

JIS

JAPANESE
INDUSTRIAL
STANDARD

Translated and Published by
Japanese Standards Association

JIS Z 8402-6 : 1999
(ISO 5725-6 : 1994)

**Accuracy (trueness and precision) of
measurement methods and results
— Part 6: Use in practice of
accuracy values**

ICS 03.120.30 ; 17.020

Descriptors : measurement, datum, precision, accuracy, reproducibility

Reference number : JIS Z 8402-6 : 1999 (E)

Z 8402-6 : 1999 (ISO 5725-6:1994)

Foreword

This translation has been made based on the original Japanese Industrial Standard established by the Minister of International Trade and Industry through deliberations at the Japanese Industrial Standards Committee in accordance with the Industrial Standardization Law. By this establishment, **JIS Z 8402 : 1991** was withdrawn and replaced with this Standard.

This Standard has been prepared based on the first edition of **ISO 5725-6** issued in 1994.

JIS Z 8402 : 1999 consists of the following 6 parts under the title "*Accuracy (trueness and precision) of measurement methods and results*".

- Part 1: General principles and definitions
- Part 2: Basic method for the determination of repeatability and reproducibility of a standard measurement method
- Part 3: Intermediate measures of the precision of a standard measurement method
- Part 4: Basic methods for the determination of the trueness of a standard measurement method
- Part 5: Alternative methods for the determination of the precision of a standard measurement method
- Part 6: Use in practice of accuracy values

Date of Establishment: 1999-05-20

Date of Public Notice in Official Gazette: 1999-05-20

Investigated by: Japanese Industrial Standards Committee
Divisional Council on Basic Items

JIS Z 8402-6 : 1999, First English edition published in 2001-01

Translated and published by: Japanese Standards Association
4-1-24, Akasaka, Minato-ku, Tokyo, 107-8440 JAPAN

In the event of any doubts arising as to the contents,
the original JIS is to be the final authority.

© JSA 2001

All rights reserved. Unless otherwise specified, no part of this publication may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying and microfilm, without permission in writing from the publisher.

Printed in Japan

Contents

	Page
Introduction.....	1
1 Scope.....	1
2 Normative references.....	2
3 Definitions	3
4 Determination of limits	3
4.1 Repeatability and reproducibility limits.....	3
4.2 Comparisons based on more than two values.....	4
5 Methods for checking the acceptability of test results and determining the final quoted result	5
5.1 General.....	5
5.2 Methods for checking the acceptability of test results obtained under repeatability conditions	6
5.3 Methods for checking the acceptability of test results obtained under both repeatability and reproducibility conditions	12
6 Method for checking the stability of test results within a laboratory.....	15
6.1 Background.....	15
6.2 Methods for checking stability	17
7 Use of repeatability and reproducibility standard deviations in assessing laboratories.....	29
7.1 Assessment method	29
7.2 Evaluation of the use of a measurement method by a laboratory not previously assessed.....	30
7.3 Continued assessment of previously approved laboratories.....	34
8 Comparison of alternative measurement methods	40
8.1 Origin of alternative measurement methods	40
8.2 Purpose of comparing measurement methods.....	40
8.3 Method B is a candidate for an alternative standard method	41
8.4 Accuracy experiment	41
8.5 Method B is a candidate for a routine method	48
Annex A (normative) Symbols and abbreviations used in JIS Z 8402.....	50

Accuracy (trueness and precision) of measurement methods and results — Part 6: Use in practice of accuracy values

Introduction This Japanese Industrial Standard has been prepared based on **ISO 5725-6 Accuracy (trueness and precision) of measurement methods and results — Part 6 : Use in practice of accuracy values** issued in 1994 as the first edition without changing the technical contents.

The “Informative references” underlined with dots in this Standard show the matters not specified in the original International Standard.

0.1 JIS Z 8402 uses two terms “trueness” and “precision” to describe the accuracy of a measurement method. “Trueness” refers to the closeness of agreement between the arithmetic mean of a large number of test results and the true or accepted reference value. “Precision” refers to the closeness of agreement between test results.

0.2 The need to consider “precision” arises because tests performed on presumably identical materials in presumably identical circumstances do not, in general, yield identical results. This is attributed to unavoidable random errors inherent in every measurement procedure; the factors that influence the outcome of a measurement cannot all be completely controlled. In the practical interpretation of measurement data, this variability has to be taken into account. For instance, the difference between a test result and some specified value may be within the scope of unavoidable random errors, in which case a real deviation from such a specified value has not been established. Similarly, comparing test results from two batches of material will not indicate a fundamental quality difference if the difference between them can be attributed to the inherent variation in the measurement procedure.

0.3 Parts 1 to 5 of JIS Z 8402 discuss the background to, and given methods for, the assessment of the precision (in terms of the repeatability standard deviation and the reproducibility standard deviation) and the trueness (in terms of the various components of bias) of measurements produced by a standard measurement method. Such assessment would, however, be pointless if there were no practical uses to which the results could be put.

0.4 Given that the accuracy of a measurement method has been established, this part of JIS Z 8402 applies that knowledge in practical situations in such a way as to facilitate commercial transactions and to monitor and improve the operational performance of laboratories.

