A multi-product approach for detecting subjects' and objects' covariates in consumer preferences

Stefania Capecchi[1], Isabella Endrizzi[2]

[1] Department of Political Sciences, University of Naples Federico II, Italy

[2] FEM, Agrifood Quality Department, S. Michele (Trento), Italy

stefania.capecchi@unina.it; isabella.endrizzi@iasma.it

Abstract

It is important in the analysis of consumer preference exploits both information deriving from subjects' and products' characteristics and this is particularly relevant in Sensometrics where decision to buy depends both on the personal habits and the organoleptic properties of food and beverages. Multivariate analysis copes with these needs by several methods based on factorial approaches and some applications have been performed in order to cluster consumers with respect to products. In this paper a different framework based on a parametric version of the process generating the hedonic scores is adopted. More precisely, a probability distribution for the ordinal responses is proposed as a mixture of feeling and uncertainty components and both of them are related to subjects' and products' characteristics. The approach is made effective by inferential methods based on maximum likelihood and asymptotic inference. A real case study is discussed and some advantages are considered.

Key Words: Preference data, Subjects' covariates, Objects' covariates, CUB models

1 Introduction

Consumer preference analysis is a multi-facet area of investigations where different disciplines intersect their objectives and tools: Psychology, Sociology, Economics and related fields. More specifically, when products are food and beverage potential buyers' behaviour is carefully examined both from Sensometric and from Marketing point of view. To simplify, sensory analysis is mainly devoted to check if and how products' physical and chemical characteristics affect the consumer reaction; on the other hand, marketing studies detect people characteristics to pick out clusters of consumers and to relate preferences or disliking towards the products with subjects' covariates.

This dichotomy has been recently debated since it is a common evidence that the selection of a product is the result of a complex human decision where personal, family and social habits interfere with price, packaging, advertising and organoleptic variables of the product. Thus, a strategy where both subjects' and object's covariates are included entails a more complete analysis of the consumer behaviour.

In this regard, different approaches have been proposed according to the objectives of the study. Recent research focusses on the so-called L-structured data where both kind of information are exploited with statistical methods. In addition, Partial Least Squares (PLS) with clustering algorithm is an effective tool to achieve important results in the study of food quality and preference as shown by Vigneau and Qannari (2003) and Endrizzi *et al.* (2010), among others.

A different paradigm is pursued in this paper and it derives from the awareness that consumers express their preference towards a set of products on the basis of a psychological mechanism where both feeling and uncertainty are always present with different weights; in addition, these components may be consistently related to both subjects' and objects' covariates which are to be explicitly stated in the modelling exercise.

The paper is organized as follows: in the next Section, the structure of data and the specification of the selected model are formally assessed. Then, a real case study is presented in Section 3 whereas in Section 4 the main results of the approach are shown and commented, also by means of graphical tools. A discussion and some concluding remarks end the work.

2 Structure of data and model specification

In this section, we extend the standard modelling approach when products compared by the same group of respondents have to be jointly examined for exploring how *choices' covariates* and *choosers' covariates* affect the stated preferences (Agresti, 2010). After a brief mention to the classical approach generally applied in sensory analysis, we introduce multi-objects CUB models. Hereafter, we adhere to standard multivariate notation, which expresses subjects' measurements along rows and covariates along columns.

When panellists express preferences on several products, the effect of the sensory characteristics of each object (generally, obtained by a samples of qualified assessors and or technical instruments) and socio-demographic, attitude and habits variables of subjects (generally, obtained by questionnaires or oral interviews) on the ordered responses (preferences) have been analyzed by using the concept of *L-structured data*. Both subjects' and objects' covariates account for the responses in a different manner: personal characteristics vary among n subjects whereas products' characteristics vary along K objects and so the relationship follows an L-shaped structure.

We assume a matrix $\boldsymbol{R}(n \times K)$ of hedonic preferences expressed by n panellists on K food products, a matrix $\boldsymbol{X}(n \times s)$ concerning the values of s covariates on the n subjects and a matrix $\boldsymbol{Z}\,(K \times H)$ which contains a synthesis of H physical and/or chemical measurements on the K objects.

In sensometric analysis, such information are often presented in the following format (Endrizzi, 2008):

$$\begin{array}{ccc} \boxed{\boldsymbol{Z}\,'} & & \\ (H \times K) & & \\ \Downarrow & & \\ \boxed{\boldsymbol{R}} & \Longleftarrow & \boxed{\boldsymbol{X}} \\ (n \times K) & & (n \times s) \end{array}$$

which justifies the name of *L-shaped structure*.

