# INTERNATIONAL STANDARD

# ISO 4120

Second edition
2004-06-15

# Sensory analysis — Methodology — Triangle test

*Analyse sensorielle — Méthodologie — Essai triangulaire*

Reference number
ISO 4120:2004(E)

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

Published in Switzerland

# Contents

Page

# Foreword

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

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

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

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. ISO shall not be held responsible for identifying any or all such patent rights.

ISO 4120 was prepared by Technical Committee ISO/TC 34, *Food products*, Subcommittee SC 12, *Sensory analysis*.

This second edition cancels and replaces the first edition (ISO 4120:1983), which has been technically revised.

# Sensory analysis — Methodology — Triangle test

## 1 Scope

This International Standard describes a procedure for determining whether a perceptible sensory difference or similarity exists between samples of two products. The method is a forced-choice procedure. The method is applicable whether a difference exists in a single sensory attribute or in several attributes.

The method is statistically more efficient than the duo-trio test (described in ISO 10399), but has limited use with products that exhibit strong carryover and/or lingering flavours.

The method is applicable even when the nature of the difference is unknown [i.e. it determines neither the size nor the direction of difference between samples, nor is there any indication of the attribute(s) responsible for the difference]. The method is applicable only if the products are fairly homogeneous.

The method is effective for

a) determining that

— either a perceptible difference results (triangle testing for difference), or

— a perceptible difference does not result (triangle testing for similarity) when, for example, a change is made in ingredients, processing, packaging, handling or storage;

b) or for selecting, training and monitoring assessors.

## 2 Normative references

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

ISO 5492:1992, *Sensory analysis — Vocabulary*

ISO 8589:1988, *Sensory analysis — General guidance for the design of test rooms*

## 3 Terms and definitions

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

**3.1**
**alpha-risk**
**$\alpha$-risk**
probability of concluding that a perceptible difference exists when one does not

NOTE This is also known as Type I error, significance level or false positive rate.

**3.2**
**beta-risk**
***β*-risk**
probability of concluding that no perceptible difference exists when one does

NOTE This is also known as Type II error or false negative rate.

**3.3**
**difference**
situation in which samples can be distinguished based on their sensory properties

NOTE The proportion of assessments in which a perceptible difference is detected between the two products is given the symbol $p_d$.

**3.4**
**product**
material to be evaluated

**3.5**
**sample**
unit of product prepared, presented and evaluated in the test

**3.6**
**sensitivity**
general term used to summarize the performance characteristics of the test

NOTE In statistical terms, the sensitivity of the test is defined by the values of $\alpha$, $\beta$ and $p_d$.

**3.7**
**similarity**
situation in which any perceptible differences between the samples are so small that the products can be used interchangeably

**3.8**
**triad**
those three samples given to an assessor in the triangle test

NOTE In the triangle test, each sample is marked with a different code. Two of the samples are alike (i.e. from one product) and one is different (i.e. from the other product).

## 4 Principle

The number of assessors is chosen based on the sensitivity desired for the test. (See 6.2 and the discussion in A.3.)

Assessors receive a set of three samples (i.e. a triad) and are informed that two of the samples are alike and that one is different. The assessors report which sample they believe to be different, even if the selection is based only on a guess.

The number of correct responses is counted and the significance is determined by reference to a statistical table.

## 5 General test conditions and requirements

**5.1** Clearly define the test objective in writing.

**5.2** Carry out the test under conditions that prevent communication among assessors until all the evaluations have been completed, using facilities and booths that comply with ISO 8589.

**5.3** Prepare the samples out of sight of the assessors and in an identical manner (i.e. same apparatus, same vessels, same quantities of product).

**5.4** Assessors shall not be able to identify the samples from the way in which they are presented. For example, in a taste test, avoid any differences in appearance. Mask any irrelevant colour differences using light filters and/or subdued illumination.

**5.5** Code the vessels containing the samples in a uniform manner, preferably using three-digit numbers, chosen at random for each test. Each triad is composed of three samples, each with a different code. Preferably, different codes should be used for each assessor during a session. However, the same three codes may be used for all assessors within a test, provided that each code is used only once per assessor during a test session (e.g. if several triangle tests on different products are being conducted in the same session).

**5.6** The quantity or volume served shall be identical for the three samples in each triad, just as that of all the other samples in a series of tests on a given type of product. The quantity or volume to be evaluated may be imposed. If it is not, the assessors should be told to take quantities or volumes that are always similar whatever the sample.

**5.7** The temperature of the three samples in each triad shall be identical, just as that of all the other samples in a series of tests on a given type of product. It is preferable to present the samples at the temperature at which the product is generally consumed.

**5.8** The assessors shall be told whether or not they are to swallow the samples or whether they are free to do as they please. In this latter case, they shall be requested to proceed in the same manner for all the samples.

**5.9** During the test sessions, avoid giving information about product identity, expected treatment effects, or individual performance until all testing is completed.

## 6 Assessors

### 6.1 Qualification

All assessors should possess the same level of qualification, this level being chosen on the basis of the test objective (see ISO 8586-1 and ISO 8586-2 for guidance). Experience and familiarity with the product may improve the performance of an assessor and, therefore, may increase the likelihood of finding a significant difference. Monitoring the performance of assessors over time may be useful for increased sensitivity.

