Testing for Sensory Threshold of Added Substances

Morten C. Meilgaard, *International Consultant, 2938 Moon Lake Drive, West Bloomfield, MI 48323*

ABSTRACT

The Ascending Method of Limits test of the American Society for Testing and Materials (ASTM) was adopted for inclusion in the Methods of Analysis of the ASBC in 1980. The ASTM recently relegated it to the status of rapid method and adopted the procedure described in this article as its reference standard practice. In the Ascending Method of Limits test, a panelist receives only one three-alternative forced-choice (3-AFC) set per concentration; in the new ASTM practice, the panelist receives a minimum of five 3-AFC sets per concentration. The new practice permits the use of iterative computer programs and estimates the threshold itself rather than a number somewhere near the threshold. This review discusses the merits of adopting the new practice as a reference procedure and the Ascending Method of Limits test as a rapid procedure.

Keywords: Sensory analysis, Threshold, Ascending Method of Limits, ASTM standard practice E 1432

Sensory thresholds are difficult to define; a good determination requires hundreds of taste tests (7), and results are not very reproducible (2). Yet the determination of thresholds is the only method by which a number of flavor problems can be investigated to determine, for example, the following: which of hundreds of substances found in beer will influence its flavor, the magnitude of such activity for substances found to be flavor active (9), whether instrumental detection limits are low enough, and the relative sensitivity of individuals to a given flavor-active substance.

In 1980, after five years of testing and considering more than a dozen methods (2), the ASBC Technical Committee adopted for inclusion in Methods of Analysis (1) the Ascending Method of Limits test of the American Society for Testing and Materials (ASTM), which the ASTM itself had adopted as a standard practice in 1979 (3). Because threshold methods are so laborious, they are always a compromise falling more or less short of the desired accuracy. A choice must be made as to how many comparisons (sets) can be served and among how many panelists these sets will be distributed. A principal reason for the choice (2) was the observation, made during collaborative testing, that variations in sensitivity among individuals are so large that the group threshold and its confidence limits are of relatively little interest. Of much greater interest in practice is the frequency distribution of the threshold values of the individual panelists. In the Ascending Method of Limits test, each panelist receives a single 3-AFC set (consisting of two controls and one test) at each of approximately six concentrations. Thus, relatively little experimental effort is devoted to each panelist, and the resulting Best Estimate Thresholds are not very accurate. However, many more panelists can be tested, and the shape of the frequency distribution can be obtained with good accuracy because the individual inaccuracies tend to balance each other.

During the 10 years since the adoption of the Ascending Method of Limits test, the method has been used with satisfactory results by members of both the ASTM and the ASBC. Criticism has been expressed by investigators and in publications (11,12) that—depending on chance, the choice of concentration scale steps, and the location of the ultimate threshold on the scale of concentrations presented—the results for group thresholds may occasionally be biased by a factor of 1.5–3 or more, both upward and downward. The chance of bias was found to be larger the further the threshold was from the center of the range of scale steps presented. As a result of such criticism, the ASTM Practice E 679 failed its regular five-year reballotting process. On May 23, 1990, ASTM Task Group E 18.04.25 recommended that Practice E 679 be relegated to the status of rapid method and a new practice (4) (see Appendix) adopted in its place.

DISCUSSION

Definition of Sensory Threshold

The two methods use different definitions. The Ascending Method of Limits test determines a "best estimate threshold" (BET) for the individual panelist as "the geometric mean of the highest concentration missed and the next higher (adjacent) concentration" (1). The group threshold is the geometric mean of the individual BETs. The new ASTM practice (4), on the other hand, determines a "statistical" threshold for each panelist, namely, "that concentration for which the probability of detection is 0.5." The group threshold is the number that best represents the central tendency of the distribution of individual panelist thresholds.

Ascending Method of Limits Test (1)

Figure 1 is the example used in ASBC method Sensory Analysis 9 (1); 24 additional examples are included in Meilgaard et al (10). No use is made of the frequency of correct results below the highest concentration missed. Panelists 24 and 38 in Figure 1 are both assigned a BET of 57 ppb, yet panelist 38 missed only one of six sets and is likely to be more sensitive than panelist 24, who missed three. This means that good sensory data are not used. The group threshold is invariably calculated as the geometric mean, whether or not the distribution of individual BET values fits the expected (log-normal) pattern.

Procedure: ASBC Sensory Analysis-9 (Ascending Method of Limits Test) (1)
Sample: Dimethyl sulfide, MCB catalog MX 1445
Purification: 4-Adsorbent procedure
Medium Used: Stroh Bohemian lager, 11.4°P, 14 BU, 50 µg/L DMS
Number of scale steps: 6 *Dilution factor per step:* 2.0
No. of assessors: 16 *Form of test used:* simple ☑ extended ☐

Panelist	Concentrations presented, µg/L 10	20	40	80	160	320	Best Estimate Threshold µg/L	Log_{10}
12	0	0	+	0	0	+	226	2.35
17	0	+	+	+	+	+	14	1.15
24	0	0	0	+	+	+	57	1.75
31	+	+	+	+	+	+	7.1	0.85
38	+	+	0	+	+	+	57	1.75
44	0	0	+	+	+	+	28	1.45
55	+	0	0	+	+	+	57	1.75
59	0	+	0	+	0	0	453	2.65
64	0	0	+	+	+	+	28	1.45
68	0	+	+	+	+	+	14	1.15
72	0	+	0	+	+	+	57	1.75
77	0	0	+	+	+	+	28	1.45
83	0	+	+	+	+	+	14	1.15
87	0	0	+	+	+	+	28	1.45
91	+	+	+	+	+	+	7.1	0.85
94	0	0	+	+	+	+	28	1.45
							Sum →	24.40
				Group BET, geometric mean, µg/L			33.5 ←	1.525
				Log standard deviation = 0.471				

