INTERNATIONAL STANDARD

ISO 5725-5

First edition
1998-07-15

Accuracy (trueness and precision) of measurement methods and results —

Part 5:
Alternative methods for the determination of the precision of a standard measurement method

Exactitude (justesse et fidélité) des résultats et méthodes de mesure —

Partie 5: Méthodes alternatives pour la détermination de la fidélité d'une méthode de mesure normalisée

Reference number
ISO 5725-5:1998(E)

Contents

International Organization for Standardization
Case postale 56 • CH-1211 Genève 20 • Switzerland
Internet iso@iso.ch

Printed in Switzerland

Foreword

ISO (the International Organization for Standardization) is a world-wide 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 organisations, 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.

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.

ISO 5725-5 was prepared by Technical Committee ISO/TC 69, *Applications of statistical methods*, Subcommittee SC 6, *Measurement methods and results*.

ISO 5725 consists of the following parts, under the general 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*

Parts 1 to 6 of ISO 5725 together cancel and replace ISO 5725:1986, which has been extended to cover trueness (in addition to precision) and intermediate precision conditions (in addition to repeatability conditions and reproducibility conditions).

Annex A forms an integral part of this part of ISO 5725. Annexes B, C and D are for information only.

Introduction

0.1 This part of ISO 5725 uses two terms *trueness* and *precision* to describe the accuracy of a measurement method. *Trueness* refers to the closeness of agreement between the average value 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 General consideration of these quantities is given in ISO 5725-1 and so is not repeated here. This part of ISO 5725 should be read in conjunction with ISO 5725-1 because the underlying definitions and general principles are given there.

0.3 ISO 5725-2 is concerned with estimating, by means of interlaboratory experiments, standard measures of precision, namely the repeatability standard deviation and the reproducibility standard deviation. It gives a basic method for doing this using the uniform-level design. This part of ISO 5725 describes alternative methods to this basic method.

a) With the basic method there is a risk that an operator may allow the result of a measurement on one sample to influence the result of a subsequent measurement on another sample of the same material, causing the estimates of the repeatability and reproducibility standard deviations to be biased. When this risk is considered to be serious, the split-level design described in this part of ISO 5725 may be preferred as it reduces this risk.

b) The basic method requires the preparation of a number of identical samples of the material for use in the experiment. With heterogeneous materials this may not be possible, so that the use of the basic method then gives estimates of the reproducibility standard deviation that are inflated by the variation between the samples. The design for a heterogeneous material given in this part of ISO 5725 yields information about the variability between samples which is not obtainable from the basic method; it may be used to calculate an estimate of reproducibility from which the between-sample variation has been removed.

c) The basic method requires tests for outliers to be used to identify data that should be excluded from the calculation of the repeatability and reproducibility standard deviations. Excluding outliers can sometimes have a large effect on the estimates of repeatability and reproducibility standard deviations, but in practice, when applying the outlier tests, the data analyst may have to use judgement to decide which data to exclude. This part of ISO 5725 describes robust methods of data analysis that may be used to calculate repeatability and reproducibility standard deviations from data containing outliers without using tests for outliers to exclude data, so that the results are no longer affected by the data analyst’s judgement.

Accuracy (trueness and precision) of measurement methods and results —

Part 5:
Alternative methods for the determination of the precision of a standard measurement method

1 Scope

This part of ISO 5725

— provides detailed descriptions of alternatives to the basic method for determining the repeatability and reproducibility standard deviations of a standard measurement method, namely the split-level design and a design for heterogeneous materials;

— describes the use of robust methods for analysing the results of precision experiments without using outlier tests to exclude data from the calculations, and in particular, the detailed use of one such method.

This part of ISO 5725 complements ISO 5725-2 by providing alternative designs that may be of more value in some situations than the basic design given in ISO 5725-2, and by providing a robust method of analysis that gives estimates of the repeatability and reproducibility standard deviations that are less dependent on the data analyst's judgement than those given by the methods described in ISO 5725-2.

2 Normative references

The following standards contain provisions which, through reference in this text, constitute provisions of this part of ISO 5725. At the time of publication, the editions indicated were valid. All standards are subject to revision, and parties to agreements based on this part of ISO 5725 are encouraged to investigate the possibility of applying the most recent editions of the standards indicated below. Members of IEC and ISO maintain registers of currently valid International Standards.

ISO 3534-1:1993, *Statistics — Vocabulary and symbols — Part 1: Probability and general statistical terms*.

ISO 3534-3:1985, *Statistics — Vocabulary and symbols — Part 3: Design of experiments*.

ISO 5725-1:1994, *Accuracy (trueness and precision) of measurement methods and results — Part 1: General principles and definitions*.

ISO 5725-2:1994, *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.*

3 Definitions

For the purposes of this part of ISO 5725, the definitions given in ISO 3534-1 and in ISO 5725-1 apply.

The symbols used in ISO 5725 are given in annex A.

4 Split-level design

4.1 Applications of the split-level design

4.1.1 The uniform level design described in ISO 5725-2 requires two or more identical samples of a material to be tested in each participating laboratory and at each level of the experiment. With this design there is a risk that an operator may allow the result of a measurement on one sample to influence the result of a subsequent measurement on another sample of the same material. If this happens, the results of the precision experiment will be distorted: estimates of the repeatability standard deviation σ_r will be decreased and estimates of the between-laboratory standard deviation σ_L will be increased. In the split-level design, each participating laboratory is provided with a sample of each of two similar materials, at each level of the experiment, and the operators are told that the samples are not identical, but they are not told by how much the materials differ. The split-level design thus provides a method of determining the repeatability and reproducibility standard deviations of a standard measurement method in a way that reduces the risk that a test result obtained on one sample will influence a test result on another sample in the experiment.

4.1.2 The data obtained at a level of a split-level experiment may be used to draw a graph in which the data for one material are plotted against the data for the other, similar, material. An example is given in figure 1. Such graphs can help identify those laboratories that have the largest biases relative to the other laboratories. This is useful when it is possible to investigate the causes of the largest laboratory biases with the aim of taking corrective action.

4.1.3 It is common for the repeatability and reproducibility standard deviations of a measurement method to depend on the level of the material. For example, when the test result is the proportion of an element obtained by chemical analysis, the repeatability and reproducibility standard deviations usually increase as the proportion of the element increases. It is necessary, for a split-level experiment, that the two similar materials used at a level of the experiment are so similar that they can be expected to give the same repeatability and reproducibility standard deviations. For the purposes of the split-level design, it is acceptable if the two materials used for a level of the experiment give almost the same level of measurement results, and nothing is to be gained by arranging that they differ substantially.

In many chemical analysis methods, the matrix containing the constituent of interest can influence the precision, so for a split-level experiment two materials with similar matrices are required at each level of the experiment. A sufficiently similar material can sometimes be prepared by spiking a material with a small addition of the constituent of interest. When the material is a natural or manufactured product, it can be difficult to find two products that are sufficiently similar for the purposes of a split-level experiment: a possible solution may be to use two batches of the same product. It should be remembered that the object of choosing the materials for the split-level design is to provide the operators with samples that they do not expect to be identical.

4.2 Layout of the split-level design

4.2.1 The layout of the split-level design is shown in table 1.

The p participating laboratories each test two samples at q levels.

The two samples within a level are denoted a and b, where a represents a sample of one material, and b represents a sample of the other, similar, material.

4.2.2 The data from a split-level experiment are represented by:

y_{ijk}

where

subscript i represents the laboratory (i = 1, 2, ..., p);

subscript j represents the level (j = 1, 2, ..., q);

subscript k represents the sample (k = a or b).

4.3 Organization of a split-level experiment

4.3.1 Follow the guidance given in clause 6 of ISO 5725-1:1994 when planning a split-level experiment.

Subclause 6.3 of ISO 5725-1:1994 contains a number of formulae (involving a quantity denoted generally by A) that are used to help decide how many laboratories to include in the experiment. The corresponding formulae for the split-level experiment are set out below.

NOTE — These formulae have been derived by the method described in NOTE 24 of ISO 5725-1:1994.

To assess the uncertainties of the estimates of the repeatability and reproducibility standard deviations, calculate the following quantities.

For repeatability

$$A_r = 1{,}96\sqrt{1/\left[2(p-1)\right]} \qquad (1)$$

For reproducibility

$$A_R = 1{,}96\sqrt{\left\{\left[1+2\left(\gamma^2-1\right)\right]^2+1\right\}\Big/\left[8\gamma^4(p-1)\right]} \qquad (2)$$

with $\gamma = \sigma_R/\sigma_r$.

If the number n of replicates is taken as two in equations (9) and (10) of ISO 5725-1:1994, then it can be seen that equations (9) and (10) of ISO 5725-1:1994 are the same as equations (1) and (2) above, except that sometimes $p - 1$ appears here in place of p in ISO 5725-1:1994. This is a small difference, so table 1 and figures B.1 and B.2 of ISO 5725-1:1994 may be used to assess the uncertainty of the estimates of the repeatability and reproducibility standard deviations in a split-level experiment.

To assess the uncertainty of the estimate of the bias of the measurement method in a split-level experiment, calculate the quantity A as defined by equation (13) of ISO 5725-1:1994 with n = 2 (or use table 2 of ISO 5725-1:1994), and use this quantity as described in ISO 5725-1.

To assess the uncertainty of the estimate of a laboratory bias in a split-level experiment, calculate the quantity A_w as defined by equation (16) of ISO 5725-1:1994 with n = 2. Because the number of replicates in a split-level experiment is, in effect, this number of two, it is not possible to reduce the uncertainty of the estimate of laboratory bias by increasing the number of replicates. (If it is necessary to reduce this uncertainty, the uniform-level design should be used instead.)

4.3.2 Follow the guidance given in clauses 5 and 6 of ISO 5725-2:1994 with regard to the details of the organization of a split-level experiment. The number of replicates, n in ISO 5725-2, may be taken to be the number of split-levels in a split-level design, i.e. two.

The a samples should be allocated to the participants at random, and the b samples should also be allocated to the participants at random and in a separate randomization operation.

It is necessary in a split-level experiment for the statistical expert to be able to tell, when the data are reported, which result was obtained on material *a* and which on material *b*, at each level of the experiment. Label the samples so that this is possible, and be careful not to disclose this information to the participants.

Table 1 — Recommended form for the collation of data for the split-level design

Laboratory	Level									
	1		2			*j*			*q*	
	a	*b*	*a*	*b*		*a*	*b*		*a*	*b*
1										
2										
i										
p										

4.4 Statistical model

4.4.1 The basic model used in this part of ISO 5725 is given as equation (1) in clause 5 of ISO 5725-1:1994. It is stated there that for estimating the accuracy (trueness and precision) of a measurement method, it is useful to assume that every measurement result is the sum of three components:

$$y_{ijk} = m_j + B_{ij} + e_{ijk} \tag{3}$$

where, for the particular material tested,

m_j represents the general average (expectation) at a particular level j = 1, ..., q;

B_{ij} represents the laboratory component of bias under repeatability conditions in a particular laboratory i = 1, ..., p at a particular level j = 1, ..., q;

e_{ijk} represents the random error of test result k = 1, ..., n , obtained in laboratory i at level j, under repeatability conditions.

4.4.2 For a split-level experiment, this model becomes:

$$y_{ijk} = m_{jk} + B_{ij} + e_{ijk} \tag{4}$$

This differs from equation (3) in 4.4.1 in only one feature: the subscript k in m_{jk} implies that according to equation (4) the general average may now depend on the material *a* or *b* (k = 1 or 2) within the level j.

The lack of a subscript k in B_{ij} implies that it is assumed that the bias associated with a laboratory i does not depend on the material *a* or *b* within a level. This is why it is important that the two materials should be similar.

4.4.3 Define the cell averages as:

$$y_{ij} = (y_{ija} + y_{ijb}) / 2 \tag{5}$$

and the cell differences as:

$$D_{ij} = y_{ija} - y_{ijb} \tag{6}$$

4.4.4 The general average for a level j of a split-level experiment may be defined as:

$$m_j = (m_{ja} + m_{jb}) / 2 \quad (7)$$

4.5 Statistical analysis of the data from a split-level experiment

4.5.1 Assemble the data into a table as shown in table 1. Each combination of a laboratory and a level gives a "cell" in this table, containing two items of data, y_{ija} and y_{ijb}.

Calculate the cell differences D_{ij} and enter them into a table as shown in table 2. The method of analysis requires each difference to be calculated in the same sense

$$a - b$$

and the sign of the difference to be retained.

Calculate the cell averages y_{ij} and enter them into a table as shown in table 3.

4.5.2 If a cell in table 1 does not contain two test results (for example, because samples have been spoilt, or data have been excluded following the application of the outlier tests described later) then the corresponding cells in tables 2 and 3 both remain empty.

4.5.3 For each level j of the experiment, calculate the average D_j and standard deviation s_{Dj} of the differences in column j of table 2:

$$D_j = \sum D_{ij} / p \quad (8)$$

$$s_{Dj} = \sqrt{\sum \left(D_{ij} - D_j\right)^2 / (p - 1)} \quad (9)$$

Here, Σ represents summation over the laboratories i = 1, 2, ..., p.

If there are empty cells in table 2, p is now the number of cells in column j of table 2 containing data and the summation is performed over non-empty cells.

4.5.4 For each level j of the experiment, calculate the average y_j and standard deviation s_{yj} of the averages in column j of table 3, using:

$$y_j = \sum y_{ij} / p \quad (10)$$

$$s_{yj} = \sqrt{\sum \left(y_{ij} - y_j\right)^2 / (p - 1)} \quad (11)$$

Here, Σ represents summation over the laboratories i = 1, 2, ..., p.

If there are empty cells in table 3, p is now the number of cells in column j of table 3 containing data and the summation is performed over non-empty cells.

4.5.5 Use tables 2 and 3 and the statistics calculated in 4.5.3 and 4.5.4 to examine the data for consistency and outliers, as described in 4.6. If data are rejected, recalculate the statistics.

4.5.6 Calculate the repeatability standard deviation s_{rj} and the reproducibility standard deviation s_{Rj} from:

$$s_{rj} = s_{Dj} \Big/ \sqrt{2} \tag{12}$$

$$s_{Rj}^2 = s_{yj}^2 + s_{rj}^2 \Big/ 2 \tag{13}$$

4.5.7 Investigate whether s_{rj} and s_{Rj} depend on the average y_j, and, if so, determine the functional relationships, using the methods described in subclause 7.5 of ISO 5725-2:1994.

Table 2 — Recommended form for tabulation of cell differences for the split-level design

Laboratory	Level					
	1	2		*j*		*q*
1						
2						
i						
p						

Table 3 — Recommended form for tabulation of cell averages for the split-level design

Laboratory	Level					
	1	2		*j*		*q*
1						
2						
i						
p						

4.6 Scrutiny of the data for consistency and outliers

4.6.1 Examine the data for consistency using the *h* statistics, described in subclause 7.3.1 of ISO 5725-2:1994.

To check the consistency of the cell differences, calculate the *h* statistics as:

$$h_{ij} = \left(D_{ij} - D_j\right) \Big/ s_{Dj} \tag{14}$$

To check the consistency of the cell averages, calculate the h statistics as:

$$h_{ij} = \left(y_{ij} - y_j\right) \Big/ s_{yj} \quad (15)$$

To show up inconsistent laboratories, plot both sets of these statistics in the order of the levels, but grouped by laboratory, as shown in figures 2 and 3. The interpretation of these graphs is discussed fully in subclause 7.3.1 of ISO 5725-2:1994. If a laboratory is achieving generally worse repeatability than the others, then it will show up as having an unusually large number of large h statistics in the graph derived from the cell differences. If a laboratory is achieving results that are generally biased, then it will show up as having h statistics mostly in one direction on the graph derived from the cell averages. In either case, the laboratory should be asked to investigate and report their findings back to the organizer of the experiment.

4.6.2 Examine the data for stragglers and outliers using Grubbs' tests, described in subclause 7.3.4 of ISO 5725-2:1994.

To test for stragglers and outliers in the cell differences, apply Grubbs' tests to the values in each column of table 2 in turn.

To test for stragglers and outliers in the cell averages, apply Grubbs' tests to the values in each column of table 3 in turn.

The interpretation of these tests is discussed fully in subclause 7.3.2 of ISO 5725-2:1994. They are used to identify results that are so inconsistent with the remainder of the data reported in the experiment that their inclusion in the calculation of the repeatability and reproducibility standard deviations would affect the values of these statistics substantially. Usually, data shown to be outliers are excluded from the calculations, and data shown to be stragglers are included, unless there is a good reason for doing otherwise. If the tests show that a value in one of tables 2 or 3 is to be excluded from the calculation of the repeatability and reproducibility standard deviations, then the corresponding value in the other of these tables should also be excluded from the calculation.

4.7 Reporting the results of a split-level experiment

4.7.1 Advice is given in subclause 7.7 of ISO 5725-2:1994 on:

— reporting the results of the statistical analysis to the panel;

— decisions to be made by the panel; and

— the preparation of a full report.

