Computational Statistics & Data Analysis 19 (1995) 203–222
North-Holland

Representation, synthesis, variability and data preprocessing of a three-way data set

Alfredo Rizzi and Maurizio Vichi
Università degli studi di Roma "La Sapienza", Roma, Italy

Received October 1992
Revised March 1993

1. Representation of the three-way data set

Let $X = \{x_{ijh} : i \in I, j \in J, h \in H\}$ be the three-way or three-mode data set, where x_{ijh} is a real value of the variable j-th observed on the i-th unit, according to the h-th situation, and $I = \{1, \ldots, n\}$, $J = \{1, \ldots, k\}$ and $H = \{1, \ldots, r\}$ are the sets of indices of modes i (units), j (variables) and h (occasions) respectively. We now analyze the representation of the three-way data set according to data types and different data structures forming X.

1.1. Matrix representation of X

The three-way data set X can be represented as a set of matrices:
(a) r units-variables matrices, or k-variates, called *frontal slabs* (or *slices*) $X_{\cdot\cdot h} \equiv \{x_{ijh} : i \in I, j \in J\}$ $(h \in H)$ (Harshman and Lundy, 1984). They can be considered the result of the observation, on a set of n units, of k variables; observation repeated for r situations. With this data representation of X we can evaluate: the relations between different occasions, i.e., the association or dependence between sets of variates (Ramsay *et al.*, 1984; Vichi, 1989) examined on the same units in different occasions; or the dissimilarity between studies (Escoufier, 1987) that is between sets of units on which the same variables have been measured in different occasions;

Correspondence to: M. Vichi, Università la Sapienza, Dip. Di Statistica e Probabilitá, Piazale Aldo Moro 5, I-00185, Roma, Italy.

SSDI 0167-9473(93)E0057-B

(b) k units-occasions matrices or r-variates, defined *lateral slices* $\boldsymbol{X}_{\cdot j \cdot} \equiv \{x_{ijh}: i \in \boldsymbol{I}, h \in \boldsymbol{H}\}$ $(j \in \boldsymbol{J})$. They can be seen as the result of the observation, on a set of n units, of r times the variable j; observation repeated for k different variables. With this data representation we can evaluate the relations between different variables repeatedly observed, i.e., by the relations between set of variates each set compound by the same variable repeatedly observed in different occasions.
(c) n variables-occasions matrices, called *horizontal slices*, $\boldsymbol{X}_{i \cdot \cdot} \equiv \{x_{ijh}: h \in \boldsymbol{H}, j \in \boldsymbol{J}\}$ $(i \in \boldsymbol{I})$. They can be considered as the result of the observation of k variables in r situations, on the i-th unit; observation repeated for n different units. With this data representation of $\boldsymbol{X}$ we can evaluate the dissimilarities between different units on the base of different variables examined on several occasions (for example k time series related to n units).

The three-way data set can be the result of the straight collection, or calculus – repeated in r occasions – of a prefixed proximity measure between couples of units or variables. In this case the three-way data matrix $\boldsymbol{X}$ is a two mode matrix represented generally as a set of:
(d) r similarity or dissimilarity square matrices of dimension n or k. Note that using the dissimilarity and similarity definition adopted by Gower and Legendre (1986), the dispersion and the correlation matrices can be considered similarity matrices of order k.

1.2. Vector representation of X

The three-way set $\boldsymbol{X}$ can be seen as a set of equal vectors (or fibers): nr row (k-elements) vectors $\boldsymbol{X}_{i.h}$ $(i \in \boldsymbol{I}, h \in \boldsymbol{H})$, each one associated to the i-th unit examined in the h-th occasion; kr column (n-elements) vectors $\boldsymbol{X}_{\cdot jh}$ $(j \in \boldsymbol{J}, h \in \boldsymbol{H})$, each one associated to the j-th variable observed in the h-th occasion; nk lateral (r-elements) vectors (tubes) $\boldsymbol{X}_{ij\cdot}$ $(i \in \boldsymbol{I}, j \in \boldsymbol{J})$, each one associated to the i-th unit on which variable j has been observed; also called fibers in direction of modes i, j and h respectively (Harshman and Lundy, 1984).

1.3. Pooled representation of X

In order to make multivariate analysis of a three-way matrix possible, it is often useful to represent $\boldsymbol{X}$ as a large pooled two mode matrix; for instance juxtaposing the frontal slabs:

$$\hat{\boldsymbol{X}} \equiv \begin{bmatrix} \boldsymbol{X}_{\cdot \cdot 1} \\ \boldsymbol{X}_{\cdot \cdot 2} \\ \vdots \\ \boldsymbol{X}_{\cdot \cdot 3} \end{bmatrix}. \tag{1}$$

There are twelve different ways to represent a three-mode matrix as a pooled matrix. The first six may be defined through the column vectorization of the

slabs or the transposes of the slabs of the three-way matrix, while the other six are the transposes of the first six. We therefore have:

1) *The column frontal pooled matrix* *:

$$\boldsymbol{X}_F = {}_{kn}\boldsymbol{X}_r \equiv (\text{vec } \boldsymbol{X}_{\cdot\cdot 1} \text{ vec } \boldsymbol{X}_{\cdot\cdot 2} \ldots \text{vec } \boldsymbol{X}_{\cdot\cdot r}), \tag{2}$$

where vec $\boldsymbol{X}_{\cdot\cdot h}$ is the column vectorization of the frontal slab $\boldsymbol{X}_{\cdot\cdot h}$, i.e., the k column (n-elements) vectors $\boldsymbol{X}_{\cdot jh}$ ($j \in \boldsymbol{J}$), forming columns of $\boldsymbol{X}_{\cdot\cdot h}$, are strung out. $\boldsymbol{X}_F$ has kn rows and r columns.

D'Alessio (1986) and Rizzi (1989) principal matrices are detected via Principal Component Analysis (PCA) of $\boldsymbol{X}_F$. The interstructure analysis of STATIS (L'Hermier des Plantes, 1976 Escoufier, 1977, 1987), is actually the PCA of the column frontal pooled matrix $\{\text{vec } \boldsymbol{X}_{\cdot\cdot h}\boldsymbol{X}'_{\cdot\cdot h}: h \in \boldsymbol{H}\}$. Also the interstructure analysis, based on a three-way similarity matrix (D'Ambra, e Marchetti, 1986; Coppi, 1986) and computed according the types of variables observed, is the PCA of the column frontal pooled matrix of the three-way similarity matrix.

2) *The row frontal pooled matrix*:

$$\boldsymbol{X}_{F'} = {}_{nk}\boldsymbol{X}_r \equiv (\text{vec } \boldsymbol{X}'_{\cdot\cdot 1} \text{ vec } \boldsymbol{X}'_{\cdot\cdot 2} \ldots \text{vec } \boldsymbol{X}'_{\cdot\cdot r}), \tag{3}$$

where the n row (k-elements) vectors $\boldsymbol{X}_{i\cdot h}$ ($i \in \boldsymbol{I}$), forming the rows of $\boldsymbol{X}_{\cdot\cdot h}$, are strung out. Matrix $\boldsymbol{X}_{F'}$ has nk rows and r columns.

3) *The column lateral pooled matrix*:

$$\boldsymbol{X}_L = {}_{rn}\boldsymbol{X}_k \equiv (\text{vec } \boldsymbol{X}_{\cdot 1\cdot} \text{ vec } \boldsymbol{X}_{\cdot 2\cdot} \ldots \text{vec } \boldsymbol{X}_{\cdot k\cdot}), \tag{4}$$

where r column (n-elements) vectors $\boldsymbol{X}_{\cdot jh}$ ($h \in \boldsymbol{H}$), forming the columns of $\boldsymbol{X}_{\cdot j}$, are strung out. Matrix $\boldsymbol{X}_L$ is matrix (1), and has rn rows and k columns. Levin (1966) generalized PCA is essentially equivalent to the PCA on the pooled matrix $\boldsymbol{X}_L$. Also simultaneous Component Analysis, based on a gradient method (Millsap and Meredith, 1988) or on alternating least square algorithm (Kiers and Ten Berge, 1989; 1991; Kiers 1990) – consists of an improved version of PCA on $\boldsymbol{X}_L$.

