INTERNATIONAL STANDARD

ISO 11056

First edition
1999-12-15

Sensory analysis — Methodology — Magnitude estimation method

Analyse sensorielle — Méthodologie — Méthode d'estimation de la grandeur

Reference number
ISO 11056:1999(E)

PDF disclaimer

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

Adobe is a trademark of Adobe Systems Incorporated.

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

© ISO 1999

All rights reserved. Unless otherwise specified, no part of this publication may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying and microfilm, without permission in writing from either ISO at the address below or ISO's member body in the country of the requester.

ISO copyright office
Case postale 56 • CH-1211 Geneva 20
Tel. + 41 22 749 01 11
Fax + 41 22 734 10 79
E-mail copyright@iso.ch
Web www.iso.ch

Printed in Switzerland

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 3.

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 International Standard may be the subject of patent rights. ISO shall not be held responsible for identifying any or all such patent rights.

International Standard ISO 11056 was prepared by Technical Committee ISO/TC 34, *Agricultural food products*, Subcommittee SC 12, *Sensory analysis*.

Annexes A and B of this International Standard are for information only.

Introduction

Magnitude estimation is a psychophysical scaling technique where assessors assign numerical values to the estimated magnitude of an attribute. The only constraint placed upon the assessor is that the values assigned should conform to a ratio principle; i.e. if the attribute appears to be twice as strong in sample B in comparison with sample A, the value assigned to sample B has to be twice that assigned to sample A. Attributes such as intensity, pleasantness or acceptability may be assessed using magnitude estimation.

Magnitude estimation can offer advantages over other scaling methods, in particular when the number of assessors and the time available for training are limited. Magnitude estimation offers a high degree of flexibility for both the panel leader and the assessor. Once trained in magnitude estimation, assessors are generally able to apply their skills to a wide range of samples and attributes, with minimal additional training.

Magnitude estimation method is also less susceptible to "end-effects" than those methods which employ continuous or discontinuous response scales. These "end-effects" occur when the assessors are unfamiliar with the extent of the sensations being presented. Then assessors might assign one of the initial samples to a category which is too close to one of the ends of the scale. Consequently, they then find themselves short of graduations and are obliged to classify samples perceived as being different into the same category. This should not occur with magnitude estimation since, in theory, there are an infinite number of categories.

Allowing each assessor to start the process at any numerical value, i.e. to use their own scale, gives rise to a particularly important "assessor" effect. However, there are various ways of solving this problem:

- the analysis of variance (ANOVA) allows the "assessor" effect and the interactions to be taken into account;
- the assessors can be forced to a common scale by use of a reference sample to which a value has been assigned;
- the data supplied by each assessor can be reduced to a common scale by applying one of numerous rescaling methods.

It is up to the experimenter to choose the most appropriate approach based on the circumstances.

The magnitude estimation method is not the most efficient technique for determining small differences between stimuli or for conducting assessments in the vicinity of a detection threshold.

Sensory analysis — Methodology — Magnitude estimation method

1 Scope

This International Standard describes a method for applying magnitude estimation to the evaluation of sensory attributes. The methodology specified covers the training of assessors, and obtaining magnitude estimations as well as their statistical interpretation.

2 Normative references

The following normative documents contain provisions which, through reference in this text, constitute provisions of this International Standard. For dated references, subsequent amendments to, or revisions of, any of these publications do not apply. However, parties to agreements based on this International Standard are encouraged to investigate the possibility of applying the most recent editions of the normative documents indicated below. For undated references, the latest edition of the normative document referred to applies. Members of ISO and IEC maintain registers of currently valid International Standards.

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

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

ISO 4121, *Sensory analysis — Methodology — Evaluation of food products by methods using scales.*

ISO 5492, *Sensory analysis — Vocabulary.*

ISO 6658, *Sensory analysis — Methodology — General guidance.*

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

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

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

3 Terms and definitions

For the purposes of this International Standard, the terms and definitions given in ISO 3534-1, ISO 3534-3, ISO 5492 and the following apply.

3.1
magnitude estimation
process of assigning values to the intensity of an attribute of the products or to their hedonic value in such a way that the ratio between the assigned values and the assessors' perception of the attributes are the same

3.2
external reference
sample which is presented to the assessor to which a numerical value is pre-assigned by the panel leader

NOTE It is the first sample of the series in relation to which all subsequent samples are then assessed.

3.3
modulus

numerical value assigned to the external reference which can be defined by the person conducting the test (fixed modulus) or left to the assessor to choose (non-fixed modulus)

3.4
rescaling

process which consists of multiplying the raw data supplied by each assessor by a factor which reduces the data of all the assessors to a common scale

NOTE Adding the logarithm of this factor to the logarithm of the raw data is an equivalent process.

3.5
internal reference

reference sample introduced into the test series and presented to the assessor as if it were a test sample

NOTE The value assigned to an internal reference may be used for rescaling an assessor's data. If an external reference is used, the internal reference(s) is(are) in principle identical to it.