4.7.2 Recommendations on the form of a published statement of the repeatability and reproducibility standard deviations of a standard measurement method are given in subclause 7.1 of ISO 5725-1:1994.

4.8 Example 1: A split-level experiment — Determination of protein

4.8.1 Table 4 contains the data from an experiment [5] which involved the determination by combustion of the content of protein in feeds. There were nine participating laboratories, and the experiment contained 14 levels. Within each level, two feeds were used having similar mass fraction of protein in feed.

4.8.2 Tables 5 and 6 show the cell averages and differences, calculated as described in clause 4.5.1, for just Level 14 (j = 14) of the experiment.

Using equations (8) and (9) in 4.5.3, the differences in table 5 give:

D_{14} = 8,34 %

s_{D14} = 0,436 1 %

and applying equations (10) and (11) in 4.5.4 to the averages in table 6 gives:

y_{14} = 85,46 %

s_{Y14} = 0,453 4 %

so the repeatability and reproducibility standard deviations are, using equations (12) and (13) in 4.5.6:

s_{r14} = 0,31 %

s_{R14} = 0,50 %

Table 7 gives the results of the calculations for the other levels.

4.8.3 Figure 1 shows the results for samples a from table 4 plotted against the corresponding results for samples b, for Level 14, in a "Youden plot". Laboratory 5 gives a point in the bottom left-hand corner of the graph, and Laboratory 1 gives a point in the top right-hand corner: this indicates that the data from Laboratory 5 have a consistent negative bias over samples a and b, and that the data from Laboratory 1 have a consistent positive bias over the two samples. It is common to find this sort of pattern when plotting the data from a split-level design as in figure 1. The figure also shows that the results for Laboratory 4 are unusual, as the point for this laboratory is some distance from the line of equality for the two samples. The other laboratories form a group in the middle of the plot. This figure thus provides a case for investigating the causes of the biases at the three laboratories.

NOTE — For further information on the interpretation of "Youden plots", see references [7] and [8].

4.8.4 The values of the h statistics, calculated as described in 4.6.1, are shown in tables 5 and 6, for only Level 14. The values for all levels are plotted in figures 2 and 3.

In figure 3, the h statistics for cell averages show that Laboratory 5 gave negative h statistics at all levels, indicating a consistent negative bias in their data. In the same figure, Laboratories 8 and 9 gave h statistics that are nearly all positive, indicating consistent positive biases in their data (but smaller than the negative bias in Laboratory 5). Also, the h statistics for Laboratories 1, 2 and 6 indicate a bias that changes with level in each of these laboratories. Such interactions between the laboratories and the levels may provide clues as to the causes of the laboratory biases.

Figure 2 does not reveal any noteworthy pattern.

4.8.5 Values of the Grubbs' statistics are given in table 8. These tests again indicate that the data from Laboratory 5 are suspect.

4.8.6 At this point in the analysis, the statistical expert should initiate an investigation at Laboratory 5 of the possible causes of the suspect data, before proceeding with the analysis of the data. If the cause cannot be identified, there is a case in this instance for excluding all data from Laboratory 5 from the calculation of the repeatability and reproducibility standard deviations. The analysis would then continue with an investigation of possible functional relationships between the repeatability and reproducibility standard deviations and the general average. This does not raise any issues that have not already been covered in ISO 5725-2, so it will not be considered here.

Table 4 — Example 1: Determination of mass fraction of protein in feed, expressed as a percentage

Laboratory	Level									
	1		2		3		4		5	
	a	*b*	*a*	*b*	*a*	*b*	*a*	*b*	*a*	*b*
1	11,11	10,34	10,91	9,81	13,74	13,48	13,79	13,00	15,89	15,26
2	11,12	9,94	11,38	10,31	14,00	13,12	13,44	13,06	15,69	15,10
3	11,26	10,46	10,95	10,51	13,38	12,70	13,54	13,18	15,83	15,73
4	11,07	10,41	11,66	9,95	13,01	13,16	13,58	12,88	15,08	15,63
5	10,69	10,31	10,98	10,13	13,24	13,33	13,32	12,59	15,02	14,90
6	11,73	11,01	12,31	10,92	14,01	13,66	14,04	13,64	16,43	15,94
7	11,13	10,36	11,38	10,44	12,94	12,44	13,63	13,06	15,75	15,56
8	11,21	10,51	11,32	10,84	13,09	13,76	13,85	13,49	15,98	15,89
9	11,80	11,21	11,35	9,88	13,85	14,46	13,96	13,77	16,51	15,72
Laboratory	**Level**									
	6		7		8		9		10	
	a	*b*	*a*	*b*	*a*	*b*	*a*	*b*	*a*	*b*
1	20,14	19,78	20,33	20,06	46,45	44,42	52,05	49,40	65,84	59,14
2	19,25	20,25	20,36	19,94	46,69	44,62	51,94	48,81	66,31	59,19
3	20,48	19,86	20,56	20,11	46,90	44,56	52,18	48,90	66,06	58,52
4	21,54	20,06	20,64	20,46	47,13	45,29	51,73	48,56	65,93	58,93
5	19,90	19,66	20,56	19,24	45,83	43,73	50,84	47,91	64,19	57,94
6	20,31	20,27	20,85	20,63	46,86	43,96	52,18	49,03	65,73	58,77
7	20,00	20,56	20,25	20,19	46,25	44,31	52,25	49,44	66,06	59,19
8	20,43	20,69	20,85	20,27	47,11	44,40	52,44	48,81	65,66	59,38
9	20,64	21,01	20,78	20,89	47,09	45,15	52,19	48,46	66,33	59,47
Laboratory	**Level**									
	11		12		13		14			
	a	*b*	*a*	*b*	*a*	*b*	*a*	*b*		
1	84,16	80,86	85,38	81,71	87,64	88,23	90,24	82,10		
2	84,50	81,06	85,56	82,44	88,81	88,38	89,88	81,44		
3	82,26	79,43	85,26	82,15	88,58	88,12	89,48	81,67		
4	84,39	80,08	85,20	81,76	88,47	87,98	90,04	80,73		
5	81,71	79,01	83,58	79,74	86,43	86,19	88,59	80,46		
6	82,85	81,16	84,44	80,90	87,78	86,89	89,40	80,88		
7	86,25	81,00	84,88	81,44	88,06	88,00	89,31	81,38		
8	84,59	81,16	84,96	81,71	88,50	87,98	89,94	81,56		
9	83,05	80,93	84,73	81,94	88,24	88,05	89,75	81,35		

Table 5 — Example 1: Cell differences for Level 14

Laboratory	Cell difference %	h statistic
1	8,14	– 0,459
2	8,44	0,229
3	7,81	– 1,215
4	9,31	2,224
5	8,13	– 0,482
6	8,52	0,413
7	7,93	– 0,940
8	8,38	0,092
9	8,40	0,138

Table 6 — Example 1: Cell averages for Level 14

Laboratory	Cell difference %	h statistic
1	86,170	1,576
2	85,660	0,451
3	85,575	0,263
4	85,385	– 0,156
5	84,525	– 2,052
6	85,140	– 0,696
7	85,345	– 0,244
8	85,750	0,649
9	85,550	0,208

Table 7 — Example 1: Values of averages, average differences, and standard deviations calculated from the data for all 14 levels in table 4

Level	Number of laboratories	General average	Average difference	Standard deviations			
j	p	y_j %	D_j %	s_{yj} %	s_{Dj} %	s_{rj} %	s_{Rj} %
1	9	10,87	0,73	0,35	0,21	0,15	0,36
2	9	10,84	1,05	0,36	0,43	0,30	0,42
3	9	13,41	0,13	0,44	0,55	0,39	0,52
4	9	13,43	0,50	0,30	0,21	0,15	0,32
5	9	15,66	0,27	0,39	0,40	0,29	0,44
6	9	20,27	0,06	0,40	0,73	0,52	0,54
7	9	20,39	0,38	0,30	0,41	0,29	0,37
8	9	45,60	2,21	0,44	0,37	0,26	0,47
9	9	50,40	3,16	0,44	0,35	0,25	0,47
10	9	62,37	6,84	0,53	0,40	0,28	0,57
11	9	82,14	3,23	1,01	1,08	0,77	1,15
12	9	83,17	3,45	0,74	0,46	0,33	0,77
13	9	87,91	0,30	0,69	0,41	0,29	0,72
14	9	85,46	8,34	0,45	0,44	0,31	0,50

Table 8 — Example 1: Values of Grubbs' statistics

Grubbs' statistics for differences				
Level	**One smallest**	**Two smallest**	**Two largest**	**One largest**
1	1,653	0,5081	0,3139	2,125
2	1,418	0,3945	0,4738	1,535
3	1,462	0,3628	0,5323	1,379
4	1,490	0,5841	0,4771	1,414
5	2,033	0,3485	0,6075	1,289
6	1,456	0,5490	0,3210	1,947
7	1,185	0,6820	0,1712	2,296* (5)
8	0,996	0,7571	0,1418* (6; 8)	1,876
9	1,458	0,5002	0,3092	1,602
10	1,474	0,3360	0,4578	1,737
11	1,422	0,5089	0,2943	1,865
12	1,418	0,6009	0,2899	1,956
13	2,172	0,2325	0,6326	1,444
14	1,215	0,6220	0,2362	2,224* (4)
Grubbs' statistics for cell averages				
Level	**One smallest**	**Two smallest**	**Two largest**	**One largest**
1	1,070	0,6607	0,1291* (6; 9)	1,832
2	1,318	0,6288	0,2118	2,165
3	1,621	0,4771	0,4077	1,680
4	1,591	0,5339	0,3807	1,429
5	1,794	0,4018	0,5009	1,333
6	1,291	0,4947	0,4095	1,386
7	1,599	0,5036	0,4391	1,470
8	1,872	0,3753	0,4536	1,404
9	2,328* (5)	0,1317* (4; 5)	0,7417	1,025
10	2,456** (5)	-	-	1,000
11	1,756	0,2469	0,5759	1,472
12	2,037	0,1063* (5; 6)	0,7116	1,130
13	2,308* (5)	0,0733** (5; 6)	0,7777	0,994
14	2,052	0,2781	0,5486	1,576

NOTE — Numbers in brackets indicate the laboratories that give rise to the stragglers or outliers.

The critical values of the Grubbs' test statistics for 9 laboratories, whether applied to the differences or the cell averages, are as follows.

	Straggler (*)	Outlier (**)
Grubbs' test for a single outlier	2,215	2,387
Grubbs' test for a pair of outliers	0,149 2	0,085 1

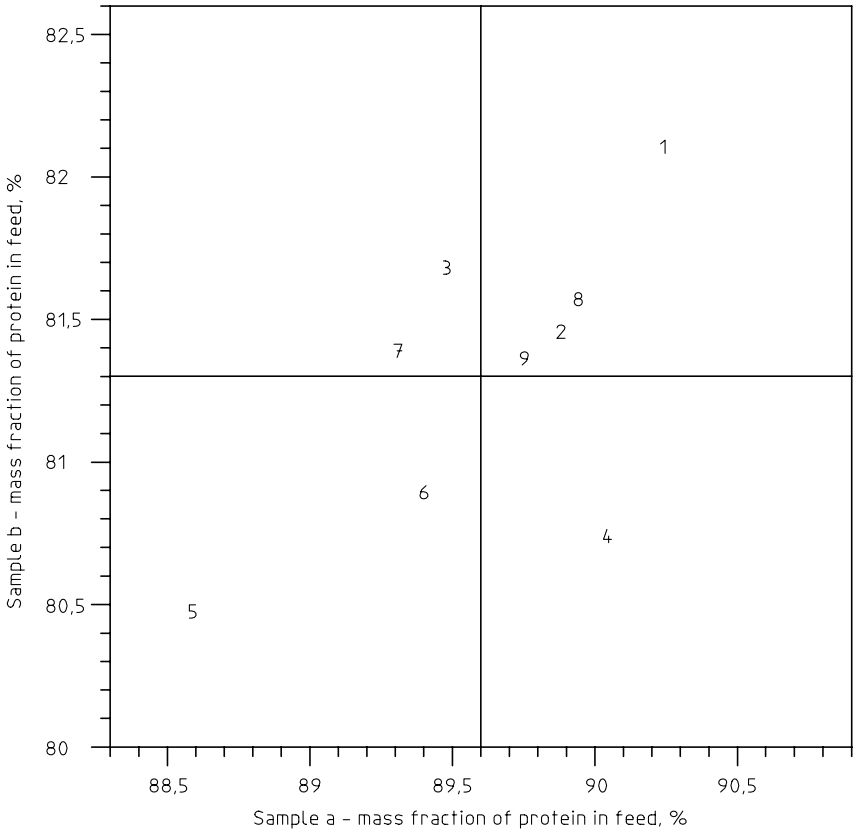


Figure 1 — Example 1: Data obtained at Level 14

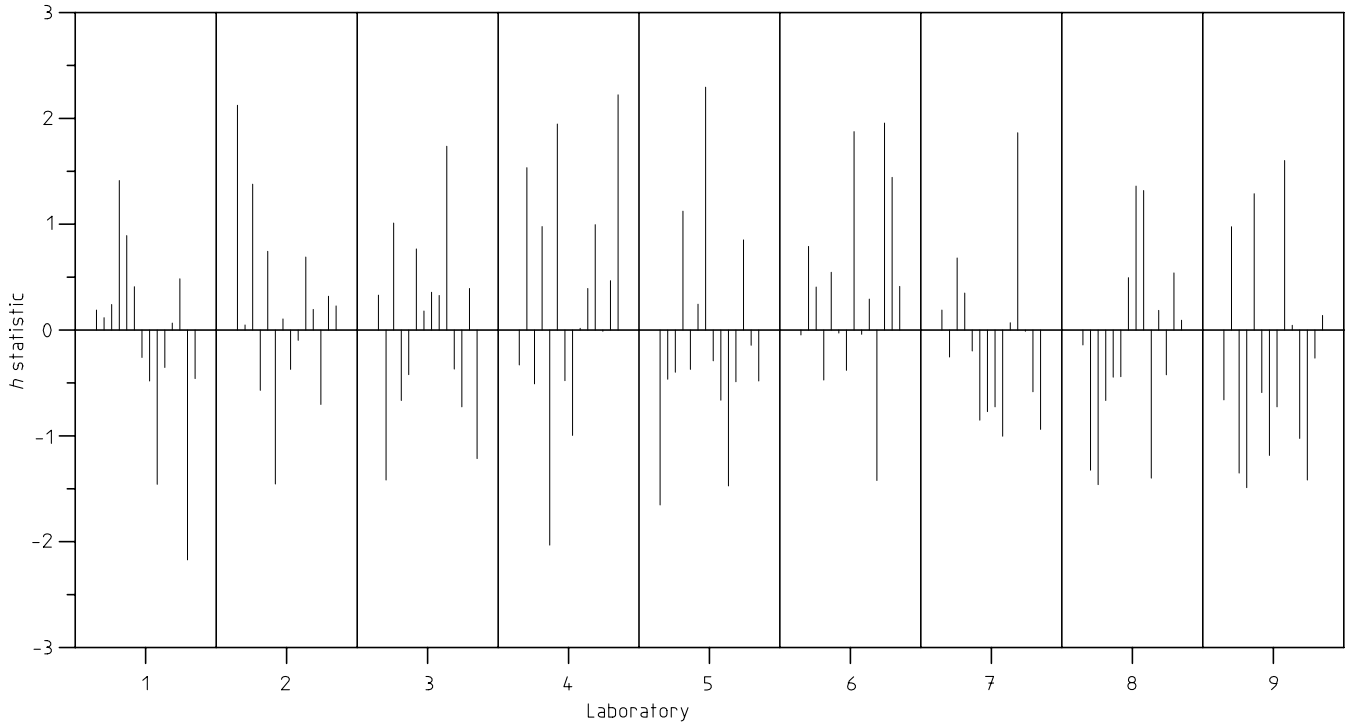


Figure 2 — Example 1: Consistency check on cell differences (grouped by laboratory)