4) *The row lateral pooled matrix*:

$$\boldsymbol{X}_{L'} = {}_{nr}\boldsymbol{X}_k \equiv (\text{vec } \boldsymbol{X}'_{\cdot 1\cdot} \text{ vec } \boldsymbol{X}'_{\cdot 2\cdot} \ldots \text{vec } \boldsymbol{X}'_{\cdot k\cdot}), \tag{5}$$

where the n tube (r-elements) vectors $\boldsymbol{X}_{ij\cdot}$ ($i \in \boldsymbol{I}$), forming the lateral slab $\boldsymbol{X}_{\cdot j}$, are strung out. Matrix $\boldsymbol{X}_{L'}$ is formed by nr rows and k columns.

5) *The column horizontal pooled matrix*:

$$\boldsymbol{X}_H = {}_{kr}\boldsymbol{X}_n \equiv (\text{vec } \boldsymbol{X}_{1\cdot\cdot} \text{ vec } \boldsymbol{X}_{2\cdot\cdot} \ldots \text{vec } \boldsymbol{X}_{n\cdot\cdot}), \tag{6}$$

where the k tube (r-elements) vectors $\boldsymbol{X}_{ij\cdot}$ ($j \in \boldsymbol{J}$), forming the columns of the horizontal slab $\boldsymbol{X}_{i\cdot\cdot}$, are strung out. Matrix $\boldsymbol{X}_H$ has kr rows and n columns.

* In the second notation used for the pooled matrix the subscript on left of $\boldsymbol{X}$ indicates the number of its rows and the subscript on the right of $\boldsymbol{X}$ the number of its columns; i.e., ${}_{kn}\boldsymbol{X}_r$ is a pooled matrix with kn rows and r columns.

Table 1
Parameter estimation of the Tucker's models I, II and III trough frontal, lateral and horizontal pooled matrices. The numbers on the left side – reported in the table (excluding the zero) – correspond to the original steps indicated in the work of Tucker (1966)

The TUCKER MODEL for X with elements x_{ijh} has the form:

$$x_{ijh} = \sum_{m=1}^{M} \sum_{p=1}^{P} \sum_{q=1}^{Q} a_{im} b_{jp} c_{hq} g_{mpq} + e_{ijh}$$

where a_{im}, b_{jp}, c_{kq} are the elements of matrices A, B and C; while g_{mpq} are the entries of the three-way core matrix G

PARAMETER ESTIMATION OF THE MODEL ACCORDING TO	
	TUCKER MODEL I
0.	given the transposes pooled matrices: X_H', X_L' and X_F';
1.-2.	compute PCA of X_H', X_L', X_F' defining principal component coefficient matrices A, B, C;
3.	compute the core matrix $G = A'X_H'(B \otimes C)$.
	TUCKER MODEL II
0.	given the pooled matrix X_L';
1.-2.	compute PCA of X_L' defining the principal component coefficient matrix B;
3.	compute principal component matrix $Y_L' = B'X_L'$;
4.-5.	rewrite Y_L' as the transpose of the column frontal pooled matrix Y_F' and compute PCA of Y_F', defining principal component coefficient matrix C;
6.	compute principal component matrix $Z_F' = C'Y_F'$;
7.	rewrite Z_F' as the transpose of the column horizontal pooled matrix Z_H';
8.-9.	compute V which columns are eigenvectors associated to eigenvalues appearing in the diagonal matrix S of $Z_H Z_H'$, then compute $G = VS^{1/2}$ and $A = Z_H' S^{-1/2}$.
	TUCKER MODEL III
0.	given the column horizontal pooled matrix: X_H
1.	compute matrix $R = X_H X_H'$
2.-3.	compute PCA of X_F and X_L defining principal component coefficient matrices C and B;
4.-5.	compute matrix $(B' \otimes C')R(B \otimes C)$, then compute their eigenvectors - arranged on columns of V - and eigenvalues, appearing on diagonal matrix S;
6.-7.	compute matrices: $G = VS^{1/2}$ and $A = X_H'(B \otimes C)VS^{-1/2}$

6) *The two horizontal pooled matrix*:

$$X_{H'} = {}_{rk}X_n \equiv (\text{vec } X_{1..}' \text{ vec } X_{2..}' \ldots \text{vec } X_{n..}'), \tag{7}$$

where the r row (k-elements) vectors $X_{i.h}$ ($h \in H$), forming the rows of $X_{i..}$, are strung out. Matrix $X_{H'}$ has rk rows and n columns. The transpose of Matrix $X_{H'}$ can be also obtained placing frontal slices one beside the other. Escofier and Pages (1984, 1989) multiple factorial analysis is based on the transpose of the row horizontal pooled matrix.

Also the solutions of the parameters of the three fundamental models I, II and III, defined by Tucker (1966, 1972), are computed trough PCA on frontal, lateral and horizontal pooled matrices, as we have summarized in Table 1.

2. Synthesis of a three-way data set

The information given by a three-way data set can be synthesized according to different data structures or types of representation forming the three-way data set (fibers, matrices and 3-way arrays).

Many authors have considered data structures, principally matrices, which from their point of view average the three-way data set. These matrices hold different properties so we have to clarify what can be properly considered a synthesis or mean structure of $\boldsymbol{X}$.

A fixed data structure $\boldsymbol{X}^*$ is a *representative synthesis* (*mean*) of $\boldsymbol{X}$ if: (a) $\boldsymbol{X}^*$ minimizes differences between itself and the data structures of the same dimension of $\boldsymbol{X}^*$ compounding $\boldsymbol{X}$; (b) $\boldsymbol{X}^*$ satisfies the internality property, that is, each element of $\boldsymbol{X}^*$ is internal to the interval formed by the least and the largest corresponding elements of the data structures forming $\boldsymbol{X}$. We can therefore define: one-way or fiber synthesis, two-way or slab synthesis, three-way or global synthesis.

Note that we have to distinguish a synthesis of $\boldsymbol{X}$ from what it can be called a *factorial data structure*, (that is a factor (one-way structure) or a factorial matrix) which does not average $\boldsymbol{X}$, but it insure a data structure with minimum loss of information of $\boldsymbol{X}$, (according to a prefixed measure of loss of information). Examples of one-way factorial structures are the principal components of each slab forming $\boldsymbol{X}$, while the compromise matrix (Escoufier, 1980, 1987) principal matrices (Rizzi, 1989) and factorial matrices (Vichi, 1990) are examples of two-way factorial structures.

2.1. One-way synthesis of $\boldsymbol{X}$

One-way synthesis or fiber mean of $\boldsymbol{X}$ is the mean of the elements of each: column-vector $\boldsymbol{X}_{\cdot jh}$ $(j \in \boldsymbol{J}, h \in \boldsymbol{H})$; or row-vector $\boldsymbol{X}_{ij\cdot}$ $(i \in \boldsymbol{I}, j \in \boldsymbol{J})$; or tube-vectors $\boldsymbol{X}_{i\cdot h}$ $(i \in \boldsymbol{I}, h \in \boldsymbol{H})$. These means may be arranged into matrices that are *synthesis matrices* of $\boldsymbol{X}$, according to the definition given in paragraph 2.

The matrix ${}_p\bar{\boldsymbol{X}}_{\cdot\cdot F}$, which elements are the power mean of tube-vectors $\boldsymbol{X}_{ij\cdot}$ elements.

$$ {}_p\bar{\boldsymbol{X}}_{\cdot\cdot F} \equiv \left\{ {}_p\bar{x}_{ij\cdot} = \left(\frac{1}{r} \sum_{h=1}^{r} x_{ijh}^{p} \right)^{1/p} : i \in \boldsymbol{I},\ j \in \boldsymbol{J} \right\}, \tag{8} $$

is a synthesis matrix, named, *frontal synthesis matrix*, or *frontal mean matrix* of $\boldsymbol{X}$, since defined a dissimilarity between a frontal slab $\boldsymbol{X}_{\cdot\cdot h}$ and ${}_p\bar{\boldsymbol{X}}_{\cdot\cdot F}$,

$$ d_p\left(\boldsymbol{x}_{\cdot\cdot h}, {}_p\bar{\boldsymbol{X}}_{\cdot\cdot F}\right) = \sum_{i=1}^{n} \sum_{j=1}^{k} \left(x_{ijh}^{p} - \bar{x}_{ij\cdot}^{p} \right)^2, \tag{9} $$

i.e., the square of the Euclidean norm (Frobenius norm) of matrix $\{x_{ijh}^{p} - \bar{x}_{ij\cdot}^{p}\}$, ${}_p\bar{\boldsymbol{X}}_{\cdot\cdot F}$ satisfies the following two properties:

(a) ${}_p\bar{\boldsymbol{X}}_{\cdot\cdot F}$ minimizes $\Sigma_h\ d(\boldsymbol{X}_{\cdot\cdot h}, \bar{\boldsymbol{X}}_{\cdot\cdot F})$, over the sum of dissimilarities between $\boldsymbol{X}_{\cdot\cdot h}$ and any real matrix with same dimensions of $\boldsymbol{X}_{\cdot\cdot h}$;

(b) ${}_p\bar{\boldsymbol{X}}_{..F}$ satisfies the internality property of a matrix: i.e.

$$\min \boldsymbol{X}_{..h} \le {}_p\bar{\boldsymbol{X}}_{..F} \le \max \boldsymbol{X}_{..h}, \qquad \text{where}$$

$$\min \boldsymbol{X}_{..h} \equiv \left\{\min(x_{ijh}: h \in \boldsymbol{H}): i \in \boldsymbol{I}, j \in \boldsymbol{J}\right\},$$

$$\max \boldsymbol{X}_{..h} \equiv \left\{\max(x_{ijh}: h \in \boldsymbol{H}): i \in \boldsymbol{I}, j \in \boldsymbol{J}\right\}$$

and $\boldsymbol{A} \le \boldsymbol{B}$ means that $a_{ij} \le b_{ij} \forall i, j$.