3.6
Stevens's equation
psychophysical power function

relationship expressed as follows:

$$R = KS^n$$

where

R is the assessor's response (perceived intensity);

K is a constant which reconciles the units of measurement used for R and S;

S is the stimulus (concentration of a chemical substance or physical variable);

n is the exponent of the power function and the slope of the regression curve for R and S when they are expressed in logarithmic units

NOTE In practice, Stevens's equation is generally transformed into decimal or natural logarithms:

$$\ln R = \ln K + n \ln S$$

4 Principle

Samples are presented successively to assessors, who are requested to record the intensity of an attribute of each sample by complying with the ratio principle.

The values are assigned by referring to the value of the first sample (reference) of the series. For this first sample, either each assessor is free to assign a value to it, or the value is fixed by the person conducting the test. The latter case is called "fixed modulus".

5 General test conditions

For the general test conditions, such as those concerning the facilities, preparation, presentation and coding of samples, refer to International Standards on general methodology, in particular ISO 6658 and ISO 8589, or to those describing the methods which use scales and categories, such as ISO 4121.

6 Selection and training of assessors

6.1 General conditions for selection and training

These shall be in accordance with ISO 8586-1 and with ISO 8586-2.

As in all other sensory analysis methods, it is the responsibility of the panel leader to judge the required level of proficiency of the assessors. The objectives of the test, the availability of the assessors, the costs incurred by recruiting additional assessors, as well as their training, are to be taken into account when planning a training programme. Assessors are generally able to use the magnitude estimation methodology after three or four training tests.

6.2 Training specific to the magnitude estimation method

6.2.1 The assessment of surface areas of geometric shapes has been proved to be particularly suited for introducing assessors into the basic concepts of magnitude estimation. The following set of 18 shapes (see Table 1) comprising six circles, six equilateral triangles and six squares ranging in size from approximately 2 cm^2 to 200 cm^2 has been used successfully for training assessors.

Table 1 — Dimensions and areas of the training exercise shapes

Circles		**Triangles**		**Squares**	
Radius cm	**Surface area** cm^2	**Side** cm	**Surface area** cm^2	**Side** cm	**Surface area** cm^2
1,4	6,2	2,2	2,1	3,2	10,2
2,5	19,6	4,1	7,3	4,2	17,6
3,7	43,0	7,6	25,0	8,5	72,3
5,4	91,6	12,2	64,4	11,1 [a]	123,2
6,8	145,3	15,5	104,0	11,1 [a]	123,2
8,3	216,4	19,2	159,6	14,2	201,6

[a] Two 11,1 cm squares are introduced into the series in order to be able to evaluate the reproducibility of the assessors.

6.2.2 Prior to presenting the shapes to the assessors, instruct them in the principles of the method. This instruction shall include, but is not necessarily limited to, the following three points:

- the values shall be assigned on a ratio basis; if the attribute is twice as intense, a value twice as high shall be assigned to it;
- there is no upper limit to the scale;
- the value 0 shall be assigned only in the exceptional case where the attribute is not perceived.

Warn the assessors, at the time of training, that the general tendency is often to use round numbers (such as 5, 10, 20, 25, etc.) but that with this method all numbers are permitted and may be used. As assessors are also influenced by the ratios mentioned during training, always take care to suggest to them the use of different ratios, e.g. 3/1; 1/3; 7/5; 5/6 without limiting oneself to 2/1 or 1/2.

6.2.3 Assign codes to the shapes and present the shapes separately by placing them in the centre of a sheet of white paper of approximately A4 size (21 cm × 29,7 cm).

Instruct each assessor to conduct the magnitude estimation beginning the series with the presentation of the 8,5 cm square (external reference). Record the responses.

Depending on the procedure adopted for the test phase, train the assessors with or without a fixed modulus. With a fixed modulus, the person conducting the test assigns a value between 30 and 100 to the 8,5 cm square.

Without a fixed modulus, leave the assessors to assign the value of their choice to the first figure, but advise them not to choose too small a value.

Randomly present the geometric shapes prior to each test, so that their shapes and dimensions do not form a particular pattern.

6.2.4 After completing the assessment of the set of shapes, allow the assessors to compare their results with the average results of the group. If this is not practical, carry out this comparison with respect to the results obtained by a previous group.

NOTE The objective is to provide positive feedback to reassure the panelists that they understand the exercise. Care should be taken not to create the impression that there is a "right" answer. Unless their results are very different, departures from the group results should be explained as order effects; that is, their responses are affected by the order in which they evaluate the samples. They should be reassured that despite individual order effects, the group’s results will be accurate.

If the results of some assessors are very different, explain once again to these assessors the principles of the method.

