BS ISO 5725-3:2023

BSI Standards Publication

Accuracy (trueness and precision) of measurement methods and results

Part 3: Intermediate precision and alternative designs for collaborative studies

Documents
Document Sharing Hub

Full document details: BS ISO 5725-3:2023

National foreword

This British Standard is the UK implementation of ISO 5725-3:2023. It supersedes BS ISO 5725-3:1994, which is withdrawn.

The UK participation in its preparation was entrusted to Technical Committee SS/6, Statistical procedures for measurement methods and results.

A list of organizations represented on this committee can be obtained on request to its committee manager.

Contractual and legal considerations

This publication has been prepared in good faith, however no representation, warranty, assurance or undertaking (express or implied) is or will be made, and no responsibility or liability is or will be accepted by BSI in relation to the adequacy, accuracy, completeness or reasonableness of this publication. All and any such responsibility and liability is expressly disclaimed to the full extent permitted by the law.

This publication is provided as is, and is to be used at the recipient's own risk.

The recipient is advised to consider seeking professional guidance with respect to its use of this publication.

This publication is not intended to constitute a contract. Users are responsible for its correct application.

© The British Standards Institution 2023
Published by BSI Standards Limited 2023

ISBN 978 0 539 01791 5

ICS 17.020

Compliance with a British Standard cannot confer immunity from legal obligations.

This British Standard was published under the authority of the Standards Policy and Strategy Committee on 31 July 2023.

Amendments/corrigenda issued since publication

Date	Text affected

Documents
Document Sharing Hub

Full document details: BS ISO 5725-3:2023

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)

Documents
Document Sharing Hub
© ISO 2023

Full document details: BS ISO 5725-3:2023

COPYRIGHT PROTECTED DOCUMENT

© ISO 2023

All rights reserved. Unless otherwise specified, or required in the context of its implementation, no part of this publication may be reproduced or utilized otherwise in any form or by any means, electronic or mechanical, including photocopying, or posting on the internet or an intranet, without prior written permission. Permission can be requested from either ISO at the address below or ISO's member body in the country of the requester.

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

© ISO 2023 – All rights reserved

Documents
Document Sharing Hub

Full document details: BS ISO 5725-3:2023

Contents

Page

Foreword ... **v**
Introduction ... **vi**
1 Scope ... **1**
2 Normative references ... **2**
3 Terms and definitions ... **2**
4 Symbols ... **3**
5 General requirements ... **4**
6 Intermediate measures of the precision of a standard measurement method ... **5**
6.1 Factors and factor levels ... 5
6.1.1 Definitions and examples ... 5
6.1.2 Selection of factors of interest ... 6
6.1.3 Random and fixed effects ... 6
6.1.4 Statistical model ... 7
6.2 Within-laboratory study and analysis of intermediate precision measures ... 9
6.2.1 Simplest approach ... 9
6.2.2 Alternative method ... 10
6.2.3 Effect of the measurement conditions on the final quoted result ... 10
7 Nested design ... **11**
7.1 Balanced fully-nested design ... 11
7.2 Staggered-nested design ... 12
7.3 Balanced partially-nested design ... 13
7.4 Orthogonal array design ... 14
8 Design for heterogeneous material ... **16**
8.1 Applications of the design for a heterogeneous material ... 16
8.2 Layout of the design for a heterogeneous material ... 17
8.3 Statistical analysis ... 17
9 Split-level design ... **17**
9.1 Applications of the split-level design ... 17
9.2 Layout of the split-level design ... 19
9.3 Statistical analysis ... 19
10 Design across levels ... **19**
10.1 Applications of the design across levels ... 19
10.2 Layout of the design across levels ... 20
10.3 Statistical analysis ... 20
11 Reliability of interlaboratory parameters ... **20**
11.1 Reliability of precision estimates ... 20
11.2 Reliability of estimates of the overall mean ... 21
11.2.1 General ... 21
11.2.2 Balanced fully-nested design (2 factors) ... 21
11.2.3 Staggered nested design (2 factors) ... 21
11.2.4 Balanced partially-nested design ... 21
11.2.5 Orthogonal array design ... 21
11.2.6 Split-level design ... 22
Annex A (informative) **Fully- and partially-nested designs** ... **23**
Annex B (informative) **Analysis of variance for balanced fully-nested design** ... **25**
Annex C (informative) **Analysis of variance for staggered design** ... **30**
Annex D (informative) **Analysis of variance for the balanced partially-nested design (three factors)** ... **88**

© ISO 2023 – All rights reserved

AVA Documents
Document Sharing Hub
Full document details: BS ISO 5725-3:2023

Annex E (informative) **Statistical model for an experiment with heterogeneous material** 41
Annex F (informative) **Analysis of variance for split-level design** 42
Annex G (informative) **Example for split-level design** 44
Annex H (informative) **Design across levels** 47
Annex I (informative) **Restricted maximum likelihood (REML)** 48
Annex J (informative) **Examples of the statistical analysis of intermediate precision experiment** 49
Annex K (informative) **Example for an analysis across levels** 55
Bibliography 57

© ISO 2023 – All rights reserved
Documents
Document Sharing Hub
Full document details: BS ISO 5725-3:2023

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.

