BS ISO 11136:2014

BSI Standards Publication

Sensory analysis — Methodology — General guidance for conducting hedonic tests with consumers in a controlled area

bsi.

...making excellence a habit.™

National foreword

This British Standard is the UK implementation of ISO 11136:2014.

The UK participation in its preparation was entrusted to Technical Committee AW/12, Sensory analysis.

A list of organizations represented on this committee can be obtained on request to its secretary.

This publication does not purport to include all the necessary provisions of a contract. Users are responsible for its correct application.

ISBN 978 0 580 72500 5

ICS 67.240

Compliance with a British Standard cannot confer immunity from legal obligations.

This British Standard was published under the authority of the Standards Policy and Strategy Committee on 31 December 2014.

Amendments issued since publication

Date	Text affected

BS ISO 11136:2014

INTERNATIONAL STANDARD

ISO 11136

First edition
2014-12-01

Sensory analysis — Methodology — General guidance for conducting hedonic tests with consumers in a controlled area

Analyse sensorielle — Méthodologie — Lignes directrices générales pour la réalisation d'épreuves hédoniques effectuées avec des consommateurs dans un espace contrôlé

Reference number
ISO 11136:2014(E)

COPYRIGHT PROTECTED DOCUMENT

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 ... **v**

1 Scope ... **1**

2 Normative references ... **1**

3 Terms and definitions ... **1**

4 Preliminary considerations ... **3**
4.1 Study proposal ... 3
4.2 Specification of the target population ... 4
4.3 Further specifications ... 4

5 The consumer sample ... **5**
5.1 General ... 5
5.2 Recruiting the consumers ... 5
5.3 Assembling a consumer sample ... 7
5.4 Sample segmentation ... 7

6 Size of the consumer sample ... **8**
6.1 Decisions of measurement ... 8
6.2 Determining the sample size ... 8
6.3 Working with subgroups ... 9
6.4 Number of product samples and experimental design ... 9
6.5 Safety margin in recruitment ... 9

7 Test area ... **10**

8 Products ... **10**
8.1 Anonymous presentation of the products ... 10
8.2 Preparation and presentation of the products ... 10

9 Procedures ... **11**
9.1 Two groups of hedonic tests ... 11
9.2 Rating tests ... 12
9.3 Paired comparison tests ... 12
9.4 Ranking test ... 13
9.5 Additional questions ... 14
9.6 Instructions for the consumers ... 14

10 Organization of test sessions ... **15**
10.1 General ... 15
10.2 Plans ... 15
10.3 Number of products evaluated in a single session ... 16
10.4 Nature of the products evaluated within a session ... 17
10.5 Time of the session ... 17
10.6 Repetition of a product within a test ... 17
10.7 Pre-test ... 17

11 Analysis of the results ... **18**
11.1 General ... 18
11.2 Inferential analysis ... 18
11.3 Ratings (see 9.2) ... 18
11.4 Paired comparisons (9.3) ... 18
11.5 Ranking (9.4) ... 18

12 Study report **19**
12.1 General 19
12.2 Products 19
12.3 Test procedure 20
12.4 Consumers 20
12.5 Results 21
12.6 Annexes to the report 21
Annex A (informative) **Example of a recruitment questionnaire** **22**
Annex B (informative) **Methods for evaluating the data of segmentation subgroups** **26**
Annex C (informative) **Plans for assigning products to consumers** **28**
Annex D (informative) **Influence of the sample size on the accuracy of one mean** **29**
Annex E (informative) **Minimum sample size for tests using rating** **30**
Annex F (informative) **Minimum sample size for tests using ranking** **34**
Annex G (informative) **Examples of data analysis for ratings** **36**
Annex H (informative) **Examples of data analysis for rankings** **41**
Bibliography **43**

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 34, *Food products*, Subcommittee SC 12, *Sensory analysis*.

INTERNATIONAL STANDARD ISO 11136:2014(E)

Sensory analysis — Methodology — General guidance for conducting hedonic tests with consumers in a controlled area

1 Scope

This International Standard describes approaches for measuring, within a controlled area, the degree to which consumers like or relatively like products.

It uses tests based on collecting consumers' responses to questions, generally on paper or via a keyboard or a touch screen. Tests of a behavioural nature (such as recording quantities consumed ad libitum by the consumers) do not fall within the scope of this International Standard.

The hedonic tests dealt with in this International Standard, can be used as a contribution to the following:

— comparing a product with competitor products;

— optimizing a product so that it obtains a high hedonic rating or is liked by a large number of consumers;

— helping to define a range of products to correspond to a particular consumer target population;

— helping to define a best-before date;

— assessing the impact of a product formulation change on the pleasure given by the product;

— studying the impact of sensory characteristics of a product on degree to which it is liked, independently of the product's extrinsic characteristics, such as brand, price, or advertising;

— studying the effect of a commercial or presentation variable, such as packaging.

The methods are effective for determining

— whether or not, a perceptible preference exists (difference in degree of liking), or

— whether or not, no perceptible preference (paired similarity test) exists.

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 4121, *Sensory analysis — Guidelines for the use of quantitative response scales*

ISO 5492, *Sensory analysis — Vocabulary*

ISO 5495, *Sensory analysis — Methodology — Paired comparison test*

ISO 8587, *Sensory analysis — Methodology — Ranking*

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

ISO 29842, *Sensory analysis — Methodology — Balanced incomplete block designs*

3 Terms and definitions

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

3.1
access panel
sample database of potential respondents who declare that they will cooperate for future data collection if selected

[SOURCE: ISO 20252]

Note 1 to entry: Respondents can be drawn on a more or less regular basis.

3.2
closed-ended question
question requiring respondents to select from a predetermined list of possible answers

3.3
commissioning party
person, body, or service who commissions a third party to conduct a study

Note 1 to entry: The concept of commissioning party corresponds to that of purchaser or customer in quality control. It contrasts with that of the supplier, service provider, or laboratory

Note 2 to entry: Usually, the commissioning party and the service provider belong to different enterprises. Where they belong to two departments of the same enterprise, the respective responsibilities of the two departments shall be carefully differentiated and explicitly clarified in the study proposal

3.4
consumer
respondent
person who uses a product

Note 1 to entry: For the purposes of this International Standard, the meaning is restricted to a person who is neither an expert, an expert sensory assessor nor a selected assessor in conformity with ISO 5492.

3.5
consumer sample
group of consumers recruited on the basis of criteria specified by the nature of the target population

3.6
controlled area
location providing

— controlled preparation and presentation of the products,

— comfortable conditions for consuming the products and for questioning the consumers, and

— absence of communication (verbal and non-verbal) between the consumers, guaranteeing independent responses

3.7
hedonic perception
amount of pleasure given by the sensory properties of a product

3.8
hedonic test
test aimed at measuring the overall hedonic perception of a product by consumers

3.9
open-ended question
question that does not provide ready-made alternatives for respondents but asks them to answer in their own words

Note 1 to entry: It can concern each of the products presented to the consumers or only some of them (e.g. when the question asks why a particular product has appealed most, or least, to an consumer).

3.10
position bias
bias related to the location of a product in a set of products being evaluated

3.11
product family
group of products that can replace one another under normal conditions of consumption

Note 1 to entry: Product families may vary according to culture and consumption habits.

3.12
representativeness
<of a sample> degree to which attributes of the sample agree with relevant attributes of the population from which it is drawn

3.13
sequential bias
bias related to the temporal or sequential position of a product in a set of products being evaluated

3.14
session
meeting of consumers working simultaneously according to the same protocol during a defined time period

Note 1 to entry: A session can be divided into several subsessions separated by a defined rest time. For example, a session lasting 1 h 30 min can be divided into two subsessions with a 10-min break between them.

3.15
target population
population whose hedonic response to a product is to be evaluated

4 Preliminary considerations

4.1 Study proposal

The study proposal is drawn up by the service-providing laboratory based on the information (especially concerning the objective of the study and target population) given by the commissioning party, as a written document that binds the laboratory and the commissioning party contractually. It includes the following elements:

— the objective of the study;

— the products to be evaluated (and the procedures for their procurement, especially when the laboratory has to procure one or more products);

— the target population for the study;

— the size of the consumer sample (sample size) and the way it is to be recruited;

— any limitations on consumers' prior history of participation in consumer studies;

— any limitations on other products to be evaluated in the same testing session;

— the evaluation conditions;

— the type of interview (self-completion forms or interviewers);

— the questionnaire;

— the place of evaluation;

— the number of products to be evaluated per session in the proposed study, together with an indication of other products to be evaluated where the session comprises several product families (where these products are not known at the time of drawing up the proposal, the laboratory should be committed to communicate them as soon as the laboratory knows the nature of the products);

— the conditions of preparation specific to the products to be evaluated (if they are known at this stage);

— any instructions specific to the products to be evaluated (if they are known at this stage) together with any specific information to be given about the products or the test at the start of the session;

— the name of the statistical tests to be used for interpreting the results; the type of risks used and their tolerable values; the magnitude of the minimum difference to be detected in a difference of liking test or of the maximum tolerable difference in a similarity of liking test;

— any envisaged segmentation of the consumer population, based either on characteristics known before obtaining judgment data or from the responses obtained;

— the method of examining and summarizing open-ended questions;

— any recommendations the commissioning party wishes the laboratory to draft on the basis of the results;

— the dates envisaged for the tests and for submission of the report.

When any of the above elements are unknown at the time of the study proposal, the proposal includes for each element concerned the following statement: "This element will be specified by the commissioning party by... [specifying the stage of the study]."

NOTE If a specification already exists, the study proposal may make reference to it without repeating technical points described in an operational manner in the specification.

4.2 Specification of the target population

In order to specify the target population, it is necessary to answer the following questions. These questions shall be considered for each test to be performed because the consumer target group can be different from one test to the next, even for the same product.

— Has the product to be tested already been introduced on the market? If so, is it possible to distinguish between actual and potential consumers?

— Are the consumers of interest those who actually use the product, potential consumers who have not been using the product or both groups?

— Are results from specified subgroups to be examined and compared?

— Are the results of the total consumer sample to be analysed with a view identifying subgroups?

— Are differences between the results of individuals of interest?

4.3 Further specifications

Subsequent to specifying the target population, the following aspects shall be specified:

— place of test (see Clause 7);

— expected precision of measurement (see 6.1);

— test method (see Clause 9);

— product presentation plan for the test methods selected (see 10.2);

— specifications for the hypothesis to be tested (see 6.1);

For the hypothesis to be tested, the following aspects shall then be specified:

— the composition of the consumer sample(see Clause 5);

— the size of the consumer sample (see Clause 6).

5 The consumer sample

5.1 General

A hedonic test aims to determine the acceptability of the products and/or to determine preferences among two or more products by the specified consumer population.

The composition of the consumer sample is decisive for any hedonic test as regards whether or not the result can answer the questions of interest to the commissioning party.

The consumers shall be volunteers, which does not imply that it is not permitted for them to receive expenses or a reward for participation.

It is essential to obtain spontaneous reactions.

Trained consumers shall be excluded from these tests since they are likely to differ systematically from the target population.

It is usually inadvisable to recruit a consumer sample from the personnel of a company manufacturing the products to be tested. The factors most likely to distort the results when an in-house consumer panel is used are

— a risk of the products to be tested being recognized,

— a tendency to judge in favour of the recognized products, and

— excessive familiarity with the products.

These factors carry a serious risk of the results not being representative for the target population of consumers.