6.2.5 When an assessor has successfully completed the area estimation exercise, further training should be given based on the product or type of substance which will be assessed in the actual test. This gives the assessor experience in applying magnitude estimation to attributes characterizing the test substance. The panel leader may need to design exercises for training panelists to identify correctly the attributes to be evaluated. This training may be drawn up using the general guidelines given in ISO 8586-1 and in ISO 8586-2.

7 Number of assessors

7.1 General

As for the other methods employing scales, the required number of assessors depends on:

- how close together the various test products are in the attribute being evaluated;
- the training received by the assessors;
- the importance which will be attached to the decision following the test results (see ISO 8586-1 and ISO 8586-2);
- the objectives which can be identified in terms of statistical power.

In the absence of such distinctly identifiable objectives, refer to the recommendations given in 7.2 and 7.3.

7.2 Analytical and research panels

The panels shall be made up as given in Table 2.

Table 2 — Formation of panels

Types of assessors	Minimum number of assessors	Recommended number
Experienced assessors, highly trained in the product and in the assessment of the attribute being studied	5	10
Experienced assessors, trained in the product and in the assessment of the attribute being studied	15	20 to 25
Newly trained assessors	20	20 and over

Issues of statistical power need to be resolved on the variance of individual evaluations and the magnitude of the differences which need to be detected.

7.3 Consumer panels

The magnitude estimation method can also be used with consumer panels or for conducting market research studies. The number of persons to be selected shall then be determined on the basis of the population sampling requirements connected with these types of tests. The use of the magnitude estimation method does not offer any particular advantage in terms of the number of assessors required, and this number shall be the same as for a typical consumer type test, namely at least 50 persons, and often much more.

NOTE For consumers, the training will be limited to area estimations.

8 Procedure

8.1 Presentation of samples

All the samples shall be presented in an identical manner (i.e. identical serving vessels, same quantity of product).

The vessels containing the samples shall be coded, preferably using randomly selected three-digit numbers.

8.2 External reference sample

The external reference sample shall be the same for all the assessors. It is desirable that the reference sample has, for the attribute being studied, an intensity which is close to that of the geometric mean of all of the products under test.

NOTE A reference which might present an extreme value for the attribute would introduce a distortion, as far as the difference between products permit, and would reduce the sensitivity of the method.

One or more randomly coded reference samples may be included in the test series. This allows assessment of the repeatability of the assessor within the session.

8.3 Order of presentation of samples

The samples shall be presented all at once or in a sequential way to the assessors. The assessors shall follow the order indicated. As in all sensory tests, this order differs from one assessor to the other, the ideal situation being that the orders of the samples are balanced. The panel leader can refer to tables proposed in reference [1] which uses Latin squares to balance the design for order and carryover effects. Where this is not possible, use random order.

8.4 Magnitude estimations

8.4.1 General

Carry out the test in accordance with one of the techniques described in 8.4.2 to 8.4.4.

8.4.2 Without fixed modulus for the external reference

Each assessor evaluates the reference and assigns a value to it. Advise the assessors not to choose too small a value.

The assessor then evaluates each subsequent coded sample, comparing it with the reference, and assigns to it a value in relation to that which he/she has previously assigned to the reference.

8.4.3 With fixed modulus for the external reference

The panel leader specifies to the assessor that the reference sample has a value of, for example, 30, 50, 100 or whatever seems appropriate to the panel leader.

The leader instructs the assessor to make his or her subsequent judgements relative to the value assigned to the reference (fixed modulus).

8.4.4 Without external reference

It is possible to use magnitude estimation methodology without using any external reference sample. Due to the limits of the sensory systems (adaptation), it may be difficult for the assessors to refer systematically to the first sample. There are two possible cases, as follows.

a) The assessors are not forced to re-evaluate the first sample prior to evaluating each of the subsequent samples.

   It is then advisable to encourage the assessors to memorize the degree of the attribute being studied for this reference sample and to re-evaluate this reference if it appears necessary to them.

   NOTE The problem which arises then is that, for a given assessor, the variances of the differences between any sample and the first sample are twice as small as the variances of the differences between any two samples.

   It is therefore possible:

   — **prior to the test**, to choose the presentation design in such a way that the first samples are not the same for each of the assessors; ideally, each sample should be used as a reference for an equal number of assessors; the variances of the mean differences between samples will therefore be equal;

   — **at the time of analysis**, to apply an arbitrarily very high (theoretically infinite) weighting to the evaluations of the first samples of each assessor, so that the variances of the differences are correctly estimated.

b) The assessors are asked to evaluate each sample by comparing it to the immediately preceding sample.

NOTE The problem which then arises is that for each assessor the evaluation errors are autocorrelated; the variance of the difference of two successive samples will be smaller than the variance of the difference of two non-successive samples.

It is therefore possible:

— **prior to the test,** to choose the presentation design in such a way that all the possible permutations of the samples are presented to an equal number of assessors; if this is not possible, try to propose orders which approximate best this ideal model; the variances of the mean differences between samples will then be equal or, at least, fairly close;

— **at the time of analysis**, to employ autocorrelated error models, the methodology of which is, however, slightly more complicated.

