2}{3} \sum_i w_{i(2)}^2$$

$$SSe = \frac{1}{2} \sum_i w_{i(1)}^2$$

The methods described in ISO 5725-2:2019, 8.3 should be applied to check the data for consistency and outliers.

The Cochran test should be applied to check

the variances $s_{iB}^2 = \frac{1}{2} w_{i(1)}^2$, $i = 1, \ldots, p$, for repeatability outliers

the variances $s_{iA}^2 = \frac{1}{2} w_{i(2)}^2$, $i = 1, \ldots, p$, for intermediate precision outliers

The Grubbs test should be applied to check the laboratory mean values $\bar{y}_i$, $i = 1, \ldots, p$ for outliers.

Since the degrees of freedom for the sums of squares $SS0$, $SS1$ and SSe are $p-1$, p and p, respectively, the ANOVA table is composed as shown in Table C.1.

Table C.1 — ANOVA table for a two-factor staggered experiment

Source	Sum of Squares	Degrees of freedom	Mean square	Expected mean square
0	$SS0$	$p-1$	$MS0 = SS0/(p-1)$	$\sigma_r^2 + \frac{5}{3}\sigma_{(1)}^2 + 3\sigma_0^2$
1	$SS1$	p	$MS1 = SS1/p$	$\sigma_r^2 + \frac{4}{3}\sigma_{(1)}^2$
Residual	SSe	p	$MSe = SSe/p$	σ_r^2
Total	SST	$3p-1$		

The unbiased estimates s_0^2, $s_{(1)}^2$ and s_r^2 for σ_0^2, $\sigma_{(1)}^2$ and σ_r^2, respectively, can be obtained from the mean squares $MS0$, $MS1$ and MSe as:

$$s_0^2 = \frac{1}{3} MS0 - \frac{5}{12} MS1 + \frac{1}{12} MSe$$

$$s_{(1)}^2 = \frac{3}{4} MS1 - \frac{3}{4} MSe$$

$$s_r^2 = MSe\,.$$

The estimates of the repeatability, intermediate and reproducibility variances are, respectively, as follows:

$$s_r^2$$

$$s_{I(1)}^2 = s_r^2 + s_{(1)}^2$$

$$s_R^2 = s_r^2 + s_{(1)}^2 + s_0^2\,.$$

NOTE Strictly speaking, the outlier tests given here and in ISO 5725-2 are valid under the assumption that the underlying distributions are approximately normal; in practice, they work for most distributions provided they are unimodal.

C.2 Three-factor staggered design

The analysis of variance described here shall be carried out separately for each level of the test included in the collaborative study. For the sake of readability, the subscript indicating the level of the test has been suppressed.

NOTE The subscript j is used in this annex for factor 1 and factor 2 (factor 0 being the laboratory), while in the other parts of ISO 5725 it is used for the level of the test.

The exact analysis of the data can be very complicated when some of the test results from a laboratory are missing. If it is decided that some of the test results from a laboratory are stragglers or outliers and should be excluded from the analysis, then it is recommended that all the data from that laboratory (at the level affected) should be excluded from the analysis. Alternatively, the statistical technique REML (Annex I) can be applied.

The outlier tests and the computation of the precision estimates for the three-factor staggered experiment will now be described.

For each level, the data obtained in the experiment are denoted y_{ijkl} (with i representing factor 0, i.e. laboratory, $i = 1, \ldots, p$; j representing factor 1; k representing factor 2; and l representing the replicate.

The basic model used in this document is described in 6.1.4 as Formula (1). For a staggered design with three factors, this model is expanded to Formula (C.2)

$$y_{ijk} = m + B_i + B_{ij} + B_{ijk} + e_{ijkl} \tag{C.2}$$

The term B_i represents the effect corresponding to factor 0 (laboratory) with variance σ_0^2.

The term B_{ij} represents the effect corresponding to factor 1 with variance $\sigma_{(1)}^2$.

The term B_{ijk} represents the effect corresponding to factor 2 with variance $\sigma_{(2)}^2$.

The term e_{ijkl} represents the residual or repeatability effect with variance σ_r^2.

For the sake of simplicity, for each level, the data obtained in the experiment within laboratory i can also be denoted y_{ij}, $j = 1, 2, 3, 4$ (see Table 5 in 7.2). This simplified notation will be used in the following.