Histogram of Individual BE Thresholds:

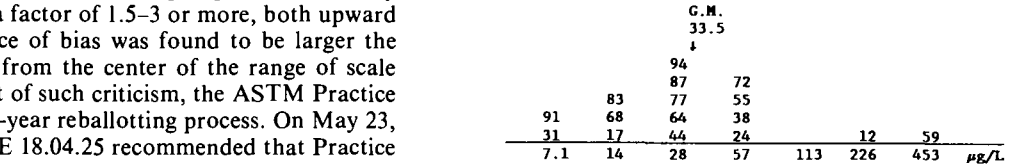


Fig. 1. Example of calculation and presentation of results, from ASBC method Sensory Analysis 9, Ascending Method of Limits Test (1). + = correct choice, 0 = incorrect choice.

The New ASTM Standard Practice (4)

The procedure adopted for individual thresholds can be illustrated by an example. If a watch is held away from the ear, a distance can be found at which part of the time one can hear it ticking and part of the time one cannot. Moving it farther away, a distance can be found at which the watch can be heard perhaps 30% of the time; moving it closer, a distance can be found at which it can be heard 90% of the time. The threshold is the distance at which the watch can be heard 50% of the time. Likewise, panelist 12 in Figure 1 might have a 12% chance of detecting an added substance at 80 ppb, perhaps a 40% chance at 160 ppb, and a 90% chance at 320 ppb. The new practice is designed to determine these probabilities. Figure 2 continues the hypothetical example. By chance alone, a 3-AFC test delivers 33% correct responses. Assuming that the responses approximate a normal distribution, the task is to fit a sigmoid curve from 0 to 100% above chance, as shown in Figure 2. In this case, all the collected data are used.

The new practice allows five ways of fitting the sigmoid (see Appendix, "Calculate the threshold for each panelist"). Except for the inaccuracies introduced by fitting a straight line by eye, the first three ways produce the same result: the use of probability paper, the use of probits, and the use of tables of the normal standard deviate. In these three methods, the "percent correct above chance" (Fig. 2) is converted to a probability measure. One major source of bias is that results of 100 and 0% cannot be accommodated, and certainly not results of less than 0% above chance. These are arbitrarily handled by substituting 99.5 for 100%, 0.5 for 0%, and 0.1 for less than 0%.

The fourth and fifth methods employ computer curve fitting using an iterative program such as SAS Proc Nlin (13). This is a nonarbitrary way of fitting the sigmoid. The dilemma of such results as 100 or 0% is resolved by assigning higher weights to results near 50% and lower weights to those at the extremes. It matters little whether the program uses logits of the logistic model (6) or probits, the formula for the normal probability curve (8), as 5,000 optimally located data points are required to distinguish between them empirically (A. W. MacRae, School of Psychology, University of Birmingham, U.K., *personal communication* to ASTM task group E-18.04.25). (Probits are standard normal deviates plus five, to avoid negative numbers.) The normal probability curve requires a more complex program. A few other formulas are used by toxicologists (14). Results in practice are indistinguishable at the threshold but vary when the formulas are used to predict responses at extremes of the frequency curve, e.g., the concentration of an added substance that can be detected 1% of the time.

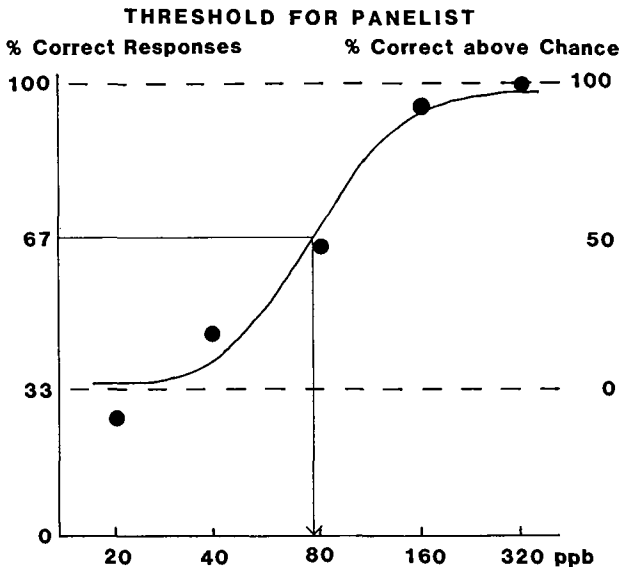


Fig. 2. Example of test situation according to ASTM Standard Practice E 1432 (4). Determination of threshold for a panelist who received five or more three-alternative forced-choice sets at each of the five concentrations shown. The task was to fit a sigmoid approaching 0% correct above chance at very low concentrations and approaching 100% correct above chance at concentrations higher than those shown. The threshold is the concentration that corresponds to 50% correct above chance.

Population Thresholds, Confidence Limits

Is the distribution normal? According to Snedecor and Cochran (15), this question is best answered after converting the distribution of individual thresholds into normal deviates, for which they present two formulas. According to French Standard NF X 43-101 (5), from which Figure A4 in the appendix was adapted, "If a straight line can be drawn through the points, conclude that the individual thresholds are normally distributed and representative of the population of which the panel is a sample. The panel threshold is the concentration corresponding to the 50% rank position. Approximate panel confidence limits are those concentrations corresponding to the 16 and 84% rank positions." The French standard uses a third, slightly different conversion formula (there is no one accepted form). It also contains text stating, in effect, that if the point of interest is not the 50% point but, as in the case of an off flavor, the concentration that can be detected by 10% or even 5% of the population, then that number can be read off the graph.

