Designation: E2139 – 05 (Reapproved 2011)

Standard Test Method for Same-Different Test[1]

This standard is issued under the fixed designation E2139; the number immediately following the designation indicates the year of original adoption or, in the case of revision, the year of last revision. A number in parentheses indicates the year of last reapproval. A superscript epsilon (ε) indicates an editorial change since the last revision or reapproval.

1. Scope

1.1 This test method describes a procedure for comparing two products.

1.2 This test method does not describe the Thurstonian modeling approach to this test.

1.3 This test method is sometimes referred to as the simple-difference test.

1.4 A same-different test determines whether two products are perceived to be the same or different overall.

1.5 The procedure of the test described in this test method consists of presenting a single pair of samples to each assessor. The presentation of multiple pairs would require different statistical treatment and it is outside of the scope of this test method.

1.6 This test method is not attribute-specific, unlike the directional difference test.

1.7 This test method is not intended to determine the magnitude of the difference; however, statistical methods may be used to estimate the size of the difference.

1.8 This test method may be chosen over the triangle or duo-trio tests where sensory fatigue or carry-over are a concern, or where a simpler task is needed.

1.9 *This standard may involve hazardous materials, operations, and equipment. This standard does not purport to address all of the safety concerns, if any, associated with its use. It is the responsibility of the user of this standard to establish appropriate safety and health practices and determine the applicability of regulatory limitations prior to use.*

2. Referenced Documents

2.1 *ASTM Standards:*[2]

E253 Terminology Relating to Sensory Evaluation of Materials and Products
E456 Terminology Relating to Quality and Statistics
E1871 Guide for Serving Protocol for Sensory Evaluation of Foods and Beverages

2.2 *ASTM Publications:*[2]

Manual 26 Sensory Testing Methods, 2nd Edition
STP 758 Guidelines for the Selection and Training of Sensory Panel Members
STP 913 Guidelines for Physical Requirements for Sensory Evaluation Laboratories

2.3 *ISO Standard:*[3]

ISO 5495 Sensory Analysis—Methodology—Paired Comparison

3. Terminology

3.1 For definition of terms relating to sensory analysis, see Terminology E253, and for terms relating to statistics, see Terminology E456.

3.2 *Definitions of Terms Specific to This Standard:*

3.2.1 *α (alpha) risk*—probability of concluding that a perceptible difference exists when, in reality, one does not (also known as Type I Error or significance level).

3.2.2 *β (beta) risk*—probability of concluding that no perceptible difference exists when, in reality, one does (also known as Type II Error).

3.2.3 *chi-square test*—statistical test used to test hypotheses on frequency counts and proportions.

3.2.4 *Δ (delta)*—test sensitivity parameter established prior to testing and used along with the selected values of α, β, and an estimated value of p_1 to determine the number of assessors needed in a study. Delta (Δ) is the minimum difference in proportions that the researcher wants to detect, where the difference is $\Delta = p_2 - p_1$. Δ is not a standard measure of sensory difference. The same value of Δ may correspond to different sensory differences for different values of p_1 (see 9.5 for an example).

3.2.5 *Fisher's Exact Test (FET)*—statistical test of the equality of two independent binomial proportions.

[1] This test method is under the jurisdiction of ASTM Committee E18 on Sensory Evaluation and is the direct responsibility of Subcommittee E18.04 on Fundamentals of Sensory.

Current edition approved Aug. 1, 2011. Published August 2011. Originally approved in 2005. Last previous edition approved in 2005 as E2139–05. DOI: 10.1520/E2139-05R11.

[2] For referenced ASTM standards, visit the ASTM website, www.astm.org, or contact ASTM Customer Service at service@astm.org. For *Annual Book of ASTM Standards* volume information, refer to the standard's Document Summary page on the ASTM website.

[3] Available from American National Standards Institute (ANSI), 25 W. 43rd St., 4th Floor, New York, NY 10036, http://www.ansi.org.

3.2.6 p_1—proportion of assessors in the population who would respond *different* to the matched sample pair. Based on experience with using the same-different test and possibly with the same type of products, the user may have *a priori* knowledge about the value of p_1.

3.2.7 p_2—proportion of assessors in the population who would respond *different* to the unmatched sample pair.

3.2.8 *power 1-*β *(beta) risk*—probability of concluding that a perceptible difference exists when, in reality, one of size Δ does.

3.2.9 *product*—material to be evaluated.

3.2.10 *sample*—unit of product prepared, presented, and evaluated in the test.

3.2.11 *sensitivity*—term used to summarize the performance characteristics of this test. The sensitivity of the test is defined by the four values selected for α, β, p_1, and Δ.

4. Summary of Test Method

4.1 Clearly define the test objective in writing.

4.2 Choose the number of assessors based on the sensitivity desired for the test. The sensitivity of the test is in part related to two competing risks: the risk of declaring a difference when there is none (that is, α-risk), and the risk of not declaring a difference when there is one (that is, β-risk). Acceptable values of α and β vary depending on the test objective. The values should be agreed upon by all parties affected by the results of the test.

4.3 The two products of interest (A and B) are selected. Assessors are presented with one of four possible pairs of samples: A/A, B/B, A/B, and B/A. The total number of *same* pairs (A/A and B/B) usually equals the total number of *different* pairs (A/B and B/A). The assessor's task is to categorize the given pair of samples as *same* or *different*.

4.4 The data are summarized in a two-by-two table where the columns show the type of pair received (*same* or *different*) and the rows show the assessor's response (*same* or *different*). A Fisher's Exact Test (FET) is used to determine whether the samples are perceptibly different. Other statistical methods that approximate the FET can sometimes be used.

5. Significance and Use

5.1 This overall difference test method is used when the test objective is to determine whether a sensory difference exists or does not exist between two samples. It is also known as the simple difference test.

5.2 The test is appropriate in situations where samples have extreme intensities, give rapid sensory fatigue, have long lingering flavors, or cannot be consumed in large quantities, or a combination thereof.

5.3 The test is also appropriate for situations where the stimulus sites are limited to two (for example, two hands, each side of the face, two ears).

5.4 The test provides a measure of the bias where judges perceive two same products to be different.

5.5 The test has the advantage of being a simple and intuitive task.

6. Apparatus

6.1 Carry out the test under conditions that prevent contact between assessors until the evaluations have been completed, for example, booths that comply with STP 913.