The frontal mean matrix of a three-way, three mode matrix $\boldsymbol{X}$, can be used to average the influence of the occasions and to apply multivariate statistical analysis to the so obtained two-way data set. For instance when the occasions are the months of a year, the frontal mean matrix averages the influence of the seasonality of the observed data.

With the same procedure we can define the *lateral synthesis matrix*, or the *lateral mean matrix* of $\boldsymbol{X}$, which elements are the power mean of the row-vectors $\boldsymbol{X}_{i.h}$

$${}_p\bar{\boldsymbol{X}}_{.L.} \equiv \left\{{}_p\bar{x}_{i.h} = \left(\frac{1}{k}\sum_{j=1}^{k} x_{ijh}^{p}\right)^{1/p} : i \in \boldsymbol{I}, h \in \boldsymbol{H}\right\}, \tag{10}$$

which satisfies:

(a) ${}_p\bar{\boldsymbol{X}}_{.L.}$ minimizes the sum of dissimilarities:

$$d_p\left(\boldsymbol{X}_{.j}, {}_p\bar{\boldsymbol{X}}_{.L.}\right) = \sum_{i=1}^{n}\sum_{h=1}^{r}\left(x_{ijh}^{p} - \bar{x}_{i.h}^{p}\right)^2, \tag{11}$$

over all sums of dissimilarities between each $\boldsymbol{X}_{.j.}$ and any real matrix with same dimensions of $\boldsymbol{X}_{.j}$;

(b) $\min \boldsymbol{X}_{.j.} \le {}_p\bar{\boldsymbol{X}}_{.L.} \le \max \boldsymbol{X}_{.j.}$, where $\min \boldsymbol{X}_{.j.} \equiv \{\min(x_{ijh}: j \in \boldsymbol{J}): i \in \boldsymbol{I}, h \in \boldsymbol{H}\}$,

$$\max \boldsymbol{X}_{.j.} \equiv \left\{\max(x_{ijh}: h \in \boldsymbol{H}): i \in \boldsymbol{I}, h \in \boldsymbol{H}\right\},$$

Note that when $\boldsymbol{X}$ is a three-mode matrix the lateral mean matrix has no statistical meaning since the mean of the elements of the row vectors of $\boldsymbol{X}$ mixes modalities generally with different unit of measure. Such mean matrix is useful when: $\boldsymbol{X}$ is a proximity three-way matrix (two mode matrix), or to define ipsative data.

Finally we can define the *horizontal synthesis matrix*, or the *horizontal mean matrix* of $\boldsymbol{X}$, which elements are the power mean of the column-vectors $\boldsymbol{X}_{.jh}$

$${}_p\boldsymbol{X}_{H..} \equiv \left\{{}_p\bar{x}_{.jh} = \left(\frac{1}{n}\sum_{i=1}^{n} x_{ijh}^{2}\right)^{1/p} : h \in \boldsymbol{H}, j \in \boldsymbol{J}\right\}, \tag{12}$$

which satisfies:

(a) ${}_p\bar{\boldsymbol{X}}_{H..}$ minimizes the sum of dissimilarities

$$d_p\left(\boldsymbol{x}_{i..}, {}_p\bar{\boldsymbol{x}}_{H..}\right) = \sum_{j=1}^{k}\sum_{r=1}^{h}\left(x_{ijh}^{p} - \bar{x}_{.jh}^{p}\right)^2, \tag{13}$$

over all possible real matrices with the same dimensions of $\boldsymbol{X}_{i..}$;
(b) $\min \boldsymbol{X}_{\mathrm{i}..} \leq \bar{\boldsymbol{X}}_{\mathrm{H}..} \leq \max \boldsymbol{X}_{\mathrm{i}..}$, where

$$\max \boldsymbol{X}_{i..} \equiv \left\{\min(x_{ijh}: i \in \boldsymbol{I}): h \in \boldsymbol{H}, j \in \boldsymbol{J}\right\},$$

$$\max \boldsymbol{X}_{i..} \equiv \left\{\max(x_{ijh}: i \in \boldsymbol{I}): h \in \boldsymbol{H}, j \in \boldsymbol{J}\right\},$$

The horizontal mean matrix of a three-way, three mode matrix $\boldsymbol{X}$, can be used to average the different units and to define a variable x occasion matrix in order to apply appropriate classical statistical analysis. For instance when the occasions are different instants of time, the horizontal mean matrix is formed by time series on which we can apply time series analysis.

In Table 2, for different values of p, some remarkable frontal, lateral and horizontal mean matrices of $\boldsymbol{X}$ are reported.

2.2. Two-way synthesis of X

Two-way synthesis or slab mean of $\boldsymbol{X}$ is the mean of the elements of frontal slabs $\boldsymbol{X}_{..h}$ $(h \in \boldsymbol{H})$; or lateral slabs $\boldsymbol{X}_{.j.}$ $(j \in \boldsymbol{J})$; or horizontal slabs $\boldsymbol{X}_{i..}$ $(i \in \boldsymbol{I})$; indicated with $\bar{x}_{..h}$, $\bar{x}_{.j.}$ *and* $\bar{x}_{i..}$, respectively.

2.3. Three-way synthesis of X

Three-way synthesis or global mean of $\boldsymbol{X}$ is the mean of all elements of $\boldsymbol{X}$, which is indicated with $\bar{x}$.

3. Variability of a three-way data set

The variability of a three way data set can be analyzed according to three different levels: one way or fiber variability; two-way or slab variability; and three-way variability.

3.1. One-way variability of X

One-way variability of $\boldsymbol{X}$ is the fiber or univariate variability of vectors $\boldsymbol{X}_{.jh}$ $(j \in \boldsymbol{J}, h \in \boldsymbol{H})$; $\boldsymbol{X}_{ij.}$ $(i \in \boldsymbol{I}, j \in \boldsymbol{J})$ and $\boldsymbol{X}_{i.h}$ $(i \in \boldsymbol{I}, h \in \boldsymbol{H})$ forming $\boldsymbol{X}$, therefore classical statistical univariate analysis can be applied.

The standard deviations of $\boldsymbol{X}_{.jh}$, $\boldsymbol{X}_{ij.}$ and $\boldsymbol{X}_{i.h}$ are named *column standard deviation* $\sigma_{.jh}$ $(j \in \boldsymbol{J}, h \in \boldsymbol{H})$, *tube standard deviation* $\sigma_{ij.}$ $(i \in \boldsymbol{I}, j \in \boldsymbol{J})$ and *row standard deviation* $\sigma_{i.h}$ $(i \in \boldsymbol{I}, h \in \boldsymbol{H})$ respectively. Note that when $\boldsymbol{X}$ is a three-mode matrix the row standard deviation has no statistical meaning. It is used in the case of similarity or dissimilarity three-way matrices.

