Designation: E 1697 – 95 (Reapproved 1999)

Standard Test Method for Unipolar Magnitude Estimation of Sensory Attributes[1]

1. Scope

1.1 This test method describes a procedure for the application of unipolar magnitude estimation to the evaluation of the magnitude of sensory attributes. The test method covers procedures for the training of panelists to produce magnitude estimations and statistical evaluation of the estimations.

1.2 Magnitude estimation is a psychophysical scaling technique in which panelists assign numeric values to the magnitude of an attribute. The only constraint placed upon the panelist is that the values assigned should conform to a ratio principle. For example, if the attribute seems twice as strong in sample B when compared to sample A, sample B should receive a value which is twice the value assigned to sample A.

1.3 The intensity of attributes such as pleasantness, sweetness, saltiness or softness can be evaluated using magnitude estimation.

1.4 Magnitude estimation may provide advantages over other scaling methods, particularly when the number of panelists and the time available for training are limited. With approximately 1 h of training, a panel of 15 to 20 naive individuals can produce data of adequate precision and reproducibility. Any additional training that may be required to ensure that the panelists can properly identify the attribute being evaluated is beyond the scope of this test method.

2. Referenced Documents

2.1 *ASTM Publications:*

STP 434 Manual on Sensory Testing Methods[2]

STP 758 Guidelines for the Selection and Training of Sensory Panel Members[2]

STP 913 Physical Requirement Guidelines for Sensory Evaluation Laboratories[2]

2.2 *ISO Standards:*

ISO 3534:1977 Statistics—Vocabulary and Symbols[3]

ISO 3534-3:1985 Statistics—Vocabulary and Symbols—Part 3: Experimental Design[3]

ISO 4121:1987 Sensory Analysis—Methodology—Evaluation of Food Products by Methods Using Scales[3]

ISO/DIS 5492:1990 Sensory Analysis—Vocabulary (1)[3]

ISO 6658:1985 Sensory Analysis—Methodology—General Guidance[3]

ISO/DIS 8586-1:1989 Sensory Analysis—Methodology—General Guide for Selection, Training and Monitoring Subjects—Part 1: Qualifying Subjects (1)[3]

ISO 8589:1988 Sensory Analysis—General Guidance for the Design of Test Rooms[3]

3. Terminology

3.1 *Definitions:*

3.1.1 *external modulus*—a number assigned by the panel leader to describe the intensity of the external reference sample or the first sample of the sample set. The external modulus is sometimes referred to as a "fixed modulus" or just the "modulus." In this case the reference is said to be modulated.

3.1.2 *external reference sample for magnitude estimation*—a sample designated as the one to which all others are to be compared, or to which the first sample of a set is to be compared, when each subsequent sample in the set is compared to the preceding sample. This sample is normally the first sample to be presented.

3.1.3 *internal modulus*—a number assigned by the panelist to describe the intensity of the external reference sample or the first sample of the sample set. The internal modulus is sometimes referred to as a "non-fixed modulus." When an internal modulus is used, the reference is sometimes said to be unmodulated.

3.1.4 *internal reference sample for magnitude estimation*—a sample present in the experimental set, which is presented to the panelist as if it were a test sample. The value assigned to this sample(s) can be used for normalizing panelists' data. If an external reference is used, the internal reference(s) are normally identical to it.

3.1.5 *magnitude estimation*—the process of assigning values to the intensities of an attribute of products in such a way that the ratios of the values assigned and the panelist's perceptions of the attribute are the same.

---

[1] This test method is under the jurisdiction of ASTM Committee E18 on Sensory Evaluation of Materials and Products and is the direct responsibility of Subcommittee E18.03 on Sensory Theory and Statistics.

Current edition approved March 15, 1995. Published May 1995.

[2] Available from ASTM Headquarters.

[3] Available from American National Standards Institute, 11 W. 42nd St., 13th Floor, New York, NY 10036.

3.1.6 *normalizing*—The process of multiplying each panelist's raw data by, or adding to the logarithm of each panelist's raw data, a value which brings all the data onto a common scale. Also referred to as rescaling.

3.1.7 *Stevens' Equation or the Psychophysical Power Function*—

$$R = KS^n \tag{1}$$

where:

R = the panelist's response (the perceived intensity),
K = a constant that reconciles the units of measurement used for R and S,
S = the stimulus (chemical concentration or physical force), and
n = the exponent of the power function and the slope of the regression curve for R and S when they are expressed in logarithmic units.

