CompR : package for uni or multivariate preference data analysis using paired comparison model

Michel Séménou[1]

[1] *Unité de Sensométrie et Chimiométrie*
Oniris, site de la Géraudière
CS82225
44322 NANTES Cedex 03 France
E-mail : *michel.semenou@oniris-nantes.fr*

Abstract
Paired comparisons tests are used in marketing studies or on sensory evaluation. The subjects are successively submitted to pairs of stimuli and have to choose an element of the pair presented according to a given criterion. Multivariate paired comparisons allow to compare pairs of stimuli for various criteria (different context, several criteria or varied sensory attributes). The package CompR supplies tools allowing to deal with such data. It is also possible to take into account a possible heterogeneity of the panel by means of a segmentation of the subjects. An adaptation of the latent class approach is implemented, on the basis of Bradley-Terry-Luce model.

Keywords: Paired comparison; segmentation; Bradley model

Résumé en Français
Les épreuves de comparaisons par paires sont employées en études marketing ou en évaluation sensorielle. Les sujets sont soumis successivement à des paires de stimuli et doivent choisir un élément de la paire présentée selon le critère que l'on souhaite étudier. Les épreuves multivariées permettent de comparer des paires de stimuli pour différents critères (contextes différents, plusieurs critères de préférence ou attributs sensoriels variés). Le package CompR fournit un certain nombre d'outils permettant de traiter des données issues de comparaisons par paires uni ou multivariées et de prendre en compte une hétérogénéité éventuelle du panel à l'aide d'une segmentation des sujets. Une adaptation de l'approche segmentation en classes latentes, basée sur le modèle de Bradley-Terry-Luce, est développée, permettant d'étudier les différences perçues entre les produits par les panélistes et de caractériser les relations entre les attributs pour les différentes classes.

Mots-clés : Comparaison par paires, segmentation, modèle de Bradley-Terry-Luce

1. Introduction

Paired comparison tests are used on marketing studies or on sensory evaluation. These tests are particularly suitable when we want to evaluate products in a subjective way. The subjects are successively submitted to pairs of stimuli and have to choose an element of the pair presented according

to the studied criterion. Multivariate paired comparison experiments allow to compare pairs of stimuli for various criteria (different contexts, several criteria rather or varied sensory attributes).
Paired comparison experiments supply an interesting alternative to tests based on hedonic scale for which the consumers meet a number of difficulties with the scale or to rankings experiments which are limited in term of number of possible products. From a practical point of view, paired comparison tests are easy to implement and give a good quality of discrimination of products. Some difficulties could appear for this kind of experiment when the number of item increases. One can overcome this problem by using suitable balanced incomplete design (Bradley 1976, David 1988).

2. Method

2.1 Bradley-Terry-Luce Model

Suppose that *n* items are evaluated by a set of *H* subjects using paired comparison experiments. Each respondent is asked to choose for each presented pair between two stimuli (first or second item). The response of the individual h for the pair (*i*, *j*) of items is noted:

$$y_{ij,h} = \begin{cases} 0 \text{ if item } j \text{ is chosen by the individual } h \\ 1 \text{ if item } i \text{ is chosen by the individual } h \end{cases}$$

$y_{ij,h}$ is an observation of a Bernoulli variable $\boldsymbol{y}_{ij,h}$. Let $\pi_{ij,h}$ be the probability of item *i* to be chosen by the individual *h* over the item *j*. Using the Bradley-Terry-Luce model (Bradley and Terry, 1952), this probability can be written in the following form:

$$\pi_{ij,h} = \pi_{ij} = \frac{\pi_i}{\pi_i + \pi_j}$$

under the constraint:

$$\pi_i \in]0;1[\; et \sum_{i=1}^{n} \pi_i = 1$$

In others words, the Bradley-Terry-Luce model formulates the choice proportions $\pi_{ij,h}$ in terms of scale values π_i (Bradley's scores) for the *n* items.

Bradley's model allows to express the odds ratio to choose the product *i* when it is compared to product *j* by:

$$\frac{\pi_{ij}}{\pi_{ji}} = \frac{\pi_{ij}}{1 - \pi_{ij}} = \frac{\pi_i}{\pi_j}$$

Maximum likelihood estimate of $\boldsymbol{\pi}$, vector of the Bradley's scores, can be obtained by solving iteratively the likelihood equation (Dykstra 1956).

2.2 Segmentation of the panel

The previous approach supposes homogeneity of the panel. It means that we consider that all the panelists perceive products in a similar way for all the criteria. This assumption may be unrealistic, in particular when dealing with consumers' data.

If we consider consumers' preferences for various criteria (smell, color, taste, …), the question of homogeneity of the preferences is justifiable. In that case, it could be interesting to perform a segmentation of the panel, in order to investigate differences between individual perceptions. We would like, thus, to segment the consumers' panel in homogeneous classes for all the criteria and estimate the configurations of the products for the different criteria (Bradley's scores) for each segment. It is thus necessary to take into account all the criteria to realize the segmentation of the panel. The objective is, on the basis of Bradley's model, to be capable to identify the different classes of homogeneous perceptions on the whole criteria and to estimate the configurations of the products, in each of these classes.

The package CompR (Séménou, 2015) allows on one hand to segment the panel in homogeneous classes for all the studied criteria and on the other hand to estimate the configurations of products (Bradley's scores) for each of the classes.

The approach is based on a mixture of distribution within the framework of Bradley's model. The situation with only one criterion was already developed by Courcoux and Séménou (1997).