Any criteria to be followed when selecting the consumers (e.g. whether or not they may be recruited from the staff of the manufacturing company or its competitors or from a particular socio-economic sector) shall be agreed with the client ordering the test irrespective of the recruiting method used.

5.2 Recruiting the consumers

5.2.1 General

WARNING The following shall be observed.

— **Confidentiality shall be observed and relevant legislation shall be complied with when creating and documenting files containing personal data.**

— **Legislation with respect to age restrictions shall be adhered to.**

Consumers can be recruited either on a task-specific basis (where consumers are freshly recruited for a particular study) or from an access panel (a list of potential consumers having known characteristics). Task-specific recruitment has often been considered a means of recruiting consumers who have not participated much in consumer studies, but this idea might have become erroneous with the multiplication of locations where hedonic tests are organized in a quasi-permanent manner.

5.2.2 Task-specific recruitment

Task-specific recruitment can be conducted in a public place, at a sales point, by telephone, by mail or electronic mail or by advertisement in a local newspaper or radio station, or by social network or other internet ways. In many cases, it is necessary to reach a prior agreement with the corporation or organization responsible for the area in which the recruitment occurs.

The manner of recruiting and the time of approaching the consumers (in particular when making contact at public places or by phone and when distributing information through the media) can have a strong effect on the composition of the consumer sample.

5.2.3 Recruitment from an access panel

For efficiency in recruiting, it is also possible to assemble an access panel in the form of a file of consumer descriptions. Prior to each test and subsequent to specifying the target group, a pre-selection can be made based on criteria recorded in the file.

A consumer description may include consumer data such as the following:

— identity details of the person concerned, contact data;

— age, gender;

— marital status, household composition;

— social situation, profession, income;

— possible health problems (e.g. on allergies, diabetes, vision impairments, etc.) relevant to the products to be tested;

— other information relevant to the products to be tested (e.g. eating or drinking habits, religious prohibitions, lifestyle, household equipment);

— information on the person's availability for participation in tests;

— information on previous participation(s) in consumer tests.

Every consumer description shall be updated at regular intervals.

Such a database allows an efficient dialogue between the commissioning party and the service provider insofar as it enables the latter to ensure, a priori, that the initial desires of the commissioning party regarding target consumers can be satisfied.

5.2.4 Frequency of employing the consumers

In comparison to task-specific recruitment, there is a greater risk of "consumer professionalizing" when recruiting from an existing pool of consumers.

It is important to avoid a training effect for the individual consumer in relation to the product group concerned. Therefore, the same consumers shall not be employed too often. As guidance, an interval of at least three months between individual tests on the same product is recommended.

The laboratory shall keep records of each consumer's frequency of participation in tests involving each product and each product family. The participation history of the consumers shall be included in the study report, possibly including information about the product families involved if requested by the commissioning party.

5.3 Assembling a consumer sample

5.3.1 Selection of consumers by means of recruitment questionnaires

When consumers are selected by means of a recruitment questionnaire relating to criteria stipulated for the composition of the consumer sample, it shall not be possible for the consumer to identify which criteria in the questionnaire are decisive for selection to participate in the test.

The following are examples of possible criteria:

— previous participation in testing the product group concerned;

— frequency of using the product ;

— the product brand usually used, product description, place of purchase, etc.;

— habits in the use or consumption of the product.

Other criteria, such as age, sex, social stratum, occupational group, or geographical location can be relevant. For reasons of practicability, classes should be specified for each of these criteria. For each class, a percentage should be stated by which this class is to be represented in the consumer sample.

For each consumer selected to take part in the test, the recruitment questionnaire shall be filled out completely.

The required recruitment details may be verified.

An example of a recruitment questionnaire is given in Annex A.

The recruitment questionnaire shall be preserved, or information arising therefrom shall be stored to ensure the traceability of recruitment parameters of the study.

5.3.2 Representativeness of the consumer sample

Representativeness with respect to the target population is the foremost quality required of a sample. Defining this population generally results from a concerted action between the study-commissioning party who proposes a target population and the laboratory which evaluates the feasibility of implementing this proposal.

To select consumers who are representative of the target population, criteria such as the following shall be used:

— frequency of use or consumption of the product (this criterion is often considered fundamental on account of its supposed influence on the discriminating power of the study);

— age, gender, socio-economic status, occupation, geographical location;

— brand of product usually used or consumed;

— place of purchase of the product.

5.4 Sample segmentation

There are two ways in which the consumer sample can be segmented:

a) Segmentation at the outset in order to obtain answers to questions such as the following:

   1) Do women respond in the same way as men?

   2) Are responses different for different age groups?

   3) Do responses depend on consumption habits?

In this case, each subgroup is analysed separately and the results are compared.

b) Segmentation may be derived from the data collected if there is evidence (such as multimodality) of different subgroups.

In Annex B, guidance is given as to how to deal with segmentation data.

6 Size of the consumer sample

6.1 Decisions of measurement

The numerical precision of any averages or estimates obtained from the results increases with sample size. However, increased precision of measurement does not in itself ensure the relevance and validity of conclusions reached. Appropriate choices of consumers and procedures are also essential.

Annex D contains some considerations on the influence of the consumer sample size on the precision of measurement.

Precision also depends on the variability of the consumers in their responses to a single product. The larger the dispersion, the larger the sample size shall be to achieve a given precision. Knowledge of this parameter comes from the experience of both the laboratory and the commissioning party.

The precision of the results depends also on the test procedure. For example, all other things being equal, the difference between two products is established more precisely if each consumer assesses both products rather than if they are judged by different consumers.

The required precision depends on the purpose of the test. With all other things being equal, it is influenced by the following:

— the smallest difference between two mean pleasure ratings that is considered large enough to be useful when the purpose is to demonstrate a difference (the smaller it is, the larger the sample size has to be);

— the greatest difference between two mean pleasure ratings that is considered small enough to be ignored, when the purpose is to demonstrate equivalence (the smaller it is, the larger the sample size has to be);

— the risk, α, that the commissioning party is willing to accept of mistakenly concluding that there is a difference (the smaller the risk, α, the larger the sample size has to be);

— the power, P, that the commissioning party desires for the test. ($P = 1 - \beta$, where β is the risk of mistakenly concluding that there is no difference); the larger the desired power, the larger the sample size has to be);

— the directionality of the intended statistical test. If the commissioning party is interested only in reassurance that product B is not inferior to product A, a directional test is appropriate and the power of the test is greater than if the conclusions that B is inferior to A or that B is superior to A would be of equal interest.

6.2 Determining the sample size

Determining the sample size is a critical step in the development of a test and demands the assistance of a statistician. Annex E gives expressions (accompanied by examples) which can be used for the calculations for either a difference test or a similarity test based on ratings. Annex F does the same for tests based on preferences.

In all cases, the size of a consumer sample conforming to this International Standard shall never be less than 60.

Where the laboratory does not have the information required by the methods described in Annex E or Annex F, this International Standard fixes the minimum sample size at 100.

Where the laboratory uses the methods described in Annex E or Annex F, the sample size is given by the calculation result. However, this number

— can be lowered to 100 when the calculated number of consumers is over 100, but the laboratory and/or the commissioning party does not have the means to question a larger number of consumers. If this occurs, the laboratory shall calculate the estimated power of the test for a number of consumers equal to 100 and obtain the agreement of the commissioning party that this lower power is acceptable, and

— shall be increased to 60 when the calculated number of consumers is below 60.

The values of 100 and 60 suppose that the commissioning party does not hope for conclusions segmented according to consumer characteristics such as age or gender. Where a segmentation is decided on prior to the test, the minimum number of consumers per segment is 60.

6.3 Working with subgroups

6.3.1 General

Subgroups of the consumer sample can be formed before collecting test data or by making use of test data. For any of the options listed below, statistical advice should be obtained at the design stage.

6.3.2 Separate tests using pre-formed consumer groups

The commissioning party may wish to study separate subgroups formed by taking account of consumer characteristics (such as gender or age), which are known before conducting the test. One option is to analyse their results independently. If this is done, each subgroup shall consist of at least 60 consumers.

6.3.3 Unified test using pre-formed consumer groups

The procedure in 6.3.2 is less efficient than balancing presentations over the whole consumer sample, followed by an integrated analysis of all the results that takes account of subgroup membership. The same degree of confidence will normally be achieved with fewer than 60 consumers in each subgroup. This style of analysis is at its most efficient when all subgroups are of the same, or of very similar, size.

6.3.4 Segmentation of the consumers

The membership of subgroups may be an outcome derived from the results of a test on a large consumer sample. The subgroups are then classes resulting from the evaluation of all the data on hand (measurement results and/or socio-demographic data). This segmentation serves to describe the consumer sample and can be used for developing new hypotheses. The number of groups formed, their characteristics, and their sizes all depend on the amount and nature of the results obtained.

6.4 Number of product samples and experimental design

The number of samples and the experimental design affect the size of the consumer sample. If a complete plan is used for product presentation, the number of consumers needed for the test is equal to the number of responses per product to be obtained. If an incomplete plan is used, the number of consumers must be larger.

6.5 Safety margin in recruitment

The number of responses actually obtained is usually smaller than the number of consumers recruited since some consumers selected for the sample might be absent and some who participate might not use the questionnaires correctly. To compensate for this, a safety margin should be included when calculating the number of consumers required.

7 Test area

This International Standard concerns hedonic measurements conducted within a controlled area as defined in 3.6. This is distinguished from hedonic tests conducted, for instance, at a point of sale or in consumers' own homes, which are outside the scope of this International Standard. Three examples of relevant test areas are described.

a) A permanent, sensory evaluation laboratory specifically equipped in accordance with ISO 8589 for conducting sensory analysis tests. It has a reception area for welcoming the consumers, a room for preparing the products and a room equipped with test booths, temperature control, and a ventilation system for periodic air renewal. This area provides the best product-preparation, product-presentation, and response-logging conditions.

b) A mobile sensory evaluation laboratory installed in a vehicle, which has been specifically converted for the conduct of sensory tests. It is likely to have much less space than a stationary laboratory. Zones reserved for welcoming consumers and for preparing products are generally very limited.

c) A room or rooms equipped temporarily in a "task-specific" way for assessing products that require little preparation. The area is arranged in two separate parts, one dedicated to conducting the tests and the other to preparing and coding the products. The layout of the test area allows the consumers to be physically isolated from each other. The test conditions (temperature, room lighting, ventilation, and storage and preparation of the samples) can be monitored. Unlike the stationary or mobile laboratories, the product preparation and presentation system is redefined each time the room is adapted for a specific task.

8 Products

8.1 Anonymous presentation of the products

Products to be tested shall be presented to the consumers after being made as anonymous as possible, the only means of identification being an arbitrary code (such as a three-digit random number) assigned by the laboratory. All references to a brand or to a quality mark shall be deleted or masked, except in the two following cases:

a) when deletion or masking is impossible;

b) when the objective is to determine the impact of the brand or quality mark.

NOTE In this International Standard, by quality mark is meant any mark which the consumer might interpret to influence expectation. For example, the type of packaging can constitute a quality mark.

8.2 Preparation and presentation of the products

8.2.1 General

The products to be evaluated shall be prepared according to the procedures recommended by the manufacturers or, when these procedures are not known, according to a method formalized following preliminary tests. Allowing for duly justified exceptions, the preparation conditions shall approximate the usual conditions of use and consumption.

The same rule applies to the presentation of the products (temperature, associated products, tableware and serving method).