In the first step, mean values and ranges are computed as follows:

$$\bar{y}_{i(1)} = \frac{1}{2}(y_{i1} + y_{i2})$$

$$\bar{y}_{i(2)} = \frac{1}{3}(y_{i1} + y_{i2} + y_{i3})$$

$$\bar{y}_{i(3)} = \frac{1}{4}(y_{i1} + y_{i2} + y_{i3} + y_{i4})$$

$$\bar{\bar{y}} = \frac{1}{p}\sum_i \bar{y}_{i(3)}$$

$$w_{i(1)} = |y_{i1} - y_{i2}|$$

$$w_{i(2)} = \left|\bar{y}_{i(1)} - y_{i3}\right|$$

$$w_{i(3)} = \left|\bar{y}_{i(2)} - y_{i4}\right|$$

where p is the number of laboratories which have participated in the collaborative study.

The Cochran test should be applied to check

the variances for repeatability outliers

the variances for intermediate precision outliers regarding factor 1 and factor 2

The Grubbs test should be applied to check the laboratory mean values $\bar{y}_i$, $i = 1,\ldots,p$ for outliers.

The ANOVA table is composed as shown in Table C.2.

Table C.2 — ANOVA table for a three-factor staggered experiment

Source	Sum of squares	Degrees of freedom	Mean square	Expected mean square
0	$4\sum_i \left(\bar{y}_{i(3)}\right)^2 - 4p\left(\bar{\bar{y}}\right)^2$	$p-1$	$SS0/(p-1)$	$\sigma_r^2 + \frac{3}{2}\sigma_{(2)}^2 + \frac{5}{2}\sigma_{(1)}^2 + 4\sigma_0^2$
1	$\frac{3}{4}\sum_i w_{i(3)}^2$	p	$SS1/p$	$\sigma_r^2 + \frac{7}{6}\sigma_{(2)}^2 + \frac{3}{2}\sigma_{(1)}^2$
2	$\frac{2}{3}\sum_i w_{i(2)}^2$	p	$SS2/p$	$\sigma_r^2 + \frac{4}{3}\sigma_{(2)}^2$
Residual	$\frac{1}{2}\sum_i w_{i(1)}^2$	p	SSe/p	σ_r^2
Total	$\sum_i\sum_j \left(y_{ij} - \bar{\bar{y}}\right)^2$	$4p-1$		

NOTE Strictly speaking, the outlier tests given here and in ISO 5725-2 are valid under the assumption that the underlying distributions are approximately normal; in practice, they work for most distributions provided they are unimodal.

C.3 Four-factor staggered design

The analysis of variance described here shall be carried out separately for each level of the test included in the collaborative study. For the sake of readability, the subscript indicating the level of the test has been suppressed.

NOTE The subscript j is used in this Annex for factor 1, factor 2, and factor 3 (factor 0 being the laboratory), while in the other parts of ISO 5725 it is used for the level of the test.

The exact analysis of the data can be very complicated when some of the test results from a laboratory are missing. If it is decided that some of the test results from a laboratory are stragglers or outliers and should be excluded from the analysis, then it is recommended that all the data from that laboratory (at the level affected) should be excluded from the analysis. Alternatively, the statistical technique REML (Annex I) can be applied.

The outlier tests and the computation of the precision estimates for the four-factor staggered experiment will now be described.

For each level, the data obtained in the experiment are denoted y_{ijklm} (with i representing factor 0, i.e. laboratory, $i=1,\ldots,p$; j representing factor 1, j; k representing factor 2; l representing factor 3; m representing the replicate.

The basic model used in this document is described in 6.1.4 as Formula (1). For a staggered design with four factors, this model is expanded to Formula (C.3):

$$y_{ijklm} = m + B_i + B_{ij} + B_{ijk} + B_{ijkl} + e_{ijklm} \tag{C.3}$$

The term B_i represents the effect corresponding to factor 0 (laboratory) with variance σ_0^2.

The term B_{ij} represents the effect corresponding to factor 1 with variance $\sigma_{(1)}^2$.

The term B_{ijk} represents the effect corresponding to factor 2 with variance $\sigma_{(2)}^2$.

The term B_{ijkl} represents the effect corresponding to factor 3 with variance $\sigma_{(3)}^2$.