Table 2
frontal, lateral and horizontal synthesis matrices of X, according to the value p of dissimilarities (9), (11) and (13) respectively.

p	expression	p	expression
FRONTAL MEAN MATRICES			
	frontal arithmetic mean matrix		frontal square mean matrix
$p=1$	${}_1\bar{X}_{..F}=\{\bar{x}_{ij.}=\frac{1}{r}\sum_{h=1}^{r}x_{ijh}: i\in I, j\in J\}$	$p=2$	${}_2\bar{X}_{..F}=\{{}_2\bar{x}_{ij.}=(\frac{1}{r}\sum_{h=1}^{r}x_{ijh}^2)^{1/2}: i\in I, j\in J\}$
	frontal harmonic mean matrix		frontal geometric mean matrix
$p=-1$	${}_{-1}\bar{X}_{..F}=\{{}_{-1}\bar{x}_{ij.}=r/(\sum_{h=1}^{r}\frac{1}{x_{ijh}}): i\in I, j\in J\}$	$p\to 0$	${}_0\bar{X}_{..F}=\{{}_0\bar{x}_{ij.}=(\prod_{h=1}^{r}x_{ijh})^{1/r}: i\in I, j\in J\}$
LATERAL MEAN MATRICES			
	lateral arithmetic mean matrix		lateral square mean matrix
$p=1$	${}_1\bar{X}_{.L.}=\{\bar{x}_{i.h}=(\frac{1}{k}\sum_{j=1}^{r}x_{ijh}): i\in I, h\in H\}$	$p=2$	${}_2\bar{X}_{.L.}=\{{}_2\bar{x}_{i.h}=(\frac{1}{k}\sum_{j=1}^{k}x_{ijh}^2)^{1/2}: i\in I, h\in H\}$
	lateral harmonic mean matrix		lateral geometric mean matrix
$p=-1$	${}_{-1}\bar{X}_{.L.}=\{{}_{-1}\bar{x}_{i.h}=k/(\sum_{j=1}^{k}\frac{1}{x_{ijh}}): i\in I, h\in H\}$	$p\to 0$	${}_0\bar{X}_{.L.}=\{{}_0\bar{x}_{i.h}=(\prod_{j=1}^{k}x_{ijh})^{1/k}: i\in I, h\in H\}$
HORIZONTAL MEAN MATRICES			
	horizontal arithmetic mean matrix		horizontal square mean matrix
$p=1$	${}_1\bar{X}_{H..}=\{\bar{x}_{.jh}=(\frac{1}{n}\sum_{i=1}^{n}x_{ijh}): h\in H, j\in J\}$	$p=2$	${}_2\bar{X}_{H..}=\{{}_2\bar{x}_{.jh}=(\frac{1}{n}\sum_{i=1}^{n}x_{ijh}^2)^{1/2}: h\in H, j\in J\}$
	horizontal harmonic mean matrix		horizontal geometric mean matrix
$p=-1$	${}_{-1}\bar{X}_{H..}=\{{}_{-1}\bar{x}_{.jh}=n/(\sum_{i=1}^{n}\frac{1}{x_{ijh}}): h\in H, j\in J\}$	$p\to 0$	${}_0\bar{x}_{H..}=\{{}_0\bar{x}_{.jh}=(\prod_{i=1}^{n}x_{ijh})^{1/n}: h\in H, j\in J\}$

3.2. Two-way variability of X

Two-way variability of X is the slab or multivariate variability of each data set $X_{..h}$ $(h \in H)$, $X_{.j.}$ $(j \in J)$, $X_{i..}$ $(i \in I)$, forming X.

A multivariate measure of variability of a given data matrix A may be defined through a norm of the dispersion matrix of A, as was noted by Mathai (1967), Lunetta (1973) and Amato (1981).

Indicated with:

$$\Sigma_{..h} \equiv \left\{\sigma(X_{jh}, X_{uh})\ j, u \in J\right\}, \qquad (h \in H);$$

$$\Sigma_{.j.} \equiv \left\{\sigma(X_{jh}, X_{jm})\ h, m \in H\right\}, \qquad (j \in J);$$

$$\Sigma_{i..} \equiv \left\{\sigma(X_{ij}, X_{iu.})\ j, u \in J\right\}, \qquad (i \in I);$$

the dispersion matrices of $X_{..h}$, $X_{.j.}$ and $X_{i..}$ respectively, and using in the following notation for the subscripts the letter a in order to unify the treatment, we can consider the indices:

$\sigma_1(X_a) = \sqrt{\operatorname{tr}(\Sigma_a)}$; the total variation of X_a;

$\sigma_2(X_a) = \sqrt{\operatorname{tr}(\Sigma_a \Sigma_a)}$; the Euclidean norm of X_a;

$\sigma_3(X_a) = |\Sigma_a|$; the determinant of X_a or generalized variance of Wilks;

where, for $a = ..h$, $a = .j.$, $a = i..$ we have multivariate variability indices respectively for the frontal, lateral and horizontal slabs of X.

3.3. Three-way variability of X

The analysis of the conjoint variability (or dispersion) of the three-way data set X is here faced with an axiomatix approach, giving desiderata properties that an index $\sigma(X)$ of variability of X should held. The variability of X is evaluated through slabs and fibers forming X. Here, for reasons of space, we discuss in depth the case of frontal slabs and column fibers. However the considerations stated below remain valid for the other cases (variability of X evaluated through lateral slabs and column fibers; or horizontal slabs and tube fibers), for which we give, at the end of this paragraph, the corresponding measures.

In order to introduce an index $\sigma(X)$, we have first to guarantee that it is the logical extension of both an univariate and a multivariate well known variability measures, with the following properties:

(a) $\sigma(X)$ is a function of the univariate measures of variability of each variate $X_{.jh}$, $(j \in J, h \in H)$ belonging to X;

(a') $\sigma(X)$ is a function of the multivariate measures of variability of the k-variate associated to $X_{..h}$, $(h \in H)$ of X;

(b) $\sigma(\boldsymbol{X})$ is reduced to an univariate measure of variability when $\boldsymbol{X}$ degenerates into a variate $\boldsymbol{X}_{.jh}$ (one-way data set, i.e. $n > 1$, $k = 1$, $r = 1$);
(b′) $\sigma(\boldsymbol{X})$ is a multivariate measure of variability when $\boldsymbol{X}$ degenerates to a k-variate $\boldsymbol{X}_{..h}$ (two-way data set; i.e. $n > 1$, $k > 1$, $r = 1$).

An index $\sigma(\boldsymbol{X})$ has also to satisfy the following conditions before it can be considered a useful measure of variability of a three-way data set:
(1) $\sigma(\boldsymbol{X}) \geq 0$; *it is non negative*;
(2) $\sigma(\boldsymbol{X}) = \sigma(\boldsymbol{X}_t)$, where $\boldsymbol{X}_t \equiv \{\boldsymbol{X}_{..h} - {}_n\mathbf{1}_k\mathbf{1}'\boldsymbol{C}_h\colon h \in \boldsymbol{H}\}$ is the three-way matrix translated according to matrices $\boldsymbol{C}_h = \operatorname{diag}(c_{1h}, \ldots, c_{kh})$, $c_{ih} \in \mathbb{R}$ $(h \in \boldsymbol{H})$ and ${}_m\mathbf{1}$ $(m = n, k)$ is a m-vector of unitary elements; *it is invariant under translation of each variate* $\boldsymbol{X}_{.jh}$ *of* $\boldsymbol{X}$;
(2′) $\sigma(\boldsymbol{X}) = \sigma(\overset{*}{\boldsymbol{X}})$, where $\overset{*}{\boldsymbol{X}} \equiv \{({}_n\boldsymbol{I} - {}_n\mathbf{1}_n\mathbf{1}'(1/n))\boldsymbol{X}_{..h}\colon h \in \boldsymbol{H}\}$ is the tree-way matrix fiber-centered, i.e. determined one-centering each variate of $\boldsymbol{X}$ (see section 4.4), and ${}_n\boldsymbol{I}$ is the identity matrix of order n; *it is invariant under one-way centering of each variate of* $\boldsymbol{X}$;
(3) $\sigma(\boldsymbol{X}) = 0$ if and only if $\overset{*}{\boldsymbol{X}} = 0 \Leftrightarrow \boldsymbol{X} \equiv \{x_{ijh} = c_{jh}\colon i \in \boldsymbol{I}, j \in \boldsymbol{J}, h \in \boldsymbol{H}\}$, $c_{jh} \in \mathbb{R}$ (real); *it is null when all elements of the three-way (one-way) centered matrix* $\overset{*}{\boldsymbol{X}}$ *are zero.*

Properties (a) and (a′) are required since it is useful to know the total degree of variability of a three-way data set, but such index might be considered too synthetic for a so complex and large data set, therefore, what we also need is what we can call a *super-index* i.e. an index which summarizes the variability of the structures of $\boldsymbol{X}$ (vectors or slabs forming $\boldsymbol{X}$).
Properties (b) and (b′) are also necessary since when the three-way matrix degenerates into a one or two way data matrix, then $\sigma(\boldsymbol{X})$ has to be one of the univariate or multivariate indices of variability already introduced in the statistical literature.

