International Standard 5725

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION•МЕЖДУНАРОДНАЯ ОРГАНИЗАЦИЯ ПО СТАНДАРТИЗАЦИИ•ORGANISATION INTERNATIONALE DE NORMALISATION

Precision of test methods — Determination of repeatability and reproducibility by inter-laboratory tests

Fidélité des méthodes d'essai — Détermination de la répétabilité et de la reproductibilité par essais interlaboratoires

First edition — 1981-04-01

UDC 519.2 : 620.1

Ref. No. ISO 5725-1981 (E)

Descriptors : tests, reproducibility, statistical analysis.

Price based on 42 pages

Foreword

ISO (the International Organization for Standardization) is a worldwide federation of national standards institutes (ISO member bodies). The work of developing International Standards is carried out through ISO technical committees. Every member body interested in a subject for which a technical committee has been set up has the right to be represented on that committee. International organizations, governmental and non-governmental, in liaison with ISO, also take part in the work.

Draft International Standards adopted by the technical committees are circulated to the member bodies for approval before their acceptance as International Standards by the ISO Council.

International Standard ISO 5725 was developed by Technical Committee ISO/TC 69, *Applications of statistical methods*, and was circulated to the member bodies in October 1977.

It has been approved by the member bodies of the following countries :

Australia	Hungary	Romania
Austria	India	South Africa, Rep. of
Belgium	Israel	Spain
Brazil	Italy	Sweden
Bulgaria	Japan	Switzerland
Canada	Kenya	Turkey
Czechoslovakia	Mexico	United Kingdom
Denmark	Netherlands	USSR
France	New Zealand	Yugoslavia
Germany, F. R.	Poland	

The member body of the following country expressed disapproval of the document on technical grounds :

USA

Printed in Switzerland

Contents

Precision of test methods — Determination of repeatability and reproducibility by inter-laboratory tests

0 Introduction

0.1 Tests performed on presumably "identical materials" (see 4.2) in presumably identical circumstances do not, in general, yield identical results. This is attributed to unavoidable random errors inherent in every test procedure; the factors that may influence the outcome of a test cannot all be completely controlled. In the practical interpretation of test data, this variability has to be taken into account. For instance, the difference between a test result and a value specified by contract may be within the scope of unavoidable random errors, in which case a true deviation from specification has not been established. Similarly, comparing test results from two batches of material will not indicate a fundamental quality difference if the difference between them can be attributed to inherent variation in the test procedure.

0.2 Many different factors (apart from error due to a lack of homogeneity of samples) may contribute to the variability of a test procedure, for example

a) the operator;

b) the instruments and equipment used;

c) the calibration of the equipment;

d) the environment (temperature, humidity, air pollution), etc.

The variability will be larger when the tests to be compared have been performed by different operators and/or with different instruments than when they have been carried out by a single operator using the same instruments. Hence, many different measures of variability are conceivable according to the circumstances under which the tests have been performed.

0.3 However, two extreme measures of variability, termed repeatability and reproducibility, have been found sufficient to deal with most practical cases. Repeatability refers to tests performed at short intervals (see 4.3) in one laboratory by one operator, using the same equipment, while reproducibility refers to tests performed in different laboratories, which implies different operators and different equipment. Under repeatability conditions, factors a) to d) listed in 0.2 are considered as constants and do not contribute to the variability, while under reproducibility conditions they vary and contribute to the variability of the test results.

1 Scope

This International Standard provides practical numerical definitions for the repeatability r and the reproducibility R of the results of a standard test method.

It discusses the implications of these definitions, and presents some practical rules for the interpretation of r and R.

It also describes the organization and analysis of inter-laboratory experiments for the numerical determination of r and R.

2 Field of application

This International Standard is exclusively concerned with test methods the results of which are expressed quantitatively.

This International Standard is primarily intended to be applied to test methods that have previously been standardized and that are used in different laboratories.

With slight modifications this International Standard may also be applied to test methods in use within a single laboratory (see 3.1.5) but this case has not been dealt with in this document.

Only the simplest type of experiment needed for estimating r and R is considered. This consists of tests made on samples of identical material sent to a number of different laboratories for testing.

This International Standard does not provide measures of the errors in estimating r and R (see 3.5).