6.2 For food and beverage tests, sample preparation and serving sizes should comply with Practice E1871, or see Refs (**1**) or (**2**).[4]

7. Definition of Hypotheses

7.1 This test can be characterized by a two-by-two table of probabilities according to the sample pair that the assessors in the population would receive and their responses, as follows:

		Assessor Would Receive	
		Matched Pair (AA or BB)	Unmatched Pair (AB or BA)
Assessor's Response	Same:	$1 - p_1$	$1 - p_2$
	Different:	p_1	$p_2 = (= p_1 + \Delta)$
	Total:	1	1

where p_1 and p_2 are the probabilities of responding *different* for those who would receive the matched pairs and the unmatched pairs, respectively.

7.2 To determine whether the samples are perceptibly different with a given sensitivity, the following one-sided statistical hypothesis is tested:

H_o: $p_1 = p_2$
H_a: $p_1 < p_2$

7.3 The hypothesis test can be expressed in terms of the minimum detectable difference Δ (H_o: $\Delta = 0$ versus H_a: $\Delta > 0$). Delta (Δ) will equal 0 and p_1 will equal p_2 if there is no detectable difference between the samples. This test addresses whether or not Δ is greater than 0. Thus, the hypothesis is one-sided because it is not of interest in this test to consider that responding *different* to the matched pair could be more likely than responding *different* to the unmatched pair.

8. Assessors

8.1 All assessors must be familiar with the mechanics of the same-different test (the format, the task, and the procedure of evaluation). Greater test sensitivity, if needed, may be achieved through selection of assessors who demonstrate above average individual sensitivity (see STP 758).

8.2 In order to perform this test, assessors do not require special sensory training on the samples in question. For example, they do not need to be able to recognize any specific attribute.

8.3 The assessors must be sampled from a homogeneous population that is well-defined. The population must be chosen on the basis of the test objective. Defining characteristics of the population can be, for example, training level, gender, experience with the product, and so forth.

[4] The boldface numbers in parentheses refer to the list of references at the end of this standard.

9. Number of Assessors

9.1 Choose all the sensitivity parameters that are needed to choose the number of assessors for the test. Choose the α-risk and the β-risk. Based on experience, choose the expected value for p_1. Choose Δ, $p_2 - p_1$, the minimum difference in proportions that the researcher wants to detect. The most commonly used values for α-risk, β-risk, p_1 and Δ are α = 0.05, β = 0.20, p_1 = 0.3, and Δ = 0.3. These values can be adjusted on a case-by-case basis to reflect the sensitivity desired versus the number of assessors.

9.2 Having defined the required sensitivity (α-risk, β-risk, p_1, and Δ), determine the corresponding sample size from Table A1.1 (see Ref (**9**)). This is done by first finding the section of the table with a p_1 value corresponding to the proportion of assessors in the population who would respond *different* to the matched sample pair. Second, locate the total sample size from the intersection of the desired α, p_2 (or Δ), and β values. In the case of the most commonly used values listed in 9.1, Table A1.1 indicates that 84 assessors are needed. The sample size *n* is based on the number of same and different samples being equal The sample sizes listed are the total sample size rounded up to the nearest number evenly divisible by 4 since there are four possible combinations of the samples. To determine the number of same and different pairs to prepare, divide *n* by two.

9.3 If the user has no prior experience with the same-different test and has no specific expectation for the value of p_1, then two options are available. Either use p_1 = 0.3 and proceed as indicated in 9.2, or use the last section of Table A1.1. This section gives samples sizes that are the largest required, given α, β, and Δ, regardless of p_1.

9.4 Often in practice, the number of assessors is determined by practical conditions (for example, duration of the experiment, number of available assessors, quantity of product, and so forth) However, increasing the number of assessors increases the likelihood of detecting small differences. Thus, one should expect to use larger numbers of assessors when trying to demonstrate that products are similar compared to when one is trying to demonstrate that they are different.

9.4.1 When the number of assessors is fixed, the power of the test (1-β) may be calculated by establishing a value for p_1, defining the required sensitivity for α-risk and the Δ, locating the number of assessors nearest the fixed amount, and then following up the column to the listed β-risk.

9.5 If a researcher wants to be 90 % certain of detecting response proportions of p_2 = 60 % versus the expected p_1 = 40 % with an α-risk of 5 %, then Δ = 0.60 − 0.40 = 0.20 and β = 0.10 or 90 % power. The number of assessors needed in this case is 232 (Table A1.1). If a researcher wants to be 90 % certain of detecting response proportions of p_2 = 70 % versus the expected p_1 = 50 % with an α-risk of 5 %, then Δ = 0.70 − 0.50 = 0.20 and β = 0.10 or 90 % power. The number of assessors needed in this case is 224 (Table A1.1).

10. Procedure

10.1 Determine the number of assessors needed for the test as well as the population that they should represent (for example, assessors selected for a specific sensory sensitivity).

10.2 It is critical to the validity of the test that assessors cannot identify the samples from the way in which they are presented. One should avoid any subtle differences in temperature or appearance, especially color, caused by factors such as the time sequence of preparation. It may be possible to mask color differences using light filters, subdued illumination or colored vessels. Prepare samples out of sight and in an identical manner: same apparatus, same vessels, same quantities of product (see Practice E1871). The samples may be prepared in advance; however, this may not be possible for all types of products. It is essential that the samples cannot be recognized from the way they are presented.

10.3 Prepare serving order worksheet and ballot in advance of the test to ensure a balanced order of sample presentation of the two products, A and B. One of four possible pairs (A/A, B/B, A/B, and B/A) is assigned to each assessor. Make sure this assignment is done randomly. Design the test so that the number of *same* pairs equals the number of *different* pairs. The presentation order of the *different* pairs should be balanced as much as possible. Serving order worksheets should also include the identification of the samples for each set.

10.4 Prepare the response ballots in a way consistent with the product you are evaluating. For example, in a taste test, give the following instructions: *(1)* you will receive two samples. They may be the same or different; *(2)* evaluate the samples from left to right; and *(3)* determine whether they are the same or different.

10.4.1 The researcher can choose to add an instruction to the ballot indicating whether the assessor may re-evaluate the samples or not.

10.4.2 The ballot should also identify the assessor and date of test, as well as a ballot number that must be related to the sample set identification on the worksheet.