Often, it is not easy to decide whether a straight line can be drawn, or how it should be drawn. Figure 3 is a rank-probability plot of the data in Figure 1. The solid line represents all 16 panelists; the dashed line, which shows a much better fit, omits the two panelists with thresholds at 226 and 453 ppb as outliers. The two lines show only a small difference in the threshold but a larger difference in the confidence limits and a much larger one again at the extremes. The choice between the two lines may depend on the purpose of the test. In most cases, the logical

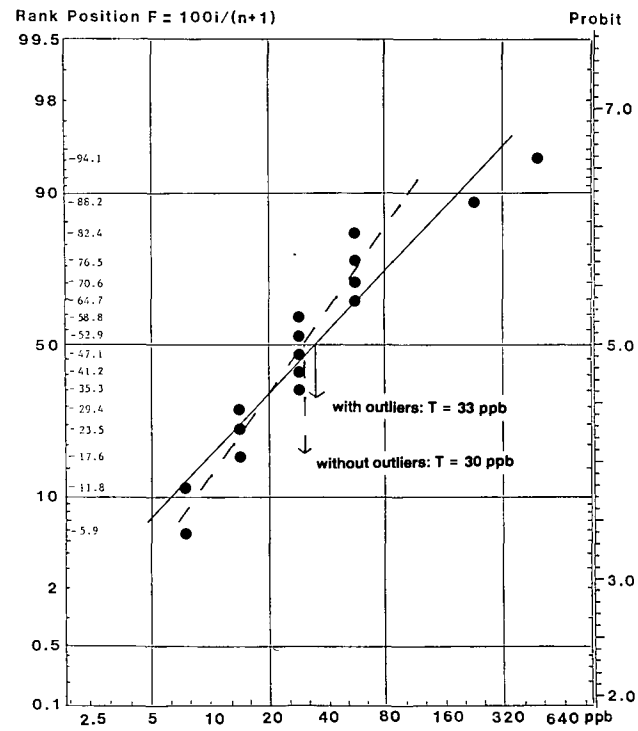


Fig. 3. Rank-probability plot of the data for the panelists in Figure 1. The group threshold (T) is the concentration that corresponds to the 50% point. The solid line represents all 16 panelists; the dashed line omits as outliers the two panelists with thresholds at 226 and 453 ppb (μg/L).

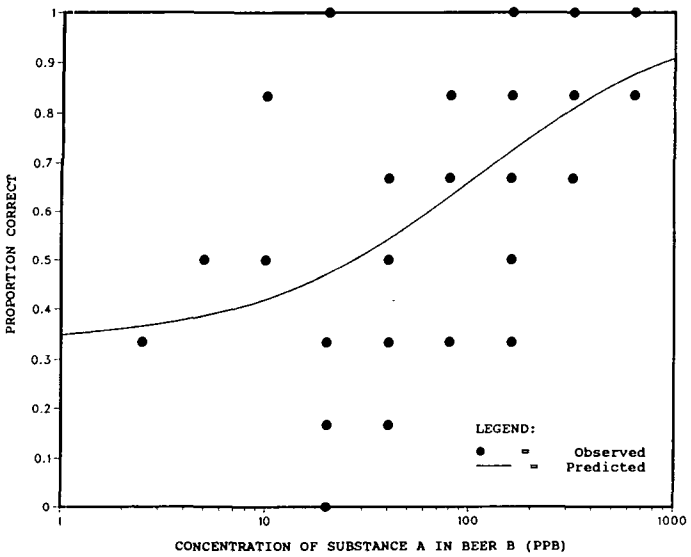


Fig. 4. Plot of the observed and predicted proportion of correct responses (from example 1 in the appendix), using iterative regression by the SAS Proc Nlin program and the logistic method. The 35 data points in the example are treated here as if they came from a single panelist. This procedure is not permitted unless previous experimentation indicates that the panel members' thresholds are normally distributed.

decision is to go with the dashed line and regard the resulting threshold and confidence limits as representative of a main population that excludes a certain number of outliers. However, if the purpose is to learn how difficult it may be for some people to detect a substance, then the solid line may be chosen.

Other test series produce different results, and it may be advisable to run a collaborative series before adopting the procedure and interpretation suggested by the French standard.

When is it permissible to combine the data from all panelists into a single computational procedure? The computation in Figure 4 combines the 35 data points for the six panelists in example 1 of the new practice (see Appendix). One might argue that this constitutes a way to calculate the panel threshold in a single step using all the data. However, doubt remains as to the degree to which such calculations are permissible and represent the population. The ASTM subcommittee voted 27 to 2 against allowing the procedure "unless it has been proven by prior experimentation that the panel members' thresholds are normally distributed." Again, a collaborative series is desirable as a source of actual examples.

ACKNOWLEDGMENTS

The author is grateful to ASTM Subcommittee E-18.04 on Sensory Analysis of Products and Materials for permission to cite work in progress, to B. Thomas Carr for performing the calculations in Figures 4 (in the article) and A3 (in the appendix), and to the Stroh Brewery Co. for the use of its facilities.

