INTERNATIONAL STANDARD

ISO 8196-2

IDF 128-2

Second edition
2009-10-01

Milk — Definition and evaluation of the overall accuracy of alternative methods of milk analysis —

Part 2: Calibration and quality control in the dairy laboratory

Lait — Définition et évaluation de la précision globale des méthodes alternatives d'analyse du lait —

Partie 2: Calibrage et contrôle qualité dans les laboratoires laitiers

Reference numbers
ISO 8196-2:2009(E)
IDF 128-2:2009(E)

ISO copyright office
Case postale 56 • CH-1211 Geneva 20
Tel. + 41 22 749 01 11
Fax + 41 22 749 09 47
E-mail copyright@iso.org
Web www.iso.org

International Dairy Federation
Diamant Building • Boulevard Auguste Reyers 80 • B-1030 Brussels
Tel. + 32 2 733 98 88
Fax + 32 2 733 04 13
E-mail info@fil-idf.org
Web www.fil-idf.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.

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 8196-2|IDF 128-2 was prepared by Technical Committee ISO/TC 34, *Food products*, Subcommittee SC 5, *Milk and milk products*, and the International Dairy Federation (IDF). It is being published jointly by ISO and IDF.

This second edition of ISO 8196-2|IDF 128-2 cancels and replaces the first edition (ISO 8196-2:2000), which has been technically revised.

ISO 8196|IDF 128 consists of the following parts, under the general title *Milk — Definition and evaluation of the overall accuracy of alternative methods of milk analysis*:

— *Part 1: Analytical attributes of alternative methods*

— *Part 2: Calibration and quality control in the dairy laboratory*

— *Part 3: Protocol for the evaluation and validation of alternative quantitative methods of milk analysis*

Foreword

IDF (the International Dairy Federation) is a non-profit organization representing the dairy sector worldwide. IDF membership comprises National Committees in every member country as well as regional dairy associations having signed a formal agreement on cooperation with IDF. All members of IDF have the right to be represented at the IDF Standing Committees carrying out the technical work. IDF collaborates with ISO in the development of standard methods of analysis and sampling for milk and milk products.

The main task of Standing Committees is to prepare International Standards. Draft International Standards adopted by the Action Teams and Standing Committees are circulated to the National Committees for voting. Publication as an International Standard requires approval by at least 50 % of IDF National Committees casting a vote.

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

ISO 8196-2|IDF 128-2 was prepared by the International Dairy Federation (IDF) and Technical Committee ISO/TC 34, *Food products*, Subcommittee SC 5, *Milk and milk products*. It is being published jointly by ISO and IDF.

All work was carried out by the Joint IDF-ISO Action Team on *Automated methods* of the Standing Committee on *Quality assurance, statistics of analytical data and sampling* under the aegis of its project leader, Mr. O. Leray (FR).

This edition of ISO 8196-2|IDF 128-2, together with ISO 8196-1|IDF 128-1 and ISO 8196-3|IDF 128-3, cancels and replaces IDF 128:1985, which has been technically revised.

ISO 8196|IDF 128 consists of the following parts, under the general title *Milk — Definition and evaluation of the overall accuracy of alternative methods of milk analysis*:

— *Part 1: Analytical attributes of alternative methods*

— *Part 2: Calibration and quality control in the dairy laboratory*

— *Part 3: Protocol for the evaluation and validation of alternative quantitative methods of milk analysis*

Introduction

The main purpose of this part of ISO 8196|IDF 128 is to provide practical details and recommendations for the calibration of instruments and quality control in routine dairy laboratories, including the checking of compliance with a specification value or limit.

ISO 8196-1|IDF 128-1 is mainly intended for users to assess alternative methods of analysis and gives guidance for routine laboratories using these methods.

This part of ISO 8196|IDF 128 relates directly to ISO 8196-1|IDF 128-1 for the definition of the relevant performance characteristics, for the quantitative evaluation of the overall accuracy and the establishment of relevant standard limit values to comply with in analytical quality assurance as described. The general concepts apply to all analytical methods, but special emphasis is given to rapid physicochemical methods which are currently in use for the compositional testing of milk.

ISO 8196|IDF 128 (all parts) only specifies the single linear regression model as a simplified approach to allow users to determine equivalence of an alternative method with a reference method. However, the linear regression approach is valid as a determination of method equivalence only in limited circumstances or if a high correlation between the results of the reference method and the routine method is achieved. If a high correlation is not achieved, recourse should be made to other data handling and measurement error modelling techniques. Although these techniques are referred to, they are not specified in ISO 8196|IDF 128 (all parts).

Milk — Definition and evaluation of the overall accuracy of alternative methods of milk analysis —

Part 2:
Calibration and quality control in the dairy laboratory

1 Scope

This part of ISO 8196|IDF 128 gives guidelines for the calibration of instruments and quality control procedures for milk analysis in dairy laboratories.

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 8196-1|IDF 128-1:2009, *Milk — Definition and evaluation of the overall accuracy of alternative methods of milk analysis*

ISO 8196-3|IDF 128-3:2009, *Milk — Definition and evaluation of the overall accuracy of alternative methods of milk analysis — Part 3: Protocol for the evaluation and validation of alternative quantitative methods of milk analysis*

3 Terms, definitions, and symbols

3.1 Terms and definitions

For the purposes of this document, the terms and definitions given in ISO 8196-1|IDF 128-1 and the following apply.

3.1.1
standardization of an instrument
experimental evaluation of the exactness of the calibration of an instrument by reference to the true values given either by a reference method or by standard materials or a standard instrument

3.1.2
calibration of an instrument
adjustment of the signal from an instrument so that, at each level of the component, the mean of individual test results given by the instrument closely approximates the true value of the component concentration

IMPORTANT — Even if the term "calibration" is often used for both the standardization and calibration of instruments (see Clause 4), the use of these words according to the definitions in 3.1.1 and 3.1.2 is strongly recommended.

3.2 Symbols

b	linear regression coefficient or slope
$\overline{d}$	average of differences d_i or mean bias
d_i	difference between means of duplicates of x_i and y_i
$\overline{d}_{\text{rel}}$	relative mean bias
$L_{\overline{d}}$	limit of the mean bias $\overline{d}$
$L_{\overline{d}\text{rel}}$	limit of the relative mean bias
m_0	reference value for the control chart
n	number of replicates
P_{xy}	sum of products of x and y
q	number of samples
r	repeatability limit
r_{xy}	correlation coefficient
R	reproducibility limit
S_d	sum of squares of d
S_x	sum of squares of x
S_y	sum of squares of y
s_b	standard deviation of slope
s_{x_0}	standard error of estimate x_0
s_y	estimate of standard deviation of y
s_{yx}	estimate of the standard deviation of accuracy
t_{obs}	observed value of the Student t-test
$t_{1-\alpha/2}$	t-value of the Student distribution for a two-sided probability $1-\alpha$
$t_{1-\alpha}$	t-value of the Student distribution for a one-sided probability $1-\alpha$
$u_{1-\alpha/2}$	value of the standard normal distribution for a two-sided probability $1-\alpha$
$\overline{x}$	arithmetic average of $\overline{x}_i$
$\overline{x}_i$	mean of duplicates of x_i
$\overline{y}$	arithmetic average of $\overline{y}_i$
$\overline{y}_i$	mean of duplicates of y_i
$\overline{y}(\overline{x})$	predicted values from $\overline{x}$ by linear regression.
ν	degree of freedom
σ_r	standard deviation of repeatability
σ_R	standard deviation of reproducibility
σ_y	standard deviation of y
σ_{yx}	standard deviation of accuracy
$\sigma_{yx\text{rel}}$	relative standard deviation of accuracy

4 Calibration of instruments

4.1 General principles

This clause only considers the general principles of calibration which apply to any alternative method of milk analysis.

Detailed and specific instructions for the calibration procedure, as well as for the preliminary checks concerning each group of methods, shall be given in specific documentation.

The calibration here described is based on the assumption of an existing linear relationship between the alternative method and the reference method. The utilization of an ordinary least-squares (OLS) linear regression model with the reference method as dependent variable (y-axis) is recommended as practical and reliable:

a) the measured residual standard deviation is structurally the lowest and is independent of the exactness of calibration, hence it can be used for calibration control as a method accuracy characteristic (see ISO 8196-1|IDF 128-1);

b) prediction of true values from alternative method values and calculation of associated errors is simplified;

c) the differences in measurement results as compared to applying other linear regression models are not significant.