The term e_{ijklm} represents the residual or repeatability effect with variance σ_r^2.

For the sake of simplicity, for each level, the data obtained in the experiment within laboratory i can also be denoted y_{ij}, $j=1,\ 2,\ 3,\ 4,\ 5$ (see Table 5). This simplified notation will be used in the following.

In the first step, mean values and ranges are computed as follows:

$$\bar{y}_{i(1)} = \frac{1}{2}(y_{i1} + y_{i2})$$

$$\bar{y}_{i(2)} = \frac{1}{3}(y_{i1} + y_{i2} + y_{i3})$$

$$\bar{y}_{i(3)} = \frac{1}{4}(y_{i1} + y_{i2} + y_{i3} + y_{i4})$$

$$\bar{y}_{i(4)} = \frac{1}{5}(y_{i1} + y_{i2} + y_{i3} + y_{i4} + y_{i5})$$

$$\bar{\bar{y}} = \frac{1}{p}\sum_i \bar{y}_{i(4)}$$

$$w_{i(1)} = |y_{i1} - y_{i2}|$$

$$w_{i(2)} = \left|\bar{y}_{i(1)} - y_{i3}\right|$$

$$w_{i(3)} = \left|\bar{y}_{i(2)} - y_{i4}\right|$$

$$w_{i(4)} = \left|\bar{y}_{i(3)} - y_{i5}\right|$$

where p is the number of laboratories which have participated in the collaborative study.

The Cochran test should be applied to check

the variances for repeatability outliers

the variances for intermediate precision outliers regarding factor 1, factor 2, and factor 3

The Grubbs test should be applied to check the laboratory mean values $\bar{y}_i$, $i = 1, \ldots, p$ for outliers.

The ANOVA table is composed as shown in Table C.3.

Table C.3 — ANOVA table for a four-factor staggered experiment

Source	Sum of squares	Degrees of freedom	Mean square	Expected mean square
0	$5\sum_i \left(\bar{y}_{i(4)}\right)^2 - 5p(\bar{\bar{y}})^2$	$p-1$	$SS0/(p-1)$	$\sigma_r^2 + \frac{7}{5}\sigma_{(3)}^2 + \frac{11}{5}\sigma_{(2)}^2 + \frac{17}{5}\sigma_{(1)}^2 + 5\sigma_0^2$
1	$\frac{4}{5}\sum_i w_{i(4)}^2$	p	$SS1/p$	$\sigma_r^2 + \frac{11}{10}\sigma_{(3)}^2 + \frac{13}{10}\sigma_{(2)}^2 + \frac{8}{5}\sigma_{(1)}^2$
2	$\frac{3}{4}\sum_i w_{i(3)}^2$	p	$SS2/p$	$\sigma_r^2 + \frac{7}{6}\sigma_{(3)}^2 + \frac{3}{2}\sigma_{(2)}^2$
3	$\frac{2}{3}\sum_i w_{i(2)}^2$	p	$SS3/p$	$\sigma_r^2 + \frac{4}{3}\sigma_{(3)}^2$
Residual	$\frac{1}{2}\sum_i w_{i(1)}^2$	p	SSe/p	σ_r^2
Total	$\sum_i\sum_j \left(y_{ij} - \bar{\bar{y}}\right)^2$	$5p-1$		

NOTE Strictly speaking, the outlier tests given here and in ISO 5725-2 are valid under the assumption that the underlying distributions are approximately normal; in practice, they work for most distributions provided they are unimodal.

C.4 Five-factor staggered design

The analysis of variance described here shall be carried out separately for each level of the test included in the collaborative study. For the sake of readability, the subscript indicating the level of the test has been suppressed.

NOTE The subscript j is used in this Annex for factor 1, factor 2, factor 3 and factor 4 (factor 0 being the laboratory), while in the other parts of ISO 5725 it is used for the level of the test.

The exact analysis of the data can be very complicated when some of the test results from a laboratory are missing. If it is decided that some of the test results from a laboratory are stragglers or outliers and should be excluded from the analysis, then it is recommended that all the data from that laboratory (at the level affected) should be excluded from the analysis. Alternatively, the statistical technique REML (Annex I) can be applied.

The outlier tests and the computation of the precision estimates for the five-factor staggered experiment will now be described.

