TECHNICAL REPORT

ISO/TR 22914

First edition
2020-10

Statistical methods for implementation of Six Sigma — Selected illustration of analysis of variance

Méthodes statistiques pour la mise en œuvre du Six Sigma - Exemples choisis d'application de l'analyse de la variance

Reference number
ISO/TR 22914:2020(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 of the voluntary nature of standards, the meaning of ISO specific terms and expressions related to conformity assessment, as well as information about ISO's adherence to the World Trade Organization (WTO) principles in the Technical Barriers to Trade (TBT), see www.iso.org/iso/foreword.html.

This document was prepared by Technical Committee ISO/TC 69, *Applications of statistical methods*, Subcommittee SC 7, *Applications of statistical and related techniques for the implementation of Six Sigma*.

Any feedback or questions on this document should be directed to the user's national standards body. A complete listing of these bodies can be found at www.iso.org/members.html.

Introduction

Analysis of variance (ANOVA) is a collection of statistical models used to analyse the differences among group means and their associated procedures (such as "variation" among and between groups), developed by statistician and evolutionary biologist Ronald A. Fisher. In the ANOVA setting, the observed variance in a particular variable is partitioned into components attributable to different sources of variation. In its simplest form, ANOVA provides a statistical test of whether or not the means of several groups are equal, and therefore generalizes the t-test to more than two groups. ANOVA models are useful for comparing (testing) three or more means (groups or variables) for statistical significance. It is conceptually similar to multiple two-sample t-tests, but is more conservative (it results in less type I error) and is therefore suited to a wide range of practical problems. In Six Sigma, ANOVA is used to find out if there are differences in the performances of different groups, and ultimately to find out if these differences count, or are important enough that a significant change or adjustment should be made. It serves as a guide on which aspect(s) of a process improvements can, or should, be made.

ANOVA is the synthesis of several ideas and it is used for multiple purposes. As a consequence, it is difficult to define concisely or precisely. Classical ANOVA for balanced data does the three following things at once.

1) As exploratory data analysis, an ANOVA is an organization of an additive data decomposition, and its sums of squares indicate the variance of each component of the decomposition (or, equivalently, each set of terms of a linear model).

2) Comparisons of mean squares, along with an F-test allow testing of a nested sequence of models.

3) Closely related to the ANOVA is a linear model fit with coefficient estimates and standard errors.

In short, ANOVA is a statistical tool used in several ways to develop and confirm an explanation for the observed data. Additionally:

1) it is computationally elegant and relatively robust against violations of its assumptions;

2) it provides industrial strength by (multiple sample comparison) statistical analysis;

3) it has been adapted to the analysis of a variety of experimental designs.

As a result, ANOVA has long enjoyed the status of being the most used (some would say abused) statistical technique in psychological research. "ANOVA "is probably the most useful technique in the field of statistical inference. ANOVA is difficult to teach, particularly for complex experiments, with split-plot designs being notorious.

There are three main assumptions:

1) independence of observations — this is an assumption of the model that simplifies the statistical analysis;

2) normality — the distributions of the residuals are normal;

3) equality (or "homogeneity") of variances, called homoscedasticity — the variance of data in groups is expected to be the same.

If the populations from which data to be analysed by a one-way analysis of variance (ANOVA) were sampled violate one or more of the one-way ANOVA test assumptions, the results of the analysis can be incorrect or misleading. For example, if the assumption of independence is violated, then the one-way ANOVA is simply not appropriate, although another test (perhaps a blocked one-way ANOVA) can be appropriate. If the assumption of normality is violated, or outliers are present, then the one-way ANOVA is not necessarily the most powerful test available. A nonparametric test or employing a transformation can result in a more powerful test. A potentially more damaging assumption violation occurs when the population variances are unequal, especially if the sample sizes are not approximately equal (unbalanced). Often, the effect of an assumption violation on the one-way ANOVA result depends on the extent of the violation (such as how unequal the population variances are, or how heavy-tailed one or another population distribution is). Some small violations can have little practical effect on the analysis, while other violations can render the one-way ANOVA result uselessly incorrect or uninterpretable. In particular, small or unbalanced sample sizes can increase vulnerability to assumption violation.

