TECHNICAL SPECIFICATION

ISO/TS 17503

First edition
2015-11-01

Statistical methods of uncertainty evaluation — Guidance on evaluation of uncertainty using two-factor crossed designs

Méthodes statistiques d'évaluation de l'incertitude — Lignes directrices pour l'évaluation de l'incertitude des modèles à deux facteurs croisés

Reference number
ISO/TS 17503:2015(E)

International Organization for Standardization

Contents

Page

Foreword

ISO (the International Organization for Standardization) is a worldwide federation of national standards bodies (ISO member bodies). The work of preparing International Standards is normally carried out through ISO technical committees. Each member body interested in a subject for which a technical committee has been established has the right to be represented on that committee. International organizations, governmental and non-governmental, in liaison with ISO, also take part in the work. ISO collaborates closely with the International Electrotechnical Commission (IEC) on all matters of electrotechnical standardization.

The procedures used to develop this document and those intended for its further maintenance are described in the ISO/IEC Directives, Part 1. In particular the different approval criteria needed for the different types of ISO documents should be noted. This document was drafted in accordance with the editorial rules of the ISO/IEC Directives, Part 2 (see www.iso.org/directives).

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

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

For an explanation on the meaning of ISO specific terms and expressions related to conformity assessment, as well as information about ISO's adherence to the WTO principles in the Technical Barriers to Trade (TBT) see the following URL: Foreword - Supplementary information

The committee responsible for this document is ISO/TC 69, *Applications of statistical methods*, Subcommittee SC 6, *Measurement methods and results*.

Introduction

Uncertainty estimation usually requires the estimation and subsequent combination of uncertainties arising from random variation. Such random variation may arise within a particular experiment under repeatability conditions, or over a wider range of conditions. Variation under repeatability conditions is usually characterized as repeatability standard deviation or coefficient of variation; precision under wider changes in conditions is generally termed intermediate precision or reproducibility.

The most common experimental design for estimating the long- and short-term components of variance is the classical balanced nested design of the kind used by ISO 5725-2. In this design, a (constant) number of observations are collected under repeatability conditions for each level of some other factor. Where this additional factor is 'Laboratory', the experiment is a balanced inter-laboratory study, and can be analysed to yield estimates of within-laboratory variance, σ_{r}^2, the between-laboratory component of variance, σ_{L}^2, and hence the reproducibility variance, $\sigma_{\mathrm{R}}^2 = \sigma_{\mathrm{L}}^2 + \sigma_{\mathrm{r}}^2$. Estimation of uncertainties based on such a study is considered by ISO 21748. Where the additional grouping factor is another condition of measurement, however, the between-group term can usefully be taken as the uncertainty contribution arising from random variation in that factor. For example, if several different extracts are prepared from a homogeneous material and each is measured several times, analysis of variance can provide an estimate of the effect of variations in the extraction process. Further elaboration is also possible by adding successive levels of grouping. For example, in an inter-laboratory study the repeatability variance, between-day variance and between-laboratory variance can be estimated in a single experiment by requiring each laboratory to undertake an equal number of replicated measurements on each of two days.

While nested designs are among the most common designs for estimation of random variation, they are not the only useful class of design. Consider, for example, an experiment intended to characterize a reference material, conducted by measuring three separate units of the material in three separate instrument runs, with (say) two observations per unit per run. In this experiment, unit and run are said to be 'crossed'; all units are measured in all runs. This design is often used to investigate variation in 'fixed' effects, by testing for changes which are larger than expected from the within-group or 'residual' term. This particular experiment, for example, could easily test whether there is evidence of significant differences between units or between runs. However, the units are likely to have been selected randomly from a much larger (if ostensibly homogeneous) batch, and the run effects are also most appropriately treated as random. If the mean of all the observations is taken as the estimate of the reference material value, it becomes necessary to consider the uncertainties arising from both run-to-run and unit-to-unit variation. This can be done in much the same way as for the nested designs described previously, by extracting the variances of interest using two-way analysis of variance. In the statistical literature, this is generally described as the use of a random-effects or (if one factor is a fixed effect) mixed-effects model.

Variance component extraction can be achieved by several methods. For balanced designs, equating expected mean squares from classical analysis of variance is straightforward. Restricted (sometimes also called residual) maximum likelihood estimation (REML) is also widely recommended for estimation of variance components, and is applicable to both balanced and unbalanced designs. This Technical Specification describes the classical ANOVA calculations in detail and permits the use of REML.

Note that random effects rarely include all of the uncertainties affecting a particular measurement result. If using the mean from a crossed design as a measurement result, it is generally necessary to consider uncertainties arising from possible systematic effects, including between-laboratory effects, as well as the random variation visible within the experiment, and these other effects can be considerably larger than the variation visible within a single experiment.

This present Technical Specification describes the estimation and use of uncertainty contributions using factorial designs.

Statistical methods of uncertainty evaluation — Guidance on evaluation of uncertainty using two-factor crossed designs

1 Scope

This Technical Specification describes the estimation of uncertainties on the mean value in experiments conducted as crossed designs, and the use of variances extracted from such experiments and applied to the results of other measurements (for example, single observations).

This Technical Specification covers balanced two-factor designs with any number of levels. The basic designs covered include the two-way design without replication and the two-way design with replication, with one or both factors considered as random. Calculations of variance components from ANOVA tables and their use in uncertainty estimation are given. In addition, brief guidance is given on the use of restricted maximum likelihood estimates from software, and on the treatment of experiments with small numbers of missing data points.

Methods for review of the data for outliers and approximate normality are provided.

The use of data obtained from the treatment of relative observations (for example, apparent recovery in analytical chemistry) is included.

2 Normative references

The following documents, in whole or in part, are normatively referenced in this document and are indispensable for its application. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

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

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

3 Terms and definitions

For the purposes of this document, the terms and definitions in ISO 3534-1, ISO 3534-3 and the following apply.

3.1
factor
predictor variable that is varied with the intent of assessing its effect on the response variable

Note 1 to entry: A factor may provide an assignable cause for the outcome of an experiment.

Note 2 to entry: The use of factor here is more specific than its generic use as a synonym for predictor variable.

