Available online at www.sciencedirect.com

SCIENCE DIRECT®

Food Quality and Preference 17 (2006) 257–265

Food Quality and Preference

www.elsevier.com/locate/foodqual

A cluster approach to analyze preference data: Choice of the number of clusters

Karin Sahmer *, Evelyne Vigneau, El Mostafa Qannari

ENITIAA/INRA Unité de Sensométrie et Chimiométrie, Rue de la Géraudière, BP 82 225, 44322 NANTES cedex 03, France

Received 25 September 2004; received in revised form 18 February 2005; accepted 5 March 2005
Available online 26 April 2005

Abstract

We consider the clustering of a panel of consumers according to their scores of liking. The procedure is based on a cluster of variables approach proposed by Vigneau et al. [Vigneau, E., Qannari, E. M., Punter, P. H., & Knoops, S. (2001). Segmentation of a panel of consumers using clustering of variables around latent directions of preference. *Food Quality and Preference, 12*, 259–363]. We aim at setting up a hypothesis-testing framework in order to determine the appropriate number of clusters. The procedure consists of two steps. Firstly, a cluster tendency test determines if there is more than one cluster. Secondly, a hierarchical algorithm is performed and cluster validity tests at the different levels of the hierarchy indicate the appropriate number of clusters. Once the number of clusters is determined, a partitioning algorithm is implemented by considering as a starting point the partition obtained from the hierarchical algorithm. We illustrate the method on preference data from a European sensory and consumer study on coffee [ESN (1996). *A European sensory and consumer study: A case study on coffee*. European Sensory Network] and we undergo a simulation study in order to assess the efficiency of the procedure.

Keywords: Bootstrap; Clustering; Cluster tendency; Cluster validity; Preference data

1. Introduction

Preference studies where a panel of consumers score different products according to their liking are very popular and very useful for product development. The results can be used to put on the market products that are liked very much. However, the consumers differ in their preferences and it is of prime interest to detect the existence of segments among the panel of consumers. Very often, preference data are analyzed by means of an internal preference mapping (Greenhoff & MacFie, 1994). Different groups of consumers can be determined by a visual inspection of the loading plot. However, this becomes very cumbersome and time consuming when the number of consumers is large and when the percentage of the total variance of the preference data explained by the first two principal components is small. An automatic tool to identify homogenous clusters of consumers is needed (Helgesen, Solheim, & Næs, 1997).

Vigneau, Qannari, Punter, and Knoops (2001) proposed a clustering of variables approach that directly gives clusters of consumers and, for each cluster, a latent variable that represents the preferences of the cluster. It consists of a partitioning algorithm that takes as a starting point the outcomes of a hierarchical clustering procedure. A plot of the clustering criterion in the hierarchy can help to decide the number of clusters. For different clustering methods, Hardy (1996) showed that this graphical method gives interesting results. However, the graphical identification seems to be very subjective. Duda, Hart, and Stork (2001) propose a significance test. Their approach concerns the clustering of individuals

* Corresponding author. Tel.: +33 251 78 54 35; fax: +33 251 78 54 38.
E-mail address: sahmer@enitiaa-nantes.fr (K. Sahmer).

0950-3293/$ - see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.foodqual.2005.03.007

and assumes a normal distribution. For the clustering of variables, the VARCLUS procedure of the SAS software contains a stopping rule (SAS/STAT, 1999). This stopping rule depends on the options chosen by the user and does not have a hypothesis testing background. We propose a hypothesis testing approach to determine how many clusters of variables (consumers) there are in the data at hand. In a first step, the cluster tendency is assessed, that is, we test whether or not there are different clusters of variables. In a second step, the number of clusters is determined by cluster validity tests at each level of the hierarchy.

2. Clustering of variables approach for segmenting a panel of consumers

The method of analysis aims at segmenting a panel of consumers around a small number of latent components that represent the main directions of preference in the panel (Vigneau et al., 2001). The scores of acceptability of p consumers for n products are arranged in a $(n \times p)$-matrix denoted by $\boldsymbol{X}$. We assume that there are K (to be defined) groups $(G_1, \ldots, G_K)$ of consumers such that the consumers in one group have the same liking pattern, while consumers of different groups differ in their preferences. The preferences of the whole panel can be summarized by K latent variables $\boldsymbol{c}^{(1)}, \boldsymbol{c}^{(2)}, \ldots, \boldsymbol{c}^{(K)}$, each of them representing one segment. If consumer j belongs to group G_k, his or her score for product i is given by

$$z_{ij} = c_i^{(k)} + \varepsilon_{ij}. \tag{1}$$

However, it is well known that the consumers use the scale differently. Each consumer has a shift a_j and a scaling factor b_j. Therefore, we do not observe the values z_{ij} but a transformation

$$x_{ij} = a_j + b_j(c_i^{(k)} + \varepsilon_{ij}). \tag{2}$$

As a first step of the analysis, the values in $\boldsymbol{X}$ are standardized to obtain the matrix $\boldsymbol{Z}$ whose columns have zero mean and unit variance.