10.4.3 A section soliciting comments may be included following the initial forced-choice question.

10.4.4 The example of a ballot is provided in Fig. X2.2.

10.5 When possible, present both samples at the same time, along with the response ballot. In some instances, the samples may be presented sequentially if required by the type of product or the way they need to be presented, or both. This may be the case, for example, for the evaluation of a fragrance in a room where the assessor must change rooms to evaluate the second sample.

10.6 Collect all ballots and tabulate results for analysis.

11. Analysis and Interpretation of Results

11.1 The data from the test is summarized in a two-by-two table, as illustrated in the table below.

		Assessor Received Matched Pair (AA or BB)	Unmatched Pair (AB or BA)	Total
Assessor's Response	Same:	17	9	26
	Different:	13	21	34
	Total:	30	30	60

11.1.1 Before computing any test statistic, determine if the number of *different* responses from those who received the unmatched pair is less than or equal to the number of *different* responses from those who received the matched pair. If this is the case, conclude that the hypothesis of no difference cannot be rejected. If this is not the case, the computation of a test statistic is needed to determine whether the samples are perceptibly different or not.

11.2 Analyze the data using a Fisher's Exact Test (**3, 4, 5**). The FET is widely available in industry standard software. See computation examples in X1.5.2 and X2.5.2.

11.3 Other statistical tests can also be used as an approximation to the FET, provided the data table is not sparse. A sparse table is defined as one that has at least one expected frequency less than 5. The expected frequency in row *i* and column *j* is computed as:

$$E_{ij} = \frac{(\text{Row } i \text{ Total})\,(\text{Column } j \text{ Total})}{(\text{Grand Total})} \tag{1}$$

11.3.1 For example, the expected frequency for Row 1: Column 1 (that is, same response on a matched pair) is:

$$E_{11} = \frac{(26)\,(30)}{60} = 13 \tag{2}$$

11.4 Available tests that approximate the FET include the one-tailed continuity corrected Chi-square (χ^2) (**6**), the one-tailed non-continuity corrected Chi-square (χ^2) (**7**) and the *z*-test (**8**).

11.4.1 In the case of either Chi-square test, compare the calculated statistic to the critical value of a χ^2 distribution with one degree of freedom and an α level of twice the desired level. The critical values for a number of α levels are given in Table 1. For example, the critical value for a 5 % a level is 2.71.

11.4.2 Computation examples of the one-tailed continuity, corrected Chi-square are given in X1.5.3 and X2.5.3.

11.4.3 In the case of a *z*-test, compare the calculated statistic to the one-tailed critical value of the *z* distribution for the chosen α level.

TABLE 1 Critical Values for a One Sided, 1 Degree of Freedom χ^2 Test

α Level	Critical Value (one sided[A] 1df χ^2)
0.01	5.41
0.05	2.71
0.1	1.64
0.2	0.708
0.3	0.275
0.4	0.0642

[A] A one sided value is obtained by using the χ^2 value corresponding to twice the desired a level.

12. Report

12.1 Report the test objective, the results, the conclusions, and the population to which they can be generalized. The following additional information is recommended:

12.1.1 The purpose of the test and the nature of the treatment studied;

12.1.2 Full identification of the samples: origin, method of preparation, quantity, shape, storage prior to testing, serving size, and temperature. (Sample information should communicate that all storage, handling, and preparation was done in such a way as to yield samples that differed only in the variable of interest);

12.1.3 The number of assessors, the number of selections of each sample, and the result of the statistical analysis;

12.1.4 Relevant assessor information such as age, gender, experience in sensory testing, and experience with the product and test samples. Provide all details necessary to clearly define the population represented by the assessors;

12.1.5 Any information or instructions given to the assessor in connection with the test;

12.1.6 The test environment: use of booths, simultaneous or sequential presentation, environmental conditions, whether the identity of samples was disclosed after the test and the manner in which this was done; and

12.1.7 The location and date of the test and name of the panel leader.

13. Precision and Bias

13.1 Because results of this test are a function of individual sensitivities, a general statement regarding the precision of results that is applicable to all populations of assessors cannot be made. However, adherence to the recommendations in this test method should increase the reproducibility of results and minimize bias.

14. Keywords

14.1 difference test; minimize carry-over; minimize sensory fatigue; sensory test for difference; two-sample sensory test

ANNEX

(Mandatory Information)

A1. NUMBER OF ASSESSORS REQUIRED FOR THE SAME-DIFFERENT TEST

A1.1 See Table A1.1.

TABLE A1.1 Number of Assessors Required for Same-Different Test Based on Fishers Exact Test (One-Tailed) (see Ref 9)

NOTE 1—Please note that this table is divided into sections based upon the value of p_1. The sample size specified for Δ in the table will apply only to that p_1; if p_1 changes, a different sample size may be needed even if the value of Δ remains the same.

NOTE 2—First, select the appropriate value for p_1 and then find the section of the table that corresponds to it. If you do not know your actual p_1 it is proposed that a value of $p_1 = 0.3$ is a reasonable generic starting point. Alternatively, you can use the last section of this table which gives sample sizes that are the largest required given α, β, and Δ.

NOTE 3—The values recorded in this table have been rounded to the nearest whole number evenly divisible by four to allow for equal presentation of all possible paired combinations of the same and different samples.

NOTE 4—The values in this table were determined by calculating the appropriate *N* divisible by 4 that is at least equal to the power (1-β) listed.