Note 3 to entry: A factor may be associated with the creation of blocks.

[SOURCE: ISO 3534-3:2013, 3.1.5, modified — cross-references within ISO 3534-3 omitted from Notes to entry]

3.2
level
potential setting, value or assignment of a factor

Note 1 to entry: A synonym is the value of a predictor variable.

Note 2 to entry: The term "level" is normally associated with a quantitative characteristic. However, it also serves as the term describing the version or setting of qualitative characteristics.

Note 3 to entry: Responses observed at the various levels of a factor provide information for determining the effect of the factor within the range of levels of the experiment. Extrapolation beyond the range of these levels is usually inappropriate without a firm basis for assuming model relationships. Interpolation within the range may depend on the number of levels and the spacing of these levels. It is usually reasonable to interpolate, although it is possible to have discontinuous or multi-modal relationships that cause abrupt changes within the range of the experiment. The levels may be limited to certain selected fixed values (whether these values are or are not known) or they may represent purely random selection over the range to be studied.

EXAMPLE The ordinal-scale levels of a catalyst may be presence and absence. Four levels of a heat treatment may be 100 °C, 120 °C, 140 °C and 160 °C. The nominal-scale variable for a laboratory can have levels A, B and C, corresponding to three facilities.

[SOURCE: ISO 3534-3:2013, 3.1.12]

3.3
fixed effects analysis of variance
analysis of variance in which the levels of each factor are pre-selected over the range of values of the factors

Note 1 to entry: With fixed levels, it is inappropriate to compute components of variance. This model is sometimes referred to as a model 1 analysis of variance.

[SOURCE: ISO 3534-3:2013, 3.3.9]

3.4
random effects analysis of variance
analysis of variance in which each level of each factor is assumed to be sampled from the population of levels of each factor

Note 1 to entry: With random levels, the primary interest is usually to obtain components of variance estimates. This model is commonly referred to as a model 2 analysis of variance.

EXAMPLE Consider a situation in which an operation processes batches of raw material. "Batch" may be considered a random factor in an experiment when a few batches are randomly selected from the population of all batches.

[SOURCE: ISO 3534-3:2013, 3.3.10]

4 Symbols

ν_{eff} Calculated effective degrees of freedom for a standard error calculated from a two-way factorial (crossed) experiment

σ_1 True between-level standard deviation for the first factor (if considered a random effect) in a two-way factorial (crossed) experiment

σ_2 True between-level standard deviation for the second factor (if considered a random effect) in a two-way factorial (crossed) experiment

σ_{I} True between-group standard deviation for the interaction term in a factorial experiment (where one or more of the factors is considered a random effect)

σ_{r} True standard deviation for the residual term in a classical analysis of variance for a two-way factorial (crossed) experiment

d_{ij} Residual corresponding to level *i* of one factor and level *j* of a second factor in a two-way factorial experiment without replication

M_1 Mean square for the first factor in a classical analysis of variance for a two-way factorial (crossed) experiment

M_2 Mean square for the second factor in a classical analysis of variance for a two-way factorial (crossed) experiment

M_I Mean square for the interaction term in a classical analysis of variance for a two-way factorial (crossed) experiment with replication

M_r Mean square for the residual term in a classical analysis of variance for a two-way factorial (crossed) experiment

M_{tot} Mean square calculated from the "Total" sum of squares in a classical analysis of variance for a two-way factorial (crossed) experiment

n The number of replicate observations at each combination of factor levels (that is, within each "cell") in a two-way factorial (crossed) experiment with replication

p The number of levels for the first factor in a two-way factorial (crossed) experiment

q The number of levels for the second factor in a factorial (crossed) experiment

x_{ij} Observation corresponding to level i of one factor and level j of a second factor in a two-way factorial experiment without replication

x_{ijk} k^{th} observation corresponding to level i of one factor and level j of a second factor in a two-way factorial experiment with replication

S_1 Sum of squares for the first factor in a classical analysis of variance for a two-way factorial (crossed) experiment

S_2 Sum of squares for the second factor in a classical analysis of variance for a two-way factorial (crossed) experiment

S_I Sum of squares for the interaction term in a classical analysis of variance for a two-way factorial (crossed) experiment with replication

S_r Sum of squares for the residual term in a classical analysis of variance for a two-way factorial (crossed) experiment

S_{tot} "Total" sum of squares in a classical analysis of variance for a two-way factorial (crossed) experiment

s Standard deviation of a set of independent observations

s_1 Estimated between-level standard deviation for the first factor (if considered a random effect) in a two-way factorial (crossed) experiment

s_2 Estimated between-level standard deviation for the second factor (if considered a random effect) in a two-way factorial (crossed) experiment

s_I Estimated between-group standard deviation for the interaction term in a factorial experiment (where one or more of the factors is considered a random effect)

s_r Estimated standard deviation for the residual term in a classical analysis of variance for a two-way factorial (crossed) experiment

$s_{\bar{\bar{x}}}$ Estimated standard error associated with the mean in a two-way factorial (crossed) experiment

u A standard uncertainty

$u_{\bar{\bar{x}}}$ Standard uncertainty, associated with random variation, for the mean in a two-way factorial (crossed) experiment

$\bar{x}_{i\bullet}$ The mean of all data for a particular level i of Factor 1 in a factorial design

$\bar{x}_{\bullet j}$ The mean for a particular level j of Factor 2 in a factorial design

$\bar{\bar{x}}$ The mean for all data in a given experiment

5 Conduct of experiments

It should be noted that as far as possible, observations should be collected in randomized order. Action should also be taken to remove confounding effects; for example, a design intended to investigate the effect of changes in test material matrix and different analyte concentrations on recovery in analytical chemistry should not run each different sample type in a single run on a different day.

6 Preliminary review of data — Overview

In general, preliminary review should rely on graphical inspection. The general principle is to form and fit the appropriate linear model (for balanced designs this is adequately done by estimating row, column and, if necessary, cell means in the two-way layout) and inspect the residuals.

Mandel's statistics, as presented in ISO 5725-2, are applicable to inspection of individual data points in two-way designs, by replacing the 'laboratory' in ISO 5725-2 by the 'cell' in a two-way design and are recommended.