Conditions (1), (2) and (3) can be considered an extension of those held by the univariate indices and it does not seem there are reasons not to preserve them also in the case of a three-way data set.

Conditions (2) and (2′) imply the possibility to express index $\sigma(\boldsymbol{X})$ as a function of the differences of the values x_{ijh} from their mean or from prefixed values.

Condition (3) is a rigid extension of that valid in the univariate case. However when variables are more then one in many other cases it may be believed the variability null. For instance it may be feasible to give null variability to $\boldsymbol{X}$ when its two way slices $\boldsymbol{X}_{..h}$ are linked by linear (or non liner) functions each other. Therefore condition (3) may be changed to satisfy other axioms. However here we do not treat furthermore this important problem.

The indications we set out may be formalized through the following index of *total variation of* $\boldsymbol{X}$, called *variance of* $\boldsymbol{X}$:

$$\sigma(\boldsymbol{X}) = \sum_{h=1}^{r} \sum_{j=1}^{k} \sum_{u=1}^{k} \sigma(\boldsymbol{X}_{.jh}, \boldsymbol{X}_{.uh}), \tag{14}$$

where $\sigma(X_{.jh}, X_{.uh})$ is the covariance between j-th and u-th variables of $X_{..h}$. Therefore $\sigma(X)$ is the sum of the covariances among rk^2 couples of variates $(X_{.jh}, X_{.uh})$, so that properties (a) and (b) directly follow.

Now to prove properties (a') and (b') we have first to consider the strong index of covariance between slabs $X_{..h}$ and $X_{..m}$ introduced by Vichi (1989):

$$\operatorname{cov}(X_{..h}, X_{..m}) = \sum_{j=1}^{k} \sum_{u=1}^{k} \sigma(X_{.jh}, X_{.um}), \tag{15}$$

which is in the case $X_{..h} = X_{..m}$ a multivariate measure of variability. In fact – given $X_{..h} = (X_{.1h}, \ldots, X_{.kh})$ and considered the linear combination of its variables $y_h = a_{1h}X_{.1h} + \cdots + a_{kh}X_{.kh}$, with coefficients a_{jh} – the variability of the k-variate $X_{..h}$ can be computed as the dispersion of the univariate y_h:

$$\sigma(y_h) = \sigma(a_{1h}X_{.1h} + \cdots + a_{kh}X_{.kh}) = \sum_{j=1}^{k} \sum_{u=1}^{k} \sigma(X_{.jh}, X_{.uh}) a_{jh}a_{mh}, \tag{16}$$

from (16) we have $\sigma(Y_h) = \operatorname{cov}(X_{..h}, X_{..h})$, when $a_{jh} = 1$ $(j \in \boldsymbol{J})$; hence $\operatorname{cov}(X_{..h}, X_{..h})$ is a multivariate measure of variability of $X_{..h}$.

From expression (15) the total variation of X can be written:

$$\sigma(X) = \sum_{h=1}^{r} \operatorname{cov}(X_{..h}, X_{..h}) = \operatorname{tr}(\Sigma), \tag{17}$$

that is, the sum of the strong index of covariance $\operatorname{cov}(X_{..h}, X_{..h})$ $\forall\, h \in \boldsymbol{H}$, or the trace of $\Sigma = \{\operatorname{cov}(X_{..h}, X_{..m}): h, m \in \boldsymbol{H}\}$, i.e., the dispersion matrix of X. From (17) properties (a') and (b') follow.

Conditions (1), (2) and (2') follow directly. Also condition (3) is satisfied, excluding the degenerate case $k = 2$ and $r = 1$ (two variables one occasion), i.e., the three-way data set degenerates to a bivariate (X, Y) data set. In this case $\sigma(X) = 1/2 + (1/2)r_{xy}$, (computed without loss of generality, on the standardized variables), is null also when $r_{xy} = -1$. Now for $r > 1$ and $k > 1$, that is, when we actually have a three-way matrix X, it is well known (Naddeo, 1978) that three or more variables cannot have simultaneously Pearson correlation coefficient equal to -1, therefore index (14) vanishes only for condition (3).

In the case we evaluate the variability of X through lateral slabs and column vectors; or through horizontal slabs and tube vectors, we have the indices corresponding to expression (17):

$$\sigma_L(X) = \sum_{j=1}^{k} \sum_{h=1}^{r} \sum_{m=1}^{r} \sigma(X_{.jh}, X_{.jm}), \tag{18}$$

$$\sigma_H(X) = \sum_{i=1}^{n} \sum_{j=1}^{k} \sum_{j=1}^{k} \sigma(X_{ij}, x_{iu.}). \tag{19}$$

Other measures of variability of a three-way matrix X defined through frontal slabs and column vectors; or lateral slabs and column vectors; or horizontal slabs and tube vectors are given considering the two-way variability measures exam-

ined in paragraph 3.2, which are also functions of univariate measures of variability:

$$\sigma_1(\boldsymbol{X}) = \sum_{h=1}^{r} \sigma_1^2(\boldsymbol{X}_a) = \sum_{h=1}^{r} \operatorname{tr}(\boldsymbol{\Sigma}_a), \tag{20}$$

$$\sigma_2(\boldsymbol{X}) = \sum_{h=1}^{r} \sigma_2^2(\boldsymbol{X}_a) = \sum_{h=1}^{r} \operatorname{tr}(\boldsymbol{\Sigma}_a \boldsymbol{\Sigma}_a). \tag{21}$$

Note that when $a = ..h$ (frontal slab) equation (21) is equal to the sum of the numerator of RV coefficient of Escoufier (1973) computed on matrices $\boldsymbol{X}_{..h}$.

4. Data preprocessing

Before performing a three-way analysis on the data matrix $\boldsymbol{X}$, it is often convenient to carry out some basic transformations, or data preprocessing, in order: (1) to return data appropriate for the three-way analysis; (2) to allow to extend the applicability of the adopted three-way model; (3) to turn interval scale data into ratio scale data (centering); (4) to quantify qualitative data; (5) to control the influence of different unit of measurements of the variables and equate differences of variability among variables (standardization); (6) to equate the influence of the correlation within each slab of the three-way data set.

4.1. One-way, two-way and three-way centering of X

Centering is done subtracting to the elements of each structure of the three-way data matrix $\boldsymbol{X}$ (i.e., fibers or slabs, or the three-way matrix itself), their mean so that the resulting data structures have zero mean.

According to Kruskal (1981), Harshman and Lundy (1984) centering can be done in three different ways: fiber or one-way centering, slab or two-way centering, global or three-way centering.

One-way centering is necessary when we need to turn interval scale variables of $\boldsymbol{X}$ into ratio scale data. This is generally done for many multivariate techniques such as principal component analysis, multidimensional scaling and cluster analysis. Two-way centering is useful when variables have different origins between slabs and we want to equate these origins.

4.2. Three-way or global centering

Three-way centering or global-centering transforms the three-way matrix $\boldsymbol{X}$ into a score three-way matrix, subtracting to each element of $\boldsymbol{X}$ the mean $\bar{x}$ computed over all its elements. This type of transformation makes invariant the analyses of data matrices transformed by a unique additive constant.

4.3. Two-way or slab centering of X

Slab-centering transforms each slab of X into score slab, subtracting from the elements of the slab their mean, so that we have:

$$\textit{frontal score matrix}: \quad S_{..h} = X_{..h} - \bar{x}_{..h\,n}\mathbf{1}_k\mathbf{1}' \quad (h \in H);$$

$$\textit{lateral score matrix}: \quad S_{.j.} = X_{.j.} - \bar{x}_{.j.\,n}\mathbf{1}_r\mathbf{1}' \quad (j \in J);$$

$$\textit{horizontal score matrix}: \quad S_{i..} = X_{i..} - \bar{x}_{i..\,k}\mathbf{1}_r\mathbf{1}' \quad (i \in I);$$

In the frontal, lateral and horizontal score slab the sum of their elements is equal to zero.

Even if slab-centering has been considered the most natural kind of centering for X, Harshman and Lundy (1984) noted that in order to apply PARAFAC three-linear model (Harshman, 1970), slab-centering is undesirable for three main reasons: firstly it introduces in the model unwanted constants; secondly it does not remove two-way interaction constants: thirdly it removes only one of the one way constants of PARAFAC.

