Statistical methods for use in proficiency testing by interlaboratory comparisons

Méthodes statistiques utilisées dans les essais d'aptitude par comparaisons interlaboratoires
Contents

Foreword............................................................................................................................................................. v
0 Introduction ......................................................................................................................................................... vi
  0.1 The aims of proficiency testing ................................................................................................................ vi
  0.2 ISO/IEC Guide 43 ....................................................................................................................................... vi
1 Scope .............................................................................................................................................................. 1
2 Normative references ...................................................................................................................................... 1
3 Terms and definitions .................................................................................................................................... 1
4 Statistical guidelines for the design and interpretation of proficiency tests ........................................... 2
  4.1 Action and warning signals ........................................................................................................................ 2
  4.2 Guidelines for limiting the uncertainty of the assigned value .............................................................. 3
  4.3 Guidelines for choosing the number of replicate measurements .......................................................... 3
  4.4 Homogeneity and stability of samples ...................................................................................................... 4
  4.5 Operationally defined measurement methods ....................................................................................... 4
  4.6 Reporting of data ....................................................................................................................................... 5
  4.7 Period of validity of the results of proficiency tests ............................................................................... 5
5 Determination of the assigned value and its standard uncertainty ............................................................. 5
  5.1 Choice of method of determining the assigned value ........................................................................... 5
  5.2 Formulation ............................................................................................................................................... 5
  5.3 Certified reference values ....................................................................................................................... 6
  5.4 Reference values ...................................................................................................................................... 6
  5.5 Consensus values from expert laboratories ............................................................................................. 8
  5.6 Consensus value from participants ......................................................................................................... 9
  5.7 Comparison of the assigned value ......................................................................................................... 14
  5.8 Missing values ......................................................................................................................................... 14
6 Determining the standard deviation for proficiency assessment ................................................................. 15
  6.1 Choice of method ..................................................................................................................................... 15
  6.2 Prescribed value ...................................................................................................................................... 15
  6.3 By perception .......................................................................................................................................... 15
  6.4 From a general model .............................................................................................................................. 17
  6.5 From the results of a precision experiment ............................................................................................ 17
  6.6 From data obtained in a round of a proficiency testing scheme .......................................................... 18
  6.7 Comparison of precision values derived from a proficiency test with established values .............. 18
7 Calculation of performance statistics ........................................................................................................... 18
  7.1 Estimates of laboratory bias ...................................................................................................................... 18
  7.2 Percentage differences ............................................................................................................................... 22
  7.3 Ranks and percentage ranks ..................................................................................................................... 24
  7.4 z-scores ..................................................................................................................................................... 25
  7.5 Eₙ numbers .............................................................................................................................................. 27
  7.6 z’-scores ................................................................................................................................................... 28
  7.7 Zeta-scores (ζ) ....................................................................................................................................... 29
  7.8 E₂ score .................................................................................................................................................... 30
  7.9 An example of the analysis of data when uncertainties are reported .................................................... 30
  7.10 Combined performance scores .............................................................................................................. 35
8 Graphical methods for combining performance scores for several measurands from one round of a proficiency test ..................................................................................................................... 36
  8.1 Application ............................................................................................................................................... 36
  8.2 Histograms of performance scores .......................................................................................................... 36
8.3 Bar-plots of standardized laboratory biases ................................................................. 37
8.4 Bar-plots of standardized repeatability measurements .................................................. 38
8.5 Youden Plot ..................................................................................................................... 38
8.6 Plots of repeatability standard deviations ...................................................................... 45
8.7 Split samples .................................................................................................................. 47
9 Graphical methods for combining performance scores over several rounds of a proficiency testing scheme ........................................................................................................ 52
9.1 Applications .................................................................................................................... 52
9.2 Shewhart control chart for $z$-scores ........................................................................... 52
9.3 Cusum control chart for $z$-scores ............................................................................... 55
9.4 Plots of standardized laboratory biases against laboratory averages .......................... 56
9.5 Dot plot ......................................................................................................................... 57
Annex A (normative) Symbols ............................................................................................ 59
Annex B (normative) Homogeneity and stability checks of samples ................................. 60
Annex C (normative) Robust analysis .................................................................................. 64
Bibliography ....................................................................................................................... 66
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.

International Standards are drafted in accordance with the rules given in the ISO/IEC Directives, Part 2.

The main task of technical committees is to prepare International Standards. Draft International Standards adopted by the technical committees are circulated to the member bodies for voting. Publication as an International Standard requires approval by at least 75 % of the member bodies casting a vote.

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. ISO shall not be held responsible for identifying any or all such patent rights.