	$p_1 = 0.1$					β			
α	p_2	Δ	0.5	0.4	0.3	0.2	0.1	0.05	0.01
0.4	0.2	0.1	32	44	60	88	168	224	364
0.4	0.3	0.2	16	20	28	36	52	68	124
0.4	0.4	0.3	12	16	16	24	32	40	60
0.4	0.5	0.4	8	12	12	16	20	28	40
0.4	0.6	0.5	8	8	12	12	16	20	28
0.4	0.7	0.6	8	8	8	8	12	16	20
0.4	0.8	0.7	4	8	8	8	12	12	16
0.4	0.9	0.8	4	4	8	8	8	8	12
0.3	0.2	0.1	52	68	88	136	200	276	436
0.3	0.3	0.2	16	24	40	48	68	88	144
0.3	0.4	0.3	12	16	20	28	40	48	72
0.3	0.5	0.4	8	12	12	16	28	32	48
0.3	0.6	0.5	8	8	12	12	20	24	36
0.3	0.7	0.6	8	8	8	8	12	20	24
0.3	0.8	0.7	4	8	8	8	12	12	20
0.3	0.9	0.8	4	4	8	8	8	8	12
0.2	0.2	0.1	72	96	132	180	260	348	536
0.2	0.3	0.2	28	40	48	60	88	112	172
0.2	0.4	0.3	20	20	28	32	48	60	92
0.2	0.5	0.4	16	16	20	24	32	40	56
0.2	0.6	0.5	12	12	16	16	20	28	40
0.2	0.7	0.6	12	12	12	12	16	20	28
0.2	0.8	0.7	4	8	12	12	12	16	20
0.2	0.9	0.8	4	4	8	8	12	12	16
0.1	0.2	0.1	116	156	200	264	368	464	684
0.1	0.3	0.2	44	52	68	88	116	152	216
0.1	0.4	0.3	24	32	40	48	64	76	112
0.1	0.5	0.4	16	20	24	32	40	48	68
0.1	0.6	0.5	12	16	20	24	32	36	48
0.1	0.7	0.6	12	12	16	16	20	28	36
0.1	0.8	0.7	8	8	12	16	16	20	28
0.1	0.9	0.8	8	8	8	8	12	16	20
0.05	0.2	0.1	176	216	272	348	464	576	820
0.05	0.3	0.2	64	76	92	112	148	184	256
0.05	0.4	0.3	36	40	48	60	80	96	132
0.05	0.5	0.4	24	28	32	40	52	60	84
0.05	0.6	0.5	20	20	24	28	36	40	56
0.05	0.7	0.6	12	16	20	20	24	32	40
0.05	0.8	0.7	12	12	12	16	20	24	32
0.05	0.9	0.8	12	12	12	12	16	20	24
0.01	0.2	0.1	312	372	444	540	688	824	1116
0.01	0.3	0.2	104	120	144	172	216	260	344
0.01	0.4	0.3	56	68	76	92	112	136	176
0.01	0.5	0.4	36	40	48	60	72	84	112
0.01	0.6	0.5	28	28	36	40	48	60	76
0.01	0.7	0.6	20	24	28	32	36	44	56
0.01	0.8	0.7	16	20	20	24	28	32	40
0.01	0.9	0.8	12	12	16	20	20	24	32
	$p_1 = 0.2$					β			
α	p_2	Δ	0.5	0.4	0.3	0.2	0.1	0.05	0.01
0.4	0.3	0.1	32	48	68	136	212	292	532
0.4	0.4	0.2	16	20	28	40	60	100	156
0.4	0.5	0.3	12	16	20	24	32	44	80
0.4	0.6	0.4	8	12	12	16	24	28	44
0.4	0.7	0.5	8	8	12	12	16	20	32
0.4	0.8	0.6	8	8	8	12	12	16	20
0.4	0.9	0.7	4	8	8	8	12	12	16
0.3	0.3	0.1	52	88	120	180	272	392	636
0.3	0.4	0.2	24	32	40	56	92	116	192
0.3	0.5	0.3	12	20	24	32	44	56	92
0.3	0.6	0.4	8	12	12	24	28	36	52
0.3	0.7	0.5	8	8	12	12	20	28	36

TABLE A1.1 *Continued*

0.3	0.8	0.6	8	8	8	12	12	20	28
0.3	0.9	0.7	4	8	8	8	12	12	20
0.2	0.3	0.1	92	128	172	244	376	504	792
0.2	0.4	0.2	32	44	56	80	112	144	224
0.2	0.5	0.3	20	20	32	40	56	76	112
0.2	0.6	0.4	16	16	20	28	36	44	68
0.2	0.7	0.5	12	12	16	16	28	32	44
0.2	0.8	0.6	8	12	12	16	16	20	32
0.2	0.9	0.7	4	8	12	12	12	16	20
0.1	0.3	0.1	164	216	284	372	528	680	1008
0.1	0.4	0.2	52	68	88	116	156	196	284
0.1	0.5	0.3	32	36	48	56	80	96	140
0.1	0.6	0.4	20	20	32	36	48	56	84
0.1	0.7	0.5	16	16	20	28	32	36	56
0.1	0.8	0.6	8	12	16	20	20	28	36
0.1	0.9	0.7	8	8	12	16	16	20	28
0.05	0.3	0.1	244	308	392	500	676	844	1212
0.05	0.4	0.2	76	96	116	148	196	240	344
0.05	0.5	0.3	40	48	60	72	96	120	168
0.05	0.6	0.4	24	32	36	48	60	72	96
0.05	0.7	0.5	20	20	24	32	36	48	64
0.05	0.8	0.6	12	16	20	20	24	32	44
0.05	0.9	0.7	12	12	12	16	20	24	32
0.01	0.3	0.1	444	536	644	788	1008	1212	1648
0.01	0.4	0.2	132	160	188	228	288	344	464
0.01	0.5	0.3	68	80	92	112	140	168	224
0.01	0.6	0.4	44	48	56	68	84	100	132
0.01	0.7	0.5	28	32	36	44	56	68	84
0.01	0.8	0.6	20	24	28	32	36	44	60
0.01	0.9	0.7	16	20	20	24	28	32	40
	$p_1 = 0.3$					β			
α	p_2	Δ	0.5	0.4	0.3	0.2	0.1	0.05	0.01
0.4	0.4	0.1	32	48	100	148	240	372	656
0.4	0.5	0.2	16	24	32	44	84	108	176
0.4	0.6	0.3	12	16	20	24	36	48	88
0.4	0.7	0.4	8	12	12	16	24	28	44
0.4	0.8	0.5	8	8	12	12	16	20	32
0.4	0.9	0.6	8	8	8	8	12	16	20
0.3	0.4	0.1	52	96	128	204	332	468	764
0.3	0.5	0.2	28	32	44	56	96	124	208
0.3	0.6	0.3	12	20	28	32	44	68	100
0.3	0.7	0.4	8	12	20	24	32	36	56
0.3	0.8	0.5	8	8	12	12	20	28	36
0.3	0.9	0.6	8	8	8	8	12	20	24
0.2	0.4	0.1	92	140	204	288	448	592	952
0.2	0.5	0.2	36	44	68	84	124	168	256
0.2	0.6	0.3	20	20	36	44	64	76	120
0.2	0.7	0.4	16	16	20	28	36	44	68
0.2	0.8	0.5	12	12	16	16	28	32	44
0.2	0.9	0.6	12	12	12	12	16	20	28
0.1	0.4	0.1	192	248	340	440	640	816	1224
0.1	0.5	0.2	56	76	100	124	180	224	332
0.1	0.6	0.3	32	36	48	60	80	104	148
0.1	0.7	0.4	20	20	32	36	52	60	84
0.1	0.8	0.5	16	16	20	28	32	36	56
0.1	0.9	0.6	12	12	16	16	20	28	36
0.05	0.4	0.1	288	364	464	604	816	1024	1468
0.05	0.5	0.2	88	108	136	168	224	280	388
0.05	0.6	0.3	44	52	64	84	108	132	176
0.05	0.7	0.4	24	32	36	48	64	72	104
0.05	0.8	0.5	20	20	24	32	36	48	64
0.05	0.9	0.6	12	16	20	20	24	32	40
0.01	0.4	0.1	532	644	776	952	1220	1468	2000
0.01	0.5	0.2	148	180	212	260	328	392	532
0.01	0.6	0.3	72	84	100	120	156	184	240
0.01	0.7	0.4	44	48	60	72	88	104	136
0.01	0.8	0.5	28	32	36	44	56	68	84
0.01	0.9	0.6	20	24	28	32	36	44	56
	$p_1 = 0.4$					β			
α	p_2	Δ	0.5	0.4	0.3	0.2	0.1	0.05	0.01
0.4	0.5	0.1	32	48	104	156	284	396	704
0.4	0.6	0.2	16	24	32	44	84	112	180
0.4	0.7	0.3	12	16	20	24	36	48	88
0.4	0.8	0.4	8	12	12	16	24	28	44
0.4	0.9	0.5	8	8	12	12	16	20	28
0.3	0.5	0.1	52	96	156	212	348	492	840