In practice, Stevens' equation is generally transformed to logarithms, either common or natural

$$lnR = lnK + nlnS \tag{2}$$

4. Summary of Test Method

4.1 Panelists judge the intensity of an attribute of a set of samples, presented in random order, on a ratio scale. For example, if one sample is given a value of 50 and a second sample is twice as strong, it will be given a value of 100. If it is half as strong it will be given a value of 25. There are three procedures that can be used.

4.1.1 Panelists are instructed to assign any value to describe the intensity of the first sample (external reference, which may or may not be part of the sample set). Panelists then rate the intensity of the following samples in relation to the value of the external reference.

4.1.2 The external reference is pre-assigned a value (modulus) to describe its intensity by the panel leader. Panelists rate the intensity of the following samples in relation to the external reference and the modulus.

4.1.3 Panelists rate the intensity of each subsequent sample in relation to the preceding sample. The first sample of the set may or may not have a modulus.

4.2 Individual judgments can be converted to a common scale by normalizing the data. Three normalizing methods can be used: internal standard normalizing, external calibration and, if a modulus is not used, no standard normalizing (method of averages). See 11.4 and Appendix X2-Appendix X4.

4.3 Results are averaged using geometric means. Analysis of variance or other statistical analyses may be performed after the data have been converted to logarithms.

5. Significance and Use

5.1 Magnitude estimation may be used to measure and compare the intensities of attributes of a wide variety of products.

5.2 Magnitude estimation provides a large degree of flexibility for both the experimenter and the panelist. Once trained in magnitude estimation, panelists are generally able to apply their skill to a wide variety of sample types and attributes, with minimal additional training.

5.3 Magnitude estimation is not as susceptible to end-effects as interval scaling techniques. These can occur when panelists are not familiar with the entire range of sensations being presented. Under these circumstances, panelists may assign an early sample to a category which is too close to one end of the scale. Subsequently, they may "run out of scale" and be forced to assign perceptually different samples to the same category. This should not occur with magnitude estimation, as, in theory, there are an infinite number of categories.

5.4 Magnitude estimation is one frequently used technique that permits the representation of data in terms of Stevens' Power Law.

5.5 The disadvantages of magnitude estimation arise primarily from the requirements of the data analysis.

5.5.1 Permitting each panelist to choose a different numerical scale may produce significant panelist effects. This disadvantage can be overcome in a number of ways, as follows. The experimenter must choose the approach most appropriate for the circumstances.

5.5.1.1 Experiments can be designed such that analysis of variance can be used to remove the panelist effects and interactions.

5.5.1.2 Alternatively, panelists can be forced to a common scale, either by training or by use of external reference samples with assigned values (modulus).

5.5.1.3 Finally, each panelist's data can be brought to a common scale by one of a variety of normalizing methods.

5.5.2 Logarithms must be applied before carrying out data analysis. This becomes problematic if values are near threshold, as a logarithm of zero cannot be taken. (See 11.2.1)

5.6 Magnitude estimation should be used:

5.6.1 When end-effects are a concern, for example when panelists are not familiar with the entire range of sensations being presented.

5.6.2 When Stevens' Power Law is to be applied to the data.

5.6.3 Generally, in central location testing with panelists trained in the technique. It is not appropriate for home use or mall intercept testing with consumers.

5.7 This test method is only meant to be used with panelists who are specifically trained in magnitude estimation. Do not use this method with untrained panelists or untrained consumers.

6. Conditions of Testing

6.1 The general conditions for testing, such as the location, preparations, presentation and coding of samples, and the selection and training of panelists are described in the standards for general methodology, such as ISO 6658, ISO 8586-1, ISO 8589, ASTM STPs 434, 758, 913 or those describing methods using scales and categories, for example, ISO 4121 and ASTM STP 434.

7. Selection and Training of Panelists

7.1 Refer to ISO 8586-1 or ASTM STP 758 for all the general considerations concerning the selection and training of panelists.

7.2 As is true for all methods of sensory evaluation, the panel leader will have to make judgments as to the level of proficiency required of the panelists. The objectives of the test, the availability of panelists, the costs of securing additional panelists and of additional training should all be considered in the design of a training program. Panelists generally reach a stable level of proficiency in the method itself after three to four exercises in assigning magnitudes.

7.3 Estimating the areas of geometric shapes has proven very useful for introducing panelists to the basic concepts of magnitude estimation. A set of 18 figures composed of 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 panelists (see Table 1).

7.4 Prior to presenting the figures, the panel leader instructs the candidate in the principles of the method. This instruction should include, but is not necessarily limited to the following three points.

7.4.1 If the attribute is not present, the value 0 should be assigned.

7.4.2 There is no upper limit to the scale.

7.4.3 Values should be assigned on a ratio basis: If the attribute is twice as intense, it should receive a rating twice as large.