Section one : General principles

3 Quantitative definitions of repeatability and reproducibility of a standard test method

3.1 For practical purposes, quantitative definitions are needed; those given below are according to ISO 3534.[1]

The **repeatability** r is the value below which the absolute difference between two single test results obtained with the same method on identical test material, under the same conditions (same operator, same apparatus, same laboratory, and a short interval of time), may be expected to lie with a specified probability; in the absence of other indications, the probability is 95 %.

The **reproducibility** R is the value below which the absolute difference between two single test results obtained with the same method on identical test material, under different conditions (*different operators, different apparatus, different* laboratories and/or different time), may be expected to lie with a specified probability; in the absence of other indications, the probability is 95 %.

In the above and elsewhere in this International Standard, a single test result is the value obtained by applying the standard test method fully once to a single specimen, and as such may be the mean of two or more observations or the result of a calculation from a set of observations as specified by the method.

3.1.1 The definitions apply to continuous variables. When the test result is discrete or is rounded off, r and R are each the minimum value equal to or below which the absolute difference between two single test results is expected to lie with a specified probability (95 % in the absence of other indications).

3.1.2 The symbols r and R for the repeatability and reproducibility *are already in general use.* In ISO 3534, r is recommended for the correlation coefficient and R (or w) for the range of a single series of observations. There should, however, be no confusion as, when quoted in a standard method, the full wording repeatability r and reproducibility R should be used.

3.1.3 The statistical analysis of section three aims at the determination of r and R corresponding to a 95 % probability. Values for other probabilities can easily be derived from these, as explained in section four. If required, the probability level adopted can be attached as a subscript, for example r_{95}, R_{95}, or r_{99}, R_{99}.

3.1.4 The definition of repeatability in 3.1 applies to any test method within any laboratory. When a test method has been standardized, it may be expected that the repeatability will be, at least approximately, the same for all laboratories using the standard procedure; and the main purpose of this International Standard is to establish a standard experimental method for determining the repeatability of a standard test method.

With slight modifications, however, the same type of experiment can also be used to determine the repeatability of a test method in use within a single laboratory. If so, it should always be stated clearly that the value of the repeatability obtained is only valid within the laboratory in question.

3.1.5 The terms reproducibility and repeatability as defined in 3.1 cover the conditions of maximum and minimum variation respectively. Other intermediate measures could be envisaged, for example the variability of results within a laboratory over a long period of time when recalibration may have occurred. Such measures have not been dealt with in this document.

3.2 The terms repeatability and reproducibility are used because they have been in common use for several years, but according to statistical terminology, r and R are critical differences at the 95 % probability level valid for two single test results obtained under repeatability or reproducibility conditions. Also, it is sometimes the practice to carry out two or more tests and a critical difference corresponding to the average of such tests may be preferred instead of the repeatability r or the reproducibility R as defined in 3.1. Critical differences valid under such modified conditions can all be derived from the values of r and R as defined in 3.1. The requisite formulae and conversion factors are given in section four.

3.3 r and R may be applied in a variety of ways. They can serve :

- to verify that the experimental technique of a laboratory is up to standard;
- to compare tests performed on a sample from a batch of material with a specification;
- *to compare test results obtained by a supplier and a* consumer on the same batch of material, etc.

Some of these various uses of repeatability and reproducibility are also discussed in section four.

3.4 Precision is a general term for the closeness of agreement between replicate test results. Thus, the repeatability r and the reproducibility R describe the precision of a given test method under two different circumstances of replication. A series of inter-laboratory trials organized with the specific purpose of *determining r and R will therefore be referred to in this Inter*national Standard as a precision experiment.

3.5 As a consequence of the unavoidable random errors in the test results, the values of r and R derived from a precision experiment are estimated values. The method recommended in this International Standard has, however, been found to yield values sufficiently precise to satisfy practical requirements, provided that the laboratories employing the method for normal purposes are similar to those that participated in the precision experiment. The precision estimates should be re-estimated if at some future date evidence is available that the laboratories which participated in the original precision experiment were not representative of those currently using the test method.

4 Practical implications of the definitions

4.1 Standard test method

4.1.1 The definition of reproducibility in 3.1 refers to a standard test method and, as stated in clause 2, it is for these methods that this International Standard is primarily intended. This means that there must be a standard : that is a written document that lays down in full detail how the test should be carried out, including how the test specimen should be obtained and prepared. That standard must be applied in all the tests forming part of a precision experiment. The values of r and R derived from such an experiment should always be quoted as valid only for tests carried out according to that standard.