4.4. One-way or fiber centering

Fiber-centering transforms each fiber of X in vector score, subtracting from the elements of the fiber their mean. Considering slabs and synthesis matrices (paragraph 2.1) we can define the *three-way fiber centered matrices* subtracting from the slab the corresponding (frontal, lateral or horizontal) mean matrices. However, one-way-centering can be computed also utilizing the matrices L_m $(m = n, r, k)$ which are symmetric, indempotent, have rank $m - 1$ and their rows and columns sum to zero. We therefore have three-way fiber centered matrices:

(1) *centered by tube*

$$\begin{aligned} C &= \left\{C_{..h} = X_{..h} - \bar{X}_{..F}\colon h \in H\right\} \\ &= \left\{X_{.j.}L_r = X_{.j.}\left({}_rI - (1/r)_r\mathbf{1}_r\mathbf{1}'\right)\colon j \in J\right\} \\ &= \left\{L_rX_{i..} = \left({}_rI - (1/r)_r\mathbf{1}_r\mathbf{1}'\right)X_{i..}\colon i \in I\right\}, \end{aligned} \tag{22}$$

formed respectively by: r frontal matrices $C_{..h}$, not centered; k lateral slabs centered by rows; n horizontal slabs centered by columns.

Matrix C is *centered by tube*, that is formed by *tube-vector scores*: $C_{ij.} = X_{ij.} - \bar{x}_{ij.\,r}\mathbf{1}$ $(i \in I, j \in J)$ with zero mean. This centering equates the origin of the variables among different units, turning interval scale variables j of unit i into ratio-scale variables.

(2) *centered by row*

$$\begin{aligned} \tilde{C} &= \left\{C_{.j.} = X_{.j.} - \bar{X}_{.L}\colon j \in J\right\} \\ &= \left\{X_{..h}L_k = X_{..h}\left({}_kI - (1/k)_k\mathbf{1}_k\mathbf{1}'\right)\colon h \in H\right\} \\ &= \left\{X_{i..}L_k = X_{i..}\left({}_kI - (1/k)_k\mathbf{1}_k\mathbf{1}'\right)\colon i \in I\right\}, \end{aligned} \tag{23}$$

formed respectively by: k lateral slabs not centered; r frontal slabs centered by rows; n horizontal slabs centered by rows. Matrix $\tilde{C}$ is *centered by row*, that is formed by *row-vector scores*: $C_{i.h} = X_{i.h} - \bar{x}_{i.hk}\mathbf{1}$ $(i \in I, h \in H)$, with zero mean, defining the so called ipsative data. This centering is commonly employed when data are ability tests scores for groups of people.

(3) *centered by column*

$$\begin{aligned}\tilde{\tilde{C}} &= \{C_{i..} = X_{H..} - \bar{X}_{H.}: i \in I\} \\ &= \{L_n X_{..h} = ({}_nI - (1/n)_n\mathbf{1}_n\mathbf{1}')X_{..h}: h \in H\} \\ &= \{L_n X_{.j.} = ({}_nI - (1/n)_n\mathbf{1}_n\mathbf{1}')X_{.j}: j \in J\},\end{aligned} \tag{24}$$

formed respectively by: n horizontal slabs not centered; r frontal slabs centered by columns; k lateral slabs centered by columns.

Matrix $\tilde{\tilde{C}}$ is *centered by column*, that is formed by *column-vector scores*: $C_{.jh} = X_{.jh} - \bar{x}_{.jh\,n}\mathbf{1}$ $(j \in J, h \in H)$, with zero mean.

The main objective of this centering is to turn a given interval-scale variable j of situation h, into a ratio-scale variable, so that to equate the origin of the variables within each occasion h.

Of course if the sum of the elements in each fiber is zero also the sum of the elements of each slab and also the sum of all elements of the three-way matrix is zero. Hence fiber centering implies slab centering and global centering, but not vice-versa.

It is easy to show that successively applying fiber centering across modes i, j and h removes in PARAFAC model all the constant terms due to interval scale variables, centering all the factor loading matrices associated to modes i, j and h.

Another fiber-centering of X can be computed applying double-fiber-centering of each slab, so that to define the following three-way matrices:

$$D = \{L_n X_{..h} L_k = ({}_nI - (1/n)_n\mathbf{1}_n\mathbf{1}')X_{..h}({}_kI - (1/k)_k\mathbf{1}_k\mathbf{1}'): h \in H\}, \tag{25}$$

formed by the *frontal slab double-centered matrices*;

$$\tilde{D} = \{L_r X_{.j.} L_n = ({}_rI - (1/r)_r\mathbf{1}_r\mathbf{1}')X_{..h}({}_nI - (1/n)_n\mathbf{1}_n\mathbf{1}'): j \in J\}; \tag{26}$$

formed by the *lateral slab double-centered matrices*;

$$\tilde{\tilde{D}} = \{L_r X_{i..} L_k = ({}_rI - (1/r)_r\mathbf{1}_r\mathbf{1}')X_{..h}({}_kI - (1/k)_k\mathbf{1}_k\mathbf{1}'): i \in I\}; \tag{27}$$

formed by the *horizontal slab double-centered matrices*.

Note that the double centering frontal slabs modify the correlation between frontal slabs computed through the weak or the strong index of correlation (Vichi, 1989), while this does not happen with the previous one-way centering.

In the double-centered matrices the elements of the rows and the columns sum to zero, therefore double-centering implies slab-centering, but not vice-versa. Obviously if the sum of elements in each slab is zero the sum of the total

elements of X is zero. Therefore double-centering implies global centering but not vice-versa. Double-centering is particularly useful when data are proximities (similarities or dissimilarities).

Double-centering slabs $X_{..h}$ is equivalent to center rows and columns of each $X_{..h}$. In the PARAFAC model, the factors that are constant in either mode j and h vanish.

4.5. One-way, two-way and three-way standardization

Standardization is done centering the elements of each predetermined structure of the three-way data matrix X (i.e. fibers or slabs, or the three-way matrix itself), and rendering the variability of each data structure comparable; for example having variance equal 1.

Standardization can be done in three different ways: fiber or one-way standardization, slab or two-way standardization, global or three-way standardization.

Three-way standardization is necessary when we have to compare two or more three-way data matrices within each one the unit of measurements of the variables remains invariant, but between them are different.

4.6. One-way or fiber standardization

Fiber-standardization transforms vectors of X into fiber-centered scores, which are divided by their standard deviation. Therefore we can define the *three-way fiber standardized matrices*: (1) $Z = \{Z_{..h}: h \in H\}$, i.e., r column standardized frontal slabs formed by k *standardized column vectors* or *columns z-score*

$$Z_{.jh} = \frac{1}{\sigma_{.jh}} C_{.jh} = \frac{1}{\sigma_{.jh}} \left(X_{.jh} - \bar{x}_{.jh\ n}\mathbf{1}\right) \qquad (j \in J,\ h \in H). \tag{28}$$

(2) $\tilde{Z} = \{Z_{.j.}: j \in J\}$, i.e., k column standardized lateral slabs formed by r *columns z-score*;

(3) $\tilde{\tilde{Z}} = \{Z_{i..}: i \in I\}$, i.e., n column standardized horizontal slabs formed by k *standardized tube vectors* or *tubes z-score*

$$Z_{ij.} = \frac{1}{\sigma_{ij.}} C_{ij.} = \frac{1}{\sigma_{ij.}} \left(X_{ij.} - \bar{x}_{ij.\ r}\mathbf{1}\right) \qquad (j \in J,\ i \in I). \tag{29}$$

One-way standardization is necessary when variables of X have different unit of measurements, and we wish to apply Factorial Matrices Analysis (Vichi, 1990, 1991), or Principal Matrices Analysis (paragraph 5).

In each *column z-score* $Z_{.jh}$ and *tube z-score* $Z_{ij.}$ the mean and the standard deviation of each variable j is zero and one respectively.

Of course this type of standardization forces each variable to contribute to the total variability in the same manner. To overcome this problem is necessary

to define new scores whose standard deviation is comparable among variables (independent from the unit of measurements), but not necessary equal to one:

$$\boldsymbol{U}_{.jh} = \frac{1}{\bar{x}_{.jh}}\boldsymbol{C}_{.jh} = \frac{1}{\bar{x}_{.jh}}\boldsymbol{X}_{.jh} - {}_n\mathbf{1} \qquad (j \in \boldsymbol{J},\ h \in \boldsymbol{H}), \tag{30}$$

$$\boldsymbol{U}_{ij.} = \frac{1}{\bar{x}_{ij.}}\boldsymbol{C}_{ij.} = \frac{1}{\bar{x}_{ij.}}\boldsymbol{X}_{ij.} - {}_r\mathbf{1} \qquad (j \in \boldsymbol{J},\ i \in \boldsymbol{I}). \tag{31}$$

In each *standardized column vector* or *column u-score* $\boldsymbol{U}_{jh}$, the mean is zero and the standard deviation of variable j is equal to the Pearson's coefficient of variation. Also for the *standardized tube vector* or *tube u-score* $\boldsymbol{U}_{ij}$, the mean is zero and the standard deviation of variable j is $\sigma_{ij.}/\bar{x}_{ij.}$.