Ordinary residual plots and normal probability plots are also applicable to the residuals.

Outlier tests might additionally be suggested, though they would need to be used with care; the degrees of freedom for the residuals is smaller than for the whole data set, compromising critical values. In addition, in designs for duplicate measurements, the residuals for a cell with a serious outlier typically appear as two outliers equidistant from a common mean. Residuals for the 'main effects' model as well as the model including cell means (the interaction term) may usefully be inspected separately to avoid such an effect.

7 Variance components and uncertainty estimation

7.1 General considerations for variance components and uncertainty estimation

Basic calculations are based on the two-way ANOVA tables obtained from classical ANOVA for the two-way layout. Detailed procedures are shown below. The use of software implementations of restricted maximum likelihood estimation ("REML") is permitted when normality is a realistic assumption for all random effects.

When calculating variance estimates from classical ANOVA tables negative estimates of variance can arise. In the following calculations (7.2 to 7.4), it is recommended that these estimates be set to zero. It is further recommended that terms in the initial, complete, statistical model that are associated with negative or zero estimates of variance are dropped from the model and the model recalculated when standard uncertainties and associated effective degrees of freedom are of interest.

NOTE 1 REML calculations do not return negative estimates of variance and it is then unnecessary to reduce and re-fit models unless effective degrees of freedom are of interest.

NOTE 2 Variance estimates from small data sets are highly variable from one sample to another. For example, estimated variances taken from independent samples of 10 observations drawn from a normal distribution can vary by more than a factor of two (that is, either greater or smaller) from the true variance. Variance estimates from other distributions can vary more.

7.2 Two-way layout without replication

7.2.1 Design

The experiment involves variation in two different factors (for example, test item and instrument) with a single observation per factor combination. Let p be the number of levels for the first factor of interest, and q the number of levels for the second, so that there are pq observations x_{ij}, where the subscripts denote level i of Factor 1 and level j of Factor 2.

7.2.2 Preliminary inspection

Calculate the mean $\overline{x}_{i\bullet}$ of all data for each level i of Factor 1, the mean $\overline{x}_{\bullet j}$ for each level j of Factor 2, and the mean $\overline{\overline{x}}$ for all data. Calculate the residuals d_{ij} from

$$d_{ij} = x_{ij} - \overline{x}_{i\bullet} - \overline{x}_{\bullet j} + \overline{\overline{x}} \tag{1}$$

Plot the residuals in run order and inspect for unexpected trends and outlying observations. Additionally, prepare a normal probability plot and inspect for serious departures from normality. Check and correct any aberrant values, by re-measurement if necessary. If outlying observations are found and cannot reasonably be corrected, inspect other values within the same factor levels. If values within the same level of one factor all appear discrepant (for example, if results for a particular test material appear unusually imprecise), discard all data from that factor level before estimating variances. If this affects more than one factor level, discontinue the analysis and either treat different factor levels separately or investigate the cause and repeat the experiment.

NOTE A single missing value can be removed if it is inconsistent with normal performance of the measurement, that is, it can be attributable to instrumental or other causes. Refer to 'treatment with missing values' below for further analysis.

7.2.3 Variance component estimation

Conduct an analysis of variance to obtain the ANOVA table of the form shown in Table 1.

Table 1 — ANOVA table for two-way design without replication

Factor	SS	DF	MS	Expected mean square
Factor 1	S_1	$p - 1$	$M_1 = S_1/(p - 1)$	$\sigma_r^2 + q\sigma_1^2$
Factor 2	S_2	$q - 1$	$M_2 = S_2/(q - 1)$	$\sigma_r^2 + p\sigma_2^2$
Residual	S_r	$(p - 1)(q - 1)$	$M_r = S_r/[(p - 1)(q - 1)]$	σ_r^2
Total	$S_{tot} = S_1 + S_2 + S_r$	$pq - 1$	$M_{tot} = S_{tot}/(pq - 1)$	

From the table, the variance estimates s_1^2, s_2^2 and s_r^2 for Factor 1, Factor 2 and the repeatability variance, respectively, are given by

$$s_1^2 = \frac{M_1 - M_r}{q} \text{ with } p - 1 \text{ degrees of freedom}$$

$$s_2^2 = \frac{M_2 - M_r}{p} \text{ with } q - 1 \text{ degrees of freedom}$$

$$s_r^2 = M_r$$

Where a variance component is less than zero and is of interest for uncertainty evaluation other than in the assessment of the uncertainty for the mean value from the experiment, set the estimate equal to zero.

EXAMPLE In a randomized block design used to determine a between-unit variance for a reference material, the between-unit variance is of interest for uncertainty evaluation even though the mean of the homogeneity experiment is of no importance.

7.2.4 Standard uncertainty for the mean of all observations

Where the experiment is intended to yield a mean value $\bar{\bar{x}}$ over all observations and all variance estimates are positive, the standard uncertainty arising from repeatability, *r*, and from variation in the two experimental factors F_1 and F_2 is identical to the standard error $s_{\bar{\bar{x}}}$ calculated from

$$s_{\bar{\bar{x}}} = \sqrt{\frac{s_1^2}{p} + \frac{s_2^2}{q} + \frac{s_r^2}{pq}} \qquad (2)$$

Where one or more variance estimates are negative or zero, *either* set the corresponding term in Formula (2) to zero if only the standard uncertainty in the mean is of interest *or*, if the effective degrees of freedom is also of interest, proceed as in 7.2.5.2.

7.2.5 Degrees of freedom for the standard uncertainty

7.2.5.1 All variance estimates positive

Where all variance estimates are positive:

— calculate

$$\nu_{\text{eff}} = \frac{\left(M_1 + M_2 - M_r\right)^2}{\frac{M_1^2}{p-1} + \frac{M_2^2}{q-1} + \frac{M_r^2}{(p-1)(q-1)}} \qquad (3)$$

— set the degrees of freedom ν_s for $s_{\bar{\bar{x}}}$ as

$$\nu_s = \max\left[\min\left(p-1, q-1\right), \nu_{\text{eff}}\right] \qquad (4)$$

7.2.5.2 One or more variance estimates zero or negative

Where one of the variance estimates s_1^2 or s_2^2 is zero or negative (see 7.2.3):

— remove the corresponding term from the model and recalculate as a one-way analysis of variance ("reduced model") to give a single between-group mean square M_b with degrees of freedom ν_b;

NOTE The analysis of variance will also provide a within-group mean square M_w which is not used further here).