All assessors shall be familiar with the mechanics of the triangle test (i.e. format, task and evaluation procedure).

### 6.2 Number of assessors

Choose the number of assessors so as to obtain the sensitivity required for the test (see the discussion in A.3). Using large numbers of assessors increases the likelihood of detecting small differences between the products. However, in practice, the number of assessors is often determined by material conditions (e.g. duration of the experiment, number of available assessors, quantity of product). When testing for a difference, typical numbers of assessors are between 24 and 30. When testing for no meaningful difference (i.e. similarity), twice as many assessors (i.e. approximately 60) are needed for equivalent sensitivity.

Avoid replicate evaluations by the same assessor whenever possible. However, if replicate evaluations are needed to produce a sufficient number of total evaluations, every effort should be made to have each assessor perform the same number of replicate evaluations. For example, if only ten assessors are available, have each assessor evaluate three triads to obtain a total of 30 evaluations.

NOTE Treating three evaluations performed by ten assessors as 30 independent evaluations is not valid when testing for similarity using Table A.2. However, the test for difference using Table A.1 is valid even when replicate evaluations are performed (see [9] and [10]). Recent publications (see [6] and [7]) on replicated discrimination tests suggest alternative approaches for analysing replicated evaluations in discrimination tests.

## 7 Procedure

**7.1** Prepare worksheets and scoresheets (see B.1 and B.2) in advance of the test so as to utilize an equal number of the six possible sequences of two products, A and B:

ABB AAB ABA

BAA BBA BAB

Distribute these at random in groups of six among the assessors (i.e. use each sequence once among the first group of six assessors; use each sequence once again among the next group of six assessors, etc.). This will minimize the imbalance that results if the total number of assessors is not a multiple of six.

**7.2** Present the three samples of each triad simultaneously if possible, following the same spatial arrangement for each assessor (e.g. on a line to be sampled always from left to right, in a triangle array). Within the triad, assessors are generally allowed to make repeated evaluations of each sample as desired (if, of course, the nature of the product allows for repeated evaluations).

**7.3** Instruct the assessors to evaluate the samples in the order in which they were presented. Inform the assessors that two of the samples are the same and that one is different. Each assessor should then indicate which one of the three samples is different from the other two.

**7.4** Each scoresheet should provide for a single triad of samples. If an assessor is to carry out more than one test in a session, collect the completed scoresheet and unused samples prior to serving the subsequent triad. The assessor shall not go back to any of the previous samples or change the verdict on any previous test.

**7.5** Do not ask questions about preference, acceptance or degree of difference after the initial selection of the odd sample. The selection the assessor has just made may bias the reply to any additional questions. Responses to such questions may be obtained through separate tests for preference, acceptance, degree of difference, etc. (See ISO 6658 for guidance.) A comment section asking why the choice was made may be included for the assessor’s remarks.

**7.6** The triangle test is a forced-choice procedure; assessors are not allowed the option of reporting “no difference”. An assessor who detects no difference between the samples should be instructed to randomly select one of the samples and to indicate that the selection was only a guess in the comments section of the scoresheet.

## 8 Analysis and interpretation of results

### 8.1 When testing for a difference

Use Table A.1 to analyse the data obtained from a triangle test. If the number of correct responses is greater than or equal to the number given in Table A.1 (corresponding to the number of assessors and the $\alpha$-risk level chosen for the test), conclude that a perceptible difference exists between the samples (see B.1).

If desired, calculate a confidence interval on the proportion of the population that can distinguish the samples. The method is described in B.3.

### 8.2 When testing for similarity[1)]

Use Table A.2 to analyse the data obtained from a triangle test. If the number of correct responses is less than or equal to the number given in Table A.2 (corresponding to the number of assessors, the $\beta$-risk level and the value of $p_d$ chosen for the test), conclude that no meaningful difference exists between the samples (see B.2). If results are to be compared from one test to another, then the same value of $p_d$ should be chosen for all tests.

If desired, calculate a confidence interval on the proportion of the population that can distinguish the samples. The method is described in B.3.

No conclusion should be drawn for maximum numbers of correct responses below $n/3$.

## 9 Test report

Report the test objective, the results and the conclusions. The following additional information is recommended:

— the purpose of the test and the nature of the treatment studied;

— full identification of the samples (i.e. origin, method of preparation, quantity, shape, storage prior to testing, serving size, temperature); sample information should communicate that all storage, handling and preparation was done in such a way as to yield samples that differ only due to the variable of interest, if at all;

— the number of assessors, the number of correct responses and the result of the statistical evaluation (including the values of $\alpha$, $\beta$ and $p_d$ used for the test);

— assessors: experience (in sensory testing, with the product, with the samples in the test), age and gender (see ISO 8586-1 and ISO 8586-2 for guidance);

— any information and any specific recommendations given to the assessors in connection with the test;

— the test environment (i.e. test facility used, simultaneous or sequential presentation, if the identity of samples disclosed after the test and, if so, in what manner);

— the location, date of the test and name of the panel leader.

## 10 Precision and bias

Because results of sensory discrimination tests are a function of individual sensitivities, a general statement regarding the reproducibility of results that is applicable to all populations of assessors cannot be made. Precision regarding a particular population of assessors increases as the size of the panel increases, and also with their training and with exposure to the product.

