INTERNATIONAL STANDARD

ISO 13299

First edition
2003-03-01

Sensory analysis — Methodology — General guidance for establishing a sensory profile

Analyse sensorielle — Méthodologie — Directives générales pour l'établissement d'un profil sensoriel

Reference number
ISO 13299:2003(E)

PDF disclaimer

This PDF file may contain embedded typefaces. In accordance with Adobe's licensing policy, this file may be printed or viewed but shall not be edited unless the typefaces which are embedded are licensed to and installed on the computer performing the editing. In downloading this file, parties accept therein the responsibility of not infringing Adobe's licensing policy. The ISO Central Secretariat accepts no liability in this area.

Adobe is a trademark of Adobe Systems Incorporated.

Details of the software products used to create this PDF file can be found in the General Info relative to the file; the PDF-creation parameters were optimized for printing. Every care has been taken to ensure that the file is suitable for use by ISO member bodies. In the unlikely event that a problem relating to it is found, please inform the Central Secretariat at the address given below.

ISO copyright office
Case postale 56 • CH-1211 Geneva 20
Tel. + 41 22 749 01 11
Fax + 41 22 749 09 47
E-mail copyright@iso.org
Web www.iso.org

Published in Switzerland

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.

International Standards are drafted in accordance with the rules given in the ISO/IEC Directives, Part 2.

The main task of technical committees is to prepare International Standards. Draft International Standards adopted by the technical committees are circulated to the member bodies for voting. Publication as an International Standard requires approval by at least 75 % of the member bodies casting a vote.

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.

ISO 13299 was prepared by Technical Committee ISO/TC 34, *Food products*, Subcommittee SC 12, *Sensory analysis*.

Introduction

The purpose of this International Standard is to serve as guidance on those steps that are common to all sensory profiling. Reference is given in Clause 4 to existing and planned International Standards describing a part of the process (e.g. the choice of descriptors or of scales) or describing specific types of sensory profiling (e.g. texture or flavour profiles).

A sensory profile is a descriptive analysis of a sample by a panel. The sample may be a product (e.g. a food, beverage, tobacco product, cosmetic, textile or paper). It could also be a sample of air or water being tested for pollutants. Profiling can be carried out in a number of ways. Over the years, a few of these have been formalized and codified as descriptive procedures by professional societies or by groups of producers and users for the purpose of improving communication between themselves. The purpose of this International Standard is to provide agreed guidelines for such descriptive procedures.

Sensory profiling is based on the concept that the sensory impression made by the sample consists of a number of identifiable sensory attributes (descriptors), each of which is present to a larger or smaller degree. The list of all relevant sensory descriptors, each with its intensity value, is the sensory profile. Some sensory profiles take a view across all of the senses; others (partial profiles) concentrate in detail on particular senses. Two samples may be different yet have the same partial profile. Usually the attributes are listed in the order of perception.

Three factors need particular attention when establishing a profile:

— that assessors differ in their sensitivity and thresholds by which they sense individual attributes;

— that assessors may lack awareness or cognizance of certain attributes of a sample; and

— that in most samples there exists a “complex” or “background” of attributes that are not easily identified or separated.

The impact of these factors can be greatly reduced, but not entirely eliminated, by putting more effort into the selection of descriptors, and by using larger numbers of repeat tests by larger numbers of sensitive and highly trained assessors.

Sensory analysis — Methodology — General guidance for establishing a sensory profile

1 Scope

This International Standard describes the overall process for developing a sensory profile. Sensory profiles can be established for products such as foods and beverages, and can also be useful in studies of human cognition and behaviour. Some applications of sensory profiling are as follows:

— to develop or change a product;

— to define a product, production standard or trading standard in terms of its sensory attributes;

— to study and improve shelf-life;

— to define a reference "fresh" product for shelf-life testing;

— to compare a product with a standard or with other similar products on the market or under development;

— to map a product's perceived attributes for the purpose of relating them to factors such as instrumental, chemical or physical properties, and/or to consumer acceptability;

— to characterize by type and intensity the off-odours or off-tastes in a sample of air or water (e.g. in pollution studies).

NOTE 1 Sensory profiles can also be established for non-alimentary products or samples which are evaluated by the senses of sight, odour, taste, touch or hearing.

NOTE 2 Some International Standards dealing with aspects of establishing a sensory profile are given in Clauses 2 and 4.

2 Normative references

The following referenced documents are indispensable for the application 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 4121, *Sensory analysis — Methodology — Evaluation of food products by methods using scales*

ISO 5492, *Sensory analysis — Vocabulary*

ISO 5496, *Sensory analysis — Methodology — Initiation and training of assessors in the detection and recognition of odours*

ISO 6564, *Sensory analysis — Methodology — Flavour profile methods*

ISO 6658:1985, *Sensory analysis — Methodology — General guidance*

ISO 8586-1, *Sensory analysis — General guidance for the selection, training and monitoring of assessors — Part 1: Selected assessors*

ISO 8586-2, *Sensory analysis — General guidance for the selection, training and monitoring of assessors — Part 2: Experts*

ISO 8589, *Sensory analysis — General guidance for the design of test rooms*

ISO 11035, *Sensory analysis — Identification and selection of descriptors for establishing a sensory profile by a multidimensional approach*

ISO 11036, *Sensory analysis — Methodology — Texture profile*

ISO 11056, *Sensory analysis — Methodology — Magnitude estimation method*

3 Terms and definitions

For the purposes of this document, the terms and definitions given in ISO 5492 and the following apply.

3.1
sensory profile
description of the sensory properties of a sample, comprising the sensory attributes in the order of perception, and with assignment of an intensity value for each attribute

NOTE This is a generic term for any type of profile, whether full or partial, trademarked or not.

3.2
partial sensory profile
profile comprising certain selected attributes, with their intensity values

EXAMPLES Odour profile, flavour profile and texture profile.

3.3
conventional sensory profile
profile obtained by statistical treatment of data issued from several assessors using a single list of attributes

3.4
consensus sensory profile
profile obtained by consensus after discussion by a group of assessors, each of them having assessed the product according to his/her own criteria before the discussion

3.5
free-choice sensory profile
profile in which each assessor chooses his/her own attributes to describe a group of samples and in which consensus space is derived statistically, for example by generalized Procrustes analysis

3.6
time-intensity sensory profile
profile that describes the intensity of a given attribute as it changes over a period of time, following a single application of the stimulus

4 Principles

Table 1 presents points to consider in the choice of technique.

Steps in establishing a sensory profile are common to all methods of sensory profiling and are set out in Table 2.