— calculate the standard error $s_{\bar{\bar{x}}}$ from

$$s_{\bar{\bar{x}}} = \sqrt{\frac{M_b}{pq}};$$

— set the number of degrees of freedom to the degrees of freedom associated with the between-group mean square in the reduced model.

Where the variance estimates for both of the two random factors are zero or negative, treat the complete data set as pq independent observations:

— calculate the standard deviation s in the usual way;

— calculate the standard error $s_{\bar{\bar{x}}}$ from

$$s_{\bar{\bar{x}}} = \sqrt{\frac{s^2}{pq}};$$

— set the degrees of freedom for the standard error to $pq - 1$.

7.3 Two-way balanced experiment with replication (both factors random)

7.3.1 Design

The experiment involves variation in two different factors (for example, test item and measurement run) with a single observation per factor combination. Let p be the number of levels for the first factor of interest, q the number of levels for the second, and n the number of observation per factor combination, so that there are pqn observations.

7.3.2 Preliminary inspection

Calculate cell means, subtract from the data and plot the resulting residuals in run order to check for unexpected trends or outlying values. If discrepant values are found, the discrepant values should be checked and corrected if possible. If correction is not possible, and if the discrepancy can be attributed to instrumental error or other identifiable cause, remove the data point and refer to 'treatment with missing values.

Inspect a normal probability plot of the residuals to check for significant departures from normality as above.

Optionally, calculate Mandel's statistics for cells and plot as in ISO 5725-2. Check extreme cell means (Mandel's h) or extreme standard deviation (Mandel's k) and if necessary correct any aberrant data.

NOTE In experiments conducted in duplicate, individual outliers in duplicate data will usually appear as pairs of outlying values equidistant from the mean for the cell.

7.3.3 Variance component extraction

a) Conduct an analysis of variance with interactions. This will yield a table of the form shown in Table 2.

Table 2 — ANOVA table for two-way design with replication, both effects random

Factor	SS	DF	MS	Expected mean square
Factor 1	S_1	$p - 1$	$M_1 = S_1/(p - 1)$	$\sigma_r^2 + n\sigma_I^2 + qn\sigma_1^2$
Factor 2	S_2	$q - 1$	$M_2 = S_2/(q - 1)$	$\sigma_r^2 + n\sigma_I^2 + pn\sigma_2^2$
Interaction	S_I	$(p - 1)(q - 1)$	$M_I = S_I/[(p - 1)\,(q - 1)]$	$\sigma_2^2 + n\sigma_I^2$
Residual[a]	S_r	$pq(n - 1)$	$M_r = S_r/[pq(n - 1)]$	σ_r^2
Total	$S_{tot} = S_1 + S_2 + S_I + S_r$	$pqn - 1$	$M_{tot} = S_{tot}/(pqn - 1)$	

[a] The residual term in two-way analysis of variance with replication is sometimes called the ‘within-group’ term.

b) Calculate the variance estimates s_1^2, s_2^2, s_I^2 and s_r^2 for Factor 1, Factor 2, the interaction term and the repeatability variance, respectively, as follows:

$$s_1^2 = \frac{M_1 - M_I}{qn} \text{ with } p - 1 \text{ degrees of freedom}$$

$$s_2^2 = \frac{M_2 - M_I}{pn} \text{ with } q - 1 \text{ degrees of freedom}$$

$$s_I^2 = \frac{M_I - M_r}{n} \text{ with } (p — 1)(q - 1) \text{ degrees of freedom}$$

$$s_r^2 = M_r$$

Where a variance component is less than zero and is itself of interest for uncertainty evaluation other than determining the uncertainty associated with the mean value for the experiment, set the estimate equal to zero.

7.3.4 Standard uncertainty for the mean of all observations

Where the experiment is intended to yield a mean value $\overline{\overline{x}}$ over all observations and all variance estimates are positive, the standard uncertainty arising from repeatability, *r*, and from variation in the two experimental factors F_1 and F_2 and the interaction term *I*, is identical to the standard error $s_{\overline{\overline{x}}}$ calculated from

$$s_{\overline{\overline{x}}} = \sqrt{\frac{s_1^2}{p} + \frac{s_2^2}{q} + \frac{s_I^2}{pq} + \frac{s_r^2}{npq}} \qquad (5)$$

Where one or more variance estimates are negative or zero, either set the corresponding term in Formula (5) to zero if only the standard uncertainty in the mean is of interest or, if the effective degrees of freedom is also of interest, proceed as in 7.3.5.2.

NOTE It can be useful to calculate and inspect *F* statistics and associated p-values to determine whether particular factors are important. Where the interaction term is not significant compared to the within-group (residual) term, the individual factor effects can be estimated by two-way analysis of variance without replication, applied to the cell means, or by forming an analysis of variance table for main effects only.

7.3.5 Degrees of freedom for the standard uncertainty

7.3.5.1 All variance estimates positive

Where all variance estimates are positive:

— calculate the effective degrees of freedom, ν_{eff}, as:

$$\nu_{eff} = \frac{\left(M_1 + M_2 - M_I\right)^2}{\frac{M_1^2}{p-1} + \frac{M_2^2}{q-1} + \frac{M_I^2}{(p-1)(q-1)}} \tag{6}$$

— set the degrees of freedom ν_s for $s_{\bar{\bar{x}}}$ as:

$$\nu_s = \max\left[\min\left(p-1, q-1\right), \nu_{eff}\right] \tag{7}$$

where max(.) denotes the maximum of terms enclosed in parentheses and min(.) denotes the minimum.

7.3.5.2 Interaction variance zero or negative

If the variance estimate s_I^2 for the interaction term is negative or zero:

— recalculate the ANOVA table using a 'main effects only' model to give an analysis of variance of the form of Table 3.