4.2 General procedure

4.2.1 Preliminary checks

4.2.1.1 Instrument checks: all functional checks and adjustments (zero setting) of the instrument specified by the relevant International Standard or the manufacturer should be carried out prior to the analysis.

4.2.1.2 Linearity: unless otherwise stated, the relationship between the instrumental signal readings and the component concentration is linear within the specified range of concentration. Normally, for alternative methods, it is only necessary to check and adjust linearity on new instruments or whenever major parts (e.g. cell or servo-system) are serviced or replaced.

4.2.1.3 Interference corrections (inter-corrections): where instrumental compensations for interfering components are applied to optimize accuracy, these should be checked and eventually adjusted prior to the analysis.

4.2.1.4 Repeatability and accuracy checks: the repeatability and accuracy of the instrument should comply with the specifications of the relevant International Standard.

4.2.1.5 Linearity and interference corrections on a single matrix. In order to check both linearity (4.2.1.2) and interference corrections (4.2.1.3), a set of samples can be prepared from a single milk matrix in such a way as to cover the concentration range of a component while maintaining others constant. For linearity, the component range relates to the measurand and the alignment of measurement results is checked against the theoretical mixing ratio or the relevant concentration calculated from prior analyses. For interference correction, the range concerns the interfering component, and the absence of related induced bias for the measurand requires checking.

4.2.2 Standardization of the instrument

4.2.2.1 Test samples

4.2.2.1.1 General requirements

Milk samples collected especially for that purpose or, when available, standard materials with similar characteristics may be used to standardize the instrument.

In order to obtain the most accurate estimate of the calibration line, the following two major requirements shall be fulfilled:

a) the samples cover the whole range of concentration of the component;

b) the residual standard deviation from the regression is minimal.

Depending on the measurand, this can best be obtained by pooling individual milk samples selected at different levels of concentration with a maximum of samples chosen for values close to the extremities of the range or by separation and recombination of milk components or by spiking and/or aqueous dilution. Prior to use, the equivalence of the alternative calibration samples for the component should have been validated by comparison with usual milk samples.

4.2.2.1.2 Nature of samples

For any calibration, it is of utmost importance that calibration samples are representative of the samples routinely tested, that is same types of milk (i.e. milk from individual animals, herd bulk milk, silo milk, processed milk), submitted to the same treatment and preservation, and originating from the same collection areas at the same periods of time. The period during which the calibration samples remain valid is to be established experimentally, taking into account changes in milk production factors related to season and regional animal feeding practices.

Discard milk of evident poor physical quality so as to guarantee homogeneous intakes and proper functioning of the instrument (suitable milk flow).

The reference method should be applied to the fresh calibration sample, i.e. before eventually treating the samples in the form of sub-samples.

4.2.2.1.3 Range of measurand

Samples collected should cover the whole range of concentrations for the component of interest. The larger the range used for calibration, the more accurate the adjustment of the calibration line. In that respect, it is recommended that the standard deviation of reference results, s_y, complies with Condition (1):

$$s_y \geqslant 5 s_{yx} \tag{1}$$

or, equivalently, the correlation coefficient, r_{xy}, complies with Condition (2)

$$r_{xy} \geqslant 0{,}98 \tag{2}$$

Alternatively, where the accuracy and/or the measurand range cannot meet the recommendation, appropriate alternative calibration samples (4.2.2.1.1) might be considered.

NOTE Equivalent to sample pooling (4.2.2.1.1), pooling of individual data can serve to produce a suitable virtual calibration data set where sufficient correlation cannot be achieved by physical means (see ISO 8196-3|IDF 128-3:2009, 5.2.2.2.3.2).

4.2.2.1.4 Number of samples

4.2.2.1.4.1 The number of samples depends on the objective of the laboratory, the accuracy of the method and the heterogeneity of the sample population for the measurand.

4.2.2.1.4.2 Select the number of samples with respect to composition representativeness for main components.

Define the number of samples needed with respect to the various origins of the test samples and the degree of their representativeness reached by assembly. For instance, in the analysis of composition, the numbers of representative samples recommended for use as a minimum are listed in Table 1.

Table 1 — Origin and number of representative samples

Milk type	No. representative samples
Individual animal	100
Herd bulk	40
Bulk	6
Processed	Set of milk samples produced from the original bulk milk or the bulk material used in the process, for appropriate levels of the measurand

Otherwise, a set of alternative calibration samples of equivalent representativeness prepared from at least nine equidistant measurand concentrations over a range relevant for the commingled milk samples can be used.

NOTE A minimum number of nine concentration levels for alternative calibration samples allows for a regular coverage for main milk components (i.e. fat, protein and lactose) and at the same time sufficient component combinations to check instrument fittings (linearity, interferences).

4.2.2.1.4.3 Ensure the exactness of calibration throughout the range of concentration, i.e., as specified in 4.2.2.1.4.4 and 4.2.2.1.4.5.

4.2.2.1.4.4 For the mean level, choose the minimum sample number so that any calibration bias (so-called mean bias $\bar{d}$) exceeding the limits $\pm L_{\bar{d}}$ set by the user is statistically significant for a risk α of error. From this prerequisite it follows that the limits of uncertainty of the estimated bias should not be larger than the calibration limits stated.

From the standard deviation of accuracy, σ_{yx}, estimated by s_{yx} in a former evaluation (see ISO 8196-1|IDF 128-1 and ISO 8196-3|IDF 128-3) and the previously defined limits of calibration bias $\pm L_{\bar{d}}$, the number q should fulfil Conditions (3):

$$L_{\bar{d}} \geqslant \frac{u_{1-\alpha/2}\,\sigma_{yx}}{\sqrt{q}} \quad \Leftrightarrow \quad q \geqslant \frac{u^{2}_{1-\alpha/2}\,\sigma^{2}_{yx}}{L^{2}_{\bar{d}}} \quad \Leftrightarrow \quad q \geqslant \frac{3{,}84 \times \sigma^{2}_{yx}}{L^{2}_{\bar{d}}} \tag{3}$$

or using the limits $\pm L_{\bar{d}\text{rel}}$ of the relative mean bias $\bar{d}_{\text{rel}}$ and the relative standard deviation of accuracy $\sigma_{yx\text{rel}}$:

$$q \geqslant \frac{u^{2}_{1-\alpha/2}\,\sigma^{2}_{yx\text{rel}}}{L^{2}_{\bar{d}\text{rel}}} \quad \Leftrightarrow \quad q \geqslant \frac{3{,}84 \times \sigma^{2}_{yx\text{rel}}}{L^{2}_{\bar{d}\text{rel}}} \tag{4}$$

where

$u_{1-\alpha/2}$ is the value of standard normal distribution for a probability level $1-\alpha$;

with

$$\overline{d} = \overline{x} - \overline{y} = \overline{x} - \overline{y}(\overline{x})$$

$$\overline{d}_{\text{rel}} = \overline{d} \times 100/\overline{y}$$

$$\sigma_{yx\text{rel}} = (\sigma_{yx}/\overline{y}) \times 100$$

$$\alpha = 0{,}05$$

EXAMPLE

Fat:	σ_{yx} = 0,07 % mass fraction fat	with $L_{\overline{d}}$ = 0,02 % mass fraction fat accepted	$\Rightarrow q \geqslant 49$
Somatic cell count:	$\sigma_{yx\text{rel}}$ = 10 %	with $L_{\overline{d}\text{rel}}$ = 3 % accepted	$\Rightarrow q \geqslant 43$