LITERATURE CITED

1. American Society of Brewing Chemists. *Methods of Analysis,* 7th ed. Sensory Analysis 9 Threshold of added substances—Ascending method of limits test (international method). The Society, St. Paul, MN, 1976.
2. American Society of Brewing Chemists. Report of subcommittee on sensory analysis. *Journal* 38:99-107, 1980.
3. American Society for Testing and Materials. E 679-79 *Standard practice for determination of odor and taste thresholds by a forced-choice ascending concentration series method of limits.* The Society, Philadelphia, PA, 1979.
4. American Society for Testing and Materials. E 1432 *Standard practice for defining and calculating individual and group sensory thresholds from forced-choice data sets of intermediate size.* The Society, Philadelphia, PA, 1991.
5. Association Française de Normalisation: *French standard NF X 43-101 Air quality: Method of measuring the odor intensity of gaseous effluents. Determination of the dilution factor to perception threshold.* The Association, Paris, 1986.
6. Bishop, Y., Fienberg, F., and Holland, P. *Discrete Multivariate Analysis,* MIT Press, Cambridge, MA, pp. 357-358, 1980.
7. Brown, D. C. W., Clapperton, J. F., Meilgaard, M. C., and Moll, M. Flavor thresholds of added substances. *J. Am. Soc. Brew. Chem.* 36:73-80, 1978.
8. Finney, D. J. *Probit Analysis,* 3rd ed. Cambridge University Press, Cambridge, U.K., 1971.
9. Meilgaard, M. C. Prediction of flavor differences between beers from their chemical composition. *J. Agric. Food Chem.* 30:1009-1017, 1982.
10. Meilgaard, M. C. Reid, P. S., and Wyborski, K. A. Reference standards for beer flavor terminology. *J. Am. Soc. Brew. Soc.* 40:119-128, 1982.
11. Morrison, G. R. Flavour thresholds for added substances. *J. Inst. Brew.* 88:167-169, 1982.
12. Morrison, G. R. Measurement of flavour thresholds. *J. Inst. Brew.* 88:170-174, 1982.
13. SAS Institute. *User's Guide: Statistics,* Version 5 edition. The Institute, Cary, NC, pp. 575-606, 1985.
14. Salsburg, D. *Statistics for Toxicologists.* Marcel Dekker, New York, 1986.
15. Snedecor, C. W., and Cochran, W. C., *Statistical Methods,* 7th ed. Iowa State University Press, Ames, pp. 59-63, 1980.

APPENDIX

ASTM STANDARD PRACTICE (E 1432)

The instructions below were rewritten in ASBC style; example 1 was adapted to fit determination of the threshold for a substance added to beer. The figures are reproduced with permission of ASTM Subcommittee E-18.04. This practice is not an official ASBC method.

Scope

The purpose of this practice is to determine individual sensory thresholds and, when appropriate, calculate group thresholds. The starting point of the practice is any sensory threshold data set of at least 20–40 presentations per individual, collected by a forced-choice procedure. The usual procedure of presentation is the three-alternative forced-choice (3-AFC).[1] For small group data sets totaling 50–100 3-AFC presentations, use ASBC method Sensory Analysis 9, Threshold of Added Substances—Ascending Method of Limits Test (International Method) (1).

It is recognized that precise thresholds for a given substance in a given medium do not exist. The probability of detection by a given panelist is typically 0.0 at low concentrations of a substance added to beer and 1.0 at high concentrations. Also, there is a range of concentrations in which the probability is between these limits. Threshold is defined as the concentration at which an individual can detect an added substance with a probability of 0.5 (i.e., 50% above chance under the conditions of the test). The method used to calculate the group threshold depends on the distribution of the individual thresholds.

The user must keep in mind the large degree of random error associated with estimating the probability of detection from, say, 100–300 3-AFC sets.

[1]The 3-AFC procedure is one of the set of *n*-AFC procedures in which the panelist receives *n* samples, one of which contains the added substance. Any such procedure can be used to measure sensory thresholds, as can the duo-trio, the triangular, and the two-out-of-five procedures. This practice does not apply to free-choice procedures.

The reliance placed on the results can be greatly increased by enlarging the panel and replicating the tests to encompass 300–1,000 3-AFC sets. The results and their relevance are greatly influenced by the composition as well as the size of the panel and by the degree of training the individuals have had with the added substance (discussed later).

Principle of the Practice

Obtain from each panelist 3-AFC data sets taken at five or more concentration scale steps (usually six or seven), each step usually differing from the previous one by a factor of 2–4 (typically 3.0). The practice presupposes that the range of concentrations has been selected by pretesting so as to ensure that the individual's threshold falls neither outside nor near the ends of the range but well within it. At each concentration step, test the individual several times (typically five or more times). Calculate the threshold for each panelist by one of the methods described below.

To obtain the threshold for one panelist graphically or by linear regression, first calculate the proportion correct above chance by deducting from the proportion of correct choices the proportion (1/3) that would have been selected by chance in the absence of the stimulus. From that, calculate the concentration that has a 0.5 probability of being detected under the test conditions either graphically, using probability paper, or by linear regression, after converting the proportion correct above chance into probits or into standard normal deviates.

Alternatively, obtain the panelist's threshold directly from the proportion of correct choices by nonlinear iterative regression using a computer program. Programs using the probit model (the normal distribution) or the logistic model are suitable.

Always report the individual thresholds of the panelists. Depending on the purpose for which a threshold is required (see below) and on the distribution found, a group threshold may be calculated in a second step as the arithmetic or geometric mean, the median, or another measure of central tendency of the individual thresholds; or, it may be concluded that a group threshold cannot be calculated.

Pooling data sets from panel members to produce a single-step calculation of the panel threshold is not permitted unless it has been proved by previous experimentation that the panel members' thresholds are normally distributed.

Purpose of Test, Composition of Panel

Purpose of test. Choose the panel variables according to the purpose for which the resulting threshold is needed. The important variables are 1) the number of tests per panelist, 2) the number of panelists, 3) the selection of panelists to represent a given population, and 4) the panelists' degree of training. Regarding the purpose of testing, it is useful to distinguish three situations:

1. Comparing an individual's threshold with a literature value. This situation, which is the simplest, requires at least 20–40 3-AFC presentations to the panel member in question. A number of training sessions may be required to establish the range of concentrations that will be used and to make certain that the individual is fully familiar with the stimulus to be detected and with the mechanics of the test.

