Multiple Factor Analysis
MFA

Analisi Statistica dei Dati Multidimensionali[1]

[1]Corso di Laurea in Scienze Statistiche e Attuariali

Facoltà di Scienze Economiche e Aziendali
Università degli Studi del Sannio

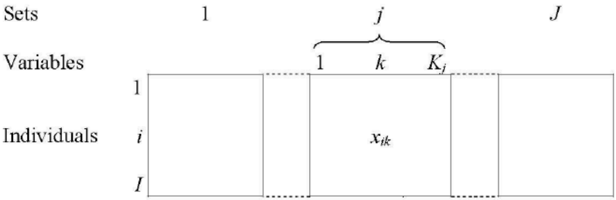
Fonte: Hervé Abdi, Dominique Valentin, "Multiple Factor Analysis"
In: Neil Salkind (Ed.) (2007). Encyclopedia of Measurement and Statistics.
Thousand Oaks (CA): Sage.

- Multiple factor analysis (MFA, see Escofier and Pagès, 1990, 1994) analyzes observations described by several "blocks" or sets of variables.
- MFA seeks the common structures present in all or some of these sets.
- MFA is performed in two steps.

1. First a principal component analysis (PCA) is performed on each data set which is then "normalized" by dividing all its elements by the square root of the first eigenvalue obtained from of its PCA.
2. Second, the normalized data sets are merged to form a unique matrix and a global PCA is performed on this matrix. The individual data sets are then projected onto the global analysis to analyze communalities and discrepancies.

- MFA is used in very different domains such as sensory evaluation, economy, ecology, and chemistry.

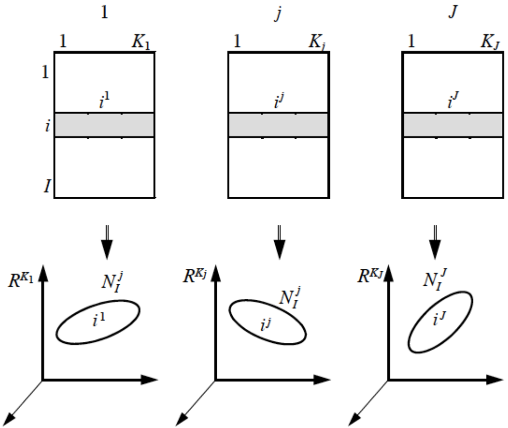
Groups of variables (MFA)



Groups of variables are quantitative and/or qualitative

Objectives :
- study the link between the sets of variables
- balance the influence of each group of variables
- give the classical graphs but also specific graphs: groups of variables - partial representation

Examples :
- Genomic: DNA, protein
- Sensory analysis: sensorial, physico-chemical
- Comparison of coding (quantitative / qualitative)

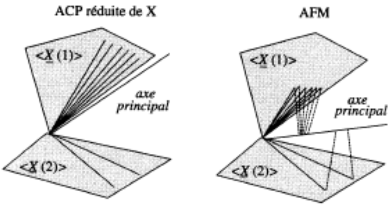

1
j
J
1
K_1
1
K_j
1
K_J
1
i^1
i^j
i^J
i
I
R^{K_1}
N_I^j
i^1
R^{K_j}
N_I^j
i^j
R^{K_J}
N_I^J
i^J

- The goal of MFA is then to integrate different groups of variables describing the same observations.
- In order to do so, the first step is to make these groups of variables comparable.
- Such a step is needed because the straightforward analysis obtained by concatenating all variables would be dominated by the group with the strongest structure.
- A similar problem can occur in a non-normalized PCA: without normalization, the structure is dominated by the variables with the largest variance.
- For PCA, the solution is to normalize (i.e., to use Z-scores) each variable by dividing it by its standard deviation.

- The solution proposed by MFA is similar to PCA: To compare groups of variables, each group is normalized by dividing all its elements by a quantity called its first singular value which is the matrix equivalent of the standard deviation.
- Practically, This step is implemented by performing a PCA on each group of variables.
- The first singular value is the square root of the first eigenvalue of the PCA.
- Note that because the subtables have previously been centered and normalized with their first singular value, the global matrix $\mathbf{X}$ is centered but it is not normalized (i.e., columns from different subtables have, in general, different norms).

- 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.
- This weighting can be easily interpreted: considering the two clouds 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.