It is to be noted that even if one proceeds as proposed in the case a) (systematic comparison with the first sample in order to carry out the evaluation), an autocorrelation term, linked to the evaluation of the preceding sample, however small it may be, very probably remains (this is also true, incidentally, for the tests with reference described in 8.4.2 and 8.4.3). The advice given earlier that the orders of the samples are balanced is therefore valid in all cases.

9 Analysis of data

9.1 Choice of data analysis method

See Table 3.

9.2 Presentation of raw results

The results may be presented in the form of a dual-entry table, placing horizontally the responses of the assessors after logarithmic transformation, and vertically the different samples.

When all the assessors have given a score the same number of times for each of the samples, a complete balanced plan is obtained and the model together with the assessor effect is orthogonal. If certain products have not been evaluated the same number of times by all the assessors, an incomplete plan is obtained and the model together with the assessor effect is non-orthogonal.

NOTE Since one cannot take the logarithm of zero, any zero response causes a problem. Different approaches have been used to deal with zeros. Zero values should be replaced by very small values. The specific value chosen should take into account the scale used by each panelist (e.g. half of the smallest value assigned by that panelist).

9.3 Establishment of product differences

An analysis of the variance (ANOVA) which explicitly accounts for all blocking factors (including unbalanced or non-orthogonal factors) and is carried out on logarithmically transformed data is the most accurate method. However, not all statistical software allow analysis of unbalanced designs. Nevertheless, in practice, it is not always possible to conduct an experiment leading to a complete design where all critical factors are balanced and orthogonal. For example, when a project extends over multiple sessions, it may be not possible to assemble exactly the same group of panelists at each session. It is always advisable to ask a statistician to set up the best possible experimental design. When it is not possible to conduct an analysis of the variance taking into account all the factors of the experimental design, it is possible to use one of the rescaling techniques indicated in 9.5. In such cases, the option consisting of carrying out a one-way analysis of variance on the rescaled data using treatment as the only factor is a less desirable but workable solution which is not rigorous but may possibly constitute a rough approximation of the results. If the experimental design is not a standard one, the best solution is, however, to consult a statistician in order to develop an appropriate form of the ANOVA model.

When significant differences between products are revealed by analysis of variance (ANOVA), one of the usual tests of multiple comparison of means is then carried out.

An example of comparison of products without rescaled data on a complete design is given in A.1.

Table 3 — Presentation of the data analysis methods

No. of replicates	Plan	Processing of data	Model	Search for differences between products			Regression		
				Sources of variation	**Degrees of freedom**	**Example**	**Sources of variation**	**Degrees of freedom**	**Example**
1	Complete $n = st$	Logarithm	Orthogonal	Assessor	$s - 1$	A.1	Assessor	$s - 1$	A.5
				Treatment	$t - 1$		Log concentration	1	
				Error	$(s - 1)(t - 1)$		Assessor–log concentration	$s - 1$	
							Error	$s(t - 2)$	
⩾ 2	Complete $n = str$	Logarithm	Orthogonal	A simple possibility is to calculate first of all the mean of the logarithms of the values given by each assessor for a same treatment. One thus comes down to the case $r = 1$ described above. (It is also possible to use more complicated models taking into account random effects, testing assessor–treatment interactions, or making assumptions on the variance/covariance matrix of the replicates of a same assessor.)		A.6	Error	$s(t - 2)$	
1	Incomplete $n \neq st$	Logarithm	Non-orthogonal	Assessor	$s - 1$				
				Treatment	$t - 1$				
				Error	$n - t - s + 1$				
1	Incomplete All assessors evaluated the same sub-group of samples	Total rescaling then logarithm	Orthogonal	Treatment	$t - 1$	A.2			
				Error (after correction)	$n - t - s + 1$				
1	Incomplete or complete The assessors evaluated a reference identified as such or a hidden reference	Rescaling in relation to the reference, then logarithm	Orthogonal	Treatment	$t - 1$	A.3			
				Error (after correction)	$n - t - s + 1$				
1	Incomplete The assessors assigned the scores to a verbal scale	External rescaling then logarithm	Orthogonal	Treatment	$t - 1$	A.4			
				Error (after correction)	$n - t - s + 1$				

NOTE The symbols are as follows:

t is the number of evaluated treatments;

n is the total number of measurements;

s is the number of assessors having participated in the evaluation;

r is the number of replicates conducted.

9.4 Regression

In the case where the values of a related variable S (such as concentration, physical quantity) are known as being capable of relating to the response R, it is possible to assume that Stevens's law is followed and to estimate its parameters by carrying out the linear regression of the sensory responses regarding this physical or chemical variable, according to the equation:

$$\ln R = \ln K + n \ln S$$

In such an analysis, the parameter which is of greatest interest is the slope which corresponds to the value n in Stevens's equation.