2. A population threshold is required, e.g., the flavor threshold of consumers for a given contaminant. In this situation, recourse to the rules of sampling from a population is necessary (5) and requires the following: 1) The population must be accurately defined and delimited, 2) the sample drawn must be truly random, (i.e., every member of the population must have a known chance of being selected), and 3) the degree of variation occurring within the population must be known or can be acquired while formulating the plan of sampling.

In practice, the cost and availability of panelists place serious limitations on the degree to which one can cover population factors affecting thresholds, e.g., age-group, gender, ethnic origin, health status, smoking status, and degree of training. The experimenter usually is limited to panels of 5–15 members, each receiving 20–40 3-AFC presentations, for a total of 100–600 presentations. If the resulting thresholds are to be valid for the population, the experimenter should calculate and tabulate the thresholds for each individual, repeat the test for those who fall well beyond the range of the rest of the panel (outliers), adjust the range and repeat the test for those whose threshold at first did not fall well within the range of samples presented, and (if necessary) repeat the test series with a second or third panel sampled from the same population to obtain the desired level of precision.

One approach to the question of a value that is generalizable to the population is to try and estimate the threshold of healthy 20-year olds. According to the Institute for Olfactory Studies (J. E. Amoore, Richmond, CA, *personal communication*), odor threshold concentrations double approximately for each 22 years of age in persons aged 20–65.

3. The distribution of thresholds in the population is required. For example, to determine which proportion of the population is affected by a given level of a contaminant, or (conversely) to determine which concentrations of a contaminant will affect a given percentage of a population. The requirements for testing are the same as discussed above, except that it is even more important to cover the range well, i.e., to repeat the tests for individuals whose thresholds fall in thinly populated parts of the panel range. Consideration should be given to increasing the number of presentations per concentration from 5–7 to 7–10 for such panel members. If the individual thresholds are plotted as in Figure 3 of the preceding article, any sector needing study will be apparent from the graph.

Trained Versus Casual Observers. Thresholds should normally be determined for observers trained by repeated exposure to detect

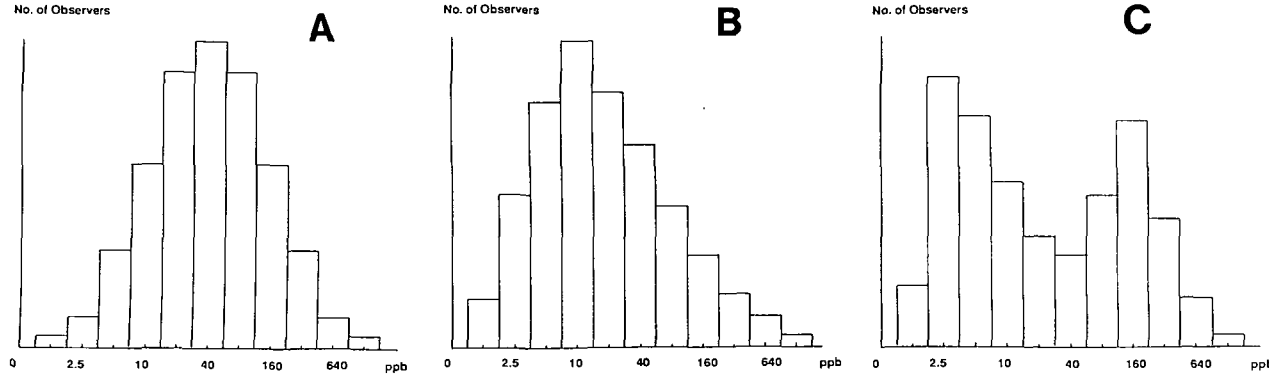


Fig. A1. Examples of frequency distributions of individual thresholds: **A**, symmetrical, bell-shaped (log-normal); **B**, skewed; **C**, bimodal. Concentrations in μg/L.

the stimulus in question whenever it is present, but if the threshold sought is that of a casual observer, then naive panelists and mild distraction (such as noise) may be used (2).

Choosing the Measure of Central Tendency. The report should contain a table or graph giving the individual thresholds of each observer. If a group threshold is required, the measure of central tendency chosen should be that which best represents the distribution obtained. In a few cases, the results form a symmetrical, bell-shaped distribution (Fig. A1, A), and the arithmetic mean may be used. With sensory data, the distribution is typically skewed (Fig. A1, B), and the median may be used, or the distribution may be normalized by converting the concentration units to log concentration, which is equivalent to converting the arithmetic mean into the geometric mean. If, as is often the case, the distribution is irregular but approaches normal, the 50% point of a log-probability graph is the appropriate measure (see Fig. 3 in the preceding article). Converting the concentration scale into double logarithms (log of log) is occasionally necessary to normalize a distribution. However, if the data show a bimodal (Fig. A1, C) or multimodal distribution (not shown), indicating

TABLE I
Results of Taste Tests: Example 1, Threshold of Substance A Added to Beer [a]

Concentrations Presented (ppb)	Number Correct Panelist					
	1	2	3	4	5	6
640	5	6	6	...	...	...
320	4	5	6	6	...	...
160	3	4	2	5	6	6
80	2	2	2	4	4	5
40	1	3	3	2	4	4
20	2	2	1	2	0	6
10	...	...	...	...	3	5
5	...	...	...	...	...	3
2.5	...	...	...	...	...	2

[a] Each panelist received six three-alternative forced-choice presentations at each of the concentrations indicated.