WARNING — Any product is likely to vary from batch to batch and as a consequence of storage. This variation may be of interest to the commissioning party and in some cases will be the focus of the investigation. In all other cases, variation should be minimized by using products that all come from the same batch and have had identical storage durations and conditions.

8.2.2 Quantity of products presented

The amount of product presented to an consumer shall correspond to the portion of product normally consumed or to a portion specified by the commissioning party. The consumer shall be informed of the minimum amount of the sample to be tested and, if necessary, also of the maximum amount to be tested or consumed. Care shall be taken to ensure that the test portion is neither too small (risking only a fleeting, initial impression) nor too great (risking repletion or aversion).

The quantity presented may need to be adjusted to be less than the quantity usually consumed in cases where the consumers consume several products of the same family. However, the quantity presented shall always be sufficient to avoid the appraisal being conducted on too small a quantity and thereby provoking only an initial, fleeting sensory impression.

When presenting a ready-made dish, the portion presented to each consumer shall proportionally include all of the constituents of the product.

8.2.3 Form of the presented products

Sometimes, there are two options for presenting a product: the product as purchased and the product in a form ready to be consumed.

EXAMPLE 1 Examples include bread (whole or sliced), sausage (whole or sliced), cheese (whole or cut), a pizza (whole or in portions), instant coffee (in a cup prior to the water being poured in or after the water has been poured in).

The laboratory and the commissioning party may consider that both forms have the same objective: to best approximate the pleasure elicited by the product. There is no need to know if any halo effect is provoked by one of the forms on the other. If so, the two forms are presented simultaneously (whole/pieces) or successively (instant coffee). The consumer gives a single appraisal rating for both presentation forms, without it having been specified to the consumers that the product is presented in two forms. This approach corresponds to the normal behaviour of a consumer who, in everyday life, does not need to split up his/her perceptions.

EXAMPLE 2 The wording of the instructions and of the question might be as follows: You are provided with a sausage bearing the code 148. Look at it, sniff it, consume a few pieces of it and then report what you think of it using the response scale below.

If the laboratory and the commissioning party agree that both forms contribute to the pleasure given by the product, the two forms need not be treated as different products. Instead, the two forms can be presented together or successively and receive a single appraisal rating.

NOTE This International Standard does not deal with the case where the laboratory and the commissioning party consider it is necessary to know the expectation effect created by the visual information. This situation falls within the scope of special experimental designs involving specific concepts and analysis methods.

9 Procedures

9.1 Two groups of hedonic tests

Two main groups of hedonic tests set different tasks to the consumer:

a) **acceptability tests**, used to measure the intensity of enjoyment at consumption. In this International Standard, the only type of acceptability test described is the **rating test**.

b) **preference tests**, used to measure the order of liking for different products (e.g. “Which sample do you like best?” or “Please rank the products presented from least liked to most liked.”). The information obtained in a test assessing momentary preference is of a relative nature. It does not say anything about the acceptability of the products because it is possible for one product to be preferred over another without either of them being acceptable. Preference tests are subdivided as follows:

   — **paired comparison tests**, when comparing two samples;

— **ranking tests**, when dealing with more than two samples.

9.2 Rating tests

9.2.1 General

Rating tests are distinguished

— by the response scale: the latter can be structured or non structured, numerical, semantic, or pictorial, as specified in ISO 4121, and

— by the method of presentation of the products.

When the number of products is two or more, three product presentation methods are possible:

a) **strict monadic presentation (single assessment)**, where each consumer evaluates a single product;

b) **sequential monadic presentation (incomplete or complete plan)**, where several products are evaluated by an consumer in one or several sessions. The consumer receives one product at a time and no information is given about products already assessed or responses made to them; it shall be ensured that it is impossible for the consumer to revert to the assessment of an earlier product;

c) **comparative presentation (incomplete or complete plan)**, where several products are presented simultaneously to the consumer and consumers are allowed to review the ratings they have given to other products.

The most common form of presentation is sequential monadic. It is less expensive than the first, but it demands mastering fully how to assign the products to the consumers (see Annex C). Comparative presentation is little used because it tends to exaggerate differences between products and also makes comparison between studies difficult when the test conditions are not strictly identical.

9.2.2 Sample presentation plans

The samples shall be presented to consumers according to the presentation plan (see 10.2).

9.3 Paired comparison tests

9.3.1 General

Paired comparison tests shall be monitored as specified in ISO 5495.

The principle of the paired comparison test is to compare products presented in pairs. The consumer judges the samples in a specified order and states which of them is preferred.

A paired comparison test on preference between two products can be carried out in two different ways, namely with forced choice (option A below) or with a "no preference" response permitted (option B below).

a) **Option A**: The consumers respond to a closed question of the following type:

"Taste the two products presented to you, beginning with the product on the left, and then mark with a cross the code of the product you prefer."

Product xxx	Product yyy
□	□

Here, the consumers shall decide in favour of one or other of the products.

b) **Option B**: The consumers respond to a closed question of the following type:

"Taste the two products presented to you, beginning with the product on the left, and then mark with a cross the code of the product you prefer. If products are similarly liked for you, mark with cross « no preference » "

Product xxx	No preference	Product yyy
□	□	□

In a test with only one pair of samples, option A has the advantage of greater potential for discrimination than option B.

The questionnaires shall not contain any questions capable of influencing consumers to decide in favour of one or other of the products or to judge the products in respect of particular attributes. For instance, if they are asked to comment on the flavour of the products they are likely to concentrate on this aspect in particular when assessing acceptability.

It is, however, permitted to ask an open-ended question at the end of the test, giving consumers an opportunity to comment on what was liked or disliked about the products.

9.3.2 Sample presentation plans

The order of presentation of the samples A and B shall be balanced by presenting the pairs AB and BA equally often and allocating them to the consumers at random.

9.4 Ranking test

9.4.1 General

Ranking tests shall be conducted as specified in ISO 8587.

The principle of the ranking test is to present several test samples simultaneously and ask the consumer to sort them in order of acceptability.

The rank-order test is rather demanding for the consumer and its complexity increases with the number of samples to be compared since the consumer has to test the products several times in order to sort them into an appropriate order.

It is permissible to specify the order of testing the samples (usually from left to right) only for the consumer's initial evaluation of each sample. After that, the consumer is free to try the products in any order. Moreover, the consumer's method of deciding on the rank order shall not be specified.

The rank-order test can be performed either as a forced-choice test or by allowing the consumer to classify two or more products as equal. However, analysis of the second of these options is outside the scope of this Standard.

9.4.2 Questionnaires

Consumers receive instructions of the following type:

"Taste the four products, one after another, beginning with the product farthest to the left and ending with the product farthest to the right. You may then taste any of the products again as often as you wish

until you have made up your mind. Then, please write down the codes of the products in order from the one you like least to the one you like best"

Liked least			Liked best

9.5 Additional questions

9.5.1 General

In addition to any of these test procedures, the commissioning party might wish to receive additional information to assist in understanding the overall appraisal of the products being assessed or the characteristics of the consumers. For example, with recruitment from a database a laboratory may wish to check that all the consumers did indeed come from the target population. To do so, a question can be asked about a particular consumption habit.

9.5.2 Questions about consumers

The choice of questions and of the stage at which they are asked demands close attention by the laboratory.

If the question is asked after the session, beware that the blind tests that have just been carried out can affect the responses. For example, following a test involving a tomato sauce accompanied by rice, the consumers may be more inclined to declare that sauce with rice is a dish that they do eat.

On the other hand, questions asked before the session can predispose consumers to consider the products in a particular way or to pay particular attention to some aspects.

9.5.3 Questions about products

Open-ended or closed-ended questions may be used for this purpose. Such questions shall be asked after all questions relating to overall appraisal and on a sheet (or screen) different from those concerning the overall appraisal to reduce the risk of consumers answering in ways intended to make their responses appear consistent.

The number of such questions should be as small as possible.

NOTE 1 Additional questions are not intended to call into question the results of the main appraisal. If there is any discrepancy between the overall appraisal and reported intention to consume, the reason for this is intended to be sought, possibly by means of an additional study.

NOTE 2 Additional questions often provide little information. The head of the sensory evaluation laboratory is intended to inform the commissioning party that there exist methods (such as preference mapping or conjoint analysis) which have been developed specifically for combining hedonic data and sensory profile data and for determining why consumers like or dislike a product.

9.6 Instructions for the consumers

Instructions may be given in writing or orally. Their importance for the relevance of the results shall not be underestimated. Particular attention shall be paid to them and they shall form the subject of a specific paragraph in the study report. They shall focus on the following:

a) the minimum and maximum quantities that each consumer is to assess before giving an appraisal. These quantities depend on the type of product, the number of products evaluated during the course of the session and the wishes of the commissioning party. The laboratory shall not only give instructions concerning the minimum and maximum quantities, but also define an operational protocol to ensure that the instructions are followed;

b) what is to be done with the questionnaire once completed. When conducting a sequential monadic evaluation, the laboratory shall use a system that prevents consumers from referring to responses to a previously evaluated product;

c) the duration of the breaks and any activity during the breaks (cleansing of the sensory receptors using bread, apple or water, leaving the evaluation room, reading a magazine and so on);

d) the format and meaning of the questions asked.

NOTE The wording of questions is never neutral. For example, to suggest to the consumers that they may explain their general appraisal of the product by appearance, odour, taste, aroma or texture-related reasons leads the consumers to more rationalization of the responses than the presence of a simple heading: *Comment(s) if any*.

The person conducting the test usually gives the consumers at the start of the session information about the number of products, the quantities to be consumed and the questions to be answered. He/she can also give other information in order to render the evaluation less impersonal or in order to meet other objectives of the study. As such information can create expectations in the consumers, the way it is expressed shall be considered carefully and be included verbatim in the study report. Where the test consists of several sessions, the phrasing shall be strictly the same in every session.

10 Organization of test sessions

10.1 General

Test sessions are organized to achieve optimal balance among a number of objectives whose requirements conflict with each other to some extent. These include the following:

— ensuring that comparisons among products are valid, being derived from equivalent consumers;

— ensuring that comparisons among products are valid, being subject to minimal position and sequential biases;

— ensuring that test conditions are as representative as possible of normal consumption behaviour;

— ensuring that consumers are not overloaded, which would result in more random responses.

To accommodate these different objectives, there are three ways in which consumers may be related to products:

a) **complete plan**: each consumer judges all products;

b) **incomplete plan**: each consumer judges a subgroup of products;

c) **single assessment**: each consumer judges only one product.

10.2 Plans

10.2.1 Complete plan

Evaluating all the products ensures that all the products are evaluated by the same consumers (so there is no problem of equivalence), but carries the greatest risk of overloading the consumers.

There is also a danger of order and transfer effects in which the position in a sequence in which a product is presented affects the mean and/or the variability of its assessment.

The order of presentation of the products to each consumer plays a crucial role. One way is to use a **strictly** randomized order, randomizing independently for each consumer. Alternatively, a systematic presentation plan can be adopted to counterbalance order and transfer effects. However, if transfer effects are an issue, then use of washout periods or other methods to reduce the likelihood of transfer effects will generally be most important

10.2.2 Incomplete plan

Order and transfer effects and the measures to deal with them are similar to complete plans. There is less danger of overloading the consumers but the disadvantages are possible context effects and additional complexity in the design and analysis of the trial. Special designs exist for balanced incomplete plans (see Reference [9]).