The equality of the slopes of the regression between the different assessors can also be tested.

Table 3 presents the models for variance and regression analysis in different cases.

9.5 Rescaling methods

NOTE As previously stated, these methods are only of interest if the design is incomplete. Rescaling is a method for estimating the assessor effect.

9.5.1 Total rescaling

Total rescaling is equivalent to the usual estimation of the assessor effect in the analysis of variance if the design is balanced.

This procedure can only be applied if all the assessors have evaluated the same set of samples (all of the samples in a balanced design, a sub-set in the unbalanced design).

The reasoning on which this method is based is as follows. Since each assessor has evaluated the same set of samples, the total magnitude of the response for this set of samples should be identical for each assessor. Therefore, the scale for each assessor is brought to the same total magnitude for a set of common samples.

If this method is used, the number of degrees of freedom of the error should be decreased by the number of assessors minus 1. The procedure is as follows.

For all of the samples evaluated by all of the assessors:

- calculate the mean of the logarithm of the estimations of each assessor;
- calculate the general mean for all the assessors;
- for each assessor, calculate the correction value which, once added to its mean, will make it equal to the mean of the group;
- add its correction value to all the estimations of each assessor.

See an example of carrying out total rescaling in annex A, Tables A.3 and A.4.

9.5.2 Rescaling in relation to the reference sample

If one or more reference samples, randomly coded, have been incorporated into the test series, first calculate for each assessor the mean of the estimations relating to the reference samples [first sample and hidden reference(s), if any]. Then, calculate the correction value which would bring this mean to a fixed value. In order to rescale the data obtained for the test samples, multiply each estimation of an assessor by the correction value calculated from the reference sample(s).

NOTE It is advisable to note that the global analysis of variance as well as the procedure for total rescaling gives rise to a smaller mean square error than the procedure of rescaling in relation to the reference sample. As indicated in 8.2, the reference sample should have an intensity close to the geometric mean of all the samples for the whole panel. It has been shown (see reference [3]) that the error is lower when the intensity of the reference sample is equal to the geometric mean. The closer the value for the reference sample is to the true geometric mean, the better.

9.5.3 External rescaling

Various forms of external rescaling have been reported in the literature. After evaluating the test samples, the assessor receives a verbal response scale comprising between four and eleven graduations. It consists of expressions such as:

- extremely intense;
- very intense;
- moderately intense;
- slightly intense, etc.

The panel leader requests the assessor to assign magnitude estimations to these expressions in a manner which is consistent with the scale being used when evaluating the test samples. Results given by each assessor are rescaled using a correction value calculated by applying the total rescaling method to the values assigned to the expressions of the verbal response scale.

See an example of external rescaling in annex A, Tables A.6, A.7 and A.8.

10 Test report

The test report shall indicate:

- the objectives of the study;
- the test results;
- the number of samples;
- recourse, if any, to a reference sample and, if used, the nature of this sample;
- replication, if any, of the tests;
- the number of assessors and their level of qualification;
- the general conditions of testing, such as test environment, date, time;
- any other information allowing the overall validity of the tests to be evaluated;
- the reference to the number of this International Standard, together with an indication of adjustments, if any, made to the method;
- the name of the person in charge of the tests.

Annex A
(informative)

Examples of data analysis

A.1 Data analysis and interpretation using analysis of variance (ANOVA)

NOTE No replication; without rescaled data.

A.1.1 Case presented

Table A.1 presents the results obtained by seven experienced assessors having evaluated the intensity of the bitterness of six samples of a beverage containing various quantities of caffeine.

Natural logarithms (ln) have been calculated and are presented in the table in parentheses.

Table A.1 — Table of data relative to the six samples

Assessor	Treatment codes					
	561	**274**	**935**	**803**	**417**	**127**
	Concentrations (mg/100 ml)					
	9	18	36	40	72	144
	Magnitude estimations (logarithms)					
1	10 (2,303)	20 (2,996)	35 (3,555)	40 (3,689)	70 (4,248)	140 (4,942)
2	8 (2,079)	20 (2,996)	38 (3,638)	44 (3,784)	85 (4,443)	160 (5,075)
3	8 (2,079)	20 (2,996)	36 (3,584)	40 (3,689)	75 (4,317)	150 (5,010)
4	7 (1,946)	15 (2,708)	32 (3,466)	37 (3,611)	70 (4,248)	135 (4,905)
5	12 (2,485)	25 (3,219)	38 (3,638)	40 (3,689)	75 (4,317)	145 (4,977)
6	12 (2,485)	22 (3,091)	35 (3,555)	40 (3,689)	80 (4,382)	160 (5,075)
7	9 (2,197)	18 (2,890)	35 (3,555)	40 (3,689)	74 (4,304)	145 (4,977)
Mean log	2,225	2,985	3,570	3,691	4,323	4,994