Since a forced-choice procedure is used, results obtained by this method are bias-free, provided that the precautions in Clause 7 are fully observed.

1) In this International Standard, "similar" does not mean "identical". Rather, "similar" means that the two products are sufficiently alike to be used interchangeably. It is not possible to prove that two products are identical. However, it can be demonstrated that any difference that does exist between two products is so small as to have no practical significance.

# Annex A
(normative)

# Tables

**A.1** Values given in Table A.1 are the minimum number of correct responses required for significance at the stated $\alpha$-risk level (i.e. column) for the corresponding number of assessors, $n$ (i.e. row). Reject the assumption of "no difference" if the number of correct responses is greater than or equal to the value in Table A.1.

**Table A.1 — Minimum number of correct responses needed to conclude that a perceptible difference exists based on a triangle test**

| $n$ | $\alpha$ | | | | | $n$ | $\alpha$ | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | **0,20** | **0,10** | **0,05** | **0,01** | **0,001** | | **0,20** | **0,10** | **0,05** | **0,01** | **0,001** |
| 6 | 4 | 5 | 5 | 6 | — | 27 | 12 | 13 | 14 | 16 | 18 |
| 7 | 4 | 5 | 5 | 6 | 7 | 28 | 12 | 14 | 15 | 16 | 18 |
| 8 | 5 | 5 | 6 | 7 | 8 | 29 | 13 | 14 | 15 | 17 | 19 |
| 9 | 5 | 6 | 6 | 7 | 8 | 30 | 13 | 14 | 15 | 17 | 19 |
| 10 | 6 | 6 | 7 | 8 | 9 | | | | | | |
| | | | | | | 31 | 14 | 15 | 16 | 18 | 20 |
| 11 | 6 | 7 | 7 | 8 | 10 | 32 | 14 | 15 | 16 | 18 | 20 |
| 12 | 6 | 7 | 8 | 9 | 10 | 33 | 14 | 15 | 17 | 18 | 21 |
| 13 | 7 | 8 | 8 | 9 | 11 | 34 | 15 | 16 | 17 | 19 | 21 |
| 14 | 7 | 8 | 9 | 10 | 11 | 35 | 15 | 16 | 17 | 19 | 22 |
| 15 | 8 | 8 | 9 | 10 | 12 | | | | | | |
| | | | | | | 36 | 15 | 17 | 18 | 20 | 22 |
| 16 | 8 | 9 | 9 | 11 | 12 | 42 | 18 | 19 | 20 | 22 | 25 |
| 17 | 8 | 9 | 10 | 11 | 13 | 48 | 20 | 21 | 22 | 25 | 27 |
| 18 | 9 | 10 | 10 | 12 | 13 | 54 | 22 | 23 | 25 | 27 | 30 |
| 19 | 9 | 10 | 11 | 12 | 14 | 60 | 24 | 26 | 27 | 30 | 33 |
| 20 | 9 | 10 | 11 | 13 | 14 | 66 | 26 | 28 | 29 | 32 | 35 |
| | | | | | | | | | | | |
| 21 | 10 | 11 | 12 | 13 | 15 | 72 | 28 | 30 | 32 | 34 | 38 |
| 22 | 10 | 11 | 12 | 14 | 15 | 78 | 30 | 32 | 34 | 37 | 40 |
| 23 | 11 | 12 | 12 | 14 | 16 | 84 | 33 | 35 | 36 | 39 | 43 |
| 24 | 11 | 12 | 13 | 15 | 16 | 90 | 35 | 37 | 38 | 42 | 45 |
| 25 | 11 | 12 | 13 | 15 | 17 | 96 | 37 | 39 | 41 | 44 | 48 |
| 26 | 12 | 13 | 14 | 15 | 17 | 102 | 39 | 41 | 43 | 46 | 50 |

NOTE 1 Values in the table are exact because they are based on the binomial distribution. For values of $n$ not in the table, compute approximate values for the missing entries based on the normal approximation to the binomial as follows. Minimum number of responses ($x$) = nearest whole number greater than

$$x = (n/3) + z\sqrt{2n/9}$$

where

$z$ varies with the significance level as follows: 0,84 for $\alpha = 0{,}20$; 1,28 for $\alpha = 0{,}10$; 1,64 for $\alpha = 0{,}05$; 2,33 for $\alpha = 0{,}01$; 3,09 for $\alpha = 0{,}001$.

NOTE 2 Values of $n < 18$ are usually not recommended for a triangle test for a difference.

NOTE 3 Adapted from Reference [11].

**A.2** Values given in Table A.2 are the maximum number of correct responses required for "similarity" at the chosen levels of $p_d$, $\beta$ and $n$. Accept the assumption of "no difference" at the 100(1−$\beta$) % level of confidence if the number of correct responses is less than or equal to the value in Table A.2.