10.2.3 Single assessment

To eliminate order transfer and context effects completely, a separate consumer sample can be assembled for each product. This approach leads to the simplest procedures and simplest analyses; however, it is imperative for the samples to be assembled with particular care to ensure their equivalence and their representativeness of the target population.

Moreover, since this design makes it impossible for the analysis to take account of individual differences, the precision of estimates is poorer than the other designs. Equivalently, with other things being equal the consumer samples must be larger to achieve the same degree of confidence and precision in the results.

10.3 Number of products evaluated in a single session

10.3.1 General

The number of products that can be evaluated effectively within a single session depends on the following:

— the nature of the products being presented. Products do not all provide the same feeling of satiety; some products (aggressive ones) can modify the sensory receptors; others (such as those containing alcohol) can affect cortical integration; some are composed of a single element, others (such as ready-made dishes) of various elements;

— the duration of the session. A long-lasting laboratory session can be divided into sub-sessions; but consumers recruited without an appointment usually have very little time available;

— the quantity to be consumed of each product;

— the senses used to evaluate the product (many more can be evaluated by appearance or tactile feel than by taste);

— the number of questions asked, particularly if any require a new presentation of the products.

Consequently, this International Standard does not set specific limits. It only requires that the laboratory be capable of justifying its choices to the commissioning party.

10.3.2 Too many products for a single session

Where the number of products is too large for all to be evaluated in a single session, two approaches are possible:

a) organize several test sessions, each consumer taking part in all of the sessions (complete plan);

b) organize a single session, each consumer receiving only some of the products (incomplete plan or single assessment).

10.3.3 Several sessions

If it is possible to arrange several sessions, all the consumers can evaluate all of the products over the course of the whole study. Ideally, all the products are presented in each session, but with the consumers in a particular session not all receiving the same products.

If constraints related to the preparation of the products or the quantities available make this impossible, the products are randomly distributed over the sessions. All the consumers receive the same subset of

products during a session, but in different orders. This choice will normally require a more sophisticated statistical analysis.

It is desirable to present, at the beginning of each session, an identical product (known as the "warm-up product") to all the consumers so that in every session all the consumers have the same reference.

Two sessions using the same consumers shall not occur on the same day.

10.3.4 Single session

A balanced incomplete block (BIB) design, allows each session to use fewer than the total number of products but aims to compare each pair of products with equal precision. It may also be desired to balance out order and transfer effects over the study as a whole. This choice will normally require a larger number of consumers and a more sophisticated analysis (see Reference [9]). Statistical advice should be sought at the design stage if this option is considered.

The balanced incomplete block design method shall be conducted as specified in ISO 29842.

10.4 Nature of the products evaluated within a session

Products from different families may be evaluated in the same session if the following conditions are met:

— the combination of products corresponds to well-established food practices;

— the target consumer is the same for all the families.

Consumers known in advance not to belong to the target population for a product will normally not be selected to evaluate it.

The laboratory shall ensure the confidentiality of the evaluations, irrespective of the persons attending the evaluation. A commissioning party shall be able, if so wished, to attend a session without knowing the name of the other commissioning parties involved in the session.

10.5 Time of the session

It is recommended to arrange the session at a time corresponding to the usual consumption of the products so the consumers are under conditions close to what they do in actual practice.

10.6 Repetition of a product within a test

If the evaluation of a product is to be repeated in the same test, the number of products is considered as being increased by one. The test plan then includes, instead of product A, products A1 and A2, which are treated as two different products, both for determining order of presentation and for analysis of the results.

10.7 Pre-test

Prior to the main test, a pre-test can be performed with only a few consumers in order

— to try out the questionnaires to check, for example, that the relevant product attributes are understood and perceived and that the response scales are appropriate,

— to try out arrangements for administering the test, and

— to ensure that test duration and the numbers of questions and of samples are appropriate.

Consumers for the pre-test shall be drawn from the same consumer population as the main test but shall not be part of the main test's consumer sample.

11 Analysis of the results

11.1 General

Every data analysis shall begin by checking the internal consistency of the data and calculating elementary descriptive statistics. There shall be at least one dispersion parameter (standard deviation, range of variation, etc.) for every location parameter (mean, median, etc.) appearing in the report.

Graphical methods are very useful. For each product sample, a histogram of the assessments given for it should be created. This shows whether or not the distribution is unimodal and whether or not there are large differences between samples.

To obtain a visual representation of the agreements and disagreements among consumers, it is possible to use Principal Components Analysis (PCA) or Ascending Hierarchical Classification (AHC).

Additional inferential statistics may be required to answer questions such as: "Did the consumer sample perceive a significant difference between the products they assessed?"

It is possible to use parametric or non-parametric methods, with non-parametric methods being especially appropriate when normal distributions are not found.

11.2 Inferential analysis

Inferential analysis extrapolates results obtained from a sample of consumers to the population it is drawn from. The method to be used depends on the test's objective and on the number of products tested. If some of the products have a special status, such as being reference samples, this too affects the analysis.

Using a presentation plan designed to balance biases does not remove the biases but only ensures that no product has been unduly favoured or penalized by the order of presentation. However, an inferential analysis can take advantage of a systematic plan to improve the way that biases are estimated and allowed for and thus give increased confidence in the main conclusions. However, this analysis in general is neither easy nor is it straightforward (see References[4] and[5]).

Standard computer packages provide all the analyses likely to be required, but the selection and application of an appropriate analysis requires statistical advice.

11.3 Ratings (see 9.2)

With rating tests, some form of analysis of variance (ANOVA) is likely to be useful.

Annex G describes several different data analysis procedures.

11.4 Paired comparisons (9.3)

If the consumers are allowed to report "no preference" in a paired comparison test (option B), one way to analyse the data is to report the number of responses of "no preference" and carry out a significance test that uses only the preferences that were expressed. For the reduced sample size, the same test methods can be used as in the case of forced choice. These methods are given in ISO 5495.

11.5 Ranking (9.4)

Various analyses addressing different aspects of the data, which are described in ISO 8587, may be used (see also Annex H).

It is possible to use parametric as well as non-parametric methods whereas non-parametric methods are practical in cases in which normality of the responses is not plausible. The methods given in Table 1 may be used.

Table 1 — Methods of data analysis

Situation		Method	
		Non-parametric	Parametric
Two products	Each of the products was assessed by a different group of consumers.	Mann-Whitney's *U*-test (also known as Wilcoxon-Mann-Whitney test)	*t*-test for independent samples
	Each of the two products was assessed by all the consumers.	Wilcoxon signed-rank test	*t*-test for paired samples
More than two products	Each product was assessed by a different consumer group.	Kruskal-Wallis test	Analysis of variance (ANOVA)
	Every consumer has assessed at least two of these products and, ideally, all of them.	Friedmann test (ANOVA by ranks), if necessary in a variant for incomplete blocks	Analysis of variance (ANOVA) with consumer effect and product effect

12 Study report

12.1 General

The first page of the report can be a brief executive summary of the results and the most important conclusions.

The report shall be drawn up so as to be intelligible without reference to the study proposal. In addition, it is recommended that every table and diagram of results be numbered and have a title and sufficient other explanation to allow it to be understood on its own.

Recommendations based on the test and requested by the commissioning party may be set out in a section identified as "recommendations" or in a separate document.

In addition to the results, the report can include the details listed below:

— title of the study and references;

— dates of conducting the study;

— date of the report;

— identification of the commissioning party;

— complete identification of the laboratory and of the person in charge of the study (together with identification of subcontractors, if any);

— objective of the test (summarizing the study proposal);

— procedures for the commissioning party to get access to the raw data;

— uniform identification of the report and of every page, including the total number of pages;

— a reference to this International Standard, i.e. ISO 11136:2014.

12.2 Products

The product information shall contain the following:

— a description of the products, such as their full or partial composition, if known by the testing laboratories. (The description may be accompanied by a photograph of each product.);

— all information supplied by the commissioning party about the samples submitted for the test;

— date of production of the samples, ‘best before’ or ‘use before’ date, the manufacturer’s lot number or batch number;

— date of receipt at the laboratory; the temperature, duration and conditions of storage (especially in the case of fresh and frozen products);

— the sampling procedure when sampling has been carried out by the service provider. (If control or reference samples are selected, the method of doing so shall be described in an extremely precise manner.);

— the method of preparing the products for assessment;

— the temperature of the products when presented to the consumers;

— the quantity presented to each consumer and the instructions concerning the minimum and maximum quantities to be consumed, including a report of any failure to comply with these instructions;

— associated products or disposals;

— all other products evaluated during the session and their order of presentation in the session.

12.3 Test procedure

The test procedure details shall contain the following:

— the test procedure used, making reference to the relevant standard or, in the case of non-standardized procedures, giving a full description;

— the questions asked and response scales used;

— the data acquisition method;

— operating procedures for the study;

— the number of sessions or subsessions;

— the dates, times and durations of sessions;

— ambient conditions, such as environment, temperature or lighting;

— the tableware used;

— the instructions and information given to the consumers and the method of their delivery, such as on paper, on a screen or orally.

12.4 Consumers

12.4.1 General

Information on consumers shall contain the following:

— a description of the target consumer population;

— a description of the consumer sample, including its size and the numbers in each of any categories specified by the commissioning party;

— a statement of whether recruitment was task-specific or from a consumer pool.

12.4.2 Task-specific recruitment

Information on the recruitment shall contain the following:

— the place or method of recruitment;

— the procedures followed;

— the method of selecting appropriate consumers.

12.4.3 Recruitment from an access panel

If the recruitment is made from a consumer pool the report shall contain:

— a description of the frequency of participation in consumer tests by the consumer sample;

— a table showing all product families that have previously been assessed by members of the pool;

— a table showing the frequency of participation in previous tests involving the product under evaluation or its product family.

All the information required in this subclause relates to the preceding 12 months. If the frequencies are zero for all consumers, this fact is all that needs to be reported.

12.5 Results

The results of the test shall contain the following:

— indication of provision of the raw data and description of their presentation. Summary of the raw data by the mean of graphics or table;

— numerical summaries of the results, including the precision of any estimates or averages;

— methods used to analyse and interpret the results;

— statistical inferences drawn from the results;

— conclusions drawn from the results with reference to the objectives of the study.

12.6 Annexes to the report

The following additional information may be attached as annexes:

— a specimen of the questionnaire or response form used for the test;

— a specimen of any questionnaire used for recruiting the consumers;

— verbatim transcripts of responses to any open-ended questions.