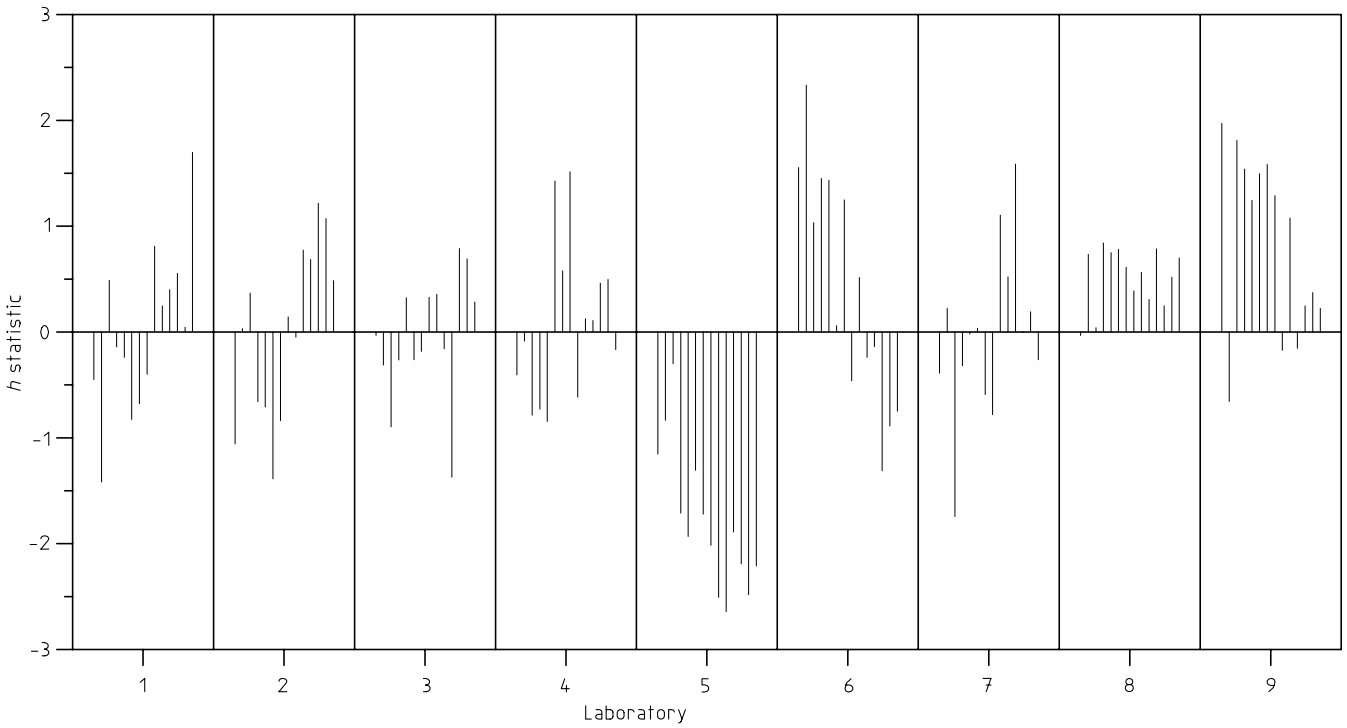


Figure 3 — Example 1: Consistency check on cell averages (grouped by laboratory)

5 A design for a heterogeneous material

5.1 Applications of the design for a heterogeneous material

5.1.1 An example of a heterogeneous material is leather; no two hides are the same, and the properties of leather vary substantially within a hide. A common test that is applied to leather is a tensile strength test in BS 3144 [3]. This is performed on dumbell-shaped specimens (BS 3144 specifies the number of such specimens to cut from a hide, and also their position and orientation within the hide, so the natural definition of a “sample” to use when testing leather is a complete hide). If a precision experiment is performed using the uniform level design described in ISO 5725-2, in which each laboratory is sent one hide at each level of the experiment, and two test results are obtained on each hide, variation between hides will add to the between-laboratory variation, and so increase the reproducibility standard deviation. However, if each laboratory is sent two hides at each level, and two test results are obtained on each hide, the data can be used to estimate the variation between hides and to calculate a value for the reproducibility standard deviation of the test method from which the variation between-hides has been removed.

5.1.2 Another example of a heterogeneous material is sand (that might be used, for example, for making concrete). This is laid down, by the action of wind or water, in strata that always contain graduations in particle size, so when sand is used the particle size distribution is always of interest. In concrete technology the particle size distribution of sand is measured by sieve testing (e.g. BS 812-103 [1]). To test a sand, a bulk sample is taken from the product, then one or more test portions are produced from the bulk sample. Typically, the bulk sample will be about 10 kg in mass, and the test portions will be about 200 g. Because of the natural variability of the material, there will always be some variability between bulk samples of the same product. Hence, as with leather, if a uniform level experiment is performed in which each laboratory is sent one bulk sample at each level, the variability between the bulk samples will increase the calculated reproducibility standard deviation of the test method, but if laboratories are sent two bulk samples at each level, then values for the reproducibility standard deviation can be calculated that exclude this variation.

5.1.3 The above examples also highlight another characteristic of heterogeneous materials: because of the variability of the material, the specimen or test portion preparation can be an important source of variation. Thus with leather, the process of cutting specimens from a hide can have a large influence on the measured tensile

strength, and with sieve tests on sand the process of preparing test portions from bulk samples is usually the major source of variability in the test method. If specimens or test portions are prepared for a precision experiment in a way that does not correspond to normal practice (in an attempt to produce identical “samples”) then the values of repeatability and reproducibility standard deviations produced by the experiment will not be representative of the variability experienced in practice. There are situations in which it can be desirable to produce identical “samples” by some special process designed to eliminate, as far as possible, the variability of the material (for example, for a proficiency test, or when a precision experiment is used as part of a programme of work during the development of a measurement method). However, when the aim of the precision experiment is to discover the variability that will be experienced in practice (for example, when vendors and purchasers test samples of the same product) then it is necessary for the variability arising as a consequence of the heterogeneity of the material to be included in the measures of the precision of the measurement method.

Care should also be taken to ensure that each test result in an experiment is obtained by carrying out the test procedure independently of other tests. This will not be so if some stages of the specimen preparation are shared by several specimens, so that a bias or deviation introduced by the preparation will have a common influence on the test results derived from these specimens.

5.1.4 The design for heterogeneous materials proposed in this clause yields information about the variability between samples that is not obtainable from the uniform level design described in ISO 5725-2. There is, inevitably, a cost associated with obtaining extra information: the proposed design requires more samples to be tested. This extra information may be valuable. In the leather example discussed in 5.1.1, information about the variability between hides could be used to decide how many hides to test when assessing the quality of a consignment, or to decide between testing more hides with fewer specimens per hide or fewer hides with more specimens per hide. In the sand example discussed in 5.1.2, information about the variability between bulk samples could be used to decide if the procedure for taking bulk samples is satisfactory or in need of improvement.

5.1.5 The design described in this clause is applicable to experiments involving three factors arranged in a hierarchy: with a factor “laboratories” at the highest level in the hierarchy, a factor “samples within laboratories” as the next level in the hierarchy, and a factor “test results within samples” as the lowest level of the hierarchy. Another case that can be encountered in practice is of a three-factor hierarchy with “laboratories” at the highest level, “test results within laboratories” as the next level, and “determinations within test results” as the lowest level. This would arise if the participating laboratories in a precision experiment were each sent a single sample of a homogeneous material, were asked to carry out two (or perhaps more) tests per sample, and if each test involved a number of determinations and the test result is calculated as the average of the determinations. The formulae given in 5.5, 5.6 and 5.9 may be applied to data obtained in such an experiment, but the repeatability and reproducibility standard deviations have to be calculated in a slightly different manner to that given here (see NOTE 2 to 5.5.5). It is also necessary to specify the number of determinations that are to be averaged to give a test result, because this affects the values of the repeatability and reproducibility standard deviations.

5.2 Layout of the design for a heterogeneous material

5.2.1 The layout of the design for a heterogeneous material is shown in table 9.

The p participating laboratories are each provided with two samples at q levels, and obtain two test results on each sample. Thus each cell in the experiment contains four test results (two test results for each of two samples).

It is possible to generalize this simple design, by allowing for more than two samples per laboratory per level, or for more than two test results per sample. The calculations required by the more general design are much more complicated than those required by the case with two test results per sample and two samples per laboratory per level. However, the principles of the more general design are the same as for the simple design, so the calculations will be set out in detail here for the simple design. Formulae for calculating values for the repeatability and reproducibility standard deviations for the general design are given below in 5.9, and an example of their application in 5.10.

5.2.2 The data from the design for a heterogeneous material are represented by:

y_{ijtk}

where

subscript i represents the laboratory (i = 1, 2, ..., p');

subscript j represents the level (j = 1, 2, ..., q);

subscript t represents the sample (t = 1, 2, ..., g);

subscript k represents the test result (k = 1, 2, ..., n).

Usually, g = 2 and n = 2. In the more general design, g or n or both are greater than two.

NOTE — In ISO 5725-1 and ISO 5725-2, p is used both as the number of laboratories, and as an index in tables of critical values for Cochran's test: with the uniform-level experiment the two numbers are the same. With the design for a heterogeneous material, the index for Cochran's test can be a multiple of the number of laboratories, so p' is used here for the number of laboratories, and p for the index for Cochran's test.

5.3 Organization of an experiment with a heterogeneous material

5.3.1 Follow the guidance given in clause 6 of ISO 5725-1:1994 when planning an experiment with a heterogeneous material. An additional question has to be considered:

how many samples should be prepared for each laboratory at each level?

Usually, because of considerations of cost, the answer will be two.

The formulae, tables and figures in clause 6 and annex B of ISO 5725-1:1994 may be used to help choose the number of laboratories, samples and replicates, but with the modifications set out in 5.3.2 to 5.3.5.

5.3.2 The uncertainty of the estimate of the repeatability standard deviation, derived from an experiment on a heterogeneous material, may be assessed by calculating the quantity A_r introduced in subclause 6.3 of ISO 5725-1:1994 as:

$$A_r = 1{,}96\sqrt{1/\left[2p'g(n-1)\right]} \qquad (16)$$

instead of as defined by equation (9) of ISO 5725-1:1994. However, the above equation may be derived by replacing p in equation (9) of ISO 5725-1:1994 by $p' \times g$. Hence, in ISO 5725-1:1994, figure B.1 and the entries for repeatability under A_r in table 1 may be used by entering the figure or the table with $p = p' \times g$. Thus in the common case when g = 2 samples are to be prepared for each laboratory at each level, enter the table or the figure in ISO 5725-1 within $p = 2p'$.

NOTE — The formulae for A_r above (and for A_R below) have all been derived by the method described in NOTE 24 of ISO 5725-1:1994.

5.3.3 The uncertainty of the estimate of the reproducibility standard deviation, derived from an experiment on a heterogeneous material, may be assessed by calculating the quantity A_R introduced in subclause 6.3 of ISO 5725-1:1994 as:

$$A_R = 1{,}96\sqrt{\left(D_1 + D_2 + D_3\right)/\left(2\gamma^4\right)} \qquad (17)$$

instead of as defined by equation (10) of ISO 5725-1:1994. Here

$$D_1 = \left[\left(\gamma^2 - 1\right) + \left(\Phi^2/g\right) + 1/ng\right]^2 \Big/ \left(p' - 1\right)$$

$$D_2 = \left[\left(\Phi^2/g\right) + (1/ng)\right]^2 \Big/ \left[p'(g-1)\right]$$

$$D_3 = 1/\left[p'g(n-1)\right]$$

$\Phi = \sigma_H/\sigma_r$ (σ_H is defined later, in 5.4.1)

$$\gamma = \sigma_R/\sigma_r \qquad (18)$$

The values of Φ and γ may be derived from preliminary estimates of the standard deviations σ_H, σ_R and σ_r obtained during the process of standardizing the measurement method.

5.3.4 Follow the guidance given in clauses 5 and 6 of ISO 5725-2:1994 with regard to the details of the organization of an experiment with a heterogeneous material.

Subclause 5.1.2 of ISO 5725-2:1994 contains requirements for "the group of n tests" or "the group of n measurements" (for example, that the group of n tests should be carried out under repeatability conditions). In an experiment with a heterogeneous material, these requirements relate to the group of $g \times n$ tests in a cell, i.e. to all the tests in a laboratory at one level.

In an experiment with a heterogeneous material, the number of samples that have to be prepared at each level is $p' \times g$ (i.e. $2p'$ in the usual case when g = 2). It is important that these $p' \times g$ samples are allocated to the participating laboratories at random.

5.4 Statistical model for an experiment with a heterogeneous material

5.4.1 The basic model used in this part of ISO 5725 is reproduced in 4.4.1 as equation (3). For an experiment with a heterogeneous material, this model is expanded to become:

$$y_{ijtk} = m_j + B_{ij} + H_{ijt} + e_{ijtk} \qquad (19)$$

The terms m, B and e have the same meaning as in equation (3) in 4.4.1, but equation (19) contains an extra term H_{ijt} that represents the variation between samples, and a subscript t representing samples within laboratories (the meaning of the other subscripts is given in 5.2.2).

It is reasonable to assume that the variation between samples is random, does not depend on the laboratory, but may depend on the level of the experiment, so the term H_{ijt} has a zero expectation, and a variance:

$$\text{var}\left(H_{ijt}\right) = \sigma^2_{Hj} \qquad (20)$$

5.4.2 For the usual case with two samples per laboratory, and two test results per sample (g = n = 2), define:

a) the sample average, and the between-test-result range, for laboratory i, level j and sample t (t = 1 or 2)

$$y_{ijt} = \left(y_{ijt1} + y_{ijt2}\right)/2 \qquad (21)$$

$$w_{ijt} = \left|y_{ijt1} - y_{ijt2}\right| \qquad (22)$$

b) the cell average, and the between-sample range, for laboratory i, and level j

$$y_{ij} = \left(y_{ij1} + y_{ij2}\right)/2 \qquad (23)$$

$$w_{ij} = \left|y_{ij1} - y_{ij2}\right| \qquad (24)$$

c) the general average, and standard deviation of cell averages, for level j

$$y_j = \sum_{j=1}^{q} y_{ij} \Big/ p' \tag{25}$$

$$s_{yj} = \sqrt{\sum_{j=1}^{q} \left(y_{ij} - y_j\right)^2 \Big/ \left(p' - 1\right)} \tag{26}$$

where the summation is over the laboratories i = 1, 2, ... , p'.

5.5 Statistical analysis of the data from an experiment with a heterogeneous material

5.5.1 The usual case when two samples are prepared for each laboratory at each level, and two test results are obtained on each sample, is considered in detail here. (The general case is considered in 5.9 and 5.10.)

Assemble the data into a table as shown in table 9. Each combination of a laboratory and a level gives a “cell” in this table, containing four test results.

Using equations (21) to (26) given in 5.4.2:

a) calculate the between-test-result ranges and enter them into a table as shown in table 10;

b) calculate the between-sample ranges and enter them into a table as shown in table 11;

c) calculate the cell averages and enter them into a table as shown in table 12;

Record ranges as positive values (i.e. ignore the sign).

Table 9 — Recommended form for collation of data from the design for a heterogeneous material

Laboratory	**Sample**	**Level 1**		**Level 2**			**Level j**			**Level q**	
		Test result number									
		1	2	1	2		1	2		1	2
1	1 2										
2	1 2										
i	1 2										
p'	1 2										

Table 10 — Recommended form for tabulation of between-test-result ranges for the design for a heterogeneous material

Laboratory	Sample	Level 1	Level 2		Level j		Level q
1	1						
	2						
2	1						
	2						
i	1						
	2						
p'	1						
	2						

Table 11 — Recommended form for tabulation of between-sample ranges for the design for a heterogeneous material

Laboratory	Level 1	Level 2		Level j		Level q
1						
2						
i						
p'						

Table 12 — Recommended form for tabulation of cell averages for the design for a heterogeneous material

Laboratory	Level 1	Level 2		Level j		Level q
1						
2						
i						
p'						

5.5.2 If a cell in table 9 contains less than four test results (e.g. because samples have been spoilt, or data have been excluded following the application of the outlier tests described below) then:

a) either use the formulae for the general case given later,

b) or ignore all the data in the cell.

Option a) is to be preferred. Option b) wastes data, but allows the simple formulae to be used.

5.5.3 For each level j of the experiment, calculate the following.

a) The sum of squared between-test-result ranges in column j of table 10 (summing over p' laboratories and over two samples):

$$SS_{rj} = \sum_{i=1}^{p'} \sum_{t=1}^{2} w_{ijt}^2 \tag{27}$$

b) The sum of squared between-sample ranges in column j of table 11 (summing over p' laboratories):

$$SS_{Hj} = \sum_{i=1}^{p'} w_{ij}^2 \tag{28}$$

c) The average and standard deviation of the cell averages in column j of table 12, using equations (25) and (26) in 5.4.2.

5.5.4 Use tables 10, 11 and 12, and the statistics calculated in 5.5.3 to examine the data for consistency and outliers, as described in 5.6. If any data are rejected, recalculate the statistics.

5.5.5 Calculate the repeatability standard deviation s_{rj} and the reproducibility standard deviation s_{Rj} from:

$$s_{rj}^2 = SS_{rj} \big/ (4p') \tag{29}$$

$$s_{Rj}^2 = s_{yj}^2 + \left(SS_{rj} - SS_{Hj}\right) \big/ (4p') \tag{30}$$

If this gives

$$s_{Rj} < s_{rj} \tag{31}$$

then set

$$s_{Rj} = s_{rj} \tag{32}$$

Calculate an estimate of the standard deviation that measures variation between samples s_{Hj} as:

$$s_{Hj}^2 = SS_{Hj} \big/ (2p') - SS_{rj} \big/ (8p') \tag{33}$$

NOTES

1 It may be of interest to perform a significance test to see if the variation between samples is statistically significant, however, this is not a necessary part of the analysis. It is incorrect to use such a test to decide if the variation between samples can be ignored in the analysis (so that the test results in each cell are treated as if they are all obtained on the same sample). This would introduce a bias into the estimation of the repeatability standard deviation, because finding that the variation between samples is not statistically significant does not prove that the variation between samples is negligible.