**Table A.2 — Maximum number of correct responses needed to conclude that two samples are similar, based on a triangle test**

| $n$ | $\beta$ | $p_d$ | | | | | $n$ | $\beta$ | $p_d$ | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | **10 %** | **20 %** | **30 %** | **40 %** | **50 %** | | | **10 %** | **20 %** | **30 %** | **40 %** | **50 %** |
| 18 | 0,001 | 0 | 1 | 2 | 3 | 5 | 66 | 0,001 | 14 | 18 | 22 | 26 | 31 |
| | 0,01 | 2 | 3 | 4 | 5 | 6 | | 0,01 | 16 | 20 | 25 | 29 | 34 |
| | 0,05 | 3 | 4 | 5 | 6 | 8 | | 0,05 | 19 | 23 | 28 | 32 | 37 |
| | 0,10 | 4 | 5 | 6 | 7 | 8 | | 0,10 | 20 | 25 | 29 | 33 | 38 |
| | 0,20 | 4 | 6 | 7 | 8 | 9 | | 0,20 | 22 | 26 | 31 | 35 | 40 |
| 24 | 0,001 | 2 | 3 | 4 | 6 | 8 | 72 | 0,001 | 15 | 20 | 24 | 29 | 34 |
| | 0,01 | 3 | 5 | 6 | 8 | 9 | | 0,01 | 18 | 23 | 28 | 32 | 38 |
| | 0,05 | 5 | 6 | 8 | 9 | 11 | | 0,05 | 21 | 26 | 30 | 35 | 40 |
| | 0,10 | 6 | 7 | 9 | 10 | 12 | | 0,10 | 22 | 27 | 32 | 37 | 42 |
| | 0,20 | 7 | 8 | 10 | 11 | 13 | | 0,20 | 24 | 29 | 34 | 39 | 44 |
| 30 | 0,001 | 3 | 5 | 7 | 9 | 11 | 78 | 0,001 | 17 | 22 | 27 | 32 | 38 |
| | 0,01 | 5 | 7 | 9 | 11 | 13 | | 0,01 | 20 | 25 | 30 | 36 | 41 |
| | 0,05 | 7 | 9 | 11 | 13 | 15 | | 0,05 | 23 | 28 | 33 | 39 | 44 |
| | 0,10 | 8 | 10 | 11 | 14 | 16 | | 0,10 | 25 | 30 | 35 | 40 | 46 |
| | 0,20 | 9 | 11 | 13 | 15 | 17 | | 0,20 | 27 | 32 | 37 | 42 | 48 |
| 36 | 0,001 | 5 | 7 | 9 | 11 | 14 | 84 | 0,001 | 19 | 24 | 30 | 35 | 41 |
| | 0,01 | 7 | 9 | 11 | 14 | 16 | | 0,01 | 22 | 28 | 33 | 39 | 45 |
| | 0,05 | 9 | 11 | 13 | 16 | 18 | | 0,05 | 25 | 31 | 36 | 42 | 48 |
| | 0,10 | 10 | 12 | 14 | 17 | 19 | | 0,10 | 27 | 32 | 38 | 44 | 49 |
| | 0,20 | 11 | 13 | 16 | 18 | 21 | | 0,20 | 29 | 34 | 40 | 46 | 51 |
| 42 | 0,001 | 6 | 9 | 11 | 14 | 17 | 90 | 0,001 | 21 | 27 | 32 | 38 | 45 |
| | 0,01 | 9 | 11 | 14 | 17 | 20 | | 0,01 | 24 | 30 | 36 | 42 | 48 |
| | 0,05 | 11 | 13 | 16 | 19 | 22 | | 0,05 | 27 | 33 | 39 | 45 | 52 |
| | 0,10 | 12 | 14 | 17 | 20 | 23 | | 0,10 | 29 | 35 | 41 | 47 | 53 |
| | 0,20 | 13 | 16 | 19 | 22 | 24 | | 0,20 | 31 | 37 | 43 | 49 | 55 |
| 48 | 0,001 | 8 | 11 | 14 | 17 | 21 | 96 | 0,001 | 23 | 29 | 35 | 42 | 48 |
| | 0,01 | 11 | 13 | 17 | 20 | 23 | | 0,01 | 26 | 33 | 39 | 45 | 52 |
| | 0,05 | 13 | 16 | 19 | 22 | 26 | | 0,05 | 30 | 36 | 42 | 49 | 55 |
| | 0,10 | 14 | 17 | 20 | 23 | 27 | | 0,10 | 31 | 38 | 44 | 50 | 57 |
| | 0,20 | 15 | 18 | 22 | 25 | 28 | | 0,20 | 33 | 40 | 46 | 53 | 59 |
| 54 | 0,001 | 10 | 13 | 17 | 20 | 24 | 102 | 0,001 | 25 | 31 | 38 | 45 | 52 |
| | 0,01 | 12 | 16 | 19 | 23 | 27 | | 0,01 | 28 | 35 | 42 | 49 | 56 |
| | 0,05 | 15 | 18 | 22 | 25 | 29 | | 0,05 | 32 | 38 | 45 | 52 | 59 |
| | 0,10 | 16 | 20 | 23 | 27 | 31 | | 0,10 | 33 | 40 | 47 | 54 | 61 |
| | 0,20 | 18 | 21 | 25 | 28 | 32 | | 0,20 | 36 | 42 | 49 | 56 | 63 |
| 60 | 0,001 | 12 | 15 | 19 | 23 | 27 | 108 | 0,001 | 27 | 34 | 41 | 48 | 55 |
| | 0,01 | 14 | 18 | 22 | 26 | 30 | | 0,01 | 31 | 37 | 45 | 52 | 59 |
| | 0,05 | 17 | 21 | 25 | 29 | 33 | | 0,05 | 34 | 41 | 48 | 55 | 63 |
| | 0,10 | 18 | 22 | 26 | 30 | 34 | | 0,10 | 36 | 43 | 50 | 57 | 65 |
| | 0,20 | 20 | 24 | 28 | 32 | 36 | | 0,20 | 38 | 45 | 52 | 60 | 67 |