4.1.2 The existence of a standard implies the existence of a standardizing authority (such as ISO), within which there is a standards panel or working group responsible for the establishment of the standard under consideration.

4.1.3 In this International Standard, an essential distinction is made between standardization experiments carried out by the standards panel in order to establish the standard, and a precision experiment organized in order to determine the repeatability and reproducibility once the standard has been established.

The standardization experiments may provide information on the value of the repeatability and reproducibility but this information will not be used in the final determination of precision. It is assumed that r and R have to be estimated exclusively from the data resulting from a precision experiment specially organized for this purpose.

It is further assumed that the planning and organization of a precision experiment is a separate task to be entrusted to a precision panel. There is no reason why this should not be the same as the standards panel.

4.1.4 A precision experiment usually requires the co-operation of a larger number of laboratories and the collection of a larger number of test results than is needed in a standardization experiment. Hence, the standard test method is tried out on a larger scale than before and a precision experiment must also be considered as a final test concerning the adequacy of the standard. In particular, pronounced differences between the results reported by different laboratories may indicate that the standard is not yet sufficiently detailed and can possibly be improved. If so, this should be reported to the standards panel with a request for further investigation. [See 9.6, 17.2 b) and c), and 17.3.]

4.2 Identical material

4.2.1 According to the definitions of 3.1, tests to determine the repeatability and the reproducibility must be made on identical material. In most cases, the material on which a test is performed is either destroyed or undergoes a change. In reality, identical material therefore means that the tests are performed on samples taken from a homogeneous batch of material. The degree of homogeneity of the batch from which these samples are taken is then of great importance.

4.2.2 A fluid or a fine powder can be satisfactorily homogenized by stirring. If the material to be tested consists of a mixture of powders of different relative density or of different grain size, some care is needed because segregation may result from shaking, for instance during transport.

4.2.3 When the tests have to be performed on specimens of solid materials which cannot be homogenized — such as metals, rubber or textile fabrics — and when the tests cannot be repeated on the same test piece, then the variability among test pieces due to the heterogeneity of the material will be inseparable from the error variability of the test equipment, and will form an inherent part of both the repeatability and the reproducibility.

4.2.4 In practice, r and R are often used in order to compare batches of commercial material with a specification, or to make a comparison between two batches of material. It is then essential that any heterogeneity in such batches of commercial material be incorporated in the values of r and R. Whether or not this is the case will depend on the way the samples used in the precision experiment are prepared. The point should be carefully considered in planning these experiments.

4.2.5 When the tests have to be performed on discrete objects which are not altered by testing, the tests could, in principle at least, be carried out using the same set of objects in different laboratories. This, however, would necessitate circulating the same set of objects around many laboratories often situated far apart, in different countries or continents, with a considerable risk of loss or damage during transport.

4.2.6 In some circumstances, many of the details of this International Standard may need to be modified, but in a large proportion of cases, it should be possible for the essentials of this International Standard to be complied with.

4.2.7 Special precautions should be taken where samples are unstable. (See 9.3.)

4.3 Short intervals of time

According to the definition of 3.1, tests for the determination of repeatability have to be made under constant operating conditions. This must be interpreted as meaning that, during the time the tests are made, such factors as listed in 0.2 can be kept constant. In practice, tests under repeatability conditions should be conducted in as short a time as possible in order to minimize changes in these factors, which, particularly in the case of 0.2 d), cannot always be guaranteed constant. (See 10.4.1.)

5 Statistical model

5.1 Definition

For estimating the precision of a test method, it is useful to assume that every single test result y is the sum of three components :

$$y = m + B + e \qquad \ldots (1)$$

where, for the particular material tested, m is the average, B is a term representing the deviation of the laboratories from m and e is a random error occurring in every test.

Other models are sometimes used but it is considered that equation (1) will cover the majority of practical cases. (See 5.6.)

5.2 Average, m

5.2.1 The average m of the material tested will be called the level of the test property; different materials (for example different compositions of concrete) will correspond to different levels.