Annex A
(informative)

Example of a recruitment questionnaire

No.	Contact interview project XY 04/05		Filter	Code
1	Hello. We're conducting a survey today, and I'd like you to answer a few questions. Have you ever taken part in an interview for market research purposes before?			
	Yes	1	▶ 2	
	No	2	▶ 3	
2	When was the last time you took part in a survey?		▶ 3	
	Information to interviewer: IF IT WAS LESS THAN HALF A YEAR AGO ▶ END; **OTHERWISE, CONTINUE WITH QUESTION 3**			
3	Do you or any of your friends or relatives work in one of the following industries? ***Information to interviewer: PRESENT THE "INDUSTRIES" LIST***			
	Industry 1	1	▶ 4	
	Industry 2	2	▶ 4	
	Market research	3	END	
	Industry 3	4	▶ 4	
	Industry 4	5	END	
	Industry 5	6	▶ 4	
	Advertising	7	END	
	Journalism	8	END	
	Marketing	9	END	
	None of the industries listed above	10	▶ 4	

No.	Contact interview project XY	04/05	Filter	Code
4	Might I ask your age?		QUOT.	
	20 years to 39 years	1	▶ 5	
	40 years to 60 years	2	▶ 5	
	Your sex:			
	Male	1	▶ 5	
	Female	2	▶ 5	
5	I have on this list different food products for daily use. Which of these food products have you bought lately? ***Information to interviewer: PRESENT THE "FOOD PRODUCTS" LIST***			
	Food product 1	1	▶ 6	
	Food product 2	2		
	Food product 3	3		
	Food product 4	4		
	Food product 5	5		
	Food product 6	6		
	Food product 7	7		
	Food product 8	8		
	Information to interviewer: IF "3" IS NOT MENTIONED ▶ ***END***			
6	You just said that you also use food product No. 3. Approximately how often do you use food product No. 3? ***Information to interviewer:*** **PRESENT THE "FREQUENCY OF USAGE" LIST**			
	More than once per day	1	▶ 7	
	Daily	2		
	More than once per week	3		
	About once a week	4		
	Approximately once every fortnight	5		
	About once a month	6	▶ END	
	About once every three months	7		
	Less frequent than that	8		

No.	Contact interview project XY 04/05	Filter	Code
7	Which brands of food product No. 3 do you buy for your household regularly, i.e. at least once a month? ***Information to interviewer: ASK FOR THE MANUFACTURER AND THE PRODUCT!IF ONLY THE MANUFACTURER IS MENTIONED, ASK AGAIN!*** Brand/Manufacturer ▶ Product 1) _____ ▶ 1) _____ 2) _____ ▶ 2) _____ 3) _____ ▶ 3) _____ 4) _____ ▶ 4) _____ 5) _____ ▶ 5) _____	▶ 8A	
8A	I have different brands of food product No. 3 on this list. Which of these brands do you use in your household, even if only occasionally?		
	Information to interviewer: PRESENT THE "BRANDS" LIST AND CIRCLE 8A IN THE SCHEME!	▶ 8B	
8B	Would you, please, tell me how often you use each of these brands?		
	Information to interviewer: PRESENT THE "FREQUENCY OF USAGE" LIST AND, ONLY IF CIRCLED FOR QUESTION 8A, MAKE AN ENTRY IN THE SCHEME UNDER QUESTION 8B!	▶ 8C	
8C	Which of these brands do you use most frequently?		
	Information to interviewer: ONLY ONE CHOICE IS POSSIBLE; CIRCLE YOUR CHOICE IN THE SCHEME UNDER QUESTION 8C)!		
	8A 8B 8C Brand 1 1 _____ 1 Brand 2 2 _____ 2 Brand 3 3 _____ 3 Brand 4 4 _____ 4 Brand 5 5 _____ 5 Brand 6 6 _____ 6 Brand 7 7 _____ 7 Brand 8 8 _____ 8 Brand 9 9 _____ 9 Brand 10 10 _____ 10 Brand 11 11 _____ 11	▶ 9	
	Other information, namely _____ 12 _____ 12		
9	Do you have a food intolerance?		
	Yes 1	▶ END	
	No 2	▶ 10	

No.	Contact interview project XY	04/05	Filter	Code
10	Do you suffer from diabetes?			
	Yes	1	▶ END	
	No	2	OK	

Annex B
(informative)

Methods for evaluating the data of segmentation subgroups

B.1 General

Depending on the aim of the study, it can be useful to compare subgroups of the target population. For instance, it might be of interest to determine whether there are any differences between sub-groups and, if so, at which levels of acceptability they occur.

Based on the entire consumer sample, two ways of segmentation are possible:

a) **Population subgroups**: the subgroups are sample classes for which hypotheses of interest shall be verified or falsified. The classes should be specified before recruiting the consumers. For each of the specified criteria and for each of the sample classes, the size shall be specified so that the individual classes are sufficient (consisting of at least 60 consumers) to obtain the desired accuracy of measurement when evaluating the data.

   NOTE Depending on the selection criteria the classes do not necessarily have to be of equal size.

b) **Segmentation of the consumers**: the subgroups shall be formed from the evaluation of all the data on hand (measurement results and/or socio-demographic data). This segmentation serves to describe the consumer sample and may be used for developing new hypotheses. Subsequent segmentation is sensible only if the consumer sample was of sufficient size.

In both cases, the necessary number of consumers in the study varies according to the number of segments selected or to be expected, respectively, and according to the variance to be assumed within the individual segments.

B.2 Population subgroups

The comparison of subgroups is used to gain information with regard to questions of the following type.

— "Do women have the same preferences as men?"

— "Are the preferences different depending on the age group?"

— "Are the preferences different depending on the country?"

— "Are the preferences different depending on the consumption habits?"

In case that analysis of variance is used and each of the consumers has assessed each of the samples the following model applies:

Response = group + consumer (group) + sample × group + error

The subgroups are tested against the factor consumer nested in group, whereas the other factors are tested with respect to the error [it is equal to the interaction sample × consumer (group)].

It is possible to include the order effect and, if necessary, also the transfer effect in this model. If the order effect is included, the model appears as follows:

Response = group + consumer(group) + sample + order + sample × group + order × group + sample × order + error

By means of this analysis, it can be determined whether the different groups position the samples differently (interaction sample × group) when the order effect is of the same size for all the groups (interaction order × group) and for all the samples (interaction sample × order).

If there is a significant difference between the groups with respect to the samples (interaction sample × group), each of the subgroups is evaluated separately.

B.3 Segmentation of the consumers

Subsequent segmentation aims to explain the preference types detected, if any.

This is particularly of use if, in checking the histograms, bi-modalities were detected.

Internal preference mapping affords the possibility of analysing data without taking into account the means of samples over the consumers. This is, in fact, a principal component analysis of the grades matrix in which the lines represent the products and the columns the consumers.

NOTE Principal components analysis: a method for analysing multivariate data with the aim of expressing their variability in a minimum number of principal components or linear combinations of the original and partly correlated variables.

The “simultaneous” representation enables the products to be superposed to the consumers.

However, the information provided by the first few components is valid only if these components explain a high percentage of the original variation. This procedure is, in general, applicable only when using complete plans (see Reference [11]). However, a procedure has been proposed for incomplete test plans as well (see Reference [10]).

Thus, when examining the first components of the principal component analysis visually, consumer groups (clusters) can be discerned, if present. In order to identify consumer clusters, classification methods can be used.

Annex C
(informative)

Plans for assigning products to consumers

A completely valid and even recommendable proposal for assigning products to consumers is a randomized complete block design. This would mean that each consumer tests each product exactly once. The order of presentation is randomized independently for each consumer. The advantage of this proposal is that it validates the analysis in the simple block model.

There are, however, reasons why in many cases the experimenter might want to use another, more complicated, design.

— Incomplete block designs: If the number of samples is too large, it might not be possible for each consumer to evaluate each product. The number of samples, which can be tested by a consumer, depends on the nature of the products. In such a case, it is preferable to use a **balanced incomplete block design**, see ISO 29842. Again, the order of the products will normally be randomized, independently for each consumer

— Order effects: Often, experimenters are worried about the presence of order effects. It is well known that products which are tasted earlier will generally be rated better than products which come later. One of the reasons the complete block design and the balanced incomplete block design require randomization is this **order effect**: Randomization of the order of presentation avoids a possible bias due to order effects. Randomization guarantees that each product has the same chance of coming early

  If the experimenter fears that the order effects might be very strong, a valid alternative to the randomized complete block design is a generalized Latin square. Here, the order of the presentation of the products to the consumers is arranged in such a way that each product comes equally often in the first, second, etc. position.

  In the case of an incomplete block design, it is also possible to arrange the order of presentation of the products in such a way that each product comes equally often in each position. Designs of that structure are called generalized Youden designs.

  It should be noted, however, that the use of a generalized Latin square or a generalized Youden design does not validate the analysis in the simple block model. Instead, a row- column model should be used, which takes account of order effects.

— Carryover effects: In spite of the use of washout periods, it may be possible that a product may have an effect on the taste of the product which is tasted in the next period. If the experimenter is worried about **carryover effects**, it may make sense to use a so-called neighbour balanced design, see e.g. Reference [2]. Some possibilities for the analysis of neighbour balanced designs are outlined in References [2] and [9].

— Organizational matters: in some cases, there are **organizational matters** that demand a more complicated design structure. For instance, it may be the case that there are eight products to be compared, but there are only four heating places, such that in any session only four products can be presented. Situations like this require a more complicated, non-standard, design, and, correspondingly, a more complicated, non-standard, analysis. In such cases, the experimenter should seek the help of a statistician

Annex D
(informative)

Influence of the sample size on the accuracy of one mean

Every consumer group has to have a certain minimum size so that differences can become evident in the assessment of the products and, most importantly, so as to enable the mean of the assessments given by a consumer group for a product to be estimated with a certain accuracy. The same applies to the total consumer sample or to a class of samples or to a segment revealed to be existent when examining the individual differences.

The accuracy depends on the sample size and on the variability of the assessment grades over the panel.

Assuming the assessment grades given for a product to be normally distributed with a standard deviation, σ, the minimum size of the consumer sample required to estimate the mean with a desired accuracy, A_c, can be determined from Formula (D.1):

$$A_c = \frac{2\sigma}{\sqrt{n}} \tag{D.1}$$

where A_c is equal to half of the confidence interval at $P = 0{,}95$ of the mean.

For example, $A_c = 0{,}2$ means that the confidence interval of the mean comprises between $(m - 0{,}2)$ and $(m + 0{,}2)$.

Table D.1 gives the relationship between σ, n, and A_c.

Table D.1 — Sample size required to obtain a desired accuracy

Accuracy A_c	**Standard Deviation** σ		
	1,5	**2,0**	**2,5**
0,2	225	400	625
0,3	100	178	278
0,4	56	100	156
0,5	36	64	100

σ and A_c shall be expressed in the same units.

But to determine the number of consumers to be recruited, it is recommended that the β-risk be taken into account (see Annexes E and F).

NOTE Statistical power analysis should be conducted when deciding the required sample size prior to test: n will not be determined unless effect size (ES) has been fixed (Cohen[8]).

Annex E
(informative)

Minimum sample size for tests using rating

E.1 General

A sensible requirement for the sample size of studies with consumers would be to get at least 60 assessments, and preferably around 100 assessments, per product. In the case that each consumer tests just one product; this would require $n = 100 \times p$ consumers, where p is the number of products in the study. If, however, each consumer tests k products, this rule of thumb would mean that $n = 100 \times (p/k)$ consumers are needed.

But this rule of thumb is poor. There are more sophisticated approaches. If the experimenter wants to use them, he needs to answer a number of questions.

— What is the aim of the study? Demonstrate that there are some differences between one product and one reference value or between two or more products? Or show that there is a similarity between one product and one reference or between two or more products?

— How many products are there in the study? One product? Two products? Three products or more?

— How many products will be evaluated by each consumer?

— What is the chosen α-risk?

— What is the desired power, P, or the β-risk for the test? [or: What is the acceptable β-risk? ($P = 1 - \beta$-risk)]

— What is the value of the difference δ to be detected for a difference test? What is the unacceptable value of the difference, Δ, for a similarity test?

— What is the within-product(s) dispersion of the ratings?

— Is the alternative hypothesis one-tailed or two-tailed?

— What is the relationship between the ratings attributed to various products by the consumers when the consumers rate several products?