4.2.2.1.4.5 For the calibration line, the limits for the relative uncertainty of slope b should not exceed the maximum value δb_{rel} defined beforehand by the user for the slope bias, so that any error in excess of this is significant. Thus, with a sample population already known with regard to σ_{yx} and σ_y or the usual correlation coefficient, r_{xy} , the number q should fulfil Conditions (5):

$$q \geqslant u^2_{1-\alpha/2} \times 100^2 \left(\frac{\sigma^2_{yx}}{\sigma^2_y - \sigma^2_{yx}} \right) \frac{1}{\delta b^2_{\text{rel}}} \quad \Leftrightarrow \quad q \geqslant 38\,400 \left(\frac{\sigma^2_{yx}}{\sigma^2_y - \sigma^2_{yx}} \right) \frac{1}{\delta b^2_{\text{rel}}} \tag{5}$$

or equivalently

$$q \geqslant u^2_{1-\alpha/2} \times 100^2 \left(\frac{1}{r^2_{xy}} - 1 \right) \frac{1}{\delta b^2_{\text{rel}}} \quad \Leftrightarrow \quad q \geqslant 38\,400 \left(\frac{1}{r^2_{xy}} - 1 \right) \frac{1}{\delta b^2_{\text{rel}}} \tag{6}$$

with

$$\delta b_{\text{rel}} \geqslant u_{1-\alpha/2}\ \sigma_b \times 100/b = u_{1-\alpha/2} \times 100\ [\sigma_{yx}/(\sigma_y \sqrt{q})]$$

$$\alpha = 0{,}05$$

EXAMPLE

Fat:	σ_y = 0,5 % mass fraction	σ_{yx} = 0,07 % mass fraction	(r_{xy} = 0,990 1)	with a limit δb_{rel} = 4 % $\Rightarrow q \geqslant 48$
Free fatty acids:	σ_y = 0,5 mmol/100 g	σ_{yx} = 0,15 mol/100 g	(r_{xy} = 0,953 9)	with a limit δb_{rel} = 5 % $\Rightarrow q \geqslant 152$

4.2.2.1.5 Number of replicates

Perform sample analysis at least in duplicate.

Where the standard deviation of repeatability of the alternative method is significantly larger than the standard deviation of repeatability of the reference method and the standard deviation of accuracy, the number of replicates can be increased so as to get standard deviations of mean results that at minimum are equivalent for both methods.

Thus the number of required replicates is obtained from Condition (7):

$$n_{\text{alt}} \geqslant n_{\text{ref}} \left(\frac{\sigma_{\text{alt}}}{\sigma_{\text{ref}}} \right)^2 \tag{7}$$

where

n_{ref} is the number of replicates with the reference method;

n_{alt} is the number of replicates with the alternative method;

σ_{ref} is the standard deviation of repeatability of the reference method;

σ_{alt} is the standard deviation of repeatability of the alternative method.

By increasing the number of replicates, the residual standard deviation of the regression is reduced and the correlation coefficient rises concurrently. When the correlation coefficient is greater than or equal to 0,98, more replicates are not necessary. Further improvement can be obtained by increasing the number of samples.

4.2.2.2 Statistical analysis

4.2.2.2.1 Principle

The approach presented is based on an OLS linear regression model with the underlying prerequisite of an approximate constancy of the distribution of the residuals throughout the range of calibration. Generally, this is the case with physicochemical methods covering no more than 1 log scale. Otherwise, transform the data so as to equalize the residual variance throughout the range or split the range into shorter sub-ranges where proper specific calibrations are acceptable. They should be considered separately.

Prior to any calculation, plot the individual mean results of the alternative method, $\bar{x}_i$, and the reference method, $\bar{y}_i$, of the q samples in an xy-axis diagram according to ISO 8196-1|IDF 128-1:2009, Figure A.1. Check the distribution of the data points, which should appear linear, regular, and homogeneous.

Do not take into account samples appearing clearly out of the dot cloud, i.e. deviating significantly from the general expected linear tendency. Consider samples lying outside the confidence belt, regression line $\pm 2{,}58\, s_{yx}$ (which constitute the limits of 99 % of the population of residuals), as suspect. For these samples, check whether their deletion significantly influences the calculated values for slope and intercept. If so, delete them.

4.2.2.2.2 Calculation

4.2.2.2.2.1 Calculate the regression equation, $\bar{y}_i = b\bar{x}_i + a$, from data analysis, using the least-squares method, where $\bar{y}_i$ refers to the reference method, and $\bar{x}_i$ to the instrument (see ISO 8196-1|IDF 128-1).

Calculate the residual standard deviation from the regression, s_{yx}. That value shall be within limits of the specification given by the relevant International Standard for accuracy.

Calculate the correlation coefficient, r_{xy}

$$r_{xy} = \frac{P_{xy}}{(S_x S_y)^{1/2}} \tag{8}$$

That value should not be lower than 0,98.

Then carry out the steps in 4.2.2.2.2.2 to 4.2.2.2.2.4.

4.2.2.2.2.2 Test whether the slope differs statistically from 1,000 by calculating the standard deviation of b:

$$s_b = \left(\frac{s_{yx}^2}{S_x} \right)^{1/2} \tag{9}$$

The slope is correct if in agreement with one of Expressions (10) or (11):

$$b - t_{1-\alpha/2}\, s_b \leqslant 1{,}000 \leqslant b + t_{1-\alpha/2}\, s_b \tag{10}$$

$$t_{\text{obs}} = \frac{|b-1|}{s_b} \leqslant t_{1-\alpha/2} \tag{11}$$

where t is the value of the Student distribution (see ISO 8196-1|IDF 128-1, ISO 3534-1[1] and Table 2). It shows both the random variable, t, and a particular or observed value of this variable, t_{obs}.

4.2.2.2.2.3 Test the null hypothesis that the regression line goes through the centre of gravity of the sample population by calculating the standard deviation of the regression equation, $\overline{y}(\overline{x}) = b\overline{x} + a$, by using Equation (12):

$$s_{\overline{y}(\overline{x})} = \frac{s_{yx}}{\sqrt{q}} \tag{12}$$

where q is the number of samples.

The mean adjustment is correct if in agreement with one of Expressions (13) to (15):

$$\overline{y}(\overline{x}) - t_{1-\alpha/2}\, s_{\overline{y}(\overline{x})} \leqslant \overline{x} \leqslant \overline{y}(\overline{x}) + t_{1-\alpha/2}\, s_{\overline{y}(\overline{x})} \tag{13}$$

$$\left[\overline{x} - \overline{y}(\overline{x})\right] - t_{1-\alpha/2}\, s_{\overline{y}(\overline{x})} \leqslant 0 \leqslant \left[\overline{x} - \overline{y}(\overline{x})\right] + t_{1-\alpha/2}\, s_{\overline{y}(\overline{x})} \tag{14}$$

$$t_{\text{obs}} = \frac{\left|\overline{x} - \overline{y}(\overline{x})\right|}{s_{\overline{y}(\overline{x})}} \leqslant t_{1-\alpha/2} \tag{15}$$

with $q - 2$ degrees of freedom and $\alpha = 0{,}05$.

Adjustment of the calibration is necessary if the slope and/or the mean bias of the regression are found to be different from 1,000 and 0, respectively.

Table 2 — Table of critical values of Student t_p for a probability p and ν degrees of freedom