TABLE A1.1 *Continued*

0.3	0.6	0.2	28	32	44	56	100	128	216
0.3	0.7	0.3	12	20	28	32	44	68	100
0.3	0.8	0.4	8	12	12	24	28	36	52
0.3	0.9	0.5	8	8	12	12	20	24	36
0.2	0.5	0.1	112	140	232	320	492	644	1032
0.2	0.6	0.2	36	44	68	84	128	172	260
0.2	0.7	0.3	20	20	36	44	64	76	120
0.2	0.8	0.4	16	16	20	28	36	44	68
0.2	0.9	0.5	12	12	16	16	20	28	40
0.1	0.5	0.1	192	272	348	472	684	872	1332
0.1	0.6	0.2	56	76	104	136	180	228	336
0.1	0.7	0.3	32	36	48	60	80	104	148
0.1	0.8	0.4	20	20	32	36	48	56	84
0.1	0.9	0.5	12	16	20	24	32	36	48
0.05	0.5	0.1	312	400	508	644	892	1116	1592
0.05	0.6	0.2	88	108	136	172	232	284	408
0.05	0.7	0.3	44	52	64	84	108	132	176
0.05	0.8	0.4	24	32	36	48	60	72	96
0.05	0.9	0.5	20	20	24	28	36	40	56
0.01	0.5	0.1	572	696	844	1040	1336	1604	2164
0.01	0.6	0.2	156	184	224	264	344	412	556
0.01	0.7	0.3	72	84	100	120	156	184	240
0.01	0.8	0.4	44	48	56	68	84	100	132
0.01	0.9	0.5	28	28	36	40	48	60	76
	p_1 = 0.5					β			
α	p_2	Δ	0.5	0.4	0.3	0.2	0.1	0.05	0.01
0.4	0.6	0.1	32	48	104	156	284	396	704
0.4	0.7	0.2	16	24	32	44	84	108	176
0.4	0.8	0.3	12	16	20	24	32	44	80
0.4	0.9	0.4	8	12	12	16	20	28	40
0.3	0.6	0.1	52	96	156	212	348	492	840
0.3	0.7	0.2	28	32	44	56	96	124	208
0.3	0.8	0.3	12	20	24	32	44	56	92
0.3	0.9	0.4	8	12	12	16	28	32	48
0.2	0.6	0.1	112	140	232	320	492	644	1032
0.2	0.7	0.2	36	44	68	84	124	168	256
0.2	0.8	0.3	20	20	32	40	56	76	112
0.2	0.9	0.4	16	16	20	24	32	40	56
0.1	0.6	0.1	192	272	348	472	684	872	1332
0.1	0.7	0.2	56	76	100	124	180	224	332
0.1	0.8	0.3	32	36	48	56	80	96	140
0.1	0.9	0.4	20	20	24	32	40	48	68
0.05	0.6	0.1	312	400	508	644	892	1116	1592
0.05	0.7	0.2	88	108	136	168	224	280	388
0.05	0.8	0.3	40	48	60	72	96	120	168
0.05	0.9	0.4	24	28	32	40	52	60	84
0.01	0.6	0.1	572	696	844	1040	1336	1604	2164
0.01	0.7	0.2	148	180	212	260	328	392	532
0.01	0.8	0.3	68	80	92	112	140	168	224
0.01	0.9	0.4	36	40	48	60	72	84	112
	p_1 = 0.6					β			
α	p_2	Δ	0.5	0.4	0.3	0.2	0.1	0.05	0.01
0.4	0.7	0.1	32	48	100	148	240	372	656
0.4	0.8	0.2	16	20	28	40	60	100	156
0.4	0.9	0.3	12	16	16	24	32	40	60
0.3	0.7	0.1	52	96	128	204	332	468	764
0.3	0.8	0.2	24	32	40	56	92	116	192
0.3	0.9	0.3	12	16	20	28	40	48	72
0.2	0.7	0.1	92	140	204	288	448	592	952
0.2	0.8	0.2	32	44	56	80	112	144	224
0.2	0.9	0.3	20	20	28	32	48	60	92
0.1	0.7	0.1	192	248	340	440	640	816	1224
0.1	0.8	0.2	52	68	88	116	156	196	284
0.1	0.9	0.3	24	32	40	48	64	76	112
0.05	0.7	0.1	288	364	464	604	816	1024	1468
0.05	0.8	0.2	76	96	116	148	196	240	344
0.05	0.9	0.3	36	40	48	60	80	96	132
0.01	0.7	0.1	532	644	776	952	1220	1468	2000
0.01	0.8	0.2	132	160	188	228	288	344	464
0.01	0.9	0.3	56	68	76	92	112	136	176
	p_1 = 0.7					β			
α	p_2	Δ	0.5	0.4	0.3	0.2	0.1	0.05	0.01
0.4	0.8	0.1	32	48	68	136	212	292	532
0.4	0.9	0.2	16	20	28	36	52	68	124
0.3	0.8	0.1	52	88	120	180	272	392	636
0.3	0.9	0.2	16	24	40	48	68	88	144