This annex might not deal with all potential cases. It only deals with the assays composed of two products. For other cases, more simple (with only one product) or more complicated (with three products or more), the reader should seek advice from a statistician or use appropriate software. The reader may also consult AFNOR NF V09-500:2012[2].

This Annex deals with four cases: two cases for a difference test and two cases for a similariy test.

E.2 Case 1: Aiming to demonstrate a difference with assessor rating only product A or product B

The number, n, of consumers for each product is given by Formula (E.1):

$$n = 2 \times \left[\frac{\left(u_{\alpha/2} + u_{\beta} \right) \times \sigma_{\text{within products}}}{\delta} \right]^2 \quad \text{(E.1)}$$

where

$u_{\alpha/2}$ is the $(1 - \alpha/2)$ quantile of the normal distribution for the α-risk for a two tailed alternative hypothesis ; it is equal to 1,960 for an α-risk = 5 %. For an one tailed alternative hypothesis, $u_{\alpha/2}$ is replaced by u_{α} ; for an α-risk = 5 %; it is equal to 1,645;

u_{β} is the $(1 - \beta)$- quantile of the normal distribution for the tolerated β (desired power $P = 1\text{-}\beta$);

$\sigma_{\text{within products}}$ is the within products standard deviation; standard deviation is assumed to be homogeneous for the two products;

δ is the difference to be detected between the two products.

NOTE $\sigma_{\text{within products}}$ and δ are expressed in the same scale, for example the 1 to 9 scale. The 1 to 9 categorical scale is supposed to lead to ratings approximately according to a normal distribution.

EXAMPLE For the following conditions

α	$u_{\alpha/2}$	B	$(P = 1\text{-}\beta)$	u_{β}	$\sigma_{\text{within products}}$	δ
0,05	1,960	0,10	(0,90)	1,282	1,50	0,50

Formula (E.1) leads to $n = 189{,}2$, use 190 consumers per product and 380 consumers for the two products.

NOTE For a one tailed hypothesis with A > B, $u_{\alpha} = 1{,}645$ (for $\alpha = 5$ %) and Formula (E.1) leads to $n = 154{,}2$, let 155 consumers.

E.3 Case 2: Aiming to demonstrate a difference with assessor rating both product A and product B

The number of consumers is given by Formula (E.2):

$$n = 2 \times \left[\frac{\left(u_{\alpha/2} + u_{\beta} \right) \times \sigma_{\text{within products}} \times \sqrt{1 - \rho}}{\delta} \right]^2 \quad \text{(E.2)}$$

where

$u_{\alpha/2}$ is the $(1 - \alpha/2)$ quantile of the normal distribution for the α-risk for a two tailed alternative hypothesis ; it is equal to 1,960 for an α-risk = 5 %. For an one tailed alternative hypothesis, $u_{\alpha/2}$ is replaced by u_{α} ; for an α-risk = 5 %, u_{α} is equal to 1,645;

u_{β} is the $(1 - \beta)$- quantile of the normal distribution for the tolerated β-risk (desired power, $P = 1\text{–}\beta$);

$\sigma_{\text{within products}}$ is the within products standard deviation; standard deviation is assumed to be homogeneous for the two products;

δ is the difference to be detected between the two products;

ρ is the correlation in the population between the ratings attributed to the two products by the consumers.

NOTE $\sigma_{\text{within products}}$ and δ are expressed in the same scale, for example the 1 to 9 scale.

As

$$\sigma_{\text{d}} = 2 \times \sigma_{\text{within products}} \times \sqrt{(1-\rho)}$$

where σ_{d} is the standard deviation of differences between the ratings attributed to the two products by each consumer of the population, Formula (E.2) becomes

$$n = \left[\frac{(u_{\alpha/2} + u_{\beta}) \times \sigma_{\text{d}}}{\delta} \right]^2 \tag{E.3}$$

Probably, Formula (E.3) is the most frequently used expression.

EXAMPLE For the following conditions

A	$u_{\alpha/2}$	β	$(P = 1 - \beta)$	u_{β}	σ_{d}	δ
0,05	1,960	0,10	(0,90)	1,282	1,70	0,50

Formula (E.3) leads to n = 121, 5; use 122 consumers.

NOTE For a one-tailed hypothesis with $A > B$, u_{α} = 1,645 (for α = 5 %) and Formula (E.3) leads to n = 99, 0.

E.4 Case 3: Aiming to demonstrate that product A is not inferior to product B with each assessor rating only product A or product B

The number of consumers for each product is given by Formula (E.4):

$$n = 2 \times \left[\frac{(u_{\alpha} + u_{\beta}) \times \sigma_{\text{withinproducts}}}{\Delta} \right]^2 \tag{E.4}$$

where

u_{α} is the $(1 - \alpha)$ quantile of the normal distribution for the α-risk for a one tailed alternative hypothesis (non-inferiority test);

u_{β} is the $(1 - \beta)$ quantile of the normal distribution for the tolerated β-risk (desired power $P = 1 - \beta$);

$\sigma_{\text{within product}}$ is the the within product standard deviation; standard deviation is assumed to be homhomogeneous for the two products;

Δ is the difference unacceptable to consider that the two products are similar (interchangeable).

NOTE $\sigma_{\text{within products}}$ and Δ are expressed in the same scale, for example the 1 to 9 scale.

EXAMPLE For the following conditions:

α	u_{α}	β	$(P = 1 - \beta)$	u_{β}	$\sigma_{\text{within products}}$	Δ
0,05	1,645	0,10	(0,90)	1,282	1,40	0,30

Formula (E.4) leads to $n = 373,2$; use 374 consumers per product and 748 consumers for the two products.

E.5 Case 4: Aiming to demonstrate that product A is not inferior to product B with each assessor rating both product A and product B

The number of consumers is given by Formula (E.5):

$$n = \left[\frac{\left(u_\alpha + u_\beta\right) \times \sigma_d}{\Delta} \right]^2 \quad \text{(E.5)}$$

where

u_α is the $(1 - \alpha)$ quantile of the normal distribution for the α-risk for a one tailed alternative hypothesis (non-inferiority test);

u_β is the $(1 - \beta)$- quantile of the normal distribution for the β-risk (desired power $P = 1 - \beta$);

σ_d is the standard deviation of differences between the ratings attributed to the two products by each consumer of the population;

Δ is the difference unacceptable to consider that the two products are similar (interchangeable).

NOTE σ_d and Δ are expressed in the same scale, for example the 1 to 9 scale.

EXAMPLE For the following conditions:

α	u_α	β	$(P = 1 - \beta)$	u_β	σ_d	Δ
0,05	1,645	0,10	(0,90)	1,282	1,70	0,30

Formula (E.5) leads to $n = 275,1$; use 276 consumers.

NOTE Cases 3 and 4 are non-inferiority tests. Strict similarity tests, where the laboratory has to prove that the product A is not inferior or superior to product B, are not addressed. The calculation of the number of consumers for such tests is difficult (there is no expression that directly gives this number) and it leads to some high numbers of consumers. For example, the number of consumers is equal to 480 per product for the conditions of the example of case 3 (960 consumers for the two products) and to 698 for the example of case 4, respectively. For tests of strict similarity, it is intended that the laboratory consults appropriate software. Note that ASTM E1958–07[3] deals with strict similarity tests only when they are made by comparing pairs and when each consumer receives both products A and B.

Annex F
(informative)

Minimum sample size for tests using ranking

F.1 General

Ranking tests may comprise two or more products. By analogy with Annex E concerning the ratings tests, Annex F is limited to presenting the expressions used for calculating the number of consumers in the case of ranking tests concerning two products. For ranking tests comprising three or more products, the laboratory may consult statisticians.

In the field of hedonic measurements, the ranking test on two products is called a preference test. Each consumer receives the two products, A and B; he/she shall determine the product that he/she likes more: A or B.

F.2 Case 1: Demonstrate that there is a preference between the two products — A and B

The number of consumers, n, is given by Formula (F.1):

$$n = \frac{\left[u_{\alpha/2} \times \sqrt{p_0 \times (1-p_0)} + u_{\beta} \times \sqrt{p_a \times (1-p_a)}\right]^2}{(\delta/2)^2} \qquad \text{(F.1)}$$

where

$u_{\alpha/2}$ is the $(1 - \alpha/2)$ quantile of the normal distribution corresponding to the risk α chosen in the case of a two tailed alternative hypothesis. For $\alpha = 5\ \%$, $u_{\alpha/2} = 1{,}960$. In the case of a one tailed hypothesis, $u_{\alpha/2}$ is replaced by u_α (where: $u_\alpha = 1{,}645$ for $\alpha = 5\ \%$);

u_β is the $(1 - \beta)$ quantile of the normal distribution corresponding to the power P required, where: $P = 1 - \beta$;

p_0 is the probability of preference corresponding to the null hypothesis H_0; it is always equal to 0,50 in a preference test;

p_a is the probability of preference of the more liked product corresponding to the alternative hypothesis Ha; it is equal to $p_0 + \delta/2$;

δ is the difference in preference between the two products considered as important.

NOTE 1 Formula (F.1) is based on the approximation of a binomial distribution by a normal distribution; this approximation is valid as soon as the two quantities $n \times p_a$ and $n \times (1 - p_a)$ are at least equal to 5.

NOTE 2 In this International Standard, the value of δ is equal to the difference between the probabilities of preference of the more liked and less liked products. Some authors express the value of δ in relation to the mean of these two probabilities, i.e. in relation to p_0. The reader is expected to be careful not to confuse these two modes of expression.

EXAMPLE For the following conditions:

p_0	δ	p_a	A	u_α	β	Power	u_β
0,50	0,20	1,14	0,05	1,960	0,10	0,90	1,282

Formula (F.1) leads to n = 261,8, i.e. 262 consumers.

If the laboratory questions 262 consumers and there is a difference in frequency of preference equal to 0,20 between the two products A and B, it is certain to conclude a difference of preference on average in 9 out of 10 tests.

NOTE 3 In the case of a unilateral hypothesis with A > B, Formula (F.1) leads to n = 235,6. If the laboratory questions 236 consumers and if the true preference for A is equal to 0,60, it is certain to conclude a significant preference on average in 9 out of 10 tests.

F.3 Case 2: Demonstrate that the frequency of preference for product A is not inferior to that of product B

The number of consumers is given by Formula (F.2):

$$n = \frac{\left[u_\alpha \times \sqrt{p_a \times (1-p_a)} + u_\beta \times \sqrt{p_0 \times (1-p_0)}\right]^2}{(\Delta/2)^2} \qquad \text{(F.2)}$$

where

u_α is the $(1 - \alpha)$- quantile of the centred and reduced normal distribution, corresponding to the risk, α, chosen. In a non-inferiority test, the hypothesis is one tailed, therefore α = 5 %, u_α = 1,645;

u_β is the $(1 - \beta)$- quantile of the normal distribution corresponding to the power, P, required for the test, where $\beta = 1 - P$;

Δ is the difference in probability of preference which does not enable product A to be considered as not inferior to product B;

p_a is the probability of preference for product A in the case of a null hypothesis of a non-inferiority test; it is equal to $p_0 - \Delta/2$;

p_0 is the probability of preference for both products when there is no difference in preference between them; it is equal to 0,50.

NOTE 1 Formula (F.2) is based on the approximation of a binomial distribution by a normal distribution.

NOTE 2 In this International Standard, the Δ value is equal to the difference of the two probabilities p_a and $(1 - p_a)$. Some authors express the Δ value in relation to the mean of these two quantities, i.e. in relation to the probability p_0. The reader is expected to be careful not to confuse these two modes of expression.