7.5 Panelists have a tendency to use "round numbers" such as 5, 10, 20, 25, etc. This should be pointed out explicitly during training. Panelists should be encouraged, "given permission," to use all numbers. Panelists are also influenced by the ratios mentioned in training. Therefore, care should be taken to mention a variety of different ratios, for example, 3:1 ⅓ , 7.5, 2.4, not just 2:1 and ½ .

7.6 *Assigning Codes to the Figures*— The figures are presented singly, centered on an 8.5 × 11 in. sheet of white paper. The panelist states his magnitude estimate; the estimation is recorded. The 8.5-cm square is presented first with the instruction to assign it a value between 30 and 100. The balance of the geometric figures should be shuffled prior to each test so that the type of geometric figure and the size of the areas do not form a particular pattern.

7.7 *Comparing the Results*—After completing the full set of shape estimates, panelists should be allowed to compare their results with the averaged results of the group. If this is not practical, the results from a previous group can also be used. The objective is to provide positive feedback, that is, 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.

7.8 If the panelists' results are very different, review the principles of the method again. If the panel leader judges that a panelist cannot be trained in the method, the training should be discontinued at this point and the panelist excused.

7.9 Once the panel has successfully completed the area estimation exercise, further training should be carried out with the commodity or type of test substance to be used in the main trial(s). This gives the panelist experience in applying magnitude estimation to attributes characterizing the test sample.

7.10 The panel leader may need to design exercises for training panelists to properly identify the attributes to be evaluated. The need for this will depend on the objectives and requirements of the test.

TABLE 1 Training Exercise Shapes

NOTE 1—Two 11.1 cm squares are included as a measure of reproducibility.

Dimensions/Areas (cm/cm^2)					
Circles		Triangles		Squares	
Radius	Area	Edge	Area	Edge	Area
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	123.2
6.8	145.3	15.5	104.0	11.1	123.2
8.3	216.4	19.2	159.6	14.2	201.6

8. Number of Panelists Required

8.1 As is true for other forms of scaling, the number of panelists necessary for a given task depends on the complexity of the task, how close together the various test samples are in the attribute being evaluated, the amount of training the panelists have received, and the importance to be attached to the decision based on the test results (c.f., ISO 8586-1). Issues of statistical power need to be resolved based on the variance associated with a particular evaluation and the magnitude of the differences that need to be detected.

9. Reference Samples

9.1 *External References*—The panel leader specifies to the panelists 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 panelists to make his or her subsequent judgments relative to the value assigned.

9.2 The reference should have an intensity close to the geometric mean for the whole panel. A reference that represents an extreme value of the attribute will distort the data due to a contrast effect and reduce the sensitivity of the method.

9.3 Magnitude estimation does not impose any specific restrictions on sample presentation. However, the external reference sample, if used, is presented to the panelist first with the specification that the sample is to have a particular value. The value chosen should be between 30 and 100. In most instances, when the initial value is in this range, the panelist will not need to use decimals in order to conform to the ratio principle. Some panelists find it more difficult to use decimals and most will avoid using them unless specifically instructed to do so.

10. Procedure—Assigning Magnitude Estimations

10.1 Magnitude estimation imposes no special restrictions on the method or order of sample presentation. As in all sensory experiments, the order of sample presentation should be randomized and balanced across all panelists.

10.2 In the modalities of olfaction and gustation, the problems of adaptation and fatigue must be carefully considered when encouraging or requiring repeated evaluations of previous samples. When only a limited number of samples can be evaluated, it may be necessary to sacrifice statistical rigor to the known limitations of the sensory systems.

10.3 *Without an External Reference Sample*—The panelist evaluates the first sample and assigns a magnitude estimate. The panelist is instructed to be careful not to assign a value that is too small. It has generally been suggested that the first sample be assigned a value in the range of 30–100 (see 9.3).

10.3.1 The panelist is then instructed to rate each sample relative to its immediately preceding sample or to the first sample.

10.4 *With an External Reference Sample*— The panelist is presented the reference sample and is informed of its assigned value or allowed to assign a value of his own. The panelist next evaluates the first coded sample and assigns it a value relative to the reference sample. All subsequent samples are rated relative to either the identified reference or to its immediately preceding sample.

10.5 The procedure of rating each sample relative to its immediate predecessor can produce scale drift due to an accumulation of errors. In addition, the random error associated with each evaluation is no longer independent from the preceding evaluations (see Section 11).