2 In the case described in 5.1.5 (when the three factors are “laboratories”, “tests within laboratories” and “determinations within tests”) the repeatability and reproducibility standard deviations should be calculated as:

$$s_{rj}^2 = SS_{Hj} \big/ (2p')$$

$$s_{Rj}^2 = s_{yj}^2 + SS_{Hj} \big/ (4p')$$

These formulae apply when the test result is calculated as the average of two determinations.

5.5.6 Investigate whether s_{rj} and s_{Rj} depend on the general average y_j and, if so, determine the functional relationships, using the methods described in subclause 7.5 of ISO 5725-2:1994.

5.6 Scrutiny of the data for consistency and outliers

5.6.1 Examine the data for consistency using the h and k statistics, described in subclause 7.3.1 of ISO 5725-2:1994.

To check the consistency of the cell averages, calculate the h statistics as:

$$h_{ij} = \left(y_{ij} - y_j\right) \big/ s_{yj} \tag{34}$$

Plot these statistics to show up inconsistent laboratories, by plotting the statistics in the order of the levels, but grouped by laboratory.

To check the consistency of the between-sample ranges, calculate the k statistics as:

$$k_{ij} = w_{ij} \big/ \sqrt{SS_{Hj} / p'} \tag{35}$$

Plot these statistics to show up inconsistent laboratories, by plotting the statistics in the order of the levels, but grouped by laboratory.

To check the consistency of the between-test-result ranges, calculate the k statistics as:

$$k_{ijt} = w_{ijt} \big/ \sqrt{SS_{rj} / 2p'} \tag{36}$$

Plot these statistics to show up inconsistent laboratories, by plotting the statistics in the order of the levels, but also grouped by laboratory.

The interpretation of these graphs is discussed fully in subclause 7.3.1 of ISO 5725-2:1994. If a laboratory is reporting generally biased results, then most of the h statistics for the laboratory, in the graph derived from the cell averages, will be large and in the same direction. If a laboratory is not carrying out the tests within levels under repeatability conditions (and allowing extraneous factors to increase the variation between the samples) then unusually large k statistics will be seen in the graph that is derived from the between-sample ranges. If a laboratory is achieving poor repeatability then it will give unusually large k statistics in the graph that is derived from the between-test-result ranges.

5.6.2 Examine the data for stragglers and outliers using Cochran’s and Grubbs’ tests as described in subclauses 7.3.3 and 7.3.4 of ISO 5725-2:1994.

To test for stragglers and outliers in the between-test-result ranges, calculate Cochran’s statistic for each level j as:

$$C = w_{\max}^2 \big/ SS_{rj} \tag{37}$$

where $w_{\max}$ is the largest of the between-test-result ranges w_{ijt} for level j.

To use the table of critical values in subclause 8.1 of ISO 5725-2:1994, enter the table at the row corresponding to $p = 2p'$ in the left-hand margin and in the column headed $n = 2$.

To test for stragglers and outliers in the between-sample ranges, calculate Cochran's statistic for each level j as:

$$C = w_{\max}^2 / SS_{HJ} \qquad (38)$$

where $w_{\max}$ is now the largest of the between-sample ranges w_{ij} for level j.

To use the table of critical values in ISO 5725-2:1994, enter the table at the row corresponding to $p = p'$ in the left-hand margin and in the column headed $n = 2$.

To test for stragglers and outliers in the cell averages, calculate Grubbs' statistics from the cell averages as shown in subclause 7.3.4 of ISO 5725-2:1994 for each level j (where s in ISO 5725-2:1994 is the s_{yj} defined by equation (26) in 5.4.2).

The interpretation of these tests is discussed fully in subclause 7.3.2 of ISO 5725-2:1994. In an experiment on a heterogeneous material, the results of applying these tests should be acted on in the following order. First, Cochran's test should be applied to the between-test-result ranges. If it is decided, on the basis of this test, that a between-test-result range is an outlier and is to be excluded, then the two test results that give the outlying range should be excluded from the calculation of the repeatability and reproducibility standard deviations (but the other test results in the cell should be retained). Next apply Cochran's test to the between-sample ranges, and finally apply Grubbs' tests to the cell averages. If it is decided that a between-sample range is an outlier, or that a cell average is an outlier, and that the results that give rise to such an outlier are to be excluded, then exclude all the test results for the appropriate cell from the calculation of the repeatability and reproducibility standard deviations.

5.7 Reporting the results of an experiment on a heterogeneous material

The references given in 4.7 apply equally to an experiment on a heterogeneous material.

5.8 Example 2: An experiment on a heterogeneous material

5.8.1 Aggregates that are used in the materials (cement-bound or bituminous-bound) that form the surfaces of airfields and highways have to be able to withstand wetting and freezing. A method that is used to measure their ability to do this is the magnesium sulfate soundness test [2], in which a test portion of aggregate is subjected to a number of cycles of soaking in saturated magnesium sulfate solution, followed by drying. Initially the test portion is prepared so that it is all retained on a 10,0 mm sieve. The treatment in the test causes the particles to degrade, and the test result is the mass fraction of the test portion that passes the 10,0 mm sieve at the end of the test. A high result in the test (in excess of 10 % to 20 %) indicates an aggregate with poor soundness.

5.8.2 The data shown in table 13 were obtained in an experiment in which pairs of samples of eight aggregates were sent to eleven laboratories, and two magnesium sulfate soundness test results were obtained on each sample. The samples were approximately 100 kg in mass (they were used for a number of other tests), and the test portions were approximately 350 g in mass.

5.8.3 Tables 14, 15 and 16 show the between-test-result ranges, between-sample ranges, and cell averages, calculated using equations (21) to (24) in 5.4.2, for only Level 6 of the experiment.

Using equations (27) and (28) in 5.5.3, the between-test-result ranges in table 14 and the between-sample ranges in table 15 give:

$$SS_{r6} = 381{,}66\ (\%)^2 \qquad SS_{H6} = 160{,}530\ 0\ (\%)^2$$

Applying equations (25) and (26) in 5.4.2 to the cell averages in table 16 gives:

$y_6 = 19{,}0\ \%$ (the general average)

$s_{y6} = 5{,}03\ \%$

so, using equations (29) to (33) in 5.5.5, the repeatability and reproducibility standard deviations, and the standard deviation that measures variation between samples, are:

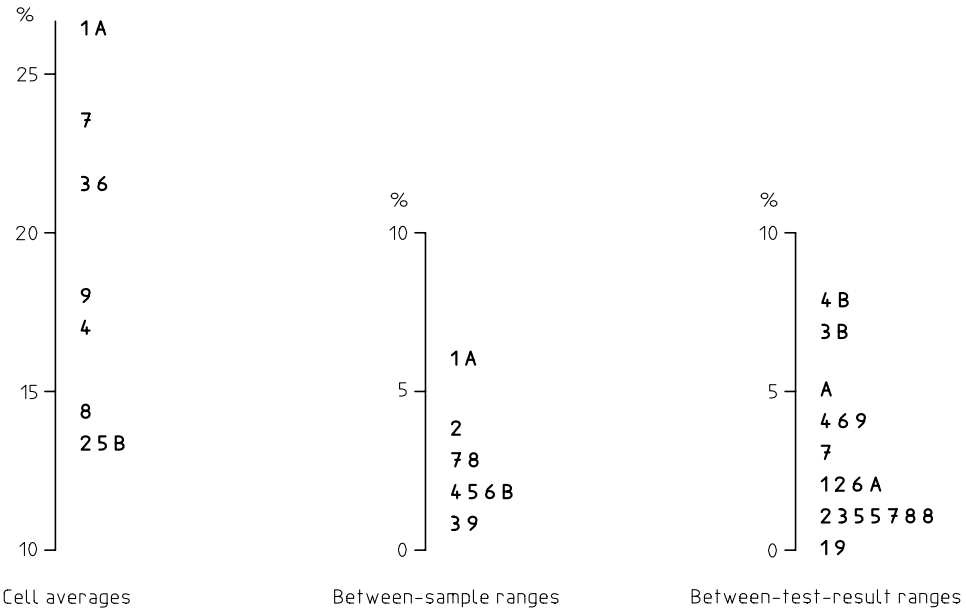
$s_{r6} = 2{,}95\ \%$ $s_{R6} = 5{,}51\ \%$ $s_{H6} = 1{,}72\ \%$

Table 17 gives the results of the calculations for the other levels.

5.8.4 Figure 4 shows histograms of between-test-result ranges, between-sample ranges, and cell averages, for Level 6. Graphs of this type give an easily understood picture of the amount of variation arising from the different sources (between test results, between samples, and between laboratories). Figure 4 shows that, in this experiment, at Level 6, there is wide variation between the cell averages, so that, if the test method were to be used in a specification, it is likely that disputes would arise between vendors and purchasers because of differences in their results. The between-sample ranges are smaller than the between-test-result ranges, suggesting that variation between samples is not important, at Level 6.

5.8.5 Values of the h and k statistics, calculated as described in 5.6.1, are also shown in tables 14, 15 and 16, for Level 6. The values for all levels are plotted in figures 5 to 7. (In these figures, the levels have been re-arranged so that the general averages are in increasing order — as shown in table 17.) Figure 5 shows that Laboratory 6 obtained several high k statistics for between-test-result ranges, indicating that it has poorer repeatability than the other laboratories. Figure 6 shows that three laboratories (1, 6 and 10) obtained high k statistics for between-sample ranges: this could be because they did not strictly follow the recommended procedures for preparing test portions from the bulk samples. Figure 7 shows consistent positive or negative h statistics in most laboratories (with Laboratories 1, 6 and 10 again achieving the largest values). This is strong evidence that there are consistent biases in most laboratories, indicating that the test method is not adequately specified.

5.8.6 Applying Cochran's test and Grubbs' test to the data, as described in 5.6.2, gives the results shown in table 18. Two outliers are identified. In the absence of other information, the data responsible for these would be excluded, and the calculations repeated. The analysis could then be continued with an investigation of functional relationships, in the same way as with the uniform level design considered in ISO 5725-2.



A, B = Laboratories 10 and 11

Figure 4 — Example 2: Histograms of ranges and averages from tables 14, 15 and 16, for Level 6

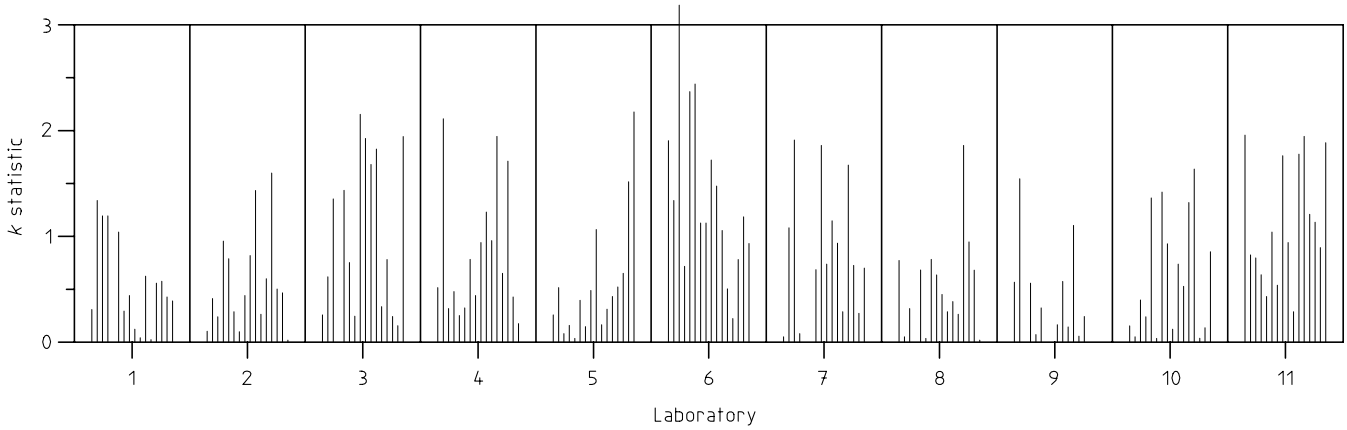


Figure 5 — Example 2: Consistency check on between-test-result ranges
(grouped by laboratory)

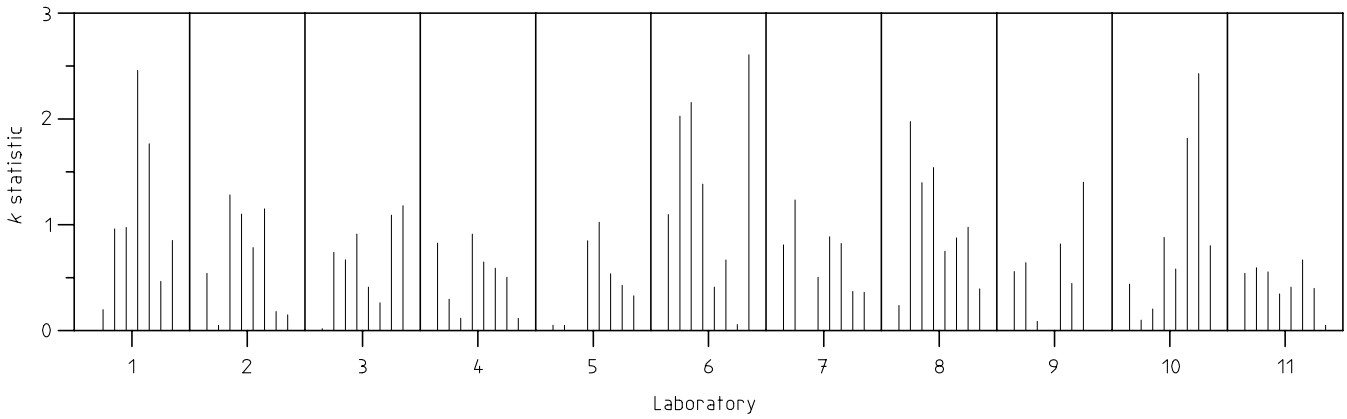


Figure 6 — Example 2: Consistency check on between-sample ranges
(grouped by laboratory)

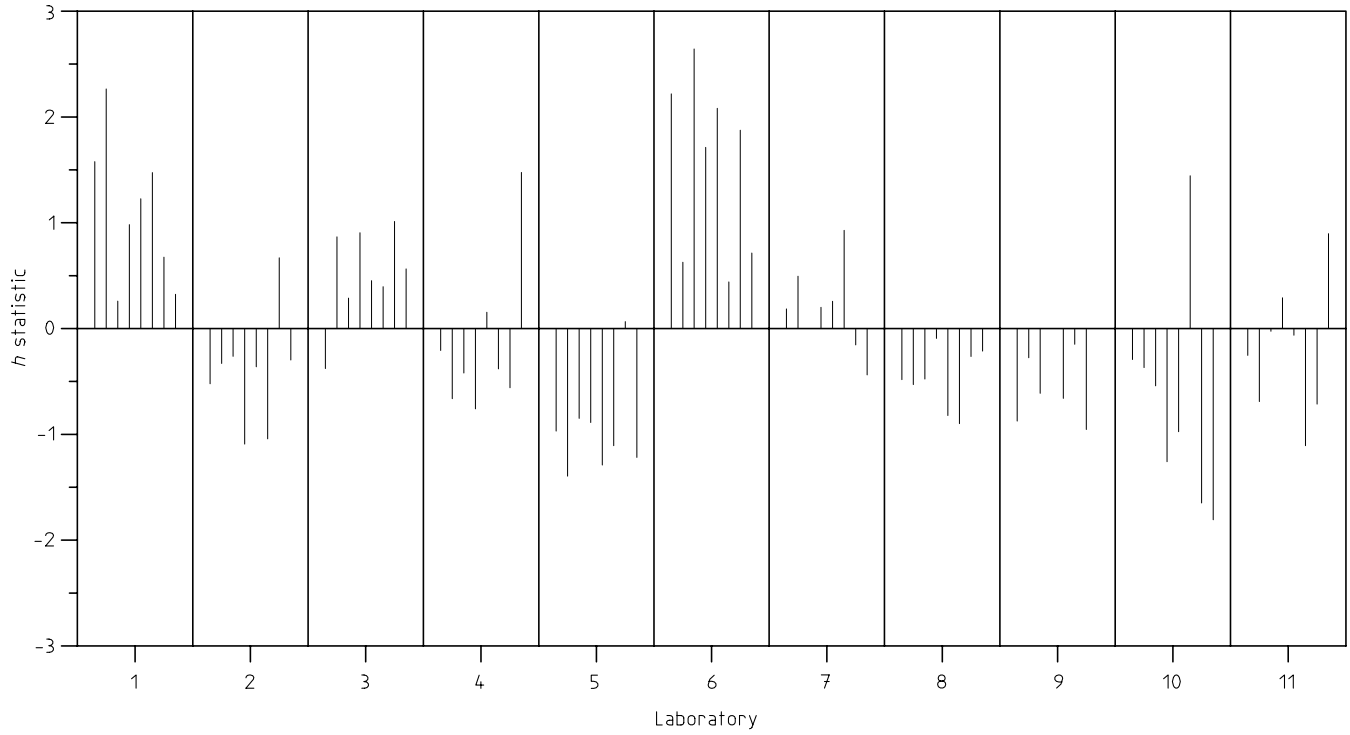


Figure 7 — Example 2: Consistency check on cell averages
(grouped by laboratory)

