Multiple Factor Analysis: main features and application to sensory data

Jérôme Pagès

AGROCAMPUS Rennes - 65 rue de St Brieuc - CS 84215 - F 35042 Rennes Cédex

pages@agrocampus-rennes.fr

Abstract. Data table in which a single set of individuals is described by several groups of variables are frequently encountered. In the factor analysis framework, taking into account different groups of variables in a unique analysis firstly raises the problem of balancing the different group. This problem being solved, beyond classical outputs from factor analysis, it is necessary to have at one's disposal specific tools in order to compare the structure upon individuals induced by the different groups of variables. That is the aim of Multiple Factor Analysis (MFA), factor analysis devoted to such data table. This paper presents the method, its main properties and an application to sensory data.

Keywords. Factor analysis, principal components analysis, canonical analysis

1. Data table: denotations, examples

Multiple Factor Analysis (MFA; Escofier & Pagès 1988-1998 ; Pagès 2002) deals with data table in which a set of individuals is described by several sets of variables. Within one set, variables must belong to the same type (quantitative or categorical) but, even for active ones, groups of variables can belong to different types. We focus the hereafter presentation on quantitative variables, with only some comments about qualitative ones.

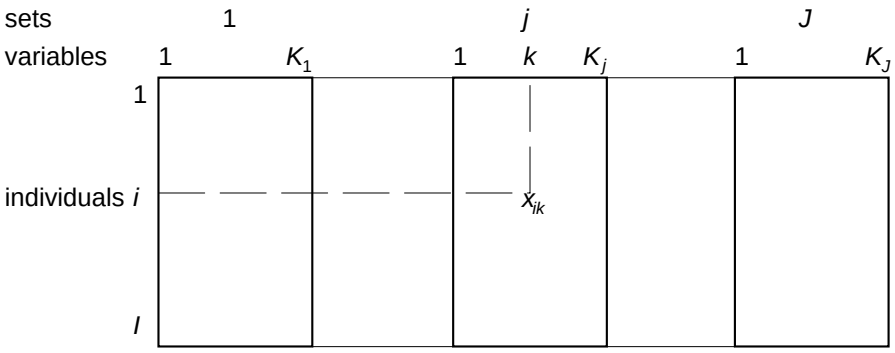

Figure 1. *Data table.*
x_{ik} : value of variable k for individual i. If k is a continue variable, x_{ik} is a real number ; if k is a categorical variable, x_{ik} is a number of category. The j^{th} set is denoted by j or K_j.

If we consider the whole table: individuals are denoted i (i =1, I); they constitute the cloud N_I lying in the K-dimensional space R^K ; the K variables constitute the cloud N_K lying in the I-dimensional space R^I.

In we consider the only (sub-)table j : individuals are denoted i^j (i =1, I); they constitute the cloud N_I^j lying in the K_j-dimensional space R^{Kj} ; the K_j variables constitute the cloud N_K^j lying in the I-dimensional space R^I.

Examples

Survey. An individual is a person ; a variable is a question. Questions are gathered according to the different themes of the questionnaire. Each theme defines one set.

Sensory analysis. An individual is a food product. A first set of variables includes sensory variables (sweetness, bitterness, etc.); a second one includes chemical variables (pH, glucose rate, etc.).

Ecology. An individual is a observation place. A first set of variables describes soil characteristics ; a second one describes flora.

Times series. Several individuals are observed at different dates. In such a case, there is often two ways of defining sets of variables : generally, each set gathers variables observed at one date ; but, when variables are the same from one date to the other, each set can gather the different dates for one variable.

2. General factor analysis : denotations and main relationships

General Factor Analysis (GFA) is here understood according to Lebart et al (1997). Main features and denotations of GFA can be summed up by 6 points.

1) Given a table X having I rows and K columns, two clouds are considered : the one of rows (N_I lying in R^K) ; the one of columns (N_K lying in R^I) ; to simplify this short synthesis, weights of individuals and weights of variables are supposed uniform.

2) The maximum inertia directions of N_I and N_K are highlighted : let u_s (resp. z_s) a unit vector along the rank-s principal direction of N_I (resp. N_K) in R^K (resp. R^I). These vectors satisfy (λ_s being the rank s eigenvalue of $X'X$):

$$X'Xu_s = \lambda_s u_s \qquad \|u_s\| = 1 \qquad\qquad XX'z_s = \lambda_s z_s \qquad \|z_s\| = 1$$

3) N_I and N_K are projected onto their maximum inertia directions u_s and z_s; co-ordinates of N_I (resp. N_K) onto axis s constitute the I-factor of rank s (resp. K-factor) denoted F_s (resp. G_s):

$$F_s = Xu_s \qquad\qquad G_s = X'z_s$$

In PCA, F_s is named *principal component*.

4) The inertia directions of N_I and N_K are related (that is named *duality*) by:

$$z_s = \frac{F_s}{\|F_s\|} = \frac{F_s}{\sqrt{\lambda_s}} \qquad u_s = \frac{G_s}{\|G_s\|} = \frac{G_s}{\sqrt{\lambda_s}}$$

z_s is often named *standardised I-factor* (in PCA: *standardised principal component*).

5) The projection of row *i* (resp. column *k*) onto rank *s* axis in R^K (resp. R^I) can be calculated from the co-ordinates of N_K (resp. N_I) onto rank *s* axis in R^I (resp. R^K) by the way of the transition formulae:

$$F_s = \frac{1}{\sqrt{\lambda_s}} XG_s \quad G_s = \frac{1}{\sqrt{\lambda_s}} X'F_s$$

$$F_s(i) = \frac{1}{\sqrt{\lambda_s}} \sum_k x_{ik} G_s(k) \quad G_s(k) = \frac{1}{\sqrt{\lambda_s}} \sum_i \frac{1}{I} x_{ik} F_s(i)$$

Since Benzecri (1973), these formulae are especially known in the case of correspondence analysis (often under the name of barycentric properties); they are seldom quoted in the case of PCA but are implicit during the interpretation.

6) Principal Components Analysis (PCA), Correspondence analysis (CA) and Multiple Correspondences Analysis (MCA) can be viewed as particular cases of GFA.

3. Usual methods or MFA ?

In the context of factor analysis (PCA or MCA according to the type of variables), to study the kind of data table described figure 1, usual practice consists in introducing only one set of variables as active, the others being illustrative. This ensures homogeneity of active variables, characteristic which goes hand in hand with a clear two-steps problematic: **1)** looking for main factors describing data variability according to one theme (the one corresponding to the active variables) **2)** relating each of the illustrative variables to the previous factors.