A.1.2 Determination of the existence of significant differences

A two-way analysis of variance was applied to the above logarithms. The results are given in Table A.2

Table A.2 — Results of two-way analysis of variance

Source of variation	Degrees of freedom	Sum of squares	Mean square	*F* value	*P* > *F*
Assessor	6	0,240	0,040	4,55	0,002 1
Treatment	5	33,177	6,635	754,69	0,000 1
Error after correction	30	0,264	0,009	—	—

Analysis of variance shows that the effect induced by the treatment is significant. Tukey's test is one of the numerous multiple comparison tests likely to be used for determining which samples differ significantly from one another. In this test, the least significant difference is determined as:

$$\sqrt{s^2 \times \frac{1}{2}\left(\frac{1}{n_\mathrm{i}}+\frac{1}{n_\mathrm{j}}\right)} \times c$$

where

s^2 is the mean square of error after correction;

n_i is the number of observations used for the calculation of the 1st mean;

n_j is the number of observations used for the calculation of the 2nd mean;

c is a factor; a function of the degrees of freedom of error after correction, of the total number of treatments, at the α-level; obtained from Table A.6 of reference [2].

In this example (six treatments and 30 degrees of freedom for error), the least significant difference is the standard error of the mean $\left(\sqrt{0{,}009/7}\right)$ multiplied by 4,30; i.e. 0,154.

The only two samples not differing in any significant manner are 803 and 935. Their means differ by only 0,12.

A.2 Data analysis and interpretation using analysis of variance (ANOVA) in the case of total rescaling

When all the assessors have evaluated samples 561, 935, 803 and 127, the total rescaling can be carried out on this sub-group of samples.

Determine the rescaling factors by calculating initially for each assessor the mean logarithm of the estimations for the four common samples (Table A.3). Next calculate the mean logarithm of the estimations for the whole panel. Finally, calculate the correction value for each assessor by subtracting the mean per assessor from the mean of the group.

Table A.3 — Logarithms of the estimations provided by the assessors and calculation of the correction value

Assessor	Treatment codes						Mean of the common sub-group	Correction factor
	561	274	935	803	417	127		
	Magnitude estimations (logarithms)							
1	2,303	2,996	3,555	3,689	4,248	4,942	3,622	– 0,002
2	2,079		3,638	3,784	4,443	5,075	3,644	– 0,024
3	2,079	2,996	3,584	3,689		5,011	3,591	+ 0,029
4	1,946		3,466	3,611	4,248	4,905	3,482	+ 0,138
5	2,485	3,219	3,638	3,689		4,977	3,697	– 0,077
6	2,485		3,555	3,689	4,382	5,075	3,701	– 0,081
7	2,197	2,890	3,555	3,689		4,977	3,604	+ 0,016
Group							3,620	

After correction, one obtains the values given in Table A.4.

Table A.4 — Log (estimations) after correction

Assessor	**Treatment codes**					
	561	**274**	**935**	**803**	**417**	**127**
	Magnitude estimations (logarithms)					
1	2,301	2,994	3,553	3,687	4,246	4,940
2	2,055		3,614	3,760	4,419	5,051
3	2,108	3,025	3,613	3,718		5,040
4	2,084		3,604	3,749	4,386	5,043
5	2,408	3,142	3,561	3,612		4,900
6	2,404		3,474	3,608	4,301	4,994
7	2,213	2,906	3,571	3,705		4,993
Mean log	2,225	3,017	3,570	3,691	4,338	4,994

Analysis of variance (ANOVA) applied to this data gives the results shown in Table A.5.

Table A.5 — Results of analysis of variance (ANOVA)

Source of variation	**Degrees of freedom**	**Sum of squares**	**Mean square**	***F* value**	*P* > *F*
Treatment	5	30,414	6,083	608,30	0,000 1
Error after correction	24	0,237	0,010		

For the calculation of the least significant difference, it is necessary to take into account the fact that the means of the different treatments are not obtained with the same number of observations (four or seven assessors as appropriate). The Tukey-Kramer method (see reference [4]) may be used.

Thus the least significant difference is equal to:

$$\sqrt{0{,}010 \times \frac{1}{2}\left(\frac{1}{4}+\frac{1}{4}\right)} \times 4{,}37 = 0{,}219$$

for the pair (274, 417), which concerns treatments for which four measurements are available. (The constant 4,37 is obtained in the same table as previously with a corrected number of degrees of freedom.)

The least significant difference is equal to:

$$\sqrt{0{,}010 \times \frac{1}{2}\left(\frac{1}{7}+\frac{1}{4}\right)} \times 4{,}37 = 0{,}194$$

for the pairs (561, 274), (935, 274), (803, 274), (127, 274), (561, 417), (935, 417), (803, 417) and (127, 417) for which one treatment formed the subject of seven measurements, and the other treatment only four measurements.

