1 Introduction

1.1 Sensometrics

This book is about sensometrics, focusing on sensory discrimination tests and measurements in the domain of sensory analysis. Sensometrics is a subfield of, or an area related to, sensory and consumer science. According to Brockhoff (2011), "Sensometrics is the scientific area that applies mathematical and statistical methods to model data from sensory and consumer science." It is similar to psychometrics in psychology, biometrics in biology, chemometrics in chemistry, econometrics in economy, politimetrics in macropolitics, environmetrics in environmental sciences, and so on. Sensometrics has experienced rapid growth in both academia and industry within the last 2 or 3 decades. It plays an important role in modern sensory analysis and consumer research, especially in the coming Big Data era.

1.2 Sensory tests and measurements

The basic functions of sensory analysis are to provide reliable sensory measurements and to conduct valid tests. Statistical hypothesis testing is the theoretical basis of sensory tests. Statistical tests include both difference tests and similarity (equivalence) tests. The Thurstonian model (Thurstone 1927) and Signal Detection Theory (SDT) (Green and Swets 1966, Macmillan and Creelman 2005) are the theoretical basis for sensory effect measurement. Psychometric functions provide invariable indices that are independent of the methods used for measurements. Notably, the Thurstonian discriminal distance δ (or d') (ASTM 2012) and the area (R-index) under the receiver operating characteristic (ROC) curve in SDT have been widely accepted and are used in both food and sensory fields. Daniel M. Ennis (1993, 1998, 2003) and Michael O'Mahony (1979, 1992), among others, should be particularly thanked for their insight and foresight in introducing the methodologies into these fields and for tirelessly promoting their research and application over recent decades.

Sensory measurement takes on a broad range of meanings and contents. Besides sensory effect measurement using Thurstonian discriminal distance and area under ROC curve, the following measurements can also be regarded as different types of sensory measurement: sensory threshold measurement, sensory risk measurement, time intensity measurement, sensory shelf life measurement, trained sensory panel/panelist performance measurement, consumer emotions and psychographics measurement, and attribute relative importance

Sensory Discrimination Tests and Measurements: Sensometrics in Sensory Evaluation, Second Edition. Jian Bi.

Companion Website: www.wiley.com/go/bi/SensoryDiscrimination

measurement. Specific statistical methodologies are used for different types of sensory measurement.

1.3 A brief review of sensory analysis methodologies

Sensory analysis can be divided into two types: laboratory sensory analysis and consumer sensory analysis. In laboratory sensory analysis, a trained panel is used as an analytical instrument to measure the sensory properties of products. In consumer sensory analysis, a sample of a specified consumer population is used to test and predict consumer responses to products. These have different goals and functions, but share some methodologies.

Discriminative analysis and descriptive analysis are the main classes of methodology for both laboratory and consumer sensory analyses. Discriminative analysis includes discrimination tests and measurements. In this book, discrimination tests are used to determine whether a difference exists between treatments for confusable sensory properties of products (difference test), or whether a difference is smaller than a specified limit (similarity/equivalence test), usually using a two-point scale or a rating or ranking scale. Discrimination measurements are used to measure, on a particular index, the extent of the difference/similarity. There are two sources of sensory differences: intensity and preference. A discrimination test is used when testing difference/similarity of intensity; a preference test is used when testing difference/similarity of preference. Descriptive analysis is used to determine, on a rating scale, how much of a specific characteristic difference exists among products (quantitative descriptive analysis) or to characterize a product's sensory attributes (qualitative descriptive analysis). Quantitative descriptive analysis for preference is also called "acceptance testing."

Acceptance or preference testing is of very limited value for a laboratory panel (Amerine *et al.* 1965) but is valuable in a consumer analysis setting. Laboratory discrimination testing, using a trained panel under controlled conditions, is referred to as "Sensory Evaluation I," while consumer discrimination testing, using a sample of untrained consumers under ordinary consumption (eating) conditions, is referred to as "Sensory Evaluation II" (O'Mahony 1995). Confusion of the two will lead to misleading conclusions. Controversy over whether the consumer can be used for discrimination testing ignores the fact that laboratory and consumer discrimination tests have different goals and functions.

