Food Quality and Preference 41 (2015) 159–162

Contents lists available at ScienceDirect

Food Quality and Preference

journal homepage: www.elsevier.com/locate/foodqual

A Thurstonian model for the degree of difference protocol

Daniel M. Ennis *, Benoît Rousseau

The Institute for Perception, Richmond, VA, USA

ARTICLE INFO

Article history:
Received 7 July 2013
Received in revised form 16 November 2014
Accepted 17 November 2014
Available online 2 December 2014

Keywords:
Degree of difference
Thurstonian model
Delta
Tau criterion

ABSTRACT

In the degree of difference methodology (DOD), subjects are presented with pairs of samples, either identical or different, and must indicate how different the samples are using a t-point category rating scale. In this article, the Thurstonian model for the DOD assuming independent assessments is derived. The model permits the estimation of the size of the underlying difference between the products (δ), the variance of this estimate, as well as the sizes of the $t-1$ τ criteria (scale boundaries on the difference perceptual distribution). This model expands further the collection of Thurstonian models already available for many discrimination, rating and ranking methodologies.

Introduction

The degree of difference method (DOD, sometimes called "difference from control test") involves the comparison of two stimuli that can be either putatively identical or different and has been used for over three decades. If two products X and Y are compared, identical stimuli (XX or YY) or different stimuli (XY) are evaluated. The task of the subject is to rate their similarity on a t-category scale, where the lowest score usually corresponds to the samples being identical and the highest score corresponds to the samples being very different. While this protocol has been used for decades, most notably as a data collection device for multidimensional scaling (Shepard, 1962a, 1962b), it has also been used in difference testing (Aust, Gacula, Beard, & Washam 1985; Bi, 2002; PME Method 720, 1984). The theory underlying the Thurstonian model discussed in this paper was derived in 1995 by one of the authors (D.M. Ennis) and incorporated into the IFPrograms collection of software distributed by the Institute for Perception, Richmond, VA, USA (Ennis, 2003). It has also recently become available in the sensR package (Christensen & Brockhoff, 2011) which includes Thurstonian models for sensory discrimination in R. A main application of the DOD is for the investigation of product differences in the presence of lot-to-lot product variability (Pecore et al., 2006; Young et al., 2008). The method has also been used for same–different judgments with sureness ratings (Rousseau, Meyer, & O'Mahony, 1998).

One particularity of the method is the need to evaluate only two samples which can be an advantage in situations involving samples with high fatigue or carryover effects (e.g., red chili peppers) or samples that require a recovery period between evaluations (e.g., chewing gum). The two sample evaluation has also been shown to provide a benefit compared to the three-sample triangle method due to its lower memory requirements (Lau, O'Mahony & Rousseau, 2004; Rousseau et al., 1998; Rousseau & O'Mahony, 2000, 2001). However, in terms of power, the DOD was found not to be as powerful as the tetrad method, assuming that the perceptual noise is similar in both protocols (Ennis & Christensen, 2014).

The same–different method is a special case of the degree of difference method assuming a 2-point scale. Tables have been developed for the method (Kaplan, Macmillan, & Creelman, 1978; Macmillan & Creelman, 2005; O'Mahony and Rousseau, 2003). The model involves the consideration of a τ criterion (Ennis & Ashby, 1993; Ennis, Palen, & Mullen, 1988), the perceptual size of a difference above which two samples will be called "different" (otherwise "same"). The degree of difference model expands the same–different model to more than two categories, with the use of $t-1$ τ criteria for a t-point scale. The objective of this manuscript is to describe a Thurstonian model for the DOD that has been in use for over two decades and that permits the estimation of the size of product differences (delta) as well as a measure of the size of the subjects' response bias (tau) for independent data.

Illustrative example

In 1984, Philip Morris International adopted the degree of difference method and prepared an internal procedure manual to describe its use (PME Method 720, 1984). This document provided

* Corresponding author at: The Institute for Perception, 7629 Hull Street Road, Richmond, VA 23235, USA. Tel.: +1 (804) 675 2980.

E-mail address: Daniel.M.Ennis@ifpress.com (D.M. Ennis).

http://dx.doi.org/10.1016/j.foodqual.2014.11.011
0950-3293/

the data collection procedure and statistical analysis approach. In the method, a same and different pair were evaluated by each subject on a 7-point degree of difference scale where 1 = "no difference" and 7 = "very big difference." The resulting data were then analyzed using a one-tailed paired t-test. Note that at the time the power of the DOD had not yet been investigated and that the sample size used here would only allow the detection of fairly large sensory differences (as described later on in this manuscript). The work of Ennis and Christensen (2014) provide insights on the power of the DOD methodology.

A problem with this approach is the implausibility of its assumptions. While one could reasonably assume that perceived intensities follow a normal distribution, distances between them, under that assumption, cannot, as required by the method. Nevertheless, this was the assumption adopted when the paired t-test was used. A better alternative is to consider that perceived intensities are normally distributed and that the perceived distances are compared to criteria that define the scale boundaries for the 7-point scale. This could be considered a Thurstonian approach to modeling the data based on Thurstone's categorical judgment model (Thurstone, 1927) but applied to distances instead of intensities and not assuming that distances follow a normal distribution. This model provides a way to link the degree of difference method to other difference testing methods based on the same theoretical framework. The parameter estimate of the degree of difference between products, d', can be related to the same parameter estimate obtained by other methods and thus provides an opportunity to study the convergent validity of d' as a measure of sensory difference.