11. Data Analysis

11.1 An analysis of variance (ANOVA), which explicitly accounts for all blocking factors and is carried out on logarithmically transformed data, will provide results of the highest precision. However, as a practical matter, it is not always possible to design and execute experiments in a manner that is consistent with an ANOVA model which contains all of the critical factors. For example, when a project extends over multiple sessions, it may not be possible to assemble exactly the same group of panelists at each session. In other cases it may be necessary to combine samples from multiple projects into a single session. If your design does not conform to standard experimental design, every effort should be made to consult a statistician to develop an appropriate form of the ANOVA model. If this is not an option, a less desirable but workable solution may be to employ a one-way ANOVA using treatments as the only factor. Finally, when investigating the dose-response relationship between some physical parameter and a sensory attribute, regression analysis is appropriate.

11.1.1 It should be noted, that both normalizing and instructing the panelists to rate each sample relative to the immediately preceding sample cause certain theoretical problems in the statistical analysis. When these techniques are employed, the statistical probabilities arising from the analyses should be regarded as approximate. The statistical approaches to dealing with these problems are beyond the scope of this test method.

11.2 *Log Transformations*—Present knowledge indicates that magnitude estimations conform to a log-normal distribution, and that more precise results are obtained when analyses are carried out on logarithmically transformed data.

11.2.1 *Dealing with Zeros*—Since one cannot take the logarithm of zero, any zero response causes a problem. Different investigators have used different approaches to dealing with zeros. It is recommended that the zero values should be replaced by very small values. The specific value chosen should take into account the scale used by each panelist (for example, half of the smallest value assigned by that panelist).

11.3 *Product-Panelist Interactions*:

11.3.1 *External Reference/Experienced Panelists*—An external reference should force panelists to use a common scale. With experienced panelists, this often eliminates product-panelist interactions. (When this is the case, the data require no special processing to remove this interaction.)

11.3.2 *No External Reference/Inexperienced Panelists*—With panelists who have just been trained, or when no external reference is used, or both, product-panelist interactions may still occur. In this case, the methods discussed below can be used to reduce, or eliminate, this interaction.

11.4 *Normalizing*—Product-panelist interactions should first be removed by normalizing. This significantly improves the sensitivity of subsequent analyses. "Internal Standard Normalizing," "No Standard Normalizing" and "External Calibration" have been used for this purpose. The most precise of these methods is "Internal Standard Normalizing." It is recommended that this method be used wherever possible.

11.4.1 *Internal Standard Normalizing*— This approach can be used whether or not an external reference is used. It requires that one or more unidentified internal reference samples be included in the test set.

11.4.1.1 When replicate internal reference samples have been included, one first averages a panelist's estimates for these samples.

11.4.1.2 If no external reference has been used, one then calculates the value which would bring the average of the internal reference samples to some predetermined, fixed value.

11.4.1.3 When an external reference has been used, one calculates the value that would bring the average of the internal reference samples to the value given to the external reference.

11.4.1.4 To normalize the test sample data, one simply multiplies each estimate by the value calculated above.

11.4.2 *No Standard Normalizing*—Also known as the "Method of Averges" and "Equalization of Means." This method is recommended for use with sets of 10 or more samples. This number of samples is necessary to provide data that approximates a normal distribution and will minimize the effect due to the loss of degrees of freedom in an ANOVA. With 10 samples, the normalization factors and scales will be more stable and the results will be more reliable. If it is not possible to evaluate at least 10 samples in one session, this method should not be used as it may not be reliable. Please note that less than 10 samples have been used in the examples in the appendices for ease of presentation.

11.4.2.1 Calculate the mean of the logarithm of each panelist's estimates.

11.4.2.2 Calculate the grand mean across all panelists.

11.4.2.3 For each panelist, calculate the value which when added to his mean makes it equal to the groups' mean.

11.4.2.4 Add to each panelist's estimates his value.

11.4.2.5 The rationale for this method is as follows: Each panelist has experienced the same set of stimuli. Therefore, the total magnitude of their responses should be identical. Therefore, one brings each panelist's scale to the same total magnitude.

11.4.2.6 When using this method, it has been suggested that for each value calculated, one degree of freedom must be lost from the total for the experiment. However, when following the recommendation to use 15 or more panelists and at least 10 determinations for each value calculated, the difference in the error term will be at most 6 %.

11.4.3 *External Calibration*—Various forms of external calibration have been used in the literature. After evaluating the test samples, the panelist receives a verbal scale of from four to eleven points. It will consist of terms such as "Extremely Intense,"" Very Intense," "Moderately Intense," "Slightly Intense," etc.

11.4.3.1 The panel leader instructs the panelist to assign magnitude estimates to these terms in a way that is consistent with the scale used for evaluating the test samples.