TABLE II
Example 1, Conversion of Results to Percent Correct Above Chance

Concentrations Presented (ppb)	Percent Correct Above Chance Panelist					
	1	2	3	4	5	6
640	75	100	100	...	...	...
320	50	75	100	100	...	...
160	25	50	0	75	100	100
80	0	0	0	50	50	75
40	−25	25	25	0	50	50
20	0	0	−25	0	−50	100
10	...	...	...	...	25	75
5	...	...	...	...	...	25
2.5	...	...	...	...	...	0

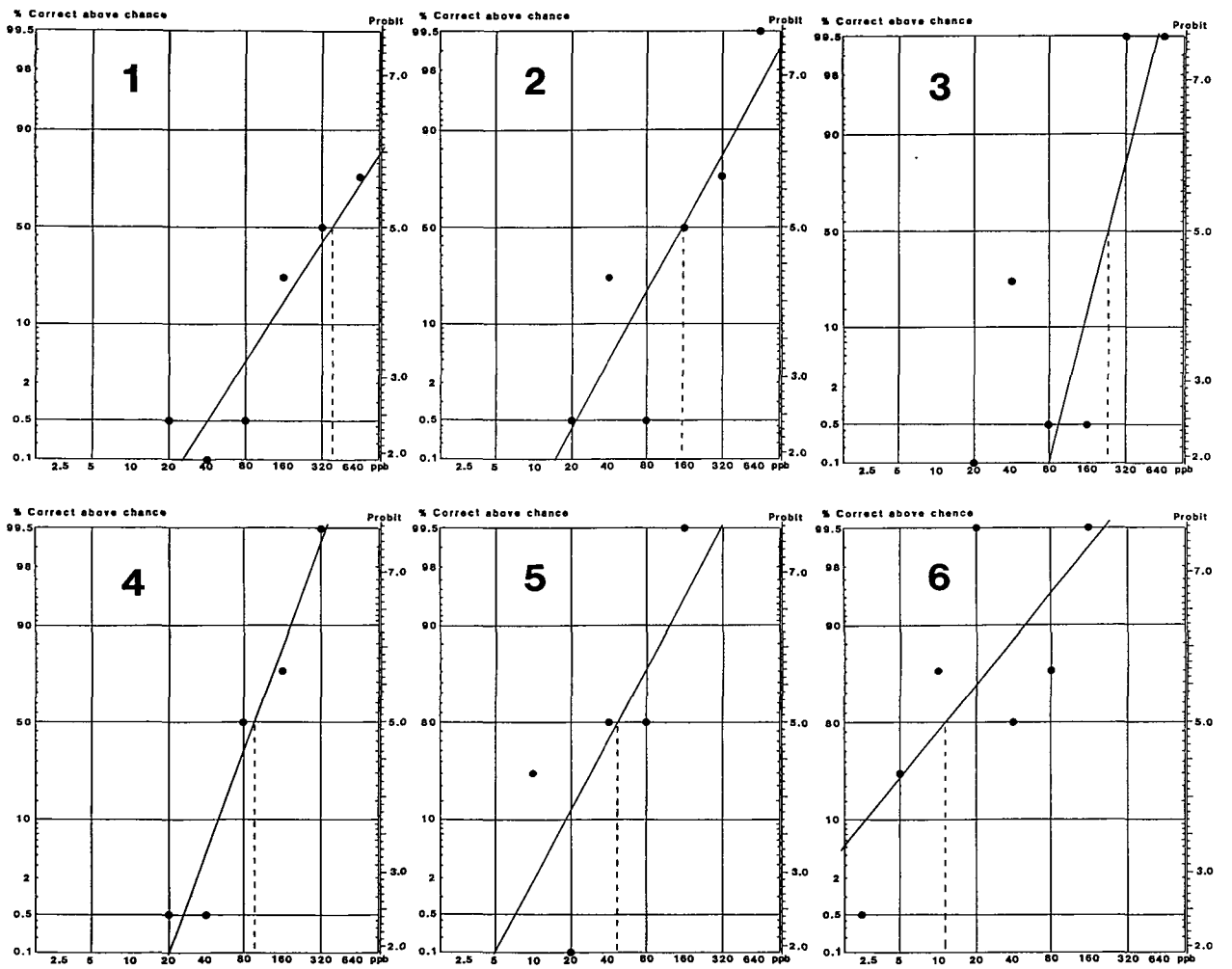


Fig. A2. Determination of threshold (using probability paper) for the six panelists in example 1. The results are: Panelist P1 = 400 ppb, P2 = 150 ppb, P3 = 220 ppb, P4 = 100 ppb, P5 = 50 ppb, P6 = 12 ppb. Concentrations in μg/L.

```
** PURPOSE:  Fit logistic models P = (1/3 + EXP[K])/(1 + EXP[K]),
**           where K = B(T - LOG[X]),
**           P is the proportion of correct identifications,
**           B is the slope,
**           X is the actual concentration (ppb) of Substance A
**             in Beer B,
**      and  T is the threshold value in log(ppb).

PROC NLIN    Method=DUD     Data=Input;     by panelist;
               PARMS B=-4  T=2
                  K   =  B*(T - LOG10(K));
                  E   =  EXP(K);
                  N   =  (1/3 + E);
                  D   =  (1 + E);
             MODEL    P =  N/D
          TITLE2  'Logistic Regression Models';
RUN;
```

OUTPUT FOR PANELIST 4:

Logistic Regression of Threshold Data Using SAS PROC NLIN

Logistic Regression Models

NON-LINEAR LEAST SQUARES ITERATIVE PHASE

DEPENDENT VARIABLE: P METHOD: DUD

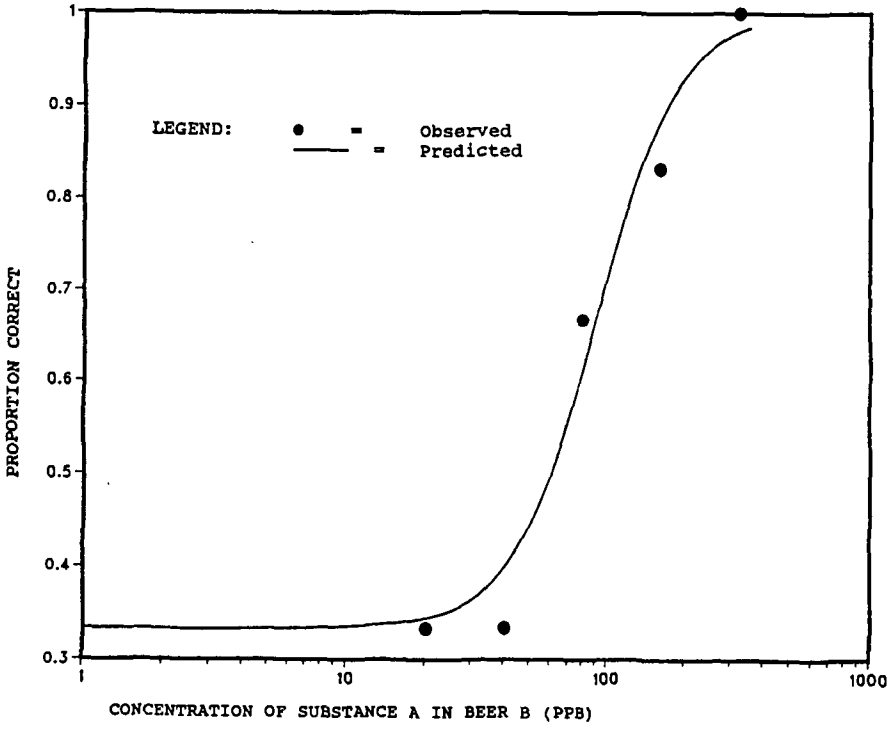
ITERATION	B	T	RESIDUAL SS
-3	-4	2.000000000	0.025885700365
-2	-4.4	2.000000000	0.02054415598
-1	-4	2.200000000	0.084958944779
0	-4.4	2.000000000	0.02054415598
1	-5.852958	1.961443385	0.010812277188
2	-6.259745	1.967823308	0.010766524899
3	-6.189164	1.951938036	0.010504941622
4	-6.283542	1.954261395	0.010481402394
5	-6.280162	1.954257276	0.010481361251
6	-6.281544	1.954068199	0.010481219887
7	-6.277816	1.953905805	0.010481193047
8	-6.280506	1.953919346	0.010481176612
9	-6.281737	1.953896400	0.010481176219
10	-6.281715	1.953899496	0.010481176193

CONVERGENCE CRITERION MET.

NON-LINEAR LEAST SQUARES SUMMARY STATISTICS DEPENDENT VARIABLE P

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE
REGRESSION	2	2.3500748238	1.1750374119
RESIDUAL	3	0.0104811762	0.0034937254
UNCORRECTED TOTAL	5	2.3605560000	
(CORRECTED TOTAL)	4	0.3558448000	

PARAMETER	ESTIMATE	ASYMPTOTIC STD. ERROR	ASYMPTOTIC 95% CONFIDENCE INTERVAL LOWER	UPPER
B	-6.281714751	1.6824126163	-11.635992903	-0.9274366000
T	1.953899496	0.0473965533	1.803059965	2.1047390264



PREDICTED VALUES

P	C	LOG[X]	X
.967	.95	2.4226	265
.933	.90	2.3037	201
.833	.75	2.1288	135
.667	.50	1.9539	90
.500	.25	1.7790	60
.400	.10	1.6041	40
.367	.05	1.4770	30

Fig. A3. Fitting sigmoids to the data in example 1 by iterative regression using SAS Proc Nlin and the logistic method. (Text explains how the data are entered and gives output for panelist 4.) Using method DUD, the results are: Panelist P1 = 330 ppb, P2 =178 ppb, P3 = 233 ppb, P4 = 90 ppb, P5 = 64 ppb, P6 = 7.8 ppb. Concentrations in μg/L.

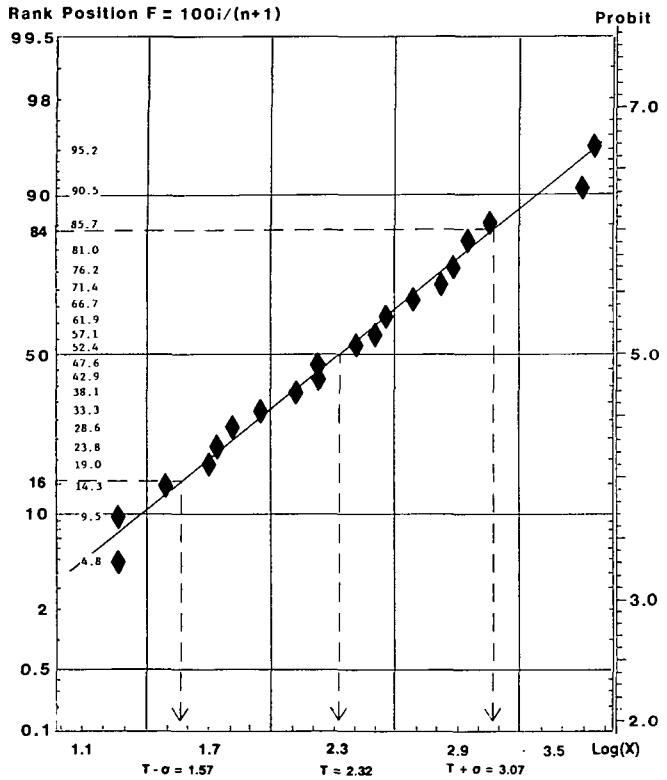


Fig. A4. Rank-probability graph for 20 panelists (example 2, adapted from French Standard NF X 43-101). Result: A straight line can be fitted through the points; consequently, the panelists are normally distributed with $T = 2.32$ and $\sigma = 0.75$ in log[x] units, where x is in $\mu g/L$.

the existence of subpopulations with different thresholds, the distribution cannot be normalized. Instead, an attempt may be made to estimate the size and group threshold of each subpopulation.