5.2.2 In some situations, the concept of a true value μ of the test property may hold good, for example the true concentration of a solution that is being titrated. The level m is, however, not necessarily equal to the true value; a difference $(m - \mu)$, when it exists, is called the bias of the test method.

When r or R is used to test the difference between two test results, a bias will have no influence and can be ignored. But when r or R are used to compare test results with a value specified in a contract or in a standard, a bias will have to be taken into account if the specification refers to the true value μ and not to the level m. If a true value exists and is known, the analysis of a precision experiment may indicate that there is a bias. (See 19.2.5.)

5.2.3 In many technical situations, however, the level of the test property is exclusively defined by the test method and the notion of an independent true value does not apply.

5.3 Term B in the model (5.1)

5.3.1 This term is considered to be constant during any series of tests performed under repeatability conditions, but to behave as a random variable in a series of tests performed under reproducibility conditions. The distribution of this variable is assumed to be approximately normal but, in practice, it is sufficient that it is unimodal. Its variance will be denoted by

$$\text{var}(B) = \sigma_L^2$$

and called the between-laboratory variance.

σ_L^2 includes the between-operator and between-equipment variabilities.

5.3.2 In general, B can be considered as the sum

$$B = B_0 + B_s$$

of a random component B_0 and a systematic component B_s.

5.3.3 In a single laboratory, such factors as listed in 0.2 cannot be kept completely constant in the long run. Hence, within laboratories, long-term variabilities will exist larger than those accounted for by the repeatability. These long term variations will contribute a random component B_0 to B.

5.3.4 In addition, there may exist permanent systematic differences between laboratories. Serious systematic differences may result from misreading of the standard for the test method or from the use of inadequate equipment. They should be investigated and corrected, and are not considered as included in the term B. Unavoidably, however, some systematic differences will remain between different laboratories. These may be due to the use of different measuring instruments or working in different climatic conditions, but even without such gross differences, variation can arise from operator technique and also from one instrument to another of the same make due to manufacturing variations. These will all contribute a systematic component B_s to B.

5.3.5 If there are in all N laboratories likely to use the method at any time, B_s will take only N discrete values and the term B in the model (5.1) can only be considered as a random variable if either the systematic differences B_s are so small that they can be ignored, or else if the test results from which the reproducibility criterion is obtained were carried out by laboratories that can be considered as selected at random from all the laboratories likely to use the method.

5.3.6 Therefore, some caution is needed when the test results to be compared are always performed by the same two laboratories. The example on the determination of the softening point of pitch given in clause 22 provides an illustration of this in that the results from laboratory 11 are consistently lower, by about 4 °C, than those from laboratory 1.

5.4 Error term e in the model (5.1)

5.4.1 This term represents a random error occurring in every single test result. The distribution of this variable is assumed to be approximately normal but, in practice, it is sufficient that the distribution is unimodal. Within a single laboratory, its variance

$$\text{var}(e) = \sigma_w^2$$

is called the within-laboratory variance.

5.4.2 It may be expected that σ_w^2 will vary between laboratories due to differences in the skills of the operators or in the quality of the equipment used. This International Standard assumes, however, that when a test method has been properly standardized, the differences between laboratories should be small so that it is justifiable to establish a common value for the within-laboratory variance valid for all laboratories using the standard method.

5.4.3 This common value, which is an average of the variances taken over the laboratories participating in the precision experiment, will be called the repeatability variance and be designated by

$$\overline{\text{var}}(e) = \sigma_r^2$$

5.5 Relation between the model (5.1), r and R

When the model (5.1) is adopted, the repeatability r and the reproducibility R are given by

$$r = f\sqrt{2}\,\sigma_r \qquad \ldots (2)$$

$$R = f\sqrt{2}\sqrt{\sigma_L^2 + \sigma_r^2} = f\sqrt{2}\,\sigma_R \qquad \ldots (3)$$

where $(\sigma_R^2 = \sigma_L^2 + \sigma_r^2)$ is called the reproducibility variance.

The coefficient $\sqrt{2}$ is derived from the fact that r and R refer to the difference between two single test results, and f is a factor whose value depends both on the number of test results available for estimating the variances σ^2 and σ_R^2, and on the shape of the distributions of the random components B and e in the model. However, if these distributions are approximately normal (in practice unimodal), the number of test results is not too small, and if the probability level is 95 %, the factor f will never differ much from the value 2 and the use of this value throughout is therefore recommended in this International Standard. (Taking into account variations in f would lead to considerable complications that would not effectively contribute to the practical value of r and R.)