NOTE 1 Values in the table are exact because they are based on the binomial distribution. For values of $n$ not in the table, compute the 100(1–$\beta$) % upper confidence limit for $p_d$ based on the normal approximation to the binomial as:

$$\left[1{,}5(x/n) - 0{,}5\right] + 1{,}5\, z_\beta \sqrt{(nx - x^2)/n^3}$$

where

$x$ is the number of correct answers; $n$ is the number of assessors; $z_\beta$ varies as follows: 0,84 for $\beta = 0{,}20$; 1,28 for $\beta = 0{,}10$; 1,64 for $\beta = 0{,}05$; 2,33 for $\beta = 0{,}01$; 3,09 for $\beta = 0{,}001$.

If the computed value is less than the selected limit for $p_d$, then declare the samples similar at the $\beta$ level of significance.

NOTE 2 Values of $n < 30$ are usually not recommended for a triangle test for similarity.

NOTE 3 Adapted from Reference [11].

**A.3** Table A.3 shows a statistical approach for determining the number of assessors. The statistical sensitivity of the test is a function of three values: the $\alpha$-risk, the $\beta$-risk and the maximum allowable proportion of distinguishers, $p_d$[2]. Prior to conducting the test, select values for $\alpha$, $\beta$ and $p_d$ using the following guidelines.

As a rule of thumb, a statistically significant result at

— an $\alpha$-risk of 10 % to 5 % (0,10 to 0,05) indicates slight evidence that a difference was apparent,

— an $\alpha$-risk of 5 % to 1 % (0,05 to 0,01) indicates moderate evidence that a difference was apparent,

— an $\alpha$-risk of 1 % to 0,1 % (0,01 to 0,001) indicates strong evidence that a difference was apparent, and

— an $\alpha$-risk below 0,1 % ($<$ 0,001) indicates very strong evidence that a difference was apparent.

For $\beta$-risks, the strength of the evidence that a difference was not apparent is assessed using the same criteria as above (substituting "was not apparent" for "was apparent").

The maximum allowable proportion of distinguishers, $p_d$, falls into three ranges:

— $p_d < 25$ % represent small values;

— 25 % $< p_d <$ 35 % represent medium sized values; and

— $p_d > 35$ % represent large values.

Choose the number of assessors so as to obtain the level of sensitivity required for the test. Enter Table A.3 in the section corresponding to the selected value of $p_d$ and the column corresponding to the selected value of $\beta$. The minimum required number of assessors is found in the row corresponding to the selected value of $\alpha$. Alternatively, Table A.3 may be used to develop a set of values for $p_d$, $\alpha$ and $\beta$ that provide acceptable sensitivity while maintaining the number of assessors within practical limits. The approach is presented in detail in Reference [12].

Values given in Table A.3 are the minimum number of assessors required to execute a triangle test with a specified sensitivity determined by the values of $p_d$, $\alpha$ and $\beta$. Enter the table in the section corresponding to the chosen value of $p_d$ and the column corresponding to the chosen value of $\beta$. Read the minimum number of assessors from the row corresponding to the chosen value of $\alpha$.

2) In this International Standard, the probability of a correct response, $p_c$, is modelled as $p_c = p_d + (1/3)(1-p_d)$, where $p_d$ is the proportion of the population of assessors who can distinguish between the two products. A psychometrical model of the assessor's decision process, such as the Thurstone-Ura model (see [8]), could also be applied in a triangle test.

**Table A.3 — Number of assessors needed for a triangle test**

| $\alpha$ | $p_d$ | $\beta$ | | | | |
|---|---|---|---|---|---|---|
| | | **0,20** | **0,10** | **0,05** | **0,01** | **0,001** |
| 0,20 | 50 % | 7 | 12 | 16 | 25 | 36 |
| 0,10 | | 12 | 15 | 20 | 30 | 43 |
| 0,05 | | 16 | 20 | 23 | 35 | 48 |
| 0,01 | | 25 | 30 | 35 | 47 | 62 |
| 0,001 | | 36 | 43 | 48 | 62 | 81 |
| 0,20 | 40 % | 12 | 17 | 25 | 36 | 55 |
| 0,10 | | 17 | 25 | 30 | 46 | 67 |
| 0,05 | | 23 | 30 | 40 | 57 | 79 |
| 0,01 | | 35 | 47 | 56 | 76 | 102 |
| 0,001 | | 55 | 68 | 76 | 102 | 130 |
| 0,20 | 30 % | 20 | 28 | 39 | 64 | 97 |
| 0,10 | | 30 | 43 | 54 | 81 | 119 |
| 0,05 | | 40 | 53 | 66 | 98 | 136 |
| 0,01 | | 62 | 82 | 97 | 131 | 181 |
| 0,001 | | 93 | 120 | 138 | 181 | 233 |
| 0,20 | 20 % | 39 | 64 | 86 | 140 | 212 |
| 0,10 | | 62 | 89 | 119 | 178 | 260 |
| 0,05 | | 87 | 117 | 147 | 213 | 305 |
| 0,01 | | 136 | 176 | 211 | 292 | 397 |
| 0,001 | | 207 | 257 | 302 | 396 | 513 |
| 0,20 | 10 % | 149 | 238 | 325 | 529 | 819 |
| 0,10 | | 240 | 348 | 457 | 683 | 1 011 |
| 0,05 | | 325 | 447 | 572 | 828 | 1 181 |
| 0,01 | | 525 | 680 | 824 | 1 132 | 1 539 |
| 0,001 | | 803 | 996 | 1 165 | 1 530 | 1 992 |
| NOTE Adapted from Reference [12]. | | | | | | |