Statistical methods for implementation of Six Sigma — Selected illustration of analysis of variance

1 Scope

This document describes the necessary steps of the one-way and two-way analyses of variance (ANOVA) for fixed effect models in balanced design. Unbalanced design, random effects and nested design patterns are not included in this document.

This document provides examples to analyse the differences among group means by splitting the overall observed variance into different parts. Several illustrations from different fields with different emphasis suggest the procedure of the analysis of variance.

2 Normative references

The following documents are referred to in the text in such a way that some or all of their content constitutes requirements of this document. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

ISO 3534-1:2006, *Statistics — Vocabulary and symbols — Part 1: General statistical terms and terms used in probability*

ISO 3534-3:2013, *Statistics — Vocabulary and symbols — Part 3: Design of experiments*

3 Terms and definitions

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

ISO and IEC maintain terminological databases for use in standardization at the following addresses:

— ISO Online browsing platform: available at https://www.iso.org/obp

— IEC Electropedia: available at http://www.electropedia.org/

3.1
response variable
variable representing the outcome of an experiment

[SOURCE: ISO 3534-3:2013, 3.1.3, modified — the notes have been removed.]

3.2
predictor variable
variable that can contribute to the explanation of the outcome of an experiment.

[SOURCE: ISO 3534-3:2013, 3.1.4, modified — the notes have been removed.]

3.3
model
formalized representation of outcomes of an experiment

[SOURCE: ISO 3534-3:2013, 3.1.2, modified — the notes and examples have been removed.]

3.4
analysis of variance
ANOVA
technique which subdivides the total variation of a response variable into components associated with defined sources of variation

[SOURCE: ISO 3534-3:2006, 3.3.8, modified — the notes and examples have been removed.]

3.5
degree of freedom
DF
number of linearly independent effects that can be estimated

[SOURCE: ISO 3534-3:2013, 3.1.32, modified — the symbol ν has been replaced with the abbreviated term DF, and the notes have been removed.]

3.6
factor
feature under examination as a potential cause of variation

[SOURCE: ISO 3534-3:2013, 3.1.5, modified — the notes have been removed.]

3.7
fixed effects analysis of variance
analysis of variance (3.4) in which the *factor levels* (3.8) of each *factor* (3.6) are preselected over the range of values of the factors

[SOURCE: ISO 3534-3:2013, 3.3.9, modified — the note has been removed.]

3.8
factor level
setting, value or assignment of a *factor* (3.6)

[SOURCE: ISO 3534-3:2013, 3.1.12, modified — the notes and the example have been removed.]

3.9
factor effect
factor (3.6) that influences the response variable

[SOURCE: ISO 3534-3:2013, 3.1.14, modified — the note has been removed.]

3.10
main effect
factor effect (3.9) applicable in the context of linearly structured *models* (3.3) with respect to expectation

Note 1 to entry: The main effect can be estimated by averaging the response variable over all other runs provided the experiment is fully balanced.

[SOURCE: ISO 3534-3:2013, 3.1.15, modified — Notes 1 and 3 have been removed; Note 2 has been renumbered as Note 1 to entry.]

3.11
one-way analysis of variance
analysis of variance (3.4) in which a single *factor* (3.6) is investigated

3.12
two-way analysis of variance
analysis of variance (3.4) in which two distinct *factors* (3.6) are simultaneously investigated for possible effects on the response variable

3.13
balanced data
set of data in which sample sizes are kept equal for each treatment combination

3.14
F-test
statistical test in which the test statistic has an F-distribution under the null hypothesis

3.15
***p*-value**
probability of observing the observed test statistic value or any other value at least as unfavourable to the null hypothesis

[SOURCE: ISO 3534-1:2006, 1.49, modified — the example and the notes have been removed.]

3.16
crossed classification
classification according to more than one attribute at the same time

Note 1 to entry: Crossed classification can be illustrated in Figure 1.