Table 1 — Suggested area of application of different profiling techniques

Technique	Principle	Area of application	Advantages	Disadvantages	Illustrative examples
Conventional profiling	Assessors, seated in booths, score each sample on a preselected set of attributes and scales.	The most-used technique. Suitable for routine applications and for research, e.g. in development and quality control of consumer goods. Various procedures exist for choosing descriptors, see Table 3.	Generally the most reliable technique. Profiles are reproducible within a panel and over time. If sufficient training is given and enough reference standards are used, profiles are also reproducible between panels.	Relatively expensive because fairly large panels and good booth areas are required. Panel selection and training are relatively time-consuming.	Standard textbooks, e.g. references [1] to [5] in Bibliography.
Consensus profiling	Through consensus discussion the panel, seated around a table, develops its own terminology and scores pertaining to the sample set presented.	Suitable for routine sensory evaluation of a variety of miscellaneous products such as those offered in a supermarket. Also useful for non-recurring sets of e.g. 3 to 10 similar items.	Many samples can be tested at relatively low cost in samples and assessors' time.	Profiles obtained are unique to a given panel and set of samples. The quality of the data is very dependent on the experience and ability of the panel leader.	See reference [1].
Free-choice profiling	Assessors, seated in booths, are free to each choose their own terminology and scale. A profile is derived statistically, e.g. via generalized Procrustes analysis.	Can be used with experienced assessors as a preliminary step to develop terms for consideration as descriptors. Alternative procedure in consumer testing of a group of products, avoiding the need with naive assessors to develop a set of attributes and scales.	Minimum training is required. No need to spend time on the development of reference standards.	Profiles obtained are unique to a given panel. Profiles are expressed in terms of multivariate equations that require interpretation by the panel leader. Preparation and data processing are time consuming for the panel leader.	See reference [6] for a description and review.
Time-intensity profiling	Assessors, seated in booths, record the intensity of an attribute over time.	Recommended in studies of attributes that change with time in the mouth, on the skin, etc., after the moment of ingestion or application.	The only technique available to describe attributes that change with time, e.g. lingering aftertastes.	Only one attribute (or at most two) can be studied. Training of assessors required. Technique is time-consuming.	See reference [7] for a description and review.

Table 2 — Steps in establishing a sensory profile and relevant International Standard

Step No.	Action	Relevant International Standard
1. Establish a sensory facility	Establish booth area, preparation area, etc.	ISO 8589 (Design of test rooms)
2. Select products that illustrate the range of attributes to be encountered	One or two technical experts obtain many products and select approx. 6 to 10 if possible.	Use experts trained per ISO 8586-2 (Expert assessors)
3. Select and train assessors for the project	Panel leader assembles a group of candidates and trains them, using the products from step 2.	ISO 8586-1 (Selected assessors) ISO 5496 (Recognition of odours)
4. Select descriptors (attributes) suitable for the application (can be combined with step 3)	Panel leader chooses from existing terms, or assessors evaluate the products from step 2 and propose a number of descriptors. Selection is made by consensus or by multivariate analysis. Suitable reference standards are selected with which the descriptors can be demonstrated.	ISO 5492 (Vocabulary) ISO 6564 (Flavour profiles) ISO 11035 (Identification of descriptors) ISO 11036 (Texture profile)
5. Determine the order of perception of the attributes in the profile (if needed)	Panel leader works with the assessors during steps 3 and 4.	
6. Select the scale or scales of intensity to be used with the descriptors	Panel leader selects the most appropriate scale(s).	ISO 4121 (Evaluation using scales) ISO 11056 (Magnitude estimation)
7. Train the assessors to use the selected descriptors and scale(s)	Panel leader works with the assessors to improve their sensitivity, repeatability and the agreement between them (the latter does not apply in free-choice profiling).	ISO 8586-1 (Selected assessors) and ISO 1586-2 (Expert assessors)
8. Conduct the test(s)	Assessors evaluate the test samples.	ISO 6658 (General guidance) ISO 6564 (Flavour profiles)
9. Report the results	Analyse the data statistically and present the results in the form of tables and diagrams; draw the conclusion(s).	See 7.5 ISO 6564 (Flavour profiles) ISO 11036 (Texture profile)

5 General test conditions

5.1 Test room

Carry out sensory profiling in booths under the conditions as described in ISO 8589. For consensus profiling and for the initial phases of the identification and selection of descriptors, arrange to have assessors seated around a central table, on which reference substances may be available; see ISO 8589.

5.2 Apparatus and sampling

Select the number of samples and their mode of presentation so as not to bias the results in any way; see ISO 6658.

5.3 Preliminary discussion and test

See ISO 6658:1985, Clause 4. Ascertain that the assessors are fully familiar with any particular characteristic to be studied and with the mechanics of the test. If necessary, arrange a preliminary general discussion concerning the test problem and the nature of the samples. Present and discuss a few samples typical of the series to be analysed. Limit the number to two or three. If the test concerns the description of off-flavours, include in the preliminary test a sample free from any off-flavours, and/or a demonstration of the off-flavour to be described. Present typical samples and direct the discussion in a manner that will not influence future judgements.

5.4 Number of assessors

Descriptive panels usually have 8 to 12 assessors, or maybe more. They may have as few as 4 (e.g. in consensus profiling). They may have as many as 20 or 30 when the purpose includes testing for taints to which only a minority may be sensitive. Specific instructions regarding panel size are not appropriate because of the many factors that have to be considered. Larger panels are required if there are more than slight differences between the panel members as regards sensitivity and/or training.

6 Selection, training and monitoring of the assessors

See ISO 8586-1 and ISO 8586-2 for descriptions of the selection, training and monitoring of assessors. Recruit candidates through talks, circulars or personal contact. Interview and screen two to three times the number of assessors required. The important characteristics are the following:

- interest and motivation;
- ability to memorize and communicate sensory impressions;
- availability for panel sessions;
- capacity to concentrate and honesty in reporting sensations;
- promptness;
- good health;
- ability to discriminate the specific characteristics studied;
- engagement for the duration of the study.

Sensory acuity is less important provided panel members balance each other's strengths and weaknesses, which is often the case with panels of 10 to 15 or more.

Include in screening procedures the products or samples for which the panel will be used. Follow with a training programme using the product itself and the terminology selected according to 7.1. Include samples chosen, produced or modified to cover the range of each attribute likely to be encountered in future panel work. Note that the training programme and the selection of terminology may to some extent be combined into a single activity. A panel may be trained for more than one type of product, but it is important, for each type, that the panel has a good appreciation of the range of attributes that it may encounter.

Monitor the performance of selected assessors at appropriate regular intervals (e.g. by giving them coded samples of known characteristics or by introducing replicates). It is the panel leader's responsibility to organize retraining of any assessor whose performance has dropped from the level which permitted admission to the panel.

7 Procedure

7.1 Choosing the optimal attributes (descriptors)

7.1.1 General

At the discretion of the panel leader, use one of the three approaches set out in Table 3 or any combination thereof for this important step. The purpose is to identify and select a set of non-overlapping attributes that, as far as possible, permit a complete descriptive analysis of the samples under study.