4.7. Two-way or slab standardization

Slab-standardization transforms slabs (multivariate variables) of $\boldsymbol{X}$ into fiber-standardized (u-scores) matrices which elements are divided by their multivariate measures of variability. We can define the *three-way slab standardized matrices*:

$$(1)\ \boldsymbol{V} = \left\{ \boldsymbol{V}_{..h} = \frac{1}{\sigma_1(\boldsymbol{X}_{..h})}\boldsymbol{U}_{..h} \colon h \in \boldsymbol{H} \right\}, \tag{32}$$

compound by r frontal slabs formed by *columns v-score*

$$\boldsymbol{V}_{.jh} = \frac{1}{\sigma_1(\boldsymbol{X}_{..h})}\boldsymbol{U}_{.jh} = \frac{1}{\sqrt{\sum_{j=1}^{k} \sigma_{.jh}^2}} \left(\frac{\boldsymbol{X}_{.jh}}{\bar{x}_{.jh}} - 1 \right) \qquad (j \in \boldsymbol{J},\ h \in \boldsymbol{H}), \tag{33}$$

with zero mean and variance equal to $\sigma_{.jh}^2 / \sum_{j=1}^{k} \sigma_{.jh}^2$.

$$(2)\ \tilde{\boldsymbol{V}} = \left\{ \boldsymbol{V}_{.j.} = \frac{1}{\sigma_1(\boldsymbol{X}_{.j.})}\boldsymbol{U}_{.j.} \colon j \in \boldsymbol{J} \right\}, \tag{34}$$

compound by k lateral slabs formed by *columns v-score*

$$\tilde{\boldsymbol{V}}_{.jh} = \frac{1}{\sigma_1(\boldsymbol{x}_{..h})}\boldsymbol{U}_{.jh} = \frac{1}{\sqrt{\sum_{k=1}^{z} \sigma_{.jh}^2}} \left(\frac{\boldsymbol{X}_{.jh}}{\bar{x}_{.jh}} - 1 \right) \qquad (j \in \boldsymbol{J},\ h \in \boldsymbol{H}), \tag{35}$$

with zero mean and variance equal to $\sigma_{.jh}^2 / \sum_{h=1}^{r} \sigma_{.jh}^2$.

$$(3)\ \tilde{\tilde{\boldsymbol{V}}} = \left\{ \boldsymbol{V}_{i..} = \frac{1}{\sigma_1(\boldsymbol{X}_{i..})}\boldsymbol{U}_{i..} \colon i \in \boldsymbol{I} \right\}, \tag{36}$$

consisted of n horizontal slabs formed by *tube v-scores*

$$\boldsymbol{V}_{ij.} = \frac{1}{\sigma_1(\boldsymbol{x}_{i..})}\boldsymbol{U}_{ij.} = \frac{1}{\sqrt{\sum_{j=1}^{k} \sigma_{ij.}^2}} \left(\frac{\boldsymbol{X}_{ij.}}{\bar{x}_{ij.}} - 1 \right) \qquad (j \in \boldsymbol{J},\ i \in \boldsymbol{I}), \tag{37}$$

with zero mean and variance equal to $\sigma_{ij}^2 / \sum_{j=1}^{k} \sigma_{ij}^2$.
Slab standardization is recommended when in each slab variables have equal unit of measurement, but different between slabs. In this case two-way standardization equate the variability between slabs.

4.8. *Two-way full standardization*

Often, when we consider three-way three mode matrices $\boldsymbol{X}$, it is necessary to examine the relations between matrices that remain invariant under the elimination of the internal relations between variables of each frontal slab. This was one of the aim of Hotelling (1936) when he proposed canonical correlation. In these cases we have to apply a specific transformation (rotation).

Given a slab $\boldsymbol{X}_a$, where $a = ..h$ (frontal), $a = .j.$ (lateral), $a = i..$ (horizontal), *full standardization* of $\boldsymbol{X}_a$ is the procedure that allows to define the *fully standardized slab* $\boldsymbol{R}_a$ with: mean of the elements of the column-vectors of $\boldsymbol{R}_a$ null; variance unitary; and correlation coefficient $\tau_{ju} = 0$ for $j \neq u$ $(j, u \in \boldsymbol{J})$.

The fully standardized slab can be defined:

$$\boldsymbol{R}_a = \boldsymbol{Z}_a \boldsymbol{Q}_a \boldsymbol{\Omega}_a^{-1/2}, \qquad \text{where} \tag{38}$$

$\boldsymbol{Q}_a$ is the matrix the normalized eigenvectors of $\boldsymbol{Z}_a' \boldsymbol{Z}_a$;
$\boldsymbol{\Omega}_a$ is the diagonal matrix with diagonal elements equal to the eigenvalues of $\boldsymbol{Z}_a' \boldsymbol{Z}_a$;

Note that given two fully standardized frontal slabs $\boldsymbol{R}_{..h}$ and $\boldsymbol{R}_{..m}$ for the RV coefficient of Escoufier we have: $\mathrm{RV}(\boldsymbol{R}_{..h}, \boldsymbol{R}_{..m}) = 1$.
In fact, the dispersion matrix of $\boldsymbol{R}_{..v}$ is $(1/n)\boldsymbol{R}_{..v}' \boldsymbol{R}_{..v} = \boldsymbol{I}$, $v = h$, m, and hence $\mathrm{RV} = \mathrm{tr}(\boldsymbol{I})/\mathrm{tr}(\boldsymbol{I}) = 1$, therefore such transformation, according to RV maximize the similarity between the two configurations corresponding to $\boldsymbol{R}_{..h}$ and $\boldsymbol{R}_{..m}$. Differently for the weak and strong correlation coefficients (Vichi, 1989) the full standardization does not necessary imply that $\mathrm{cor}(\boldsymbol{R}_{..h}, \boldsymbol{R}_{..m}) = 1$.

5. Principal matrices analysis

We now show that Principal Matrices Analysis in the case of quantitative variables (Rizzi, 1989), defines factorial data structures insuring minimum loss of information according to a three-way variability index considered in paragraph 3.3.

When a large set of k-variates $\boldsymbol{X}_{..h}$ $(h = 1, \ldots, r)$, arranged into a three-way matrix $\boldsymbol{X}$ are studied, it may be interesting to inquire whether this set can be replaced by a smaller not directly observable set of k-variates $\boldsymbol{Y}_{..h}$ $(h = 1, \ldots, r';\ r' < r)$, forming the three-way matrix $\boldsymbol{Y}$, so that to insure the minimum loss of variability of $\boldsymbol{X}$.

This problem is equivalent to find $\boldsymbol{Y}_{..g}$ $(g = 1, \ldots, r)$, which explain, one after the other, the maximum variability, provided that the variability of $\boldsymbol{Y}$ (when

$r' = r$) is equal to the variability of X (see section 3.3 for the three-way variability of X). It is also useful to require that the information summarized by each $Y_{..g}$ be independent on the others $Y_{..h}$, that is to have $Y_{..g}$ and $Y_{..l}$ uncorrelated according a measure of correlation between matrices.

Rizzi (1989) defines, in the case of quantitative variables, the principal matrices:

$$Y_g = \sum_{h=1}^{r} a_{hg} X_{..h}, \qquad \text{such that} \tag{39}$$

$$\sigma_3(Y_{..g}, Y_{..g}) = \frac{1}{k^2} \sum_{h=1}^{r} \sum_{m=1}^{r} \sigma(X_{.jh}, X_{.jm}) a_{hg} a_{mg} = \max, \tag{40}$$

subject to constraints

$$a'_g a_g = 1$$

$$\sigma_3(Y_{..f}, Y_{..l}) = \sum_{h=1}^{r} \sum_{m=1}^{r} \operatorname{cov}(X_{..h}, X_{..m}) a_{hf} a_{ml} = 0 \text{ for } f \neq l;\ f, l = 1, \ldots, g \tag{41}$$

Now $\sigma_3(Y_{..g}, Y_{..g})$ is the weak index of covariance computed on matrix $Y_{..g}$. It coincides with the strong index of covariance $\sigma(Y_{..g})$ – that has been already proved to be an index of multivariate variability (paragraph 3.3) – when $\sigma(X_{.jh}, X_{.um}) = 0\ \forall\ j, u \in J\ (j \neq u)$ so that $\sigma_3(Y_{..g}, Y_{..g})$ measures an amount of total variation of X when the covariance between different variables of $X_{..h}$ and $X_{..m}$ $\forall h, m$ are null.