Hence :

$$r = 2{,}83\,\sigma_r$$

$$R = 2{,}83\,\sigma_R$$

As the values of repeatability variance (σ_r^2) and the reproducibility variance (σ_R^2) are not known, their estimates s_r^2 and s_R^2 are used instead.

5.6 Suitability of the model (5.1)

It is clear that the model presented in 5.1 is an approximation that, by extensive experience, is known to satisfy practical requirements as a working hypothesis for designing the experiments and analysing the data. The point of view adopted in this International Standard is that the model is an acceptable approximation so long as the experimental requirements of section two are heeded and the statistical tests of section three do not yield significant results and thereby indicate its unsuitability. What action should be taken when these statistical tests indicate that the model is unsuitable will be discussed together with these tests.

6 Design of a precision experiment

6.1 One layout is as follows : samples from q batches of material, representing q different levels of the test property, are sent to p different laboratories, which are instructed to perform n tests under repeatability conditions at each level. These n tests are thus made on identical material and this type of experiment will be called a uniform-level experiment.

6.2 An alternative, preferred in certain cases (see 10.4.2), is the split-level experiment; each level is split into two sub-levels A and B, which are only slightly different. Each laboratory receives one sample from each of these sub-levels for testing.

6.3 These layouts are fully exemplified by the case studies in section five and will not be discussed here. Practical considerations in planning and execution are deferred to section two.

7 Analysis of the data

7.1 The analysis of the data produced by a precision experiment must be considered as a statistical problem to be entrusted to a statistical expert. (See 8.2 and 9.2.)

7.2 Three successive stages can be recognized, namely

a) a critical examination of the data in order to identify and treat outliers or other irregularities, and contingently to test the suitability of the model;

b) computation of preliminary values of r and R for each level separately;

c) establishment of final values of r and R including the establishment of a relation between r, R and m when the analysis indicates that they depend on the level m. If r and/or R are judged to be independent of m, the final values taken are the simple average over the levels.

7.2.1 As detailed in sub-clauses 14.8 to 14.11, the analysis of a precision experiment recommended in this International Standard first computes for each level estimates, s_r^2 and s_L^2, of the repeatability variance and the between-laboratory variance, as defined in 5.4.3 and 5.3.1 respectively, and from these derives the values of the repeatability r and the reproducibility R.

7.3 The analysis, especially stage a) described in 7.2, includes a systematic application of statistical tests. A great variety of statistical tests that might be used for the purpose of this International Standard is available from the literature.

In order to standardize the statistical analysis as far as possible, a judicious choice had to be made, and only a limited number of statistical tests, as explained in section three, has been incorporated in this International Standard.

Section two : Organization of an inter-laboratory precision experiment

General remark

The methods of operation within different organizations are not expected to be identical. Therefore, the contents of this section are only intended as a guide to be appropriately modified to cater for a particular situation.

8 Personnel requirements

8.1 Panel

The actual planning of the experiment should be the task of a panel of experts familiar with the test method and its application.

8.2 Statistical expert

At least one member of the panel should have experience in the statistical design and analysis of experiments.

8.3 Executive officer

The actual organization of the experiment should be entrusted to a single laboratory, and a member of the staff of that laboratory shall take full responsibility. He will be the executive officer.

8.4 Supervisors

A staff member in each of the participating laboratories should be made responsible for organizing the actual performance of the tests in keeping with instructions received from the executive officer, and for reporting the test results.

8.5 Operators

In each laboratory, the tests shall be carried out by one operator selected as representative of those likely to perform the tests in normal operations. He should be instructed by the supervisor as to the dates on which, and the order in which, the tests have to be carried out, but the instructions should not amplify the test method itself.