Table 13 — Example 2: Determination of magnesium sulfate soundness (%)

Laboratory	Sample	Level 1		Level 2		Level 3		Level 4	
		Test result number							
		1	2	1	2	1	2	1	2
1	1	69,2	67,0	7,4	8,0	4,1	3,5	10,4	10,1
	2	69,7	71,7	6,6	5,7	10,5	13,1	13,9	13,8
2	1	66,5	64,1	1,9	2,1	3,0	3,2	8,7	6,7
	2	65,7	65,8	4,2	3,3	1,9	1,1	8,3	4,8
3	1	68,7	69,5	6,3	5,8	2,4	2,9	11,7	7,0
	2	67,7	77,7	9,7	5,3	2,1	3,3	7,9	12,0
4	1	77,5	75,3	2,0	3,6	2,4	1,4	9,4	7,1
	2	76,3	77,2	4,7	3,8	6,4	2,3	10,7	7,7
5	1	55,4	63,2	3,8	4,1	1,3	0,8	3,7	6,3
	2	65,9	54,7	2,1	3,1	0,7	1,7	3,3	3,7
6	1	64,8	70,9	8,4	6,1	6,0	9,7	16,5	12,3
	2	78,2	73,4	8,3	10,6	12,4	9,8	13,2	16,8
7	1	64,8	63,4	4,3	5,7	2,9	3,0	7,5	9,3
	2	67,0	63,4	7,7	3,9	4,3	6,4	11,1	8,3
8	1	64,9	68,4	4,4	2,8	1,3	2,8	5,7	6,8
	2	65,4	65,5	5,4	6,7	2,7	2,8	4,8	5,5
9	1	—	—	—	—	1,1	0,0	6,6	7,0
	2	—	—	—	—	0,7	3,7	4,9	6,3
10	1	57,0	57,7	3,3	0,4	2,1	2,4	5,5	5,8
	2	57,1	52,7	4,2	2,3	3,6	3,5	3,9	5,7
11	1	70,6	75,2	5,3	6,4	5,7	1,9	9,5	7,2
	2	77,9	68,2	3,5	7,1	1,4	3,0	8,1	7,4
Laboratory	**Sample**	**Level 5**		**Level 6**		**Level 7**		**Level 8**	
		Test result number							
		1	2	1	2	1	2	1	2
1	1	8,9	7,4	31,1	28,5	38,7	41,7	4,2	4,1
	2	7,6	9,1	23,0	23,1	44,2	41,1	7,3	4,4
2	1	3,2	3,5	16,5	15,4	36,6	45,2	3,2	5,4
	2	2,8	4,0	10,3	12,8	43,2	40,5	1,7	2,5
3	1	4,4	6,1	24,3	16,7	38,9	43,1	3,7	7,7
	2	6,0	6,0	20,8	22,2	46,1	47,4	3,5	5,6
4	1	2,7	3,1	20,2	16,2	32,0	35,5	2,9	2,2
	2	2,3	2,9	20,0	11,9	26,5	35,7	3,2	2,3
5	1	1,3	1,4	13,8	15,1	36,7	39,5	1,1	1,2
	2	1,5	1,3	11,5	13,3	37,6	34,1	0,6	1,7
6	1	8,2	4,2	20,3	24,7	49,4	50,6	11,9	18,5
	2	3,7	4,6	21,0	18,9	48,2	52,4	14,9	8,1
7	1	3,1	5,5	27,2	23,3	38,9	29,9	—	1,7
	2	5,6	5,5	21,5	22,7	34,4	38,3	2,2	5,0
8	1	1,8	2,2	13,6	12,0	27,0	37,0	0,3	2,2
	2	4,0	4,0	15,6	16,7	39,7	34,6	3,6	3,7
9	1	3,8	3,8	17,7	17,1	33,4	33,1	1,8	2,0
	2	3,5	2,8	21,4	16,8	26,5	25,2	2,5	1,6
10	1	3,5	3,0	21,7	23,9	35,3	26,5	0,5	4,3
	2	3,2	3,5	27,0	32,5	18,0	18,2	2,0	2,1
11	1	3,5	2,5	11,0	18,4	27,0	33,5	5,1	3,9
	2	2,0	2,8	16,4	8,1	35,4	29,3	2,1	5,0

Table 14 — Example 2: Between-test-result ranges for Level 6

Laboratory	Sample	Between-test-result range %	k statistic
1	1	2,6	0,624
	2	0,1	0,024
2	1	1,1	0,264
	2	2,5	0,600
3	1	7,6	1,825
	2	1,4	0,336
4	1	4,0	0,960
	2	8,1	1,945
5	1	1,3	0,312
	2	1,8	0,432
6	1	4,4	1,056
	2	2,1	0,504
7	1	3,9	0,936
	2	1,2	0,288
8	1	1,6	0,384
	2	1,1	0,264
9	1	0,6	0,144
	2	4,6	1,104
10	1	2,2	0,528
	2	5,5	1,320
11	1	7,4	1,777
	2	8,1	1,945

Table 15 — Example 2: Between-sample ranges for Level 6

Laboratory	Between-sample range %	k statistic
1	6,75	1,767
2	4,40	1,152
3	1,00	0,262
4	2,25	0,589
5	2,05	0,537
6	2,55	0,668
7	3,15	0,825
8	3,35	0,877
9	1,70	0,445
10	6,95	1,819
11	2,55	0,668

Table 16 — Example 2: Cell averages for Level 6

Laboratory	Cell average %	h statistic
1	26,425	1,475
2	13,750	−1,043
3	21,000	0,397
4	17,075	−0,382
5	13,425	−1,108
6	21,225	0,442
7	23,675	0,929
8	14,475	−0,899
9	18,250	−0,149
10	26,275	1,445
11	13,425	−1,108

Table 17 — Example 2: Values of averages, sums of squared ranges, and standard deviations, calculated from the data for all 8 levels in table 13 (excluding cells with missing data)

Level	Number of laboratories	General average	Sums of squared ranges		Standard deviations			
j	p'	y_j %	SS_{rj} %^{2}	SS_{Hj} %^{2}	s_{yj} %	s_{rj} %	s_{Rj} %	s_{Hj} %
3	11	3,7	82,99	96,3725	2,62	1,37	2,56	1,85
5	11	4,0	34,70	11,2550	1,88	0,89	2,01	0,34
8	10	4,1	155,39	29,4225	3,49	1,97	3,92	0,00
2	10	5,0	83,51	25,2375	1,95	1,44	2,29	0,47
4	11	8,2	131,07	23,5775	3,10	1,73	3,47	0,00
6	11	19,0	381,66	160,5300	5,03	2,95	5,51	1,72
7	11	36,5	636,19	305,4775	7,28	3,80	7,78	2,58
1	10	67,4	529,71	92,9225	6,23	3,64	7,05	0,00

Table 18 — Example 2: Values of Cochran's and Grubbs' statistics

Level j	Number of laboratories p'	Cochran's statistics for between-test-result ranges	Cochran's statistics for between-sample ranges
3	11	0,203	0,664* (1)
5	11	0,461** (6)	0,374
8	10	0,298	0,465
2	10	0,232	0,238
4	11	0,169	0,550
6	11	0,172	0,301
7	11	0,157	0,536
1	10	0,237	0,680* (6)

Grubbs' statistics for cell averages

Level j	Number of laboratories p'	One smallest	Two smallest	Two largest	One largest
3	11	0,970	0,791	0,098** (1; 6)	2,219
5	11	1,396	0,709	0,302	2,266
8	10	0,849	—	—	2,643** (6)
2	10	1,259	0,614	0,466	1,713
4	11	1,290	0,681	0,294	2,082
6	11	1,108	0,700	0,479	1,475
7	11	1,649	0,562	0,453	1,875
1	10	1,808	0,345	0,590	1,476

NOTE — Numbers in brackets indicate the laboratories that gives rise to the stragglers or outliers. The critical values are as follows:

Statistical test	Applied to	Number of laboratories p'	Index in table in ISO 5725-2 p	Straggler (*)	Outlier (**)
Cochran's test	Between-test-result ranges	10	20	0,389	0,480
		11	22	0,365	0,450
Cochran's test	Between-sample ranges	10	10	0,602	0,718
		11	11	0,570	0,684
Grubbs' test for a single outlier	Cell averages	10	10	2,290	2,482
		11	11	2,355	2,564
Grubbs' test for a pair of outliers	Cell averages	10	10	0,186 4	0,115 0
		11	11	0,221 3	0,144 8

5.9 General formulae for calculations with the design for a heterogeneous material

Calculate the following statistics at each level j.

a) The general average (with summation over i, t and k):

$$m_j = \sum\sum\sum y_{ijtk} / n_j \tag{39}$$

where n_j is the number of test results included in the sum.

b) The laboratory effects for each i (with summation over t and k):

$$B_{ij} = \sum\sum \left(y_{ijtk} - m_j\right) / n_{ij} \tag{40}$$

= laboratory average – general average

where n_{ij} is the number of test results included in the sum.

c) The sample effects for each i and t (with summation over k):

$$H_{ijt} = \sum \left(y_{ijtk} - m_j - B_{ij}\right) / n_{ijt} \tag{41}$$

= sample average – laboratory average

where n_{ijt} is the number of test results included in the sum.

d) The residuals for each i, t and k:

$$z_{ijtk} = y_{ijtk} - m_j - B_{ij} - H_{ijt} \tag{42}$$

= test result – sample average

e) The sum of squares for samples (with summation over i):

$$SS_{Lj} = \sum n_{ij}\ B_{ij}^2 \tag{43}$$

f) The sum of squares for samples (with summation over i and t):

$$SS_{Hj} = \sum\sum n_{ijt}\ H_{ijt}^2 \tag{44}$$

g) The sum of squares for repeatability (with summation over i, t and k):

$$SS_{rj} = \sum\sum\sum z_{ijtk}^2 \tag{45}$$

h) Degrees of freedom:

$$v_{Lj} = p'_j - 1 \qquad v_{Hj} = g_j - p'_j \qquad v_{rj} = n_j - g_j \tag{46}$$

where

p'_j is the number of laboratories that report at least one test result;

g_j is the number of samples for which at least one test result is reported;

n_j is the total number of test results.

i) Factors for each i (with summation over t):

$$n_{ij} = \sum n_{ijt} \tag{47}$$

$$K_{ij} = \sum n_{ijt}^2 \tag{48}$$

j) Factors (with summation over i):

$$K_j = \sum n_{ij}^2 \tag{49}$$

$$K'_j = \sum K_{ij} \tag{50}$$

$$K''_j = \sum K_{ij} / n_{ij} \tag{51}$$

k) The repeatability standard deviation s_{rj}, between-samples standard deviation s_{Hj}, between-laboratory standard deviation s_{Lj}, and reproducibility standard deviation s_R using:

$$s_{rj}^2 = SS_{rj} / \nu_{rj} \tag{52}$$

$$s_{Hj}^2 = \left[SS_{Hj} - \nu_{Hj} \times s_{rj}^2 \right] \Big/ \left(n_j - K''_j \right) \tag{53}$$

$$s_{Lj}^2 = \left[SS_{Lj} - \left(K''_j - K'_j / n_j \right) \times s_{Hj}^2 - \nu_{Lj} \times s_{rj}^2 \right] \Big/ \left(n_j - K_j / n_j \right) \tag{54}$$

$$s_{Rj}^2 = s_{rj}^2 + s_{Lj}^2 \tag{55}$$

NOTE — The above formulae were derived using statistical theory set out by Scheffé [4].

5.10 Example 3: An application of the general formulae

5.10.1 The data of Example 2, Level 4, are used to provide an example of the application of the general formulae by omitting some of the test results (see table 19). The formulae presented in 5.9 give the general average shown under table 19 and the sums of squares, degrees of freedom and factors shown under tables 20, 21 and 22.

5.10.2 Applying equations (52) to (55) in step k) of 5.9 then gives:

$$s_{rj}^2 = SS_{rj} / \nu_{rj}$$

$$= 36{,}895\,0 / 16 \ \%^2$$

so

$$s_{rj} = 1{,}52 \ \%$$

and

$$s_{Hj}^2 = \left[SS_{Hj} - \nu_{Hj} \times s_{rj}^2 \right] \Big/ \left(n_j - K''_j \right)$$

$$= \left[29{,}907\,5 - 9 \times 1{,}518\,5^2 \right] \Big/ (36 - 19{,}667) \ \%^2$$

so

$$s_{Hj} = 0{,}75\ \%$$

and

$$s_{Lj}^2 = \left[SS_{Lj} - \left(K_j'' - K_j^{'}/n_j\right) \times s_{Hj}^2 - v_{Lj} \times s_{rj}^2\right] \Big/ \left(n_j - K_j/n_j\right)$$

$$= \left[378{,}853\,1 - (19{,}666\,7 - 68/36) \times 0{,}748\,7^2 - 10 \times 1{,}518\,5^2\right] \Big/ (36 - 130/36)$$

so

$$s_{Lj} = 3{,}27\ \%$$

and

$$s_{Rj} = \sqrt{1{,}52^2 + 3{,}27^2}$$

$$= 3{,}61\ \%$$

Table 19 — Example 3: Determinations of magnesium sulfate soundness for Level 4

Laboratory i	**Sample** t	**Test result** $k = 1$ %	**Test result** $k = 2$ %
1	1	—	10,1
	2	13,9	13,8
2	1	—	—
	2	8,3	4,8
3	1	—	7,0
	2	—	12,0
4	1	9,4	—
	2	—	—
5	1	3,7	6,3
	2	3,3	3,7
6	1	16,5	12,3
	2	13,2	16,8
7	1	7,5	9,3
	2	11,1	8,3
8	1	5,7	6,8
	2	4,8	5,5
9	1	6,6	7,0
	2	4,9	6,3
10	1	5,5	5,8
	2	3,9	5,7
11	1	9,5	7,2
	2	8,1	7,4
General average: m_j = 8,111 1 %			
Number of test results: n_j = 36			

Table 20 — Example 3: Calculation of the sum of squares for laboratories

Laboratory i	Laboratory average %	Number of test results n_{ij}	Laboratory effect B_{ij} %	Factor K_{ij}
1	12,600	3	4,488 9	5
2	6,550	2	– 1,561 1	4
3	9,500	2	1,388 9	2
4	9,400	1	1,288 9	1
5	4,250	4	– 3,861 1	8
6	14,700	4	6,588 9	8
7	9,050	4	0,938 9	8
8	5,700	4	– 2,411 1	8
9	6,200	4	– 1,911 1	8
10	5,225	4	– 2,886 1	8
11	8,050	4	– 0,061 1	8
Sum of squares for laboratories:		SS_{Lj} = 378,853 1 %^{2}		
Degrees of freedom for laboratories:		ν_{Lj} = 11 – 1 = 10		
Factors:		K_j = 130	K'_j = 68	K''_j = 19,666 7

Table 21 — Example 3: Calculation of the sum of squares for samples

Laboratory i	Sample t	Sample average %	Number of test results n_{ijt}	Sample effect H_{ijt} %
1	1	10,10	1	– 2,500
	2	13,85	2	1,250
2	1	—	0	—
	2	6,55	2	0,000
3	1	7,00	1	– 2,500
	2	12,00	1	2,500
4	1	9,40	1	0,000
	2	—	0	—
5	1	5,00	2	0,750
	2	3,50	2	– 0,750
6	1	14,40	2	– 0,300
	2	15,00	2	0,300
7	1	8,40	2	– 0,650
	2	9,70	2	0,650
8	1	6,25	2	0,550
	2	5,15	2	– 0,550
9	1	6,80	2	0,600
	2	5,60	2	– 0,600
10	1	5,65	2	0,425
	2	4,80	2	– 0,425
11	1	8,35	2	0,300
	2	7,75	2	– 0,300
Sum of squares for samples:		SS_{Hj} = 29,907 5 %^{2}		
Degrees of freedom for samples:		ν_{Hj} = 20 – 11 = 9		

Table 22 — Example 3: Calculation of the sum of squares for repeatability

Laboratory i	Sample t	Test result $k = 1$ %	Test result $k = 2$ %
1	1	—	0,00
	2	0,05	– 0,05
2	1	—	—
	2	1,75	– 1,75
3	1	—	0,00
	2	—	0,00
4	1	0,00	—
	2	—	—
5	1	– 1,30	1,30
	2	– 0,20	0,20
6	1	2,10	– 2,10
	2	– 1,80	1,80
7	1	– 0,90	0,90
	2	1,40	– 1,40
8	1	– 0,55	0,55
	2	– 0,35	0,35
9	1	– 0,20	0,20
	2	– 0,70	0,70
10	1	– 0,15	0,15
	2	– 0,90	0,90
11	1	1,15	– 1,15
	2	0,35	– 0,35
Sum of squares for repeatability:		SS_{rj} = 36,895 %^{2}	
Degrees of freedom for repeatability:		v_{rj} = 36 – 20 = 16	

6 Robust methods for data analysis

6.1 Applications of robust methods of data analysis

6.1.1 In ISO 5725-2, it is recommended that two tests for outliers (Cochran's test and Grubbs' test) should be applied to the data obtained in a precision experiment, and any data that give rise to a test statistic in one or other of these tests that exceeds the critical value of the test at the 1 % significance level should be discarded (unless the statistician has good reason to retain the data). In practice, this procedure is often not easy to apply. Consider the results of the outlier tests in Example 1 of 4.8. These are given in table 8. Laboratory 5 gives only one cell average (at Level 10) that is sufficiently extreme to be classed as an outlier by Grubbs' test, but gives three other stragglers, and a strong indication in figure 3 that something is amiss in this laboratory. The statistician, in this situation, has to decide between:

a) retaining all the data for Laboratory 5;

b) discarding only the data from Level 10 for Laboratory 5;

c) discarding all the data for Laboratory 5.