1 Scope

1.1 The purpose of this part of JIS Z 8402 is to give some indications of the way in which accuracy data can be used in various practical situations by:

- a) giving a standard method of calculating the repeatability limit, the reproducibility limit and other limits to be used in examining the test results obtained by a standard measurement method;

- b) providing a way of checking the acceptability of test results obtained under repeatability or reproducibility conditions;
- c) describing how to assess the stability of results within a laboratory over a period of time, and thus providing a method of "quality control" of the operations within that laboratory;
- d) describing how to assess whether a given laboratory is able to use a given standard measurement method in a satisfactory way;
- e) describing how to compare alternative measurement methods.

1.2 This part of **JIS Z 8402** is concerned exclusively with measurement methods which yield measurements on a continuous scale and give a single numerical figure as the result, although this single figure may be the outcome of a calculation from a set of observations.

1.3 It is assumed that the estimates of trueness and precision for the method have been obtained in accordance with parts 1 to 5 of **JIS Z 8402**.

1.4 Any additional information regarding the field of application will be given at the beginning of each particular application.

2 Normative references The following standards contain provisions which, through reference in this Standard, constitute provisions of this Standard. If the indication of the year of coming into effect is given to these referred standards, only the edition of indicated year constitutes the provision of this Standard but the revision and amendment made thereafter are not applied. The normative references without the indication of the year of coming into effect apply limiting only to the most recent edition (including the amendment).

JIS Q 0033 : 1997	<i>Uses of certified reference materials</i>
JIS Q 0035 : 1997	<i>Certification of reference materials—General and statistical principles</i>
JIS Z 8101-1	<i>Statistics—Vocabulary and symbols—Part 1: Probability and general statistical terms</i>
JIS Z 8101-2	<i>Statistics—Vocabulary and symbols—Part 2: Statistical quality control</i>
JIS Z 8402-1 : 1999	<i>Accuracy (trueness and precision) of measurement methods and results—Part 1: General principles and definitions</i>
JIS Z 8402-2 : 1999	<i>Accuracy (trueness and precision) of measurement methods and results—Part 2: Basic method for the determination of repeatability and reproducibility of a standard measurement method</i>
JIS Z 8402-3 : 1999	<i>Accuracy (trueness and precision) of measurement methods and results—Part 3: Intermediate measures of the precision of a standard measurement method</i>
JIS Z 8402-4 : 1999	<i>Accuracy (trueness and precision) of measurement methods and results—Part 4: Basic methods for the determination of the trueness of a standard measurement method</i>
JIS Z 9021 : 1998	<i>Shewhart control charts</i>

JIS Z 9325 : 1996 *General requirements for the competence of calibration and testing laboratories*

3 Definitions For the purposes of this part of JIS Z 8402, the definitions given in JIS Z 8101 and JIS Z 8402-1 apply.

The symbols used in JIS Z 8402 are given in Annex A.

4 Determination of limits

4.1 Repeatability and reproducibility limits

4.1.1 In JIS Z 8402-2, attention has been focussed on estimating the standard deviations associated with operations under repeatability or reproducibility conditions. However, normal laboratory practice requires examination of the differences (s) observed between two (or more) test results, and for this purpose some measure akin to a critical difference is required, rather than a standard deviation.

4.1.2 When a quantity is based on sums or differences of n independent estimates each having a standard deviation σ , then that resultant quantity will have a standard deviation $\sigma\sqrt{n}$. The reproducibility limit (R) or repeatability limit (r) are for differences between two test results, so the associated standard deviation is $\sigma\sqrt{2}$. In normal statistical practice, for examining the difference between these two values the critical difference used is f times this standard deviation, i.e. $f\sigma\sqrt{2}$. The value of f (the critical range factor) depends on the probability level to be associated with the critical difference and on the shape of the underlying distribution. For the reproducibility and repeatability limits, the probability level is specified as 95 %, and throughout the analysis in JIS Z 8402 the assumption is made that the underlying distribution is approximately normal. For a normal distribution at 95 % probability level, f is 1.96 and $f\sqrt{2}$ then is 2.77. As the purpose of this part of JIS Z 8402 is to give some simple "rule of thumb" to be applied by non-statisticians when examining the results of tests, it seems reasonable to use a rounded value of 2.8 instead of $f\sqrt{2}$.

4.1.3 As has been stated, the process of estimating precision leads to estimates of the true standard deviations while the true standard deviations remain unknown. Therefore in statistical practice they should be denoted by s rather than σ . However, if the procedures given in JIS Z 8402-1 and JIS Z 8402-2 are followed, these estimates will be based on an appreciable number of test results, and will give the best information we are likely to have of the true values of the standard deviations. In other applications that follow, for estimates of these standard deviations based on more limited data, the symbol s (estimate of a standard deviation) is used. Therefore it seems best to use the symbol σ to denote the values obtained from a full precision experiment, and treat these as true standard deviations with which other estimates (s) will be compared.

4.1.4 In view of 4.1.1 to 4.1.3, when examining two single test results obtained under repeatability or reproducibility conditions, the comparison shall be made with the repeatability limit

$$r = 2.8\sigma,$$

or the reproducibility limit.

$$R = 2.8\sigma_R$$

4.2 Comparisons based on more than two values

4.2.1 Two groups of measurements in one laboratory If, in one laboratory under repeatability conditions, two groups of measurements are performed with the first group of n_1 test results giving an arithmetic mean of $\bar{y}_1$ and the second group of n_2 test results giving an arithmetic mean of $\bar{y}_2$, then the standard deviation of $(\bar{y}_1 - \bar{y}_2)$ is

$$\sigma = \sqrt{\sigma_r^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}$$

and the critical difference for $|\bar{y}_1 - \bar{y}_2|$ is

$$CD = 2.8\sigma_r \sqrt{\frac{1}{2n_1} + \frac{1}{2n_2}}$$

at the 95 % probability level.

