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)

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, $\sigma_R^2 = \sigma_L^2 + \sigma_r^2$. 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 $\bar{x}_{i\bullet}$ of all data for each level i of Factor 1, the mean $\bar{x}_{\bullet j}$ for each level j of Factor 2, and the mean $\bar{\bar{x}}$ for all data. Calculate the residuals d_{ij} from

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

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	$S_{tot} = S_1 + S_2 + S_r$	$pq - 1$	$M_{tot} = S_{tot}/(pq - 1)$	