The least significant difference is equal to:

$$\sqrt{0{,}010 \times \frac{1}{2}\left(\frac{1}{7}+\frac{1}{7}\right)} \times 4{,}37 = 0{,}165$$

for all the other pairs which concern treatments for which seven measurements are available.

Thus, only the treatments 803 and 935 are not significantly different.

A.3 Data analysis and interpretation using analysis of variance (ANOVA) in the case of external rescaling

After carrying out the main experiment, the assessors are requested to assign magnitude values to a verbal calibration scale. For illustrating this, a five-point scale ranging from "extremely bitter" to "slightly bitter" has been created. The panel leader requests the assessor to assign magnitude estimations to these expressions in a manner which is consistent with the scale being used when evaluating the test samples. The hypothetical results for this exercise are presented in Table A.6.

Table A.6 — Hypothetical results of ratings for the verbal scale

Assessor	Slightly bitter	Bitter	Moderately bitter	Very bitter	Extremely bitter	Rescaling factor [a]
1	5	25	50	100	150	– 0,002 5
2	5	30	60	100	160	+ 0,088 0
3	5	25	50	100	150	– 0,002 5
4	5	20	45	90	140	+ 0,098 0
5	5	25	50	100	150	– 0,002 5
6	3	30	55	110	170	0
7	5	25	50	100	150	– 0,002 5

[a] Calculated by the total rescaling method.

First calculate the correction values using the total rescaling method applied to the calibration scores. Next correct the data of Table A.7 by adding to the data of each assessor his/her correction value.

Table A.7 — Log (estimations) given by the assessors

Assessor	Treatment codes					
	561	274	935	803	417	127
1	2,303	2,996	3,555	3,689	4,248	4,942
2		2,996	3,638	3,784	4,443	5,075
3	2,079		3,584	3,689	4,317	5,011
4	1,946	2,708		3,611	4,248	4,905
5	2,485	3,219	3,638		4,317	4,977
6	2,485	3,091	3,555	3,689		5,075
7	2,197	2,890	3,555	3,689	4,304	

Table A.8 is then obtained.

Table A.8 — Log (estimations) after correction

Assessor	**Treatment codes**					
	561	**274**	**935**	**803**	**417**	**127**
1	2,300	2,993	3,553	3,686	4,246	4,939
2		2,908	3,550	3,696	4,355	4,987
3	2,077		3,581	3,686	4,315	5,008
4	2,044	2,806		3,709	4,346	5,003
5	2,482	3,216	3,635		4,315	4,974
6	2,485	3,091	3,555	3,689		5,075
7	2,195	2,888	3,553	3,686	4,302	
Group	2,264	2,984	3,571	3,692	4,313	4,998

Table A.9 — Results of analysis of variance (ANOVA)

Source of variation	**Degrees of freedom**	**Sum of squares**	**Mean square**	F **value**	$P > F$
Treatment	5	27,773	5,555	427,31	0,000 1
Error after correction	24	0,322	0,013		

The least significant difference calculated as described in A.1.2 for six treatments and 24 degrees of freedom amounts to 0,203 and the only treatments which do not differ significantly are 935 and 803.

A.4 Examination of the slope of the regression curve

If the caffeine concentration of the samples increases and the contents are known, it is possible to apply a linear regression to both the logarithms of the concentrations and of the magnitude estimations in order to assess the slope of the regression curve. If the magnitude estimations have not been rescaled, it is necessary to provide for different ordinates for the different assessors.

It is recommended that the assessor–log concentration interaction is tested to see whether the slopes are equal for each assessor.

The analysis of variance (ANOVA) gives the results shown in Table A.10.

Table A.10 — Results of analysis of variance (ANOVA) for assessor–log concentration

Source of variation	**Degrees of freedom**	**Sum of squares**	**Mean square**	F **value**	$P > F$
Assessor	6	0,240	0,040	8,09	0,000 1
Log concentration	1	33,129	33,129	6692,26	0,000 1
Assessor–log concentration	6	0,173	0,029	5,81	0,000 5
Error	28	0,139	0,005	—	—

As the assessor–log concentration effect is highly significant, it is necessary to perform a regression for each subject in order to calculate individual slopes. The values given in Table A.11 are thus obtained.

Table A.11 — Regression curve

Assessor	Slope		r^2
	Estimation	**Standard error**	
1	0,94	0,02	0,998
2	1,07	0,03	0,995
3	1,04	0,03	0,995
4	1,08	0,02	0,998
5	0,87	0,04	0,988
6	0,93	0,05	0,986
7	1,00	0,01	0,999

It should be remarked that although the interaction is significant, there is little difference in the slopes. The mean is calculated to be 0,99 with a standard error of 0,03.

A.5 Data analysis and interpretation using analysis of variance (ANOVA) in the case of replications

The Table A.12 presents a list of the replicated magnitude estimations assigned by the same members of the panel as those whose results are given in Table A.1.