TABLE A1.1 *Continued*

0.2	0.8	0.1	92	128	172	244	376	504	792
0.2	0.9	0.2	28	40	48	60	88	112	172
0.1	0.8	0.1	164	216	284	372	528	680	1008
0.1	0.9	0.2	44	52	68	88	116	152	216
0.05	0.8	0.1	244	308	392	500	676	844	1212
0.05	0.9	0.2	64	76	92	112	148	184	256
0.01	0.8	0.1	444	536	644	788	1008	1212	1648
0.01	0.9	0.2	104	120	144	172	216	260	344
	p_1 = 0.8					β			
α	p_2	Δ	0.5	0.4	0.3	0.2	0.1	0.05	0.01
0.4	0.9	0.1	32	44	60	88	168	224	364
0.3	0.9	0.1	52	68	88	136	200	276	436
0.2	0.9	0.1	72	96	132	180	260	348	536
0.1	0.9	0.1	116	156	200	264	368	464	684
0.05	0.9	0.1	176	216	272	348	464	576	820
0.01	0.9	0.1	312	372	444	540	688	824	1116
Note: This section differs from the others in that it is not for a specific p_1 value; rather, it defines p_1 and p_2 as functions of Δ, and gives the maximum sample size required for a given α, β, and Δ.									
	$p_1 = ½ (1 − Δ)$[A]					β			
α	p_2	Δ	0.5	0.4	0.3	0.2	0.1	0.05	0.01
0.4	0.55	0.1	32	48	104	156	296	400	708
0.4	0.60	0.2	16	24	32	44	84	112	180
0.4	0.65	0.3	12	16	20	24	36	48	88
0.4	0.70	0.4	8	12	12	16	24	28	44
0.4	0.75	0.5	8	8	12	12	16	20	32
0.3	0.55	0.1	72	96	156	212	352	496	848
0.3	0.60	0.2	28	32	44	56	100	128	216
0.3	0.65	0.3	12	20	28	32	44	68	100
0.3	0.70	0.4	8	12	20	24	32	36	56
0.3	0.75	0.5	8	8	12	12	24	28	36
0.2	0.55	0.1	112	140	232	320	496	648	1040
0.2	0.60	0.2	36	44	68	84	128	172	260
0.2	0.65	0.3	20	20	36	44	64	80	120
0.2	0.70	0.4	16	16	20	28	36	44	68
0.2	0.75	0.5	12	12	16	16	28	32	44
0.1	0.55	0.1	192	272	348	472	684	904	1340
0.1	0.60	0.2	56	76	104	136	180	228	336
0.1	0.65	0.3	32	36	48	60	80	104	148
0.1	0.70	0.4	20	20	32	36	52	60	84
0.1	0.75	0.5	16	16	20	28	32	36	56
0.05	0.55	0.1	312	404	508	644	892	1120	1600
0.05	0.60	0.2	88	108	136	172	232	284	408
0.05	0.65	0.3	44	52	64	84	108	132	188
0.05	0.70	0.4	24	32	36	48	64	72	104
0.05	0.75	0.5	20	20	24	32	36	48	64
0.01	0.55	0.1	572	700	844	1044	1340	1608	2200
0.01	0.60	0.2	156	184	224	264	344	412	556
0.01	0.65	0.3	72	84	100	120	156	184	240
0.01	0.70	0.4	44	48	60	72	88	104	136
0.01	0.75	0.5	32	36	36	44	56	68	88

[A] This calculation for p_1 generates the largest required sample sizes given α, β, and Δ.

APPENDIXES

(Nonmandatory Information)

X1. SAME-DIFFERENT TEST FOR DIFFERENCE: TESTING FOR DIFFERENCES BETWEEN TWO FORMULATIONS OF LEMONADES

X1.1 Background

X1.1.1 A juice company has recently developed a new formulation for their brand of lemonade, with a slightly higher sucrose concentration. They wish to know if there is a perceivable difference between their existing product and the one with the new formulation before moving to consumer preference testing.

X1.2 Objective

X1.2.1 To confirm perceivable differences between the existing product and the new product with an increased level of sucrose. If this is the case, consumer preference testing can then be conducted.

X1.3 Number of Assessors

X1.3.1 Since the product developers wish to strongly protect themselves from stating there are perceivable differences when there are none, a Type I error rate of 5 % is decided upon ($\alpha = 0.05$). They also want to be 80 % certain of detecting a 30 % difference between the proportion of assessors who would correctly identify the unmatched and the proportion who would say the matched pair was *different* in the population. Prior experience indicates that about 30 % of the population will declare a matched pair of samples as *different*. Consulting Table A1.1 for $\alpha = 0.05$, $\beta = 0.2$, $p_1 = 0.3$ and $\Delta = 0.3$, we find that 84 assessors are required. These 84 assessors were chosen at random from a database of consumers.

X1.4 Test Procedure

X1.4.1 The samples were prepared in the same manner in advance of the testing. The appearance of samples A and B (existing lemonade and new formulation, respectively) were identical. ASTM serving protocols were observed.

X1.4.2 An equal number of matched and unmatched pairs were presented (42 of each), with each of the four possible pairs seen an equal number of times. Pairs were randomly assigned to assessors upon arrival. The components of each pair were presented simultaneously for testing.

X1.4.3 Data were collected using a computerized data collection system. Sample set identification, assessor information, and results were related by a unique registration code given to each assessor.

X1.4.4 Panelists were given instructions on the task and the evaluation procedure before testing. Panelists entered their response by selecting the appropriate box on the screen using a light pen. Data were automatically collected.

X1.5 Analysis and Interpretation of Results

X1.5.1 Data were tabulated as shown in Section 11 (see Table X1.1). Of those 42 panelists who received a matched pair, 21 responded *same* and 21 responded *different*. Of those 42 panelists who received an unmatched pair, 9 responded *same* and 33 responded *different*.

