Available online at www.sciencedirect.com
SCIENCE DIRECT®
Food Quality and Preference 17 (2006) 179–187
Food Quality and Preference
www.elsevier.com/locate/foodqual

Visualizing individual differences in pairwise comparison data ☆

Ulf Böckenholt *

Faculty of Management, McGill University, 1001 Sherbrooke Street West, Montreal, QC, Canada H3A 1G5

Received 25 September 2004; received in revised form 23 June 2005; accepted 5 July 2005
Available online 19 August 2005

Abstract

Methods of comparative judgments such as paired comparisons and rankings share one common problem: They do not allow recovering the origin of the stimulus evaluations. One stimulus may be judged more positively than another but this result does not allow any conclusions about whether either of the stimuli are attractive or unattractive. This article discusses the implications of this limitation for the interpretation of individual differences in multiple comparative judgment data. It is shown that because of the comparative nature of the judgments, distances instead of covariances between stimuli should be interpreted and a graphical method is presented that facilitates understanding the underlying similarity relationships among the stimuli. One consumer-test application illustrates the benefits of the proposed graphical approach for understanding individual differences in preference judgments even when the scale origin cannot be identified.
© 2005 Elsevier Ltd. All rights reserved.

Keywords: Choice data; Individual differences; Multi-level regression models; Paired comparisons

1. Introduction

In the method of paired comparisons, judges are presented with pairs of stimuli and, for each pair, they are asked to choose the preferred one. Because this task imposes minimal constraints on the response behavior of a judge, paired comparison judgments are collected in a wide range of applications ranging from sensory testing to investigations of preference and choice behavior (David, 1988). Especially, when differences between stimuli are small, it is desirable to compare them in isolation and to free the judgment process as much as possible from context effects caused by the presence of other stimuli.

Since in many paired comparison studies, respondents are sampled randomly from a population of judges, it is appropriate to model individual differences in the stimulus evaluations with a multi-level framework (Bock & Jones, 1968; Gabrielsen, 2001; Takane, 1987). Taking into account both within- and between-judge effects in the evaluation of stimuli, this approach has three major advantages over the single-judgment paired comparison methods proposed by Thurstone (1927), Bradley and Terry (1952), and Luce (1959). First, a multi-level approach facilitates detailed tests of how stimulus parameters vary from person to person and whether this variation can be explained by concomitant variables. Second, multi-level paired comparison models are consistent with both a monotonic and non-monotonic response function. Thus, a researcher does not need to specify whether individual differences are described more adequately by a quadratic (ideal-point) or by by a linear (vector) model (Roberts & Laughlin, 1996; Van Schuur & Kiers, 1994). Mis-specifications of the response functions and the resulting difficulties in the interpretation of the results (for a recent example, see Brazill & Grofman, 2002) are therefore less likely to

☆ This research was supported partially by the Social Sciences and Humanities Research Council of Canada.
* Fax: +1 514 370 3909.
E-mail address: ulf.bockenholt@mcgill.ca

0950-3293/$ - see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.foodqual.2005.07.007

occur. Third, multi-level paired comparison model facilitate systematic investigations of the extent to which inconsistent responses are reflective of the stochastic nature of the judgmental process or a result of systematic deviations from the modeling assumptions. Such analyses are useful in understanding whether judges discriminate and assess the stimuli on one- or multi-dimensional continua (Tversky, 1969).

Although the method of paired comparison has a number of distinct advantages that secured its place in the toolbox of experimental researchers (Duineveld, Arents, & King, 2000), it has one limitation that requires special care in applications of the method: Based on pairwise judgments it is not possible to recover the origin of the stimulus evaluations. One stimulus may be judged more positively than another but this result does not allow any conclusions about whether either of the stimuli are attractive or unattractive. Thus, although one judge may find both stimuli attractive and another judge may dislike both stimuli, their comparative judgments can be identical. The objective of this article is to show that these difficulties in analyzing individual differences can be overcome by representing comparative judgments as distances in a multi-dimensional space. As a result, individual differences can be displayed graphically using a classical multi-dimensional scaling approach (Torgerson, 1952). The proposed representation is easy to interpret and helpful in communicating how judges vary in their assessment of the stimuli.