Table A.12 — Estimations for replications 1 and 2

Assessor	Replicate	Treatment codes					
		561	**274**	**935**	**803**	**417**	**127**
		Magnitude estimations (logarithms)					
1	1	10 (2,303)	20 (2,996)	35 (3,555)	40 (3,689)	70 (4,248)	140(4,942)
	2	15 (2,708)	25 (3,219)	35 (3,555)	38 (3,638)	70 (4,248)	135 (4,905)
2	1	8 (2,079)	20 (2,996)	38 (3,638)	44 (3,784)	85 (4,443)	160 (5,075)
	2	8 (2,079)	15 (2,708)	35 (3,555)	45 (3,807)	90 (4,500)	180 (5,193)
3	1	8 (2,079)	20 (2,996)	36 (3,584)	40 (3,689)	75 (4,317)	150(5,011)
	2	10 (2,303)	20 (2,996)	35 (3,555)	35 (3,555)	70 (4,248)	145 (4,977)
4	1	7 (1,946)	15 (2,708)	32 (3,466)	37 (3,611)	70 (4,248)	135 (4,905)
	2	10 (2,303)	20 (2,996)	35 (3,555)	38 (3,638)	65 (4,174)	130 (4,868)
5	1	12 (2,485)	25 (3,219)	38 (3,638)	40 (3,689)	75 (4,317)	145 (4,977)
	2	10 (2,303)	25 (3,219)	35 (3,555)	40 (3,689)	80 (4,382)	150 (5,011)
6	1	12 (2,485)	22 (3,091)	35 (3,555)	40 (3,689)	80 (4,317)	160 (5,075)
	2	10 (2,303)	20 (2,996)	35 (3,555)	40 (3,689)	80 (4,382)	160 (5,075)
7	1	9 (2,197)	18 (2,890)	35 (3,555)	40 (3,689)	74 (4,304)	145 (4,977)
	2	10 (2,303)	15 (2,708)	35 (3,555)	38 (3,638)	70 (4,248)	140 (4,942)
Mean log		2,277	2,981	3,562	3,678	4,313	4,995

One possibility is to calculate the mean for each assessor and each treatment and to conduct the analysis of variance as in A.1. Another approach is to use a model where the assessor effect and the assessor–treatment interaction are considered as random factors.

Table A.13 — Results of analysis of variance (ANOVA) for assessor–treatment interaction

Source of variation	Degrees of freedom	Sum of squares	Mean square	F value	$P > F$
Assessor	6	0,134	0,022	2,46	0,047
Treatment	5	32,121	6,424	706,26	0,000 1
Error	30	0,273	0,009	—	—

The least significant difference for the six samples and 30 degrees of freedom amounts to 0,155. The only two samples which do not differ significantly are 803 and 935.

Annex B
(informative)

Questionnaire models

B.1 Questionnaire model without fixed modulus for the reference sample

Surname: .... Date: ....

First name: ....

1) A reference sample of orange juice coded "R" is presented to you

You are requested to taste it and to rate the intensity of its acid taste with the aid of a number of your choice:

Response :

Memorize well the intensity of its acidity.

2) Six orange drinks are presented to you

You are requested to evaluate them in the order given below.

For each of the samples, assign a value to the intensity of the acid taste, in proportion to the value of the reference "R".

You have to retaste the reference before each sample.

Sample 561 ☐

Sample 274 ☐

Sample 935 ☐

Sample 803 ☐

Sample 417 ☐

Sample 127 ☐

B.2 Questionnaire model with fixed modulus for the reference sample

Surname: .... Date: ....

First name: ....

1) A reference sample of orange juice coded "R" is presented to you

The value assigned to its "acidity" is equal to 50.

Taste this sample and memorize the intensity of its acidity.

2) Six orange drinks are presented to you

You are requested to evaluate them in the order indicated below.

For each of the samples, assign a value to the intensity of the acid taste, in proportion to the value (50) given to the reference "R".

You have to retaste the reference before each sample.

Sample 561 ☐

Sample 274 ☐

Sample 935 ☐

Sample 803 ☐

Sample 417 ☐

Sample 127 ☐

Bibliography

[1] MACFIE, H.J., BRATCHELL, N., GREENHOFF, K. and VALLIS, L.V. Designs to balance the effect of order of presentation and first-order carry-over effects in hall tests. *Journal of Sensory Studies*, **4**, 2, 1989, pp. 129-148.

[2] GACULA, M.C. Jr. and SINGH, J. *Statistical methods in food and consumer research*. Academic Press, Orlando, USA, 1984.

[3] POWERS, J.J., WARE, G.O. and SHINHOLSTER, K.J. Magnitude estimation with and without rescaling. *Journal of Sensory Studies*, **5**, 1990, pp. 105-116.

[4] KRAMER, C.Y. Extension of multiple range tests to group means with unequal numbers of replications. *Biometrics*, **12**, 1956, pp. 307-310.

ICS 67.240

Price based on 20 pages