Note 1 If n_1 and n_2 are both unity, this reduces to $r = 2.8\sigma_r$, as expected.

4.2.2 Two groups of measurements in two laboratories If the first laboratory obtains n_1 test results giving an arithmetic mean of $\bar{y}_1$ while the second laboratory obtains n_2 test results giving an arithmetic mean of $\bar{y}_2$, in each case under repeatability conditions, then the standard deviation of $(\bar{y}_1 - \bar{y}_2)$ is

$$\begin{aligned} \sigma &= \sqrt{\sigma_L^2 + \frac{1}{n_1}\sigma_r^2 + \sigma_L^2 + \frac{1}{n_2}\sigma_r^2} \\ &= \sqrt{2\sigma_L^2 + \sigma_r^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)} \\ &= \sqrt{2(\sigma_L^2 + \sigma_r^2) - 2\sigma_r^2 \left(1 - \frac{1}{2n_1} - \frac{1}{2n_2} \right)} \end{aligned}$$

and the critical difference for $|\bar{y}_1 - \bar{y}_2|$ is

$$CD = \sqrt{(2.8\sigma_R)^2 - (2.8\sigma_r)^2 \left(1 - \frac{1}{2n_1} - \frac{1}{2n_2} \right)}$$

at the 95 % probability level.

Note 2 If n_1 and n_2 are both unity, this reduces to $R = 2.8\sigma_R$, as expected.

4.2.3 Comparison with a reference value for one laboratory If n test results are obtained under repeatability conditions within one laboratory which give an arithmetic mean of $\bar{y}$, then the comparison with a given reference value μ_0 shall be made, in the absence of specific knowledge of the laboratory component of bias, using a standard deviation for $(\bar{y} - \mu_0)$ of

$$\begin{aligned}\sigma &= \sqrt{\sigma_L^2 + \frac{1}{n}\sigma_r^2} \\ &= \frac{1}{\sqrt{2}} \sqrt{2(\sigma_L^2 + \sigma_r^2) - 2\sigma_r^2 \left(1 - \frac{1}{n}\right)} \\ &= \frac{1}{\sqrt{2}} \sqrt{2(\sigma_L^2 + \sigma_r^2) - 2\sigma_r^2 \left(\frac{n-1}{n}\right)}\end{aligned}$$

and the critical difference for $|\bar{y} - \mu_0|$ is

$$CD = \frac{1}{\sqrt{2}} \sqrt{(2.8\sigma_R)^2 - (2.8\sigma_r)^2 \left(\frac{n-1}{n}\right)}$$

4.2.4 Comparison with a reference value for more than one laboratory If p laboratories have obtained n_i test results giving arithmetic means of $\bar{y}_i$ (in each case under repeatability conditions) and the grand mean $\bar{\bar{y}}$ is computed by

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

and this grand mean is to be compared with a reference value μ_0 , then the standard deviation for $(\bar{\bar{y}} - \mu_0)$ is

$$\begin{aligned}\sigma &= \sqrt{\frac{1}{p}\sigma_L^2 + \frac{1}{p^2}\sigma_r^2 \sum \frac{1}{n_i}} \\ &= \frac{1}{\sqrt{2p}} \sqrt{2(\sigma_L^2 + \sigma_r^2) - 2\sigma_r^2 + \frac{2\sigma_r^2}{p} \sum \frac{1}{n_i}} \\ &= \frac{1}{\sqrt{2p}} \sqrt{2(\sigma_L^2 + \sigma_r^2) - 2\sigma_r^2 \left(1 - \frac{1}{p} \sum \frac{1}{n_i}\right)}\end{aligned}$$

and the critical difference for $|\bar{\bar{y}} - \mu_0|$ is

$$CD = \frac{1}{\sqrt{2p}} \sqrt{(2.8\sigma_R)^2 - (2.8\sigma_r)^2 \left(1 - \frac{1}{p} \sum \frac{1}{n_i}\right)}$$

at the 95 % probability level.

4.2.5 Quoting the results of a comparison When the absolute difference exceeds the appropriate limit as given in the preceding clauses, then the difference shall be considered as suspect, and therefore all measurements that have given rise to this difference shall be considered as suspect and subject to further investigation.

5 Methods for checking the acceptability of test results and determining the final quoted result

5.1 General

5.1.1 The checking method described in this clause should be applied only to the case where the measurement was carried out according to a measurement method which has been

standardized and whose standard deviations σ_r and σ_R are known. Therefore, when the range of N test results exceeds the appropriate limit as given in clause 4, it is considered that one, two or all of the N test results is or are aberrant. It is recommended that the cause of the aberrant result (s) should be investigated from the technical point of view. However, it may be necessary for commercial reasons to obtain some acceptable value, and in such cases the test results shall be treated according to the stipulations of this clause.

5.1.2 This clause has been prepared on the assumptions that the test results were obtained under repeatability and reproducibility conditions, and that the probability level to be used is 95 %. If intermediate conditions (see JIS Z 8402-3) were in force, then it is necessary to replace σ_r by the appropriate intermediate measure.