His/her decision will have a substantial influence on the calculated values for the repeatability and reproducibility standard deviations. It is a common experience when analysing data from precision experiments to find data that are on the borderline between stragglers and outliers, so that judgements may have to be made that affect the results of the calculation. This may be unsatisfactory. The robust methods described in this clause allow the data to be analysed in such a way that it is not required to make decisions that affect the results of the calculations. Thus if

there is reason to expect that the results of a precision experiment will contain outliers, robust methods may be preferred.

6.1.2 The basic model discussed in clause 5 of ISO 5725-1:1994 contains the assumption that establishing a common value for the repeatability standard deviation for all laboratories using the measurement method is justifiable. In practice it is often found that some laboratories achieve poorer repeatability than others. See, for example, figure 5 for Example 2 in 5.8. Laboratory 6 obviously achieved a much worse repeatability than Laboratory 9 in this experiment, so that the assumption that they achieve similar repeatability does not appear to be true in this instance. Some participants in a precision experiment may achieve poor repeatability when a measurement method is subjected to a precision experiment for the first time, or when they have little experience of the measurement method, and these are situations when the use of robust methods will be particularly appropriate.

6.1.3 The object of using robust methods[6], when analysing the data from a precision experiment, is to calculate values for the repeatability and reproducibility standard deviations in such a way that they are not influenced by outlying data. If the participants in the experiment can be considered to be divided into two classes, those that produce good-quality data, and those that produce poor-quality data, then the robust methods should yield values for the repeatability and reproducibility standard deviations that are valid for the good-quality-data class, and not affected by the poor-quality data (provided that the poor-quality-data class is not too large).

6.1.4 The use of robust methods of data analysis does not affect the planning, organization or execution of a precision experiment. The decision as to whether to use robust methods or methods that require outliers to be discarded should be made by the statistical expert, and reported to the panel. When robust methods are used, the outlier tests and consistency checks described in ISO 5725-2 or ISO 5725-5 should be applied to the data, and the causes of any outliers, or patterns in the h and k statistics, should be investigated. However, data should not be discarded as a result of these tests and checks.

6.1.5 The denominators in the h and k statistics are standard deviations that, according to the methods of calculating these statistics described in ISO 5725-2, are calculated from the data as reported. If outliers are present in the data then these will inflate the denominators and produce a distorted effect in the graphs of the statistics. For example, if, at one level of an experiment, one laboratory gives a cell average that is an outlier, and much more extreme than any other outlier at that level, then it will show up in a graph of h statistics as giving an exceptionally large h for that level. However, the h statistics for all the other laboratories for that level will be small, even if some of these other laboratories give outliers. The use of the overall average in the calculation of the h statistics can give rise to a similar effect. The use of robust estimates of the standard deviations as the denominators in the h and k statistics, and of robust estimates of the overall averages in the calculation of the h statistics, avoids this distortion. It is therefore recommended that they are used for this purpose.

6.1.6 The data from a precision experiment allow two types of statistic to be calculated:

a) cell averages, from which a standard deviation can be calculated that gives a measure of between-laboratory variation;

b) standard deviations or ranges (or differences in a split-level design) within cells that are combined to give a measure of within-laboratory variation.

The robust methods described here do not replace these cell averages, standard deviations or ranges or differences, but provide alternative ways of combining them to obtain the statistics that are used to calculate the repeatability and reproducibility standard deviations.

For example, with the data from one level of the uniform-level design considered in ISO 5725-2, the first stage of the analysis is to calculate the average and standard deviation of the measurement results in each cell. The cell averages are then used to calculate a standard deviation that is a measure of between-laboratory variation. When the robust methods of this clause are used, this calculation is performed using **'Algorithm A'** and cell averages are not excluded from the calculation as a result of applying Grubbs' test. Also in this design, the cell standard deviations are pooled to give an estimate of the repeatability standard deviation. With robust analysis, this is performed using **'Algorithm S'** and cell standard deviations are not excluded as a result of applying Cochran's test. With either approach (that described in ISO 5725-2 or that described here), the two measures are then used to calculate estimates of the repeatability and reproducibility standard deviations in the same way.

A more complicated example is the 6-factor staggered-nested design given in annex C of ISO 5725-3:1994. With this design, the first stage of the analysis is to calculate the average of the data for each laboratory (at each level), denoted $y_{i(1)}$, ... , $y_{i(5)}$, and a series of ranges, denoted $w_{i(1)}$, ... , $w_{i(5)}$, that contain information about the variability attributable to the various factors examined in the experiment. To analyse the data using the robust methods described here, **'Algorithm A'** is applied to the cell averages, and **'Algorithm S'** is applied to each series of ranges in turn. The statistics obtained by these operations are then used to obtain estimates of repeatability, intermediate precision and reproducibility standard deviations, in the same way as if the method of analysis described in ISO 5725-3 were being used.

6.1.7 The robust methods that are included in this part of ISO 5725 were chosen because they can be applied to all the experimental designs given in parts 2, 3, 4 and 5 of ISO 5725, and because the calculations they require are relatively simple. It should be noted, however, that they provide a means of combining, in a robust manner, cell averages, cell standard deviations and cell ranges. They do not combine individual test results in a robust manner, i.e. they start with the arithmetic cell averages and the cell standard deviations. There are robust methods that combine test results within cells in a robust manner, and they would be more complicated to apply in practice.

6.2 Robust analysis: Algorithm A

6.2.1 This algorithm yields robust values of the average and standard deviation of the data to which it is applied, and is applied to:

a) cell averages, in any design;

b) cell differences, in the split-level design.

6.2.2 Denote the p items of data, sorted into increasing order, by:

$$x_1,\ x_2,\ \ldots, x_i,\ \ldots,\ x_p$$

Denote the robust average and robust standard deviation of these data by x^* and s^*.

6.2.3 Calculate initial values for x^* and s^* as:

$$x^* = \text{median of } x_i \qquad (i = 1,\ 2,\ \ldots,\ p) \tag{56}$$

$$s^* = 1{,}483 \times \text{median of } |x_i - x^*| \qquad (i = 1,\ 2,\ \ldots,\ p) \tag{57}$$

6.2.4 Update the values of x^* and s^* as follows.

Calculate:

$$\varphi = 1{,}5\ s^* \tag{58}$$

For each x_i value $(i = 1,\ 2,\ \ldots,\ p)$, calculate:

$$x_i^* = \begin{cases} x^* - \varphi & \text{if} \quad x_i < x^* - \varphi \\ x^* + \varphi & \text{if} \quad x_i > x^* + \varphi \\ x_i & \text{otherwise} \end{cases} \tag{59}$$

Calculate the new values of x^* and s^* from:

$$x^* = \sum_{i=1}^{p} x_i^* / p \tag{60}$$

$$s^* = 1{,}134\sqrt{\sum_{i=1}^{p}\left(x_i^* - x^*\right)^2 \Big/ (p-1)} \qquad (61)$$

6.2.5 The robust estimates x^* and s^* may be derived by an iterative calculation, i.e. by repeating the calculations in 6.2.4 several times, until the change in the estimates of x^* and s^* from one calculation to the next is small. This is a simple method to program on a computer.

6.2.6 An alternative method that does not involve iteration and so may be easier to apply if the calculations are being done by hand, is derived by observing that equations (60) and (61) in 6.2.4 may be written:

$$x^* = x' + 1{,}5 \times \left(u_U - u_L\right)s^* \Big/ \left(p - u_L - u_U\right) \qquad (62)$$

$$(s^*)^2 = \left(p - u_L - u_U - 1\right) \times (s')^2 \Big/ \left[(p-1)/1{,}134^2 - 1{,}5^2\left(pu_L + pu_U - 4u_L u_U\right)/\left(p - u_L - u_U\right)\right] \qquad (63)$$

where

u_L is the number of data items x_i for which $x_i < x^* - \varphi$;

u_U is the number of data items x_i for which $x_i > x^* + \varphi$;

x' and s' are the average and standard deviation of the $(p - u_L - u_U)$ data items x_i for which $\left|x_i - x^*\right| \leq \varphi$.

These may be used to calculate x^* and s^* directly if u_L and u_U are known. One approach is to try the various possibilities in a systematic order (i.e. try $u_L = 0$, $u_U = 0$; then $u_L = 0$, $u_U = 1$; then $u_L = 1$, $u_U = 0$; then $u_L = 1$, $u_U = 1$; and so on) until a valid solution is found in which the actual numbers of items of data more than $1{,}5s^*$ from x^* equal the values of u_L and u_U used to calculate x^* and s^*. In practice, the analyst will be able to use histograms such as those shown in figure 4 to identify the values that are likely to differ from x^* by more than $1{,}5s^*$, and so find the solution by evaluating a small number of cases.

A further possibility is to use the iterative method to find an approximate solution, then solve equations (62) and (63) to find the exact solution. This is the approach that is used in the examples that follow.

6.3 Robust analysis: Algorithm S

6.3.1 This algorithm is applied to within-laboratory standard deviations (or within-laboratory ranges) in any design.

It yields a robust pooled value of the standard deviations or ranges to which it is applied.

6.3.2 Denote the p items of data, sorted into increasing order, by:

$$w_1, w_2, \ldots, w_i, \ldots, w_p$$

(These may be ranges or standard deviations).

Denote the robust pooled value by w^*, and the degrees of freedom associated with each w_i by v. (When w_i is a range, $v = 1$. When w_i is the standard deviation of n results, $v = n - 1$.) Obtain the values of ξ and η required by the algorithm from table 23.

6.3.3 Calculate an initial value for w^* as follows.

$$w^* = \text{median of } w_i \qquad (i = 1, 2, \ldots, p) \qquad (64)$$

6.3.4 Update the value of w^* as follows.

Calculate:

$$\psi = \eta \times w^* \tag{65}$$

For each w_i (i = 1, 2, ..., p), calculate:

$$w_i^* = \begin{cases} \psi & \text{if} \quad w_i > \psi \\ w_i & \text{otherwise} \end{cases} \tag{66}$$

Calculate the new value of w^* from:

$$w^* = \xi \sqrt{\sum_{i=1}^{p} \left(w_i^*\right)^2 \Big/ p} \tag{67}$$

6.3.5 The robust estimate w^* may be derived by an iterative calculation by repeating the calculations in 6.3.4 several times, until the change in the estimate of w^* from one calculation to the next is small. This is a simple method to program on a computer.

6.3.6 An alternative method, that does not involve iteration, and so may be easier to apply if the calculations are being done by hand, is similar to that described in 6.2.6. Equation (67) in 6.3.4 may be written:

$$\left(w^*\right)^2 = \left[\xi^2 / p\right] \times \left[\sum{}' \left(w_i^*\right)^2 + u_U \times \left(\eta w^*\right)^2\right] \tag{68}$$

where

$\sum'$ is the summation over the w_i for which $w_i \leq \psi$;

u_U is the number of w_i for which $w_i > \psi$.

This may be solved by trying u_U = 0, u_U = 1, u_U = 2, and so on in turn, until a valid solution is obtained in which the actual number of w_i that exceed $\eta \times w^*$ is equal to u_U. In practice the analyst will be able to use histograms such as those shown in figure 4 to identify the ranges that are likely to exceed $\eta \times w^*$ and find the solution by evaluating a small number of cases.

The approach that is used in the examples that follow is to use the iterative method to find an approximate solution, then solve equation (68) to find the exact solution.

Table 23 — Factors required for robust analysis: Algorithm S

Degrees of freedom ν	**Limit factor** η	**Adjustment factor** ξ
1	1,645	1,097
2	1,517	1,054
3	1,444	1,039
4	1,395	1,032
5	1,359	1,027
6	1,332	1,024
7	1,310	1,021
8	1,292	1,019
9	1,277	1,018
10	1,264	1,017
NOTE — The values of ξ and η are derived in annex B.		

6.4 Formulae: Robust analysis for a particular level of a uniform-level design

6.4.1 With a uniform-level design, a robust estimate of the repeatability standard deviation s_r for a level may be obtained by applying Algorithm S to the cell ranges or cell standard deviations for the level to derive a robust value w^* from equation (67) in 6.3.4. If Algorithm S is applied to the cell standard deviations, then

$$s_r = w^* \tag{69}$$

If there are two measurement results per cell and Algorithm S is applied to the cell ranges, then

$$s_r = w^* / \sqrt{2} \tag{70}$$

6.4.2 A robust estimate of the standard deviation of cell averages s_d for a level may be obtained by applying Algorithm A to the cell averages for the level to obtain a robust value s^* from equation (61) in clause 6.2.4, and then using

$$s_d = s^* \tag{71}$$

6.4.3 Next the between-laboratory standard deviation s_L may be derived using

$$s_L = \sqrt{s_d^2 - \left(s_r^2 / n\right)} \tag{72}$$

where n is the number of measurement results per cell.

If the expression under the square root is negative then set

$$s_L = 0 \tag{73}$$

Calculate the reproducibility standard deviation for the level as

$$s_R = \sqrt{s_L^2 + s_r^2} \tag{74}$$

6.5 Example 4: Robust analysis for a particular level of a uniform-level design

6.5.1 Example 3 in ISO 5725-2:1994 is an example of a uniform-level design in which the data contain stragglers and outliers. Level 5 in that example is of particular interest because Laboratory 1 gave a cell average that was shown to be a near-straggler by Grubbs' test, and Laboratory 6 gave a cell range that was shown to be a near-straggler by Cochran's test. These data are reproduced here in table 24.

6.5.2 If the data from all the laboratories are retained, the repeatability and reproducibility standard deviations may be estimated using the formulae in clause 7.4 of ISO 5725-2:1994, giving:

p = 9

m = 20,511

s_r = 0,585

s_d = 1,727

s_L = 1,677

s_R = 1,776

6.5.3 However, according to ISO 5725-2, the data analyst used information from other levels in the experiment, and suspicions concerning the identity of the samples tested by Laboratory 6, to justify excluding data from both Laboratories 1 and 6 from the calculation, giving:

p = 7

m = 20,412

s_r = 0,393

s_d = 0,573

s_L = 0,501

s_R = 0,637

Clearly the decision to exclude the data from the two laboratories has had a substantial effect on the estimates of the repeatability and reproducibility standard deviations.

6.5.4 The first step in the analysis is to obtain a robust estimate of the repeatability standard deviation. The calculations may be set out conveniently as shown in table 25, in which the cell ranges have been sorted into increasing order. Applying Algorithm S using iteration gives the results shown in this table. In this example the degrees of freedom of each cell range is ν = 1, so ξ = 1,097 and η = 1,645. From the four iterations shown in the table it appears that the robust value $w^* \approx 0{,}7$ and only one cell range ($w_9^* = 1{,}98$) exceeds ψ. If the calculations were being done on a computer, the process could be allowed to continue until the change in the value w^* from one iteration to the next is small.

The solution may also be derived directly, as follows. Using equation (68) in 6.3.6 with:

$u_U = 1$

$$\sum_{i=1}^{p}{}' \left(w_i^*\right)^2 \Big/ p = 0{,}249\,5$$

gives

$$(w^*)^2 = 1{,}097^2 \times 0{,}249\,5 + (1{,}097 \times 1{,}645\ w^*)^2 \Big/ 9$$

giving the solution (if the assumption u_U = 1 is correct) of:

w^* = 0,69 % creosote

It may then be confirmed that this value gives ψ = 1,645 × 0,69 = 1,14, so, as assumed, only w_9^* exceeds ψ and further that replacing w_9^* by 1,14 gives w^* = 0,63 × 1,097 = 0,69, again, so that the solution is shown to be valid.

The estimate of the repeatability standard deviation is thus:

$s_r = 0{,}69 / \sqrt{2} = 0{,}49\ \%$ creosote

This lies between the two estimates given in 6.5.2 and 6.5.3.