7.1.2 Order of perception

In addition to detecting the attributes of a sample and then scaling each attribute, panels can often detect differences among products in the order in which attributes are perceived and in the duration of the sensation. In certain products (e.g. beverages), the order of perception of certain attributes is as characteristic of the product profile as the individual flavour notes and their respective intensities. In other products, the order of perception can change, for example in accordance with textural or physical properties, or with changes during the evaluation.

EXAMPLES A piece of chocolate which melts; a facial tissue which is moistened.

The order of perception should determine the order in which attributes are evaluated. Aftertaste or afterfeel should be evaluated last; these are attributes which can still be perceived after the sample has been consumed or used. Aftertaste or afterfeel need not imply a defect or a negative characteristic. For example, the cool aftertaste of a mouthwash and the spicy afterheat of a curry are desirable flavour notes. In oral texture and in skinfeel and fabric feel applications, the order of perception can be predetermined by the way the product is handled. By controlling the manipulation (one chew, one manual squeeze), the panel leader determines which attributes are allowed to be perceived first.

7.1.3 Overall evaluation

As a part of a profiling session, usually at the end, assessors may be asked to provide one or more overall evaluations. Examples are

- overall fruitiness or spiciness,
- amplitude (overall flavour intensity),
- degree of presence of agreed defects, and
- overall non-hedonic score (e.g. by a particular system of grading or rating).

CAUTION Hedonic evaluations by trained sensory panels can be biased. If they are used to guide new product development, its aim or goal should first be set by consumer tests in which respondents are selected to represent various prospective customer groups, and are asked to grade products on a scale of acceptability or intent-to-purchase. Sensory assessors have been trained to be objective in describing food products and may, consciously or unconsciously, adopt different measures for what is high or low quality.

Table 3 — Suggested procedures for choosing optimal descriptors

No.	Principle	Method	Advantages	Disadvantages	Illustrative examples
1	Use existing terminology and reference standards.	Consult the literature and experts to make an appropriate selection. Acquire the prescribed standards and use these to teach the assessors the quality of each descriptor and, if needed, an intensity scale for that descriptor.	The accumulated experience of the experts is utilized. Profiles may be interpreted by other groups and compared to other research.	Existing terminology or reference standards may include choices that are imprecise or inappropriate for a particular set of samples. Attributes may be missed that could have been discovered during the development of new terminology.	Published terminology and many standards exist (e.g. for water [8]; dairy products [9]; fish products [10]; beer [11]; wine [12]; Scotch whisky [13]; and for attributes of texture and fabric feel [14]).
2	Use the panel in special session to develop the terminology it will use.	Using a panel of trained assessors, develop the terminology in round table discussion under the direction of an experienced panel leader. Reference standards are used which may be provided by the panel leader or the test requester, or by an assessor during the session. May be combined with method 1.	The process of terminology development is less time-consuming than method 3.	Profiles obtained are unique to a given panel and set of samples. They cannot be interpreted by other groups if no reference standards are given.	See reference [1], chapter 1; reference [4], chapter 6.
3	Use the panel in special session to develop the terminology it will use.	Consult ISO 11035 which describes a recommended method of identifying and selecting discriminating terms using a set of prepared training samples; then reduce the number of terms by stepwise elimination using statistical techniques.	A fully objective process of selection and elimination is used, thus terms based on traditional misconceptions or preconceived notions are minimized, and the selected terms will give optimal coverage of the qualities which the assessors perceive in the samples.	Profiles obtained are unique to a given panel and set of samples. They cannot be interpreted by other groups if no reference standards are given. The process is relatively time-consuming and requires a certain level of experience, especially in data analysis.	ISO 11035.

7.2 Selecting an appropriate scale

Once the distinguishing attributes have been selected, the next step is to choose an appropriate response scale by which to indicate the intensity of each attribute present in a given sample. Consult ISO 4121 which provides guidelines for the choice of response scales.

For purposes of statistical analysis, it is useful to distinguish between two uses of the term "scale":

1st sense: The scale is a mathematical relationship between the strength of the sensory impression and the numbers recorded by the assessor.	For clarification, here called MEASUREMENT SCALE. Examples: ordinal scale, interval scale, ratio scale.
2nd sense: The scale is a measuring instrument: the assessor has a scale to be used in making his or her response.	For clarification, here called RESPONSE SCALE. Examples: a ruler, a kitchen scale, a line scale.

NOTE 1 The measurement scale sets limits on the type of statistical manipulations that can be permitted. For example, it is often considered advisable to use non-parametric methods if the data are on an ordinal scale. For guidelines, see ISO 4121.

Response scales used in sensory profiling may be numerical or semantic, continuous or discontinuous, unipolar or bipolar. For time/intensity profiling, dynamic response scales may be applied.

NOTE 2 Data obtained with discontinuous response scales may suffer from so-called “end effects”: assessors are reluctant to use numbers near either end of the scale for fear that later samples may be more extreme. Continuous response scales, e.g. line scales, are less prone to produce end effects. Pictorial response scales are suitable for use with children.

7.3 Conducting the test

7.3.1 Preparation of test samples

See ISO 6658. Take care that assessors cannot draw conclusions about the nature of samples from the way they are presented. For example, use coloured glasses or coloured lights to mask differences in appearance, if needed. Standardize the preparation of samples and serve samples at uniform temperature. Code samples with three-digit random numbers, or code the order of presentation to the assessor. If the number of possible combinations becomes too large for full randomization, use an appropriate sampling plan (see reference [15]).

7.3.2 Presentation of samples

Present the samples simultaneously, if possible, or else monadically (in succession, one-by-one). Simultaneous presentation facilitates the comparison of samples and is useful especially if a comparison on many attributes is desired, but profiles can become interdependent and hence less descriptive. Use monadic presentation if this is a problem, and also if samples show differences in appearance that cannot be masked, and if samples have strong or persistent flavours (e.g. smoked or spicy foods, bitter substances, or greasy textures).

7.3.3 Preparation of scoresheets

Use preprinted scoresheets containing instructions regarding the scale to be used. Alternatively, use computer screens or tablet digitizers to record the verdicts. Leave a blank space on scoresheets and invite the assessors' comments or suggestions for additional attributes. Note that several ratings placed on a single sheet or screen may mutually influence or distort each other (i.e. halo effect: a positive or negative evaluation may carry over from one attribute to the following).

If it is not possible to use a single screen or sheet per sample and per attribute, attempt to minimize the distortion by appropriate training of the assessors and/or by including samples of known characteristics. Again, when a sheet contains several samples, order of presentation effects may occur (e.g. contrast effects, group effects or placement effects). For that reason, it is desirable to balance or randomly assign the order of samples on the sheet or screen from one assessor to another.

7.3.4 Evaluation of samples

Arrange for assessors to work alone in a booth, except in the consensus approach where the assessors sit around a table on which reference standards are placed. Do not disclose the identity of the samples until the assessors have completed all replicate assessments. Occasionally include prepared samples or repeat samples for educational purposes.