# Annex B
(informative)

# Examples

## B.1 Example 1: Triangle test to confirm that a difference exists

### B.1.1 Background

A brewery has developed a process to reduce the unwanted grainy flavour characteristic in its non-alcoholic beer. The process requires an investment in new equipment. Before proceeding to a larger scale preference test involving consumers, the head brewer wants to confirm that the experimental non-alcoholic beer is different from the company's current non-alcoholic beer. The head brewer is willing to take only a small chance of concluding that a difference exists when one does not. However, the head brewer is willing to accept a greater risk of missing a difference that does exist as there are alternative ways of operating the new process.

### B.1.2 Test objective

The objective is to confirm that the current prototype process produces a non-alcoholic beer that can be distinguished from the current non-alcoholic product in order to justify testing with consumers.

### B.1.3 Number of assessors

To protect the head brewer from falsely concluding that a difference exists, the sensory analyst proposes $\alpha = 0{,}05$. In order to balance the order of presentation of the samples, the analyst decides to use 24 assessors. Note from Table A.3 that choosing to use 24 assessors also ensures that the test has a 95 % chance [i.e. 100(1−$\beta$) %] of detecting the case in which 50 % of the assessors can detect a difference between the samples in the test. The actual entry in Table A.3 for $\alpha = 0{,}05$, $\beta = 0{,}05$ and $p_\mathrm{d} = 50$ % is $n = 23$.

### B.1.4 Conducting the test

Samples (36 glasses of “A” and 36 glasses of “B”) are coded with unique random numbers. Each of the triads ABB, BAA, AAB, BBA, ABA, and BAB is presented four times so as to cover the 24 assessors in a balanced random order. An example of the scoresheet used is shown in Figure B.1.

### B.1.5 Analysis and interpretation of results

A total of 14 assessors correctly identify the odd sample. In Table A.1, in the row corresponding to $n = 24$ assessors and the column corresponding to $\alpha = 0{,}05$, the sensory analyst finds that 14 correct responses are sufficient to conclude that the two beers are perceptibly different.

Optionally, the analyst may choose to compute a one-sided, lower confidence interval on the proportion of the population that can perceive a difference between the samples. The calculations (see also B.3) are:

$$\left[1{,}5 \times (14/24) - 0{,}5\right] - 1{,}5 \times 1{,}64\sqrt{(14/24)\left[1-(14/24)\right]/24} = 0{,}13$$

The analyst is able to conclude with 95 % confidence that at least 13 % of the population can perceive a difference between the samples.

### B.1.6 Test report and conclusions

The sensory analyst reports that the prototype could, in fact, be distinguished from the current product by the panel ($n = 24$, $x = 14$) at the 5 % level of significance. The analyst may choose to report at the 95 % confidence level that at least 13 % of the assessors could distinguish the two samples. Brewing trials using the new process should proceed to testing with consumers as proposed in B.1.2.

**Triangle test**

Assessor No. ________________ Name _________________________________ Date __________________

**Instructions**

Taste samples from left to right. Two samples are alike; one is different. Write the number of the sample that differs from the others in the space below. If you are not sure, record your best guess; you may note under Remarks that you were guessing.

The one sample that differs from the other two is:_______________

Remarks: ______________________________________________________________________________

_______________________________________________________________________________________

**Figure B.1 — Scoresheet for triangle difference test in Example 1**

## B.2 Example 2: Triangle test to confirm that two samples are similar

### B.2.1 Background

A confectionery manufacturer wants to utilize a new packaging material because it offers greater flexibility for label graphics. However, the new material must provide the same level of storage stability. The manufacturer realizes that it is impossible to prove that the two products are identical, but wants to be very sure that only a reasonably small proportion of the population will be able to detect a difference after 3-months storage if one exists. On the other hand, the manufacturer is willing to take a fairly large chance of incorrectly concluding that the products are different when they are not, because the current package is acceptable as far as graphics are concerned and storage stability is a more important criterion to the manufacturer.

### B.2.2 Test objective

The objective is to determine if product stored for 3 months in the new packaging material is the same as product stored for 3 months in the current packaging material.

### B.2.3 Number of assessors

The sensory analyst works with the manufacturer to decide on the levels of risk that are appropriate for the test. It is decided that the maximum allowable proportion of discriminators should be $p_d = 20$ %. The manufacturer is only willing to take a $\beta = 0{,}10$ chance of failing to detect that level of discriminators. Because the manufacturer is not greatly concerned with falsely concluding that a difference exists when one does not, the sensory analyst chooses $\alpha = 0{,}20$. Entering Table A.3 with $\alpha = 0{,}20$, $\beta = 0{,}10$ and $p_d = 20$ %, the sensory analyst finds that $n = 64$ assessors need to be recruited for the test.