TABLE X1.1 Results of the Same Different Test

		Assessor Received		
		Matched Pair (AA or BB)	Unmatched Pair (AB or BA)	Total
Assessor's Response	Same:	21	9	30
	Different:	21	33	54
	Total:	42	42	84

X1.5.1.1 Since the number of *different* responses from those who received the unmatched pairs (33) is more than the number of *different* responses from those who received the matched pairs (21), we need to proceed with a statistical test to determine if the difference between the formulations is perceptible.

X1.5.2 *Fisher's Exact Test:*

X1.5.2.1 Values used in the computation of the FET *p*-value come from the observed frequency table as shown in Table X1.1. Specifically, the total number receiving an unmatched pair (42), the total number of responses for *different* (54), the number of *different* responses of those given an unmatched pair (33), and the total number of responses (84) are all used in the calculation.

X1.5.2.2 The *p*-value of a one-sided FET is computed by summing the probabilities of the tables supporting the alternative hypothesis; that is, the probabilities for the tables where the number of *different* responses from the unmatched pair group is increased from its observed value to its maximum possible value. This is explained in detail below.

X1.5.2.3 Maintaining the total number of *same* responses and the total number of *different* responses (row totals) and sample sizes (column totals) from the example table above (30, 54, 42, and 42, respectively), a table where the number of *different* responses received by the unmatched pair is set as *i* is given below:

		Assessor Received	
		Matched Pair (AA or BB)	Unmatched Pair (AB or BA)
Assessor's Response	Same:	42-(54-*i*)	42-*i*
	Different:	54-*i*	*i*
	Total:	42	42

X1.5.2.4 The probability of this specific table will be denoted as Q_i since each cell of the table is a function of *i*. The maximum allowable value of *i* is the smaller of the total number of *different* responses (54) and the number of assessors receiving the unmatched pair (42). The one-sided FET *p*-value, *p*, is then computed as follows:

$$p = \sum_{i=33}^{\min(5442)} Q_i \qquad \text{(X1.1)}$$

where Q_i is given by:

$$Q_i = \frac{C(54,i)\cdot C(84-5442-i)}{C(8442)} \qquad \text{(X1.2)}$$

and where *C(n,r)* represents the number of combinations of picking *r* items from *n* objects, and is *n* factorial divided by the product of *r* factorial and *(n-r)* factorial.

X1.5.2.5 Therefore the *p*-value for Fisher's Exact Test is given by:

$$\sum_{i=33}^{\min(5442)} \frac{\binom{54}{i}\binom{84-54}{42-i}}{\binom{84}{42}} = \sum_{i=33}^{42} = \frac{\binom{54}{i}\binom{30}{42-i}}{\binom{84}{42}} \quad (X1.3)$$

$$= \frac{\binom{54}{33}\binom{30}{42-33}}{\binom{84}{42}} + \frac{\binom{54}{34}\binom{30}{42-34}}{\binom{84}{42}}$$

$$+ \ldots + \frac{\binom{54}{42}\binom{30}{42-42}}{\binom{84}{42}}$$

$$= 0.004434187 + 0.001120403 + 0.000222689 + 0.0000342796$$

$$+ 0.00000400238 + 0.000000344334 + 0.0000000209282$$

$$+ 0.000000000840865 + 0.0000000000198017$$

$$+ 0.000000000000204303$$

$$= 0.005815927$$

This *p*-value is for a one-sided alternative hypothesis.

X1.5.3 *Continuity Corrected-Chi-Square:*

X1.5.3.1 Since all of the expected frequencies are greater than 5 (see below), the table is not sparse and we may use a one-tailed continuity corrected χ^2-statistic to analyze the resulting data (this could in fact replace the above FET computation if desired).

E11 = (30)(42)/84 = 15
E12 = (30)(42)/84 = 15
E21 = (54)(42)/84 = 27
E12 = (54)(42)/84 = 27

X1.5.3.2 The computation for the calculated Chi-square is given as follows:

$$\chi^2 = \sum_{i=1}^{2}\sum_{j=1}^{2} \frac{(\max\{0,(|O_{ij}-E_{ij}|-0.5)\})^2}{E_{ij}} \quad (X1.4)$$

where O_{ij} is the observed frequency in row *i* and column *j*, and E_{ij} is the expected frequency in row *i* and column *j*. Applying the formula to the data in Table X1.1 gives:

$$\chi^2 = \frac{(|21-15|-0.5)^2}{15} + \frac{(|9-15|-0.5)^2}{15} \quad (X1.5)$$

$$+ \frac{(|21-27|-0.5)^2}{27} + \frac{(|33-27|-0.5)^2}{27}$$

$$= 6.2741$$

X1.5.3.3 Thus, the calculated Chi-square value is 6.2741. The critical value from the Chi-square distribution for a one sided test with 1 degree of freedom and α = 0.05 is 2.71. Since our calculated value exceeds this critical value, the samples are perceivably different at this level of significance (5 %).

X1.5.3.4 Note that both tests gave the same results which is not surprising since the continuity corrected χ^2 test is a good approximation to the FET when the table is not sparse.

X1.6 Report and Conclusions

X1.6.1 The test showed that untrained assessors can perceive the difference between the existing product and the new product with an increased level of sucrose. The test product is therefore suitable for a market test to determine consumer preference in comparison with the existing product.

X2. SAME-DIFFERENT TEST FOR SIMILARITY: APPROVAL OF AN ALTERNATE VENDOR FOR A LIQUEUR INGREDIENT

X2.1 Background

X2.1.1 A major beverage alcohol manufacturer sources a special ingredient for its best selling liqueur from a single world vendor. Sales volumes are increasing faster than projected and there is some doubt that the single vendor can supply enough of the ingredient to meet demand. The purchasing department has identified a potential alternate vendor who can supply the ingredient. The manufacturer wants to be sure that the ingredient supplied by the alternate vendor does not significantly change the sensory characteristics of the liqueur. The alcohol content of the liqueur is 20 %. Sensory services recommends a same-different test because, at this alcohol level, any more than two samples may lead to fatigue and decreased sensory acuity.

X2.2 Objective

X2.2.1 The objective of this same-different test is to qualify the alternate vendor by determining that the liqueur made using the ingredient supplied by the alternate vendor is not significantly different from the control liqueur made with the ingredient supplied by the current vendor.