9 Tasks and problems

9.1 The following questions should be discussed by the panel :

a) Is a satisfactory standard for the test method available?

b) What is the range of levels encountered in practice?

c) How many levels should be used in the experiment? (See 10.1.)

d) What are suitable materials to represent these levels?

e) Should the material be specially homogenized before preparing the samples or should the heterogeneity in the material be included in the values of r and R? (See 10.3.)

f) What number n of replicates should be specified and what amount of material should be sent to the laboratories? (See 10.1.)

g) Should each laboratory be sent n separate samples for each level or one sample for n replicate tests? (See 10.3.) Or is a split-level experiment desirable? (See 10.4.2.)

h) Should the laboratories be sent additional material for practical exercises before the official tests are performed? (See 10.5.4.)

j) How many laboratories should be recruited to co-operate in the experiment? (See 10.1.)

k) How should the laboratories be recruited and what requirements should they satisfy? (See 10.2.)

m) Which are the details concerning the test method when the application is difficult? What kind of precisions are given to minimize these difficulties?

n) What instructions should be issued to the supervisors concerning the execution of the tests, and to how many significant figures should the test results be reported? (See 10.4.1 and 10.5.3.)

p) What information should be requested in addition to the numerical test results? (See 10.6.)

9.2 The task of the statistical expert is to contribute his specialized knowledge in designing the experiment, to analyse the data and to write a report for submission to the panel following the instructions contained in section three.

9.3 The task of the executive officer is to organize the experiment as planned by the panel, and in particular

a) to enlist the co-operation of the requisite number of laboratories and see to it that supervisors are appointed;

b) to organize and supervise the preparation of the materials and samples, and the despatch of the samples. For each level, a certain quantity of material should be set aside as a reserve stock;

c) to give special instructions when samples are unstable;

d) to draft instructions (including the interval of time between consecutive determinations) and circulate them to the supervisors early enough in advance for them to raise comments or queries;

e) to design suitable forms for the operator to use as a working record and for reporting the test results;

f) to collect the test results and prepare a table suitable for the statistical analysis.

9.4 The task of the supervisor is

a) to hand out the samples to the operators in keeping with the instructions of the executive officer;

b) to supervise the execution of the tests. The supervisor should not take part in performing the test;

c) to collect the test results, with any anomalies or difficulties experienced, and to report them to the executive officer.

9.5 The task of the operators is to perform the tests according to the standard test method and to report any anomalies and difficulties experienced (see 10.5.2 and 10.5.5).

9.6 The final task of the panel is to discuss the report by the statistical expert, establish final values for the repeatability and reproducibility, and decide if further actions are required for improving the standard for the test method or with regard to laboratories that have been rejected as outliers (see 11.6.4).

9.7 As 9.2 and 9.6 are considered to be the final stages of the statistical analysis, further details will be given in section three.

10 Comments on clauses 8 and 9

10.1 Number of laboratories and levels

No hard and fast rules can be laid down. The number of levels in a precision experiment should be chosen in relation to the range of levels to be covered, bearing in mind the cost of performing tests.

If the range of levels is very wide, r and R can be expected to depend on the level m and the use of at least 6 levels seems desirable in order to establish the relation between these quantities in a satisfactory manner.

For the example on the determination of the softening point of pitch given in clause 22 with a range of levels from 88 to 102 °C, the use of 4 levels may be considered as more than is strictly needed.

The number of laboratories should to some extent depend on the number of levels. It is recommended that the number of laboratories should never be less than 8, and if only a single level is of interest, the number of laboratories should preferably be higher, say 15 or more.

Regarding the value of n, the recommended figure is 2 except where it is customary to make a larger number of replicates, for example with certain simple physical tests.

10.2 Recruitment of participating laboratories

10.2.1 From a statistical point of view, the laboratories participating in a precision experiment should be chosen at random out of all laboratories likely to use the test method under investigation. Volunteers may not represent a realistic cross-section of laboratories.

However, in practice, other considerations may intervene; for example, the requirement that the participating laboratories should be evenly distributed over different continents or climatic regions.

The panel should decide the recruitment policy and the requirements for the participating laboratories.

10.2.2 In enlisting the co-operation of the requisite number of laboratories, their responsibility should be clearly stated. An example of a questionnaire that may be used for this purpose follows.

Questionnaire on inter-laboratory study

Title of method (copy attached) : .....
.....

1 Our laboratory wishes to participate in the co-operative testing of this method for precision data.

YES ☐ NO ☐

2 As a participant, we understand that

a) all essential apparatus, chemicals, and other requirements specified in the method must be available in our laboratory when the program begins;

b) specified "timing" requirements (such as starting date, order of testing specimens, and finishing date) of the program must be rigidly met;

c) the method must be strictly adhered to;

d) samples must be handled in accordance with instructions;

e) a qualified operator must perform the tests.