As a general rule, offer a maximum of six samples per session and per assessor for a full descriptive analysis, or a maximum of ten samples for partial profiling with under ten descriptors. Use smaller numbers if small differences are critical, and for samples with strong or persistent flavour.

7.3.5 Duplicate tests

To increase the reliability and validity of results, present any sample or sample group two or three times or more, if possible on different days. In choosing the number of replications, be guided by the precision required, the observed dispersion of results, and by any trend (e.g. towards improved discrimination as the assessors become familiar with the samples). Replication is necessary to provide an estimate of experimental error. Note that the effect of repeating the same sample is to show the dispersion of scores among assessors, whereas repeating across batches of the product will reflect variations within the product as well. Watch for assessor × product interaction (assessors evaluate the products differently).

7.4 Time-intensity profile

In time-intensity profiling, the intensity of an attribute is recorded over time (see reference [7]). The sweetness of a sweetener or the bitterness of a beer may be recorded continuously over a period of 1 min to 3 min after ingestion. Usually the intensity is found to rise rapidly to a maximum and then to decrease more slowly. Flavour release from chewing gum may be recorded every minute for 30 min.

The response may be recorded using pencil and paper, or on a scroller chart recorder, or by any of a number of commercially available computer systems. The assessor should not see the evolving response curve being traced as this may result in bias from preconceived notions of its form. The important variables to consider are the following:

a) protocols for evaluation: type of presentation, amount of product, time to hold in the mouth or on the skin, type of manipulation, expectoration or swallowing;

b) a protocol must be established to coordinate product evaluation (sample holding) with recording of the response (data entry) so as to reduce bias from the mode of presentation;

c) assessors may require several training sessions in order for all of the protocols needed for a well-controlled study to be developed and learned.

7.5 Evaluation of results, using appropriate statistical methods

7.5.1 General

For each attribute and treatment, tabulate the results and evaluate their significance, using appropriate statistical methods. See examples in ISO 4121, ISO 8586-1, ISO 11035, ISO 11056 and Annex A.

Profile data are not a simple case for statistical evaluation, hence experiments should be designed with a view to the statistical analysis that will be used. In all but the simplest cases, it is advisable to consult a statistician about the design and analysis. Only general guidelines are given below.

Typically the analysis used is two-way (or higher) Analysis of Variance (ANOVA). In addition, multivariate methods may be used to explore the interaction often seen between the attributes.

7.5.2 Sources of variability

In an ANOVA on the profile ratings for an attribute, the sum of squares, etc. are calculated for each source of variability. Two principal sources determine the early part of the analysis: the assessor and the product. Exceptionally, as in the example in Annex A, product variability may be ignored as all samples of a given product may be assumed to be identical.

Typically, the third, fourth, etc. sources of variability to be considered are those between “treatments”. The treatments are the factor or factors which were deliberately introduced and which are to be quantified.

In the design of serving protocols for profile panels, the investigator's concern is to reduce the total variability of the profile ratings for each treatment. Total variability is the sum of the relative variabilities from each source, measured as variances. For example, if the assessor variability is quite large whereas the product is quite uniform, the investigator may wish to add more assessors. Conversely, if the product is non-uniform (e.g. hamburgers), the investigator may double or triple the number of samples while leaving the number of assessors unchanged.

7.5.3 Replication: Assessors as fixed effects or random effects

Replication, although highly desirable to improve the reliability of the results, introduces two important considerations into the ANOVA. The first is to decide whether assessors will be seen as fixed or random effects. When assessments are replicated, assessors are treated as one of the factors in a factorial experiment. Factorial effects are either random effects or fixed effects. The distinction is important because fixed effects influence the mean ratings; random effects influence the error term.

Strictly speaking, assessors should be considered a fixed effect since, because of their training, they do not represent a random sample from any existing population except perhaps that of trained assessors. However, in practice it is often the experimenter's aim to provide a conclusion that is valid not only for the panel. If it is statistically defensible, a decision may be made to consider the assessors as a random effect. A factor is random when its levels present in the experiment may be different without changing the purpose of the experiment.

7.5.4 Replication within a batch or between batches

The second point to consider is that the model for the ANOVA will differ if replications are performed on the same batch of products or not. At least three possible schemes are possible, each requiring a different model. See Figure 1.