6.5.5 The next step in the analysis is to obtain a robust estimate of the standard deviation of the cell averages. Applying Algorithm A to the cell averages gives the results shown in table 26, where now the cell averages have been sorted into increasing order. From the four iterations shown in the table it appears that the robust values are x^* = 20,412 and $s^* \approx 1{,}1$, and that only the two extreme cell averages ($x_1^* = 17{,}570$, $x_9^* = 24{,}140$) differ from x^* by

more than φ. If the calculations were being done on a computer, the process could be allowed to continue until the changes in the values x^* and s^* from one iteration to the next are small.

If the calculations are being done by hand, the data analyst should use the direct method described in 6.2.6 and try:

$u_L = u_U = 1$

This gives

$x' = 20{,}412$ and $s' = 0{,}573$ % creosote

Hence from equations (62) and (63) in 6.2.6

$$(s^*)^2 = 6 \times (0{,}573)^2 / \left[8 / 1{,}134^2 - 1{,}5^2(9+9-4)/7\right]$$

giving:

$s^* = 1{,}070$ % creosote

and

$x^* = x' = 20{,}412$ % creosote

It may then be confirmed that this value of s^* gives φ = 1,605 (so, as assumed, only x_1^* and x_9^* differ from x^* = 20,412 by more than φ), and that replacing x_1^* by 18,807 and x_9^* by 22,017 gives a new value x^* = 20,412 and a new value s^* = 0,944 × 1,134 = 1,070 again, so the solution is shown to be valid.

The estimate of the between-laboratory standard deviation is thus, from equation (72) in 6.4.3:

$$s_L = \sqrt{1{,}070^2 - \left(0{,}49^2 / 2\right)} = 1{,}012 \text{ \% creosote}$$

and the estimate of the reproducibility standard deviation is, from equation (74) in 6.4.3:

$$s_R = \sqrt{1{,}012^2 + 0{,}49^2} = 1{,}124 \text{ \% creosote}$$

Again, this lies between the two estimates given in 6.5.2 and 6.5.3.

Table 24 — Example 4: Thermometric titration of creosote oil

Laboratory i	**Data** % creosote		**Cell average** % creosote	**Cell range** % creosote
1	24,28	24,00	24,140	0,28
2	20,40	19,91	20,155	0,49
3	19,30	19,70	19,500	0,40
4	20,30	20,30	20,300	0,00
5	20,53	20,88	20,705	0,35
6	18,56	16,58	17,570	1,98
7	19,70	20,50	20,100	0,80
8	21,10	20,78	20,940	0,32
9	20,71	21,66	21,185	0,95

Table 25 — Example 4: Applying Algorithm S to the cell ranges (% creosote)
(v = 1; ξ= 1,097; η= 1,645)

Iteration	0 [1]	1	2	3	4
ψ	—	0,66	0,86	1,00	1,09
w_1^*	0,00	0,00	0,00	0,00	0,00
w_2^*	0,28	0,28	0,28	0,28	0,28
w_3^*	0,32	0,32	0,32	0,32	0,32
w_4^*	0,35	0,35	0,35	0,35	0,35
w_5^*	0,40	0,40	0,40	0,40	0,40
w_6^*	0,49	0,49	0,49	0,49	0,49
w_7^*	0,80	0,66	0,80	0,80	0,80
w_8^*	0,95	0,66	0,86	0,95	0,95
w_9^*	1,98	0,66	0,86	1,00	1,09
Pooled w New w^*	0,83 0,40 [2]	0,47 0,52	0,56 0,61	0,60 0,66	0,62 0,68

1) The column for iteration 0 is taken from table 24 after rearrangement in increasing order.

2) 0,40 is the median range [see equation (64) in 6.3.3].

Table 26 — Example 4: Applying Algorithm A to the cell averages (% creosote)

Iteration	0 [1]	1	2	3	4
φ	—	1,424	1,478	1,514	1,539
$x^* - \varphi$ $x^* + \varphi$	— —	18,876 21,724	18,909 21,865	18,893 21,921	18,872 21,950
x_1^*	17,570	18,876	18,909	18,893	18,872
x_2^*	19,500	19,500	19,500	19,500	19,500
x_3^*	20,100	20,100	20,100	20,100	20,100
x_4^*	20,155	20,155	20,155	20,155	20,155
x_5^*	20,300	20,300	20,300	20,300	20,300
x_6^*	20,705	20,705	20,705	20,705	20,705
x_7^*	20,940	20,940	20,940	20,940	20,940
x_8^*	21,185	21,185	21,185	21,185	21,185
x_9^*	24,140	21,724	21,865	21,921	21,950
Average	20,511	20,387	20,407	20,411	20,412
Standard deviation	1,727	0,869	0,890	0,905	0,916
New x^* New s^*	20,300 [2] 0,949 [2]	20,387 0,985	20,407 1,009	20,411 1,026	20,412 1,039

1) The column for iteration 0 is taken from table 24 after rearrangement in increasing order.

2) 20,300 and 0,949 are derived using equations (56) and (57) in 6.2.3.

6.6 Formulae: Robust analysis for a particular level of a split-level design

6.6.1 With a split-level design, a robust estimate of the repeatability standard deviation s_r for a level may be obtained by applying Algorithm A to the cell differences for the level to derive a robust value s^*, from equation (61), and then calculating s_r using

$$s_r = s^* \sqrt{2} \tag{75}$$

6.6.2 A robust estimate of the standard deviation of cell averages s_y for a level may be obtained by applying Algorithm A again to the cell averages for the level to derive a robust value s^* from equation (61), and then obtaining s_y as:

$$s_y = s^* \tag{76}$$

The formulae given in 4.5.6 may then be used to calculate an estimate of the reproducibility standard deviation for the level.

6.7 Example 5: Robust analysis for a particular level of a split-level design

6.7.1 The data of Example 1 in 4.8 contained a number of stragglers and an outlier (see table 8). Also, the figure 3 showed a consistent negative bias in the results from Laboratory 5. If the data analyst cannot discover the causes of these anomalies then he/she is put in the difficult position of having to decide which data, if any, to exclude from the calculation of the repeatability and reproducibility standard deviations. The data for Level 14 (see table 4) are used here to illustrate the results that are obtained by robust analysis.

6.7.2 To obtain a robust estimate of the repeatability standard deviation, apply Algorithm A to the cell differences (from table 5). This gives the results shown in table 27, in which the cell differences have been sorted into increasing order. From the four iterations shown in the table, it appears that the robust values are $x^* \approx 8{,}29$ and $s^* \approx 0{,}36$, and that only x_9^* differs from x^* by more than φ.

Applying the method described in 6.2.6 with $u_L = 0$ and $u_U = 1$ gives

$$x' = 8{,}219 \qquad \text{and} \qquad s' = 0{,}257 \text{ \% protein}$$

so equations (62) and (63) in 6.2.6 may be written:

$$x^* = 8{,}219 + 1{,}5 \times s^* / 8$$

and

$$(s^*)^2 = 7 \times (0{,}257)^2 / \left[8 / 1{,}134^2 - 1{,}5^2 (0 + 9 - 0) / 8\right]$$

giving

$$s^* = 0{,}354 \text{ \% protein}$$

and, using equation (75) in 6.6.1,

$$s_r = 0{,}354 / \sqrt{2} = 0{,}250 \text{ \% protein}$$

The robust average of the cell differences is

$$x^* = 8{,}219 + 1{,}5 \times 0{,}354 / 8 = 8{,}285 \text{ \% protein}$$

With these values of x^* and s^*,

$$\varphi = 1{,}5 \times 0{,}354 = 0{,}531$$

so

$$x^* - \varphi = 7{,}754 \quad \text{and} \quad x^* + \varphi = 8{,}816 \text{ \% protein}$$

It was assumed for the calculation of x^* and s^* that only x_9^* was outside these limits. It may be confirmed that this is indeed the case, so that a valid solution has been obtained.

6.7.3 Applying Algorithm A to the cell averages (from table 6) gives the results shown in table 28, in which the cell averages have been entered in increasing order. The situation here is similar to that found in table 26, where x_1^* and x_9^* are more than φ from x^*, and x^* is converging on the average of x_2^* to x_8^*, namely 85,486. Applying the method of 6.2.6 again, but with $u_L = u_U = 1$, the average and standard deviation of x_2^* to x_8^* are:

$$x' = 85{,}486 \quad \text{and} \quad s' = 0{,}209$$

Hence equation (63) in 6.2.6 may be used to derive s^* from:

$$(s^*)^2 = 6 \times (0{,}209)^2 / \left[8 / 1{,}134^2 - 1{,}5^2 (9 + 9 - 4)/7\right]$$

giving

$$s^* = 0{,}390 \text{ \% protein}$$

Now equation (62) in 6.2.6 may be used to give

$$x^* = 85{,}486 \text{ \% protein}$$

To check that the solution is valid, calculate

$$\varphi = 1{,}5 \times 0{,}390 = 0{,}585$$

and

$$x^* - \varphi = 84{,}901 \quad \text{and} \quad x^* + \varphi = 86{,}071 \text{ \% protein}$$

It will be seen that, as assumed, only x_1^* and x_9^* fall outside these limits.

To obtain the reproducibility standard deviation, use equation (76) in 6.6.2 to give:

$$s_y = 0{,}390 \text{ \% protein}$$

and then equation (13) in 4.5.6 to give:

$$s_R = 0{,}410 \text{ \% protein}$$

Hence, in this example, the robust method gives estimates s_r and s_R that are somewhat smaller than the values obtained using all the data as reported (given in table 7).

Table 27 — Example 5: Applying Algorithm A to the cell differences (% protein)

Iteration	0	1	2	3	4
φ	—	0,53	0,56	0,55	0,54
$x^* - \varphi$ $x^* + \varphi$	— —	7,85 8,91	7,74 8,86	7,74 8,84	7,75 8,83
x_1^*	7,81	7,85	7,81	7,81	7,81
x_2^*	7,93	7,93	7,93	7,93	7,93
x_3^*	8,13	8,13	8,13	8,13	8,13
x_4^*	8,14	8,14	8,14	8,14	8,14
x_5^*	8,38	8,38	8,38	8,38	8,38
x_6^*	8,40	8,40	8,40	8,40	8,40
x_7^*	8,44	8,44	8,44	8,44	8,44
x_8^*	8,52	8,52	8,52	8,52	8,52
x_9^*	9,31	8,91	8,86	8,84	8,83
Average	8,340	8,300	8,290	8,288	8,287
Standard deviation	0,436	0,326	0,322	0,317	0,315
New x^* New s^*	8,380 [1)] 0,356 [1)]	8,300 0,370	8,290 0,365	8,288 0,359	8,287 0,357
1) Obtained using equations (56) and (57) in 6.2.3.					

Table 28 — Example 5: Applying Algorithm A to the cell averages (% protein)

Iteration	0	1	2	3	4
φ	—	0,446	0,492	0,519	0,537
$x^* - \varphi$ $x^* + \varphi$	— —	85,104 85,996	85,009 85,993	84,971 86,009	84,950 86,024
x_1^*	84,525	85,104	85,009	84,971	84,950
x_2^*	85,140	85,140	85,140	85,140	85,140
x_3^*	85,345	85,345	85,345	85,345	85,345
x_4^*	85,385	85,385	85,385	85,385	85,385
x_5^*	85,550	85,550	85,550	85,550	85,550
x_6^*	85,575	85,575	85,575	85,575	85,575
x_7^*	85,660	85,660	85,660	85,660	85,660
x_8^*	85,750	85,750	85,750	85,750	85,750
x_9^*	86,170	85,996	85,993	86,009	86,024
Average	85,456	85,501	85,490	85,487	85,487
Standard deviation	0,453	0,289	0,305	0,316	0,324
New x^* New s^*	85,550 [1)] 0,297 [1)]	85,501 0,328	85,490 0,346	85,487 0,358	85,487 0,367
1) Obtained using equations (56) and (57) in 6.2.3.					

6.8 Formulae: Robust analysis for a particular level of an experiment on a heterogeneous material

6.8.1 With the design for a heterogeneous material, in the usual case when two samples are prepared for each of p' laboratories at each level, and two test results are obtained on each sample, robust estimates of the repeatability and reproducibility standard deviations may be obtained by three applications of Algorithms A and S as follows.

a) Apply Algorithm S to the between-test-result ranges to derive a robust value w^* from equation (67) in 6.3.4, and set

$$SS_r = 2p'(w^*)^2 \tag{77}$$

b) Apply Algorithm S to the between-sample ranges to derive another robust value w^* from equation (67), and set

$$SS_H = p'(w^*)^2 \tag{78}$$

c) Apply Algorithm A to the cell averages to derive a robust value s^* from equation (61) in 6.2.4, and set

$$s_y = s^* \tag{79}$$

These calculations may be set out conveniently in tabular form, with the first column of ranges or averages entered in increasing order, as shown in the examples that follow.

6.8.2 The formulae given in 5.5 may then be used to calculate estimates of the repeatability and reproducibility standard deviations, and of the standard deviation s_H that measures variation between the samples.

6.9 Example 6: Robust analysis for a particular level of an experiment on a heterogeneous material

6.9.1 The data for Level 6 of Example 2 in 5.8 do not contain any outliers or stragglers, so they are used here to illustrate the results that are obtained by robust analysis in such a case.

6.9.2 Applying Algorithm S to the between-test-result ranges (from table 14) gives the results shown in table 29. Here the degrees of freedom $\upsilon = 1$ so $\eta = 1{,}645$ and $\xi = 1{,}097$, and the number of data items $p = 2p' = 22$. From the four iterations shown in the table, it appears that the robust value is $w^* \approx 4{,}5$ and that w^*_{19} to w^*_{22} exceed ψ. With Σ' and u_U defined as in 6.3.6, in this case:

$u_U = 4$

$$\sum{}'\left(w_i^*\right)^2 \Big/ p = 137{,}92 / 22 = 6{,}269\,1$$

so equation (68) in 6.3.6 becomes

$$(w^*)^2 = 1{,}097^2 \times 6{,}269\,1 + 4(1{,}097 \times 1{,}645\ w^*)^2 / 22$$

giving

$$w^* = 4{,}30\ \%$$

It may then be confirmed that with this value of w^*, $\psi = 7{,}1$, and the four values w^*_{19} to w^*_{22} exceed ψ, so that this value of w^* is a valid solution.

Using equation (77) in 6.8.1 then gives:

$$SS_r = 22 \times 4{,}30^2 = 406{,}78\ \%^2$$

6.9.3 Applying Algorithm S a second time, to the between-sample ranges (from table 15) gives the results shown in table 30. From the four iterations shown in the table, it appears that the robust value is $w^* \approx 4{,}0$, and that w_{10}^* and w_{11}^* exceed ψ. With Σ' and u_L defined as in 6.3.6, in this case:

$$u_U = 2$$

$$\sum{}'\left(w_i^*\right)^2 \Big/ p' = 66{,}665 / 11 = 6{,}060\ 5$$

so equation (68) in 6.3.6 becomes

$$(w^*)^2 = 1{,}097^2 \times 6{,}060\ 5 + 2(1{,}097 \times 1{,}645\ w^*)^2 / 11$$

giving

$$w^* = 4{,}23\ \%$$

Unfortunately, this corresponds to $\psi = 1{,}645 \times 4{,}23 = 6{,}96$, so it is not valid solution because w_{10}^* and w_{11}^* do not exceed this value. This suggests that the solution may require $u_U = 1$ or $u_U = 0$.

Trying $u_U = 1$ gives

$$\sum{}'\left(w_i^*\right)^2 \Big/ p' = 112{,}227\ 5 / 11 = 10{,}202\ 5$$

so equation (68) now becomes

$$(w^*)^2 = 1{,}097^2 \times 10{,}202\ 5 + (1{,}097 \times 1{,}645\ w^*)^2 / 11$$

giving

$$w^* = 4{,}18\ \%$$

Now $\psi = 1{,}645 \times 4{,}18 = 6{,}88$, and this can be seen to be a valid solution because only w_{11}^* exceeds this value.

Using equation (78) in 6.8.1) then gives:

$$SS_H = 11 \times 4{,}18^2 = 192{,}20\ \%^2$$

6.9.4 Applying Algorithm A to the cell averages (from table 16) gives the results shown in table 31. The calculation converges after two iterations to give $s^* = 5{,}70$ (with no x_i^* more than φ from x^*).

Using equation (79) in 6.8.1 then gives:

$$s_y = 5{,}70\ \%$$

6.9.5 Combining the results obtained in 6.9.2, 6.9.3 and 6.9.4, equations (29) to (33) in 5.5.5 now give:

$$s_r^2 = 406{,}78 / 44 \ \%^2$$

$$s_R^2 = 5{,}70^2 + (406{,}78 - 192{,}20) / 44 \ \%^2$$

$$s_H^2 = 192{,}20 / 22 - 406{,}78 / 88 \ \%^2$$

so

$$s_r = 3{,}04 \ \%$$

$$s_R = 6{,}11 \ \%$$

$$s_H = 2{,}03 \ \%$$

Hence in this example, the robust method gives estimates s_r, s_R and s_H that are marginally larger than the values obtained using all the data as reported (given in 5.8.3 and table 17).