This basic methodology is excellent. But it should be noted that, in this strategy, the only multidimensional structure really handled is the one of the active variables; the illustrative variables intervene independently one to the other. According to this point of view, one can want to introduce several sets of variables simultaneously as active elements, in order to take them simultaneously into account in the definition of distance between individuals. Introducing, as active elements, several sets of variables (or, according to an other point of view, distinguishing sets among active variables) firstly implies to balance these sets and, secondly, enriches problematic, that is to say induces new questions about data.

MFA brings solutions to these problems in a way described hereafter.

4. Balancing the sets of variables

If all the sets of variables are introduced, as active elements, without balancing their influence, a single set can contribute quite by itself to the construction of the first axes. In such a case, the user want to analyse all the sets and, in fact, analyses only one of them.

Thus, the global analysis, in which several sets of variables are simultaneously introduced as active ones, requires balancing the influences of these sets. The influence of one set j derives from its structure, in the sense of its inertia distribution (of the two clouds N_I^j and N_K^j it induces) in the different space directions. For example, if a set presents a high inertia in one direction, this direction will strongly influence the first axis of the global analysis.

This suggests to normalise the highest axial inertia of each set. Technically, it is done by weighting each variable of the set j by $1/\lambda_1^j$, denoting λ_1^j the first eigenvalue of factor analysis applied to set j.

This weighting can be easily interpreted : considering the two clouds (N_I^j et N_K^j) induced by the set j of variables, MFA weighting normalises each of these two clouds by making its highest axial inertia equal to 1. This weighting does not balance total inertia of the different sets. Thus, a set having a high dimensionality will have a high global influence in that sense that this set will contribute to numerous axes. But such a set has no reason to contribute particularly to the first axes. Correlatively, a one-dimensional set can strongly contribute to only one axis, but this axis can be the first one.

5. MFA as a general factor analysis

The core of MFA is a general factor analysis applied to all active sets of variables (global analysis). MFA works with continuous variables as principal component analysis does, the variables being weighted ; MFA works with categorical variables as multiple correspondences analysis does, the variables being weighted. Weighting, which balances highest axial inertia of sets, allows to work simultaneously with continuous and categorical variables as active elements. Likewise, it is possible to introduce simultaneously as active elements, set(s) of standardised variables and set(s) of un-standardised variables.

The aim is to bring out main factors of data variability, individuals being described, in a balanced manner, by several sets of variables (those introduced as active).

According to this point of view, MFA provides classical outputs of general factor analysis, that is to say, for each axis :

- Co-ordinates, contributions and squared cosines of individuals ;
- Correlation coefficient between factors and continuous variables ;

- For each category, co-ordinate (and test-value in the sense of Spad software 2002) of the centre of gravity of individuals belonging to this category.

Remark about categories inside MFA.

In MFA, categories are represented by exact centres of gravity (as in PCA and differently from MCA). For each category, we can calculate the inertia of the corresponding centre of gravity in per cent of the total inertia ; this ratio is named contribution. It is proportional to the contribution usually defined in MCA for the active variables and possesses the following property : its sum, for all the categories of the variable *k* and for axis *s*, equals to the correlation ratio between the variable *k* and the factor *s*. This ratio can be calculated for all categorical variables (active and illustrative).

6. Superimposed representation of the *J* clouds of individuals.

Questions

To each set *j*, we associate the cloud N_I^j of individuals lying in the space R^{K_j}. This cloud, named " partial ", is the one analysed in the factor analysis restricted to set *j*; it contains " partials " individuals, denoted i^j (individual *i* according to the set *j*).

A classical question is : are there structures common to these clouds N_I^j j=1, J ? That is to say : are there some resemblances, from one cloud to the other, among distances between homologous points ?

To answer these questions, we are looking for a superimposed representation of clouds N_I^j which :

- fits well each of the clouds N_I^j ;
- highlights the resemblances between the different N_I^j, that is to say displays homologous points as close one to the other as possible (homologous i.e. referring to the same individual).

This problematic is similar to the one of Generalised Procustes Analysis (GPA ; Gower, 1975). The way used by MFA to obtain such a superimposed representation is described herafter.

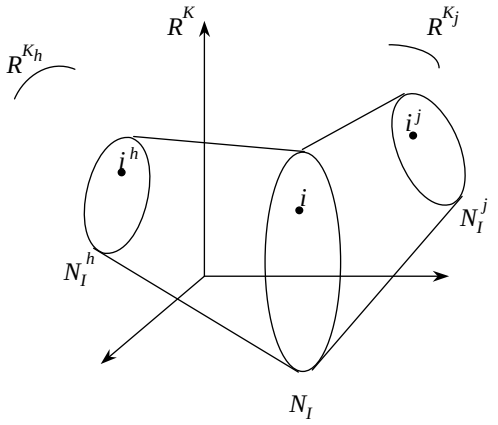

Figure 2. *Representation of the J partial clouds N_I^j in the space R^K*
i : individual described by all the variables ; i^j : individual described by the variables of the group K_j. N_I^j can be viewed as a projection of N_I onto R^{K_j}, the subspace of R^K spanned by the variables of K_j.

Principle

R^K can be viewed as the direct sum of the R^{K_j} : $R^K = \oplus R^{K_j}$. Using this property, it is possible to put all the N_I^j in the same space (cf. figure 2). In MFA, the clouds N_I^j are projected upon the axes of the global analysis, as illustrative elements (cf. figure 3). In fact these elements are not exactly illustrative since their data contribute to axes construction ; moreover, this representation is possible only for the clouds N_I^j corresponding to active sets.

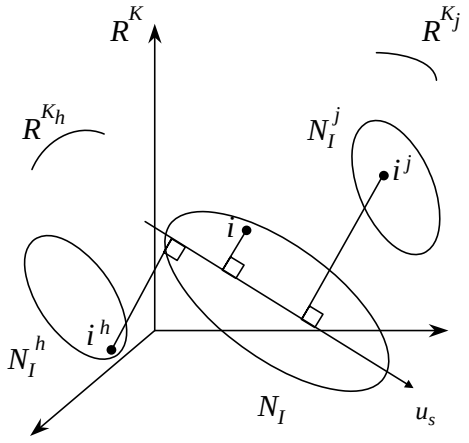

Figure 3. *Principle of the superimposed representations provided by MFA*
Each partial cloud N_I^j is projected onto main axes of the mean cloud N_I

The co-ordinate of i^j along axis s is denoted : $F_s(i^j)$. These co-ordinates can be gathered in the vector F_s^j such as : $F_s^j(i) = F_s(i^j)$.