The distinction between discriminative analysis and quantitative descriptive analysis is not absolute from the viewpoint of modern sensory analysis. The Thurstonian model and SDT (see Chapters 2–3) can be used for both discriminative analysis and quantitative descriptive analysis. The Thurstonian δ (or d'), a measure of sensory difference/similarity, can be obtained from any kind of scale used in discriminative and descriptive analysis. A rating scale, typically used in descriptive analysis, is also used in some modified discrimination tests.

The following types of analysis are the important topics and methodologies of sensory analysis: sensory threshold analysis, sensory risk analysis, time intensity analysis, sensory shelf life analysis, trained sensory panel/panelist performance analysis, consumer emotions and psychographics analysis, and sensory attribute relative importance analysis.

This book is primarily concerned with methodology, mainly from a statistical point of view, of sensory discrimination tests and measurements, including laboratory and consumer sensory analyses.

1.4 Method, test, and measurement

In this book, a distinction is made among three terms: "sensory discrimination method," "sensory discrimination test," and "sensory discrimination measurement."

In sensory discriminative analysis, certain procedures are used for experiments. These procedures are called discrimination methods (e.g., the Duo–Trio method, the Triangular method, the ratings method).

When discrimination procedures are used for statistical hypothesis testing, or when statistical testing is conducted for the data from a discrimination procedure, the procedure is called discrimination testing (e.g., the Duo–Trio test, the Triangular test, the ratings test). In this book, discrimination testing is referred to as both difference testing and similarity/equivalence testing for both preference and intensity (Chapters 4–5). Bayesian statistical tests are also discussed, in Chapter 6. In Chapter 7, some modified discrimination tests are discussed. Multiple-sample discrimination tests are discussed in Chapter 8. Replicated discrimination tests are discussed in Chapters 9–11.

When discrimination procedures are used to measure, or, in other words, when an index (e.g., Thurstonian δ (or d') or R-index) is produced using the data from a discrimination procedure, the procedure is called a discrimination measurement (e.g., Duo–Trio measurement, Triangular measurement, ratings of the A–Not A measurement). Effect measurement includes distance measure d' and area measure R-index (or Gini-index). Besides the effect measurement discussed in Chapters 2–3, other sensory measurements are discussed in Chapters 12–18. Both sensory testing and measurement are of importance and are useful. However, generally speaking, sensory measurement is more important and more useful in practice. Sensory measurements provide indices of the magnitude of sensory effects.

1.5 Commonly used discrimination methods

1.5.1 *Forced-choice methods*

(a) The Two-Alternative Forced Choice (2-AFC) method (Green and Swets 1966): This method is also called the paired comparison method (Dawson and Harris 1951, Peryam 1958). With this method, the panelist receives a pair of coded samples, A and B, for comparison on the basis of some specified sensory characteristic. The possible pairs are AB and BA. The panelist is asked to select the sample with the strongest (or weakest) sensory characteristic. The panelist is required to select one even if he or she cannot detect the difference.

(b) The Three-Alternative Forced Choice (3-AFC) method (Green and Swets 1966): Three samples of two products, A and B, are presented to each panelist. Two of them are the same. The possible sets of samples are AAB, ABA, BAA or ABB,

BAB, BBA. The panelist is asked to select the sample with the strongest or the weakest characteristic. The panelist has to select a sample even if he or she cannot identify the one with the strongest or the weakest sensory characteristic.

(c) The Four-Alternative Forced Choice (4-AFC) method (Swets 1959): Four samples of two products, A and B, are presented to each panelist. Three of them are the same. The possible sets of samples are AAAB, AABA, ABAA, BAAA or BBBA, BBAB, BABB, ABBB. The panelist is asked to select the sample with the strongest or the weakest characteristic. The panelist is required to select a sample even if he or she cannot identify the one with the strongest or weakest sensory characteristic.