5.1.3 In some cases where the procedures described in 5.2 lead to the median being quoted as the final result, it might be better to abandon the date.

5.2 Methods for checking the acceptability of test results obtained under repeatability conditions

Note 3 In 5.2.2.1 and 5.2.2.2, reference made to measurements being expensive or inexpensive should be interpreted not only in financial terms but also whether the measurement is complex, troublesome or time-consuming.

5.2.1 Single test result It is not common in commercial practice to obtain only one test result. When only one test result is obtained, it is not possible to make an immediate statistical test of the acceptability of that test result with respect to the given repeatability measure. If there is any suspicion that the test result may not be correct, a second test result should be obtained. Availability of two test results leads to the more common practice which is described below.

5.2.2 Two test results The two test results should be obtained under repeatability conditions. The absolute difference between the two test results should then be compared with the repeatability limit $r = 2.8\sigma_r$.

5.2.2.1 Case where obtaining test results is inexpensive If the absolute difference between the two test results does not exceed r , then both test results are considered acceptable, and the final quoted result should be quoted as the arithmetic mean of the two test results. If the absolute difference does exceed r , the laboratory should obtain two further test results.

If the range ($x_{\max} - x_{\min}$) of the four test results is equal to or less than the critical range at the 95 % probability level for $n=4$, $CR_{0.95}(4)$, the arithmetic mean of the four test results should be reported as the final quoted result. Critical range factors, $f(n)$, for $n=2$ to $n=40$ and selected values from $n=45$ to $n=100$ are given in table 1 to be used to calculate the critical range according to the following equation:

$$CR_{0.95}(n) = f(n) \sigma_r$$

If the range of the four test results is greater than the critical range for $n=4$, the median of the four test results should be reported as the final quoted result.

This procedure is summarized in the flowchart given in Fig. 1.

5.2.2.2 Case where obtaining test results is expensive If the absolute difference between the two test results does not exceed r , then both test results are considered acceptable, and the final quoted result should be quoted as the arithmetic means of the two test results. If the absolute difference does exceed r , the laboratory should obtain a further test result.

If the range ($x_{\max} - x_{\min}$) of the three test results is equal to or less than the critical range for $n=3$, $CR_{0.95}(3)$, the arithmetic mean of the three test results should be reported as the final quoted result.

If the range of the three test results is greater than the critical range for $n=3$, a decision on one of the following two cases shall be made.

a) Case where it is impossible to obtain a fourth test result:

The laboratory should use the median of the three test results as the final quoted result. This procedure is summarized in the flowchart given in Fig. 2.

b) Case where it is possible to obtain a fourth test result:

The laboratory should obtain the fourth test result. If the range ($x_{\max} - x_{\min}$) of the four test results is equal to or less than the critical range for $n=4$, $CR_{0.95}(4)$, the arithmetic mean of the four test results should be reported as the final quoted result. If the range of the four test results is greater than the critical range for $n=4$, the laboratory should use the median of the four test results as the final quoted result.

This procedure is summarized in the flowchart given in Fig. 3.

5.2.3 More than two test results to start with It is sometimes practical to start with more than two test results. The method for obtaining the final quoted result under repeatability conditions for the cases where $n > 2$ is similar to the case for $n=2$.

The range ($x_{\max} - x_{\min}$) of the test results is compared with the critical range $CR_{0.95}(n)$ calculated from Table 1 for the appropriate value of n . If the range does not exceed the critical range, then the arithmetic mean of all the n test results is used as the final quoted result.

If the range does exceed the critical range $CR_{0.95}(n)$, then a decision on one of the cases A, B or C given in Figs. 4 to 6 shall be made to obtain the final quoted result.

Cases A and B correspond to the cases where obtaining test results is inexpensive and expensive, respectively. Case C is an alternative which is recommended when the starting number of test results is five or more and where obtaining each test result is inexpensive, or when the starting number of test results is four or more and where obtaining each test result is expensive.

For inexpensive measurements, the difference between case A and case C is that case A requires n further measurements, whereas case C requires less than half that number of further measurements. The decision will depend on the size of n and the ease of performing the measurements.

For expensive measurements, the difference between case B and case C is that case C requires further measurements, whereas in case B no further measurements are carried out.