For a given number K of groups, the problem consists in minimizing the quantity (Vigneau, Qannari, Sahmer, & Ladiray, in press):

$$Q = \frac{1}{n} \sum_{k=1}^{K} \sum_{j \in G_k} \|z_j - \boldsymbol{c}^{(k)}\|^2. \tag{3}$$

This quantity reflects the principle of the procedure which states that consumers in each segment G_k should be as close as possible to the latent variable $\boldsymbol{c}^{(k)}$. It is clear from criterion Q that the classification procedure is similar to the k-means algorithm. However, we consider herein this criterion within a framework of cluster analysis of variables that proved to be particularly appropriate to investigate the structure of preference data and to take account of external variables such as sensory or instrumental measurements (Vigneau & Qannari, 2002). We can remark that in previous papers Vigneau et al. (2001, 2002) used a slightly different criterion based on the covariance between the scores of preference and the latent variables. Nevertheless, the clustering procedure itself is the same. The procedure for the choice of the number of clusters described below can be used with both criteria. If we know which variable belongs to which group, Q is minimized by choosing $\boldsymbol{c}^{(k)} = \bar{\boldsymbol{z}}_k$, the average of the standardized variables belonging to group G_k.

In general, the appropriate number of clusters is not known. Moreover, if there is more than one cluster, we have to determine which variable belongs to which cluster. For these purposes, we perform the following procedure. First of all, a cluster tendency test is run. It aims at deciding whether or not there is more than one cluster. If there is just one cluster, no clustering is needed: the consumers have the same preferences for the products or do not express any consistent differences of preference among the products. If there is more than one group an agglomerative hierarchical clustering algorithm is performed and at each level of the hierarchy a cluster validity test is undergone. These cluster validity tests help in choosing the number of groups of consumers. Thereafter, we cut the hierarchical tree at the chosen level and, finally, perform a partitioning algorithm by considering the partition obtained from the hierarchical algorithm as a starting point. This last step usually leads to a better partition in that sense that criterion Q becomes smaller. In what follows, we emphasize on the test procedures: the cluster tendency test and the cluster validity test.

3. Cluster tendency test

This test makes it possible to assess whether there is more than one group of consumers. If there is only one cluster, all the preferences can be summarized by one latent variable $\boldsymbol{c}$, the average over all standardized variables. Let us consider the residual matrix $\boldsymbol{E}$,

$$\boldsymbol{E} = (\boldsymbol{z}_1, \boldsymbol{z}_2, \ldots \boldsymbol{z}_p) - (\boldsymbol{c}, \boldsymbol{c}, \ldots, \boldsymbol{c}), \tag{4}$$

which is formed by the difference between each variable and the latent variable $\boldsymbol{c}$.

If there are actually K ($K \geqslant 2$) clusters, $\boldsymbol{E}$ is structured because there is more information in the data than what is reflected by $\boldsymbol{c}$ and the model $\boldsymbol{z} = \boldsymbol{c} + \varepsilon$ does not fit the data. Contrariwise, if there is not any structure in $\boldsymbol{E}$, we have an indication that all information concerning the preferences is summarized by $\boldsymbol{c}$ or that there is no structure in the preference data themselves (no well identified clusters). From this standpoint, the hypotheses can be stated as follows:

H_0: $\boldsymbol{E}$ is formed by noise: $K = 1$ or no group at all.
H_1: $\boldsymbol{E}$ is structured: $K \geqslant 2$.

The test statistic T is based on the $r(r = \min(n-1, p-1))$ non-null eigenvalues of the variance covariance matrix of the residuals $(\frac{1}{n}\boldsymbol{E}^t\boldsymbol{E})$. With $(\lambda_1, \ldots, \lambda_m, \ldots, \lambda_r)$ referring to these eigenvalues, the statistic T is defined by:

$$T = \frac{\left(\prod_{m=1}^{r}\lambda_m\right)^{1/r}}{\frac{1}{r}\sum_{m=1}^{r}\lambda_m}. \tag{5}$$

This test statistic is usually used to test the sphericity of a matrix (Basilevsky, 1994) and represents the ratio of the geometric and the arithmetic mean of the eigenvalues. It is always comprised between zero and one. If $\boldsymbol{E}$ is just formed by noise, T will be close to one. In order to decide if the observed T value is likely to be observed when H_0 is true, it is necessary to draw the distribution of T under H_0. A Bootstrap procedure is used to reach an estimation of this distribution (Efron & Tibshirani, 1993). This procedure is based on the model under H_0 where the score of consumer j to product i can be written as:

$$x_{ij} = a_j + b_j(c_i + \varepsilon_{ij}) \tag{6}$$

where the scaling factor b_j and the translation parameter a_j of each consumer j are estimated by the mean and standard deviation of the column j of the data matrix $\boldsymbol{X}$. The latent variable $\boldsymbol{c}$ is estimated by the average of the standardized scores. In order to obtain simulated data matrices $\boldsymbol{X}^*$ which conform with the null hypothesis, we use, on the one hand, the estimations of the parameters a_j and b_j for each consumer computed above and, on the other hand, we use the values of the actual matrix $\boldsymbol{E}$ (denoted by $\boldsymbol{E}_{\mathrm{OBS}}$) to generate the values of the noise. For this purpose, all the observed residuals (entries of matrix $\boldsymbol{E}_{\mathrm{OBS}}$) are considered as one set without paying attention to the row or column to which they belong. In a subsequent stage, $n * p$ values ε^* are drawn with replacement from this set. These values are independent and come from the same distribution. According to Eq. (6), the Bootstrap values are obtained by:

$$x_{ij}^* = a_j + b_j\left(c_i + \varepsilon_{ij}^*\right) \tag{7}$$

which form the entries of the matrix $\boldsymbol{X}^*$. Thereafter, the same procedure as for the observed matrix $\boldsymbol{X}$ is undergone on $\boldsymbol{X}^*$. The associated matrix of residuals is given by:

$$\boldsymbol{E}^* = (\boldsymbol{z}_1^*, \boldsymbol{z}_2^*, \ldots, \boldsymbol{z}_p^*) - (\boldsymbol{c}^*, \boldsymbol{c}^*, \ldots, \boldsymbol{c}^*) \tag{8}$$

where $\boldsymbol{z}_1^*, \boldsymbol{z}_2^*, \ldots, \boldsymbol{z}_p^*$ are obtained from $\boldsymbol{X}^*$ after centering and standardization and $\boldsymbol{c}^*$ is the latent variable (average of the $\boldsymbol{z}_j^*$). Eventually, the value T^* of the test statistic T is calculated. The procedure is repeated B times. The principle of this Bootstrap procedure is depicted in Fig. 1.

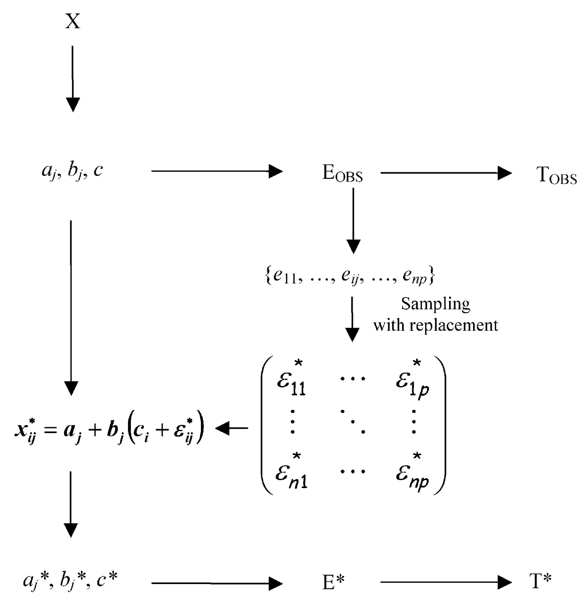

Fig. 1. Bootstrap procedure for the cluster tendency test.

The decision as whether to reject or not H_0 is based on evaluating the likelihood of obtaining a value equal or smaller than the actual value T_{OBS} of T (observed from the data) using the empirical distribution obtained from the Bootstrap procedure. More precisely, if less than α (for instance, $\alpha = 5\%$) of the simulated values T^* are smaller than T_{OBS}, we reject H_0 and decide that there is more than one cluster.

4. Cluster validity test

If the cluster tendency test leads to the rejection of the null hypothesis implying that there is more than one cluster, then we have to determine how many clusters there are. For this purpose, we consider the evolution of criterion Q in the course of the hierarchical algorithm. An equivalent expression of Q is given by:

$$Q = p - \sum_{k=1}^{K} p_k \mathrm{Var}(\bar{z}_k) \tag{9}$$

where $\mathrm{Var}(\bar{z}_k)$ denotes the variance of $\bar{z}_k$. At a given stage of the hierarchy, two clusters A and B (containing p_{A} and p_{B} variables, respectively) are merged into cluster $A \cup B$, and the criterion Q increases by:

$$\Delta Q = p_{\mathrm{A}}\mathrm{Var}(\bar{z}_{\mathrm{A}}) + p_{\mathrm{B}}\mathrm{Var}(\bar{z}_{\mathrm{B}}) - (p_{\mathrm{A}} + p_{\mathrm{B}})\mathrm{Var}(\bar{z}_{A\cup B}) \tag{10}$$

If consumers in clusters A and B have the same patterns of liking, then ΔQ will be close to zero. Contrariwise, if groups A and B are well separated groups, the variability of the centroid of the new group does not reflect the variability of the centroids of the two groups. In this case ΔQ is large. With g being the number of groups after merging, we have the hypotheses:

H_0: $\Delta Q = 0 \rightarrow$ merge without significant loss of information ($K \leqslant g$).
H_1: $\Delta Q > 0 \rightarrow$ do not merge ($K > g$).