(d) The Triangular (Triangle) method (Dawson and Harris 1951, Peryam 1958): Three samples of two products, A and B, are presented to each panelist. Two of them are the same. The possible sets of samples are AAB, ABA, BAA, ABB, BAB, and BBA. The panelist is asked to select the odd sample. The panelist is required to select one sample even if he or she cannot identify the odd one.

(e) The Duo–Trio method (Dawson and Harris 1951, Peryam 1958): Three samples of two products, A and B, are presented to each panelist. Two of them are the same. The possible sets of samples are A: AB, A: BA, B: AB, and B: BA. The first one is labeled as the "control." The panelist is asked which of the two test samples is the same as the control sample. The panelist is required to select one sample to match the "control" sample even if he or she cannot identify which is the same as the control.

(f) The Unspecified Tetrad method (Lockhart 1951): Four stimuli, two of A and two of B, are used, where A and B are confusable and vary in the relative strengths of their sensory attributes. Panelists are told that there are two pairs of putatively identical stimuli and to sort them into their pairs.

(g) The Specified Tetrad method (Wood 1949): Four stimuli, two of A and two of B, are used, where A and B are confusable and vary in the relative strengths of their sensory attributes. Panelists are told that there are two pairs of putatively identical stimuli and to indicate the two stimuli of specified A or B.

(h) The Dual Pair (4IAX) method (Macmillan *et al.* 1977): Two pairs of samples are presented simultaneously to the panelist. One pair is composed of samples of the same stimuli, AA or BB, while the other is composed of samples of different stimuli, AB or BA. The panelist is told to select the most different pair of the two pairs.

(i) The "M + N" method (Lockhart 1951): M + N samples with M sample A and N sample B are presented. The panelist is told to divide the samples into two groups, of A and B. There are two versions of the method: specified and unspecified. This is a generalization of many forced-choice discrimination methods, including the Multiple-Alternative Forced Choice (m-AFC), Triangle, and Specified and Unspecified Tetrad. The "M + N" with larger M and N can be regarded as a specific discrimination method with a new model. Unlike the conventional difference tests using the "M + N" with small M and N based on a binomial model, the "M + N" with larger M and N ($M = N > 3$) can reach a statistical significance in a single trial for only one "M + N" sample set based on a hypergeometric model. The methods that use a new model are particularly useful for assessing the discriminability of sensory panels and panelists; these are discussed in Chapter 16.5.

1.5.2 *Methods with response bias*

(a) The A–Not A method (Peryam 1958): Panelists are familiarized with samples A and Not A. One sample, which is either A or Not A, is presented to each panelist. The panelist is asked if the sample is A or Not A.
(b) The A–Not A with Remind (A–Not AR) method (Macmillan and Creelman 2005): Unlike the A–Not A, which is a single-sample presentation, a reminder (e.g., sample A) is provided before each test sample (sample A or Not A) in order to jog the panelist's memory.
(c) The Same–Different method (see, e.g., Pfaffmann 1954, Amerine *et al.* 1965, Macmillan *et al.* 1977, Meilgaard *et al.* 1991, among others, for the same method under different names): A pair of samples, A and B, is presented to each panelist. The four possible sample pairs are AA, BB, AB, and BA. The panelist is asked if the two samples that he or she received are the same or different.

The ratings methods discussed in the book include ratings of the A–Not A, A–Not AR, and Same–Different methods.

1.6 Classification of sensory discrimination methods

Sensory discrimination methods are typically classified according to the number of samples presented for evaluation, as single-sample (stimulus), two-sample, three-sample, or multiple-sample methods. This classification is natural, but it does not reflect the inherent characteristics of the methods. In this book, the discrimination methods are classified according to the decision rules and cognitive strategies they involve. This kind of classification may be more reasonable and profound. In the following chapters, we will see how methods in the same class correspond to the same types of statistical model and decision rules.