Degrees of freedom ν	Probability p						
	0,90	0,950	0,975	0,99	0,995	0,999 0	0,999 5
1	3,078	6,314	12,706	31,821	63,657	318,309	636,6
2	1,886	2,920	4,303	6,965	9,925	22,327	31,60
3	1,638	2,353	3,182	4,541	5,841	10,215	12,94
4	1,533	2,132	2,776	3,747	4,604	7,173	8,610
5	1,476	2,015	2,571	3,365	4,032	5,893	6,859
6	1,440	1,943	2,447	3,143	3,707	5,208	5,959
7	1,415	1,895	2,365	2,998	3,499	4,785	5,405
8	1,397	1,860	2,306	2,896	3,355	4,501	5,041
9	1,383	1,833	2,262	2,821	3,250	4,297	4,781
10	1,372	1,812	2,228	2,764	3,169	4,144	4,587
11	1,363	1,796	2,201	2,718	3,106	4,025	4,437
12	1,356	1,782	2,179	2,681	3,055	3,930	4,318
13	1,350	1,771	2,160	2,650	3,012	3,852	4,221
14	1,345	1,761	2,145	2,624	2,977	3,787	4,140
15	1,341	1,753	2,131	2,602	2,947	3,733	4,073
16	1,337	1,746	2,120	2,583	2,921	3,686	4,015
17	1,333	1,740	2,110	2,567	2,898	3,646	3,965
18	1,330	1,734	2,101	2,552	2,878	3,610	3,922
19	1,328	1,729	2,093	2,539	2,861	3,579	3,883
20	1,325	1,725	2,086	2,528	2,845	3,552	3,850
21	1,323	1,721	2,080	2,518	2,831	3,527	3,819
22	1,321	1,717	2,074	2,508	2,819	3,505	3,792
23	1,319	1,714	2,069	2,500	2,807	3,485	3,767
24	1,318	1,711	2,064	2,492	2,797	3,467	3,745
25	1,316	1,708	2,060	2,485	2,787	3,450	3,725
26	1,315	1,706	2,056	2,479	2,779	3,435	3,707
27	1,314	1,703	2,052	2,473	2,771	3,421	3,690
28	1,313	1,701	2,048	2,467	2,763	3,408	3,674
29	1,311	1,699	2,045	2,462	2,756	3,396	3,659
30	1,310	1,697	2,042	2,457	2,750	3,385	3,646
31	1,309	1,696	2,040	2,453	2,744	3,375	3,633
32	1,309	1,694	2,037	2,449	2,738	3,365	3,622
33	1,308	1,692	2,035	2,445	2,733	3,356	3,611
34	1,307	1,691	2,032	2,441	2,728	3,348	3,601
35	1,306	1,690	2,030	2,438	2,724	3,340	3,591
36	1,306	1,688	2,028	2,434	2,719	3,333	3,582
37	1,305	1,687	2,026	2,431	2,715	3,326	3,574
38	1,304	1,686	2,024	2,429	2,712	3,319	3,566
39	1,304	1,685	2,023	2,426	2,708	3,313	3,558
40	1,303	1,684	2,021	2,423	2,704	3,307	3,551
45	1,301	1,679	2,014	2,412	2,690	3,281	3,520
50	1,299	1,676	2,009	2,403	2,678	3,261	3,496
55	1,297	1,673	2,004	2,396	2,668	3,245	3,476
60	1,296	1,671	2,000	2,390	2,660	3,232	3,460
70	1,294	1,667	1,994	2,381	2,648	3,211	3,435
80	1,292	1,664	1,990	2,374	2,639	3,195	3,415
90	1,291	1,662	1,987	2,368	2,632	3,183	3,400
100	1,290	1,660	1,984	2,364	2,626	3,174	3,389
200	1,286	1,653	1,972	2,345	2,601	3,131	3,339
500	1,283	1,648	1,965	2,334	2,586	3,106	3,310
Inf (∞)	1,282	1,645	1,960	2,326	2,576	3,090	3,291

4.2.2.2.2.4 The test in 4.2.2.2.2.3 is equivalent to testing the mean of differences $\overline{d}_i = \overline{x}_i - \overline{y}_i$, or mean bias $\overline{d}$, versus 0 when the slope is not statistically different from 1,000, since $\overline{y}(\overline{x}) = \overline{y}$ and $\overline{d} = \overline{x} - \overline{y} = \overline{x} - \overline{y}(\overline{x})$.

In that case, the adjustment at the average level is correct when in agreement with Condition (16):

$$\overline{d} - t_{1-\alpha/2}\frac{s_d}{\sqrt{q}} \leqslant 0 \leqslant \overline{d} + t_{1-\alpha/2}\frac{s_d}{\sqrt{q}} \tag{16}$$

with $q - 1$ degrees of freedom and $\alpha = 0{,}05$.

If both tests (slope and mean adjustment) are negative, the intercept does not statistically differ from 0 when in agreement with one of Expressions (17) or (18):

$$a - t_{1-\alpha/2}\, s_a \leqslant 0 \leqslant a + t_{1-\alpha/2}\, s_a \tag{17}$$

$$t_{\text{obs}} = \frac{|a|}{s_a} \leqslant t_{1-\alpha/2} \tag{18}$$

with

$$s_a = s_{yx}\left(\frac{1}{q} + \frac{\overline{x}^2}{S_x}\right)^{1/2} \tag{19}$$

4.2.3 Calibration of the instrument

Adjust the calibration slope and intercept so that the observed calibration line fits the theoretical line. For correct calibration setting, follow the manufacturer's instructions and the relevant International Standard.

In all cases, and after the final adjustment has been made, check that the calibration is correct both in slope and intercept, i.e. at low and high levels, by using samples. Ensure that the mean values obtained with both the instrument and the reference method or standard materials are not statistically different.

4.3 Frequency of calibration control

Under routine conditions of use, the standardization of the instrument and its possible calibration readjustment are only necessary when the relationship between the instrument readings and the reference has changed. This is likely to occur:

a) after repair or servicing of the major components of the instrument (i.e. cell, servo, homogenizer, optical filters);

b) when the properties or the composition of the matrix examined has changed (e.g. due to seasonal influences).

Assuming that the composition and the properties of milk remain unchanged within a period of two weeks or one month, it is not necessary to standardize the instrument too frequently (e.g. every week). Such a procedure is costly and, if carried out with too few samples or with samples not obtained specifically for that purpose, it can introduce further errors. This would be due to a poor standardization procedure rather than real variation of the instrument calibration.

4.4 Centralized calibration

When it has been clearly demonstrated that a single calibration of the instrument can be used to analyse samples from various origins without loss in accuracy, a centralized system of calibration is possible. That can be done by a reference laboratory, which should standardize and calibrate a reference or master instrument according to the specified method. Then, by means of suitable standard materials, the calibration of the reference instrument may be transferred to identical instruments in other laboratories.

5 Quality control in a routine dairy laboratory

NOTE This clause deals with all checks that should be performed by a routine dairy laboratory to ensure the quality of its analytical results. These recommendations can be considered as good laboratory practice.

5.1 Verification of repeatability

The first and most frequent check to carry out is that on repeatability. It is the simplest test indicating whether or not the instrument is working properly.

Check that, for repeatability conditions, the standard deviation of repeatability for the instrument is in agreement with the specification given by the relevant International Standard.

Select a set of q samples ($q = 20$) of milk or milk product in good physicochemical condition covering a wide composition range. Analyse these samples consecutively twice in random order so as to take into account any carry-over effect. Record the absolute difference between duplicate results, w_i. Calculate the standard deviation of repeatability, s_r, using Equation (20):

$$s_r = \left(\frac{1}{2q} \sum_{i=1}^{q} w^2{}_i \right)^{1/2} \tag{20}$$

If using a smaller number of samples, repeat the determination at least three times. Calculate s_r by using one-way analysis of variance. The variance of error, s_e^2, or within-sample variance, is an unbiased estimator of σ_r^2.

Action should be taken if s_r is larger than the specified value.

5.2 Daily check on short-term stability of the instrument

5.2.1 Objective

This daily check is typical for any automated alternative method. Instability of the instrument signal may arise from several origins, e.g. electric drift, temperature variation, and fouling of cell walls.

Check, by analysing regularly one or more control samples, whether the results remain within accepted tolerances, while assuming that no changes in the major physicochemical characteristics of the control sample occur during the checking period. This test is useful not only for checking the instrument stability during a working day, but also from day to day between two standardizations of the instrument against the reference method.

5.2.2 Procedure

5.2.2.1 Select one sample of average composition or, preferably, two samples with a low and high content of the measurand using a selected bulk milk of good physicochemical quality. This can be commercial pasteurized milk, in which the range of the measurand has been extended by the addition of quantities of the component measured (e.g. skim milk powder, cream, or cell concentrate).

Depending on the measurand, apply, where needed, an appropriate heat treatment. Prepare carefully, under constant agitation, as many milk portions as required for one or more working days. Store them with a suitable preservative at 4 °C. As such, good quality pasteurized milk, preserved for instance with bronopol (2-bromo-2-nitropropan-1,3-diol), can safely be stored for two weeks. With instruments having a homogenizer, use homogenized milk only when the homogenization efficiency is checked separately.

Analyse control samples on a regular basis at least three times per hour or after every 100 samples, whichever is the more frequent.

To monitor the quality of the whole analytical procedure including the instrument stability, set up a control chart according to the principles set out in 5.2.2.2 to 5.2.2.4.