It is relevant to relate ΔQ to the contribution of the groups A and B to criterion Q. Thus, we consider the test statistic

$$D = \frac{\Delta Q}{p_A \mathrm{Var}(\bar{z}_A) + p_B \mathrm{Var}(\bar{z}_B)}. \tag{11}$$

In order to set up a hypothesis testing strategy, we need to draw the distribution of D under H_0. For this purpose, $(p_A + p_B)$ variables are randomly selected from all p variables and divided at random into two groups of p_A and p_B variables respectively. Each of these two groups represent a random sample of all variables. Differences between the two groups are accidental. The corresponding value of D, which we denote D^*, represents a value which is likely to be observed under H_0. If less than α of the D^* are larger than the observed D, we state at the level α that ΔQ is larger than zero.

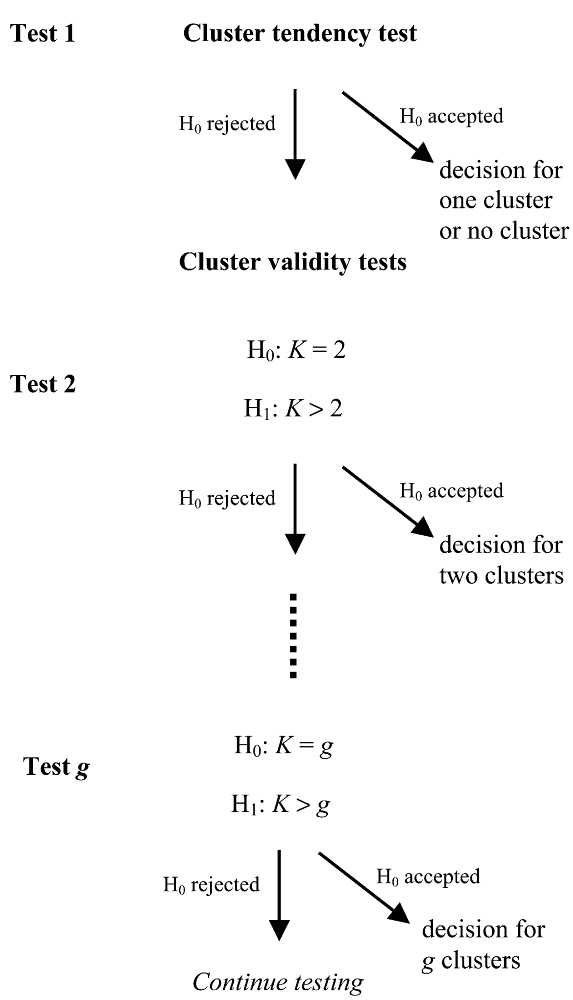

Fig. 2. Test procedure.

In order to decide the number of clusters, the cluster validity tests at different levels of the hierarchy are run in an ascending order. More precisely, after having decided (by means of the cluster tendency test) that there is more than one group, we test if there are more than two groups. If the hypothesis of two groups is rejected, we consider the following test to decide if there are three groups or more, and so forth. The procedure is depicted in Fig. 2.

5. Simulation study

In order to assess the efficiency of the test procedure, a simulation study was set up. This consists in simulating preference data with a known number of segments. Thereafter, the procedure described above is performed and we compared the number of clusters determined from this procedure with the actual number of clusters. The simulation of preference data sets is based on the model given by Eq. (2).

First of all, for each of the K groups, a "mean vector" $c^{(k)}$ is simulated. Its values are comprised between 1.1 and 8.9. This choice is in accordance with the aim to simulate individual preference scores comprised between 0 and 10. Some products (their number varied in the simulations) are considered as segmenting products. For these products, the mean score of each group is set at a minimal distance from the mean score of the other groups. In addition, the correlation between each pair of average scores $c^{(k)}$ and $c^{(k')}$ is not allowed to be larger than 0.5.

In a subsequent stage, the number of consumers per group is determined by a multinomial distribution $M(p, (1/K \ldots 1/K))$ where p denotes the number of consumers and K denotes the number of groups. The translation parameters a_j and the scaling factors b_j for each consumer are determined by random procedures. Details can be found in Callier (1996) and Sahmer (2003). The $\boldsymbol{c}^{(k)}$ are centered and the centered score of consumer j who belongs to group G_k for product i is therefore given by $x^*_{ij} = b_j(c_i^{(k)} - \bar{c}^{(k)} + \varepsilon_{ij})$. Finally, the group average and the translation parameter are added to obtain

$$x_{ij} = a_j + \bar{c}^{(k)} + b_j\left(c_i^{(k)} - \bar{c}^{(k)} + \varepsilon_{ij}\right). \tag{12}$$

The noise ε_{ij} was always simulated with a normal distribution with zero mean and variance σ^2. We wanted to assess the efficiency of the tests in an ideal situation, where the groups are well separated, as well as in a more realistic situation with a higher level of noise. For this purpose, the variance of the noise was allowed to take two different values. For the "ideal" situation, we chose $\sigma = 0.5$ and for the "realistic" situation, we chose $\sigma = 2$.