1.6.1 *Methods requiring and not requiring the nature of difference*

There are two different types of instruction in the discrimination method. One type involves asking the panelists to indicate the nature of difference in the products under evaluation; for example, "Which sample is sweeter?" (the 2-AFC and the 3-AFC methods); or "Is the sample A or Not A?" (the A–Not A method). The other type compares the distance of difference; for example, "Which of the two test samples is the same as the control sample?" (the Duo–Trio method); "Which of these three samples is the odd one out?" (the Triangular method); or "Are these two samples the 'same' or 'different'?" (the Same–Different method). The two types involve different cognitive strategies and result in different percentages of correct responses. Hence, the discrimination methods can be divided into these two types: methods using the "skimming" strategy and methods using the "comparison of distance" strategy (O'Mahony *et al.* 1994). The two types of methods can also called specified or unspecified method.

1.6.2 *Methods with and without response bias*

Response bias is a basic problem with sensory discrimination methods. Many authors hav eaddressed this problem (e.g. Torgerson 1958, Green and Swets 1966, Macmillan

and Creelman 2005, O'Mahony 1989, 1992, 1995). Sensory discrimination methods are designed for the detection and measurement of confusable sensory differences. There is no response bias if the difference is large enough, but response bias may occur when the difference between two products is so small that a panelist makes an uncertain judgment. In this situation, how large a difference can be judged as a difference may play a role in the decision process. Criterion variation (strictness or laxness of a criterion) causes response bias. A response bias is a psychological tendency to favor one side of a criterion. Response bias is independent of sensitivity. This is why the methods with response bias (e.g., the A–Not A and the Same–Different methods) can also be used for difference testing. However, response bias affects test effectiveness (power).

Forced-choice procedures can be used to stabilize decision criteria. Hence, most sensory discrimination methods are designed as forced-choice procedures. A forced-choice procedure must have at least three characteristics: (1) Two sides of a criterion must be presented. The two sides may be "strong" and "weak," if the criterion is about the nature of the difference between products. The two sides may be "same" and "different," if the criterion is about the distance of the difference between products. Because a single sample or two samples of the same type cannot contain two sides of a criterion, evaluating a single sample or the same type of sample is not a forced-choice procedure. Because a single pair of samples or a pair of samples of the same type cannot contain two sides of a criterion concerning the distance of a difference, evaluating a single sample pair or a pair of samples of the same type is not a forced-choice procedure, either. (2) A panelist should be instructed that the samples presented for evaluation contain the two sides of a criterion. (3) A response must be given in terms of one clearly defined category. The "don't know" response is not allowed.

1.6.3 *Methods using multiple sets and only one set of samples*

In conventional discrimination tests using forced-choice methods, such as the "M + N" method with small M and N, we cannot get a statistical conclusion from a response for only one set of samples, because even for the perfect response for a set of the samples, the chance probability (e.g., 1/3 in the 3-AFC) is still larger than any acceptable significance level. Hence, multiple sets of "M + N" samples are needed. A binomial model is used for analysis of the proportion of correct responses. However, we can get a conclusion based on responses in a 2×2 table for only one set of "M + N" samples with larger M and N in a Fisher's exact test.

1.6.4 *Methods with binary and ratings data*

The responses in forced-choice methods are binary. The responses in the methods with response bias may be binary or ratings. The ratings of the methods represent degrees of sureness of a judgment or different decision criteria. For example, the responses in an A–Not A test are "A"/"Not A" (i.e., 1 or 2). The responses in a ratings of the A–Not A test may be a six-point scale with (1, 2, 3, 4, 5, 6) corresponding to (A, A?, A??, N??, N?, N).

Table 1.1 describes the classifications of sensory discrimination methods.

Table 1.1 Classifications of sensory discrimination methods

		Requiring the nature of difference	Comparing distance of difference
Forced-choice methods	Based on multiple sets of samples	2-AFC 3-AFC 4-AFC Specified Tetrad	Duo–Trio Triangular Unspecified Tetrad Dual-Pair (4IAX)
	Based on one set of samples	Specified "M + N" with larger M and N	Unspecified "M + N" with larger M and N
Methods with response bias	Binary response	A–Not A	Same–Different A–Not AR
	Ratings response	Ratings of A–Not A	Ratings of Same–Different Ratings of A–Not AR