X2.3 Number of Assessors

X2.3.1 The liqueur brand is a best seller and is growing, therefore, we want to minimize the beta error, the chance of concluding that the samples are the same when, in fact, they are different. We want to be 90 % certain of detecting a 30 % difference in the proportion of assessors who would identify the unmatched pair as *different* compared to the proportion who would say the matched pair was *different*. It is expected that about 30 % of those receiving the matched pair will say that the samples are *different*. Therefore β = 0.1, p_1 = 0.3, and Δ = 0.3. Using Table A1.1 we find that for these values and for α = 0.2, 64 assessors are needed. The company has a pool of 64 employees who can be used as assessors. These employees are not trained descriptive assessors, but they have extensive experience in the use of the same-different method.

X2.4 Test Procedure

X2.4.1 There are four possible serving orders: AA, AB, BB, and BA. Each order is represented 16 times and recorded in the serving order worksheet. Each assessor receives one pair of samples to evaluate.

X2.5 Analysis and Interpretation of Results

X2.5.1 The table below indicates the results of the test in a two-by-two contingency table.

		Assessors Received		
		Matched Pair (AA or BB)	Unmatched Pair (AB or BA)	Total
Assessor's Response	Same:	17	14	31
	Different:	15	18	33
	Total:	32	32	64

X2.5.1.1 The number of *different* responses from those who received the unmatched pair (18) is greater than the number of *different* responses from those who received the matched pair (15). We need to go forward with a statistical test.

X2.5.2 *Fisher's Exact Test:*

X2.5.2.1 Using the procedure described in X1.5.2, the FET *p*-value for the above data is given by:

$$p = \sum_{i=18}^{\min(3332)} \frac{\binom{33}{i}\binom{64-33}{32-i}}{\binom{64}{32}} = \sum_{i=18}^{32} = \frac{\binom{33}{i}\binom{31}{32-i}}{\binom{64}{32}} \quad \text{(X2.1)}$$

$$= \frac{\binom{33}{18}\binom{31}{32-18}}{\binom{64}{32}} + \frac{\binom{33}{19}\binom{31}{32-19}}{\binom{64}{32}}$$

$$+ \ldots + \frac{\binom{33}{32}\binom{31}{32-32}}{\binom{64}{32}}$$

$$= 0.309$$

X2.5.2.2 With a *p*-value of 0.309, we can conclude that the samples are not perceivably different.

X2.5.3 *Continuity Corrected Chi-square:*

X2.5.3.1 Alternately, since all of the expected frequencies are greater than 5 (as shown below), the continuity corrected χ^2 test can be used instead of the FET.

Expected frequencies:
E11 = (31)(32)/64 = 15.5
E12 = (31)(32)/64 = 15.5
E21 = (33)(32)/64 = 16.5
E12 = (33)(32)/64 = 16.5

$$\chi^2 = \frac{(|17-15.5|-0.5)^2}{15.5} + \frac{(|14-15.5|-0.5)^2}{15.5} \quad \text{(X2.2)}$$

$$+ \frac{(|15-16.5|-0.5)^2}{16.5} + \frac{(|18-16.5|-0.5)^2}{16.5}$$

$$= \frac{1}{15.5} + \frac{1}{15.5} + \frac{1}{16.5} + \frac{1}{16.5}$$

$$= 0.250244379$$

X2.5.3.2 For the 80 % confidence level (α = 0.20) of a one-sided test with one degree of freedom, a Chi-square value of 0.708 or greater is significant.

X2.6 Report and Conclusions

X2.6.1 The Chi-square value of 0.250 is less than the critical value of 0.708. Thus, there is no significant difference between the liqueur made by using the ingredient supplied by the alternate vendor and the control liqueur made with the ingredient supplied by the current vendor. Sensory services recommends considering the alternate vendor as an ingredient supplier for this liqueur.

Same-Different Test for Liqueurs–Serving Order Worksheet

Assessor #																Pair Presented
1	5	9	13	17	21	25	29	33	37	41	45	49	53	57	61	AA
2	6	10	14	18	22	26	30	34	38	42	46	50	54	58	62	AB
3	7	11	15	19	23	27	31	35	39	43	47	51	55	59	63	BB
4	8	12	16	20	24	28	32	36	40	44	48	52	56	60	64	BA

FIG. X2.1 Serving Order Worksheet Example Used for the Same-Different Test

Ballot # _______________

Same-Different Test for Liqueurs

Assessor Number ___________________ Date ____________________

Instructions: You will receive two samples of a liqueur. These samples may be the same or they may be different.

Evaluated the samples in the order that they are presented from left to right.

Indicate whether the samples are the same or different by checking the appropriate line below.

The two samples are the same ______________________________

The two samples are different ______________________________

Comments:

FIG. X2.2 Assessor Ballot Example Used for the Same-Different Test

REFERENCES

(1) Herz, R. S., and Cupchik, G. C., "An Experimental Characterization of Odor-Evoked Memories in Humans," *Chemical Senses*, Vol 17, No. 5, 1992, pp. 519-528.

(2) Todrank, J., Wysocki, C. J., and Beauchamp, G. K., "The Effects of Adaptation on the Perception of Similar and Dissimilar Odors," *Chemical Senses*, Vol 16, No. 5, 1991, pp. 476-482.

(3) Steel, R. G. D., and Torrie, J. H., *Principals and Procedures of Statistics* , McGraw-Hill Inc., 1980, pp. 504-506.

(4) Lehmann, E. L., *Testing Statistical Hypotheses*, J. Wiley & Sons, New York, 1959, p. 143.

(5) Kendall, M., and Stuart, A., *The Advanced Theory of Statistics*, Vol II, Hafner Publishing Co., New York, 1961.

(6) O'Mahony, M., *Sensory Evaluation of Food, Statistical Methods and Procedures*, Marcel Dekker, Inc., New York, 1986, pp. 102-104.

(7) O'Mahony, M., *Sensory Evaluation of Food, Statistical Methods and Procedures*, Marcel Dekker, Inc., New York, 1986, pp. 101-102.

(8) O'Mahony, M., *Sensory Evaluation of Food, Statistical Methods and Procedures*, Marcel Dekker, Inc., New York, 1986, pp. 45-56.

(9) Haseman, J. K., "Exact Sample Sizes for Use with the Fisher-Irwin Test for 2·2 Tables," *Biometrics*, Vol 34, 1978, pp. 106-109.