According to the lines introduced by Wold *et al.* (1983), several researchers have been taken all these matrices into account by means of Partial Least Squares (PLS) regression methods (Martens *et al.*, 2005; Esposito Vinzi *et al.*, 2007, 2010). This approach has been specifically applied to sensory data, for clustering and classification purposes (Vigneau and Qannari, 2003), as L-PLS regression by Endrizzi (2008) and Plaehn and Lundahl (2006), among others. In this regard, most of the statistical literature is involved with algorithmic aspects of the PLS approach and its generalizations, also with reference to structural equations models with latent variables (Boari and Cantaluppi, 2011).

We show how the same data set may be effectively exploited within the logic of CUB models without modifying the existing software programs in order to obtain additional information about the structure of data. We refer here to cognitive Psychology which assessed that human decision is a very complex act where the personal history of the subject, the circumstances and context of the decision and the object to be selected interact in several ways. However, when the final choice implies to pick out select an ordinal category from a list of m prefixed values (qualitative or coded as quantitative) it is possible to focus the data generating process as mainly composed of two factors: an attraction (positive or negative) towards the item and an inherent indecision which surrounds any human choice. The model who is the kernel of the following discussion is an effective way to formalize such an approach by means of two discrete random variables.

Briefly, we assume that the response R is a random variable defined over the support $\{1, 2, \ldots, m\}$, for a given m, whose probability mass distribution is:

$$Pr\,(R = r \mid \boldsymbol{\theta}) = \pi \binom{m-1}{r-1} \xi^{m-r}\,(1-\xi)^{r-1} \,+\, (1-\pi)\,\frac{1}{m}\,, \quad r = 1, 2, \ldots, m\,. \tag{1}$$

The mixture is a convex Combination of a discrete Uniform and a shifted Binomial random variable, and this motivates the acronym CUB . Since (1) is well defined for parameter vector $\boldsymbol{\theta} = (\pi,\, \xi)$ such that $\pi \in (0, 1]$ and $\xi \in [0, 1]$, the parametric space is the (left open) unit square. The case $m = 3$ implies a saturated model where CUB models are identifiable for any $m > 3$ (Iannario, 2010).

We observe that $\pi \to 0$ implies a model with an almost totally uncertain selection of a category, whereas $\pi \to 1$ implies a drastic reduction of this component: then, $1 - \pi$ may be considered as a measure of the *uncertainty* of the selection. Similarly, when $\xi \to 0$ high values of the support are more likely whereas for $\xi \to 1$ the low values of the support are preferred: then, $1 - \xi$ may be considered as a direct measure of the positive *feeling* (attraction) towards the item. Thus, both parameters have an immediate interpretation since they summarize uncertainty and feeling of the respondents, respectively. In addition, since a CUB model is univocally defined by (π, ξ), each model is depicted in the unit square by means of a point with coordinates $(1 - \pi, 1 - \xi)$, which can be immediately associated to the weight of uncertainty and feeling, respectively. As a consequence, it will be immediate to visualize the effect of covariates on the estimated models in terms of uncertainty and feeling. Then, it is really important to introduce covariates in the model (1) by means of a logistic link with the parameters π_i ad ξ_i, which are now immediately related to respondents' characteristics.

More precisely, a CUB model with p covariates to explain uncertainty and q covariates to explain feeling is specified by:

1. A *stochastic component*:

   $$Pr\,(R_i = r \mid \boldsymbol{y}_i;\, \boldsymbol{w}_i) = \pi_i \binom{m-1}{r-1} \xi_i^{m-r} (1-\xi_i)^{r-1} + (1-\pi_i)\left(\frac{1}{m}\right),$$

   for $r = 1, 2, \ldots, m$, and for any i-th subject, $i = 1, 2, \ldots, n$.

2. Two *systematic components*:

$$logit(\pi_i) = \boldsymbol{\beta}\boldsymbol{x}_i^{(\pi)}\,; \qquad logit(\xi_i) = \boldsymbol{\gamma}\boldsymbol{x}_i^{(\xi)}\,.$$

where we denote $logit(z) = \log(z/(1-z))$ for any real $z \in (0,1)$ and $\boldsymbol{x}_i^{(\pi)}$ and $\boldsymbol{x}_i^{(\xi)}$, for $i = 1, 2, \ldots, n$, are the subjects' covariates for explaining π_i e ξ_i, respectively.

For convenience, we set $y_{i0} = w_{i0} = 1, \forall i$ and $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_p)'$, $\boldsymbol{\gamma} = (\gamma_0, \gamma_1, \ldots, \gamma_q)'$. Notice that $\boldsymbol{x}_i^{(\pi)}$ and $\boldsymbol{x}_i^{(\xi)}$ are subset of a row $\boldsymbol{x}_i$ which includes al information collected on the i-th subject, for $i = 1, 2, \ldots, n$.