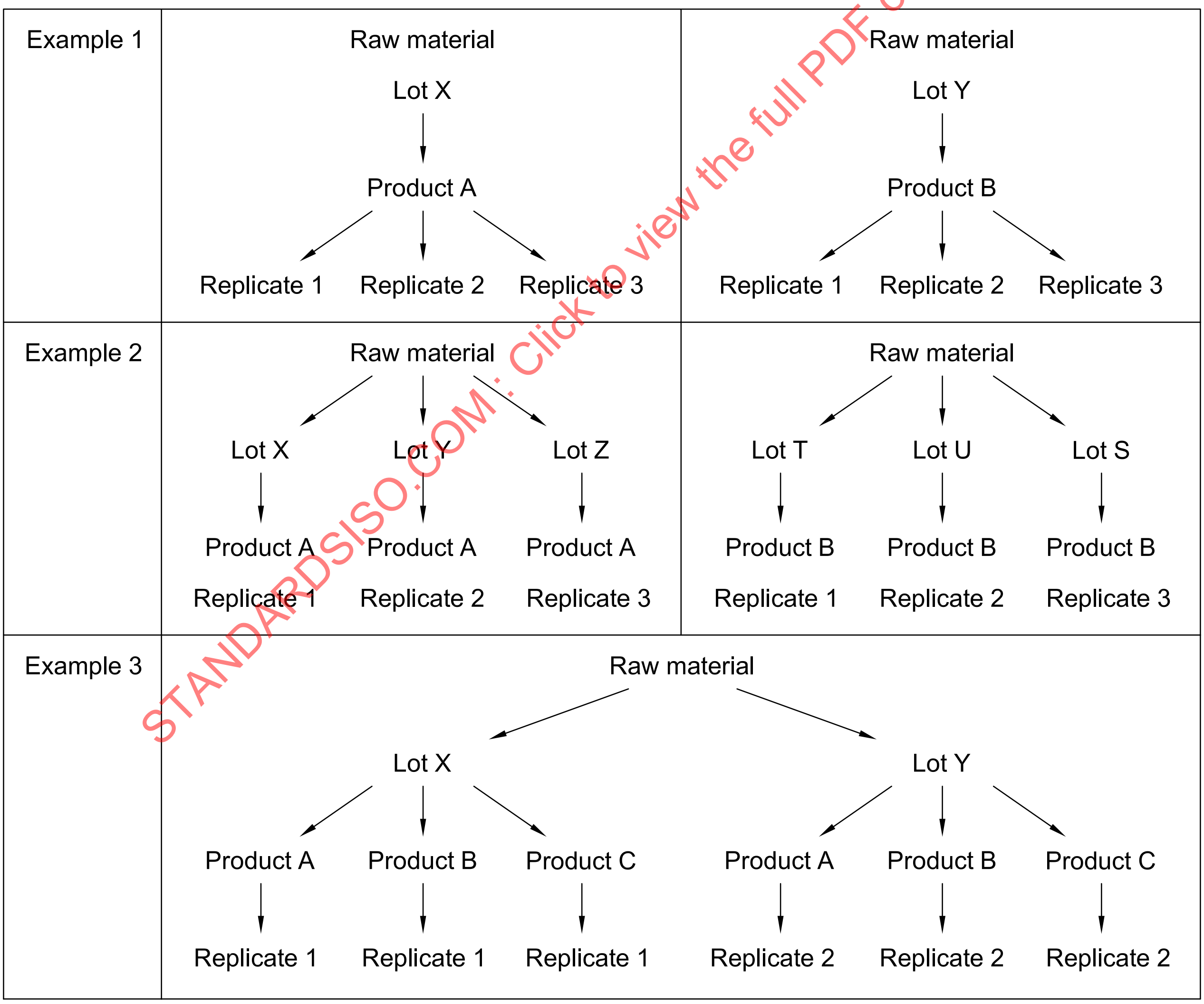


Figure 1 — Three experimental designs requiring different ANOVA models

Example 1 in Figure 1 is the commonly used repeated-measures design. Replicates 1, 2 and 3 are actually subsamples of a single batch of each product A and B. No batch-to-batch variability is present in the design. Examples 2 and 3 in Figure 1 include true replications of both products A and B. A split-plot ANOVA model is used to analyse all three examples (see Table 4). In each example the sub-plot effects include assessors and the assessor-by-product interaction. The whole-plot effects differ among the examples.

The whole-plot for Example 1 and Example 2 is a completely randomized design involving products and error (i.e. replicates within products). The whole-plot analysis for Example 3 is a randomized block design (blocking on raw-material lots) involving replicates, products and error (i.e. replicates-by-products). Although the models for Example 1 and Example 2 appear to be identical, it is important to note that the applicability of the findings differ between the two examples. In Example 1, it is only possible to determine if the one *batch* of Product A is perceptibly different from the one *batch* of Product B. While in Example 2 (and Example 3), it is possible to determine if Product A is perceptibly different from Product B.

Table 4 — ANOVA models for experimental designs in Figure 1

EXAMPLE 1 Repeated measures design				
Source	d_f	S	M	F
Total	$apr - 1$	S_T	M_T	
Product	$p - 1$	S_P	M_P	$F = M_P/M_{Ea}$
Error (A) = Rep within product	$p(r - 1)$	S_{Ea}	M_{Ea}	
Assessor	$a - 1$	S_A	M_A	$F = M_A/M_{Eb}$
Assessor × product	$(a - 1)(p - 1)$	S_{AXP}	M_{AXP}	$F = M_{AXP}/M_{Eb}$
Error (B) = Residual	$(a - 1)p(r - 1)$	S_{Eb}	M_{Eb}	
EXAMPLE 2 Completely randomized split-plot design				
Source	d_f	S	M	F
Total	$apr - 1$	S_T	M_T	
Product	$p - 1$	S_P	M_P	$F = M_P/M_{Ea}$
Error (A) = Rep within product	$p(r - 1)$	S_{Ea}	M_{Ea}	
Assessor	$a - 1$	S_A	M_A	$F = M_A/M_{Eb}$
Assessor × product	$(a - 1)(p - 1)$	S_{AXP}	M_{AXP}	$F = M_{AXP}/M_{Eb}$
Error (B) = Residual	$(a - 1)p(r - 1)$	S_{Eb}	M_{Eb}	
EXAMPLE 3 Randomized block split-plot design				
Source	d_f	S	M	F
Total	$apr - 1$	S_T	M_T	
Rep	$r - 1$	S_R	M_R	
Product	$p - 1$	S_P	M_P	$F = M_P/M_{Ea}$
Error (A) = Rep – product	$(r - 1)(p - 1)$	S_{Ea}	M_{Ea}	
Assessor	$a - 1$	S_A	M_A	$F = M_A/M_{Eb}$
Assessor × product	$(a - 1)(p - 1)$	S_{AXP}	M_{AXP}	$F = M_{AXP}/M_{Eb}$
Error (B) = Residual	$(a - 1)(p - 1)(r - 1)$	S_{Eb}	M_{Eb}	

where

a is the number of assessors

p is the number of products

r is the number of replications

d_f is the degrees of freedom

S is the sum of squares

M is the mean square

The ANOVA models in Table 4 apply when assessors are treated as a fixed effect in the analysis. If assessors are randomly selected, then the following quasi F statistic is used to test for differences among the products:

$$F = \frac{M_{\mathrm{P}}}{M_{\mathrm{Ea}} + M_{\mathrm{AXP}} - M_{\mathrm{Eb}}} \quad (1)$$

The numerator degrees of freedom for F are $d_{\mathrm{f1}} = p - 1$.

The denominator degrees of freedom are estimated by

$$d_{\mathrm{f2}} = \frac{\left(M_{\mathrm{Ea}} + M_{\mathrm{AXP}} - M_{\mathrm{Eb}}\right)^2}{\left(\frac{M_{\mathrm{Ea}}^2}{d_{\mathrm{f,Ea}}}\right) + \left(\frac{M_{\mathrm{AXP}}^2}{d_{\mathrm{f,AXP}}}\right) + \left(\frac{M_{\mathrm{Eb}}^2}{d_{\mathrm{f,Eb}}}\right)} \quad (2)$$

7.5.5 Particular case of free-choice profile

Free-choice profiling differs from the other methods described in this International Standard in that each member of the panel uses his/her individual list of terms instead of a common list. For a description and review, see reference [6].

Results are interpreted with the aid of generalized Procrustes analysis or equivalent programs. The output is in the form of a sensory map of the products.

7.6 Writing the test report

7.6.1 Report the main results as described in 7.5. If desired, calculate confidence regions either for a single session if a large panel is used, or more commonly, showing the average variability of the panel over a period of weeks or months. Make reference to the relevant International Standard and include the following information:

a) objective of the study;

b) full identification of sample(s);

c) if two or more samples were compared, criterion of comparison and procedure applied;

d) type of response scale used;

e) reference substances and form in which they were presented;

f) any other information given the assessors during the test, e.g. information regarding the type or brand of product;

g) number of samples presented and number of assessors, selected assessors or experts;

h) the operating conditions of the test, including any conditions differing from the recommendations given in this method;

i) the results, with any statistical evaluation;

j) date and time of the test;

k) name of the panel leader and the test supervisor.

If agreed, the content of the customer report may be less exhaustive.