5.2.2.2 Determine carefully the average pilot value, m_0, of the control sample and the day-to-day or within-day standard deviation of reproducibility, σ_R, of the method. If unknown, σ_R may roughly be estimated as twice the value of the standard deviation of repeatability.

5.2.2.3 The control chart (see Figure 1) shall represent:

a) a straight line in the centre corresponding to the reference value, m_0;

b) a lower and an upper "confidence belt" corresponding to the 1 – α probability of the two-sided confidence limits of the cumulative mean, m, of the control sample results — calculate these limits by using Formula (21):

$$\frac{m_0 \pm u_{1-\alpha/2}\,\sigma_R}{\sqrt{n}} \tag{21}$$

where

u is a particular value of the standardized normal random variable U, whose value depends on the probability level, 1 – α,

n is the number of control samples analysed,

α is the probability of rejecting the null hypothesis ($m = m_0$) although being correct — this would indicate the need to readjust the instrument when in fact this is not necessary — to avoid a too frequent and unnecessary adjustment, a 0,01 probability level can be used, in which case, $u_{1-\alpha/2} = 2{,}58$;

c) a lower and an upper "individual line" corresponding to the 1 – α probability of the two-sided statistical tolerance interval of individual tests — calculate these limits by using Formula (22):

$$m_0 \pm k_{(n,p,1-\alpha/2)}\,\sigma_R \tag{22}$$

where k is a coefficient determined for n number of samples and the probability level, 1 – α, that the statistical interval contains at least a proportion p of the population.

NOTE For an infinite value of n, with 1 – α = 0,99 and p = 99 %, k = 2,58.

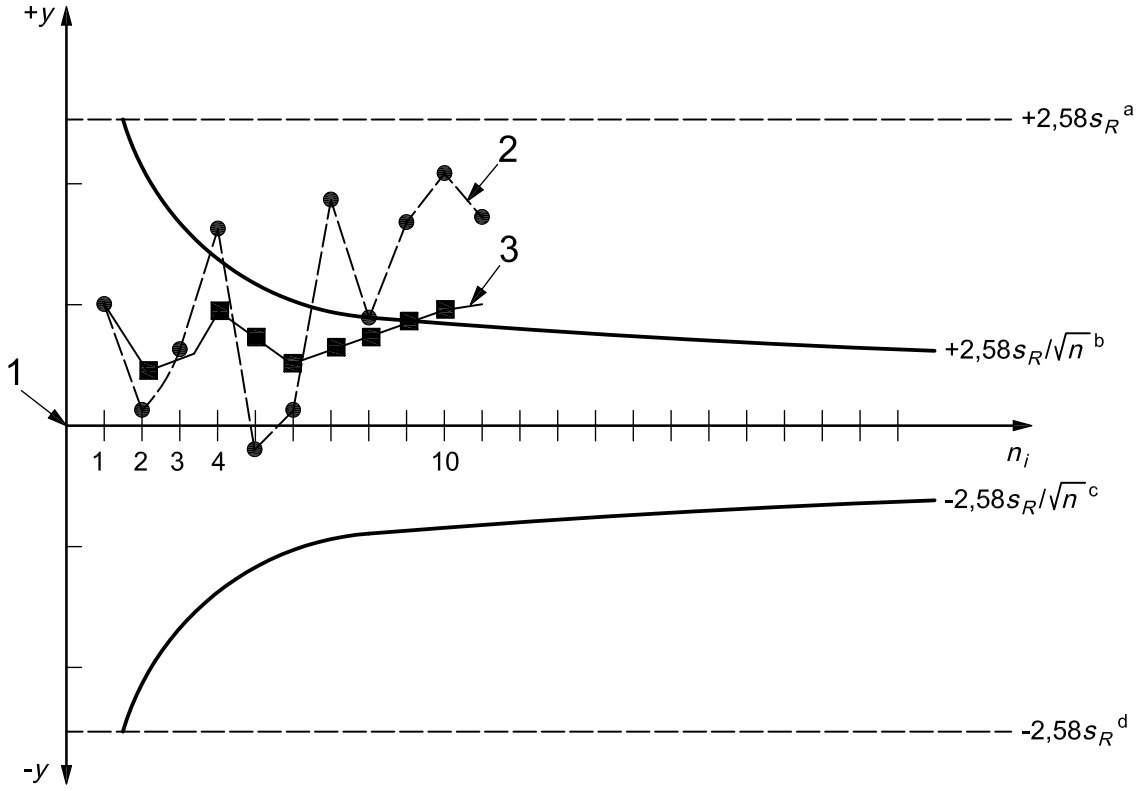
5.2.2.4 Plot on the control chart each individual result of the control sample and the arithmetic mean, m, of the results of the samples being analysed.

Take action when:

a) the arithmetic mean, m, is, for two consecutive samples, outside the same (upper or lower) confidence belts, indicating that the instrument is drifting — the deviation should normally be in the same direction as those of the individual results outside the corresponding individual line;

b) individual results fall frequently near or outside individual lines indicating a poor repeatability of the instrument or poor quality of the sample.

In each case, stop and check the instrument functions and, if necessary, readjust the calibration. The use of a microcomputer and automatic data capture system can be very helpful. However, automatic correction of results shall be exercised with caution in order to avoid giving "correct" results while the instrument is not working properly. After readjustment of the instrument, prepare a new control chart.



Key

1 reference value
2 individual results
3 arithmetic mean of results

n_i sample number in time order
y instrumental value

NOTE 1 Confidence belts relate to the arithmetic mean of the cumulative mean results, $\bar{x}_i$.

NOTE 2 Individual lines relate to the individual results.

NOTE 3 Limits contain 99 % of the population with a 99 % probability.

a Upper individual line.
b Upper confidence belt.
c Lower confidence belt.
d Lower individual line.

Figure 1 — Control chart model for alternative method of analysis

5.3 Verification of bias between laboratories

5.3.1 General

Checking differences between laboratories, regardless of their values as compared to the true value, is part of the laboratory performance checks. It is done through interlaboratory trials in which participating laboratories are asked to analyse a set of samples with the same or, sometimes, a different method of analysis.

Aside from repeatability errors, differences between laboratories originate from the theoretical reproducibility of the method and from differences in calibration. Instruments can be calibrated by means of reference samples or by comparison with a reference method, each laboratory using its own set of samples. In the latter case, checking laboratory performance through an interlaboratory study may not be very meaningful if the origin of the samples has a significant influence on the calibration of the instruments.

5.3.2 Procedure

Check laboratory bias in accordance with the method described in ISO 8196-1|IDF 128-1.

When the alternative method is almost unaffected by the origin of the samples, best results are obtained by supplying reference samples to all participants for calibration purposes.

5.4 Verification of the difference between reference and alternative method results

5.4.1 Objective

This verification is to determine the largest difference which may be accepted between results obtained by a reference method and an alternative method for a single sample or for a population of samples, provided that calibration fitting is optimized. It corresponds to the statistical tolerance limits of the method and includes both random (precision) and systematic (trueness) errors.

5.4.2 Procedure

Three sources of error may arise to explain differences between reference and alternative method results. For each kind of error, corresponding confidence intervals can be calculated as specified in 5.4.3 to 5.4.7.

5.4.3 Precision error

If n replications are performed by one laboratory, the two-sided confidence interval for the mean, $\overline{x}_i$, is:

$$\overline{x}_i \pm t_{1-\alpha/2}\left[s_R^2 - \left(1 - \frac{1}{n}\right)s_r^2\right]^{1/2} \tag{23}$$

It should be noted that s_R is the within-day standard deviation of reproducibility (5.2.2.2) determined in ISO 8196-3|IDF 128-3. Where unknown, replace it by the s_R of the method as determined in the interlaboratory study (see ISO 8196-1|IDF 128-1).

5.4.4 Accuracy error

According to the definition of accuracy, the confidence interval of the difference between $\overline{x}_i$ and the reference value $\overline{y}_i$ is:

$$(\overline{x}_i - \overline{y}_i) \pm t_{1-\alpha/2}\, s_{yx} \tag{24}$$

It should be noted that s_{yx} is the residual standard deviation estimated with routine samples. Do not use residual standard deviation values measured with alternative calibration samples here.