The logistic link has been preferred for its formal simplicity; however, any mapping from $\mathbb{R}$ to $(0,1)$ is a legitimate choice, as those generally advocated in GLM approach: see Agresti (2010) for details. Given our parameterization, the covariates $\boldsymbol{x}_i^{(\pi)}$ and $\boldsymbol{x}_i^{(\xi)}$ may be coincident, completely different or partially overlap.

In addition to previous information, let us consider the H sensory measurements on the K products collected in the $(K \times H)$ matrix:

$$\boldsymbol{Z} = \{z_{kh},\ k = 1, 2, \ldots, K; \quad h = 1, 2, \ldots, H\}\,,$$

so that $\boldsymbol{z}_k = (z_{k1}, z_{k1}, \ldots, z_{kH})$ is the row vector of the H sensory measurements available for a given k-th product, $k = 1, 2, \ldots, K$. Then, both subjects' and objects' covariates may be introduced in the framework of CUB models, as successfully experienced by Piccolo and D'Elia (2008).

In fact, we can jointly consider all K products in a unique CUB model where parameters and subjects' and objects' characteristics are linked by means of:

$$\begin{cases} \pi_{ik} = \dfrac{1}{1 + e^{-\boldsymbol{x}_i^{(\pi)}\,\boldsymbol{\beta} - \boldsymbol{z}_k\,\boldsymbol{\delta}}}; \\ \xi_{ik} = \dfrac{1}{1 + e^{-\boldsymbol{x}_i^{(\xi)}\,\boldsymbol{\gamma} - \boldsymbol{z}_k\,\boldsymbol{\eta}}}; \end{cases} \qquad i = 1, 2, \ldots, n;\ k = 1, 2, \ldots, K. \tag{2}$$

Here, $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_K)'$ and $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_K)'$ are parameter vectors which measure the impact of the product characteristics on uncertainty and feeling components, respectively. If these parameters are significant, their interpretation is relevant since they reflect the effect on the final preference of chemical and physical components for several products in a straightforward manner. In this specification, the information set is composed by $\mathcal{S}_n = (\tilde{\boldsymbol{r}} \mid \tilde{\boldsymbol{X}}^{(\pi)} \mid \tilde{\boldsymbol{X}}^{(\xi)} \mid \tilde{\boldsymbol{Z}})$ where $\tilde{\boldsymbol{r}} = vec(\boldsymbol{R})$ is the vectorized matrix of the hedonic scores of the K products, $\tilde{\boldsymbol{X}}^{(\pi)} = (\boldsymbol{1}_K \bigotimes X^{(\pi)})$

and $\tilde{\boldsymbol{X}}^{(\xi)} = (\mathbf{1}_K \bigotimes X^{(\xi)})$ are the matrices of subjects' covariates able to explain uncertainty and feeling, respectively (they are derived from $\boldsymbol{X}^{(\pi)}$ and $\boldsymbol{X}^{(\xi)}$, that is subsets of $\boldsymbol{X}$, a subjects' covariates matrix), and finally $\tilde{\boldsymbol{Z}} = (\boldsymbol{Z} \bigotimes \mathbf{1}_n)$. In this context, $\mathbf{1}_H$ and $\mathbf{1}_N$ are unit vectors of length K and $N = nK$, respectively.

More specifically, according to (2), $1 - \pi_{ik}$ $(1 - \xi_{ik})$ is related to *uncertainty* (*feeling*) expressed by the i-th subject, whose profile is specified by $\boldsymbol{x}_i^{(\pi)}$ (by $\boldsymbol{x}_i^{(\xi)}$) when faced to the k-th object, whose characteristics are specified by $\boldsymbol{z}_k$. It should be noted that the "intercepts" β_0 and γ_0 of the model (2) contain the joint level effect of the i-th subject and k-th object with regard to uncertainty and feeling, respectively.

With respect to the multivariate approach of *L-structured data*, this framework allows for an explicit modelling formulation of the probability of different answers given the profiles of respondents and the contents of several products.

3 A real case study

The data we will study come from the INTERBERRY project, a multidisciplinary research which aims to improve the quality and marketability of soft fruit, carried out by FEM (Edmund Mach Foundation, Italy) whose details are discussed in Endrizzi *et al.* (2009). In order to exploit the sensory characteristics and nutritional advantages of soft fruits, 25 juice prototypes obtained without any pasteurization treatment were designed and selected using a focus group of 10 people involved in several aspects of juice production: marketing, product development and research. Each one of the five berry fruits examined (strawberries, raspberries, blackberries, redcurrants and blueberries) was offered in five different formulations: freshly squeezed berry fruit (20%) was mixed with each of the five different base juices (apple, orange, blood orange, pineapple and pomegranate) in order to find out which base juice could enhance the sensory characteristics of each berry fruit. These juices have been firstly analyzed in terms of chemical compositional parameters; then, they have been assessed in terms of overall liking in five different sessions by a consumer panel, recruited from FEM staff members and students who claimed to like berry fruits. Information about consumer characteristics and habits regarding fruit and fruit juices purchased and consumed was also collected. The chemical analysis procedure, the design and execution of consumer test, the list of questionnaire items and the variable description can be found in Endrizzi (2008) who describes also the statistical analysis conducted to cluster the panel of consumers by PLS methods and by taking into account the overall information on consumer preferences for food products, the chemical descriptors of products and the socio-demographic characteristics, consumption and purchasing behaviour of consumers. Two-step L-PLSR has been to describe the global relationship among the three data tables and to perform a comparison between consumer classifications which come from the two approaches (Endrizzi *et al.*, 2010).