7.6.2 Illustrate the results with graphs and diagrams. Be guided by the examples in Annex B, and those in ISO 6564. Show consensus profiles as semicircular plots. Show other profiles as histograms or as spiderweb plots, or if multivariate analysis is used, as Principal Component diagrams or canonical plots. Show confidence regions as narrow extensions to the bars of histograms, or as bands at the points of spiderweb plots.

Annex A
(informative)

Comparison of eight model cola systems

A.1 The randomized (complete) block design (RCBD) used here is appropriate when there is negligible sample-to-sample variability within each of the products in the study, and the number of samples is sufficiently small to be presented in a single session. Implicit in the RCBD is the assumption that only a single sample of each product is necessary to completely represent the entire population. The only relevant source of variability that need be considered, other than that between products, is that arising from differences among assessors (i.e. assessors are treated as a random effect in this example).

A.2 Eight model cola systems were evaluated once by eight trained assessors. Ten attributes (R1 to R10) were scored using a 0-to-15 numerical scale to represent the perceived intensities of the attributes. Table A.1 shows the results.

The physical differences among the eight samples were deemed to be highly reproducible at the bench level. Therefore it was decided (exceptionally) that no replication of the samples was needed. Differences that the trained assessors perceived among the eight samples would be assumed to hold for any future production of the sample products. Also it was determined by an external assessor tracking system that the assessors were consistent one-to-the-other in reporting perceived differences among cola products. Therefore, no repeated evaluations of the same samples were performed.

Table A.1 — Descriptive profile panel — Model cola study — Listing of input data

Obs.	Treatment	Assessor	R1	R2	R3	R4	R5	R6	R7	R8	R9	R10
1	Sys 1	15	3,0	3,0	4,0	7,5	2,0	0,0	5,0	6,0	0,0	0
2	Sys 1	16	4,0	3,0	3,0	8,5	5,5	3,0	7,0	3,0	0,0	3
3	Sys 1	31	3,0	3,0	2,0	8,5	1,5	1,0	10,5	6,0	0,5	2
4	Sys 1	32	2,0	3,0	1,0	12,0	1,0	1,0	1,0	4,0	2,0	0
5	Sys 1	33	3,0	4,0	1,5	7,5	1,5	0,0	7,0	4,5	3,0	2
6	Sys 1	34	3,0	2,0	2,0	10,0	0,0	0,0	6,5	3,0	0,0	0
7	Sys 1	36	4,0	3,0	1,0	10,0	0,0	0,0	8,0	2,0	1,0	0
8	Sys 1	38	3,0	2,0	1,0	12,0	5,0	1,0	5,0	5,0	2,0	0
9	Sys 2	15	3,0	2,0	2,0	7,5	2,0	0,0	4,0	4,0	2,0	0
10	Sys 2	16	5,0	4,0	4,0	9,0	4,0	4,0	6,0	6,0	2,0	4
11	Sys 2	31	4,0	3,5	3,0	10,5	2,0	1,0	9,5	9,5	0,0	2
12	Sys 2	32	2,0	3,0	1,0	10,0	1,0	1,0	1,0	1,0	2,0	0
13	Sys 2	33	2,0	2,5	1,5	10,0	1,5	0,0	4,0	4,0	2,0	4
14	Sys 2	34	3,0	2,0	2,0	8,0	2,5	2,0	4,0	4,0	0,0	3
15	Sys 2	36	4,0	3,0	1,0	10,0	0,0	0,0	8,0	8,0	1,0	0
16	Sys 2	38	2,5	2,0	1,0	11,0	4,0	1,5	4,0	4,0	3,0	0
17	Sys 3	15	2,0	2,0	2,0	7,5	4,0	0,0	4,0	5,0	4,0	0
18	Sys 3	16	4,0	5,0	5,0	10,0	2,5	3,0	4,0	4,0	0,0	4
19	Sys 3	31	4,5	3,5	3,0	10,5	2,0	1,5	8,5	5,0	0,0	3
20	Sys 3	32	2,0	3,0	2,0	10,0	1,0	1,0	1,0	3,0	2,0	0
21	Sys 3	33	1,5	1,5	0,0	11,0	1,0	0,0	4,0	3,0	2,0	4
22	Sys 3	34	2,0	2,0	2,0	7,0	1,5	0,0	3,0	0,0	0,0	0
23	Sys 3	36	4,0	3,0	1,0	10,0	0,0	0,0	8,0	2,0	1,0	0
24	Sys 3	38	2,0	1,5	0,5	10,0	3,0	1,5	2,0	6,0	3,0	0
25	Sys 4	15	0,0	2,0	2,0	7,0	4,0	0,0	4,0	5,0	4,0	0
26	Sys 4	16	5,0	6,0	4,0	11,0	3,0	2,0	3,0	4,0	0,0	4
27	Sys 4	31	4,0	3,5	3,0	10,5	1,5	1,5	8,5	5,0	0,5	3
28	Sys 4	32	3,0	3,0	1,0	10,0	1,0	1,0	1,0	3,0	2,0	0
29	Sys 4	33	2,5	2,5	0,0	12,0	1,0	1,0	4,0	3,0	3,0	4
30	Sys 4	34	2,0	2,0	2,0	7,0	1,0	0,0	2,0	0,0	0,0	0
31	Sys 4	36	4,0	3,5	1,5	9,5	0,0	0,0	7,0	2,0	1,0	0
32	Sys 4	38	2,0	1,5	0,5	8,0	2,0	1,5	1,0	6,0	3,0	0
33	Sys 5	15	2,0	3,0	0,0	8,0	4,0	0,0	5,0	5,0	3,0	0
34	Sys 5	16	3,0	2,0	5,0	11,0	6,0	2,0	4,0	3,0	3,0	0
35	Sys 5	31	4,0	3,0	1,5	13,5	0,5	0,5	9,5	5,5	1,0	1
36	Sys 5	32	2,0	2,0	3,0	10,0	1,0	1,0	1,0	4,0	2,0	0
37	Sys 5	33	5,5	2,0	2,0	9,0	0,0	0,0	4,0	1,0	2,0	0
38	Sys 5	34	1,0	2,0	2,0	9,0	2,0	2,0	5,0	3,0	0,0	0
39	Sys 5	36	3,0	3,0	2,0	7,5	0,0	0,0	8,0	2,0	1,0	0
40	Sys 5	38	2,5	1,5	2,0	14,0	4,0	2,0	3,0	3,0	2,0	1
41	Sys 6	15	2,0	2,0	0,0	8,0	3,0	0,0	4,0	5,0	3,0	0
42	Sys 6	16	2,0	3,0	6,0	10,0	6,5	3,0	4,0	4,0	2,0	0
43	Sys 6	31	5,0	4,0	2,0	11,5	1,0	0,0	8,5	5,5	2,0	2
44	Sys 6	32	2,0	1,0	2,0	10,0	1,0	0,0	0,0	2,0	2,0	0
45	Sys 6	33	4,0	1,0	1,0	7,0	0,0	0,0	4,0	1,0	3,0	0
46	Sys 6	34	0,0	2,0	2,0	10,0	0,0	0,0	4,0	2,0	0,0	0
47	Sys 6	36	4,0	2,5	2,0	7,5	0,0	0,0	8,0	2,0	1,0	0
48	Sys 6	38	2,0	1,5	1,5	12,0	4,0	2,0	2,0	3,0	3,0	2
49	Sys 7	15	2,0	2,0	0,0	8,0	2,0	0,0	3,0	4,0	4,0	0
50	Sys 7	16	4,0	3,0	7,0	9,0	4,0	5,0	2,0	1,0	1,0	0
51	Sys 7	31	5,0	4,0	2,0	11,5	2,0	0,0	6,0	3,5	2,0	2
52	Sys 7	32	2,0	0,0	3,0	10,0	1,0	0,0	0,0	2,0	2,0	1
53	Sys 7	33	3,0	1,0	1,0	7,0	0,0	0,0	3,0	1,0	3,0	0
54	Sys 7	34	1,0	2,0	2,0	9,0	2,0	2,0	3,0	0,0	0,0	0
55	Sys 7	36	4,0	2,5	2,0	9,0	0,0	0,5	8,0	2,0	1,0	0
56	Sys 7	38	2,0	1,5	1,0	10,0	4,0	2,0	2,0	3,5	3,0	3
57	Sys 8	15	2,0	2,0	0,0	8,0	1,0	0,0	2,0	4,0	4,0	0
58	Sys 8	16	5,0	3,0	6,0	7,5	6,0	3,5	0,0	2,0	2,0	0
59	Sys 8	31	5,0	4,5	2,0	11,5	2,0	0,0	5,5	3,5	2,5	3
60	Sys 8	32	2,0	2,0	2,0	10,0	1,0	1,0	1,0	3,0	2,0	1
61	Sys 8	33	2,0	1,0	1,0	8,0	0,0	0,0	2,0	1,0	4,0	0
62	Sys 8	34	1,0	3,0	2,0	10,0	2,0	0,0	3,0	0,0	0,0	0
63	Sys 8	36	4,0	2,5	2,0	9,0	0,0	0,5	5,0	2,0	1,0	0
64	Sys 8	38	1,5	1,5	1,0	9,0	3,0	2,0	1,5	3,5	3,0	3