Restricted transition formula

The superimposed representation of MFA is not optimal, in the sense that the F_s^j do not satisfy a global criterion. But it posseses the very useful following property.

It can be easily shown that the co-ordinate $F_s(i^j)$ can be calculated from the coordinates of the variables $G_s(k)$, $k \in K_j$, by the way of the following relationship :

$$F_s^j(i) = F_s(i^j) = \frac{1}{\sqrt{\lambda_s}} \frac{1}{\sqrt{\lambda_1^j}} \sum_{k \in K_j} x_{ik} G_s(k)$$

We recognise here the usual transition formula (see § 2) but restricted to the variables of the group K_j.

Ratio to measure the global similarity between axial representations of the clouds N_I^j

When the different sets induce similar structures on individuals, homologous points $\{i^j, j=1, J\}$ are close one to the other. This global property is measured, per axis, through the ratio described below.

Let's consider all the points of all the clouds N_I^j ($j = 1, J$) and a partition of these $I \times J$ points in I classes, such as the J homologous points $\{i^j, j=1, J\}$ corresponding to the same individual i belong to the same class. When axis s brings out a structure common to the different sets of variables, the homologous points i^j, corresponding to the same individual i, are close one to the other and this partition has a low within-inertia (along axis s). The ratio (*between-inertia)* / (*total-inertia*) can be calculated for each axis. This ratio is close to 1 when the axis represents a structure common to the different sets.

Be careful: **1)** this ratio does not decrease with axis rank order since it is not the criterion optimised by MFA; **2)** it cannot be summerised for several axes.

Detailed examination of axial representations of N_I^j

The distance between each point i^j and the corresponding mean point i gives an idea about the position of i (among I) in the cloud N_I^j compared to the one in the cloud N_I. These distances can be examined visually, or by selecting the projections of i^j having the highest contributions to the within inertia. This allows to detect :

- Individuals having their homologous points close one to the other (low within inertia) ; they illustrate the common structure represented by axis s ;
- Individuals having their homologous points far one from the other (high within inertia) ; they constitute exceptions to the common structure represented by axis s.

Case of categories

In factor analysis, when the individuals are numerous, it is the case in surveys for example, they aren't studied directly but by means of categorical variables, active and/or illustrative (students, old people, etc.). Thus :

- In PCA, each category k is represented by the centre of gravity of individuals that belong to this category k ;
- In MCA, the co-ordinates of points representing the categories are only proportional to those of the corresponding centres of gravity (application of the correspondence analysis centroid property to indicator matrix)

In MFA, the categories are represented by their associated centres of gravity. This allows to work with categories as with individuals. Particularly, each category (e.g. *student* in a survey) can be represented by a global point (centre of gravity of the students) and by one partial point for each set of variables (e.g. the centre of gravity of partial points representing the *students according to set j*).

7. MFA as a multicanonical analysis

Principle of multicanonical analysis

In the simultaneous study of several sets of variables, the main question is : are there factors common to the different sets of variables ?

In the simple case of two sets, this question refers to canonical analysis (Hotelling 1936). When there are more than two groups, the reference method is multicanonical analysis. There are several multicanonical analyses. The most used is the one of CARROLL (1968), that works in two steps :

- Looking for a sequence of variables $\{z_s\ ;\ s=1,S\}$ (named general variables), normalised and not correlated one to the other, related to the sets of variables as strongly as possible ;
- For each general variable z_s and for each set j, looking for the linear combination of the variables of set j (combinations named canonical variables) related to general variable z_s as strongly as possible.

MFA can be interpreted in this framework.

Measure of relationship between one variable and a group of variable

To do this, it is necessary to firstly define a measure of relationship (denoted $\mathcal{L}_g$) between one continuous variable z and a set of variables $K_j = \{v_k,\ k=1,K_j\}$

$$\mathcal{L}_g(z,\ \{v_k,\ k=1,K_j\}) = \mathcal{L}_g(z,\ K_j) = \text{inertia of all variables } v_k \text{ projected upon } z.$$

When the v_k are reduced continuous variables, weighted by m_k :

$$\mathcal{L}_g(z,\ K_j) = \sum_k m_k[r(z,v_k)]^2$$

This measure is implicitly used in PCA : the first principal component is the linear combination (of variables) the most related to the whole set of variables (it maximises $\mathcal{L}_g(z,\ K)$).

If $\mathcal{L}_g(z,\ K_j)=0$, variable z is not correlated to any variable of the set K_j.

Due to MFA weighting, $\mathcal{L}_g(z, K_j) \leq 1$; $\mathcal{L}_g(z, K_j)=1$ when z is the first principal component of K_j.

General variables

The first factor of MFA (as defined in §5) maximises projected inertia of all the sets of variables, that is :

$$\sum_j \mathcal{L}_g(z, K_j) \text{ maximum}$$

In that sense, MFA factors (denoted F_s §6) can be considered as general variables of a multicanonical analysis (in CARROLL's method, relationship between one variable and one set of variables is measured by means of multiple correlation coefficient).

Canonical variables

The coherence between the multicanonical point of view and the superimposed representation point of view suggests to use the previously defined F_s^j as canonical variables. It can be shown (Pagès & Tenenhaus, 2001) that F_s^j is the first component in the PLS regression between the general variable z_s and the data table X_j. This result reinforces the superimosed representation: it induces that the F_s^j j=1,J must be correlated one to the other since each F_s^j expresses the same structure F_s in the group K_j.

Canonical correlation coefficients in MFA

In MFA, factors of global analysis (denoted F_s) are the common factors and factors of partial points (denoted F_s^j) represent common factors in each set j of variables. In order to judge if factors of global analysis really are common to the different sets, it is possible to calculate, for each set j and each factor s, the correlation coefficient between general variable F_s and canonical variable F_s^j. If this coefficient (named canonical correlation coefficient and always positive) is high, then the structure brought out by variable F_s does " exist " in the set j. If not, it does not. The synthesis of all these correlation coefficients shows factors common to all the sets, common to some sets, specific to only one set.

8. Global study of sets of variables

It is often interesting to globally study the sets of variables, the question being : do these sets define similar structures upon individuals (i.e. similar distances between individuals from one cloud N_i^j to the other) ? We find again the problem of superimposed representations and the one of common factors but now the investigation about the similarities between sets is global.

Here, we are looking for a display in which each set of variables is represented by a unique point. In such a display, two sets must be close one another if they induce similar structures on the individuals.