All the collected information useful for our research may be arranged in the following matrices:

— $\boldsymbol{R}(25, 72)$ contains the hedonic scores of the 72 consumers regarding the 25 juice mixes expressed, on a 9-point scale.

— $\boldsymbol{Z}(25, 15)$ contains 5 chemical compounds expressed by quantitative variables (sugar content, malic acid, citric acid, ascorbic acid, total amount of poliphenols) and the qualitative design variables coded as dummy variables: all these variables have been conveniently centred and standardized.

— $\boldsymbol{X}(72, 179)$ includes socio-demographic descriptors (gender, age class, educational background, etc.), fruit consumption and purchasing habits, juice consumption and purchasing habits, berry fruit consumption and purchasing habits, food neophobia scale scores, impressions of new foods, exotic foods, ready-to-eat foods and familiar foods and some measure of knowledge about healthy diet and antioxidants in general.

According to the framework of CUB models, different analysis will be performed on the basis of different information that will be used. First of all, the hedonic scoring on the 25 formulations obtained by mixing the 5 berry fruits examined (1:strawberry, 2:raspberry, 3:blackberry, 4:red currant, 5:blueberry) with the 5 base juices (A:apple, O:range, BO:blood orange, P:pineapple, PG:pomegranate) have been examined by modelling a CUB probability distribution for each beverage; then, the estimates are visualized in the parameter space (Figure 1). It is evident that the pineapple base is most appreciated in all combinations (especially, 2P:rasperry with pineapple) and besides with a low uncertainty. On the contrary, pomegranate receives quite low liking in all combination with high uncertainty except for 2PG:raspberry with pomegranate (which receives high scoring with high uncertainty, and this denotes a large heterogeneity in the responses). For instance, a raspberry juice supports completely different feeling and uncertainty components which vary with the presence of the basic juice. As a global evidence of this analysis, the responses are substantially more clustered with respect to base juices than with respect to berry fruits; thus, the accurate selection of the base juice is a fundamental step in forming the preference in the preparation of these kinds of juices.