ISO 13528 was prepared by Technical Committee ISO/TC 69, Applications of statistical methods, Subcommittee SC 6, Measurement methods and results.
0 Introduction

0.1 The aims of proficiency testing

Proficiency testing by interlaboratory comparisons is used to determine the performance of individual laboratories for specific tests or measurements, and to monitor the continuing performance of laboratories. The Introduction to ISO/IEC Guide 43-1:1997 should be consulted for a full exposition of the purposes of proficiency testing. In statistical language, the performance of laboratories can be described by three properties: laboratory bias, stability and repeatability. Laboratory bias and repeatability are defined in ISO 3534-1, ISO 3534-2 and ISO 5725-1. The stability of a laboratory's results is measured by intermediate precision as defined in ISO 5725-3.

Laboratory bias may be assessed by tests on reference materials, when these are available, using the procedure described in ISO 5725-4. Otherwise, proficiency testing by interlaboratory comparisons provides a generally available means of obtaining information about laboratory bias, and the use of data from proficiency tests to obtain estimates of laboratory bias is an important aspect of the analysis of such data. However, stability and repeatability will affect data obtained in proficiency tests, so that it is possible for a laboratory to obtain data in a round of a proficiency test which indicate bias that is actually caused by poor stability or poor repeatability. It is therefore important that these aspects of laboratory performance are assessed regularly.

Stability may be assessed by re-testing of retained samples, or by making regular measurements on a reference material or an in-house reference material (a stock of material established by a laboratory to use as private reference material). These techniques are described in ISO 5725-3. Stability may also be assessed by plotting estimates of laboratory bias derived from proficiency tests in control charts. This can provide information about laboratory performance that is not apparent from the examination of the results of individual rounds of proficiency testing schemes, and is another important aspect of the analysis of such data.

Data suitable for assessing repeatability may be generated by tests carried out in the normal course of the work of a laboratory, or by extra tests carried out within a laboratory specifically to assess repeatability. Consequently, the assessment of repeatability is not necessarily an important aspect of proficiency testing, although it is important that laboratories monitor their repeatability in some way. Repeatability may be assessed by plotting ranges of duplicate measurements on a control chart as described in ISO 5725-6.

The flowchart (Figure 1) illustrates how the techniques described in this International Standard are to be applied.

0.2 ISO/IEC Guide 43

ISO/IEC Guide 43-1 describes different types of proficiency testing schemes and gives guidance on the organization and design of proficiency testing schemes. ISO/IEC Guide 43-2 gives guidance on the selection and use of proficiency testing schemes by laboratory accreditation bodies. Those documents should be consulted for detailed information in those areas (the information is not duplicated here). ISO/IEC Guide 43-1 contains an annex that briefly describes the statistical methods that are used in proficiency testing schemes.

This International Standard is complementary to ISO/IEC Guide 43, providing detailed guidance that is lacking in that document on the use of statistical methods in proficiency testing. ISO 13528 is to a large extent based on a harmonized protocol for the proficiency testing of analytical laboratories [1], but is intended for use with all measurement methods.
Figure 1 — Flowchart showing the activities requiring the use of statistical methods when operating a proficiency testing scheme
Statistical methods for use in proficiency testing by interlaboratory comparisons

1 Scope

This International Standard complements ISO Guide 43 (all parts) by providing detailed descriptions of sound statistical methods for organizers to use to analyse the data obtained from proficiency testing schemes, and by giving recommendations on their use in practice by participants in such schemes and by accreditation bodies.

This International Standard can be applied to demonstrate that the measurement results obtained by laboratories do not exhibit evidence of an unacceptable level of bias.

It is applicable to quantitative data but not to qualitative data.

2 Normative references

The following referenced documents are indispensable for the application of this document. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

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

ISO 3534-2:— 1), 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


3 Terms and definitions

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

3.1 interlaboratory comparison
organization, performance and evaluation of tests or measurements on the same or similar test items by two or more laboratories in accordance with predetermined conditions

NOTE Adapted from ISO/IEC Guide 43-1.

3.2 proficiency testing
determination of laboratory testing performance by means of interlaboratory comparisons

1) To be published.
3.3 assigned value
value attributed to a particular quantity and accepted, sometimes by convention, as having an uncertainty appropriate for a given purpose

3.4 standard deviation for proficiency assessment
measure of dispersion used in the assessment of proficiency, based on the available information

3.5 z-score
standardized measure of laboratory bias, calculated using the assigned value and the standard deviation for proficiency assessment

3.6 coordinator
organization (or person) with responsibility for coordinating all of the activities involved in the operation of a proficiency testing scheme

4 Statistical guidelines for the design and interpretation of proficiency tests
(see ISO/IEC Guide 43-1:1997, 5.4.2.)

4.1 Action and warning signals

4.1.1 This International Standard describes some simple numerical or graphical criteria that should be applied to the data obtained in a proficiency test to see if they give rise to action or warning signals. Even in a well-run laboratory, with experienced staff, anomalous results may sometimes be obtained. Also, it is possible that a standardized measurement method, even though it has been validated by a precision experiment, may contain faults that become apparent only after several rounds of a proficiency testing scheme. The proficiency scheme itself may contain faults. For these reasons, the criteria given here shall not be used to condemn laboratories, as being unfit to perform the measurement method under examination. If proficiency testing is used to condemn laboratories, then it shall be necessary to devise appropriate criteria for that purpose.