Theory

Let X_1 and X_2 be identically normally distributed random variables with mean zero and unit standard deviation. Let Y_1 and Y_2 be identically normally distributed random variables with mean δ and unit standard deviation (see Fig. 1(a)). Then $X_1 - X_2$ and $Y_1 - Y_2$ are distributed as $N(0, 2)$ and $Y_1 - X_1$ (or $Y_2 - X_2$) is distributed as $N(\delta, 2)$ where $N(0, 2)$ and $N(\delta, 2)$ refer to normal distributions with means of 0 and δ, respectively, and variance 2 (see Fig. 1(b)). Suppose that there are two items to be compared perceptually and that the degree of difference between the items is scaled on a t-category degree of difference scale. It should be noted that the "degree of difference" refers mathematically to the absolute value of the difference between momentary percepts and not the difference itself which may assume negative and positive values. The subject actually responds to the absolute values and refers them to category boundaries on a 7-point distance scale with a zero origin. Let $\tau_0, \tau_1, \tau_2, \ldots, \tau_t$ represent decision boundaries on the continuum of distances between the items. Let $\tau_0 = 0$ and $\tau_t = \infty$. Let "i" be the score given on the t-point scale to a given momentary perceptual distance.

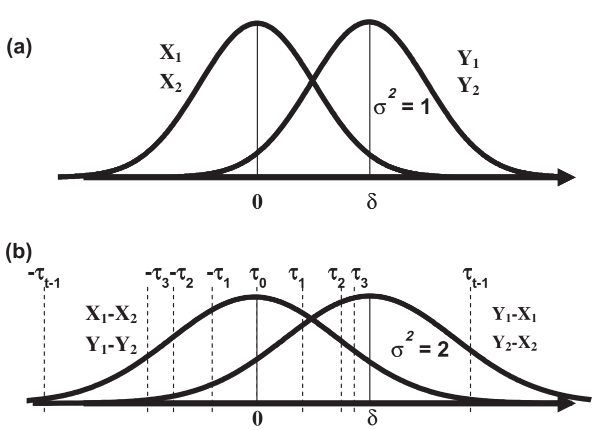

Fig. 1. Stimulus perceptual intensity (a) and difference (b) distributions.

When provided with samples from different items, we first determine the probability that a subject will choose to rate the difference "i", P_i. In order to derive these probabilities, one could consider the distribution of the absolute value of differences in normal random variables as described above and this would involve using folded normal random variables. Squaring the absolute values leads to models expressed in terms of central and non-central chi-squares. Instead of working with chi-square random variables, an exactly identical solution is to consider the distribution of normally distributed differences but refer them to negative and positive boundary values. This approach provides easier computation, although one would get the same result using the chi-square distribution functions. We use the normally distributed difference method for mathematical convenience but its use is not meant to imply that subjects actually construct negative and positive boundaries. Fig. 2 focuses on the difference distribution generated from two different samples and illustrates the method used to estimate P_i. In Fig. 2(a), P_i corresponds to the sum of the shaded areas under the difference distribution for two different samples. That sum can be calculated based on two areas:

$$P_i = \Pr(\tau_{i-1} < Y - X < \tau_i) + \Pr(-\tau_i < Y - X < -\tau_{i-1})$$

Since the difference distribution is centered at δ and has a variance of 2, P_i can be calculated using areas from the standard normal distribution function (centered at 0 with variance of 1), as shown in Fig. 2(b). Accordingly, P_i can then be expressed as:

$$P_i = \Phi\left(\frac{\tau_i - \delta}{\sqrt{2}}\right) - \Phi\left(\frac{\tau_{i-1} - \delta}{\sqrt{2}}\right) + \Phi\left(\frac{-\tau_{i-1} - \delta}{\sqrt{2}}\right) - \Phi\left(\frac{\tau_i - \delta}{\sqrt{2}}\right) \quad (1)$$

where $\Phi(a)$ is the cumulative distribution function of the standard normal from $-\infty$ to a.

When putatively identical samples are presented, a similar approach can be used to estimate Q_i, the probability that a subject will choose to rate the difference "i":

$$Q_i = 2\left[\Phi\left(\frac{\tau_i}{\sqrt{2}}\right) - \Phi\left(\frac{\tau_{i-1}}{\sqrt{2}}\right)\right] \quad (2)$$

Suppose that an experiment has been conducted and that n_i and m_i are the counts for a rating difference of "i" for the putatively different and same pairs, respectively. In order to estimate τ_i and δ the following likelihood equation should be maximized

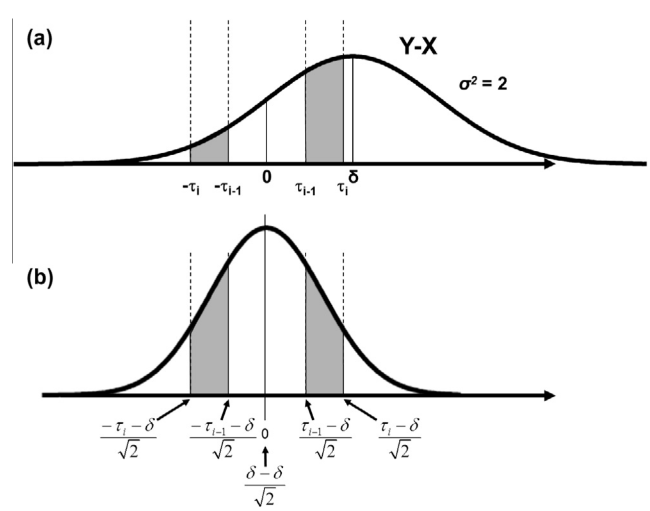

Fig. 2. Different pair difference distribution with area for "i" response and its standard transform.