5.4.5 Calibration error

The confidence interval of the calibration line for every estimate, $\overline{y}_i(\overline{x}_i) = \overline{x}_i$, is given by Expression (25):

$$\overline{x}_i \pm t_{1-\alpha/2}\, s_{yx} \left[\frac{1}{q} + \frac{(\overline{x}_i - \overline{x})^2}{S_x} \right]^{1/2} \tag{25}$$

The error corresponding to the confidence interval of the line, including the errors of both the mean bias, $\overline{d} \pm t_{1-\alpha/2}\, s_{yx}/\sqrt{q}$, and the slope, $b \pm t_{1-\alpha/2}\, s_b$, may be considered negligible in appropriate calibration conditions (4.2.2).

NOTE The symbol s_{yx} represents the residual standard deviation estimate with calibration samples. If calibration is carried out with alternative samples instead of routine-like samples, s_{yx} is then also obtained with the alternative samples.

5.4.6 Interpretation of results with single samples

Assuming that the instrument is correctly calibrated, the variance of the overall error is the sum of both the precision and accuracy errors. Then the overall confidence limits of an estimate corresponding to the mean of n replications in a calibration carried out with q samples are determined by using Formula (26):

$$\overline{x}_i \pm t_{1-\alpha/2} \left\{ \left[s_R^2 - \left(1 - \frac{1}{n} \right) s_r^2 \right] + s_{yx}^2 \left[1 + \frac{1}{q} + \frac{(\overline{x}_i - \overline{x})^2}{S_x} \right] \right\}^{1/2} \tag{26}$$

As appropriate conditions allow calibration error to be neglected, the variance of the overall error is the sum of both the precision and accuracy errors, which can then be expressed by using simplified Formula (27):

$$\overline{x}_i \pm t_{1-\alpha/2} \left\{ \left[s_R^2 - \left(1 - \frac{1}{n} \right) s_r^2 \right] + s_{yx}^2 \right\}^{1/2} \tag{27}$$

where $t_{1-\alpha/2} \approx 1{,}96$ for $\alpha = 0{,}05$ for large degrees of freedom in estimating the components of the error.

If a result falls outside these limits, the error can arise from incorrect calibration and/or abnormal composition.

NOTE In the particular case of modified calibration samples, s_{yx} value ($s_{yx,\mathrm{cal}}$) of calibration samples is lower than routine sample value ($s_{yx,\mathrm{rout}}$), thus then the confidence limits are:

$$\overline{x}_i \pm t_{1-\alpha/2} \left\{ \left[s_R^2 - \left(1 - \frac{1}{n} \right) s_r^2 \right] + s_{yx,\mathrm{rout}}^2 + s_{yx,\mathrm{cal}}^2 \left[\frac{1}{q} + \frac{(x_0 - \overline{x})^2}{S_x} \right] \right\}^{1/2} \approx$$

$$\overline{x}_i \pm t_{1-\alpha/2} \left\{ \left[s_R^2 - \left(1 - \frac{1}{n} \right) s_r^2 \right] + s_{yx,\mathrm{rout}}^2 \right\}^{1/2} \tag{28}$$

5.4.7 Interpretation of results with a population of samples

Checking the average difference between alternative method results and reference values is part of a laboratory performance check and is often considered as one of the most important.

It corresponds to checking the trueness of the instrument and can be done through a procedure similar to that being used for checking the calibration. The precision error is now taken as negligible.

Let $\overline{d}$ stand for the mean of the differences d_i between the reference and the instrument results obtained with q samples.

Test the null hypothesis that the population mean, $\bar{d}$, is equal to zero by using one of Equations (29) or (30):

$$t_{\text{obs}} = \left|\bar{d}\right| \frac{\sqrt{q}}{s_d} \tag{29}$$

$$t_{\text{obs}} = \left|\bar{d}\right| \left[\frac{q(q-1)}{\sum (d_i - \bar{d})^2} \right]^{1/2} \tag{30}$$

The hypothesis is accepted if $t_{\text{obs}} \leqslant t_{1-\alpha/2}$

Calculate the standard deviation, s_d, using Equation (31):

$$s_d = \left[\frac{\sum (d_i - \bar{d})^2}{q-1} \right]^{1/2} \tag{31}$$

The value of s_d should be less than that of σ_{yx} given by the relevant International Standard for the kind of milk product under test.

If $\bar{d}$ is non-zero (hypothesis rejected), check and readjust the calibration. If the calibration is not under the responsibility of the operator and if $\bar{d}$ is non-zero, the method is tainted with a systematic error which results in increased uncertainty of results.

If $s_d > \sigma_{yx}$ estimated from the results, the slope is incorrect or abnormal samples are present.

5.5 Verification of the compliance of alternative method results with a compositional requirement

5.5.1 Objective

The objective is to determine the critical difference, CD_α, which may be accepted between results of an alternative method for a single sample and a defined critical limit value, CL_α, of the measurand concentration with a given risk of error, α. It refers to the particular case in which the alternative method is used for sample screening in order to check its compliance with compositional requirements or specification limits for official purposes. Then the largest difference is accepted, checking the compliance with a target value, or the lowest difference is accepted when the alternative method result passes neither the upper nor the lower limit.

NOTE The uncertainty component related to the precision error of the reference method is an additional part of the overall uncertainty of analytical results not taken into account in the following statistics.

5.5.2 Procedure

5.5.2.1 Statistical basis

The standard error, s_{x_0}, of a mean result, x_0, of n replicates measured with a method calibrated with q samples (levels) distributed with $\sigma_x = (S_x/q)^{1/2}$ is given by Equation (32):

$$s_{x_0} = \left\{ \left[s_R^2 - \left(1 - \frac{1}{n}\right) s_r^2 \right] + s_{yx}^2 \left[1 + \frac{1}{q} + \frac{(x_0 - \bar{x})^2}{S_x} \right] \right\}^{1/2} \tag{32}$$

For a particular case of screening, calibration is generally optimized at the measurand level where compliance is checked. Hence the mean level, $\bar{x}$, is chosen close to the limit, L. Since in that case $x_0 - \bar{x}$ is small or zero and 1/q negligible compared to 1, Equation (32) simplifies to Equation (33):

$$s_{x_0} = \left\{ \left[s_R^2 - \left(1 - \frac{1}{n} \right) s_r^2 \right] + s_{yx}^2 \right\}^{1/2} \tag{33}$$

When n = 1 (single measurement), this simplifies further to Equation (34):

$$s_{x_0} = (s_R^2 + s_{yx}^2)^{1/2} \tag{34}$$

5.5.2.2 Compliance of the result of a single sample with a defined target value, X

The unknown true value of the sample estimated by the result, x_0, should not differ from the defined value, X, hence the latter should be found within the confidence interval of x_0. The risk, α, of making an erroneous conclusion is split into equal parts for both sides (two-sided risk, α/2) as follows:

$$x_0 - t_{1-\alpha/2}\, s_{x_0} \leqslant X \leqslant x_0 + t_{1-\alpha/2}\, s_{x_0} \quad \Leftrightarrow \quad X - t_{1-\alpha/2}\, s_{x_0} \leqslant x_0 \leqslant X + t_{1-\alpha/2}\, s_{x_0} \tag{35}$$

or

$$\mathrm{CL}_{\mathrm{U},\alpha/2} = X + t_{1-\alpha/2}\, s_{x_0}$$

and

$$\mathrm{CL}_{\mathrm{L},\alpha/2} = X - t_{1-\alpha/2}\, s_{x_0}$$

In that case, $|x_0 - X| \leqslant \mathrm{CD}_\alpha = t_{1-\alpha/2}\, s_{x_0}$ where $t_{1-\alpha/2}$ = 1,96 for α = 0,05.

Any routine values within the limits [$\mathrm{CL}_{\mathrm{L},\alpha/2}$, $\mathrm{CL}_{\mathrm{U},\alpha/2}$] — or differing from X by less than the critical difference, CD_α — are not considered to be statistically different from the defined value, X. The risk of being wrong when considering a value to differ from X if outside the limits is α.