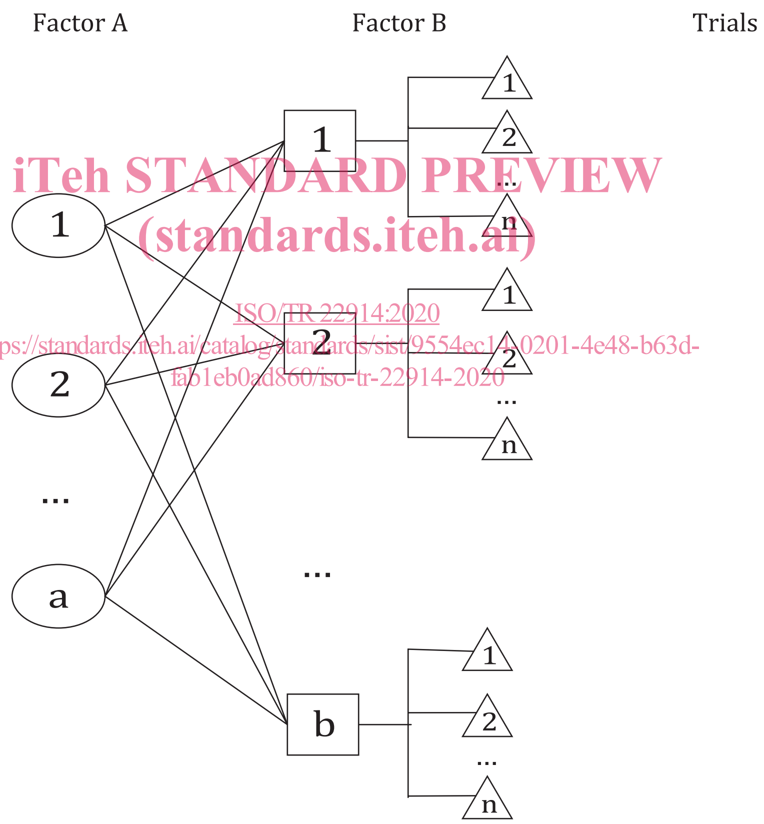

Figure 1 — Crossed classification graphic in ANOVA

3.17
interaction
influence of one *factor* (3.6) on one or more other factors' impact on the response variable

[SOURCE: ISO 3534-3:2013, 3.1.17, modified — the notes have been removed.]

3.18
replication
multiple occurrences of a given treatment combination or setting of *predictor variables* (3.2)

[SOURCE: ISO 3534-3:2013, 3.1.36, modified — the notes have been removed.]

4 Symbols and abbreviated terms

H_0	null hypothesis
H_1	alternative hypothesis
DF	degree of freedom
F	F-statistic
SS	sums of squares
MS	mean squares
Adj SS	adjusted sums of squares
Adj MS	adjusted mean squares

5 General description of one-way and two-way classifications

5.1 General

This clause provides general guidelines to conduct the one-way and two-way analysis of variances and illustrates the necessary steps. The formulae are shown in Annex F.

Five distinct applications illustrating the procedures are given in Annexes A through E. Each of these examples follows the basic structure in nine steps given in Table 1.

The (common) flowchart for one-way and two-way ANOVA is given in Figure 2.

Table 1 — General ANOVA procedure

1	Stating objectives
2	Data collection plan
3	Variables description
4	Measurement system considerations
5	Performing data collection
6	Verification of ANOVA assumptions
7	Undertaking ANOVA analysis
8	Further analysis
9	Conclusion