11.4.3.2 The ratio of the geometric mean of a panelist's calibration scale values and the geometric mean of the entire group's calibration scale values can be used as the correction factor for that panelist's scores. (See X4.2 for an example.) Alternatively, the correction factor may be calculated by dividing the geometric mean of a panelist's calibration scale values into an arbitrary value assigned by the panel leader. Another method uses each panelist's maximum calibration scale value as the correction factor, thereby transforming their estimates into percentages. The geometric mean of each panelist's calibration scale may also be used.

11.5 *Test Results*:

11.5.1 If the desire is to learn whether sample treatments differ significantly, then analysis of variance, followed by a multiple comparison procedure is the usual course of analysis followed.

11.5.2 When regression analysis is appropriate, the parameter of primary interest is usually the slope. This corresponds to the *n* of Stevens' equation.

12. Keywords

12.1 agricultural products; beverages; color; estimation; feel; food products; magnitude estimation; odors; odor or water pollution; perfumes; scaling; sensory analysis; sound; taste; tobacco

APPENDIXES

(Nonmandatory Information)

X1. DATA ANALYSIS AND INTERPRETATION USING ANOVA WITHOUT NORMALIZING (NO REPLICATION)

X1.1 Table X1.1 lists the results obtained when seven experienced panelists scaled the intensity of bitterness of six samples of a beverage containing various levels of caffeine. Natural logarithms were taken and are included in Table X1.1 in parentheses.

X1.2 *Determining Whether Significant Differences Exist*—Two-way analysis of variance was applied to the ln (magnitude estimations) in Table X1.1. The results were as follows in Table X1.2.

TABLE X1.1 Sample Data Set 1

Trt Codes	561	274	935	803	417	127
Conc (mg/100 ml)	9	18	36	40	72	144
Panelist	Magnitude Estimations (Logarithms) R1					
1	10 (2.30)	20 (3.00)	35 (3.56)	40 (3.69)	70 (4.25)	140 (4.94)
2	8 (2.08)	20 (3.00)	38 (3.64)	44 (3.78)	85 (4.44)	160 (5.08)
3	8 (2.08)	20 (3.00)	36 (3.58)	40 (3.69)	75 (4.32)	150 (5.01)
4	7 (1.95)	15 (2.71)	32 (3.47)	37 (3.61)	70 (4.25)	135 (4.91)
5	12 (2.48)	25 (3.22)	38 (3.64)	40 (3.69)	75 (4.32)	145 (4.98)
6	12 (2.48)	22 (3.09)	35 (3.56)	40 (3.69)	80 (4.38)	160 (5.08)
7	9 (2.20)	18 (2.89)	35 (3.56)	40 (3.69)	74 (4.30)	145 (4.98)
Mean Ln	2.22	2.99	3.57	3.69	4.32	4.99

TABLE X1.2 ANOVA of Data Set 1

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square	F Value
Panelist	6	0.240	0.040	4.55
Treatment	5	33.177	6.635	754.69
Error	30	0.264	0.009	

X1.3 The analysis of variance shows a significant treatment effect. Tukey's test is one of several multiple comparison tests that may be used to determine which samples differ significantly.[4] As there are six treatments and 30 degrees of freedom for error, Tukey's honestly significant difference is the standard error of the mean, ($\sqrt{0.009/7}$ = 0.035) multiplied by 4.30,[5] that is 0.154. The only two samples not differing significantly were 803 and 935. These two means differ by only 0.12.

[4] Hochberg, Y., and Tamhane, A. C., *Multiple Comparison Procedures*, John Wiley, New York, 1987.

[5] Poste, L. M., Makie, D. A., Butler, G., and Larmond, E., "Laboratory Methods for Sensory Analysis of Food," Research Branch Agriculture Canada, Publication 864/E, 1991.

X2. DATA ANALYSIS AND INTERPRETATION USING INTERNAL STANDARD NORMALIZING (NO REPLICATION)

X2.1 *Normalizing With An External Reference*—Just prior to evaluating the intensity of bitterness of the six samples, the panelists were presented with a reference sample and told that it had a designated value of 40. The six samples above were presented to the panelists in random order. Sample 803 was the same as the reference sample. To normalize the coded samples using this reference sample the following procedure was used. Panelist 1 had assigned 40 to it; thus no correction needed to be applied to his responses. Panelist 2 assigned 44 to sample 803: accordingly his values needed to be multiplied by 0.909 (or divided by 1.1) to bring the value of 44 to 40. All the other values assigned by that panelist were multiplied by the same factor. The same procedure had to be used for panelist 4 who had assigned 37 to the coded reference sample. His values had to be multiplied by 1.081 to bring the value for sample 803 up to 40. The same multiplier was used to adjust his other assigned values.