Table 3 — ANOVA table for two-way design with replication, both effects random (omitting interaction)

Factor	SS	DF	MS	Expected mean square
Factor 1	S_1	$p - 1$	$M_1 = S_1/(p - 1)$	$\sigma_r^2 + qn\sigma_1^2$
Factor 2	S_2	$q - 1$	$M_2 = S_2/(q - 1)$	$\sigma_r^2 + pn\sigma_2^2$
Residual[a]	S_r'	$pqn - p - q + 1$	$M_r' = S_r'/(pqn - p - q + 1)$	σ_r^2
Total	$S'_{tot} = S_1 + S_2 + S_r'$	$pqn - 1$	$M'_{tot} = S'_{tot}/(pqn - 1)$	

NOTE This table may be constructed from Table 2 by calculating $S_r' = S_r + S_I$ and using degrees of freedom as above.

[a] The residual term in two-way analysis of variance with replication is sometimes called the 'within-group' term.

— recalculate s_1^2, s_2^2 and s_r^2 as follows:

$$s_1^2 = \frac{M_1 - M_r'}{qn}$$ with $p - 1$ degrees of freedom

$$s_2^2 = \frac{M_2 - M_r'}{pn}$$ with $q - 1$ degrees of freedom

$s_r^2 = M'_r$ with $(pqn - p - q + 1)$ degrees of freedom

If both of the variance estimates s_1^2 and s_2^2 are positive:

— recalculate $s_{\bar{\bar{x}}}$ from

$$s_{\bar{\bar{x}}} = \sqrt{\frac{s_1^2}{p} + \frac{s_2^2}{q} + \frac{s_r^2}{npq}}$$

— recalculate the effective degrees of freedom ν_{eff} as

$$\nu_{\text{eff}} = \frac{\left(M_1 + M_2 - M_r\right)^2}{\frac{M_1{}^2}{p-1} + \frac{M_2{}^2}{q-1} + \frac{M_r{}^2}{pqn - p - q + 1}}$$

— set the degrees of freedom ν_s for $s_{\bar{\bar{x}}}$ as

$$\nu_s = \max\left[\min\left(p-1, q-1\right), \nu_{\text{eff}}\right]$$

where max(.) denotes the maximum of terms enclosed in parentheses and min(.) the minimum.

If one or both of s_1^2 or s_2^2 is zero or negative, reduce the analysis further by removing the term(s) corresponding to negative variances, and proceed as in 7.2.5.2.

7.3.5.3 One factor variance estimate zero or negative

Where either s_1^2 or s_2^2 is zero or negative, remove the corresponding term from the model and reanalyse as a nested two-factor analysis of variance following the methods of ISO/TS 21749.

7.4 Two-way balanced experiment with replication (one factor fixed, one factor random)

7.4.1 Design

The experiment involves variation in two different factors (for example, test item and measurement Run) with a single observation per factor combination. One of the factors is, however, the subject of an investigation and held to be a fixed effect; that it, the levels of the factor are not selected at random from a larger population and their effect is constant over time. For the purpose of this guide, Factor 2 is taken as the fixed effect. As before, let *p* be the number of levels for the first factor of interest, *q* the number of levels for the second, and *n* the number of observation per factor combination, so that there are *pqn* observations.

NOTE Information about the fixed factor (Factor 2) is not useful in the uncertainty experiment but can still be important and should be studied elsewhere if so.

7.4.2 Preliminary inspection

Inspection should follow the same procedure as for the two-way layout with both factors random.

Table 4 — ANOVA table for two-way design with replication, one fixed effect

Factor	SS	DF	MS	Expected mean square
Factor 1 (Random)	S_1	$p - 1$	$M_1 = S_1/(p - 1)$	$\sigma_r^2 + n\sigma_I^2 + nq\sigma_1^2$
Factor 2 (Fixed)	S_2	$q - 1$	$M_2 = S_2/(q - 1)$	$\sigma_r^2 + n\sigma_I^2 + np\sigma_2^2$ [c]
Interaction	S_I	$(p - 1)(q - 1)$	$M_I = S_I/[(p - 1)(q - 1)]$	$\sigma_r^2 + n\sigma_I^2$
Residual[b]	S_r	$pq(n - 1)$	$M_r = S_r/[pq(n - 1)]$	σ_r^2
Total	$S_{tot} = S_1 + S_2 + S_I + S_r$	$pqn - 1$	$M_{tot} = S_{tot}/(pqn - 1)$	

[a] The F statistic for the fixed effect, Factor 2, is calculated by dividing by the mean square for the interaction term because the expected mean square includes random deviations associated with the random interaction with Factor 1.

[b] The residual term in two-way analysis of variance with replication is sometimes called the 'within-group' term.

[c] Strictly, the effect of Factor 2, denoted σ_2^2 in this table, is not a variance but a function of fixed deviations from the mean.

7.4.3 Variance component extraction

a) Conduct an analysis of variance 'with interactions'. This will yield a table of form shown in Table 4.

b) Calculate the variance estimates s_1^2, s_I^2 and s_r^2 for Factor 1, the interaction term and the repeatability variance, respectively, as follows:

$$s_1^2 = \frac{M_1 - M_I}{qn} \text{ with } p - 1 \text{ degrees of freedom}$$

$$s_I^2 = \frac{M_I - M_r}{n} \text{ with } (p - 1)(q - 1) \text{ degrees of freedom}$$

$$s_r^2 = M_r$$

NOTE No variance component is calculated for Factor 2 as this is taken as a fixed effect. The interaction term is taken as random because it arises from interaction between a fixed and a random effect.

7.4.4 Standard uncertainty for the mean of all observations

Where the experiment is intended to yield a mean value $\bar{\bar{x}}$ over all observations, the standard uncertainty arising from repeatability and from variation in the two experimental factors is identical to the standard error $s_{\bar{\bar{x}}}$ calculated from

$$s_{\bar{\bar{x}}} = \sqrt{\frac{s_1^2}{p} + \frac{s_I^2}{pq} + \frac{s_r^2}{npq}}$$

NOTE 1 If the fixed effect is statistically significant, it is inappropriate to estimate a single mean value for all observations. Instead, mean values for each level of the fixed effect is estimated separately.