Therefore the g-th principal matrix is the linear combination of slabs $X_{..h}$ ($h \in H$) with normalized coefficients which summarizes the g-th widest amount of the total variation of X when we can suppose that between matrices $X_{..h}$ and $X_{..m}$ the correlation logically correct and worthy to be studied is that between the same variables while the correlation between different variables can be supposed null.

Rizzi and Vichi (1991) show that the coefficient vectors a_g $g = 1, \ldots, r$ are the normalized eigenvectors corresponding to the g-th largest eigenvalue α_g of the matrix Σ, which elements are the weak covariances between $X_{..h}$ and $X_{..m}$.

Furthermore we have that the total variation of Y, when $r' = r$ (i.e. $Y = \{Y_{..1}, \ldots, Y_{..r}\}$):

$$\sigma(Y) = \frac{1}{r} \sum_{g=1}^{r} \sigma_3(Y_{..g}, Y_{..g}) = \frac{1}{r} \sum_{g=1}^{r} \alpha_g = \frac{1}{r} tr(a_g \Sigma a'_g) = \frac{1}{r} tr(\Sigma) = \sigma(X),$$

is equal to the total variation of X, where for the last equality we suppose $\sigma_3(Y_{..g}, Y_{..g}) = \sigma(Y_{..g})$.

Finally $Y_{..g}$ is uncorrelated with each of the preceding $g - 1$ principal matrices; where the correlation between matrices $Y_{..l}$ and $Y_{..f}$ ($l, f \in H$) is measured with the weak index of covariance.

6. Summary

In the first section of this paper we describe the structures (vectors and matrices) on which a three-way data set X can be organized, and the information we can point out when using these structures. Many three-way analyses are based on pooled representations of X, that are systematically studied. The information given by a three-way data set can be synthesized according to the structures utilized to represent X. In the second section we define one-way, two-way and three-way syntheses of X. Also the variability of a three-way data set is evaluated, in section three, according three different levels: one-way or fiber variability, two-way or slab variability and three-way variability. The syntheses and the variability indices of X can be used for data preprocessing of X, which is here discussed in section four. Furthermore we discuss, in Section 5, the Principal Matrices Analysis on the base of three-way variability indices.

Acknowledgements

The Authors share the responsibility for the content of this paper. However Sections 4 and 5 are due to Alfredo Rizzi, while Sections 1, 2 and 3 are due to Maurizio Vichi.

References

Amato, V. (1981), Variabilità multidimensionale e le sue misure, *Atti del convegno 1981 della Società Italiana di Statistica*, Pavia-Salice Terme, 57–67.

Coppi, R. (1986), Analysis of Three-Way Data Matrices Based on Pairwise Relation Measures, *Procedings in Computational Statistics*, Edited by F. De Antoni, N. Lauro and A. Rizzi, Physica – Verlag, 1986.

D'Alessio, G. (1986), L'analisi di successioni di matrici di dati qualitativi, *Tesi di Dottorato di Ricerca*, Statistica Metodologica, Università La Sapienza, Roma.

D'Ambra, L. and Marchetti G.M. (1986), Un metodo per l'analisi interstrutturale di più matrici basato su misure di relazione tra le unità statistiche, *Atti della XXXIII riunione scientifica della SIS*, Cacucci, Bari.

Escofier, B. and Pages J. (1984), L'analyse factorielle multiple, Cohiers du B.U.R.O., *Serie Recherche*, **42**, *Université Pierre et Marie Curie*, Paris.

Escofier, B. and Pages J. (1989), Multiple factor analysis: results of a three-year utilization. Multiway Data Analysis, edited by R. Coppi, S. Bolasco, 277–285.

Escoufier, Y. (1977), Le traitement des variables vectorielles, *Biometrics*, **29**, 751–760.

Escoufier, Y. (1977), Operators related to a data matrix, in *Recent Developments in Statistics*, ed. J.R. Barra et al., North-Holland Publishing Company, 125–131.

Escoufier, Y. (1980), Exploratory Data Analysis when data are matrices, in *Recent developments in statistical inference and data analysis*, North Holland, Amsterdam, 45–53.

Escoufier, Y. (1987), Three-mode data analysis: the STATIS method. in Methods for Multidimensional Data Analysis. ECAS., 325–338.

Flury, B.N. (1984), Common Principal Components on k Groups. *J. American Statistical Association*, **79**, 892–898.

Gower, J.C. and Legendre, P. (1986), Metric and Euclidean properties of Dissimilarity Coefficents, *Journal of Classification*, **3**, 5–48.

Harshman, R.A. and Lundy, M.E. (1984), Data Preprocessing and the Extended PARAFAC Model, *Research Methods for Multivariate Data Analysis*, H.G. Law et al eds, New York: Praeger.

Hotelling, H. (1936), Relations between two sets of variates, *Biometrika*, **28**, 321–346.

L'Hermier des Plantes, H. (1976), Structurasion des tableaux à trois indices de la statistique, *Thêse de-eme cycle*, Université de Montpellier.

Kiers, H.A.L. (1990), A program for simultaneous components analysis of variables measured in two or more populations. University of Groningen.

Kiers, H.A.L. and Ten Berge J.M.F. (1989), Alternating least squares algorithms for Simultaneous Components Analysis with equal component weight matrices in two or more populations. *Psychometrika*, **54**, 467–473.

Kruskal, J.B. (1981), Multilinear models for data analysis. *Behaviormetrika*

Levin, J. (1966), Simultaneous factor analysis of several gramian matrices, *Psychometrika*, **31**, 413–419.

Lunetta, G. (1973), *Variabilità a più dimensioni e analisi dei gruppi (cluster analysis)*, Catania.

Lunetta, G. (1981), Su alcuni aspetti della variabilità statistica a più dimensioni, *Atti del convegno 1981 della Società Italiana di Statistica*, Pavia-Salice Terme, 37–56.

Millsap, R.E. and Meredith W. (1988), Component analysis in cross-sectional and longitudinal data. *Psychometrika*, **53**, 123–134.

Naddeo A. (1978), *Statistica di Base*, Kappa, Roma.

Pieri, L. and Vichi, M. (1990), Le matrici a tre indici e le loro sintesi rappresentative, *Rivista di Statistica Applicata*, **3**, n. 2, 145–173.

Ramsay, J.O., Ten Berge, J., Styan, G.P.H. (1984). Matrix correlation, *Psychometrika*, 403–423.

Rizzi, A. (1989), On the synthesis of three-way data matrices, *Presented at the International Meeting "Multiway'88"*, 28–30 March, Rome.

Rizzi, A. (1989a), Clustering per le matrici a tre vie, *Statistica*,

Rizzi, A. Vichi M. (1992), Relations between sets of variates of a three-way data set, *Rivista di Statistica Applicata*.

Robert, P. and Escoufier Y. (1976), A unifying tool for linear multivariate statistical methods: the Rv coefficient, *Applied Statistics*, **25**, n. 3, 257–265.

Ten Berge, J.M.F., Kiers, H.A.L., Van der Stel V. (1992), Simultaneous components analysis, *Rivista di Statistica Applicata*, **4**, n. 3.

Tucker, L.R. (1966), Some Mathematical Notes on Three-Mode Factor Analysis, *Psychometrika*, **31**, n. 3, 279–311.

Tucker, L.R. (1972), Relations between multidimensional scaling and three-mode factor analysis. *Psychometrika*, **37**, 3–27.

Vichi, M. (1988), Two way data matrix representative synthesis of a three way data matrix, *Statistica*, **1**, 2, 91–106.

Vichi, M. (1989), La connessione e la correlazione tra due matrici dei dati componenti una matrice a tre indici. *Statistica*, **1**, 225–243.

Vichi, M. (1990), L'analisi in matrici fattoriali di una matrice a tre indici, *Statistica*, **1**, n. 4, 525–546.

Vichi, M. (1991), Le tecniche che derivano dall'analisi in matrici fattoriali di una matrice a tre indici, *Statistica*, **1**, 53–77.