TECHNICAL REPORT

ISO/TR 27877

First edition

Statistical analysis for evaluating the precision of binary measurement methods and their results

PROOF/ÉPREUVE

Reference number
ISO/TR 27877:2021(E)

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 documents should be noted. This document was drafted in accordance with the editorial rules of the ISO/IEC Directives, Part 2 (see www.iso.org/directives).

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

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

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

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

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

The documents in the ISO 5725 series define the precision of quantitative measurement methods and their results, and assume that the errors follow normal distributions in their basic models. Also, they provide how to run experiments to evaluate precision measures, such as repeatability and reproducibility. Nowadays, there is also a demand for dealing with qualitative measurement methods and their results, which output binary data, categorical data, etc. However, the ISO 5725 series is not suitable mathematically for analyzing such data.

Several existing studies propose statistical methods for dealing with binary and/or categorical data, but no guidance documents are available so far. Hence, this document summaries various methods to evaluate the precision of binary measurement methods and their results, which are the most essential and frequently used methods for qualitative data.

Statistical analysis for evaluating the precision of binary measurement methods and their results

1 Scope

This document introduces five statistical methods for evaluating the precision of binary measurement methods and their results. The five methods can be divided into two types. Both types are based on measured values provided by each laboratory participating in a collaborative study. In the first type, each laboratory repeatedly measures a single sample. The samples measured by the laboratories are nominally identical. The second type is an extension of the first type, where there are several levels of samples.

For each statistical method, this document briefly summarizes its theory and explains how to estimate the proposed precision measures. Some real cases are illustrated to help the readers understand the evaluation procedures involved. For the first and second types of methods, five and three cases are presented, respectively.

Finally, this document compares the five statistical methods.

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

3 Terms and definitions, and symbols

3.1 Terms and definitions

For the purposes of this document, the terms and definitions given in ISO 3534-1 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/

3.1.1
accordance
probability that two binary measured values be identical when they are taken from the same laboratory

Note 1 to entry: The concept corresponds to the definition of “repeatability” in ISO 5725 and was originally proposed by Reference [3].

3.1.2
concordance
probability that two binary measured values be identical when they are taken from different laboratories

Note 1 to entry: The concept corresponds to the definition of "reproducibility" in ISO 5725 and was originally proposed by Reference [3].

3.1.3
ORDANOVA
statistical method for evaluating the precision of ordinal-scale measurement methods and their results based on an ordinal dispersion measure

Note 1 to entry: The concept was originally proposed by Reference [4].

3.1.4
true positive
TP
correct measured value in positive results, that is, the case where both the measured and the correct results are positive

3.1.5
true negative
TN
correct measured value in negative results, that is, the case where both the measured and the correct results are negative

3.1.6
false positive
FP
incorrect measured value in positive results, that is, the case where the measured value is positive but the correct one is negative

3.1.7
false negative
FN
incorrect measured value in negative results, that is, the case where the measured value is negative but the correct one is positive

3.1.8
confusion matrix
2×2 matrix showing the numbers of true positives, true negatives, false positives and false negatives

3.1.9
CM-accuracy
statistic for indicating the capability of two-class classifications, defined by the percentage of correct measured values

Note 1 to entry: The term CM-accuracy is identical to the term accuracy in the machine learning field, and is not generally used. However, this document uses CM-accuracy instead of accuracy in the machine learning field to distinguish between the term accuracy in ISO 5725 and the term in the machine learning field. In this document, the prefix CM stands for confusion matrixes because CM-accuracy can be calculated based on those.

3.1.10
sensitivity
statistic for indicating the capability of two-class classifications, defined by the percentage of correct positive measured values

 PROOF/ÉPREUVE

3.1.11
specificity
statistic for indicating the capability of two-class classifications, defined by the percentage of correct negative measured values

3.1.12
CM-precision
statistic for indicating the capability of two-class classifications, defined by the percentage of correct measured values in positive measured values

Note 1 to entry: The term CM-precision is identical to the term precision in the machine learning field, and is not generally used. However, this document uses CM-precision instead of precision in the machine learning field to distinguish between the term precision in ISO 5725 and the term in the machine learning field. In this document, the prefix CM stands for confusion matrixes because CM-precision can be calculated based on those.