Group Standard Deviation. To characterize the dispersion of thresholds around the population mean, the group standard deviation (σ) may be calculated as shown in the examples. This is permissible only if the distribution is normal or near normal or has been normalized.

Procedure

Proceed as outlined in the examples presented.

Example 1, threshold of substance A added to beer B. Six panelists were tested at five or more concentrations chosen in advance to bracket each person's threshold.[2] Each panelist took six tests per concentration; results are shown in Table I.

Calculate the Threshold

Graphic method or linear regression. First, convert the results to percent correct above chance (Table II), using the formula of the 3-AFC procedure:[3]

$$\text{percent correct above chance} = 100(3C - N)/2N,$$

where N = number of tests presented per panelist and concentration (six tests were used here), and C = number of correct choices.

Next, plot log stimulus intensity against percent correct above chance on probability graph paper (Fig. A2), and fit a straight line through the points by eye (method 1). Plot scores of 100% as 99.5%, 0% as 0.5%, and less than 0% as 0.1%. Read the threshold as the concentration that corresponds to 50% probability. Alternatively (method 2), convert the percent scores to probits (3), or use a table of the normal deviate (method 3), and fit the line using the method of least squares.

Using a computer package. Use a computer package that employs an iterative curve-fitting procedure and weights the data by probability. The desired S-shaped curve (ogive) may be approximated using the normal probability curve (method 4) or a logistic model (method 5):

$$P = (1/3 + e^k)/(1 + e^k) \quad k = b(t - \log[x]),$$

where P = the proportion of correct responses (i.e., C/N), b = the slope, x = the concentration (e.g., ppb), and t = the threshold (in log[ppb]). Using this formula, results are entered as percent correct (not as percent above chance). Note that the threshold is at $P = 2/3$, and that all values for C can be accommodated (also $C = N$ and $C = 0$). Figure A3 shows the results obtained for panelist 4.

Calculate the group threshold. Report the threshold obtained for each panelist. If the purpose of the test and the results themselves so require, calculate a group threshold as discussed earlier in the section "Choosing the Measure of Central Tendency." In the example shown in Table III, the geometric mean may be chosen as the best central measure.

Group threshold by rank-probability graph. Use this method when the number of panelists is 10–15 or higher and the distribution is near normal, as illustrated by example 2 (Fig. A4).

Sort the panelist thresholds by rank i and plot them in a probability graph, using as ordinate the rank position

$$F_i = 100\ i/(n + 1).$$

For example, panelist 11 of a group of 20 panelists will plot at $100 \times 11/(20 + 1) = 52.4\%$. If, as here, a straight line can be drawn through the points, consider the group normally distributed with group threshold at the 50% point and group standard deviation (one-sigma) limits at the 16 and 84% (probit 4.0 and 6.0) points. Read other points of interest from the graph, such as the concentration that only 1% of the population can detect as the 1% point or that 95% can detect as the 95% point.

TABLE III
Calculating the Group Threshold[a]

Panelist	Log [ppb]	ppb
1	2.518	330
2	2.249	178
3	2.368	233
4	1.954	90
5	1.806	64
6	0.892	7.8
	11.787	

[a]Sum = 11.787, divided by 6 = 1.964. Antilog = group threshold = group geometric mean = 92 ppb. Group standard deviation for six panelists: $\sigma = 1.39$ in log [ppb] units.

Presentation of Results

Report all test conditions, such as the nature and source of the samples, method of sampling, choice of control sample (diluent), equipment, and physical test setup under which samples were presented to the panelists, concentrations used, temperature and other conditions of the samples, and instructions and report sheets given to the panelists.

Report the composition of the panel regarding age, gender,

[2]For example, each person can be given a single test (or a few tests) of the concentrations 640, 160, 40, 10, and 2.5 ppb.

[3]The formulas of other forced-choice procedures are: paired-comparison and duo-trio, percent correct above chance = $100(2C - N)/N$; triangular, percent correct above chance = $100(3C - N)/2N$; two-out-of-five, percent correct above chance = $100(10C - N)/9N$.

and experience. Additional information may be useful, e.g., familiarity with the stimulus being evaluated, health, smoking, use of dentures, and time since last meal. No panelist should be identified by name nor should the report allow a reader familiar with the panel to refer a particular judgment to a particular panel member.

Report the number of repetitions of the presentations per panelist.

Report the individual thresholds; calculate a group threshold and a group standard deviation (as in the examples) if the results permit doing so and if the purpose of the test requires it.

LITERATURE CITED

1. American Society of Brewing Chemists. *Methods of Analysis,* 7th ed. Sensory Analysis 9 Threshold of added substances—Ascending method of limits test. The Society, St. Paul, MN, 1976.
2. Amoore, J. E. and Hautala, E., Odor as an aid to chemical safety: Odor thresholds compared with threshold limit values and volatilities for 214 industrial chemicals in air and water dilution. *J. Appl. Toxicol.* 3(6):-272-290, 1983.
3. Finney, D. J. *Probit Analysis,* 3rd ed. Cambridge University Press, 1971.
4. SAS Institute. *User's Guide: Statistics,* Version 5 edition. The Institute: Cary, NC, pp. 575-606, 1985.
5. Snedecor, G. W., and Cochran, W. G., Design and analysis of sampling. In *Statistical Methods,* 6th ed., Iowa State University Press, Ames, IA, 1967, pp. 504-539.

[Received September 19, 1990. Accepted January 11, 1991.]