NOTE 2 Pairwise, comparisons between mean values for different levels of the fixed effect allows the correlation introduced by the common effects of Factor 1. This is beyond the scope of this Technical Specification.

7.4.5 Degrees of freedom for the standard uncertainty

Degrees of freedom for the standard error $s_{\bar{\bar{x}}}$ and for the estimated standard deviation s_1 should be taken as $p - 1$.

8 Application to observations on a relative scale

Some experiments yield data in the form of relative deviations $d_i' = (x_i\text{-}x_{\text{ref}})/x_{\text{ref}}$ from a reference value x_{ref}, or as ratios $r_i = x_i/x_{\text{ref}}$. For example, in analytical chemistry, it is common to investigate the recovery of material added to a (usually blank) test material and to report the results as a fraction or percentage of the amount added. It is also sometimes convenient to examine the dispersion of relative results x_i/x_{ref} or $x_i\big/\bar{x}$ (where $\bar{x}$ is the mean of the observations) at a number of different values of the measurand in the expectation that the standard deviation is proportional to the value of the measurand to a good approximation, allowing performance to be described in the form of an approximately constant relative standard deviation.

The methods described in Clause 6 of this Technical Specification may be applied to relative observations.

NOTE 1 The variance components and standard deviations resulting from the use of relative observations are the variances and standard deviations of the relative values and it is not always safe to treat these as estimates of the relative standard uncertainties $u_i(y)/y$. This interpretation is strictly valid only when the uncertainty in the reference value is negligible compared to the dispersion of results or where the dispersion of results is small compared to the reference value and the dispersion can be shown to be proportional to measurand value to an adequate approximation in the range of interest. An adequate approximation for this purpose is an approximation showing deviations from exact values that are small compared to the corresponding uncertainties in estimated standard deviations (see 7.1).

NOTE 2 It might be possible to use $s(x_i\big/\bar{x})$ as an estimate of $u_i(y)/y$ where, for example, $s(x_i\big/\bar{x}) < 0{,}1$, but the resulting bias is to be checked

NOTE 3 For pooling a relative standard deviation over several levels (values of the measurand) it might be necessary to treat the value of the measurand as one of the (fixed) factors of interest. Some authorities also recommend taking logs before processing ratio data; where this is done, the resulting standard deviation of log values should be converted to standard uncertainties. For this purpose, the approximation $s(\ln(X))$ approximately $s(X)/\mathrm{E}[X]$ holds to approximately two significant digits if $s(X)/\mathrm{E}[X] < 0{,}1$; that is, a standard deviation of natural logs of the raw data are approximately equal to the relative standard deviation of the raw data.

9 Use of variance components in subsequent measurements

Variance components estimated as in Clause 7 may be used in subsequent experiments provided that the effect is considered to be of similar magnitude. For example, a variance derived from an instrument effect study may be used as the basis for a standard uncertainty, as defined in ISO/IEC Guide 98-3, for a measurement of mass on an instrument of closely similar type to those studied and for a mass similar to those studied.

Where such an experiment averages of the effect of n_F levels of a factor F, the uncertainty contribution u_F is calculated from

$$u_F = \sqrt{\frac{s_F^2}{n_F}} \tag{8}$$

where s_F is the standard deviation derived from the procedures above.

10 Alternative treatments

10.1 Restricted (or residual) maximum likelihood estimates

Variance component extraction by specialist software is permitted by this Technical Specification provided that the software returns restricted maximum likelihood ("REML") estimates of variance.

NOTE REML estimates are guaranteed to be non-negative.

10.2 Alternative methods for model reduction

The removal of terms from the analysis only when the corresponding variance estimates reach zero is intended to retain model terms as far as possible. This is motivated by two considerations:

a) Early removal of terms from a model based on significance tests is insufficiently conservative when the number of degrees of freedom is small, as insignificant findings are then likely even when the corresponding true variance is important;

b) There is good reason, based on prior knowledge, to include the relevant terms in the model.

Where the degrees of freedom are large or where a term has been included in the experiment as a precaution, the data analyst may adopt a less conservative methodology for model reduction. The alternative methodology recommended for this situation by this Technical Specification is to choose the model corresponding to the minimum value for Akaike's Information Criterion (AIC). For the case of classical analysis of variance assuming normality of errors, AIC comparison may be carried out by calculating the AIC criterion I_{AIC} for each model as

$$I_{\mathrm{AIC}} = N\ln(S_{\mathrm{r}}/N) + 2(N - \nu_{\mathrm{r}}) \tag{9}$$

where N is the total number of observations, S_{r} the residual (or within-group) sum of squares from the corresponding ANOVA table, and ν_{r} the corresponding residual degrees of freedom from the same table.

NOTE This simplified implementation of the AIC is sufficient for comparison between classical ANOVA models but differs by an additive constant (for a given data set) from the general formulation based on calculated log-likelihood.

11 Treatment with missing values

If values are missing from the compiled data table, either through measurement failure or rejection on technical grounds, variance components should be extracted using restricted maximum likelihood procedures implemented in software.

Annex A
(informative)

Examples

A.1 Example 1: Estimation of a between-unit term using a randomized block design over three runs

A.1.1 Overview

The experiment is intended to estimate the between-unit standard deviation for a candidate reference material. The between-unit standard deviation will form the basis for a subsequent estimate of the uncertainty associated with homogeneity in the final certified value. The between-unit term is used to estimate the contribution of inhomogeneity to the uncertainty in certified value for an individual unit provided to the end user of the material. The experiment was constructed as a randomized block design in which 10 units of the material were measured once each in each of three separate runs. The run order was randomized for each run. This layout corresponds to the two-way layout without replication described in 7.2.

A.1.2 Data

The data are from a homogeneity study on a candidate reference material for the fungicide malachite green in fish tissue. The experiment was a randomized block design, with one observation on each of 12 units of the material in each of three instrument runs, with observations taken in random order. Units were selected randomly from a test batch of 100. The data are listed in unit order in Table A.1.