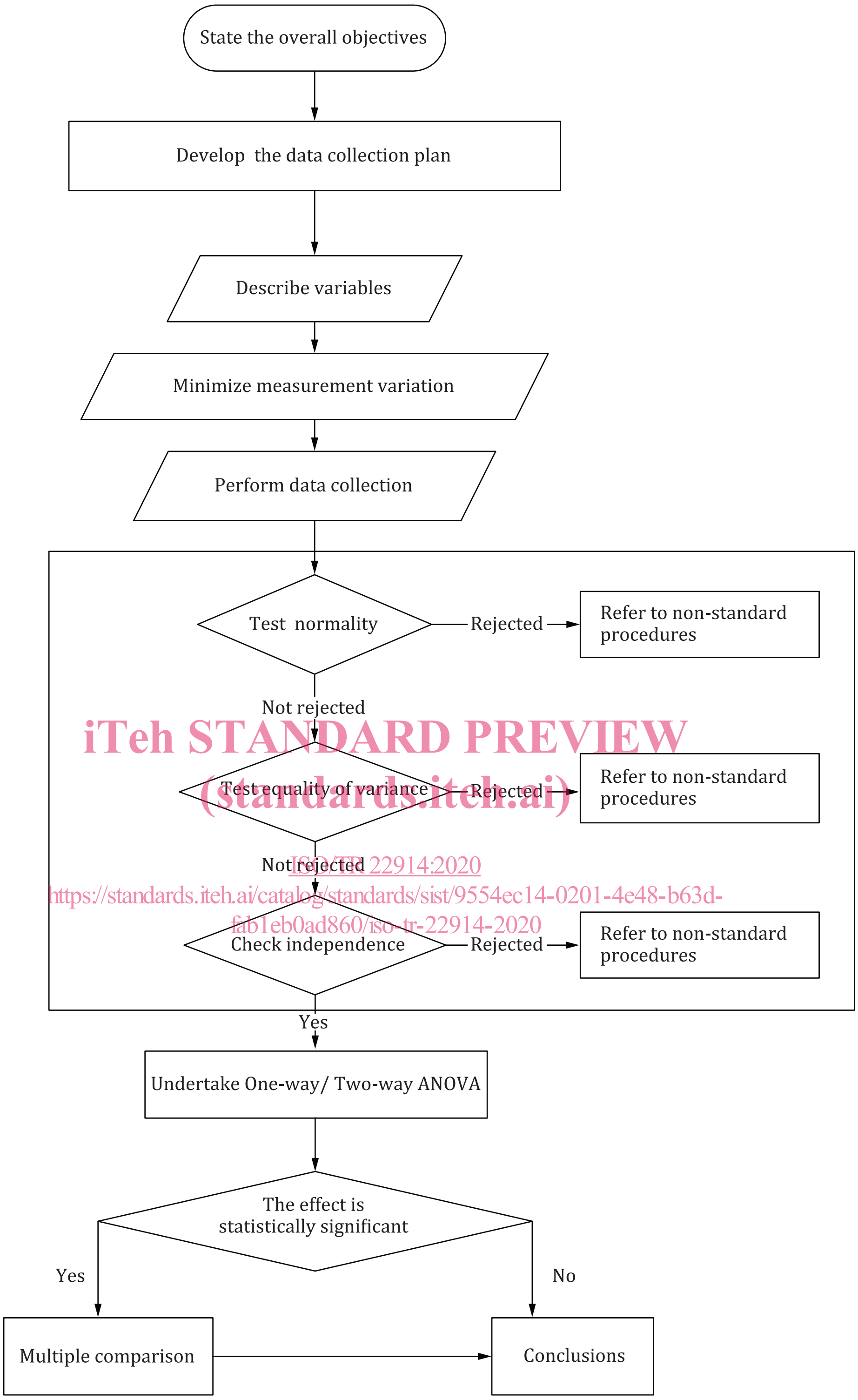

Figure 2 — Common flowchart for one-way and two-way ANOVA

5.2 Stating objectives

ANOVA is used to determine if there are differences in the mean in groups of continuous data. Analysis of variance is often used in Six Sigma projects in the 'analyse' phase of DMAIC (define, measure, analyse, improve, control) methodology. It is a statistical technique for analysing measurements depending on several kinds of effects operating simultaneously. Analysis of variance aims for deciding which kinds of factors are important and estimating the effects of them. It is likely to be one of the most common tests that will be used by a Six Sigma project.

ANOVA is conducted for a variety of reasons, which include, but are not limited to:

a) assess the need for a model to represent the data;

b) test whether a factor with several levels is effective;

c) test whether two factors have an interaction, which is only applicable for two-way ANOVA;

d) test whether there is any difference between levels of some variables.

Analysis of variance examines the influence of one or two different categorical independent variables on one continuous dependent variable. One-way ANOVA examines the equality of the means of the continuous variable for each level of a single categorical explanatory variable. The two-way ANOVA not only aims at assessing the main effect of each independent variable but also if there is any interaction between them.