Table 29 — Example 6: Applying Algorithm S to between-test-result ranges
(ν = 1; ξ = 1,097; η = 1,645)

Iteration	0	1	2	3	4
ψ	—	3,9	5,1	5,9	6,4
w_1^*	0,1	0,1	0,1	0,1	0,1
w_2^*	0,6	0,6	0,6	0,6	0,6
w_3^*	1,1	1,1	1,1	1,1	1,1
w_4^*	1,1	1,1	1,1	1,1	1,1
w_5^*	1,2	1,2	1,2	1,2	1,2
w_6^*	1,3	1,3	1,3	1,3	1,3
w_7^*	1,4	1,4	1,4	1,4	1,4
w_8^*	1,6	1,6	1,6	1,6	1,6
w_9^*	1,8	1,8	1,8	1,8	1,8
w_{10}^*	2,1	2,1	2,1	2,1	2,1
w_{11}^*	2,2	2,2	2,2	2,2	2,2
w_{12}^*	2,5	2,5	2,5	2,5	2,5
w_{13}^*	2,6	2,6	2,6	2,6	2,6
w_{14}^*	3,9	3,9	3,9	3,9	3,9
w_{15}^*	4,0	3,9	4,0	4,0	4,0
w_{16}^*	4,4	3,9	4,4	4,4	4,4
w_{17}^*	4,6	3,9	4,6	4,6	4,6
w_{18}^*	5,5	3,9	5,1	5,5	5,5
w_{19}^*	7,4	3,9	5,1	5,9	6,4
w_{20}^*	7,6	3,9	5,1	5,9	6,4
w_{21}^*	8,1	3,9	5,1	5,9	6,4
w_{22}^*	8,1	3,9	5,1	5,9	6,4
New w New w^*	4,17 2,35 [1)]	2,80 3,07	3,29 3,61	3,55 3,89	3,70 4,06

1) Obtained using equation (64) in 6.3.3.

Table 30 — Example 6: Applying algorithm S to between-sample ranges (%)
(v = 1; ξ = 1,097; η = 1,645)

Iteration	0	1	2	3	4
ψ	—	4,19	5,43	6,10	6,45
w_1^*	1,00	1,00	1,00	1,00	1,00
w_2^*	1,70	1,70	1,70	1,70	1,70
w_3^*	2,05	2,05	2,05	2,05	2,05
w_4^*	2,25	2,25	2,25	2,25	2,25
w_5^*	2,55	2,55	2,55	2,55	2,55
w_6^*	2,55	2,55	2,55	2,55	2,55
w_7^*	3,15	3,15	3,15	3,15	3,15
w_8^*	3,35	3,35	3,35	3,35	3,35
w_9^*	4,40	4,19	4,40	4,40	4,40
w_{10}^*	6,75	4,19	5,43	6,10	6,45
w_{11}^*	6,95	4,19	5,43	6,10	6,45
New w New w^*	3,82 2,55 [1)]	3,01 3,30	3,38 3,71	3,58 3,92	3,69 4,05

1) Obtained using equation (64) in 6.3.3.

Table 31 — Example 6: Applying algorithm A to cell averages (%)

Iteration	0	1	2	3	4
φ	—	10,005	8,550		
$x^* - \varphi$ $x^* + \varphi$	— —	8,245 28,255	10,450 27,550		
x_1^*	13,425	13,425	13,425		
x_2^*	13,425	13,425	13,425		
x_3^*	13,750	13,750	13,750		
x_4^*	14,475	14,475	14,475		
x_5^*	17,075	17,075	17,075		
x_6^*	18,250	18,250	18,250		
x_7^*	21,000	21,000	21,000		
x_8^*	21,225	21,225	21,225		
x_9^*	23,675	23,675	23,675		
x_{10}^*	26,275	26,275	26,275		
x_{11}^*	26,425	26,425	26,425		
Average	19,00	19,00	19,00		
Standard deviation	5,03	5,03	5,03		
New x^* New s^*	18,25 [1)] 6,67 [1)]	19,00 5,70	19,00 5,70		

1) Obtained using equations (56) and (57) in 6.2.3.

Annex A
(normative)

Symbols and abbreviations used in ISO 5725

a	Intercept in the relationship $s = a + bm$
A	Factor used to calculate the uncertainty of an estimate
b	Slope in the relationship $s = a + bm$
B	Component in a test result representing the deviation of a laboratory from the general average (laboratory component of bias)
B_0	Component of B representing all factors that do not change in intermediate precision conditions
$B_{(1)}$, $B_{(2)}$, etc.	Component of B representing factors that vary in intermediate precision conditions
c	Intercept in the relationship $\lg s = c + d \lg m$
C, C', C''	Test statistics
C_{crit}, C'_{crit}, C''_{crit}	Critical values for statistical tests
CD_P	Critical difference for probability P
CR_P	Critical range for probability P
d	Slope in the relationship $\lg s = c + d \lg m$
e	Component in a test result representing the random error occurring in every test result
f	Critical range factor
$F_p(\upsilon_1, \upsilon_2)$	p-quantile of the F-distribution with υ_1 and υ_2 degrees of freedom
G	Grubbs' test statistic
h	Mandel's between-laboratory consistency test statistic
k	Mandel's within-laboratory consistency test statistic
LCL	Lower control limit (either action limit or warning limit)
m	General mean of the test property; level
M	Number of factors considered in intermediate precision conditions
n	Number of test results obtained in one laboratory at one level (i.e. per cell)
N	Number of iterations
p	Number of laboratories participating in the interlaboratory experiment
P	Probability
q	Number of levels of the test property in the interlaboratory experiment
r	Repeatability limit
R	Reproducibility limit
RM	Reference material
s	Estimate of a standard deviation
$\hat{s}$	Predicted standard deviation
t	Number of test objects or groups
T	Total or sum of some expression
UCL	Upper control limit (either action limit or warning limit)
w	Range of a set of test results
W	Weighting factor used in calculating a weighted regression
x	Datum used for Grubbs' test
y	Test result
$\bar{y}$	Arithmetic mean of test results

$\bar{\bar{y}}$	Grand mean of test results
α	Significance level
β	Type II error probability
γ	Ratio of the reproducibility standard deviation to the repeatability standard deviation (σ_R/σ_r)
δ	Bias of the measurement method
$\hat{\delta}$	Estimate of δ
Δ	Laboratory bias
$\hat{\Delta}$	Estimate of Δ
λ	Detectable difference between two laboratory biases or the biases of two measurement methods
μ	True value or accepted reference value of a test property
υ	Number of degrees of freedom
ρ	Detectable ratio between the repeatability standard deviations of method B and method A
σ	True value of a standard deviation
τ	Component in a test result representing the variation due to time since last calibration
ϕ	Detectable ratio between the square roots of the between-laboratory mean squares of method B and method A
$\chi^2_p(\upsilon)$	p-quantile of the χ^2-distribution with υ degrees of freedom

Symbols used as subscripts

C	Calibration-different
E	Equipment-different
i	Identifier for a particular laboratory
$I(\ )$	Identifier for intermediate measures of precision; in brackets, identification of the type of intermediate situation
j	Identifier for a particular level (ISO 5725-2). Identifier for a group of tests or for a factor (ISO 5725-3)
k	Identifier for a particular test result in a laboratory i at level j
L	Between-laboratory (interlaboratory)
m	Identifier for detectable bias
M	Between-test-sample
O	Operator-different
P	Probability
r	Repeatability
R	Reproducibility
T	Time-different
W	Within-laboratory (intralaboratory)
1, 2, 3...	For test results, numbering in the order of obtaining them
(1), (2), (3)...	For test results, numbering in the order of increasing magnitude

Additional symbols and abbreviations used in ISO 5725-5

D	Within-cell difference in a split-level experiment
g	Number of samples tested in a laboratory at one level
H	Component in a measurement result representing the random error of the sample
K	Function of the numbers of measurement results in cells
p'	Number of laboratories participating in the interlaboratory experiment
SS	Sum of squares
u_L	Number of data values below a lower limit in robust analysis

u_U Number of data values above an upper limit in robust analysis

z Residual

Φ Ratio of standard deviations

φ Limit used in robust analysis (Algorithm A)

η Limit factor used in robust analysis (Algorithm S)

ψ Limit used in robust analysis (Algorithm S)

ζ Adjustment factor used in robust analysis (Algorithm S)

Additional symbols used as subscripts in ISO 5725-5

a, b Identifiers for samples in a split-level experiment

t Identifier for a sample in a laboratory i at level j

H Between-sample

Additional symbol used as a superscript in ISO 5725-5

* A robust estimate

Annex B
(informative)

Derivation of the factors used in Algorithms A and S

B.1 Introduction

The use of robust methods of analysing the data from precision experiments was proposed by the Analytical Methods Committee of the Royal Society of Chemistry, UK (see reference [6]). “Algorithm A” in this part of ISO 5725 is derived from their paper, and the factor 1,134 used to calculate s^* in Algorithm A is also taken from their paper (It is the value of $1/\sqrt{\theta}$ for the case c = 1,5 in the notation of the paper).

“Algorithm S” is similar to the procedure given in reference [6] for the special case where each laboratory reports n = 2 measurements per level. It provides a convenient method of applying robust analysis to precision experiments with more than two factors (such as the design for a heterogeneous material in clause 5 of this part of ISO 5725, and the staggered-nested designs of ISO 5725-3). The derivation of the factors used in algorithm S is set out below.

B.2 Symbols used in this annex

σ true standard deviation

s standard deviation that gives an estimate of σ

v degrees of freedom of s

ω $v + 2$

ξ adjustment factor for Algorithm S

η limit factor for Algorithm S

χ^2_v chi-squared variate with v degrees of freedom

$$s^* = \begin{cases} s & \text{if } s \le \eta\sigma \\ \phi\sigma & \text{if } s > \eta\sigma \end{cases}$$

B.3 Derivation of the limit factor η and the adjustment factor ξ

The adjustment factor ξ is defined as the value needed to adjust s^* so that $(s^*)^2$ gives an unbiased estimate of σ^2, i.e. so that:

$$E\left\{\left(\xi \times s^*\right)^2\right\} = \sigma^2 \qquad \text{(B.1)}$$

This requirement may be written:

$$E\left\{v\left(s^*/\sigma\right)^2\right\} = v/\xi^2 \qquad \text{(B.2)}$$

where the quantity in $\{\ \}$ is closely related to $v(s/\sigma)^2$, a chi-squared variate, χ^2_v.

The probability density function of the chi-squared variate χ^2_v is:

$$f(x) = \mathrm{e}^{-x/2} x^{(\nu/2-1)} 2^{-\nu/2} / \Gamma(\nu/2) \tag{B.3}$$

so

$$E\left\{\nu(s^*/\sigma)^2\right\} = \int_0^{\nu\eta^2} x f(x)\,\mathrm{d}x + \int_{\nu\eta^2}^{\infty} \nu\eta^2 f(x)\,\mathrm{d}x \tag{B.4}$$

because the limit $s \leq \eta\sigma$ is equivalent to $\nu(s/\sigma)^2 \leq \nu\eta^2$.

The second term on the right-hand side of (B.4) is

$$\nu\eta^2 \times P\left(\chi^2_\nu > \nu\eta^2\right) = \nu\eta^2 \times P(s > \eta\sigma) \tag{B.5}$$

For Algorithm S, the limit factor η is chosen so that $\eta\sigma$ is the upper 10 % point of the distribution of s, i.e. so that:

$$P(s > \eta\sigma) = 0{,}1 \tag{B.6}$$

Biometrika Tables for the chi-squared distribution give the values of η shown in table 23 of this part of ISO 5725. Equations (B.5) and B.6) imply that the second term on the right-hand side of equation (B.4) is 0,1 $\nu\eta^2$. Note that η depends on the degrees of freedom of s.

The first term on the right-hand side of equation (B.4) may be written:

$$\int_0^{\nu\eta^2} x\mathrm{e}^{-x/2} x^{\nu/2-1} 2^{-\nu/2} / \Gamma(\nu/2)\,\mathrm{d}x$$

With $\omega = \nu + 2$, a well-known property of the gamma function is that:

$$\Gamma(\omega/2) = \Gamma(\nu/2 + 1) = (\nu/2) \times \Gamma(\nu/2)$$

so this first term may be rewritten as:

$$\int_0^{\nu\eta^2} \nu\mathrm{e}^{-x/2} x^{\omega/2-1} 2^{-\omega/2} / \Gamma(\omega/2)\,\mathrm{d}x$$

$$= \nu \times P\left(\chi^2_\omega < \nu\eta^2\right)$$

$$= \nu \times z \tag{B.7}$$

Thus for a given number of degrees of freedom ν, η may be calculated as described above, and then z may be evaluated using tables of the chi-squared distribution given in Biometrika Tables again, so that both terms on the right-hand side of (B.4) may be evaluated.

Substituting equations (B.2), (B.5), (B.6) and (B.7) in (B.4), gives:

$$\nu / \xi^2 = \nu \times z + 0{,}1\,\nu\eta^2$$

or

$$\xi = 1/\sqrt{z + 0{,}1\,\eta^2} \tag{B.8}$$

This may be used to obtain the values of the adjustment factor ξ given in table 23 of this part of ISO 5725.

Annex C
(informative)

Derivation of equations used for robust analysis

Equations (62) and (63), used to calculate robust values of the average and standard deviation by the method described in 6.2.6, may be derived from equations (60) and (61) in 6.2.4 (Algorithm A) as follows.

With the notation as set out in 6.2.4 and 6.2.6:

$$x^* = \sum x_i^* / p \tag{C.1}$$

$$x' = \sum{}' x_i \left(p - u_L - u_U\right) \tag{C.2}$$

and

$$s' = \sqrt{\sum{}' (x_i - x')^2 / \left(p - u_L - u_U - 1\right)} \tag{C.3}$$

where $\sum'$ represents summation over the ($p - u_L - u_U$) data items x_i for which $\left|x_i - x_i^*\right| \le \varphi$.

Hence equation (C.1) may be written:

$$\begin{aligned} p \times x^* &= \sum x_i^* \\ &= \sum{}' x_i + u_L \times (x^* - 1{,}5\ s^*) + u_U \times (x^* + 1{,}5\ s^*) \end{aligned}$$

so

$$\left(p - u_L - u_U\right) \times x^* = \left(p - u_L - u_U\right) \times x' + \left(u_L - u_U\right) \times s^*$$

giving

$$x^* = x' + 1{,}5\left(u_U - u_L\right) s^* / \left(p - u_L - u_U\right) \tag{C.4}$$

which is equation (62) in 6.2.6.

To derive equation (63) from equation (61), note that the summation term in equation (61) may be expanded as follows:

$$\sum \left(x_i^* - x^*\right)^2 = \sum{}' (x_i - x^*)^2 + \left(u_L + u_U\right) \times (1{,}5\ s^*)^2 \tag{C.5}$$

Substituting for x^* in the summation term here using equation (C.4) gives, after some algebra:

$$\sum \left(x_i^* - x^*\right)^2 = \sum{}' (x_i - x')^2 + (1{,}5\ s^*)^2 \times \left(pu_L + pu_U - 4u_L u_U\right) / \left(p - u_L - u_U\right) \tag{C.6}$$

Using the definition of s' in equation (C.3), this may be written:

$$\sum \left(x_i^* - x^*\right)^2 = \left(p - u_L - u_U - 1\right) \times (s')^2 + (1{,}5\ s^*)^2 \times \left(pu_L + pu_U - 4u_L u_U\right) / \left(p - u_L - u_U\right) \tag{C.7}$$

Substituting equation (C.7) in equation (61), then gives equation (63).

Annex D
(informative)

Bibliography

[1] BS 812-103:1985, *Testing aggregates — Part 103: Methods for determination of particle size distribution.* British Standards Institution.

[2] BS 812-121:1989, *Testing aggregates — Part 121: Methods for determination of soundness.* British Standards Institution.

[3] BS 3144:1968, *Methods of sampling and physical testing of leather*. British Standards Institution.

[4] SCHEFFE, H. *The analysis of variance*. Wiley, New York, 1959.

[5] SWEENEY, An inter-laboratory study of the determination of protein by combustion in feeds. *Journal of the Association of Official Analytical Chemists,* **72**, 1989, pp. 770-774.

[6] ANALYTICAL METHODS COMMITTEE. Robust statistics — How not to reject outliers. Part 1: Basic concepts. Part 2: Inter-laboratory trials. *The Analyst*, **114**, 1989, pp. 1653-1697 (part 1), pp. 1699-1702 (part 2). Royal Society of Chemistry, London.

[7] YOUDEN, W.J. The Youden plot. *Industrial Quality Control,* **15**, 1959, pp. 133-137.

[8] MANDEL, J. and LASHOF, T.W. Interpretation and Generalization of Youden's Two-Sample Diagram. *Journal of Quality Technology*, **6**, 1974, pp. 22-36.

ICS 03.120.30; 17.020

Descriptors: measurement, tests, test results, accuracy, precision, reproducibility, rules of calculation, statistical analysis.

Price based on 56 pages