X2.2 The adjusted values were then transformed using natural logarithms (see Table X2.1).

X2.3 Analysis of variance was applied to these magnitude estimations (logarithms) and the means and least significant difference were calculated as in Section Appendix X1. The results were as follows in Table X2.2.

X2.4 The honestly significant difference for six samples and 36 degrees of freedom is 0.169. As before, all samples except 935 and 803 differ significantly.

X2.5 As can be seen, the first approach gives the same means but a smaller error. However, this approach avoids the use of a two-way analysis of variance and may be preferred in some cases despite the loss in precision.

TABLE X2.1 Data Normalized Using Internal Standard Normalization

Trt Code	561	274	935	803	417	127
Panelists	Magnitude Estimations (Logarithms)					
1	2.303	2.996	3.555	3.689	4.248	4.942
2	1.984	2.900	3.542	3.689	4.347	4.980
3	2.079	2.996	3.584	3.689	4.317	5.011
4	2.024	2.786	3.544	3.689	4.326	4.983
5	2.485	3.219	3.638	3.689	4.317	4.977
6	2.485	3.091	3.555	3.689	4.382	5.075
7	2.197	2.890	3.555	3.689	4.304	4.977
Mean Ln	2.22	2.98	3.57	3.69	4.32	4.99

TABLE X2.2 ANOVA of Normalized Data (Internal Standard)

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square	F Value
Treatment	5	33.177	6.635	581.86
Error	36	0.411	0.011	

X3. DATA ANALYSIS AND INTERPRETATION USING EXTERNAL CALIBRATION

X3.1 *Performing the Calibration*—After completion of the main experiment, panelists are required to assign magnitude estimates to a verbal calibration scale. For purposes of illustration a five-point scale ranging from "Extremely Bitter" to "Very Slightly Bitter" has been created. The ten sample minimum recommended for "No Standard Normalization" is not an issue in this situation because the sample set (the words) have been carefully selected to cover the entire scale and therefore should provide a stable measure.

X3.2 Panelists would be instructed to assign the "Extremely Bitter" category a value greater than or equal to that given to the most bitter sample rated. They would also be instructed to assign the "Very Slightly Bitter" category a value less than or equal to the least bitter sample evaluated. Hypothetical results for this exercise are presented in Table X3.1.

X3.3 *Normalizing to the Geometric Mean of the Calibration Scale* —First calculate the normalizing values using the method of no standard normalizing on the calibration scores. A one-way ANOVA is then carried out on the corrected ln(estimates) (Table X3.2).

X3.4 The honestly significant difference calculated as above for six treatments and 36 degrees of freedom is 0.170 and the only treatments that do not differ significantly are 935 and 803.

X3.5 *Normalizing to the Maximum of the Calibration Scale* —Divide each score by the maximum value of the calibration scale and then multiply by 100 (Table X3.4). Then perform the one-way ANOVA and multiple comparison as above.

X3.6 The honestly significant difference calculated as above for six treatments and 36 degrees of freedom is 0.163 and the only treatments that do not differ significantly are 935 and 803.

TABLE X3.1 Hypothetical External Calibration Scores

Panelists	Very Slightly Bitter	Somewhat Bitter	Moderately Bitter	Very Bitter	Extremely Bitter	Normalizing Value[A]
1	5	25	50	100	150	−0.002
2	5	30	60	100	160	−0.088
3	5	25	50	100	150	−0.002
4	5	20	45	90	140	0.098
5	5	25	50	100	150	−0.002
6	3	30	55	110	170	0.000
7	5	25	50	100	150	−0.002

[A] Calculated by the method of "No Standard Normalizing" (see 11.4.2, 11.4.3 and X4.2).

TABLE X3.2 Magnitude Estimates (LN) Corrected by the Geometric Mean of the External Scale

Trt Code	561	274	935	803	417	127
Panelist	Corrected Magnitude Estimates (Ln)					
1	2.300	2.994	3.553	3.686	4.246	4.940
2	1.991	2.908	3.550	3.696	4.355	4.987
3	2.077	2.994	3.582	3.687	4.315	5.009
4	2.044	2.806	3.564	3.709	4.346	5.003
5	2.483	3.217	3.636	3.687	4.315	4.975
6	2.485	3.091	3.555	3.689	4.382	5.075
7	2.195	2.888	3.553	3.687	4.302	4.975
Ln Means	2.22	2.98	3.57	3.69	4.32	4.99

TABLE X3.3 ANOVA of Corrected Data (Geometric Mean)

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square	F Value
Treatment	5	33.177	6.635	603.18
Error	36	0.385	0.011	