For each level, the data obtained in the experiment are denoted y_{ijklmn} (with i representing factor 0, i.e. laboratory, $i=1,\ldots,p$; j representing factor 1, j; k representing factor 2; l representing factor 3; m representing factor 4; n representing the replicate.

The basic model used in this document is described in 6.1.4 as Formula (1). For a staggered design with four factors, this model is expanded to Formula (C.4):

$$y_{ijk} = m + B_i + B_{ij} + B_{ijk} + B_{ijkl} + B_{ijklm} + e_{ijklmn} \tag{C.4}$$

The term B_i represents the effect corresponding to factor 0 (laboratory) with variance σ_0^2.

The term B_{ij} represents the effect corresponding to factor 1 with variance $\sigma_{(1)}^2$.

The term B_{ijk} represents the effect corresponding to factor 2 with variance $\sigma_{(2)}^2$.

The term B_{ijkl} represents the effect corresponding to factor 3 with variance $\sigma_{(3)}^2$.

The term B_{ijklm} represents the effect corresponding to factor 4 with variance $\sigma_{(4)}^2$.

The term e_{ijklmn} represents the residual or repeatability effect with variance σ_r^2.

For the sake of simplicity, for each level, the data obtained in the experiment within laboratory i can also be denoted y_{ij}, $j=1, 2, 3, 4, 5, 6$ (see Table 5). This simplified notation will be used in the following.

In the first step, mean values and ranges are computed as follows:

$$\bar{y}_{i(1)} = \frac{1}{2}(y_{i1} + y_{i2})$$

$$\bar{y}_{i(2)} = \frac{1}{3}(y_{i1} + y_{i2} + y_{i3})$$

$$\bar{y}_{i(3)} = \frac{1}{4}(y_{i1} + y_{i2} + y_{i3} + y_{i4})$$

$$\bar{y}_{i(4)} = \frac{1}{5}(y_{i1} + y_{i2} + y_{i3} + y_{i4} + y_{i5})$$

$$\bar{y}_{i(5)} = \frac{1}{6}(y_{i1} + y_{i2} + y_{i3} + y_{i4} + y_{i5} + y_{i6})$$

$$\bar{\bar{y}} = \frac{1}{p}\sum_i \bar{y}_{i(5)}$$

$$w_{i(1)} = |y_{i1} - y_{i2}|$$

$$w_{i(2)} = |\bar{y}_{i(1)} - y_{i3}|$$

$$w_{i(3)} = |\bar{y}_{i(2)} - y_{i4}|$$

$$w_{i(4)} = |\bar{y}_{i(3)} - y_{i5}|$$

$$w_{i(5)} = \left| \bar{y}_{i(4)} - y_{i6} \right|$$

where p is the number of laboratories which have participated in the collaborative study.

The Cochran test should be applied to check

the variances for repeatability outliers

the variances for intermediate precision outliers regarding factor 1, factor 2, factor 3, and factor 4

The Grubbs test should be applied to check the laboratory mean values $\bar{y}_i$, $i = 1, \ldots, p$ for outliers.

The ANOVA table is composed as shown in Table C.4.

Table C.4 — ANOVA table for a five-factor staggered experiment

Source	Sum of Squares	Degrees of freedom	Mean square	Expected mean square
0	$6\sum_i \left(\bar{y}_{i(5)}\right)^2 - 6p\left(\bar{\bar{y}}\right)^2$	$p-1$	$SS0/(p-1)$	$\sigma_r^2 + \frac{4}{3}\sigma_{(4)}^2 + 2\sigma_{(3)}^2 + 3\sigma_{(2)}^2 + \frac{13}{3}\sigma_{(1)}^2 + 6\sigma_0^2$
1	$\frac{5}{6}\sum_i w_{i(5)}^2$	p	$SS1/p$	$\sigma_r^2 + \frac{16}{15}\sigma_{(4)}^2 + \frac{6}{5}\sigma_{(3)}^2 + \frac{7}{5}\sigma_{(2)}^2 + \frac{5}{3}\sigma_{(1)}^2$
2	$\frac{4}{5}\sum_i w_{i(4)}^2$	p	$SS2/p$	$\sigma_r^2 + \frac{11}{10}\sigma_{(4)}^2 + \frac{13}{10}\sigma_{(3)}^2 + \frac{8}{5}\sigma_{(2)}^2$
3	$\frac{3}{4}\sum_i w_{i(3)}^2$	p	$SS3/p$	$\sigma_r^2 + \frac{7}{6}\sigma_{(4)}^2 + \frac{3}{2}\sigma_{(3)}^2$
4	$\frac{2}{3}\sum_i w_{i(2)}^2$	p	$SS4/p$	$\sigma_r^2 + \frac{4}{3}\sigma_{(4)}^2$
Residual	$\frac{1}{2}\sum_i w_{i(1)}^2$	p	SSe/p	σ_r^2
Total	$\sum_i \sum_j \left(y_{ij} - \bar{\bar{y}}\right)^2$	$6p-1$		