For both situations and by varying the number of groups between one and four, we simulated 800 data sets with 100 consumers and 800 data sets with 200 consumers evaluating 10 or 15 products. In the case of 10 products, the number of segmenting products varied between 1 and 5, and in the case of 15 products, there were 1–7 segmenting products. All the tests are performed at the level $\alpha = 5\%$.

The results for the *ideal situation* (*low noise level*, $\sigma = 0.5$) are shown in Table 1. It can be seen that the tests performed very well. When there was one group in the data, the test procedure led to correct decisions in all the cases except for 3% of the data sets with 15 products and 100 consumers. Similarly, for the data sets with 2 and 3 groups, the test procedure never failed. For the data sets with 4 groups, the results depend on the number of consumers and products. There is no problem for the data sets with 200 consumers whereas it turned out that 100 consumers are not sufficient for the identification of four groups. In the latter case, the results are better for data sets with 15 products (94% of correct decisions) than for data sets with only 10 products (83% of correct decisions).

As it can be expected, the *realistic situation* (*high noise level*, $\sigma = 2$) gave rise to difficulties which will be sketched below. The results are shown in Table 2. It can be seen that the cluster tendency test performs well even in the case of a relatively high noise level. Indeed this test indicated the existence of more than one group whenever this was the case. Contrariwise, when there was only one group the correct decision was taken in 94–97% of the cases. The percentages of wrong decisions when H_0 is true are compatible with the fixed level of $\alpha = 5\%$. Considering the wrong decisions (indication of more than one group when there was only one group), it turned out that the test procedure indicated a relatively large number of groups (between 7 and 21 groups). From this simulation study as well as from our experience in using this test procedure, we reached the conclusion that if the test indicates a large number of clusters then this should arouse suspicion that actually there is no structure in the data or that there is only one group but, for some reasons, the cluster tendency test failed to decide so.

In the case of two groups of consumers, the procedure performed better for data sets with 100 consumers (91% correct decisions for data sets with 10 products and 98% of correct decisions for data sets with 15 products) than for data sets with 200 consumers (70% of correct decisions for data sets with 10 products and 87% of correct decisions for data sets with 15 products). In the case of three groups of consumers, the results are very good for the data sets with 15 products (99% of correct decisions). However, for the data sets with 10 products, the procedure fails in 16% of the data sets with 100 consumers and in 19% of the data sets with 200 consumers. In the presence of four groups of consumers there are some problems for the data sets with 10 products and 100 consumers as only in 69% of the cases the correct number of groups is found. Good results (95% of correct

Table 1
Number of groups determined in an ideal situation ($\sigma = 0.5$)

Actual number of groups	Number of groups determined by the tests					Total (%)
	1 (%)	2 (%)	3 (%)	4 (%)	5 or more (%)	
10 products, 100 consumers						
1	**100**	–	–	–	–	100
2	–	**100**	–	–	–	100
3	–	–	**100**	–	–	100
4	–	3	14	**83**	–	100
15 products, 100 consumers						
1	**97**	–	–	–	3	100
2	–	**100**	–	–	–	100
3	–	–	**100**	–	–	100
4	–	1	5	**94**	–	100
10 products, 200 consumers						
1	**100**	–	–	–	–	100
2	–	**100**	–	–	–	100
3	–	–	**100**	–	–	100
4	–	–	1	**99**	–	100
15 products, 200 consumers						
1	**100**	–	–	–	–	100
2	–	**100**	–	–	–	100
3	–	–	**100**	–	–	100
4	–	–	1	**99**	–	100

Table 2
Number of groups determined in a realistic situation ($\sigma = 2$)

Actual number of groups	Number of groups determined by the tests					Total (%)
	1 (%)	2 (%)	3 (%)	4 (%)	5 or more (%)	
10 products, 100 consumers						
1	**97**	–	–	–	3	100
2	–	**91**	7	8	4	100
3	–	8	**84**	6	2	100
4	–	5	24	**69**	2	100
15 products, 100 consumers						
1	**94**	–	–	–	6	100
2	–	**98**	1	1	–	100
3	–	–	**99**	1	–	100
4	–	1	9	**90**	–	100
10 products, 200 consumers						
1	**97**	–	–	–	3	100
2	–	**70**	9	4	17	100
3	–	–	**81**	12	7	100
4	–	1	9	**89**	1	100
15 products, 200 consumers						
1	**95**	–	–	–	5	100
2	–	**87**	7	3	3	100
3	–	–	**99**	–	1	100
4	–	–	2	**95**	3	100

decisions) are found for the data sets with 15 products and 200 consumers.

6. Case study: Consumer preferences for coffee

We illustrate the test procedures using a data set from a study of the European sensory network (ESN, 1996). The study provides sensory data on 16 varieties of coffee and preference data on eight of them. The codes and names of the coffees used in the preference study are given in Table 3. Consumers of different European countries gave their scores of preference to these coffees. In this paper, we restrict ourselves to the data from French and Norwegian consumers.