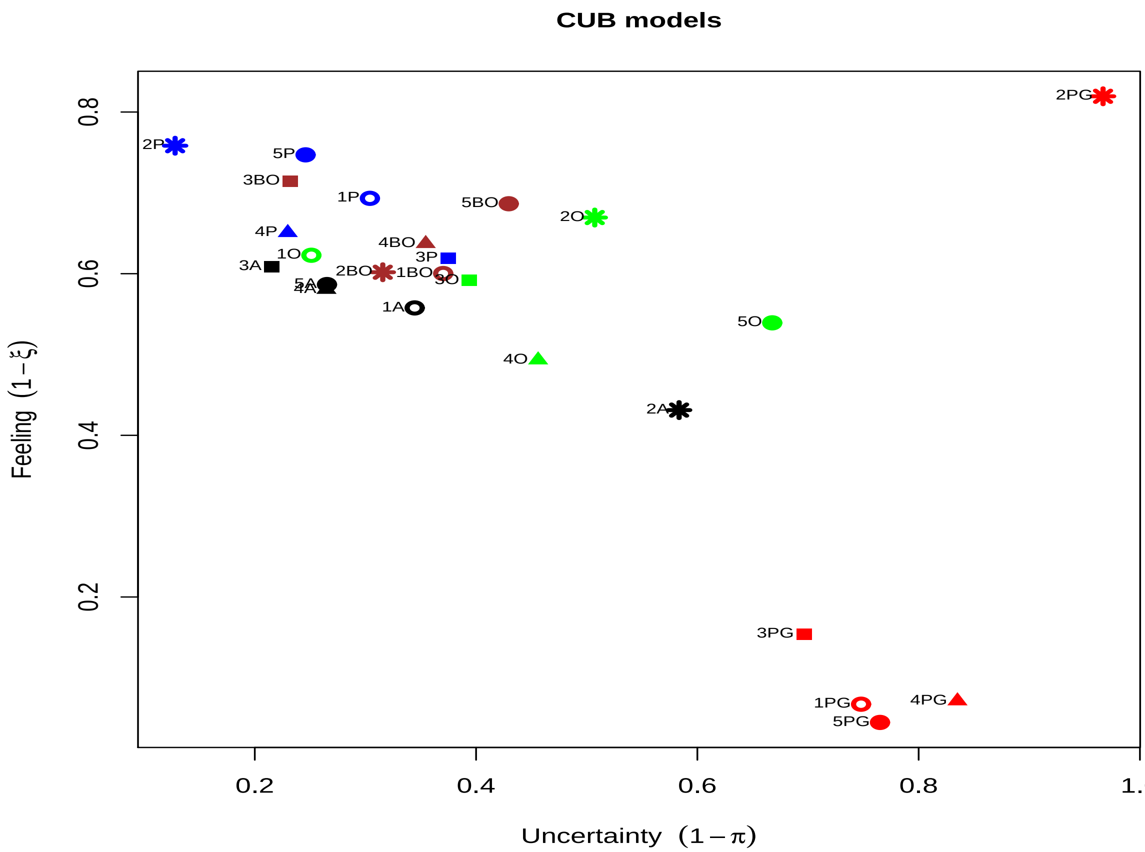

Figure 1: CUB model estimation for the hedonic scoring of the 25 mixed formulations

The next step consists in evaluating the significance of subjects' covariates for feeling and uncertainty, and this is generally obtained through a stepwise strategy based on backward and forward techniques. However, in this paper we would focus on the ability of CUB models framework to take into account the multi-item aspects of the estimation process by means of both subjects' and objects covariates. In fact, we consider the preferences of all berry fruits (with different bases) and look for significant covariates. These analyses produce several models and in Table 1 we summarize just the main results. A likelihood ratio test (LRT) –which compares the log-likelihoods of the estimated CUB models without and with the listed covariates– is also presented (degrees of freedom in parenthesis): they are all quite significant given the values $\chi^2_{(6)} = 12.592$ and $\chi^2_{(6)} = 14.067$ for a significant level of $\alpha = 0.05$.

Table 1: Multi-items CUB models and significant covariates

Berry fruits	Subjects' covariates	Objects' covariates	*LRT* (*df*)
Strawberry	staff, likejuice, raspberry, marviolet	sugar, malic	58.076 (6)
Raspberry	ageclass, likejuice, apple, marviolet	sugar, malic, ascorbic	67.157 (7)
Blackberry	staff, likejuice, orange	sugar, malic, ascorbic	92.685 (6)
Red Currant	smoke, likejuice, raspberry	sugar, malic, citric	52.671 (6)
Blueberry	children, married, likejuice, green	sugar, malic, citric	67.785 (7)
`Legenda`:			
Dichotomous:	staff, married, smoke $(0, 1)$		
Polytomous:	ageclass (from 1 to 8)		
Ordinal:	likejuice, orange, raspberry, apple, marviolet, green		
Continuous:	sugar, malic, ascorbic, citric		

From Table 1 several considerations may be summarized:

- A basic preference for juices (measured by `likejuice`, which is a normalized sum of all hedonic scores for a list of juices) is significant in any case.
- Preference of `staff` is different from students only for *Strawberry* and *Blackberry.* A difference in `age` and `smoke` affect only *Raspberry* and *Red Currant*, respectively.
- A `married` respondent with `children` modifies the preference for *Blueberry.*
- A specific preference for raspberry, apple, orange (as basic juices) affect the preference of *Strawberry* and *Red Currant*, *Raspberry*, *Blackberry*, respectively.
- A preference for colour `maroon/violet` is significant for *Strawberry* and *Raspberry* whereas `green` affects the preference towards *Blueberry.*

- `Sugar` and `malic` acid are everywhere significant; however, they affect preferences of *Raspberry* and *Blackberry* if combined with `ascorbic` acid where are important for *Red Currant* and *Blueberry* if combined with `citric` acid.

In this framework several other results may be derived by combining subjects' and objects' covariates for the berry fruits; however, we prefer to build an omnibus model which encompasses all fruits of this research in a unique structure. This result is achieved by vectorizing all hedonic scores of the 25 combinations and defining for each of them the corresponding subjects' characteristics and chemical compositions. Here, most of the variables concerning the expressed consumption and liking for fruits and juices have been normalized in $[0, 1]$ in order to get comparable effects and simpler interpretations.

The best model we get is the following one:

$$\begin{cases} \pi_{ik} \;\equiv\; \pi & = \; 0.570; \\ \\ logit\,(1-\xi_{ik}) & = \; -1.761 - 0.469\, staff_i + 2.597\, likejuice_i + 0.678\, orange_i \\ & = \; +0.745\, blueberry_i - 0.311\, green_i - 0.216\, yelloworange_i - 0.446\, impcolours_i \\ & = \; -8.283\, sugar_k - 3.297\, malic_k - 1.291\, citric_k \end{cases} \tag{3}$$

for $i = 1, 2, \ldots, n$ and $k = 1, 2, \ldots, K$.