NOTE Strictly speaking, the outlier tests given here and in ISO 5725-2 are valid under the assumption that the underlying distributions are approximately normal; in practice, they work for most distributions provided they are unimodal.

Annex D
(informative)

Analysis of variance for the balanced partially-nested design (three factors)

The analysis of variance described here shall be carried out separately for each level of the test included in the collaborative study. For the sake of readability, the subscript indicating the level of the test has been suppressed.

NOTE The subscript *j* is used in this document for factor 1 (factor 0 being the laboratory), while in the other parts of ISO 5725 it is used for the level of the test.

The exact analysis of the data can be very complicated when some of the test results from a laboratory are missing. If it is decided that some of the test results from a laboratory are stragglers or outliers and should be excluded from the analysis, then it is recommended that all the data from that laboratory (at the level affected) should be excluded from the analysis.

The outlier tests and the computation of the precision estimates for the partially-nested experiment with 3 factors and $n=2$ replicates will now be described.

NOTE The balanced partially-nested experiment with 2 factors is equivalent to the balanced nested experiment with 2 factors. Computation of precision estimates can, therefore, be computed as described in Annex B.

The data obtained in the experiment are denoted y_{ijkl} where $i=1,\ldots,p$ denotes the laboratory, $j=1,2$ and $k=1,2$ denote the intermediate factors (i.e. subsumed under laboratory), and $l=1,2$ denotes the replicate.

The basic model used in this document is described in 6.1.4 as Formula (1). For a partially-nested design with three factors, this model is expanded to

$$y_{ijkl} = m + B_i + B_{ij} + B_{ik} + B_{ijk} + e_{ijkl} \tag{D.1}$$

The term B_i represents the effect corresponding to factor 0 (laboratory) with variance σ_0^2.

The term B_{ij} represents the effect corresponding to factor 1 with variance $\sigma_{(1)}^2$.

The term B_{ik} represents the effect corresponding to factor 2 with variance $\sigma_{(2)}^2$.

The term B_{ijk} represents the effect corresponding to interaction between factors 1 and 2, with variance $\sigma_{Interaction}^2$.

The term e_{ijkl} represents the residual or repeatability effect with variance σ_r^2.

In a first step, mean values are computed as follows:

$$\bar{y}_{ijk} = \frac{1}{2}\sum_l y_{ijkl}$$

$$\bar{y}_{ij} = \frac{1}{2}\sum_k \bar{y}_{ijk}$$

$$\bar{y}_{ik} = \frac{1}{2}\sum_j \bar{y}_{ijk}$$

$$\bar{y}_i = \frac{1}{2}\sum_j \bar{y}_{ij}$$

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

The total sum of squares, SST, can be partitioned as

$$SST = \sum_i \sum_j \sum_k \sum_l \left(y_{ijkl} - \bar{\bar{y}}\right)^2 = SS0 + SS1 + SS2 + SS3 + SSe$$

where

$$SS0 = 8\sum_i \left(\bar{y}_i - \bar{\bar{y}}\right)^2$$

$$SS1 = 4\sum_i \sum_j \left(\bar{y}_{ij} - \bar{y}_i\right)^2 = 4\sum_i s_{iA}^2$$

$$SS2 = 4\sum_i \sum_k \left(\bar{y}_{ik} - \bar{y}_i\right)^2 = 4\sum_i s_{iB}^2$$

$$SS3 = 2\sum_i \sum_j \sum_k \left(\bar{y}_{ijk} - \bar{y}_i\right)^2 - SS1 - SS2 = 4\sum_i s_{iC}^2 - SS1 - SS2$$

$$SSe = \sum_i \sum_j \sum_k \sum_l \left(y_{ijkl} - \bar{y}_{ijk}\right)^2 = 4\sum_i s_{iD}^2$$

The methods described in ISO 5725-2:2019, 8.3 should be applied to check the data for consistency and outliers.