5.5.2.3 Compliance of the result of a single sample with a defined upper or lower limit

5.5.2.3.1 Upper limit, U

The unknown true value of the sample estimated by the result, x_0, should not exceed a defined upper limit, U, which means that the larger part of possible values for x_0, i.e. the 1 – α lower part, has to be lower than U with a remaining risk, α, of making an erroneous conclusion (one-sided risk, α):

$$x_0 + t_{1-\alpha}\, s_{x_0} \leqslant U \quad \Leftrightarrow \quad x_0 \leqslant \mathrm{CL}_{\mathrm{U},\alpha} = U - t_{1-\alpha}\, s_{x_0} \tag{36}$$

In that case, $U - x_0 \geqslant \mathrm{CD}_\alpha = t_{1-\alpha} s_{x0}$ where $t_{1-\alpha}$ = 1,645 for α = 0,05.

For any routine value less than or equal to the critical limit $\mathrm{CL}_{\mathrm{U},\alpha}$ — not closer to U than the critical difference CD_α — the risk of error that the unknown true value is greater than or equal to the specified maximum limit U does not exceed α.

5.5.2.3.2 Lower limit, L

The unknown true value of the sample estimated by the result, x_0, should not be lower than a defined lower limit, L, which means that the larger part of possible values for x_0, i.e. the $1 - \alpha$ upper part, has to be higher than L with a remaining risk, α, of making an erroneous conclusion (one-sided risk, α):

$$x_0 - t_{1-\alpha}\, s_{x_0} \geqslant L \quad \Leftrightarrow \quad x_0 \geqslant \mathrm{CL}_{\mathrm{L},\alpha} = L + t_{1-\alpha}\, s_{x_0} \tag{37}$$

In that case, $x_0 - L \geqslant \mathrm{CD}_\alpha = t_{1-\alpha} s_{x_0}$ where $t_{1-\alpha} = 1{,}645$ for $\alpha = 0{,}05$.

For any routine value greater than or equal to the critical limit, $\mathrm{CL}_{\mathrm{L},\alpha}$ — not closer to L than the critical difference CD_α — the risk of error that the unknown true value is less than or equal to the specified lower limit, L, does not exceed α.

6 Examples

6.1 Calculation of statistical values

Milk samples numbering 10 were analysed in duplicate with an indirect method and a reference method for fat determination. For simplification, a smaller number of samples than really required is used in this example. Data and computation are given in Table 3.

Table 3 — Results from fat determination of 10 milk samples

Sample No.	Indirect method g/l		Arithmetic mean $\bar{x}_i$	Range w_i	Reference method — mean of duplicate $\bar{y}_i$	Difference $d_i = \bar{x}_i - \bar{y}_i$
	1st test	2nd test	g/l	g/l	g/l	g/l
1	25,9	26,1	26,0	0,2	27,5	− 1,5
2	28,0	28,6	28,3	0,6	28,6	− 0,3
3	28,5	28,5	28,5	0	29,2	− 0,7
4	31,3	31,5	31,4	0,2	32,2	− 0,8
5	33,4	33,6	33,5	0,2	33,5	0
6	35,7	36,1	35,9	0,4	36,0	− 0,1
7	36,6	36,5	36,6	0,1	36,0	+ 0,6
8	40,1	40,0	40,0	0,1	38,2	+ 1,8
9	40,4	41,0	40,7	0,6	40,2	+ 0,5
10	42,8	42,8	42,8	0	41,1	+ 1,7

a) Calculations from the observed values listed in Table 3:

$$\sum \overline{x}_i = 343{,}70 \qquad \sum \overline{y}_i = 342{,}5 \qquad q = 10 \qquad \sum d_i = +1{,}2$$

$$\sum \overline{x}_i^2 = 12\ 114{,}05 \qquad \sum \overline{y}_i^2 = 11\ 942{,}43 \qquad \sum \overline{x}_i \overline{y}_i = 12\ 023{,}13 \qquad \sum d_i^2 = 10{,}22$$

$$\overline{x} = 34{,}37 \qquad \overline{y} = 34{,}25 \qquad \frac{\sum x_i \sum y_i}{q_0} = 11\ 771{,}725 \qquad \overline{d} = +0{,}12$$

$$\frac{\left(\sum x_i\right)^2}{q} = 11\ 812{,}969 \qquad \frac{\left(\sum y_i\right)^2}{q} = 11\ 730{,}625 \qquad \frac{\left(\sum d_i\right)^2}{q} = 0{,}144$$

b) Calculation of sum of squares, S, by the equations:

$$S_x = \sum (\overline{x}_i - \overline{x})^2 = \sum \overline{x}_i^2 - \frac{\left(\sum \overline{x}_i\right)^2}{q} = 301{,}081$$

$$S_y = \sum (\overline{y}_i - \overline{y})^2 = \sum y_i^2 - \frac{\left(\sum \overline{y}_i\right)^2}{q} = 211{,}805$$

$$S_d = \sum (d_i - \overline{d})^2 = \sum d_i^2 - \frac{\left(\sum d_i\right)^2}{q} = 10{,}076$$

c) Calculation of sum of products, P, of the deviation by:

$$P_{xy} = \sum (\overline{x}_i - \overline{x})(\overline{y}_i - \overline{y}) \sum \overline{x}_i \overline{y}_i - \frac{\sum \overline{x}_i \sum \overline{y}_i}{q} = 251{,}405$$

d) Calculation of the correlation coefficient, r_{xy} [see Equation (8)]:

$$r_{xy} = \frac{P_{xy}}{(S_x\ S_y)^{1/2}} = \frac{251{,}405}{(301{,}081 \times 211{,}80)^{1/2}} = 0{,}996$$

e) Calculation of the slope, b, by:

$$b = \frac{P_{xy}}{S_x} = \frac{251{,}405}{301{,}081} = 0{,}835$$

f) Calculation of the intercept, a, by:

$$a = \overline{y} - b\overline{x} = 34{,}25 - 0{,}835 \times 34{,}37 = +5{,}55$$

g) Calculation of the predicted value from $\overline{x}$ by:

$$\overline{y}(\overline{x}) = b\overline{x} + a = \overline{y} = 34{,}25$$

h) Calculation of the residual standard deviation, s_{yx}, by:

$$s_{yx} = \left[\frac{1}{q-2}\left(S_y - \frac{P_{xy}^2}{S_x}\right)\right]^{1/2} = \left[\frac{1}{8}\left(211{,}805 - \frac{251{,}405^2}{301{,}081}\right)\right]^{1/2} = 0{,}485$$

i) Calculation of the standard deviation of the differences, s_d, by one of the following equations:

$$s_d = \left[\frac{1}{q-1}\left(S_y + S_x - 2P_{xy}\right)\right]^{1/2} = \left[\frac{1}{9}\left(211{,}805 + 301{,}081 - 2 \times 251{,}405\right)\right]^{1/2} = 1{,}058$$

or

$$s_d = \left(\frac{S_d}{q-1}\right)^{1/2} = \left(\frac{10{,}076}{9}\right)^{1/2} = 1{,}058$$

6.2 Determination of repeatability

A simplified equation, ISO 8196-1|IDF 128-1:2009, Equation (3), is used:

$$s_r = \left(\frac{1}{2q}\sum_{i=1}^{q} w_i^2\right)^{1/2} = \left(\frac{1}{20} \times 1{,}02\right)^{1/2} = 0{,}226$$

Therefore, the repeatability limit of the determination, r, in grams per litre, is given by:

$$r = 2{,}83 \times 0{,}226 = 0{,}64$$

6.3 Verification of the exactness of the calibration

The correlation coefficient, $r_{xy} = 0{,}996$, assures the adequacy of the calibration data set, as this is larger than the limit 0,98.

The observed regression line is obtained by:

$$\bar{y}_i = b\bar{x}_i + a$$

Thus

$$\bar{y}_i = 0{,}835\bar{x}_i + 5{,}55$$

a) Compare the slope with the theoretical value of 1,000.

The standard deviation of b is given by [see Equation (9)]:

$$s_b = \left(\frac{s_{yx}^2}{S_x}\right)^{1/2} = \left(\frac{0{,}234\,9}{301{,}081}\right)^{1/2} = 0{,}027\,9$$

For $\alpha = 0{,}05$ and $\nu = 8$, the t-value is 2,306. Thus, the true value of the slope is within $b \pm 2{,}306\ s_b$.