© ISO 2023 – All rights reserved

Documents
Document Sharing Hub
Full document details: BS ISO 5725-3:2023

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.

© ISO 2023 – All rights reserved
Documents
Document Sharing Hub
Full document details: BS ISO 5725-3:2023

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.

Documents
Document Sharing Hub

© ISO 2023 – All rights reserved

Full document details: BS ISO 5725-3:2023

Documents
Document Sharing Hub

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.

© ISO 2023 – All rights reserved

Documents
Document Sharing Hub
Full document details: BS ISO 5725-3:2023

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.

© ISO 2023 – All rights reserved
Documents
Document Sharing Hub
Full document details: BS ISO 5725-3:2023

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)

© ISO 2023 – All rights reserved

Documents
Document Sharing Hub
Full document details: BS ISO 5725-3:2023

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.

© ISO 2023 – All rights reserved
Documents
Document Sharing Hub
Full document details: BS ISO 5725-3:2023

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.

© ISO 2023 – All rights reserved

Documents
Document Sharing Hub
Full document details: BS ISO 5725-3:2023

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.

© ISO 2023 – All rights reserved
Documents
Document Sharing Hub
Full document details: BS ISO 5725-3:2023

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

© ISO 2023 – All rights reserved

Documents
Document Sharing Hub
Full document details: BS ISO 5725-3:2023

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:

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

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

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

$Var(B)$ 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.

© ISO 2023 – All rights reserved
Documents
Document Sharing Hub
Full document details: BS ISO 5725-3:2023

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 kth 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.

AVA Documents
Document Sharing Hub

© ISO 2023 – All rights reserved

Full document details: BS ISO 5725-3:2023

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 $s_{I(O)}$ to be calculated.

6.2.2.2 Analysis

A graph of $\left(y_{jk} - \bar{y}_j\right)$ versus the material number, j, where y_{jk} is the kth test result on the jth material and $\bar{y}_j$ is the average of the n replicate test results on the jth 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):

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

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

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

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

© ISO 2023 – All rights reserved
Documents
Document Sharing Hub
Full document details: BS ISO 5725-3:2023

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

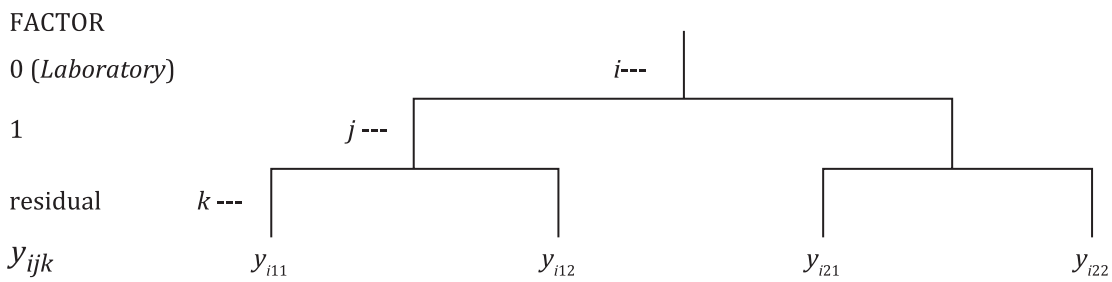
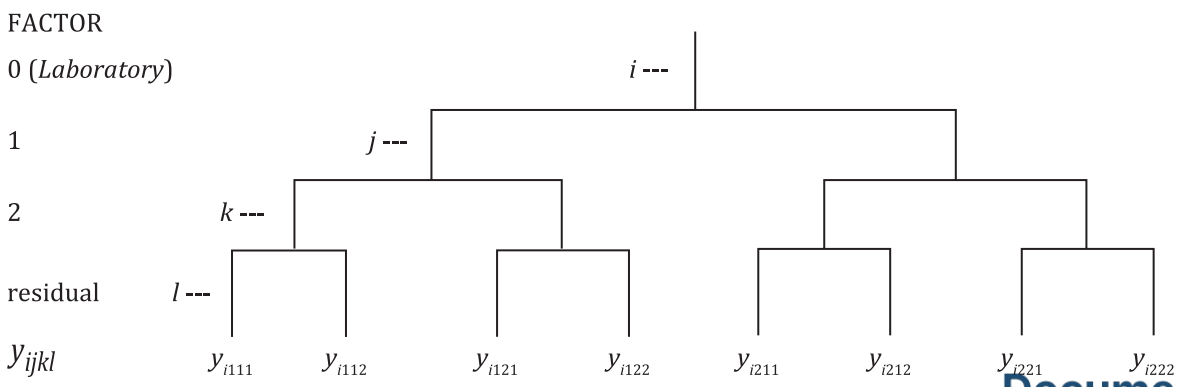


Table 4 — Schematic layout for three-factor balanced fully-nested design



© ISO 2023 – All rights reserved

AVA Documents Document Sharing Hub
Full document details: BS ISO 5725-3:2023