TABLE X3.4 Magnitude Estimates (LN) Corrected by the Maximum of the External Scale

Trt Code	561	274	935	803	417	127
Panelist	Corrected Magnitude Estimate (Ln [Estimate])					
1	6.7 (1.897)	13.3 (2.590)	23.3 (3.150)	26.7 (3.283)	46.7 (3.843)	93.3 (4.536)
2	5.0 (1.609)	12.5 (2.526)	23.8 (3.168)	27.5 (3.314)	53.1 (3.973)	100.0 (4.605)
3	5.3 (1.674)	13.3 (2.590)	24.0 (3.178)	26.7 (3.283)	50.0 (3.912)	100.0 (4.605)
4	5.0 (1.609)	10.7 (2.372)	22.8 (3.129)	26.4 (3.274)	50.0 (3.912)	96.4 (4.569)
5	8.0 (2.079)	16.7 (2.813)	25.3 (3.232)	26.7 (3.283)	50.0 (3.912)	96.7 (4.571)
6	7.0 (1.954)	12.9 (2.560)	20.6 (3.025)	23.5 (3.158)	47.0 (3.851)	94.1 (4.544)
7	6.0 (1.792)	12.0 (2.485)	23.3 (3.150)	26.7 (3.283)	49.3 (3.899)	96.7 (4.571)
Mean Ln	1.80	2.56	3.15	3.27	3.90	4.57

TABLE X3.5 ANOVA of Corrected Data (Maximum)

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square	F Value
Treatment	5	33.177	6.635	663.5
Error	36	0.362	0.010	

X4. DATA ANALYSIS AND INTERPRETATION USING NO STANDARD NORMALIZING

X4.1 When both ANOVA and internal standard normalizing are not feasible, no standard normalizing may be used on suitable data sets. While the data set in Table X1.1 does not meet the minimum standards recommended for this method, it will be used for the purpose of illustration.

X4.2 Determining the normalizing values: The first step is to calculate the mean ln(estimate) for each panelist (Table X4.1). Next calculate the overall panel mean ln(estimate). Finally, for each panelist, calculate the normalizing value by subtracting the panelist's mean from the group mean.

X4.3 *Analyzing the data*—To normalize each panelist's data, add the normalizing value to each ln(estimate) (see Table X4.2).

X4.4 When analysis of variance was applied to these data, results were as follows in Table X4.3.

TABLE X4.1 Calculation of Normalizing Values

Panelist	Sum of Ln (Estimates)	Mean of Ln (Estimates)	Normalizing Value
1	21.732	3.622	0.010
2	22.015	3.669	−0.037
3	21.676	3.613	0.019
4	20.884	3.481	0.151
5	22.324	3.721	−0.089
6	22.277	3.713	−0.081
7	21.613	3.602	0.031
Group Mean		3.632	

X4.5 In this instance six degrees of freedom (number of panelists—1) have been subtracted from the error degrees of freedom as these have been lost when the seven geometric means were estimated from and used to adjust the data. It can be seen that this analysis of variance is identical to that in Table X1.2.

TABLE X4.2 Normalized Ln(Estimates)

Treatments Panelists	561	274	935	803	417	127
	ln (Magnitude Estimations)					
1	2.313	3.006	3.565	3.699	4.258	4.944
2	2.042	2.959	3.601	3.747	4.406	5.038
3	2.098	3.015	3.603	3.708	4.336	5.030
4	2.097	2.859	3.617	3.762	4.399	5.056
5	2.396	3.130	3.549	3.600	4.228	4.888
6	2.404	3.010	3.474	3.608	4.301	4.994
7	2.228	2.921	3.586	3.720	4.335	5.008
Mean LN	2.23	2.99	3.57	3.69	4.32	4.99

TABLE X4.3 ANOVA on Normalized Data (No Standard Normalizing)

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square	F Value
Treatment	5	33.177	6.635	754.69
Error	30	0.264	0.009	

X4.6 Therefore, when ANOVA on the raw data is feasible, there is no value in the extra steps required for no standard normalizing.

X5. ADDITIONAL INFORMATION

X5.1 It should be noted that the complete ANOVA and the "no standard normalizing" result in a smaller mean squared error than internal standard normalizing. Powers et al.[6] have demonstrated that the error is less when the geometric mean is the normalizing position rather than some arbitrary point such as a designated reference sample. The reader should note from 9.2 that if a designated reference sample is used the reference should have an intensity close to the geometric mean for the whole panel. The closer the reference sample is to the actual geometric mean, the better.