To each set of variables K_j, we associate the $I \times I$ matrix W_j of scalar products between individuals ($W_j = X_j X_j'$). Each scalar product matrix W_j can be represented by one point in the I^2-dimensional Euclidean space (denoted R^{I^2}). Thus, in this space, one set is represented by one point: the J points constitute the set cloud, denoted N_J. In this cloud N_J, the distance between two points W_j and W_l decreases as the similarity between the structures (defined upon individuals) induced by the sets K_j and K_l increases. For this reason, it is interesting to get a representation of the cloud N_J.

The Statis method (Lavit, 1988) is based on such a representation, obtained by projecting N_J onto its main inertia directions. But these directions cannot be interpreted (they are linear combinations of couples of individuals) (Pagès 1996).

The representation provided by MFA is obtained by projecting N_J upon vectors (in R^{I^2}) induced by I-factors of global analysis (one factor may be considered as a set including a single variable; it is possible to associate to this set a scalar product matrix and thus a vector in R^{I^2}).

The normalised factor of rank s in R^K, previously denoted z_s, induces $w_s = z_s z_s'$ in R^{I^2}. Some properties of z_s induce corresponding properties for w_s :

$$z_s' z_t = 0 \Rightarrow \langle w_s, w_t \rangle = 0$$
$$\|z_s\| = 1 \Rightarrow \|w_s\| = 1$$

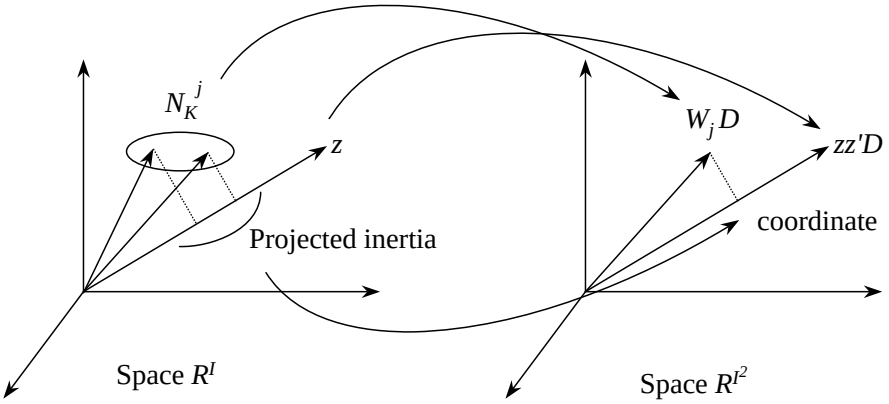

Figure 4. *The representation of the groups and its links with the one of variables*

The main interest of this projection space is that its axes (upon which N_J is projected) are interpretable and, above all, possess the same interpretation that axes of global analysis (in the same manner, due to factor analysis duality, axis of rank s upon which individuals are projected and axis of rank order s upon which variables are projected possess the same interpretation).

This representation has the following property: it ca be shown (Escofir & Pagès 1998 p 167) that co-ordinate of set j upon axe of rank s is equal to $\mathcal{L}_g(z_s, K_j)$. Thus:

- Set co-ordinates are always comprised between 0 and 1;
- A small distance between two sets along axe *s* means that these two sets include the structure expressed by factor *s* each one with the same intensity. In other words, set representations shows which ones are similar (or different) from the point of view of global analysis factors.

This representation has been introduced as an aid to the interpretation of representations of individuals and variables. But it possesses its own optimality: axes upon which N_J is projected, taking into account the usual orthogonality constraint but also the constraint to be of order 1 (i.e. induced by one direction in R^I), make the sum (and not the sum of squares) of co-ordinates maximum (for axis *s*, this sum is equal to the s^{th} eigenvalue of the global analysis). Thus, from the R^{I^2} point of view, the contribution of the set *j* to axis *s* is equal to the set *j* co-ordinate divided by the sum of co-ordinates (this contribution is equal to the sum of contributions to axis *s*, in R^I, of variables belonging to the set *j*).

The set study can be completed by squared cosines computed in R^{I^2}.

9. Relationship between global analysis and separated analysis of each set.

It is always important and interesting to relate MFA results to those of separated analysis of each set. To do this, *I*-factors of separated analysis (called " partial " factors) are projected as illustrative quantitative variables.

It is equivalent to perform MFA from variables or from all the partial factors (each one being, within its set, " pre-weighted " proportionally to its eigenvalue). Thus:

- For each partial factor, the ratio between its projected inertia along axis *s* and the axis *s* eigenvalue may be interpreted, in case of active sets, actually as a contribution to MFA axis *s*;
- MFA may be considered as a method providing an optimal representation of separated analyses axes.

This last point is useful for applications: MFA is a convenient tool in order to compare several factor analyses having the same individuals. Example: in order to compare, for the same variables, normalised PCA an un-normalised PCA, variables must appear twice in data base, firstly within a normalised set and secondly within a un–normalised set; MFA is there an optimal tool in order to display factors of the two analyses (a direct PCA is not desirable because set weighting is necessary).

10. The orange juice example

10.1 Data

Six pure orange juices (P1 to P6) were selected from the main brands on the French market. These juices were pasteurised in two ways: thus, three of them must be stored in refrigerated conditions (R) while the others can be stored at ambient temperature (A). Here is the list of the six orange juices: Pampryl at ambient temperature (P1), Tropicana at ambient temperature (P2), refrigerated Fruivita (P3), Joker at ambient temperature (P4), refrigerated Tropicana (P5), refrigerated Pampryl (P6).

Ninety-six students in food science, both trained to evaluate foodstuffs and consumers of orange juice, described each of these six products on the basis of seven attributes: intensity and typical nature of its smell, intensity of the taste, pulp content, sweetness, sourness and bitterness.

The serving order design was a juxtaposition of Latin squares balanced for carry-over effects (MacFie, Bratchell, Greenhoff and Vallis, 1989).

In addition to the sensory investigation, chemical measurements (pH, citric acid, overall acidity, saccharose, fructose, glucose, vitamin C and sweetening power - defined as : saccharose + .6 glucose + 1.4 fructose) were carried out.

The data are gathered in a table using the format shown in figure 5. The complete data table is in the appendix.

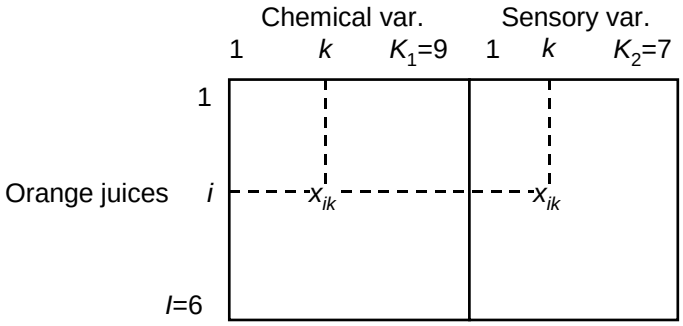

Figure 5. *The orange juices data table*
For juice i : x_{ik} is the panel average of the sensory variable k or the chemical measurement k

The ouputs described below come from Spad 2002 software.