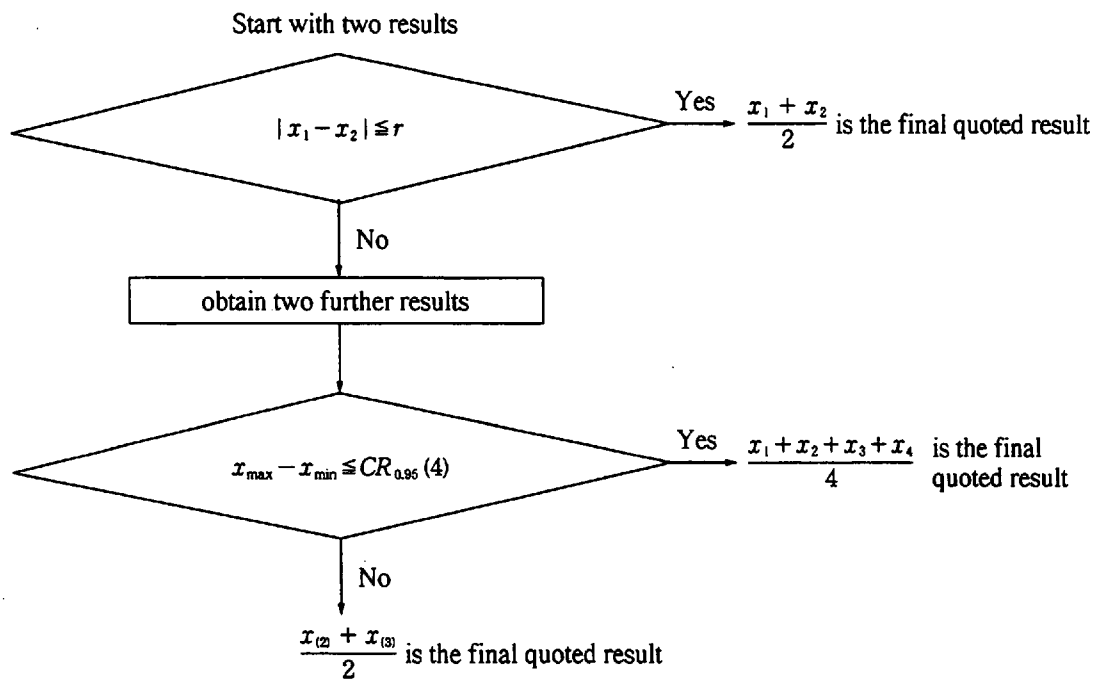
Case B shall only be considered where the performance of further measurements is so expensive as to be prohibitive.

Table 1 Critical range factors, $f(n)$

n	$f(n)$	n	$f(n)$
2	2,8	25	5,2
3	3,3	26	5,2
4	3,6	27	5,2
5	3,9	28	5,3
6	4,0	29	5,3
7	4,2	30	5,3
8	4,3	31	5,3
9	4,4	32	5,3
10	4,5	33	5,4
11	4,6	34	5,4
12	4,6	35	5,4
13	4,7	36	5,4
14	4,7	37	5,4
15	4,8	38	5,5
16	4,8	39	5,5
17	4,9	40	5,5
18	4,9	45	5,6
19	5,0	50	5,6
20	5,0	60	5,8
21	5,0	70	5,9
22	5,1	80	5,9
23	5,1	90	6,0
24	5,1	100	6,1

Note The critical range factor $f(n)$ is the 95 % quantile of the distribution of $(x_{\max} - x_{\min})/\sigma$ where $x_{\max}$ and $x_{\min}$ are the extreme values in a sample of size n from a normal distribution with standard deviation σ .

Informative reference In the table, comma (,) is
used for the decimal points.