The estimated model (3) suggests some considerations:

- The role of uncertainty in the scoring process is comparatively important and constant for all products. In fact, it is possible to relate uncertainty to some subjects' covariates but its significance disappears in presence of feeling covariates.

- The only global effect of a personal variable is the `staff` who consider more critically all compositions with respect to students.

- A positive effect of an attitude towards `juices` is significant for the final score in any case and with an important weight, and similar effect is obtained by a preference for `blueberry` and `orange`.

- A negative contribution is obtained from the personal liking of `green` and `yellow-orange` colours as well as for a general `importance given to colours`.

- On the final scoring the fundamental ingredients are `sugar` and `malic` and `citric` acids; their contribution is negative for the feeling with `sugar` definitively more important.

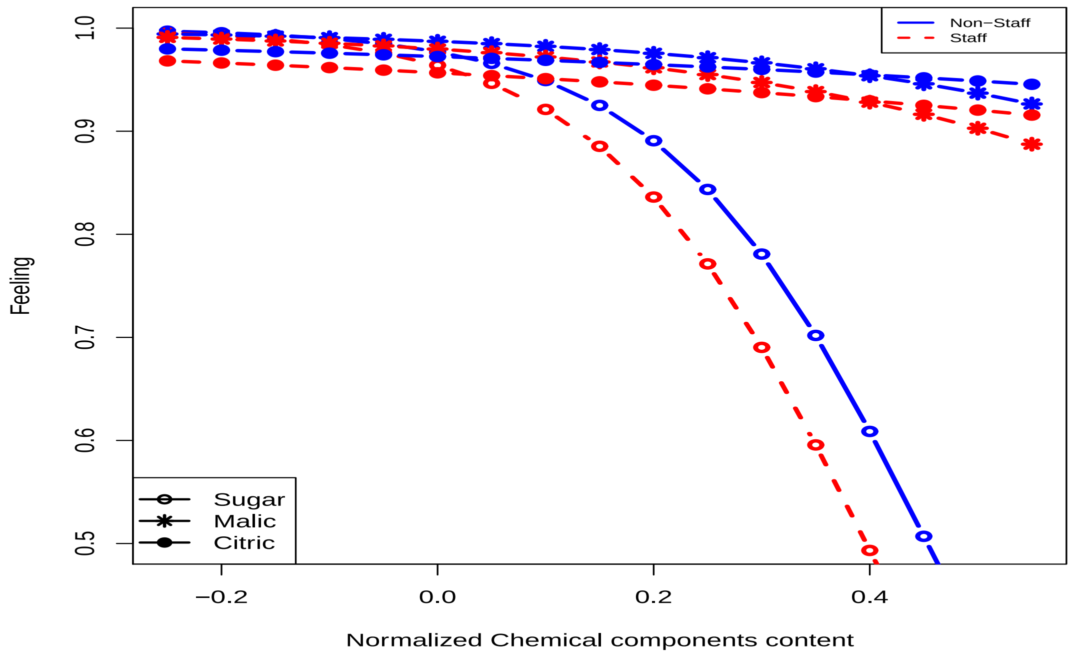

Figure 2: Feeling as a function of staff and chemical compounds

Such considerations become more evident if the graphical abilities of this approach are fully exploited. In our case, given the many covariates of the models, the pattern of some of them on the feeling may be analyzed only *ceteris paribus*, that is by letting the conditioned variables at the average values observed in the sample.

Figure 2 shows the modifications in the feeling expressed by respondents (both staff and non-staff) when the composition of chemical compounds varies. It turns out that a variation under the average is substantially is negligible whereas when these compounds (sugar, malic and citric acid) increase more than the average the effect is more significant and in the case of sugar extremely pervasive. The effect of staff/non-staff is homogeneous in all contexts with a constant decreasing of feeling when the consumer belongs to staff.

Figure 3 represents a more complex situation where the effect on the feeling of 4 variables