10.2 Separate analyses

Table 1 gathers inertia from separate and global analysis. The first eigenvalue of the separate PCA of chemical variables is slightly higher than the one of PCA of sensory variables. The balancing is useful to avoid the domination of chemical variables in the construction of the first axis.

The sequence of eigenvalues is similar from one analysis to the other: the two groups of variables have a strong first direction of inertia. Moreover, the homologous factors of the two separate PCA are correlated one another (table 2).

This data set is interesting from a methodological point of view: the similarity between the two groups of variables justifies their simultaneous analysis; the differences between the two groups are sufficiently important to justify the use of a specific method to highlight common and specific features.

	PCA chemical var.		PCA sensory var.		MFA	
Axes	Eigenvalue	%	Eigenvalue	%	Eigenvalue	%
1	6.2135	69.04	4.7437	67.77	1.7907	61.24
2	1.4102	15.67	1.3333	19.05	0.4764	16.29
3	1.0457	11.62	0.8198	11.71	0.2938	10.05
4	0.3173	03.53	0.0840	01.20	0.2009	6.87
5	0.0133	00.15	0.0192	00.27	0.1623	5.55

Table 1*: Eigenvalues (= inertia) from separate PCA and from MFA*

		PCA chemical var.	
		F1	F2
PCA sensory variables	F1	-0.78	-0.25
	F2	0.08	-0.74

Table 2*: Correlations between separate PCA factors*

10.3 Representation of individuals (= products) and variables (fig. 6 and 7)

This MFA builds a product space starting from factors common to the sensory and instrumental data, in which the influences of these two groups of variables are balanced. These MFA representations (of products and variables) can be read like those from a PCA: the co-ordinates of a product are its values for the common factors; the co-ordinates of a variable are its correlations with these factors.

The first axis is highly correlated to variables belonging to the two groups ; that was awaited due to the group balancing. It opposes the juices 1, 4 and 6 to the juices 2, 3 and 5.

According to the usual transition formulae, the juices 1, 4 and 6 have a high level of acidity (and a low pH), and a high [(glucose + fructose)/saccharose] ratio; they were perceived as sour, bitter and not very sweet. Symmetrically, the juices 2, 3 and 5 have opposite characteristics: a low level of acidity (and a high pH), and a low [(glucose + fructose)/saccharose] ratio; they were perceived as being not sour or bitter, but sweet.

The juices 2, 3 and 5 come from Florida; the first axis is linked to geographic origin.

The second bissector is more interesting than the second axis. This bissector correspond to *pulpy*. It opposes the refrigerated juices to the others.

Let us also point out:

- the opposition between fructose and glucose on the one hand and saccharose and pH on the other hand, connected with the hydrolysis of saccharose, facilitated in an acid medium;
- the correlation between acidity, pH and sourness;
- the absence of correlation between sweetening power and sweetness : a high level of sweetness is associated with a low level of sourness (this refers to the concept of gustatory balance). Thus the strong correlation between saccharose and sweetness is not due to the direct influence of saccharose but to a high pH.

10.4 Factors from separate analyses (Fig. 8)

Factors from separate analyses can be represented by the way of their correlations with factors of MFA. Figure 8 shows that the first factor of MFA is highly correlated with the first factor of each separate analysis. These factors of separate analyses are not so highly correlated one another (cf. table 2) but, in this analysis, the first MFA factor, being a kind of compromise between them, is highly correlated to the two. The same observation can be made for the second factor. Thus, the first MFA map is roughly similar to the one of each separate analysis (figure 8 gives an idea of the slight rotation to obtain one map from an other).

10.5 Superimposed representations (Fig. 9)

Figure 9 derives from figure 6 by adding partial points (C and S). Whatever the set of variables considered, the first axis opposes the products 1, 4 and 6 to the product 2, 3 and 5. This is a other way to highlight common factor.

The resemblance between the two partial clouds can be globally evaluated by two series of measures (cf. table 3 and 4). The two first factors from MFA can be considered as factors common to the two groups of variables.

	Factor from MFA				
	F1	F2	F3	F4	F5
G1 : chemical var.	0.95	0.88	0.49	0.46	0.82
G2 : sensory var.	0.95	0.90	0.45	0.90	0.26

Table 3. *Canonical correlation coefficient*
At the intersection of row j and column s: r(F_s^j , F_s) (cf. § 7)

	F1	F2	F3	F4	F5
ratio	0.90	0.80	0.22	0.51	0.35

Table 4. *Ratio [(between-inertia) / (total-inertia)]; cf. § 6*

This representation allows a precise comparison of the clouds N_I^j. Thus, the figure 9 suggests that the product 5 is highly characteristic from a sensory point of view though it is not the case from a chemical point of view; conversely, product 2 is characteristic from the chemical point of view but that does not induce a particular sensory evaluation.

This can be directly verified in the data (cf. appendix), preferably in the standardised data that can be compared from one variable to the other. Thus, the standardised data table shows that product 5 has absolute values higher for the sensory attributes than for the chemical variables. It is the reverse for he product 2.

In these data, in which the two first factors from MFA are highly correlated to the corresponding ones from each separate analysis, the superimposed representation gives a good idea of the representation from separate analysis. This can be illustrated by the comparison between the representation of partial individuals in MFA and the representation of individuals from separate PCA (cf. Fig. 10). Thus, for exemple, the opposition between the products P2 and P4 is much bigger from a chemical point of view than for a sensory point of view.

10.6 Representation of categories (Fig. 11)

In this data table, the individuals are very few : here, the interest of the representation of categories is mostly technical. But, when the individuals are numerous, as in a survey for example, this representation is essential. Each category lies in the center of gravity of the individuals which posses this category. This is applied to mean and partial point.

In Figure 11, the mean points show immediatly that the factor 1 corresponds to origin (Florida / elswhere) and that the second bissector corresponds to way of storing (refrigerated or ambient). The partial points shows that the opposition of the two origins along axis 1 is equally clear from the two points of view (chemical/sensory). Along the second bissector, the opposition of the two ways of storing mostly appears from the sensory point of view (pulpyness).