The interval of the slope is then:

$$0{,}711 \leqslant b \leqslant 0{,}899$$

and, according to 4.2.2.2, $t_{obs} = 5{,}91$, i.e. greater than t (= 2,306).

That means that the slope differs from the theoretical value of 1,000 and, therefore, the instrument calibration shall be changed accordingly.

Verify whether the regression line goes through the theoretical mean point of the sample population. The standard deviation is given by:

$$s_{\bar{y}(\bar{x})} = \frac{s_{yx}}{\sqrt{q}} = \frac{0{,}485}{\sqrt{10}} = 0{,}153$$

Using the same t-value as above, the mean of reference values is within the limits:

$$\bar{y}(\bar{x}) \pm 2{,}306\ s_{\bar{y}(\bar{x})}$$

Thus the interval of the mean is:

$$33{,}90 \leqslant \bar{y}(\bar{x}) \leqslant 34{,}60$$

and the mean value, $\bar{x} = 34{,}37,$ belongs to this interval.

Since $\bar{y}(\bar{x}) = \bar{y},$ the mean bias is within the limits:

$$(\bar{x} - \bar{y}) \pm 2{,}306\ s_{\bar{y}(\bar{x})}$$

Thus the interval of the mean bias is

$$-0{,}23 \leqslant \bar{d} \leqslant 0{,}47$$

which includes the theoretical zero value and, according to 4.2.2.2, $t_{obs} = 0{,}78$, i.e. less than t (= 2,306).

That also means that the mean of the alternative method values does not differ significantly from the mean of the reference values and the mean of differences does not differ significantly from zero.

b) Verify that the regression line goes through the origin.

The standard deviation of a is given by [see Equation (19)]:

$$s_a = s_{yx}\left(\frac{1}{q} + \frac{\bar{x}^2}{S_x}\right)^{1/2} = 0{,}485\left(\frac{1}{10} + \frac{34{,}37^2}{301{,}081}\right)^{1/2} = 0{,}973$$

Using the same t-value as mentioned above, the estimate of a is within the limits $\pm 2{,}306\ s_a$. Thus the interval of a is:

$$3{,}31 \leqslant a \leqslant 7{,}79$$

And, according to 4.2.2.2, $t_{obs} = 5{,}70$, i.e. greater than t (= 2,306).

This means that the intercept value of 5,55 differs significantly from zero.

6.4 Determination of accuracy

The actual accuracy of estimate of the instrument is within the limits $\pm 2{,}306\, s_{yx}$. Thus the accuracy limits of the determination, in grams per litre, are calculated as:

$$\pm 2{,}306 \times 0{,}485 = \pm 1{,}12$$

6.5 Checking the difference between reference and alternative method results

6.5.1 Individual samples: the confidence interval of the alternative method results with a 95 % probability is within the limits:

$$\pm 1{,}96\left[s_R^2 - \left(1 - \frac{1}{n}\right) s_r^2 + s_{yx}^2 \right]^{1/2}$$

The number of replicates, $n = 2$, the variance of repeatability, $s_r^2 = 0{,}051$, and the variance of reproducibility is estimated at $s_R^2 = 0{,}204$.

Thus the confidence interval is calculated as:

$$\pm 1{,}96\left[0{,}204 - \left(1 - \frac{1}{2}\right) 0{,}051 + 0{,}235 \right]^{1/2} = \pm 1{,}26$$

That means that if the difference between the reference values and the mean of duplicate determinations of a sample is greater than 1,26, it may be assumed with 95 % confidence that the calibration instrument is incorrect or the sample composition is abnormal.

6.5.2 Population of q samples: to check the trueness of the instrument, calculate the mean and standard deviation of the arithmetic differences, d_i, between the instrument and the reference results.

These calculated values are

$$\bar{d} = +0{,}12$$

$$s_d = 1{,}058$$

Test the null hypothesis that the mean bias is equal to 0, by [see Equation (29)]:

$$t_{\text{obs}} = \left|\bar{d}\right| \frac{\sqrt{q}}{s_d} = 0{,}12\, \frac{\sqrt{10}}{1{,}058} = 0{,}359$$

For $\alpha = 0{,}05$ and $\nu = 9$, the value of t is 2,262, i.e. greater than t_{obs} (= 0,359). Thus the null hypothesis is not rejected.

6.6 Conclusions

On average, the results from the alternative method are not statistically different from the results of the reference method. However, the standard deviation of the differences (1,058) is much higher than the accepted value of the standard deviation of accuracy (0,485). That means that the slope of the instrument is incorrect and that the actual calibration line crosses the theoretical line close to the mean reference values.

Nevertheless, the slope differs significantly from 1,000 and the standard deviation of the differences is higher than the standard deviation of the residuals, making the applied test less powerful than that in 4.2. The latter is more appropriate when b is significantly different from 1,000.

NOTE For convenience, when referring to true values, statistical parameters are written in full, i.e. “slope” and “mean bias”, and when referring to estimates, are identified by b and $\bar{d}$.

6.7 Verification of compliance with target values and upper or lower limits

6.7.1 General

The variance of repeatability, $s_r^2 = 0{,}051$, and the variance of reproducibility is estimated at $s_R^2 = 0{,}204$, assuming $s_R = 2s_r$ and that the defined value is near to the mean level of calibration.

The standard error of an estimated value, x_0, is calculated by:

$$s_{x0} = \left[s_R^2 - \left(1 - \frac{1}{n}\right) s_r^2 + s_{yx}^2 \right]^{1/2}$$

Thus, for a single measurement result, $n = 1$, the standard error is equal to:

$$s_{x0} = \left[0{,}204 - (1-1)\,0{,}051 + 0{,}235 \right]^{1/2} = 0{,}66$$

6.7.2 Compliance of x_0 with a defined value X = 35 g/l

$$x_0 - 1{,}33 \leqslant 35{,}00 \leqslant x_0 + 1{,}33 \quad \Leftrightarrow \quad 33{,}67 \leqslant x_0 \leqslant 36{,}33$$

$$\mathrm{CL}_{\mathrm{U},\alpha/2} = 36{,}33$$

$$\mathrm{CL}_{\mathrm{L},\alpha/2} = 33{,}67$$

or

$$|x_0 - 35{,}00| \leqslant \mathrm{CD}_\alpha = 1{,}33$$

where $t_{1-\alpha/2} = 1{,}96$ for $\alpha = 0{,}05$.

Any routine values that are within the limits calculated, 33,67 to 36,33 — or differing from X by less than the critical difference 1,33 — are considered not statistically different from the defined value, X = 35. The risk of being wrong in considering it different from 35 if outside the limits is less than or equal to 5 %.

6.7.3 Compliance of the result of a single sample with a defined maximum or minimum limit

6.7.3.1 Upper limit, $U = 35{,}00$ g/l

$$x_0 + 1{,}09 \leqslant 35{,}00 \quad \Leftrightarrow \quad x_0 \leqslant \mathrm{CL}_{\mathrm{U},\alpha} = 3{,}91$$

or

$$35{,}00 - x_0 \geqslant \mathrm{CD}_\alpha = 1{,}09$$

where $t_{1-\alpha} = 1{,}645$ for $\alpha = 0{,}05$.

For any routine values less than or equal to the critical limit of 33,91 the risk of error that the unknown true value is greater than or equal to the specified upper limit U does not exceed 5 %.

6.7.3.2 Lower limit, $L = 35{,}00$ g/l

$$x_0 - 1{,}09 \geqslant 35{,}00 \quad \Leftrightarrow \quad x_0 \geqslant \mathrm{CL}_{\mathrm{L},\alpha} = 36{,}09$$

or

$$x_0 - 35{,}00 \geqslant \mathrm{CD}_{\alpha} = 1{,}09$$

where $t_{1-\alpha} = 1{,}645$ for $\alpha = 0{,}05$.

For any routine values greater than or equal to the critical limit of 36,09 the risk of error that the unknown true value is less than or equal to the specified lower limit L does not exceed 5 %.

Bibliography

[1] ISO 3534-1, *Statistics — Vocabulary and symbols — Part 1: General statistical terms and terms used in probability*

ICS 67.100.10

Price based on 25 pages