The remainder of the paper is structured as follows. First, a brief review of multi-level models for the analysis of paired comparison data is presented. Next, it is shown how to obtain graphical displays of individual differences in stimulus evaluations. Although this presentation focuses on paired comparison data, it should be noted that the results apply directly to other types of comparative judgment data such as rankings. Paired comparisons are considered because they can be analyzed easily with currently available multi-level packages for logistic and probit models. The paper concludes with a discussion of the main results.

2. Modelling multi-level paired comparison judgments

In a complete paired comparison experiment a judge is presented with $\frac{J(J-1)}{2}$ pairs of J stimuli and asked to express a preference for one of them. A number of paired comparison models have been proposed in the literature to describe the underlying response process of a judge (David, 1988). Perhaps because of their computational convenience and easy interpretation, Luce's (1959) and Thurstone's (1927) models have been used most frequently over the years.

Typically, paired comparison data that are collected by asking judges to compare multiple stimulus pairs. In this case, it should be taken into account that the data contain variation between individuals as well as momentary fluctuations within each person. A two-level representation provides a flexible framework to analyze both within- and between-judge effects. The first level describes the stochastic variation in the responses of a single judge and the second level represents the judges' variability in the assessment of the stimuli (Böckenholt, 2001). The following section provides a brief review of both levels and discusses the implications of the scale origin problem on the interpretation of the between-judge model parameters.

According to Thurstone's (1927) random utility representation of the paired comparison judgment process, judges arrive at their response by first determining the underlying values of the two stimuli under consideration and then selecting the stimulus with the higher utility value. Because respondents may not always select the same stimulus in repeated comparisons, a random error term is added to the comparison process. Specifically, letting μ_{ij} and μ_{ik} denote the mean evaluations of stimuli j and k by person i, the latent judgment outcome y_{ijk} between two stimuli can be written as a difference between the respective mean evaluations:

$$y_{ijk} = \mu_{ij} - \mu_{ik} + \epsilon_{ijk}. \tag{1}$$

Fluctuations in the evaluations of the two stimuli j and k are captured by ϵ_{ijk} which is assumed to be independently distributed for all stimulus pairs. When ϵ_{ijk} follows a normal distribution with mean 0 and variance σ^2, the so-called "Case 5" Thurstone model is obtained. Specifying a logistic distribution yields the Bradley–Terry–Luce model. However, in applications there is little to choose between these two specifications. Very large sample sizes are required to distinguish a normal from a logistic distribution function which share the same mean and variance (Camilli, 1994). Differences between the Bradley–Terry–Luce and the Thurstonian model become apparent when more general cases of the Thurstonian model are considered with within-pair variances that are not equal for all or some of the pairs, i.e., $\sigma^2_{jk} \neq \sigma^2_{lm}$. Heiser and DeLeuuw (1981) discuss these cases and present a rich set of geometric representations for such models.

Since the response of a judge is binary, the latent difference judgment y_{ijk} needs to be mapped onto the discrete response scale. When $y_{ijk} > 0$, stimulus j is preferred by person i, and, otherwise, stimulus k is selected. For the Bradley–Terry–Luce model, the selection of stimulus j in a comparison between stimuli j and k can then be expressed as

$$\Pr(y_{ijk} > 0) = \Psi(\mu_{ij} - \mu_{ik}), \tag{2}$$

where $\Psi(\bullet)$ is a short form for the logistic distribution function $\frac{1}{1+\exp[-(\bullet)]}$. Similarly, for the Thurstonian model, we obtain

Download English Version:

https://daneshyari.com/en/article/4318281

Download Persian Version:

https://daneshyari.com/article/4318281

Daneshyari.com