Table A.1 — Homogeneity data for a candidate reference material

Unit	Run		
	Run 1	Run 2	Run 3
2	2,801 8	2,845 7	2,791 2
10	2,860 1	2,832 3	2,722 1
14	2,832 6	2,849 4	2,661 9
20	2,872 2	2,872 3	3,474 2
23	2,614 3	2,821 6	2,866 6
34	2,677 9	2,723 2	2,742 9
37	2,907 7	2,813 7	2,672 3
43	2,869 6	2,851 6	2,697 1
51	2,608 3	2,697 5	2,678 1
56	2,804 8	2,887 4	2,757 9
60	2,771 6	2,803 5	2,673 0
65	2,812 5	2,768 8	2,846 1
The table shows the measured malachite green content in mg kg^{-1} in reference material unit order within Runs.			

A.1.3 Review of data

The data from Table A.1 are plotted in run order in Figure A.1. The data show no strong run effects but there is a marked outlier at observation 25. Inspection of the instrument output suggested a possible instrument fault for one observation for reference material unit 20. Following 7.2.2, all observations for unit 20 were removed from the data set.

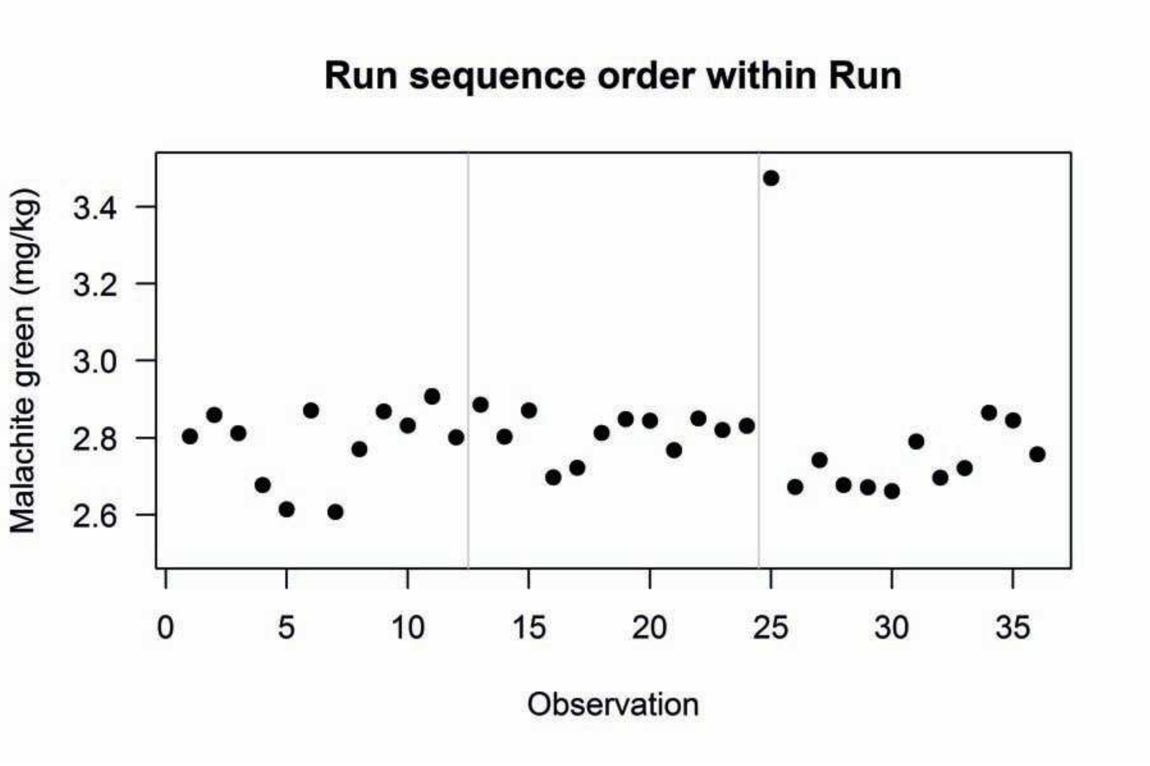

Figure A.1 — Homogeneity data for a candidate reference material (run order)

Figure A.1 shows the data from Table A.1, plotted in run sequence.

A.1.4 Variance component estimation

The analysis of variance table for the data in Table A.1 is shown as Table A.2. Following paragraph 7.2.3, the estimated variance components are

$$s_{\mathrm{Unit}}^2 = s_1^2 = \frac{0{,}007\ 21 - 0{,}005\ 77}{3} = 0{,}000\ 48$$ with 10 degrees of freedom (there were 11 units after removing unit 20)

$$s_{\mathrm{Run}}^2 = s_2^2 = \frac{0{,}014\ 13 - 0{,}005\ 77}{11} = 0{,}000\ 76$$ with 2 degrees of freedom

$$s_{\mathrm{r}}^2 = 0{,}005\ 77$$ with 20 degrees of freedom

The between-unit standard deviation is therefore 0,022 mg kg^{-1}.

NOTE The uncertainty included in the case of a certified reference material is discussed in detail in ISO Guide 35 and can be larger than the between-unit standard deviation.

Table A.2 — Analysis of variance for reference material homogeneity data

Effect	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Unit	10	0,072 1	0,007 21	1,25	0,32
Run	2	0,028 3	0,014 13	2,45	0,11
Residuals	20	0,115 4	0,005 77		

A.2 Example 2: Standard uncertainty associated with Run and Unit effects in a reference material characterization measurement

A.2.1 Overview

This Example describes the determination of a standard uncertainty associated with random variation in a two-factor crossed experiment to determine a reference value for mercury in a candidate reference material (RM).

A.2.2 Data

The experiment involved isotope dilution measurements. Three units (bottles) of the reference material were randomly selected from the production batch and measured. For each unit, duplicate measurements were made in each of three different runs. Observations were carried out in random order within each run. The results obtained are shown in Table A.3.

Table A.3 — Mercury measurements (μg kg^{-1}) on a gypsum candidate reference material

RM Unit Number	Run		
	A	B	C
77	627,247	650,980	649,989
77	632,721	655,328	638,066
87	627,170	638,822	641,432
87	613,682	634,851	643,924
127	635,729	648,628	641,972
127	638,025	657,087	651,948

A.2.3 Review of the data

The data from Table A.3 are plotted in Figure A.2, grouped by measurement run and RM Unit number. There is a strong suggestion of a difference between different runs (run A appearing consistently lower than runs B and C), and a possible difference between units. There are no severe outliers. A normal probability plot of residuals calculated as in 7.3.2 gave no reason to suspect non-normality.