The Cochran test should be applied to check

the variances s_{iA}^2, $i = 1, \ldots, p$, for outlying effects of factor 1

the variances s_{iB}^2, $i = 1, \ldots, p$, for outlying effects of factor 2

the variances s_{iC}^2, $i = 1, \ldots, p$, for outlying interaction outliers

the variances s_{iD}^2, $i = 1, \ldots, p$, for repeatability outliers

The Grubbs test should be applied to check the laboratory mean values $\bar{y}_i$, $i = 1, \ldots, p$ for outliers.

Since the degrees of freedom for the sums of squares $SS0$, $SS1$, $SS2$, $SS3$ and SSe are $p-1$, p, p, p and $4p$ respectively, the ANOVA table is composed as shown in Table D.1.

The unbiased estimates $s_0^2, s_{(1)}^2, s_{(2)}^2, s_{\text{interaction}}^2$ and s_r^2 for $\sigma_0^2, \sigma_{(1)}^2, \sigma_{(2)}^2, \sigma_{\text{interaction}}^2$ and σ_r^2, respectively, can be obtained from the mean squares $MS0$, $MS1$, $MS2$, $MS3$ and MSe as

$$s_0^2 = \frac{1}{8}(MS0 - MS1 - MS2 - MS3 + 2MSe)$$

$$s_{(1)}^2 = \frac{1}{4}(MS1 - MSe)$$

$$s_{(2)}^2 = \frac{1}{4}(MS2 - MSe)$$

$$s^2_{\text{interaction}} = \frac{1}{4}(MS3 - MSe)$$

$$s_r^2 = MSe\,.$$

Table D.1 — ANOVA table for a three-factor balanced partially-nested experiment

Source	Sum of Squares	Degrees of freedom	Mean square	Expected mean square
0	$SS0$	$p-1$	$MS0 = SS0/(p-1)$	$\sigma_r^2 + 4\sigma^2_{(1)} + 4\sigma^2_{(2)} + 4\sigma^2_{\text{interaction}} + 8\sigma^2_0$
1	$SS1$	p	$MS1 = SS1/p$	$\sigma_r^2 + 4\sigma^2_{(1)}$
2	$SS2$	p	$MS2 = SS2/p$	$\sigma_r^2 + 4\sigma^2_{(2)}$
Interaction	$SS3$	p	$MS3 = SS3/p$	$\sigma_r^2 + 4\sigma^2_{\text{interaction}}$
Residual	SSe	$4p$	$MSe = SSe/(4p)$	σ_r^2
Total	SST	$8p-1$		

The estimates of the repeatability, intermediate (for factor 1, factor 2, and factors 1 and 2 together) and reproducibility variance are, respectively, as follows:

$$s_r^2$$

$$s^2_{I(1)} = s_r^2 + s^2_{(1)}$$

$$s^2_{I(2)} = s_r^2 + s^2_{(2)}$$

$$s^2_{I(1,2)} = s_r^2 + s^2_{(1)} + s^2_{(2)} + s^2_{\text{interaction}}$$

$$s_R^2 = s_r^2 + s^2_{(1)} + s^2_{(2)} + s^2_{\text{interaction}} + s^2_0\,.$$

NOTE Strictly speaking, the computation methods given here and in ISO 5725-2 are valid under the assumption that the underlying distributions are approximately normal; in practice, they work for most distributions provided they are unimodal.

Annex E
(informative)

Statistical model for an experiment with heterogeneous material

For each level, the data obtained in the experiment are denoted y_{ijk} (with i representing the laboratory, j representing the sample, and k representing the replicate).

The basic model used in this document is described in 6.1.4 as Formula (1). For an experiment with a heterogeneous material, this model is expanded to Formula (E.1):

$$y_{ijk} = m + B_i + H_{ij} + e_{ijk} \tag{E.1}$$

The term H_{ij} represents the variation between samples. It is reasonable to assume that, for each level, the variation between samples is random and does not depend on the laboratory, so the term H_{ij} has a zero expectation and variance $var\left(H_{ij}\right) = \sigma_H^2$.