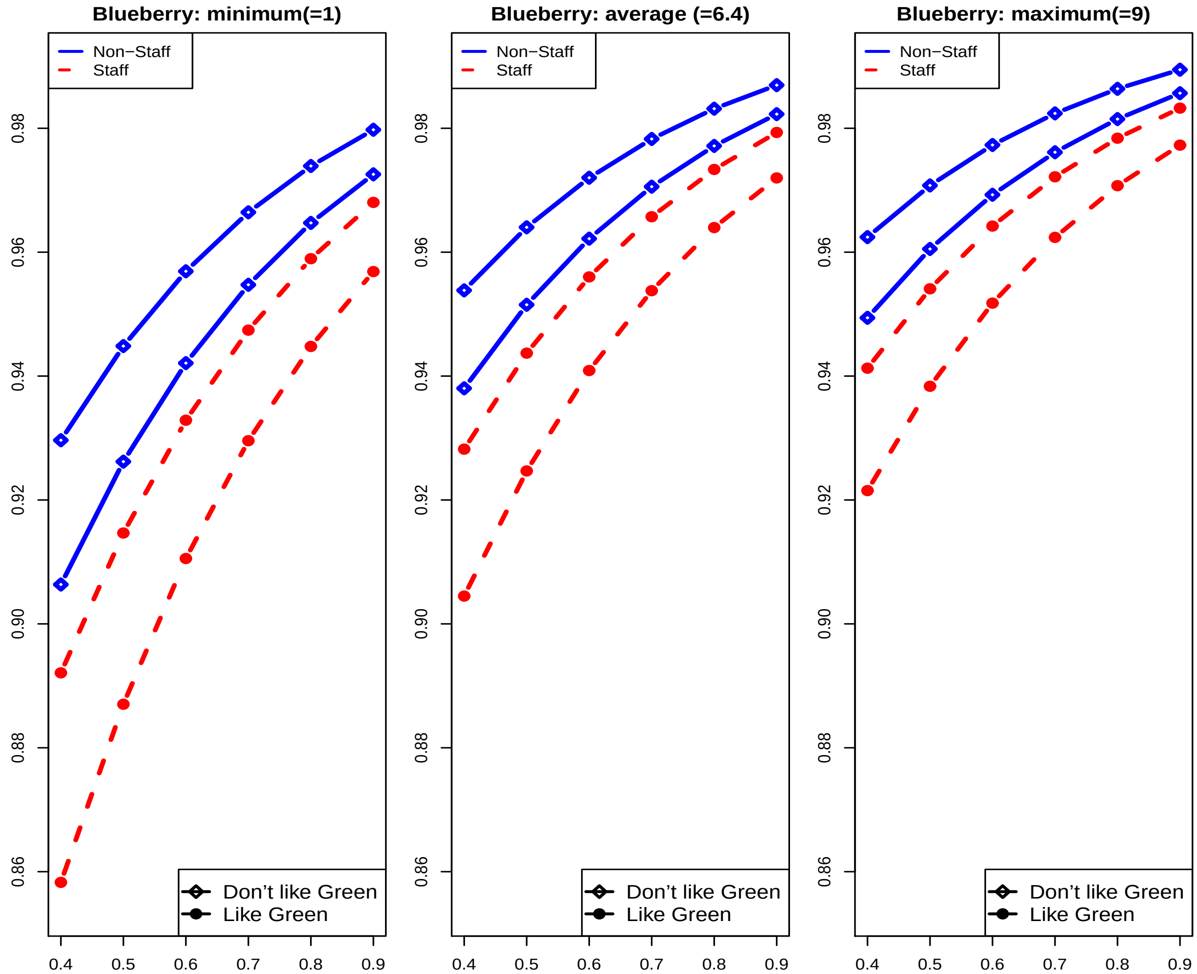

Figure 3: Feeling as a function of covariates: `likejuices, green, staff, blueberry`

(`likejuices, green, staff, blueberry`) are jointly analyzed. In fact, we consider how the feeling is modified with respect to a measure of a global `liking for all juices` (in abscissa, such a measure in the sample varies in [0.43, 0.90]) given the preference for the `green` colour and the score given to preference for `blueberry` (an important covariate of the model (3)); finally, a difference between `staff` and non-staff is also plotted. It is confirmed a more critical evaluation of staff in all situations and a negative effect on the hedonic score from people who like green colour. On the contrary, the global liking of all juice reflects on the hedonic score with a positive effect which combines with the score given to blueberry: this behaviour is even more evident when a respondent scores blueberry more than the average (= 6.4).

4 Discussion and conclusions

The usefulness to include a sensometric analysis in a parametric framework consists in obtaining a measurable effect of subjects' and products' covariate on the final hedonic scores and in the possibility to plot the main relationship by means of effective graphical tools. In addition, the statistical significance of the results is assessed in a formal manner; in fact, estimation and test are consolidated fields thanks to the asymptotical maximum likelihood theory.

In this paper we examine the joint effects of the personal characteristics and chemical contents of juice on the hedonic scores of a sample of consumer in case of the berry fruit experiments. Some relevant feature have been discussed and the main relationships graphically considered.

Although all the results here presented derive from significant estimates, a critical issue is the moderate size of the sample which consists of 20 staff people and 52 students. It is likely that with a large set of consumers more significant relationships could be derived.

Acknowledgement: This work has been realized with the support of SHAPE project within the frame of Programme STAR (CUP E68C13000020003) at University of Naples Federico II, financially supported by UniNA and Compagnia di San Paolo.