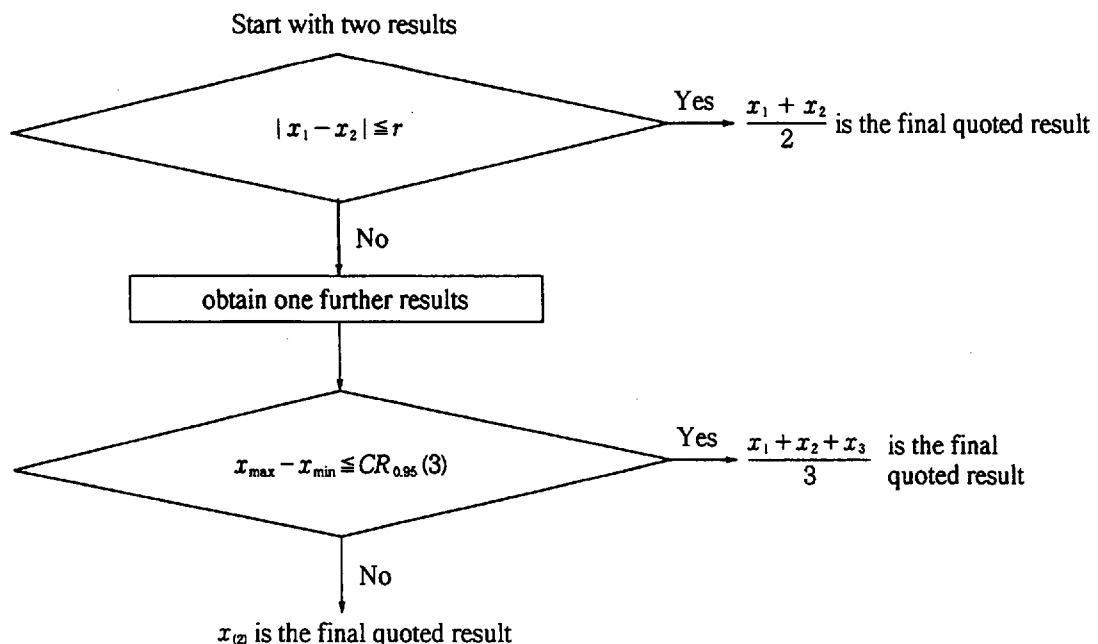


Where

$x_{(2)}$ is the second smallest result

$x_{(3)}$ is the third smallest result

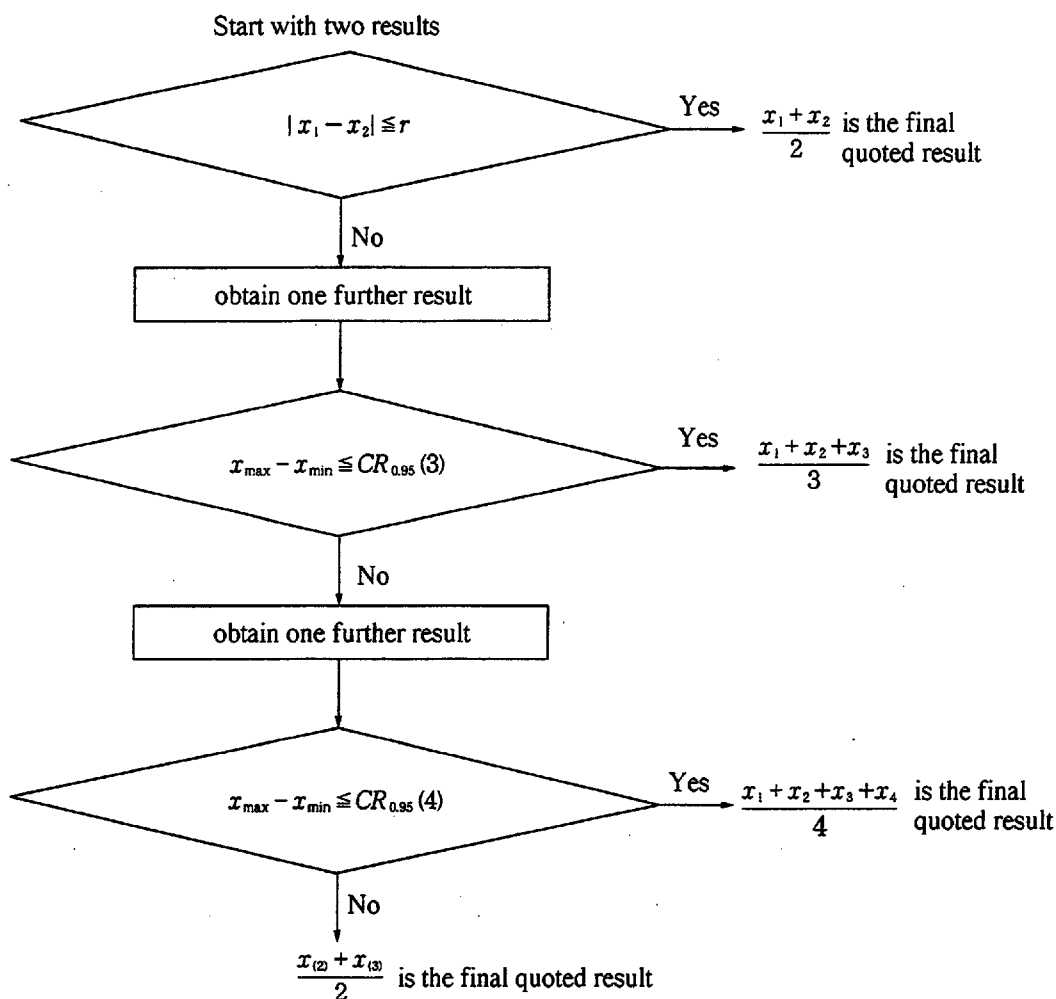
Fig. 1 Method for checking the acceptability of test results, obtained under repeatability conditions, when two test results are obtained to start with and obtaining test results is inexpensive: Case 5.2.2.1



where,

$x_{(2)}$ is the second smallest result

Fig. 2 Method for checking the acceptability of test results, obtained under repeatability conditions, when two test results are obtained to start with and obtaining test results is expensive: Case 5.2.2.2 a)



where

$x_{(2)}$ is the second smallest result

$x_{(3)}$ is the third smallest result

Fig. 3 Method for checking the acceptability of test results, obtained under repeatability conditions, when two test results are obtained to start with and obtaining test results is expensive: Case 5.2.2.2 b)

5.2.4 Example of case B: An expensive chemical analysis Expensive cases are often found in chemical analyses which consist of complicated and time-consuming procedures, requiring 2 or 3 days or more for one analysis. In such a case, it is troublesome and expensive to carry out a re-analysis when a technically questionable datum or an outlier is found in the first analysis, therefore usually three or four test results are obtained under repeatability conditions from the beginning, and test-processed according to case B. See Fig. 5.

For example, in the determination of gold and silver in ores by fire assay, although there are several methods, all of them require expensive specific equipment, highly skilled operators and a long time, usually about 2 days, for completing the entire processes and even more if the ore contains platinum-group metals or other specific co-existing elements.

The following four test results for gold content were obtained on a copper concentrate under repeatability conditions:

Au (g/t): 11.0 11.0 10.8 10.5

These test results are processed according to method B.

The method for determination of gold and silver has not been established in an International Standard, however when a value of

$$\sigma_r = 0.12 \text{ (g/t)}$$

is given for the determination of gold,

$$CR_{0.95}(4) = 3.6 \times 0.12 = 0.43 \text{ (g/t)}$$

according to Table 1, where $f(4) = 3.6$.