10.7 Representation of groups of variables (Fig. 12)

These data being composed of two groups only, the representation of groups as points of R^{I^2} (cf. § 8) has not practical interest. To enrich the technical comments, here a third group is added (as a supplementary one): it derives from the chemical group, in which we have removed the variables *pH2* (because it is quite equivalent to *pH1*) and *vitamin C* (because it is not related to sensory evaluation). Thus we get an empiric idea about the stability of the chemical group.

The three groups are strongly related to Factor 1 (in the sense of $\mathcal{L}_q$ measurement ; cf. § 7): this factor correspond to a direction having a strong inertia for the three clouds of variables (in other words, many variables of each groups are related to this factor).

Regarding the first two MFA factors, these three groups are very similar: the clouds of individuals they induce (previously denoted N_I^j) are very similar. In particular, removing *pH2* and *vitamin C* has had a weak influence.

11. Conclusion

MFA allows to take into account several sets of variables as active elements in a unique factor analysis. Its main features are:

- the balancing of the sets of variables;
- outputs specific of the partition of the variables in differents sets ; mainly 1) the superimposed representations of individuals and of categories 2) the groups representation.

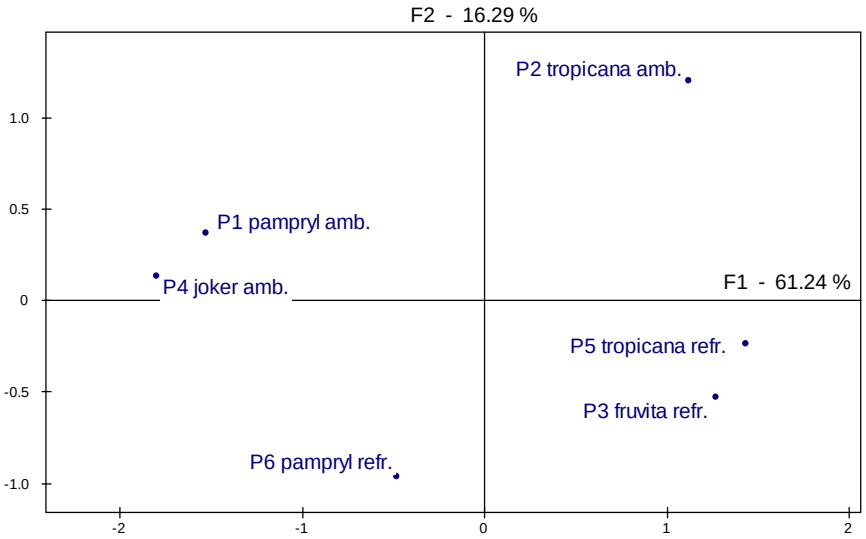

Figure 6. *First factorial map from MFA : mean individuals*
Refr. : refrigerated ; amb. : stored at ambient temperature. Tropicana and Fruvita come from Florida

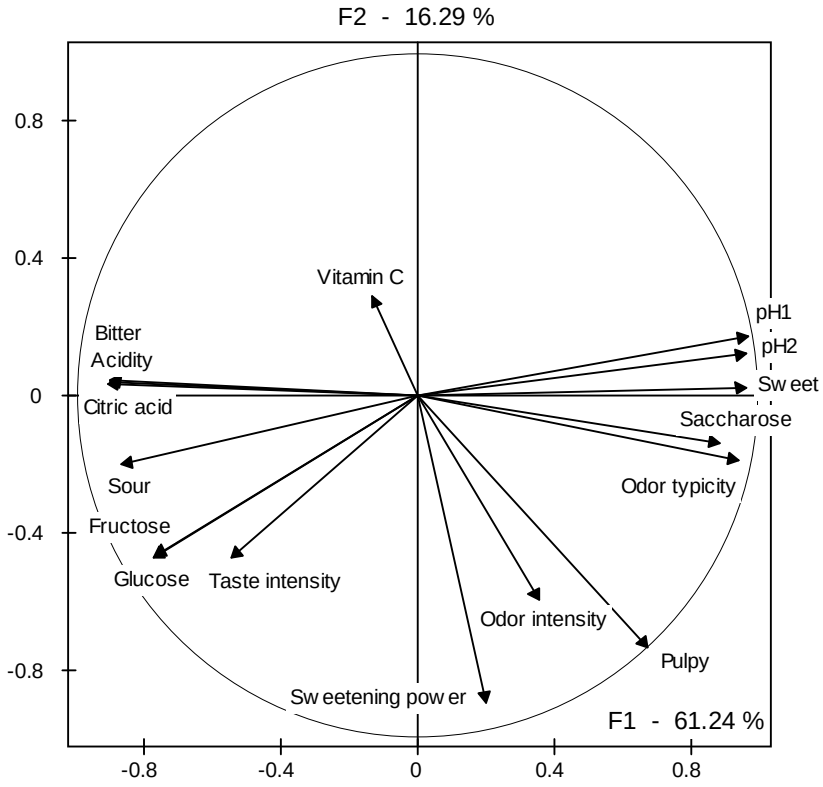

Figure 7. *First factorial map from MFA : chemical and sensory variables*

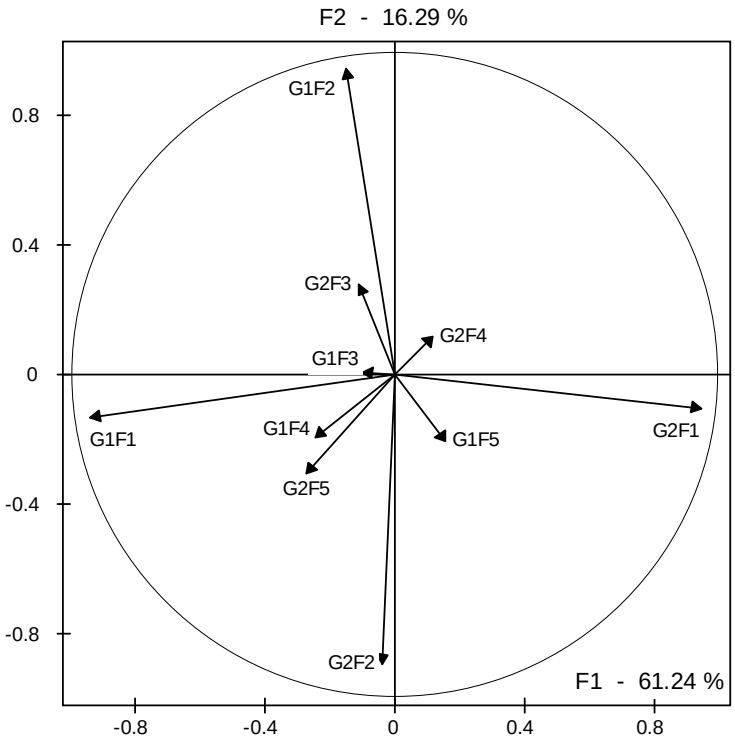

Figure 8. *Representation of the factors from separate analyses*
GjFs : rank-s factor of separate PCA of group j