3.1.13
F-measure
statistic for indicating the capability of two-class classifications, defined by the harmonic mean between sensitivity and CM-precision

3.1.14
kappa coefficient
statistic for indicating the capability of two-class classifications, defined by the ratio of CM-accuracy minus the possibility of the correctness occurring by chance to one minus the possibility of the correctness occurring by chance

Note 1 to entry: Note to entry 1: The kappa coefficient is an extended statistic of CM-accuracy, originally introduced by Reference [15], which takes into account the possibility of the correctness occurring by chance.

3.2 Symbols

L number of laboratories participating in a collaborative study

n number of repetitions in each laboratory participating in a collaborative study

i suffix describing a laboratory, and $i \in \{1, \ldots, L\}$

j suffix describing a repetition, and $j \in \{1, \ldots, n\}$

y_{ij} measured value of repetition j of laboratory i

x_i sum of the measured values y_{ij} of laboratory i for the case where $y_{ij} \in \{0,1\}$, that is,

$$x_i = \sum_{j=1}^{n} y_{ij} \quad \left(y_{ij} \in \{0,1\}\right)$$

$\bar{y}_i$ arithmetic mean of y_{ij} of laboratory i, that is, $\bar{y}_i = (1/n)\sum_{j=1}^{n} y_{ij} \quad \left(i \in \{1, \ldots, L\}\right)$

$\bar{\bar{y}}$ overall arithmetic mean of y_{ij}, that is, $\bar{\bar{y}} = \left(1/(nL)\right)\sum_{i=1}^{L}\sum_{j=1}^{n} y_{ij}$

n_i^p number of positive measured values of laboratory i

n^p sum of n_i^p with respect to $i \in \{1, \ldots, L\}$, that is, $n^p = \sum_{i=1}^{L} n_i^p$

c_{ij} (i, j)-element of a confusion matrix

m general mean (expectation) in the basic model of ISO 5725-2

B_i laboratory component of bias under repeatability conditions of laboratory i in the basic model of ISO 5725-2

e_{ij} random error occurring in every measurement under repeatability conditions in the basic model of ISO 5725-2

σ_{ri}^2 within-laboratory variance of laboratory i

σ_r^2 repeatability variance or within-laboratory variance

σ_L^2 between-laboratory variance

σ_R^2 reproducibility variance, that is, $\sigma_R^2 = \sigma_r^2 + \sigma_L^2$

$\sigma_{ri:W}^2$ within-laboratory variance of laboratory i in the ISO 5725-based method

$\sigma_{r:W}^2$ repeatability variance in the ISO 5725-based method

$\sigma_{L:W}^2$ between-laboratory variance in the ISO 5725-based method

$\sigma_{R:W}^2$ reproducibility variance in the ISO 5725-based method, that is, $\sigma_{R:W}^2 = \sigma_{r:W}^2 + \sigma_{L:W}^2$

σ^2 ordinal dispersion measure proposed by Reference [14] for the case of binary data assumed to follow a binomial distribution

$\sigma_{ri:O}^2$ within-laboratory variance of laboratory i in ORDANOVA

$\sigma_{r:O}^2$ repeatability variance in ORDANOVA

$\sigma_{L:O}^2$ between-laboratory variance in ORDANOVA

$\sigma_{R:O}^2$ reproducibility variance in ORDANOVA, that is, $\sigma_{R:O}^2 = \sigma_{r:O}^2 + \sigma_{L:O}^2$

p_i probability of obtaining a measurement value $y_{ij} = 1$ of laboratory i

p arithmetic mean of p_i, that is, $p = (1/L)\sum_{i=1}^{L} p_i$

H_0 null hypothesis of a statistical test

χ_0^2 test statistic of a chi-squared test

A accordance of Reference [3] method

C concordance of Reference [3] method

$\hat{X}$ estimate of X, where X stands for any statistic

4 Overview

This document deals with the following five methods.

a) ISO 5725-based method (proposed by Reference [13]);

b) accordance and concordance (proposed by Reference [3]);

c) ORDANOVA (proposed by Reference [4]);

d) CM-accuracy, sensitivity and specificity;

e) Kappa coefficient (proposed by Reference [15]).