Since the range of the above four test results is

$$11.0 - 10.5 = 0.5 \text{ (g/t)}$$

which is greater than $CR_{0.95}(4)$, the final quoted result is the median of the four test results, i.e.

$$\frac{11.0 + 10.8}{2} = 10.9 \text{ (g/t)}$$

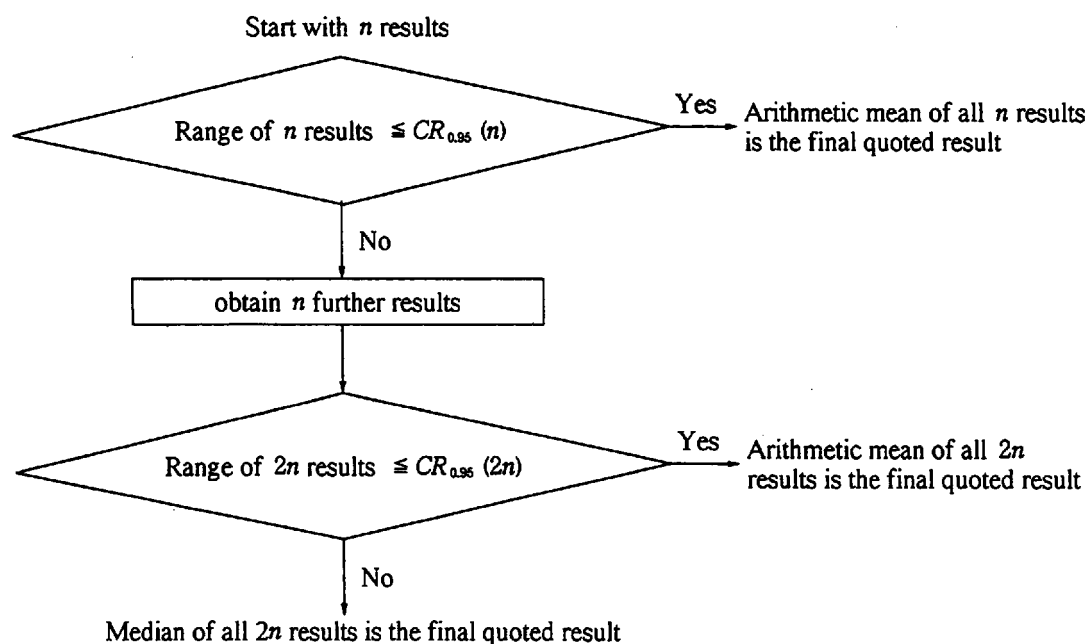


Fig. 4 Method for checking the acceptability of test results, obtained under repeatability conditions, when n test results are obtained to start with and obtaining test results is inexpensive: Case A

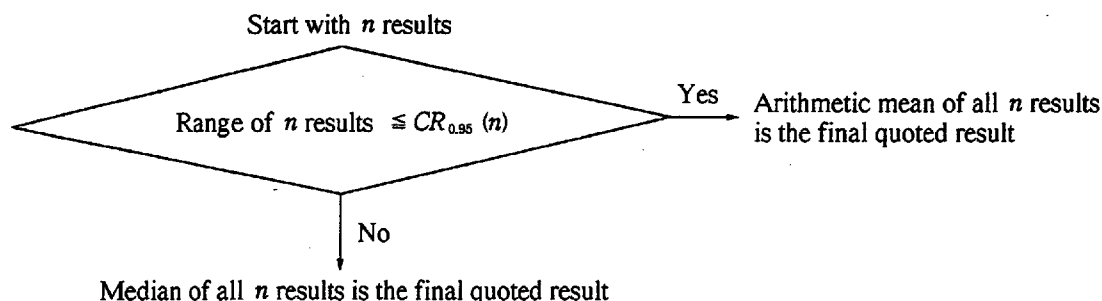
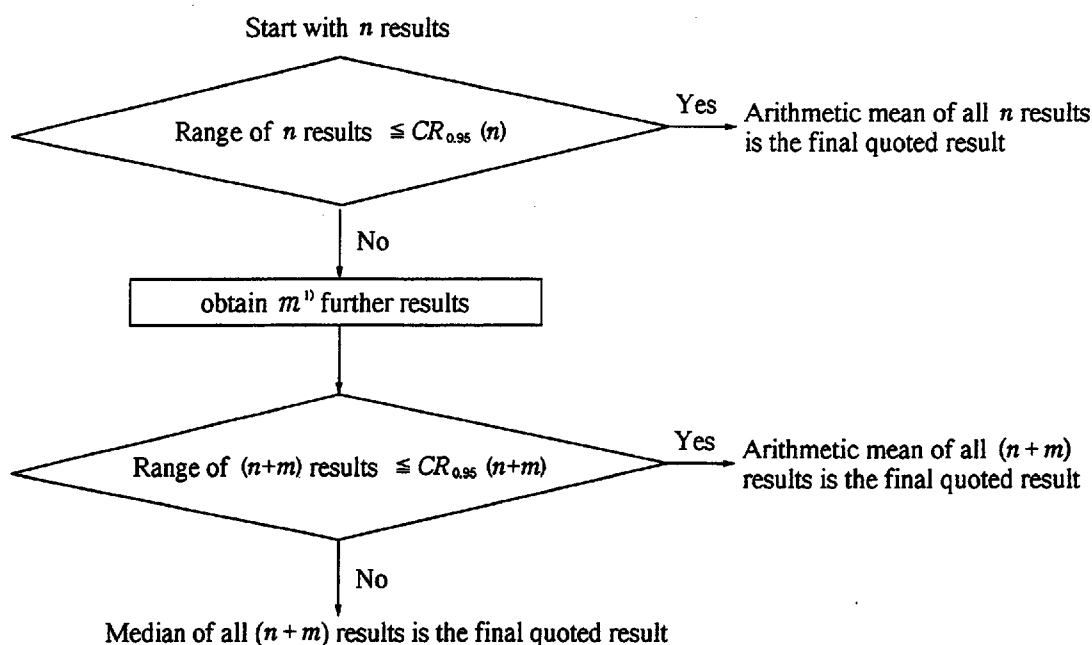


Fig. 5 Method for checking the acceptability of test results, obtained under repeatability conditions, when n test results are obtained to start with and obtaining test results is expensive: Case B

5.2.5 Note regarding precision experiment If the procedures given in 5.2.2 or 5.2.3 result frequently in values exceeding the critical values, the precision of the measurement method for this laboratory and/or precision experiment should be investigated.