Package CompR generalizes of this latent class approach for multi-criteria paired comparison experiments. Maximum likelihood estimates of within segment scale values $\boldsymbol{\pi}$ and class membership probabilities $\boldsymbol{\alpha}(t)$ can be obtained by solving iteratively the likelihood equation using EM algorithm (Dempster et al., 1977).

2.3 Determination of the number of classes

In order to determine an adequate number of classes, one can use Information criteria such as AIC, BIC, CAIC (McLachlan and Basford 1988) or perform likelihood ratio test using Monte Carlo simulations (McLachlan and Basford 1988).

3. Application

The study concerns preferences for 7 cocktails with fruit juice evaluated by 112 consumers in paired comparison. (data(Cocktail) in CompR package).

The evolution of information criteria leads us to retain the existence of two latent classes (Table 1). The results can be obtained by using the EstimBradley function in CompR Package.

Number of classes	AIC	BIC	CAIC
1	858.40	886.39	892.39
2	762.40	823.03	836.03
3	757.00	850.29	870.29

Table 1: Evolution of information criteria for T=1 to 3 classes

The weights of the different classes and products' configuration (Table 2) are available by using the EstimBradley function in CompR Package.

	Class1: Weight 0.75	Class2: Weight 0.25
P1	0.002	0.050
P2	0.329	0.054
P3	0.029	0.240
P4	0.061	0.386
P5	0.255	0.057
P6	0.023	0.160
P7	0.302	0.052

Table 2: Products' configuration and weights of the different segments

It is also possible to represent products' scores in each class by using the Piplot function in CompR package (Figures 1 and 2).

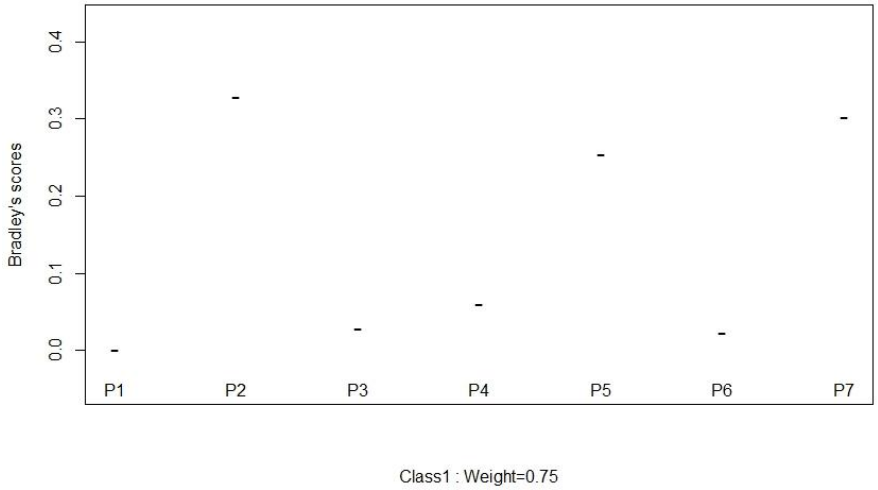

Figure 1: Products' configuration in Class 1 for the two latent classes' solution.

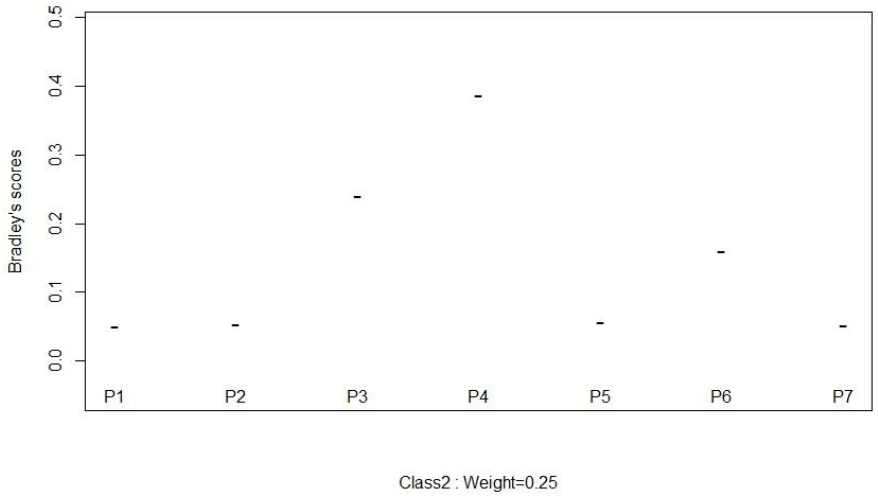

Figure 2: Products' configuration in Class 2 for the two latent classes' solution.

References

Bradley, R. A. (1976). Science, statistics and paired comparisons. Biometrics 32, 213-232.

Bradley, R. A. & Terry, M. E. (1952). Rank analysis of incomplete block designs: the method of paired comparisons. Biometrika, 39, 324-345.

David H.A. (1988). The method of paired comparisons. Charles Griffin, London, UK.

Dykstra, O. Jr. (1956). A note on rank analysis of incomplete block designs: A method of paired comparisons employing unequal repetitions on pairs. Biometrics, 12, 301-306.

Courcoux, Ph. & Séménou, M. (1997). Preference data analysis using a paired comparison model Food quality and Preference Vol. 8, No. 5/6, 353-358.

McLachlan, G. L. & Basford, K. E. (1988). Mixtures models: inference and application to clustering. Marcel Dekker, New York.

Séménou, M. (2015). CompR : Paired Comparison Data Analysis.  package version 1.0. https://cran.r-project.org/web/packages/CompR/index.html