4.1.2 The criteria given here are designed so that, when the standard deviation for proficiency assessment is based on observed performance (using one of the methods described in 6.4 to 6.6), the criteria give action signals when results are so exceptional as to merit investigation and corrective action.

4.1.3 The coordinator should have an understanding of the major sources of variability that can be anticipated in proficiency test data for the measurement in question. The first step in any analysis should be to examine the distribution of results for evidence of unanticipated sources of variability. For example, a bimodal distribution might be evidence of a mixed population of results caused by different methods, contaminated samples or poorly worded instructions. In this situation, the concern should be resolved before proceeding with analysis or evaluation. Accrediting bodies shall have policies for response to unacceptable performance in proficiency testing. Follow-up actions are determined by that policy or by the laboratory's quality procedures. However, there are generally recommended actions when a laboratory produces an unacceptable result in a proficiency test. Guidance for actions by laboratories in response to unsuccessful performance on a proficiency test is given in 4.1.4.

4.1.4 In schemes where the standard deviation for proficiency assessment is based on observed performance, when a result gives an action signal, the laboratory shall decide what investigations and corrective actions are appropriate, in consultation with the coordinator or an accreditation body if necessary. Unless there is a valid reason not to do so, the laboratory shall examine its procedures and identify one or more corrective actions that, according to staff in the laboratory, are likely to prevent the recurrence of such results. The laboratory may ask the coordinator for advice on possible causes of its problem, or may ask the coordinator to consult other experts. The laboratory shall take part in further rounds of the proficiency testing scheme to assess the effectiveness of the corrective actions. Appropriate corrective actions may be one of the following:
a) checking that staff understand and follow the measurement procedure;  
b) checking that all details of the measurement procedure are correct;  
c) checking the calibration of equipment and the composition of reagents;  
d) replacing suspect equipment or reagents;  
e) comparative tests of staff, equipment and/or reagents with another laboratory.


4.2 Guidelines for limiting the uncertainty of the assigned value

The assigned value $X$ has a standard uncertainty $u_X$ that depends on the method that is used to derive it, and also, when it is derived from tests in several laboratories, on the number of laboratories and, perhaps, on other factors. Methods for calculating the standard uncertainty of the assigned value are given in Clause 5.

The standard deviation for proficiency testing $\hat{\sigma}$ is used to assess the size of estimates of laboratory bias found in a proficiency test. Methods for obtaining the standard deviation for proficiency testing are given in Clause 6 and criteria that compare it with estimates of laboratory bias are given in Clause 7.

If the standard uncertainty $u_X$ of the assigned value is too large in comparison with the standard deviation for proficiency testing $\hat{\sigma}$, then there is a risk that some laboratories will receive action and warning signals because of inaccuracy in the determination of the assigned value, not because of any cause within the laboratories. For this reason, the standard uncertainty of the assigned value shall be established and shall be reported to laboratories participating in proficiency testing schemes (see ISO/IEC Guide 43-1:1997, A.1.4 and A.1.6).

If $u_X \leq 0.3 \hat{\sigma}$

then the uncertainty of the assigned value is negligible and need not be included in the interpretation of the results of the proficiency test.

If these guidelines are not met, then the coordinator shall consider the following.

a) Look for a method for determining the assigned value such that its uncertainty meets the above guideline.

b) Use the uncertainty of the assigned value in the interpretation of the results of the proficiency test (see 7.5 on $E_n$ numbers or 7.6 on the $z$-score).

c) Inform the participants in the proficiency test that the uncertainty of the assigned value is not negligible.

EXAMPLE Suppose that the assigned value $X$ is determined as the average $\bar{x}$ of the results of tests in 11 laboratories, and that the standard deviation for proficiency testing is determined as the standard deviation $s$ of these same 11 results, so $\hat{\sigma} = s$. As a first approximation, the standard uncertainty of the assigned value in this situation may be estimated by $u_X = s/\sqrt{11} = 0.3 s$, so that the requirement appears to be met. However, the requirement cannot be met in this situation with fewer than 11 laboratories. Further, the uncertainty of the assigned value will be larger than $s/\sqrt{11}$ if the samples suffer from non-homogeneity or instability, or if there is a factor that causes a common bias in the results of the laboratories (e.g. if they all use the same reference standard).

4.3 Guidelines for choosing the number of replicate measurements

Repeatability variation contributes to the variation between the laboratory biases in a proficiency test. If the repeatability variation is too large in comparison with the standard deviation for proficiency testing, then there