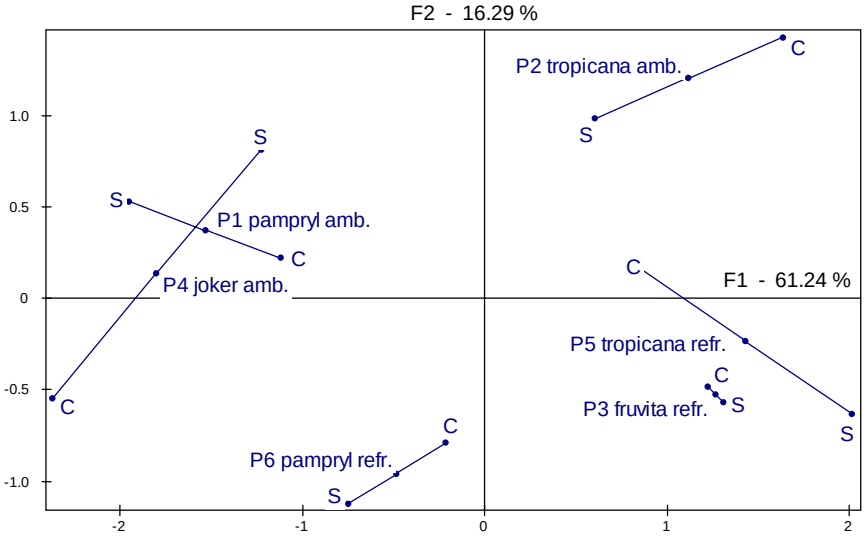

Figure 9. *Superimposed representation of partial individuals*
The mean points of figure 6 are here joined to corresponding partial points. S: sensory; C: chemical

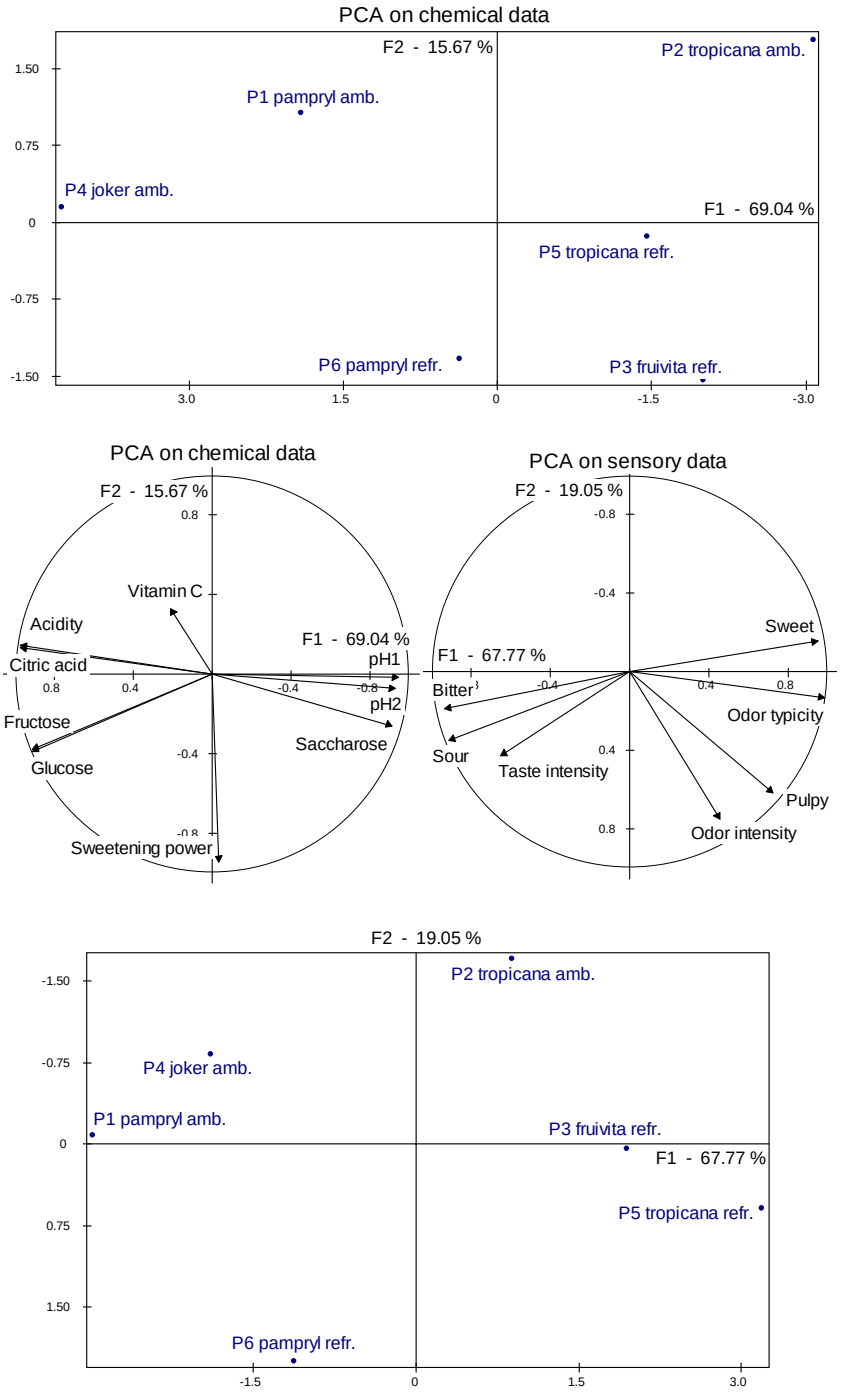

Figure 10. *Separate PCA of each group of variables*

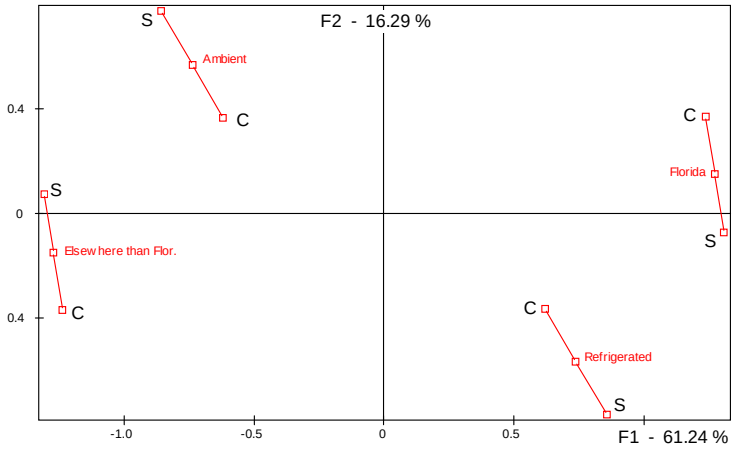

Figure 11. *Display of mean and partial categories*
Each category lies in the center of gravity of the individuals belonging to it