### B.2.4 Conducting the test

The sensory analyst uses the worksheet shown in Figure B.2 and the score sheet shown in Figure B.3 to run the test. The analyst cycles through the six possible triads: AAB, ABA, BAA, BBA, BAB and ABB ten times with the first 60 assessors. The analyst then randomly selects four triads to serve assessors 61 through 64.

### B.2.5 Analysis and interpretation of results

A total of 24 out of the 64 assessors correctly identify the odd sample in the test. Referring to Table A.2, the analyst finds that there is no entry for $n = 64$. So, the analyst uses the equation in Note 1 to Table A.2 to determine if it can be concluded that the two samples are similar. The analyst finds that:

$$\left[1{,}5 \times (24/64) - 0{,}5\right] + 1{,}5 \times 1{,}28\sqrt{\left(64 \times 24 - 24^2\right)/64^3} = 0{,}178\ 7$$

That is, the analyst can be 90 % confident that no more than 18 % of assessors can distinguish the samples. The analyst concludes that the new packaging meets the manufacturer's criterion of 90 % certainty (i.e. $\beta = 0{,}10$) that no more than $p_d = 20$ % of the population are able to detect a difference. The new packaging may be substituted for the current.

## B.3 Example 3: Confidence intervals for triangle tests

### B.3.1 Background

If desired, analysts can calculate confidence intervals on the proportion of the population that can distinguish the samples. The calculations are as follows, where $x =$ the number of correct responses and $n =$ the total number of assessors:

— $p_c$ (proportion correct) $= x/n$

— $\hat{p}_d$ (proportion distinguished) $= 1{,}5\,p_c - 0{,}5$

— $s_d$ (standard deviation of $\hat{p}_d$) $= 1{,}5\sqrt{p_c(1 - p_c)/n}$

— upper confidence limit $= \hat{p}_d + z_\alpha s_d$

— lower confidence limit $= \hat{p}_d - z_\alpha s_d$

where $z_\alpha$ is the critical value of the standard normal distribution.

For a 90 % confidence interval, $z_\alpha = 1{,}28$; for a 95 % confidence interval, $z_\alpha = 1{,}64$; and for a 99 % confidence interval, $z_\alpha = 2{,}33$.

Date: Oct. 4, 1993 Test code: 587-FF03

**Triangle test sample order and serving protocol**

Post this sheet in the area where trays are prepared. Code score sheets and serving containers ahead of time

Product type: *Candy Bars*

Sample identification:

Sample 1 = Package 4736 (current) Sample 2 = Package 3987 (new)

Code serving containers as follows:

| Panelist | Sample-code | | | Panelist | Sample-code | | |
|---|---|---|---|---|---|---|---|
| 1 | 1-108 | 1-795 | 2-140 | 33 | 2-360 | 1-303 | 1-415 |
| 2 | 1-189 | 2-168 | 1-733 | 34 | 2-134 | 2-401 | 1-305 |
| 3 | 2-718 | 1-437 | 1-488 | 35 | 2-185 | 1-651 | 2-307 |
| 4 | 2-535 | 2-231 | 1-243 | 36 | 1-508 | 2-271 | 2-465 |
| 5 | 2-839 | 1-402 | 2-619 | 37 | 1-216 | 1-941 | 2-321 |
| 6 | 1-145 | 2-296 | 2-992 | 38 | 1-494 | 2-783 | 1-414 |
| 7 | 1-792 | 1-280 | 2-319 | 39 | 2-151 | 1-786 | 1-943 |
| 8 | 1-167 | 2-936 | 1-180 | 40 | 2-423 | 2-477 | 1-164 |
| 9 | 2-689 | 1-743 | 1-956 | 41 | 2-570 | 1-772 | 2-887 |
| 10 | 2-442 | 2-720 | 1-213 | 42 | 1-398 | 2-946 | 2-764 |
| 11 | 2-253 | 1-444 | 2-505 | 43 | 1-747 | 1-286 | 2-913 |
| 12 | 1-204 | 2-159 | 2-556 | 44 | 1-580 | 2-558 | 1-114 |
| 13 | 1-142 | 1-325 | 2-632 | 45 | 2-345 | 1-562 | 1-955 |
| 14 | 1-472 | 2-762 | 1-330 | 46 | 2-385 | 2-660 | 1-856 |
| 15 | 2-965 | 1-641 | 1-300 | 47 | 2-754 | 1-210 | 2-864 |
| 16 | 2-582 | 2-659 | 1-486 | 48 | 1-574 | 2-393 | 2-753 |
| 17 | 2-429 | 1-884 | 2-499 | 49 | 1-793 | 1-308 | 2-742 |
| 18 | 1-879 | 2-891 | 2-404 | 50 | 1-147 | 2-395 | 1-434 |
| 19 | 1-745 | 1-247 | 2-724 | 51 | 2-396 | 2-629 | 1-957 |
| 20 | 1-344 | 2-370 | 1-355 | 52 | 1-147 | 2-395 | 1-434 |
| 21 | 2-629 | 1-543 | 1-951 | 53 | 2-525 | 1-172 | 2-917 |
| 22 | 2-482 | 2-120 | 1-219 | 54 | 1-325 | 2-993 | 2-736 |
| 23 | 2-259 | 1-384 | 2-225 | 55 | 1-771 | 1-566 | 2-376 |
| 24 | 1-293 | 2-459 | 2-681 | 56 | 1-585 | 2-628 | 1-284 |
| 25 | 1-849 | 1-382 | 2-192 | 57 | 2-354 | 1-526 | 1-595 |
| 26 | 1-294 | 2-729 | 1-390 | 58 | 2-358 | 2-606 | 1-586 |
| 27 | 2-165 | 1-661 | 1-336 | 59 | 2-548 | 1-201 | 2-684 |
| 28 | 2-281 | 2-409 | 1-126 | 60 | 1-475 | 2-339 | 2-573 |
| 29 | 2-434 | 1-384 | 2-948 | 61 | 1-739 | 1-380 | 2-472 |
| 30 | 1-819 | 2-231 | 2-674 | 62 | 1-417 | 2-935 | 1-784 |
| 31 | 1-740 | 1-397 | 2-514 | 63 | 2-127 | 2-692 | 1-597 |
| 32 | 1-354 | 2-578 | 1-815 | 64 | 1-157 | 2-315 | 1-594 |