The analysis of variance can be presented in terms of a linear model. The objective of ANOVA is to find the differences between the data. It provides the basis for optimizing experiment design. Additionally, in Six Sigma, ANOVA is used to find out if there are differences in the performances of different factors. It serves as a guide on which aspect(s) of a process improvement can, or should, be made.

5.3 Data collection plan

The data collection plan describes the relationship with the design of the experiment; refer to Factsheet 26 in ISO 13053-2:2011[1] for the design of the experiment. It contains the necessary steps for collecting, characterizing, categorizing, cleaning and contextualizing the data to enable its analysis.

The data collection plan also includes how to manage data quality. Data quality establishes the set of actions to be taken for ensuring the veracity of the data, such as integrity, completeness, timeliness and accuracy.

After collecting the data, it is highly recommended to check it for completeness (non-missing), errors or outliers, since these types of anomalies can distort the data.

For missing data, whether to use methods for dealing with missing data, such as imputation, or not is decided.

NOTE In statistics, imputation is the process of replacing missing data with substituted values. Once all missing values have been imputed, the data set can then be analysed using standard techniques for complete data. For more details about imputation methods, see Reference [2].

5.4 Variables description

Consists of describing the response variable and the independent factors and their relationship with the process.

5.5 Measurement system considerations

Consists of describing the measurement system analysis in place and the underlying requirements in order to minimize the measurement system variation. For details, refer to ISO 22514-6[3] or ISO/TR 12888[4].

5.6 Performing data collection

Consists of performing the data collection in accordance to the data collection plan in 5.3.

5.7 Verification of ANOVA assumptions

5.7.1 General

Analysis of variance is used to analyse the effects of factors, which can have an impact on the result of an experiment. This document focuses on fixed effects analysis of variance for data that satisfy three conditions: (1) the normality assumption; (2) the assumption of homogeneity of variances; (3) the independence of the observations.

5.7.2 Test of normality

There are two methods to test the normality of the model error: graphically and numerically, respectively relying on visual inspection and on statistics. In order to determine normality graphically, the output of normal probability plots, quantile-quantile plots (Q-Q plots) can be used. If the data are normally distributed, the Q-Q plot shows a diagonal line. If the Q-Q plot shows a line in an obvious non-linear fashion, the data are not normally distributed.

Numerically, the well-known tests of normality are the Kolmogorov-Smirnov test, the Shapiro-Wilk test, the Cramer-von Mises test and the Anderson-Darling test. They can be combined with the graphical analysis as performed in 5.8, refer to ISO 5479[5].

NOTE When the volume of data increases (which often happens nowadays), the coefficients of Pearson skewness and kurtosis can be taken into account. This is important because normality tests become powerful in this case and therefore rejects the hypothesis of normality simply for a small gap. However, the assumption of normality remains a good working hypothesis.

5.7.3 Test of homogeneity of variance

ANOVA requires that the variances of different populations be equal. This can be determined by the following approaches: comparison of graphs (box plots); comparison of variances, standard deviations.

The F-test of two sample hypothesis test of variances can be used to determine if the variances of two populations are equal.

5.7.4 Test of independence

ANOVA requires the independence of the observations. This can be determined, for example, by the following approaches:

a) it can be checked by investigating the method of data collection; a pattern that is not random suggests a lack of independence;

b) it can be evaluated by looking at the residuals against any time variables present (e.g., order of observation), any factors;

c) it can be evaluated by looking at the auto-correlation and the Durbin-Watson statistic.

NOTE Data needs be sorted in correct order for meaningful results. For example, samples collected at the same time would be ordered by time if it is suspected that results could depend on time.

5.7.5 Outliers identification

For outliers' identification and treatment, refer to ISO 16269-4:2010[6] and ISO 5725-2:1994[7].

5.7.6 How to deal with non-standard cases

In many situations, the data do not fulfil all or part of the assumptions as described in 5.7.2 to 5.7.4. In these cases, the following several options can be adopted:

— transform the data using various algorithms so that the shape of the distribution becomes normally distributed;

— choose a nonparametric test, such as the Kruskal-Wallis H Test, which does not require the assumption of normality.