Having studied the method and having made a fair appraisal of our capabilities and facilities, we feel that we will be adequately prepared for co-operative testing of this method.

3 Comments : .....
.....

Signature : .....

Company or laboratory : .....

10.3 Heterogeneity of material

When the material to be tested is not homogeneous, it is important to prepare the samples in the manner prescribed by the method, preferably starting with one batch of commercial material for each level. Some modifications may be necessary to ensure that the amount of material available is sufficient to cover the experiment and keep a certain stock in reserve. For the samples at each level, n separate containers should be used where there is any danger of the material deteriorating when the container has once been opened, for example hygroscopic material, oxidation or loss of volatile components. In the case of unstable materials, special instructions on storage and treatment should be prescribed.

In general, it is recommended that the material used in a precision experiment and the range of materials to which r and R therefore apply be clearly specified.

10.4 Actual organization of the tests

10.4.1 With q levels and n replicates, each participating laboratory has to carry out qn tests. The performance of these tests should be organized and the operators instructed as follows :

a) All qn tests should be performed by one and the same operator using the same equipment throughout.

b) Each group of n tests belonging to one level must be carried out under repeatability conditions, that is, in a short interval of time and the same operator.

c) If, in the course of the tests, the operator should drop out through illness or some other unforeseen circumstances, another operator may complete the tests, but this must be reported with the test results.

d) It is not necessary that all qn tests be performed strictly within a short interval; the q groups of n tests may be carried out on different days.

e) It is essential that a group of n tests under repeatability conditions be performed independently as if they were n tests on different materials. As a rule, however, the operator will know that he is testing identical material. If it is feared that this knowledge may influence his test results, and consequently the repeatability variance, then a split-level experiment (see 10.4.2) shall be considered as the best procedure. Randomization of qn tests could be considered if it would not affect repeatability conditions.

10.4.2 An alternative method sometimes adopted when $n = 2$ is that of using split-levels. Instead of testing two samples which the operator has been told should be identical or of performing two tests on the same specimen of material, two series of p samples are prepared at slightly different levels m_A and m_B, where $m_A - m_B$ is a small quantity, and each of the p laboratories receives one sample of series A and one of series B for testing. Adoption of this method may be considered when it is feared that the operator, when using identical samples in carrying out his second test, may be influenced by the result of his first test.

The split-level experiment requires a slight modification in the statistical analysis which will be discussed in section three. Also, it should be clearly distinguished which test result belongs to series A and which to series B; they cannot be interchanged as can tests on identical material.

The values of r and R derived from a split-level experiment will be taken to be valid for the mean level $m = (m_A + m_B)/2$.

10.4.3 It may be necessary to limit the time that should be allowed to elapse between the day the samples are received and the day the tests are performed.

10.4.4 Any preliminary checking of equipment should be as laid down in the standard method.

10.4.5 All samples should be clearly labelled with the name of the experiment and a sample identification.

10.5 Instructions to operators

10.5.1 Before performing the tests, the operators should receive no instructions other than those contained in the standard test method; these should suffice.

10.5.2 The operators should, however, be asked to comment on this standard and in particular to state whether the instructions contained in it are sufficiently unambiguous and clear. Ambiguities may, for example, creep in when the standard has been translated into a number of different languages.

10.5.3 It is desirable that all participating laboratories report their test results to the same number of significant figures and the supervisors should be instructed accordingly. In commercial practice, the test results are often rounded rather crudely and in a precision experiment, it may be advisable to use one more significant figure than is customary or prescribed in the Standard.

When r depends on the level m, different rules for rounding may be needed for different levels.

10.5.4 An operator will not as a rule achieve normal precision when he caries out a test for the first time or after a long interval. In that case, the operators should be instructed to carry out a few unofficial tests in order to gain experience before they start testing the official samples of the precision experiment. Whether this is needed should be decided by the panel or by the supervisors; material for such preliminary tests should be supplied by the executive officer.

10.5.5 The operators should be told to report any occasions on which they are not able to follow their instructions or on which they accidentally failed to keep to the instructions. They should also be told that it is better to report a mistake than to adjust the results, because one or two missing results will not spoil the experiment and may indicate a deficiency in the standard.