The statistical analysis is carried out as described in Annex B.

Annex F
(informative)

Analysis of variance for split-level design

For each level, the data obtained in the split-level experiment are denoted y_{ij} (with i representing the laboratory and j representing the sample).

The basic model used in this document is given as Formula (1) in 6.1.4. For the split-level design, this model is modified to Formula (F.1):

$$y_{ij} = m_j + B_i + e_{ij} \tag{F.1}$$

The term m_j denotes the general average (expectation) for material j ($j = a$ or b).

The term B_i represents the effect corresponding to the factor laboratory with variance σ_0^2.

The term e_{ij} denotes the random error of test result j obtained in laboratory i with variance σ_r^2.

The lack of a subscript j in B_i implies that it is assumed that the bias associated with laboratory i does not depend on the material a or b within a level. This is why it is important that the two materials should be similar.

In the following, Σ represents summation over the laboratories $i = 1, 2, ..., p$.

Step 1

Compute the laboratory differences as $D_i = y_{ia} - y_{ib}$ and the overall difference as $D = \frac{\Sigma D_i}{p}$.

NOTE The method of analysis requires each difference D_i to be calculated in the same direction $a - b$ and the sign of the difference to be retained.

The standard deviation s_D is computed as follows:

$$s_D = \sqrt{\Sigma (D_i - D)^2 / (p-1)}$$

Step 2

Compute the laboratory averages as $y_i = \frac{y_{ia} + y_{ib}}{2}$ and the overall average as $y = \frac{\Sigma y_i}{p}$.

The standard deviation s_y is computed as follows:

$$s_y = \sqrt{\Sigma (y_i - y)^2 / (p-1)}$$

Step 3

Compute the repeatability standard deviation s_r and the reproducibility standard deviation s_R:

$$s_r = s_D / \sqrt{2}$$

$$s_R^2 = s_y^2 + \frac{s_r^2}{2}.$$

The estimate for σ_0^2 is $s_R^2 - s_r^2$. If the difference is negative, set the estimate = 0.

Annex G
(informative)

Example for split-level design

Table G.1 provides test results for 10 laboratories and 4 levels obtained in accordance with the split-level design.

Table G.1 — Test results for 10 laboratories and 4 levels according to the split level design

Laboratory	Level							
	1		2		3		4	
	a	b	a	b	a	b	a	b
1	9,87	10,91	27,64	28,60	58,60	66,07	76,14	97,85
2	10,04	11,09	27,80	32,03	46,68	65,78	99,05	94,69
3	8,67	9,82	30,11	28,79	69,63	55,27	85,26	93,93
4	9,49	10,88	33,34	27,80	57,80	53,77	62,67	97,45
5	9,38	10,89	31,69	31,40	56,56	74,70	73,35	79,00
6	10,09	10,18	28,53	32,69	58,19	61,97	80,61	41,97
7	10,11	10,88	28,56	33,70	54,80	49,89	98,50	79,58
8	9,99	10,80	31,22	28,13	70,41	53,15	72,53	98,70
9	10,34	10,15	32,31	30,15	54,86	57,48	94,07	96,80
10	9,79	10,39	26,58	29,09	70,06	63,63	98,08	95,95

The computation of precision estimates is performed as follows:

a) Step 1

The laboratory- and level-specific differences ($D_i = y_{ia} - y_{ib}$) are provided in Table G.2.

Table G.2 — Laboratory- and level-specific differences

Laboratory	Level			
	1	2	3	4
1	-1,04	-0,96	-7,47	-21,71
2	-1,05	-4,23	-19,10	4,36
3	-1,15	1,32	14,36	-8,67
4	-1,39	5,54	4,03	-34,78
5	-1,51	0,29	-18,14	-5,65
6	-0,09	-4,16	-3,78	38,64
7	-0,77	-5,14	4,91	18,92
8	-0,81	3,09	17,26	-26,17
9	0,19	2,16	-2,62	-2,73
10	-0,60	-2,51	6,43	2,13

The level-specific overall differences $\left(D = \frac{\sum D_i}{p} \right)$ are provided in Table G.3.