**Figure B.2 — Worksheet for Example 2**

| **Triangle test** | | | Test code: *587-FF03* |
|---|---|---|---|
| Taster No. *21* | Name: ______________________ | | Date: ____________ |
| Type of sample: *Candy bar* | | | |
| **Instructions:** Taste the samples on the tray from left to right. Two samples are alike; one is different. Select the odd/different sample and identify it by placing a X in the corresponding box | | | |
| **Samples tray** | Indicate odd sample | Remarks | |
| **629** | ☐ | ______________________ | |
| **543** | ☐ | ______________________ | |
| **951** | ☐ | ______________________ | |
| If you wish to comment on the reasons for your choice or on the characteristics of the sample, you may do so under Remarks. | | | |

**Figure B.3 — Scoresheet for Example 2**

### B.3.2 Analysis and interpretation of results

Consider the data from Example 2, above, where $x = 24$ and $n = 64$. It follows that:

— $p_c$ (proportion correct) = 24/64 = 0,375

— $\hat{p}_d$ (proportion distinguishers) = (1,5 × 0,375) − 0,5 = 0,062 5

— $s_d$ (standard deviation of $\hat{p}_d$) $= 1{,}5\sqrt{0{,}375\,(1-0{,}375)/64} = 0{,}090\,8$

— 90 % upper confidence limit: 0,062 5 + (1,28 × 0,090 8) = 0,18

— 90 % lower confidence limit: 0,062 5 − (1,28 × 0,090 8) = − 0,05

If the analyst were testing for similarity, the analyst could be 90 % confident that the actual proportion of the population that could distinguish the samples was no more than 18 %. On the other hand, if the analyst were testing for a difference, since the lower 90 % confidence limit is negative, $p_d = 0$ % is in the interval and is therefore a possible value, thus supporting the conclusion that there is no perceptible difference between the samples.

Taken together, the confidence intervals allow for 10 % error for both the upper and lower limits, so the sensory analyst can be 80 % confident that the true proportion of distinguishers is somewhere between 0 % and 18 % of the population. Depending on the objective of the study, the researcher may choose to use the one-sided upper confidence limit, the one-sided lower confidence limit, or the combined, two-sided confidence limits.

# Bibliography


[1] ISO 3534-1, *Statistics — Vocabulary and symbols — Part 1: Probability and general statistical terms*

[2] ISO 6658, *Sensory analysis — Methodology — General guidance*

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

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

[5] ISO 10399, *Sensory analysis — Methodology — Duo-trio test*

[6] BROCKHOFF, P.B. AND SCHLICH, P. Handling replications in discrimination tests. *Food Quality and Preference*, **9** (5), 1998, pp. 303-312

[7] ENNIS, D.M. AND BI, J. The Beta-Binomial Model: Accounting for Inter-trial Variation in Replicated Difference and Preference Tests. *Journal of Sensory Studies*, **13** (4), 1998, pp. 389-412

[8] FRIJTERS, J.E.R. Three-Stimulus Procedure in Olfactory Psycholophysics: An Experimental Comparison of Thurstone-Ura and Three-Alternative Forced-Choice Models of Signal Detection Theory. *Perception & Psychophysics*, **28** (5), 1980, pp. 390-397

[9] KUNERT, J. AND MEYNERS, M. On the triangle test with replications. *Food Quality and Preference*, **10**, 1999

[10] KUNERT, J. On repeated difference testing. *Food Quality and Preference*, **12**, 2001, pp 385-391

[11] MEILGAARD, M., CIVILLE, G.V., CARR, B.T. Sensory Evaluation Techniques, 2nd Edition, CRC Press, Inc., Boca Raton, FL, 1991, p. 338

[12] SCHLICH, P. Risk Tables for Discrimination Tests. *Food Quality and Preference*, **4**, 1993, pp. 141-151

ICS 67.240

Price based on 15 pages