X5.2 Examining the slope of the regression curve: In as much as the samples progress in concentration in caffeine and the amounts are known, linear regression may be applied to the logarithms of the concentrations and to the ln (magnitude estimations) to ascertain the slope of the regression curve. If the magnitude estimations have not been normalized to a reference or internally it is necessary to allow for different intercepts for the different panelists.

X5.3 The following analysis of variance is the result. The estimate of the slope is 0.992 with a standard error of 0.016.

TABLE X5.1 ANOVA Table for Testing that the Slope Coefficient in the Regression Model is Significantly Different from Zero

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square	F Value
Panelist	6	0.240	0.040	4.37
Ln Conc.	1	33.129	33.129	3618.70
Error	34	0.311	0.009	

X5.4 The regression curves can be further examined by checking the interaction with panelists to see if each panelist has the same slope. See Table X5.2 for analysis results.

TABLE X5.2 ANOVA Table for Testing for the Equality of the Slope Coefficients from Panelist to Panelist

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square	F Value
Panelist	6	0.240	0.040	4.37
Ln Conc.	1	33.129	33.129	3618.70
P*Ln Conc.	6	0.173	0.029	5.81
Error	28	0.139	0.005	

X5.5 Once again the analysis can be done on the normalized values. In this case the panelist effect does not have to be removed. The estimate of the slope will remain the same. When normalized to a reference, the standard error of the slope is 0.018. When normalized internally with geometric means one must again take care to adjust the degrees of freedom for the error by six. The result is a standard error of 0.106, identical to the analysis described above.

[6] Power, J. J., Ware, G. O., and Shinholser, K. J., "Magnitude Estimation With and Without Rescaling," *Journal of Sensory Studies*, 1990, 5: 105-116.

REFERENCES

(**1**) Butler, G., Poste, L. M., Wolynetz, M. S., Agar, V. E., Larmond, E., “Alternative Analysis of Magnitude Estimation Data,” *Journal of Sensory Studies*, 1987, 2:243–257.

(**2**) Jounela-Eriksson, P. “Whisky Aroma Evaluated by Magnitude Estimation.” *Lebensmittel-Wissenschaft u. Technology*, 1982, 15:302–7.

(**3**) Lawless, H. T. “Logarithmic Transformation of Magnitude Estimation Data and Comparisons of Scaling Methods,” *Journal of Sensory Studies* , 1989, 4:75–86.

(**4**) Leight, R. S., and Warren, C. B. “Standing Panels Using Magnitude Estimation for Research and Product Development,” *Applied Sensory Analysis of Foods*, H. Moskowitz, ed., CRC Press, Boca Raton, Florida, USA, 1988, pp 225–249.

(**5**) McDaniel, M. R., and Sawyer, F. M. “Descriptive Analysis of Whiskey Sour Formulations: Magnitude Estimation versus a 9-point Category Scale,” *Journal of Food Sciences*, 1981, 46:178–81,189.

(**6**) McDaniel, M. R. and Sawyer, F. M. “Preference Testing of Whiskey Sour Formulations: Magnitude Estimation versus the 9-point Hedonic Scale,” *Journal of Food Sciences*, 1981, 46:182–5.

(**7**) Meilgaard, M. C. and Reid, D. S. “Determination of Personal and Group Thresholds and the Use of Magnitude Estimation in Beer Flavour Chemistry,” *Progress in Flavour Research* , D. G. Land and H. E. Nurstein, eds., Applied Sci. Publishers, London, 1979, pp 67–73.

(**8**) Moskowitz, H. R. “Magnitude Estimation: Notes on How, When, Why and Where to Use It,” *Journal of Food Quality* , 1977, 1:195–228.

(**9**) Pearce, J. H., Korth, B. and Warren, C. B. 1986. Evaluation of Three Scaling Methods for Hedonics, *Journal Sensory Studies*, 1986, 1:27–46.

(**10**) Powers, J. J., Warren, C. B., Masurat, T. “Collaborative Trials Involving Three Methods of Normalizing Magnitude Estimations,” *Lebensmittel-Wissenschaft u. Technol*. 1981, 14:86–93.

(**11**) Shand, P. J., Hawrysh, Z. J., Hardin, R. T., and Jeremiah, L. E. “Descriptive Sensory Assessment of Beef Steaks by Category Scaling, Line Scaling and Magnitude Estimation,” *Journal of Food Sciences*, 1985, 50:495–500.

(**12**) Stevens, S. S. “On the Psychophysical Law,” *Psychological Review* 1957, 64:153–181.

(**13**) Warren, C. B. “Development of Fragrances with Functional Properties by Quantitative Measurement of Sensory and Physical Parameters,”“ Odor Quality and Chemical Structure”, Symposium Series 148, American Chemical Society, Washington, D.C. USA. 1981, pp 57–77.