5.2.6 Reporting the final quoted result If only the final quoted test result is presented, both of the following points should be specified:

- the number of test results used for the computation of the final quoted result; and
- whether the arithmetic mean or the median of the test results was taken.



1) m has to be chosen as an integer satisfying the condition $n/3 \leq m \leq n/2$

Fig. 6 Method for checking the acceptability of test results, obtained under repeatability conditions, when $n \geq 5$ and obtaining test results is inexpensive, or $n \geq 4$ and obtaining test results is expensive: Case C

Informative reference Fig. 6 (case of C) should be applied to the case of an expensive measurement where the addition is allowed and when $n = 3$.

5.3 Methods for checking the acceptability of test results obtained under both repeatability and reproducibility conditions

5.3.1 General These methods cover the case where two laboratories obtain test results and there is some difference in the test results or in the arithmetic means of the test results. The reproducibility standard deviation becomes part of the statistical testing procedure as well as the repeatability standard deviation.

In all cases of obtaining test results on test samples, sufficient material should be provided to obtain the test results plus a reserve, which may be used if any re-testing becomes necessary. How large this reserve needs to be depends on the measurement method and its complexity. In any event, the surplus material should be carefully stored to protect against deterioration or adverse changes in the test material.

Test samples should be identical, that is, last-stage samples of the sample-preparing procedure should be used by both laboratories.

5.3.2 Statistical testing for agreement between test results from two laboratories

5.3.2.1 Case where only one test result is obtained in each laboratory When each laboratory has obtained only one test result, the absolute difference between the two test results should be tested against the reproducibility limit $R = 2.8\sigma_R$. If the absolute difference between the two test results does not exceed R , the two test results are considered to be in agreement and the mean of the two test results may be used as the final quoted result.

If R is exceeded, then it is necessary to discover whether the difference is due to poor precision of the measurement method and/or a difference in the test samples. To test the precision under repeatability conditions, each laboratory should follow the procedures described in 5.2.2.

5.3.2.2 Case where two laboratories obtain more than one single test result It is assumed that each laboratory will have used the procedures of 5.2 and obtained its final quoted result. Thus, it is only necessary to consider the acceptability of the two final quoted results. To verify whether the quoted results of the laboratories are in agreement, the absolute difference between the two final quoted results should be tested against the critical difference, $CD_{0.95}$, as given below.

a) $CD_{0.95}$ for two arithmetic means of n_1 and n_2 test results, respectively:

$$CD_{0.95} = \sqrt{R^2 - r^2 \left(1 - \frac{1}{2n_1} - \frac{1}{2n_2} \right)}$$

Note that in the equation above if $n_1 = n_2 = 1$, the expression reduces to R as given in 5.3.2.1.

If $n_1 = n_2 = 2$, the expression reduces to

$$CD_{0.95} = \sqrt{R^2 - \frac{r^2}{2}}$$

b) $CD_{0.95}$ for an arithmetic mean of n_1 and a median of n_2 test results, respectively:

$$CD_{0.95} = \sqrt{R^2 - r^2 \left(1 - \frac{1}{2n_1} - \frac{\{c(n_2)\}^2}{2n_2} \right)}$$

where $c(n)$ is the ratio of the standard deviation of the median to the standard deviation of the arithmetic mean. Its value is given in Table 2.

c) $CD_{0.95}$ for two medians of n_1 and n_2 test results, respectively :

$$CD_{0.95} = \sqrt{R^2 - r^2 \left(1 - \frac{\{c(n_1)\}^2}{2n_1} - \frac{\{c(n_2)\}^2}{2n_2} \right)}$$

See Table 2 for values of $c(n)$.

If the critical difference is not exceeded, both final quoted results of the two laboratories are acceptable and the grand mean of these two final quoted results can be used. If the critical difference is exceeded, then the procedures outlined in 5.3.3 should be followed.

Table 2 Values of $c(n)$

Number of test results, n	$c(n)$
1	1,000
2	1,000
3	1,160
4	1,092
5	1,197
6	1,135
7	1,214
8	1,160
9	1,223
10	1,176
11	1,228
12	1,187
13	1,232
14	1,196
15	1,235
16	1,202
17	1,237
18	1,207
19	1,239
20	1,212

Informative reference: In the table, comma (,) is used for the decimal points.

5.3.3 Resolving discrepancies between results from two laboratories The cause of discrepancies between the test results or the final quoted results of the two laboratories could be due to

- systematic differences between the two laboratories,
- difference in test samples, or
- errors in the determination of σ_r and/or σ_R .

If it is possible to exchange the test samples and/or reference standard materials, each laboratory should obtain test results using the other's test sample to determine the existence and degree of systematic error. If exchange of test samples is not possible, each laboratory