6.1. French consumers

The French panel is composed of 80 consumers. The first two axes of the internal preference mapping performed on the preference scores are shown in Fig. 3. Along the first axis, the preferences of almost all consumers are in the direction of products 2, 8, 12 and 15. Very few consumers have opposite preferences. On the second axis, we distinguish consumers who prefer product 8 from consumers who prefer product 12.

The cluster tendency test is performed in order to determine whether the observed differences can be attributed to noise or whether they indicate the existence of several groups. The observed value for the test statistic is 0.96 which is close to 1. The empirical distribution of the values T^* is shown in Fig. 4. The p-value (which is the probability that T is smaller than T_{OBS} when H_0 is true) is equal to 0.76. We conclude that there is just one cluster in the French consumers panel. One latent variable, which is depicted in Fig. 5, can be used to summarize the preferences of the French consumers. Product 4 which is a Robusta is very much rejected.

The French consumers seem to prefer products 8, 12 and 15 which are medium roasted. We can remark in

Table 3
The coffee samples for the preference tests

Code	Origin	Type	Roasting
Product 2	Brazil	Arabica new processing method—Ouro Fino/MG	Light–medium
Product 4	Indonesia	Washed Java Robusta	Medium–dark
Product 8	Ethiopia	Unwashed Arabica Djimma 5	Medium
Product 10	Various	High grown washed Arabicas decaffeinated by CO_2 method	–
Product 11	Kenya	Washed Arabica	Light
Product 12	Kenya	Washed Arabica	Medium
Product 13	Kenya	Washed Arabica	Dark
Product 15	Brazil, Costa Rica, Mexico	Blend	Medium

passing that the level of noise of the one-segment French model is 1.8. This gives some credit to the simulation study undertaken above.

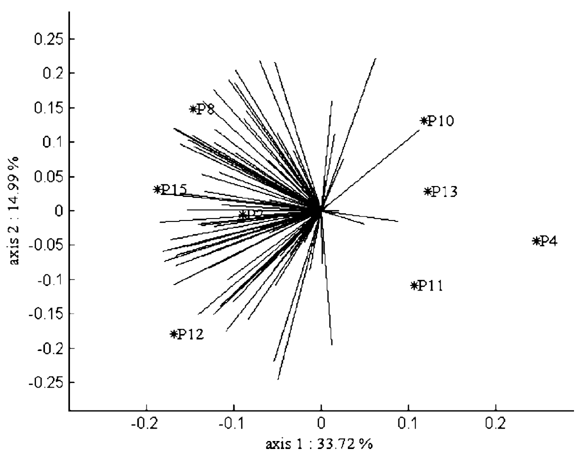

Fig. 3. Internal preference mapping of the French consumers.

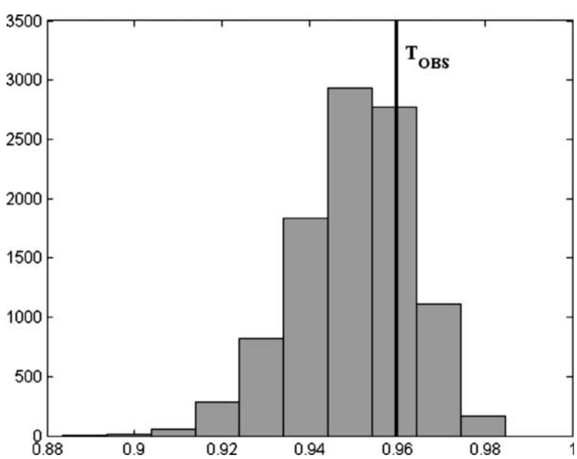

Fig. 4. French consumers: distribution of the cluster tendency test statistic T obtained by a Bootstrap procedure. The observed value is equal to 0.96.

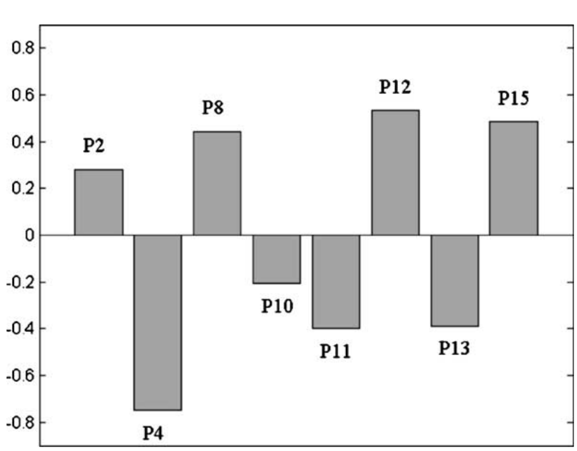

Fig. 5. Latent variable for the French consumers panel.