A.3 A two-way ANOVA was performed for each attribute with the assessor as the blocking factor and the cola systems as the experimental factor or treatment. The results of the analyses for the attributes “citrus” and “astringent” are presented in Table A.2.

Table A.2 — Descriptive profile panel — Model cola study — ANOVA of attribute data

Source	d_f	Sum of squares	Mean square	F	P-Value
Dependent variable: R1 Citrus					
Assessor	7	61,59			
Treatment	7	1,96	0,2807	0,36	0,9215
Error	49	38,32	0,7820		
Corrected total	63	101,87			
Dependent variable: R8 Astringent					
Assessor	7	107,05			
Treatment	7	27,11	3,8728	5,22	0,0002
Error	49	36,33	0,7414		
Corrected total	63	170,48			

Based on the criterion that an α-risk of $\leqslant$ 5 % is required for a significant difference, it was concluded that there was not sufficient evidence to conclude that perceptible differences existed among the samples in the “citrus” attribute. However, there *was* sufficient evidence to conclude that the samples differed in astringency. A summary of all ten attributes with significant differences annotated appears in Table A.3.

Table A.3 — Descriptive profile panel — Model cola study — Summary of attribute data (mean values)

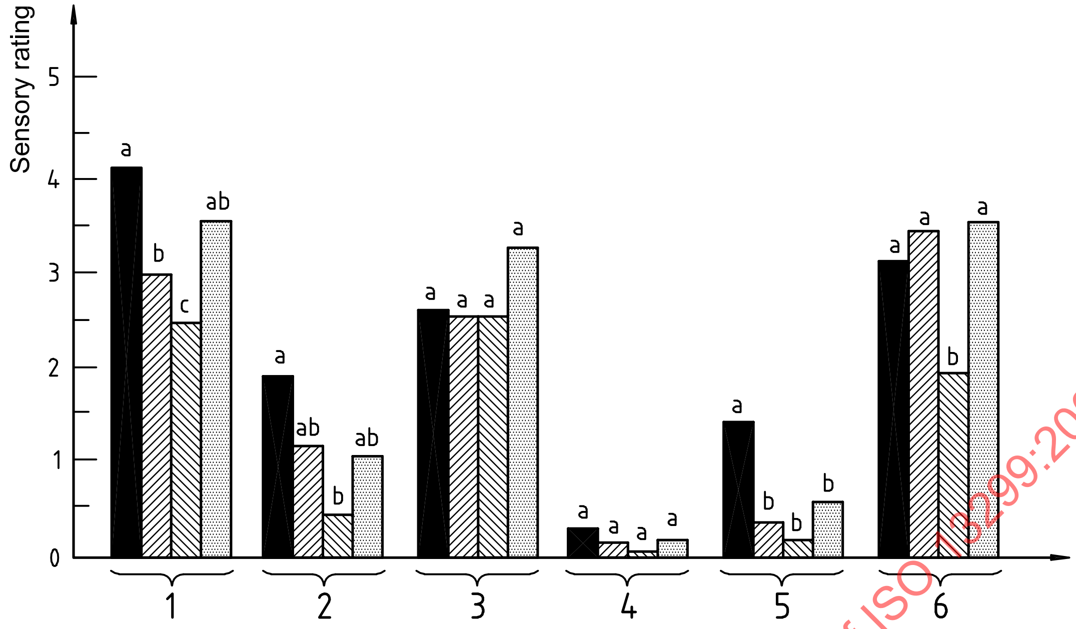
Attribute	Cola system No.								
	Sys 1	Sys 2	Sys 3	Sys 4	Sys 5	Sys 6	Sys 7	Sys 8	P-Value
Citrus	3,1	3,2	2,8	2,8	2,9	2,6	2,9	2,8	0,92
Spice	2,9	2,8	2,7	3,0	2,3	2,1	2,0	2,4	0,09
Vanilla	1,9	1,9	1,9	1,8	2,2	2,1	2,3	2,0	0,97
Sweet	9,5	9,5	9,5	9,4	10,3	9,5	9,2	9,1	0,83
Sour	2,1	2,1	1,9	1,7	2,2	1,9	1,9	1,9	0,97
Bitter	0,8	1,2	0,9	0,9	0,9	0,6	1,2	0,9	0,61
Prickle	6,3e	5,1d	4,3cd	3,8bc	4,9d	4,3cd	3,4b	2,5a	< 0,01
Astringent	4,2d	3,8cd	3,5cd	3,5cd	3,3bc	3,1bc	2,1a	2,4ab	< 0,01
Mouth dry	1,1	1,5	1,5	1,7	1,8	2,0	2,0	2,3	0,05
Mouth coat	0,9	1,6	1,4	1,4	0,3	0,5	0,7	0,8	0,23

NOTE For attributes “prickle” and “astringent”, mean values not followed by the same letter (a, b, c, d) are different at the α = 0,05 level of significance using Fisher's LSD.