The assumed data structure depends on each method. In methods a), b) and c), each laboratory measures one identical sample multiple times, and the data can be summarized as in Table 1 and/or Table 2; while methods d) and e) are based on many levels of samples, and the data can be summarized as in Table 3.

NOTE 1 Nowadays, the probability for detection (POD) approach is used to analyze binary measurement values, instead of the methods d) and e), but this document deals with only classical methods. There are some ISO documents introducing the POD approach; see References [13] and [15].

Table 1 — Data format for methods a), b) and c)

Laboratory	Rep 1		Rep j		Rep n
Lab 1	y_{11}	$\cdots$	y_{1j}	$\cdots$	y_{1n}
$\vdots$	$\vdots$		$\vdots$		$\vdots$
Lab i	y_{i1}	$\cdots$	y_{ij}	$\cdots$	y_{in}
$\vdots$	$\vdots$		$\vdots$		$\vdots$
Lab L	y_{L1}	$\cdots$	y_{Lj}	$\cdots$	y_{Ln}
NOTE y_{ij} is either 0 or 1, which means negative and positive measured values, respectively.					

Table 2 — Another expression of Table 1

Laboratory	Number of 1	Number of 0
Lab 1	x_1	$n-x_1$
$\vdots$	$\vdots$	
Lab i	x_i	$n-x_i$
$\vdots$	$\vdots$	
Lab L	x_L	$n-x_L$
NOTE n is the number of repetitions in each laboratory, and $x_i \in \{0,\ 1,\ \ldots, n\}$.		

Table 3 — Data format for methods d) and e)

Actual values	Measured values	
	1	**0**
1	c_{11}	c_{12}
0	c_{21}	c_{22}
NOTE $c_{11}, c_{12}, c_{21}, c_{22}$ are non-negative integers.		

5 Examples used in this document

5.1 Case 1: Listeria monocytogenes

This subclause deals with the results of a collaborative study on Listeria monocytogenes, which was quoted and analyzed in Reference [13]. The study consisted of ten laboratories, where each laboratory repeated five-time measurements, that is, $L = 10$, $n = 5$. The results are shown in Table 4. In this table, numbers 1 and 0 mean that Listeria monocytogenes were and were not detected, respectively; and the column Total indicates the total number of detections of Listeria monocytogenes.

Table 4 — Case 1 — The results of a collaborative study on Listeria monocytogenes

Laboratory	Measured value					
	Repetition					Total
Lab 1	1	1	1	1	1	5
Lab 2	1	1	1	1	1	5
Lab 3	1	1	1	1	1	5
Lab 4	1	1	1	1	1	5
Lab 5	0	0	1	1	1	3
Lab 6	1	1	1	1	1	5
Lab 7	0	0	1	1	1	3
Lab 8	1	1	1	1	1	5
Lab 9	1	1	1	1	1	5
Lab 10	1	1	1	1	1	5

5.2 Case 2: Human cell line activation test (h-CLAT)-1

This subclause deals with the human cell line activation test (h-CLAT)[5][6]. The h-CLAT is an alternative to an animal experiment for evaluating the skin sensitization potential.

Reference [7] conducted a collaborative study and reported the results. The study consisted of five laboratories, where each laboratory repeated three-time measurements. Each laboratory measured 21 chemicals, but this document deals with two chemicals out of the 21, denoted by chemical *A* and chemical *B*, which have different pattern results. Case 2(a) reports the results of chemical *A*, shown in Table 5, and Case 2(b) reports the results of chemical *B*, shown in Table 6.

Table 5 — Case 2(a) — Number of detections of the skin sensitization potential of chemical *A* by h-CLAT in three-time measurements

Laboratory	Number of detections in three repetitions
Lab 1	3
Lab 2	3
Lab 3	1
Lab 4	3
Lab 5	3