5.8 Undertaking ANOVA analysis

5.8.1 State hypotheses H_0 and H_1

State H_0: the equality hypothesis among subgroups.

State H_1: the inequality hypothesis among subgroups.

NOTE The hypotheses reflect the commonalities or the lack thereof among subgroups in business terms.

5.8.2 Graphical analysis

One can perform graphical analysis, i.e. histograms, box plots, to gain a better understanding of the data. Graphical analysis are linked to the business context and the data generating process.

5.8.3 Generate analysis results

A generic table for ANOVA is described, see Table 2.

Table 2 — Analysis of variance table

Variation	Cause	Source	Type	Sums of squares	Degrees of freedom	Variance estimate
Between	Assignable cause	Factor A	1-way and 2-way			
		Factor B	2-way			
		Interaction	2-way			
Within	Common cause	Error	1-way			
			2-way			
Total						

NOTE 1 The variance estimate is also known as mean squares.

NOTE 2 For the explicit formula in every case in Table 2 refer to Annex F. All the ANOVA tables can be interpreted in the same way. They allow to split the aggregate variability inside the data into two parts: assignable and common. The analysis of variance test determines whether the influence of assignable factors is statistically significant.

5.8.4 Residual analysis

Check residuals for independence, normality and auto-correlation using graphical visualisation or by quantitative methods. For graphical visualisation, it can be checked by residual plots. A residual plot is a graph that shows the residuals on one axis and the independent variable on the other axis.

The best test for auto-correlation is to look at a residual time series plot (residuals vs row number). If the plot of the residuals versus order does not show any pattern, there is no time dependence in the residuals.

The test for homogeneity of variance is to look at a plot of residuals versus predicted values. If the residuals are randomly scattered about zero and have approximately the same scatter for all fitted values, the constant variance assumption does not appear to be violated.

The best test for normally distributed errors is a normal probability plot or normal quantile plot of the residuals. If the points on the normal probability plot roughly follow a straight line, one can assume that the residuals do not deviate substantially from a normal distribution.

5.9 Further analysis

When a statistically significant effect in ANOVA exists, further analysis can be implemented. A statistically significant effect in ANOVA is often followed up with one or more different follow-up tests. This can be done in order to assess which groups are different from other groups.

Some tests such as Tukey's range test most commonly compare every group mean with every other group mean and typically incorporate some methods of controlling for Type I errors. Simple comparisons compare one group mean with one other group mean. Compound comparisons typically compare two sets of groups' means where one set has two or more groups (e.g., compare average group means of group A, B and C with group D).

For further analysis and model development, see References [8] and [9].

NOTE Tukey's range test, also known as the Tukey's test, Tukey method, Tukey's honest significance test, or Tukey's HSD (honestly significant difference) test, is a single-step multiple comparison procedure and statistical test. It can be used on raw data or in conjunction with ANOVA to find means that are significantly different from each other.

5.10 Conclusion

Based on the results of the above analysis of variance, some conclusions of the effect of the factor on the response variable can be obtained. With these findings, formulate a conclusion statement that links to the ANOVA results to the project objectives given in 5.2.

6 Description of Annexes A through E

Five distinct examples of ANOVA are illustrated in Annexes A to E, which have been summarized in Table 3 with the different aspects indicated.

Table 3 — Example summaries, by Annex

Annex	Example	ANOVA-details
A	Bond strength	Two-way ANOVA analysis: detects the factors which have effects on the bond strength. (Germany, Minitab<?> Minitab 17, R 3.0, JMP 11 and Q-DAS v12 are examples of suitable products available commercially. This information is given for the convenience of users of this document and does not constitute an endorsement by ISO of these products. 17)
B	Effect of script and training on income per sale	Two-way ANOVA analysis: detects the effect of script and training on income per sale (UK, R 3.0[1)])
C	Strength of welded joint	Two-way ANOVA analysis in DOE: detects the factors which have effects on the strength of welded joint. (India, JMP 11[1)])