6.2. Norwegian consumers

The Norwegian panel consisted of 79 consumers. The internal preference mapping of their preference scores is shown in Fig. 6. Compared to the French panel, we observe more consumers with opposite directions of preference. It does not seem possible to summarize the preferences of all Norwegian consumers by a single latent variable.

The cluster tendency test corroborates this first impression. The histogram of the values of T^* (Fig. 7) shows that the observed T is small compared to the values generated under H_0. The value of T_{OBS} is equal to 0.91, and only 0.1% of the values of the test statistic are smaller than this value. This gives a strong evidence that there are two or more clusters. In the next step, we perform the hierarchical clustering algorithm. A plot depicting the evolution of criterion ΔQ (Fig. 8) can be used to assess the number of clusters. Obviously, there is more than one cluster. However, it is difficult to decide by simply observing the plot whether the jump when

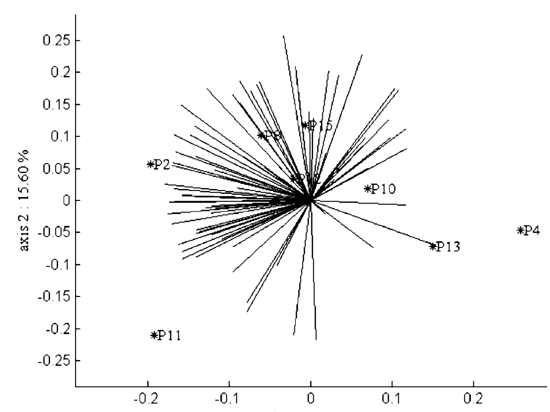

Fig. 6. Internal preference mapping for the Norwegian consumers.

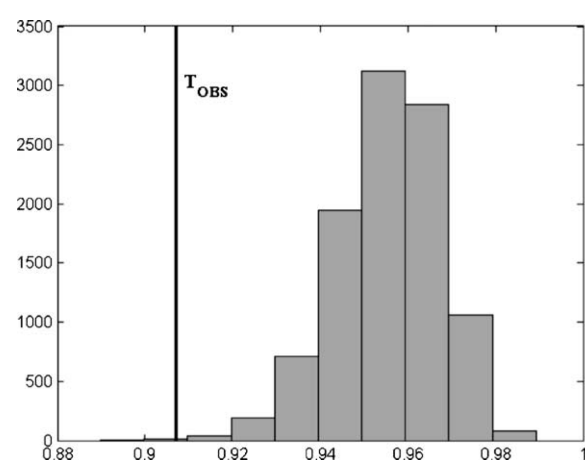

Fig. 7. Norwegian consumers: distribution of the cluster tendency test statistic T obtained by a Bootstrap procedure. The observed value is equal to 0.91.

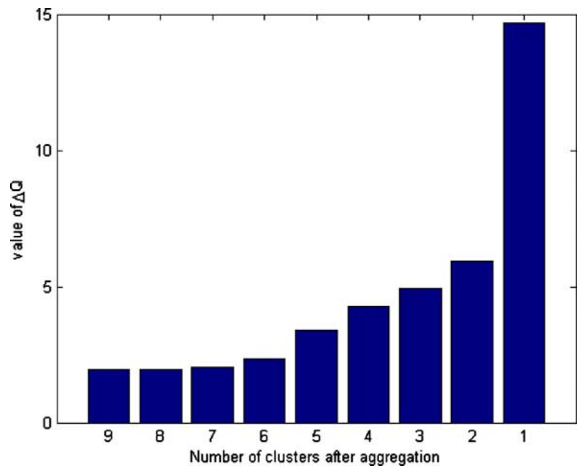

Fig. 8. Scree plot of ΔQ for the Norwegian consumers.

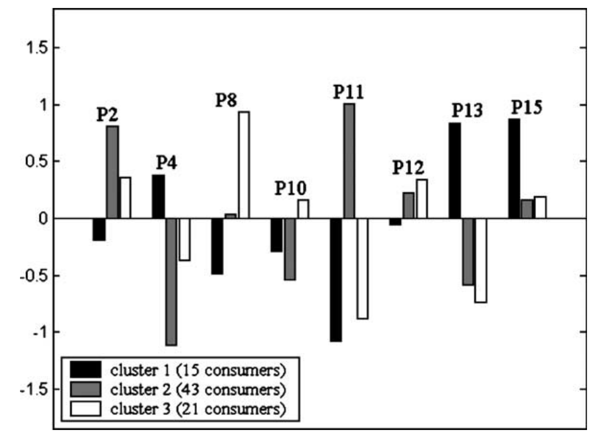

Fig. 9. Latent variables for the Norwegian consumers panel.

passing from three to two clusters is significant or not. The same uncertainty holds when passing from six to five clusters.