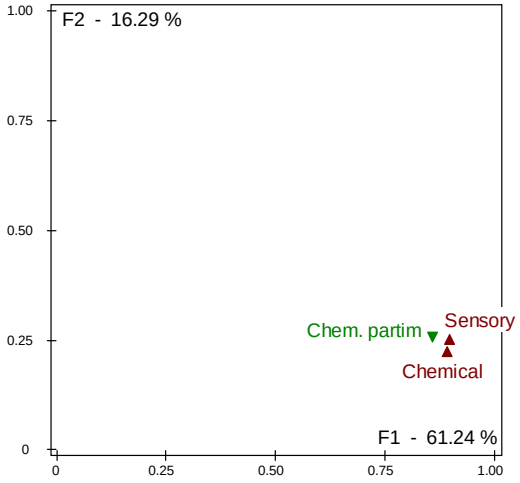

Figure 12. *Representation of groups of variables.*
In this display, a group of variables is represented by a single point. Here, appears a third group, obtained from the chemical one, by giving up two variables (pH2 and vitamin C)

12. References

CARROLL J.D. (1968). A generalization of canonical correlation analysis to three or more sets of variables, *Proc. 76th Conv. Amer. Psych. Assoc.*, pp. 227-228.

ESCOFIER B. & PAGÈS J. (1994). Multiple Factor Analysis (AFMULT package). *Computational statistics & data analysis* 18 121-140

ESCOFIER B., PAGES J. (1988-1998). Analyses factorielles simples et multiples ; objectifs, méthodes et interprétation. Dunod. Paris.

HOTELLING H. (1936). Relations between two sets of variables. *Biometrika,* 28, p 129-149.

MACFIE H. J., BRATCHELL N., GREENHOFF, VALLIS L. V. (1989). Designs to balance the effect of order of presentation and first-order carry-over effects in hall tests. *Journal of Sensory Studies* 4 129-148.

PAGÈS J. (1996). Eléments de comparaison entre l'Analyse Factorielle Multiple et la méthode STATIS. *Revue de Statistique appliquée* XLIV (4) 81-95.

PAGÈS J. (2002). Analyse factorielle multiple appliquée aux variables qualitatives et aux données mixtes. *Rev Statistique appliquée*, L (4), 5-37.

SPAD rel. 5.5, (2002). Système pour l'analyse des données. Logiciel diffusé par Decisia, Levallois-Perret, 92532 France.

Appendix : raw data and standardised data

Raw data	Pampryl amb.	Tropicana amb.	Fruivita refr.	Joker amb.	Tropicana refr.	pampryl refr.	mean	Standard deviation
	P1	P2	P3	P4	P5	P6		
Glucose	25,32	17,33	23,65	32,42	22,7	27,16	24,76	4,57
Fructose	27,36	20	25,65	34,54	25,32	29,48	27,06	4,41
Saccharose	36,45	44,15	52,12	22,92	45,8	38,94	40,06	9,16
Pouvoir sucrant	89,95	82,55	102,22	90,71	94,87	96,51	92,80	6,11
PH brut	3,59	3,89	3,85	3,6	3,82	3,68	3,74	0,12
PH après centrif.	3,55	3,84	3,81	3,58	3,78	3,66	3,70	0,11
Titre	13,98	11,14	11,51	15,75	11,8	12,21	12,73	1,62
Acide citrique	0,84	0,67	0,69	0,95	0,71	0,74	0,77	0,10
Vitamine C	43,44	32,7	37	36,6	39,5	27	36,04	5,18
Saccharose %	0,41	0,54	0,51	0,26	0,49	0,41	0,44	0,10
Intensité odeur	2,82	2,76	2,83	2,76	3,2	3,07	2,91	0,17
Typicité odeur	2,53	2,82	2,88	2,59	3,02	2,73	2,76	0,17
Pulpy	1,66	1,91	4	1,66	3,69	3,34	2,71	0,99
Intensité goût	3,46	3,23	3,45	3,37	3,12	3,54	3,36	0,14
Caractère acide	3,15	2,55	2,42	3,05	2,33	3,31	2,80	0,38
Caractère amer	2,97	2,08	1,76	2,56	1,97	2,63	2,33	0,42
Caractère sucré	2,6	3,32	3,38	2,8	3,34	2,9	3,06	0,30

Standardised Data	Pampryl amb.	Tropicana amb.	Fruivita refr.	Joker amb.	Tropicana refr.	pampryl refr.	mean	Standard deviation
	P1	P2	P3	P4	P5	P6		
Glucose	0,12	-1,63	-0,24	1,67	-0,45	0,52	0	1
Fructose	0,07	-1,60	-0,32	1,70	-0,39	0,55	0	1
Saccharose	-0,39	0,45	1,32	-1,87	0,63	-0,12	0	1
Pouvoir sucrant	-0,47	-1,68	1,54	-0,34	0,34	0,61	0	1
PH brut	-1,23	1,26	0,93	-1,15	0,68	-0,49	0	1
PH après centrif.	-1,36	1,21	0,94	-1,09	0,68	-0,38	0	1
Titre	0,77	-0,98	-0,75	1,86	-0,57	-0,32	0	1
Acide citrique	0,75	-0,98	-0,78	1,86	-0,58	-0,27	0	1
Vitamine C	1,43	-0,65	0,19	0,11	0,67	-1,75	0	1
Saccharose %	-0,28	1,11	0,82	-1,90	0,55	-0,30	0	1
Intensité odeur	-0,52	-0,87	-0,46	-0,87	1,75	0,97	0	1
Typicité odeur	-1,38	0,35	0,71	-1,03	1,54	-0,19	0	1
Pulpy	-1,06	-0,81	1,30	-1,06	0,99	0,64	0	1
Intensité goût	0,68	-0,91	0,61	0,06	-1,67	1,24	0	1
Caractère acide	0,91	-0,66	-1,00	0,65	-1,24	1,33	0	1
Caractère amer	1,52	-0,59	-1,35	0,55	-0,85	0,71	0	1
Caractère sucré	-1,50	0,87	1,06	-0,85	0,93	-0,52	0	1