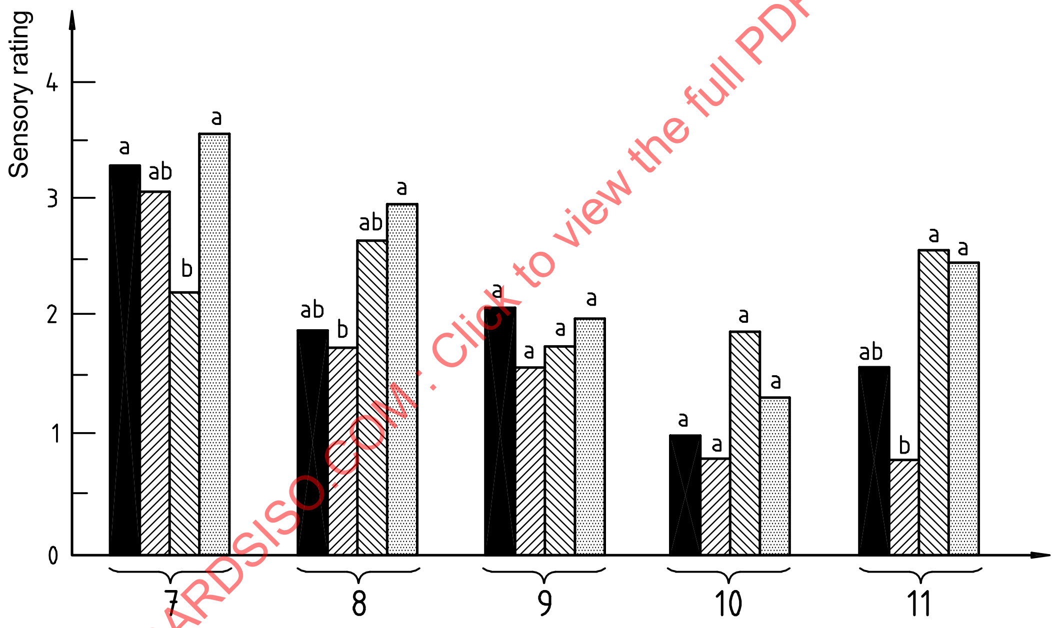
Annex B
(informative)

Recommended graphical and diagrammatic methods of representing sensory profiles

Figures B.1 to B.5 are examples from the literature. Other types of graphical representation may be equally satisfactory.



a) **Appearance and odour**



b) **Texture**

Sample

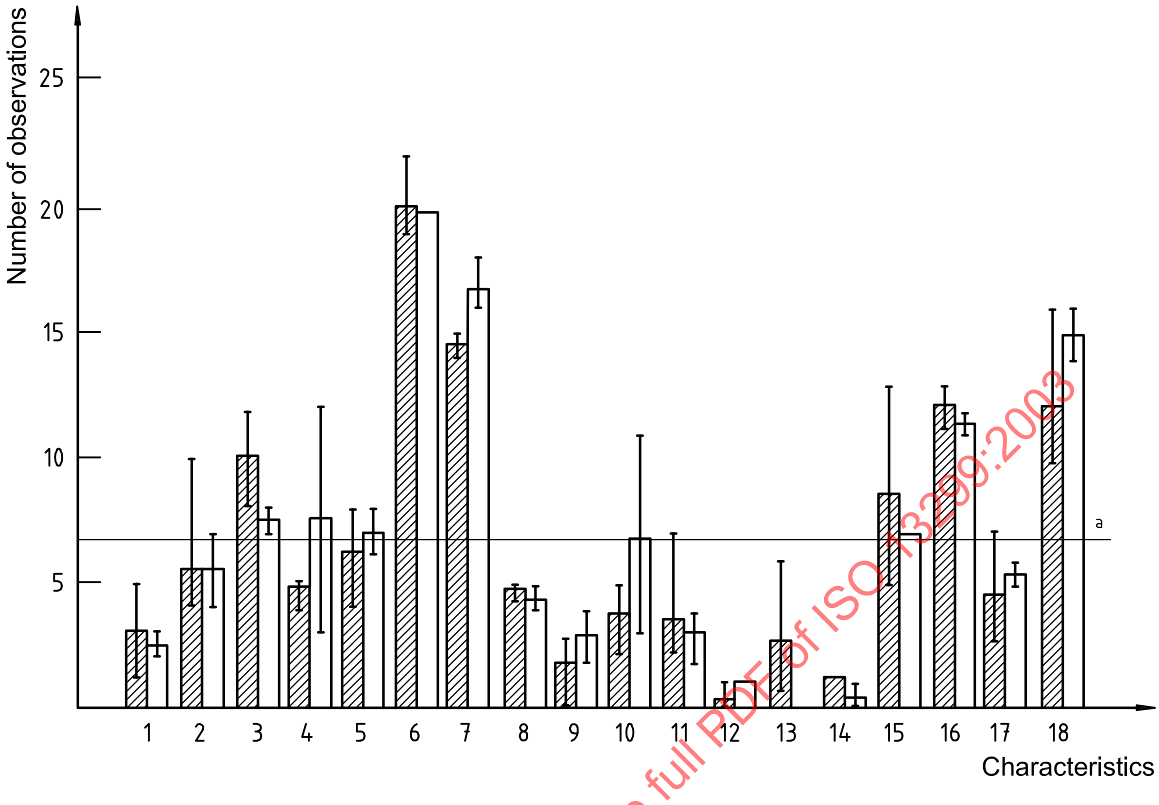
- Leg
- Mixed
- White
- Claw

Key

1 wet dry
2 ammonia
3 cooked crab
4 putrid
5 TMA
6 cereal
7 moistness
8 fibrousness
9 adhesiveness
10 chewiness
11 particle size

NOTE Meat types shown with different letters (a, b, c) above the bar are different at the $\alpha \leqslant 0{,}05$ level of significance. Taken from reference [16].

Figure B.1 — Histogram showing appearance and odour profiles and texture profiles of minced meat products from blue crab



Key

1 grass
2 fruity
3 raisin
4 honey
5 blackcurrant
6 longonberry
7 conifer
8 turpentine
9 camphor
10 tarry
11 paper
12 sulfurous, cooked cabbage
13 vanilla
14 minty, eucalyptus
15 spicy, clove
16 sweet
17 sour
18 bitter

a Line of significance: characteristics which exceed the line are considered typical characteristics of the sample.

NOTE Vertical line bars represent the variance. Taken from reference [17].

Figure B.2 — Histogram showing flavour profiles of extracts of sweet gale (bog myrtle)