EXAMPLE For the following conditions:

$\Delta = \lvert p_a - (1-p_a)\rvert$	p_a	p_0	A	u_α	β	Power	u_β
0,10	0,45	0,50	0,05	1,645	0,10	0,90	1,282

Formula (F.2) leads to n = 851,9, i.e. 852 consumers.

Annex G
(informative)

Examples of data analysis for ratings

G.1 General

This annex uses the examples of Annex E. The first two examples concern a difference test where the null hypothesis H_0, may be defined as:

> The two products, A and B, are extracted from two populations which have the same mean.

The last two examples concern a non-inferiority test where the null hypothesis H_0, may be defined as:

> The product A is inferior to the product B by a quantity at least equal to Δ; therefore the product A is not equivalent to the product B.

For the difference tests, the examples are considered in the two-tailed hypothesis. In the case of one-tailed hypothesis, a statistician shall be consulted.

G.2 Example 1 (case 1 of Annex E): to prove that two products, A and B, are liked differently; each consumer only rates one product

— After discussion with the sponsor, the laboratory questioned 230 consumers: 110 consumers for product A (n_A = 110) and 120 consumers for product B (n_B = 120). For the two products, it obtained the means and standard deviations as follows: m_A = 7,23, s_A = 1,85, m_B = 6,87, and s_B = 1,65.

— The t_{cal} value is given by Formula (G.1):

$$t_{cal} = \frac{m_A - m_B}{s_{\text{within products}} \times \sqrt{\left(\frac{1}{n_A} + \frac{1}{n_B}\right)}} \qquad \text{(G.1)}$$

where

$s_{\text{within products}}$ is the standard deviation within products

$s_{\text{within products}}$ is calculated from Formula (G.2):

$$s_{\text{within products}}{}^2 = \frac{s_A{}^2 \times (n_A - 1) + s_B{}^2 \times (n_B - 1)}{n_A - n_B - 2} \quad \text{(G.2)}$$

Hence in this example:

$$s_{\text{within products}}{}^2 = \frac{1{,}85^2 \times (110 - 1) + 1{,}65^2 \times (120 - 1)}{110 - 120 - 2}$$

and

$$s_{\text{within products}} = \sqrt{s_{\text{within products}}{}^2} + \sqrt{3{,}057} = 1{,}748$$

The t_{cal} value is:

$$t_{\text{cal}} = \frac{7{,}23 - 6{,}87}{1{,}748 \times \sqrt{\left(\frac{1}{100} + \frac{1}{120}\right)}} = \frac{0{,}36}{0{,}231} = 1{,}558$$

This value is compared with the t_{the} value given for the chosen α-risk. This t_{the} value is given by the $(1 - \alpha/2)$-quantile of the t-distribution with $n_A + n_B - 2$ degrees of freedom. For the α-risk = 5 % in this example, it is equal to 1,970.

To reject H_0, t_{cal} needs to be $>t_{\text{the}}$. As 1,558 < 1,970, this condition is not checked in this example; the laboratory cannot conclude there is a significant difference between the two means of products A and B.

— The laboratory therefore calculates the effective power of its test using Formula (G.3):

$$P = 1 - \beta \quad \text{(G.3)}$$

The value of t_β is calculated using Formula (G.3):

$$t_\beta = t_{\text{the}} - \left(\frac{\delta}{t_{\text{calden}}}\right) \quad \text{(G.4)}$$

where

t_{the} is the value leading to reject H_0;

δ is the difference to be detected between the two products;

t_{calden} is the denominator of t_{cal} [Formula (G.1)].

$t_\beta = 1{,}970 - (0{,}50/0{,}231) = -0{,}195$

When $t_\beta < 0$, the value of β is given by the distribution function of the t-distribution with $n_A + n_B - 2$ degrees of freedom at t_β. The distribution function of the t-distribution with 110 + 120 − 2 degrees of freedom at −0,195 equals 0.423.Thus, the test power is equal to $1 - \beta = 1 - 0{,}423 \approx 0{,}58$. This unsatisfactory value is much lower than the required value: 90 %.

NOTE When $t_\beta > 0$, the value of the power is directly given by the distribution function of the t-distribution at (–1) x t_β. For example, for $\delta = 0{,}30$, $t_\beta = 1{,}970 - (0{,}30/0{,}231) = + 0{,}671$; the calculation of the t-distribution function leads to $P = 0{,}251 \approx 0{,}25$.

G.3 Example 2 (case 2 of Annex E): to prove that two products, A and B, are liked differently; each consumer rates both products

The laboratory questioned 120 consumers (n = 120). It obtained m_A = 6,33, m_B = 6,66 and s_d = 1,92

where

s_d is the standard deviation of differences between the ratings attributed to the two products by each consumer.

— The t_{cal} value is given by Formula (G.5):

$$t_{cal} = \frac{m_A - m_B}{s_d \times \sqrt{\left(\frac{1}{n}\right)}}$$

$$t_{cal} = \frac{m_A - m_B}{s_d \times \sqrt{\left(\frac{1}{n}\right)}} = \frac{6,33 - 6,66}{1,92 \times \sqrt{\left(\frac{1}{120}\right)}} = \frac{0,33}{0,175} = -1,886 \quad \text{(G.5)}$$

— This value is compared with the t_{the} value for the chosen α-risk. This t_{the} value is given by the (1 – α/2) quantile of the t-distribution with n –1 degrees of freedom. In this example, for the α-risk = 0, 05 it is equal to 1,980.

To reject H_0 and conclude that there is a significant difference between A and B, the absolute value of t_{cal} needs to be > t_{the}. As this condition is not checked in this example [abs(–1,886) < 1,980], the laboratory cannot conclude a significant difference between the two ratings of products A and B.

— The laboratory therefore calculates the effective power of the test that it has just carried out using the expression:

$$t_\beta = t_{the} - \left(\frac{\delta}{t_{calden}}\right) \quad \text{(G.6)}$$

where

t_{the} is the value leading to reject H_0;

δ is the difference to be detected between the two products;

t_{calden} is the denominator of t_{cal} [Formula (G.5)]

t_β = 1,980 – (0,50/0,175) = –0,873

As the rules concerning the β calculation are the same for Formulae (G.4) and (G.6), the value of β is given by the distribution function of the t-distribution with 120 – 1 degrees of freedom at –0,873; it is equal to 0,192. The power of the test is therefore close to 1 – 0,192 = 0,808, i.e. 81 %, a value generally considered as satisfactory as it is above 80 %.

G.4 Example 3 (case 3 of Annex E): To prove that the product A is not inferior to the product B; each consumer only rates one product

The calculation made in Annex E led to n = 374 consumers per product. The laboratory only questioned 300 consumers for A (n_A = 300) and 316 consumers for B (n_B = 316). It obtained: m_A = 7,01, s_A = 1,95, m_B = 6,87 and s_B = 1,80.

— In a non-inferiority trial, the t_{cal} value is given by Formula (G.7):

$$u_{cal} = \frac{m_A - m_B + \Delta}{s_{within\,products} \times \sqrt{\left(\frac{1}{n_A} + \frac{1}{n_B}\right)}} \quad \text{(G.7)}$$

where

$$s_{within\,products}^2 = \frac{s_A^2 \times (n_A - 1) + s_B^2 (n_B - 1)}{n_A + n_B - 2}$$

In our example,

$$s_{within\,products}^2 = \frac{1{,}95^2 \times (300 - 1) + 1{,}80^2 (316 + -1)}{300 + 316 - 2} = 3{,}514 \text{ and}$$

$$s_d = \sqrt{s_{within\,products}} = \sqrt{3{,}514} = 1{,}875$$

Hence:

$$t_{cal} = \frac{7{,}01 - 6{,}87 + 0{,}30}{1{,}875 \times \sqrt{\frac{1}{300} + \frac{1}{316}}} = \frac{0{,}44}{0{,}151} = 2{,}914$$

— This value is compared with the t_{the} value for the chosen α-risk. This t_{the} value is given by the (1 – α)-quantile of the t-distribution with $n_A + n_B$ –2 degrees of freedom. For the α-risk = 5 % in this example, it is equal to 1,647.

To reject H_0 and conclude a significant non-inferiority of A in relation to B, t_{cal} needs to be > t_{the}. As this condition is checked in this example (2,914 > 1,647), the laboratory can reject H_0 and conclude that the product A is similar to the product B for non-inferiority.

— When the laboratory cannot conclude a significant similarity for non-inferiority, it shall calculate the power of its test. This is given by Formula (G.8):

$$t_\beta = t_{the} - \left(\frac{\Delta}{t_{calden}}\right) \quad \text{(G.8)}$$

where

t_{the} is the value leading to reject H_0;

Δ is the difference unacceptable to consider that two products are similar (i.e interchangeable);

t_{calden} is the denominator of t_{cal} (Formula G.7).

For example, with m_A = 6,87, s_A = 1,95, m_B = 7,01, and s_B = 1,80:

$$t_{cal} = \frac{6{,}87 - 7{,}01 + 0{,}30}{1{,}875 \times \sqrt{\frac{1}{300} + \frac{1}{316}}} = \frac{0{,}16}{0{,}161} = 1{,}060$$

This value does not lead to reject H_0 since 1,060 < 1,647. The value of t_β is:

t_β = 1,647 – (0,30/0,151) = 1,647 – 1,987 = -0,340

The rules concerning the β calculation are the same for Formulae (G.4), (G.6) and (G.8). Since $t_\beta < 0$, the value of β is given by the distribution function of the t-distribution with $n_A + n_B - 2$ degrees of freedom at t_β. The distribution function of the t-distribution with 300+316 – 2 degrees of freedom at –0,340 equals

0,367. Thus the test power is equal to $1 - \beta = 1 - 0{,}367 = 0{,}633 \approx 0{,}63$. This unsatisfactory value is much lower than the required value of 90 %.

G.5 Example 4 (case 4 of Annex E): To prove that the product A is not inferior to the product B); each consumer rates both products

— The calculation made in Annex E led to $n = 374$ consumers. But the laboratory only questioned 200 consumers ($n = 200$). It obtained: $m_A = 7{,}26$, $m_B = 7{,}31$, and $s_d = 2{,}08$

where

s_d is the standard deviation of the population of differences between the two ratings attributed by each consumer to the two products

— In a non-inferiority trial, the t_{cal} value is given by Formula (G.9):

$$t_{cal} = \frac{m_A - m_B + \Delta}{s_d \times \sqrt{\frac{2}{n}}} \tag{G.9}$$

In our example, this is

$$t_{cal} = \frac{7{,}26 - 7{,}31 + 0{,}30}{2{,}08 \times \sqrt{\frac{2}{200}}} = \frac{0{,}25}{0{,}208} = 1{,}202$$

— This value is compared with the t_{the} value for the chosen α-risk. This t_{the} value is given by the $(1 - \alpha)$-quantile of the t-distribution with n-1 degrees of freedom. In this example and for the α-risk = 0,05, it is equal to 1,653.

To reject H_0 and conclude a significant similarity between A and B, t_{cal} needs to be > t_{the}. As this condition does not hold true in this example (1,202 < 1,653), the laboratory cannot reject H_0 and conclude that product A is similar to product B for non-inferiority.