References

Agresti, A. (2010). *Analysis of Ordinal Categorical Data*, 2^{nd} edition, J.Wiley & Sons, Hoboken.

Boari G. and Cantaluppi G. (2011). PLS models, in: Kenett R. S. and Salini S. (eds.), *Modern Analysis of Customer Surveys: with applications using R*, Chapter 16, J. Wiley & Sons, Chichester, 309–331.

Endrizzi, I. (2008). *Clustering of variables around latent components: An application in consumer science*. Ph.D. Thesis, University of Bologna.

Endrizzi, I., Pirretti, G., Calò, G. and Gasperi, F. (2009). A consumer study of fresh juices containing berry fruits. Journal of Science of Food and Agriculture, **89**, 1227–1235.

Endrizzi I., Gasperi F., Calò D. G. and Vigneau E. (2010). Two-step procedure for classifying consumers in a L-structured data context, *Food Quality and Preference*, **21**, 270–277.

Esposito Vinzi V., Chin W. W. , Henseler J. and Wang H. (2010). *Handbook of Partial Least Squares*, Springer, New York.

Esposito Vinzi V., Guinot C. and Squillacciotti S. (2007). Two-step PLS regression for L-structured data. An application to cosmetic industry, *Statistical methods and applications*, **16**, 263–278.

Iannario M. (2010). On the identifiability of a mixture model for ordinal data, *Metron*, **LXVIII**, 87–94.

Iannario M. (2012a). Modelling *shelter* choices in a class of mixture models for ordinal responses, *Statistical Methods and Applications*, **21**, 1–22.

Iannario M. (2012b). CUBE models for interpreting ordered categorical data with overdispersion, *Quaderni di Statistica*, **14**, 137–140.

Iannario M. (2012). Preliminary estimators for a mixture model of ordinal data, *Advances in Data Analysis and Classification*, **6**, 163–184.

Iannario M. (2012d). Hierarchical CUB models for ordinal variables, *Communications in Statistics. Theory and Methods*, **41**, 3110–3125.

Iannario M. (2013). Modelling uncertainty and overdispersion in ordinal data, *Communications of Statistics. Theory and Methods*, **42**, 771–786.

Iannario, M., Manisera, M., Piccolo, D. and Zuccolotto, P. (2012). Sensory analysis in the food industry as a tool for marketing decisions, *Advances in Data Analysis and classification*, **6**, 303–321.

Iannario M. and Maravalle M. (2012). Parametric and nonparametric approaches to the semantic relationship among synonymy, *Quaderni di Statistica*, **13**, 83–108.

Iannario M.and Piccolo D. (2009). A program in R for CUB models inference, Version 3.0, available from *http://www.researchgate.net/publication/228987009*

Iannario M. & Piccolo D. (2010b). A New Statistical Model for the Analysis of Customer Satisfaction, *Quality Technology & Quantitative Management*, **7**, 149–168.

Iannario M. and Piccolo D. (2012a). CUB models: Statistical methods and empirical evidence, in: Kenett R. S. and Salini S. (eds.), *Modern Analysis of Customer Surveys: with applications using R*, J. Wiley & Sons, Chichester, pp. 231–258.

Martens H., Anderssen E., Flatberg A., Gidskehaug L. H., Høy M., Westad F., Thybo A. and Martens M. (2005). Regressions of a matrix on descriptors of both its rows and of its columns via latent variables: L-PLSR, *Computational Statistics & Data Analysis*, **48**, 103–123.

Piccolo D. (2003). On the moments of a mixture of uniform and shifted binomial random variables, *Quaderni di Statistica*, **5**, 85–104.

Piccolo D. and D'Elia A. (2008). A new approach for modelling consumers' preferences, *Food Quality and Preference*, **19**, 247–259.

Piccolo D. and Iannario M. (2010). A new approach for modelling consumers' preferences, *Proceedings of the* 11^{th} *European Symposium on Statistical Methods for the Food Industry*, University of Sannio, Benevento, editore Academy School, Afragola, 139–148.

Plaehn D. and Lundahl D. S. (2006). An L-PLS preference cluster analysis on French consumer hedonics to fresh tomatoes, *Food Quality and Preference*, **17**, 243–256.

Vigneau E. and Qannari E. M. (2003). Clustering of variables around latent components, *Communications in Statistics. Simulations and Computations*, **12**, 1131–1150.

Wold S., Albano C., Esbensen W. J. III, Hellberg S. and Johansson E. *et al.* (1983). Pattern recognition: Finding and using regularities in multi-variate data, in H. Martens & H. Russwurm (eds.), *Food research and data analysis*, Applied Science Publishers, 147–188.

Wold H. (1985). Partial least squares, in S. Kotz & N. l. Johnson (eds.), *Encyclopedia of Statistical Sciences*, Vol.6, J. Wiley & Sons, 581–591.