A.2.4 Variance component analysis

Two-way analysis of variance with allowance for interactions leads to the ANOVA table shown as Table A.4.

Table A.4 — Analysis of variance for mercury data in a candidate reference material

Factor	SS	DF	MS
Unit	485,08	2	242,54
Run	1182,74	2	591,37
Unit:Run	155,77	4	38,94
Residuals	285,64	9	31,74

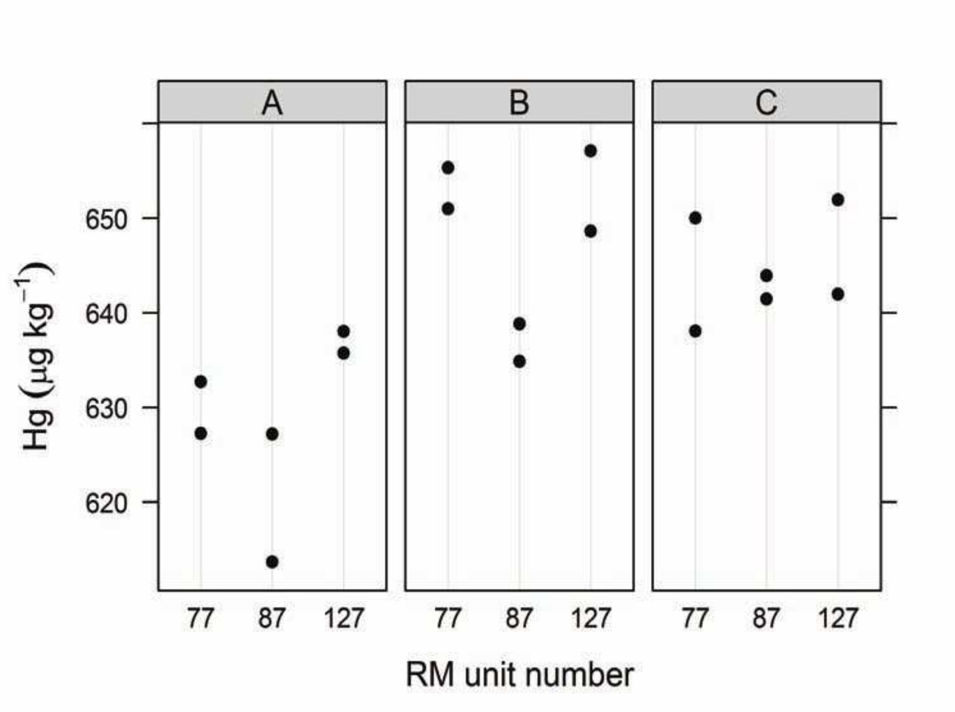

Figure A.2 — Mercury measurements on a gypsum candidate reference material

Figure A.2 shows measurements grouped by Run and RM Unit number.

Both the RM unit and the run error are random effects in the present experiment. The variance estimates s_1^2, s_2^2, s_I^2 and s_r^2 for Factor 1 (the Unit term), Factor 2 (the Run term), the interaction term and the repeatability variance, respectively, are therefore calculated from Table A.4 as follows:

$$s_1^2 = \frac{242{,}54 - 38{,}94}{3 \times 2} = 33{,}93 \text{ with } 3 - 1 = 2 \text{ degrees of freedom}$$

$$s_2^2 = \frac{591{,}37 - 38{,}94}{3 \times 2} = 92{,}07 \text{ with } 3 - 1 = 2 \text{ degrees of freedom}$$

$$s_\mathrm{I}^2 = \frac{38{,}94 - 31{,}74}{2} = 3{,}60 \text{ with } (3 - 1)(3 - 1) = 4 \text{ degrees of freedom}$$

$$s_\mathrm{r}^2 = 31{,}74$$

NOTE In this experiment, $n = 2$ and $P = q = 3$.

None of the variance components is zero or negative, so the present model is retained.

A.2.5 Standard uncertainty for the mean of all observations

The mean of the observations can be calculated from Table A.3 as 640,422 µg kg^{-1}. The standard uncertainty $u_{\bar{\bar{x}}}$ arising from variation visible in the experiment can be calculated from Formula (5), 7.3.4 as

$$u_{\bar{\bar{x}}} = s_{\bar{\bar{x}}} = \sqrt{\frac{33{,}93}{3} + \frac{92{,}07}{3} + \frac{3{,}60}{3 \times 3} + \frac{31{,}74}{18}} = 6{,}78 \text{ µg kg}^{-1}$$

NOTE Other uncertainties, including (for example) calibration uncertainties and allowances for inhomogeneity will normally be combined with the standard uncertainty calculated here to obtain the uncertainty for a certified value.

A.2.6 Degrees of freedom for the standard uncertainty

Since all variance components are positive, the effective degrees of freedom for the above standard uncertainty can be calculated from Formulae (6) and (7), 7.3.5.1. These yield

$$\nu_{\text{eff}} = \frac{\left(242,54 + 591,37 - 38,94\right)^2}{\dfrac{242,54^2}{3-1} + \dfrac{591,37^2}{3-1} + \dfrac{38,94^2}{(3-1)(3-1)}} = 3,09$$

and

$$\nu_s = \max\left[\min\left(3-1,\, 3-1\right),\, 3,09\right] = 3,09$$

The degrees of freedom associated with the standard uncertainty calculated in A.2.6 is therefore set to 3,09.

NOTE If used to calculate a confidence interval using the corresponding value of student's t from statistical tables, the calculated value of 3,09 will normally be rounded down to 3.

Bibliography

[1] 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*

[2] ISO/IEC Guide 98-3, *Uncertainty of measurement — Part 3: Guide to the expression of uncertainty in measurement (GUM:1995)*

[3] ISO/TS 21749, *Measurement uncertainty for metrological applications — Repeated measurements and nested experiments*

ICS 17.020

Price based on 19 pages