— Then it shall calculate the power of its test using Formula (G.10):

$$t_\beta - t_{the} - \left(\frac{\Delta}{t_{calden}} \right) \tag{G.10}$$

where

t_{the} is the value leading to reject H_0;

Δ is the difference unacceptable to consider that two products are similar (i.e interchangeable);

t_{calden} is the denominator of t_{cal} [Formula (G.9)]

In our example,

$t_\beta = 1{,}653 - (0{,}30/0{,}208) = 0{,}211$

The rules concerning the β calculation are the same for Formulae (G.4), (G.6), (G.8), and (G.10). Since t_β is > 0, the power is directly given by the distribution function of the t-distribution with 200-2 degrees of freedom at $-1 \times t_\beta = -0{,}211$. Hence, it equals $0{,}417 \approx 0{,}42$. This value is very unsatisfactory.

Annex H
(informative)

Examples of data analysis for rankings

H.1 General

This annex uses the examples of Annex F; it therefore only deals with the tests comprising two products.

For complete block (CB) designs with three or more samples, the laboratory can consult ISO 29842. ISO 29842 uses the Friedman test to determine whether significant differences in rankings exist between the products and in the event of a positive response, proposes a multiple comparison test for each pair. For balanced incomplete block designs, the Durbin test is used to determine whether or not significant differences in rank exist between the products.

H.2 Example 1 (case 1 of Annex F): demonstrate that there is a different preference between the two products A and B

— The laboratory wants to demonstrate that there is a preference between the two products A and B. The null hypothesis H_0, is the following:

  The two products A and B are extracted from two populations which have the same mean.

  If the experimental values lead to rejection of H_0, the laboratory shall conclude a significant difference between the means of both products A and B.

— The number of consumers to be questioned, calculated in Annex F, was equal to 262 consumers. After discussion with the sponsor, the laboratory only questioned 200 consumers. It obtained 82 responses in favour of A and 118 responses in favour of B. The relative frequency of preference is therefore equal to f_A = 0,41 for A and f_A = 0,59 for B.

— The u_{cal} value is given by Formula (H.1):

$$u_{cal} = \frac{X - 0{,}5 \times n}{\left(n \times 0{,}5 \times 0{,}5\right)^{0{,}5}} \tag{H1}$$

where

X is the highest number of responses;

n is the total number of responses.

Hence, for the example

$$u_{cal} = \frac{118 - 0{,}5 \times 200}{\left(0{,}50 \times 0{,}5 \times 200\right)^{1/2}} = \frac{18}{7{,}07} = 2{,}55$$

This value is compared with the u_{the} value for a chosen α-risk. For the α-risk = 5 % and a bilateral alternative hypothesis, it is equal to 1,960.

To conclude a significant preference for one of the two products, u_{cal} needs to be > u_{the}. As this condition is met in this example, the laboratory can conclude a significant difference in preference in favour of A.

— If this condition had not been met, the H_0 hypothesis would not have been rejected. The laboratory would therefore have calculated the power of the test as it had not questioned 264 consumers, using the Formula (F1).

H.3 Example 2 (case 2 of Annex F): demonstrate that the frequency of preference for product A is not inferior to that of product B

— The null hypothesis H_0, may be formulated in the following way:

The product A is inferior to the product B by a quantity equal to at least Δ; therefore the product A is not equivalent to the product B.

If the experimental values lead to H_0, being rejected, the similarity for non-inferiority of product A in relation to product B shall be concluded.

— The number of consumers to be questioned given in Annex F is equal to 853. This number is high. It is explained by the low value of Δ (equal to 0,10), based on which the sponsor no longer considers product A as not inferior to B.

After discussion with the sponsor, the laboratory questioned 300 consumers and obtained 147 responses in favour of A and 153 responses in favour of B. Therefore f_A = 0,49 and f_B = 0,51. Hence $f_A - f_B$ = 0,49 – 0,51 = –0,02.

— The u_{cal} value is given by Formula (H.2):

$$u_{cal} = \frac{f_A - f_B + \Delta}{s_d \times \sqrt{\frac{1}{n}}} \qquad \text{(H.2)}$$

where

Δ is the difference of frequency considered as inadmissible;

s_d is the standard deviation of the difference under H_0; it is equal to $2 \times \sqrt{(f_A \times f_B)}$;

n is the number of consumers.

In this example:

$s_d = 2\sqrt{(f_A \times f_B)} = 2\sqrt{(0,55 \times 0,45)} = 2 \times 0,497\,5 = 0,995\,0 \approx 1$

Then:

$$u_{cal} = \frac{0,49 - 0,51 + 0,10}{1 \times \sqrt{\frac{1}{300}}} = \frac{0,08}{0,057\,7} = 1,386$$

This value is compared with the u_{the} value for α = 5 % and a one tailed hypothesis, i.e. 1,645. As $u_{cal} < u_{the}$, the laboratory cannot reject H_0. Hence, there is no significant similarity for non-inferiority of product A in relation to product B.

— Since the u_{cal} value did not enable H_0 to be rejected, the laboratory should calculate the power of the test, using Formula (F.2).

Bibliography

[1] Abdi H., Edelman B., Valentin D., Dowling W.G. Experimental Design and Analysis for Psychology. Oxford University Press, 2009, pp. 1–538.

[2] AFNOR NF V09-500, *Directives Générales pour la réalisation d'épreuves hédoniques effectuées avec des consommateurs dans un espace contrôlé. 2012*

[3] ASTM E1958–12, *Guide for Sensory Claim Substantiation. 2012*

[4] Ball R. Incomplete block designs for the minimization of order and carry-over effects in sensory analysis. Food Qual. Prefer. 1997, **18** pp. 111–118

[5] Bi J. Similarity testing using paired comparison method. Food Qual. Prefer. 2007, **18** pp. 500–507

[6] Blackwelder W.C. Proving the null hypothesis» in clinical trials. Control. Clin. Trials. 1982, **3** pp. 345–353

[7] Cochran W.G., & Cox G.M. Experimental Designs. Wiley, New York, Second Edition, 1957

[8] Cohen J. Statistical Power Analysis for the Behavioural Sciences. Elsevier-Academic Press, New York, Second Edition, 1988

[9] Gacula M.C. Singh J., Bi J., Altan S., Statistical methods in food and consumer research. Academic Press, New York, Second Edition, 2009, pp. 1–853.

[10] Greehouse S.W., & Geisser S. On methods in the analysis of profile data. Psychometrika. 1959, **24** pp. 98–112

[11] Howell D.C. Statistical Methods for Psychology. Duxbury, Pacific Grove, Fifth Edition, 2002, pp. 1–802.

[12] Kunert J., & Sailer O. On Nearly Balanced Designs for Sensory Trials. Food Qual. Prefer. 2006, **17** pp. 219–227

[13] Kunert J. Sensory Experiments as Crossover Studies. Food Qual. Prefer. 1998, **9** pp. 243–253

[14] Lawless H.T., & Heymann H. Sensory Evaluation of Food – Principles and Practices. Springer, New York, Second Edition, 2010, pp. 1–596.

[15] Mac Fie H.J., Bratchell N., Greenhoff K., Vallis L.V. Designs to balance the effect of order of presentation and first-order carry-over effects in hall tests. J. Sens. Stud. 1989, **4** pp. 129–148

[16] Meilgaard M.C., Civille G.V., Carr B.T. Sensory evaluation techniques. CRC Press, Boca Raton, Fourth Edition, 2007, pp. 1–448.

[17] Meyners M. Least equivalent allowable differences in equivalence testing. Food Qual. Prefer. 2007, **18** pp. 541–547

[18] Meyners M. Equivalence tests – A review. Food Qual. Prefer. 2012, **26** pp. 231–245

[19] Naes T., Brockhoff P.B., Tomic B. Statistics for Sensory and Consumer Science. Wiley, Chichester, Second Edition, 2010, pp. 1–287.

[20] Schuirmann D. A comparison of the Two One Sided Tests procedure and the Power Approach for assessing the equivalence of average bioavailability. J. Pharmacokinet. Biopharm. 1987, **15** pp. 657–680

[21] Siegel S., & Castellan N. *Nonparametric systems for the behavioural sciences*. 2nd ed., 1988, Singapour: Mc Graw-Hill International Editions

[22] Ten Kleijf F., & Musters P.A.D Text analysis of open-ended survey responses: a complementary method to preference mapping. Food Qual. Prefer., **Vol. 14**, 2003, pp. 43–52.

[23] Wakeling I.N., & MacFie H.J.H. Designing consumer trials balanced for first and higher orders of carry-over effect when only a subset of k samples from t may be tested. Food Qual. Prefer. 1995, **6** pp. 299–308

[24] Wakeling I.N., Hasted A., Buck D. Cyclic presentation order designs for consumer research. Food Qual. Prefer. 2001, **12** pp. 39–46

[25] ISO 20252, *Market, opinion and social research — Vocabulary and service requirements*

ICS 67.240
Price based on 44 pages

This page deliberately left blank

British Standards Institution (BSI)

BSI is the national body responsible for preparing British Standards and other standards-related publications, information and services.

BSI is incorporated by Royal Charter. British Standards and other standardization products are published by BSI Standards Limited.

About us

We bring together business, industry, government, consumers, innovators and others to shape their combined experience and expertise into standards -based solutions.

The knowledge embodied in our standards has been carefully assembled in a dependable format and refined through our open consultation process. Organizations of all sizes and across all sectors choose standards to help them achieve their goals.

Information on standards

We can provide you with the knowledge that your organization needs to succeed. Find out more about British Standards by visiting our website at bsigroup.com/standards or contacting our Customer Services team or Knowledge Centre.

Buying standards

You can buy and download PDF versions of BSI publications, including British and adopted European and international standards, through our website at bsigroup.com/shop, where hard copies can also be purchased.

If you need international and foreign standards from other Standards Development Organizations, hard copies can be ordered from our Customer Services team.

Subscriptions

Our range of subscription services are designed to make using standards easier for you. For further information on our subscription products go to bsigroup.com/subscriptions.

With **British Standards Online (BSOL)** you'll have instant access to over 55,000 British and adopted European and international standards from your desktop. It's available 24/7 and is refreshed daily so you'll always be up to date.

You can keep in touch with standards developments and receive substantial discounts on the purchase price of standards, both in single copy and subscription format, by becoming a **BSI Subscribing Member**.

PLUS is an updating service exclusive to BSI Subscribing Members. You will automatically receive the latest hard copy of your standards when they're revised or replaced.

To find out more about becoming a BSI Subscribing Member and the benefits of membership, please visit bsigroup.com/shop.

With a **Multi-User Network Licence (MUNL)** you are able to host standards publications on your intranet. Licences can cover as few or as many users as you wish. With updates supplied as soon as they're available, you can be sure your documentation is current. For further information, email bsmusales@bsigroup.com.

BSI Group Headquarters

389 Chiswick High Road London W4 4AL UK

Revisions

Our British Standards and other publications are updated by amendment or revision.

We continually improve the quality of our products and services to benefit your business. If you find an inaccuracy or ambiguity within a British Standard or other BSI publication please inform the Knowledge Centre.

Copyright

Useful Contacts:

Customer Services
Tel: +44 845 086 9001
Email (orders): orders@bsigroup.com
Email (enquiries): cservices@bsigroup.com

Subscriptions
Tel: +44 845 086 9001
Email: subscriptions@bsigroup.com

Knowledge Centre
Tel: +44 20 8996 7004
Email: knowledgecentre@bsigroup.com

Copyright & Licensing
Tel: +44 20 8996 7070
Email: copyright@bsigroup.com

bsi.

...making excellence a habit.™