The results of the cluster validity tests which are shown in Table 4 indicate the existence of three groups. The null-hypothesis of two clusters against the alternative hypothesis of three or more clusters is rejected, whereas the null-hypothesis stating the existence of three clusters cannot be rejected. Therefore, we decide that the panel of Norwegian consumers can be segmented into three groups. In a subsequent stage, the tree obtained by the hierarchical clustering method is cut at the level of three clusters and the partitioning algorithm is performed using this partition as an initial solution. This led to three groups with 15, 43 and 21 consumers, respectively. Again, we can remark in passing that the level of the noise of the three-segment model is equal to 1.5 whereas it is equal to 1.9 for only one group and 1.6 for a model with two groups. The latent variables of the three groups are shown in Fig. 9. The liking of groups 2 and 3 is quite similar except for product 11 which is very likely not enough strong for consumers in group 3, whereas consumers in group 2 seem to like this product. Moreover, product 8 is more appreciated in group 3 than in group 2 whereas product 2 is more appreciated in group 2. Consumers belonging to group 1 appreciate products 13, 15 and 4 and reject product 11. These consumers seem to prefer coffees with higher bitter/burnt/roast flavour as it can be checked by referring to the sensory characterization of the products (ESN, 1996).

Table 4
The results of the cluster tendency test and the cluster validity tests for the Norwegian consumers

Test	H_0	H_1	p-Value
Cluster tendency test	1 cluster	2 or more clusters	0.0010
Cluster validity test	2 clusters	3 or more clusters	0.0004
Cluster validity test	3 clusters	4 or more clusters	0.1218

7. Conclusion

The clustering of variables approach is very appropriate to analyze preference data. This method makes it possible to form groups of consumers according to their preferences and to summarize the preferences in each group by a latent variable. Properties of the approach and graphical displays give hints to the practitioners as to the appropriate number of clusters. For instance, the evolution of the clustering criterion which can be read as a scree diagram is a useful tool in this respect. However, as the decision is not always clear, we recommend that these graphical displays should be complemented by a hypothesis testing framework. Simulations of several data sets with 10 or 15 products, 100 or 200 consumers and a known number of clusters (one to four) showed a good performance of the test procedure proposed herein. However, this performance depends upon the number of products and consumers. In particular, we found that the performance of the hypotheses tests was hindered when the number of products or consumers was not large enough to allow a good investigation of the structure of the preferences. Additional simulations with 4 products (not reported herein) led to the same conclusions. In further investigations, we will examine the connection between the number of products, the number of consumers and the performance of the procedure. Hopefully, we will be able to propose a modified version which not only works well for data sets with several products but also for data sets with a small number of products.

In practice, the preference mapping with inspection of the first axes and the cluster of variables approach appear to be complementary and give insight into the structure of consumers' preferences. Moreover, clustering of variables can be extended to the case where external data such as sensory data are used to explain the preferences (Vigneau & Qannari, 2002, 2003). Investigations are being undertaken in order to extend the hypothesis testing framework to this case.

Acknowledgement

We would like to thank the referees for their helpful remarks and for outlining objectives for further research.

References

Basilevsky, A. (1994). Statistical factor analysis and related methods. Theory and applications. New York: John Wiley & Sons, Inc.

Callier, P. (1996). La cartographie des préférences. *Son application en milieu industriel et son extension aux plans incomplets*. Ph.D. thesis: Doctorat en biostatistique, Université Montpellier II.

Duda, R. O., Hart, P. E., & Stork, D. G. (2001). Pattern classification (2nd Ed.). New York: John Wiley & Sons, Inc.

Efron, B., & Tibshirani, R. J. (1993). An introduction to the Bootstrap. New York: Chapman & Hall, Inc.

ESN (1996). *A European sensory and consumer study: A case study on coffee*. European Sensory Network.

Greenhoff, K., & MacFie, H. J. H. (1994). Preference mapping in practice. In H. J. H. MacFie & D. M. H. Thomson (Eds.), *Measurement of food preferences* (pp. 137–166). London: Blackie Academic & Professional.

Hardy, A. (1996). On the number of clusters. *Computational Statistics & Data Analysis, 23*, 83–96.

Helgesen, H., Solheim, R., & Næs, T. (1997). Consumer preference mapping of dry fermented lamb sausages. *Food Quality and Preference, 8*, 97–109.

Sahmer, K. (2003). *Classification des variables en présence de données manquantes: Application aux données de préférence*. Diplomarbeit Universität Dortmund.

SAS/STAT (1999). *User's guide, Version 8*. Cary, North Carolina: SAS Institute Inc.

Vigneau, E., & Qannari, E. M. (2002). Segmentation of consumers taking account of external data. A clustering of variables approach. *Food Quality and Preference, 13*, 515–521.

Vigneau, E., & Qannari, E. M. (2003). Clustering of variables around latent components. *Communications in Statistics—Simulation and Computation, 32*, 1131–1150.

Vigneau, E., Qannari, E. M., Punter, P. H., & Knoops, S. (2001). Segmentation of a panel of consumers using clustering of variables around latent directions of preference. *Food Quality and Preference, 12*, 259–363.

Vigneau, E., Qannari, E. M., Sahmer, K., & Ladiray, D. (in press). Classification de variables autour de composantes latentes. *Revue de Statistique Appliquée*.