Balancing the sets of variables



- After normalization, the data tables are concatenated into a data table

$$\mathbf{X} = \left[\; \frac{1}{\sqrt{\lambda_1^1}}\mathbf{X}_1, \quad \frac{1}{\sqrt{\lambda_1^2}}\mathbf{X}_2, \quad \ldots, \quad \frac{1}{\sqrt{\lambda_1^J}}\mathbf{X}_J \; \right]$$

which is submitted to PCA. This amounts to computing the singular value decomposition of the global data matrix

$$\boxed{\mathbf{X} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{V}^T}$$

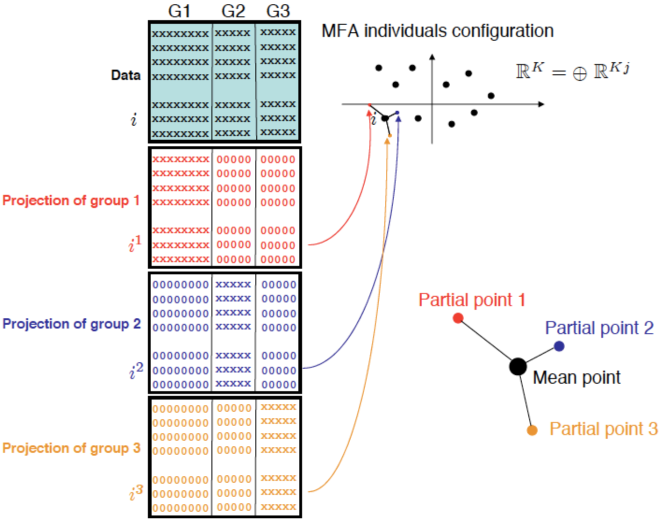
with $\mathbf{U}^T\mathbf{U} = \mathbf{I}$ and $\mathbf{V}^T\mathbf{V} = \mathbf{I}$, where $\mathbf{U}$ and $\mathbf{V}$ are the left and right singular vectors of $\mathbf{X}$ and $\boldsymbol{\Lambda}$ is the diagonal matrix of the singular values.

- The global factor scores for the statistical units (with weights $\mathbf{D}$) are obtained as:

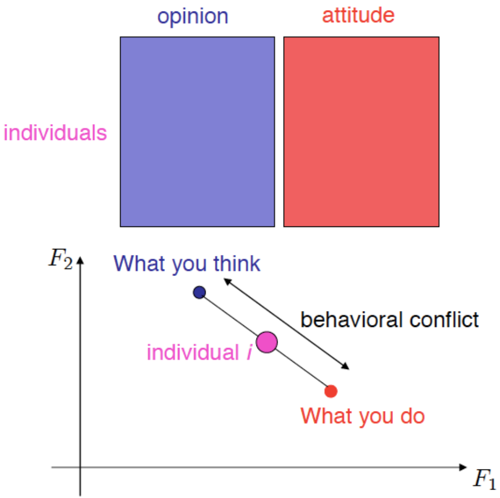
$$\mathbf{F} = \mathbf{D}^{-\frac{1}{2}}\mathbf{U}\boldsymbol{\Lambda}$$

- The global analysis reveals the common structure of the statistical units space.
- In addition, we want to see how each set of variables is related to this space.
- This is achieved by projecting the data set of each set of variables onto the global analysis.

Projection of partial points



Partial points



- This is implemented by multiplication of a cross product matrix by a projection matrix. The projection matrix is obtained by rewriting the global factor scores formula, to show that the global factor scores could be computed as

$$
\begin{aligned}
\boxed{\mathbf{F} = \mathbf{D}^{-\frac{1}{2}}\mathbf{U}\boldsymbol{\Lambda}} &= \mathbf{D}^{-\frac{1}{2}}\mathbf{U}\boldsymbol{\Lambda}\overbrace{\mathbf{V}^{T}\mathbf{V}\boldsymbol{\Lambda}\mathbf{U}^{T}\mathbf{U}\boldsymbol{\Lambda}^{-1}}^{=\mathbf{I}} \\
&= \mathbf{D}^{-\frac{1}{2}}\underbrace{\mathbf{U}\boldsymbol{\Lambda}\mathbf{V}^{T}}_{=\mathbf{X}}\underbrace{\mathbf{V}\boldsymbol{\Lambda}\mathbf{U}^{T}}_{=\mathbf{X}^{T}}\mathbf{U}\boldsymbol{\Lambda}^{-1} \\
&= \underbrace{\mathbf{D}^{-\frac{1}{2}}\mathbf{X}\mathbf{X}^{T}}_{\textit{symmetric}}\mathbf{U}\boldsymbol{\Lambda}^{-1} \\
&\boxed{= (\mathbf{X}\mathbf{X}^{T}) \times (\mathbf{D}^{-\frac{1}{2}}\mathbf{U}\boldsymbol{\Lambda}^{-1})}
\end{aligned}
$$

- This shows that $\mathbf{P} = \mathbf{D}^{-\frac{1}{2}}\mathbf{U}\boldsymbol{\Lambda}^{-1}$ is a projection matrix which forms the matrix $\mathbf{X}\mathbf{X}^{T}$ into factor scores.

- The projection matrix $\mathbf{P} = \mathbf{D}^{-\frac{1}{2}}\mathbf{U}\mathbf{\Lambda}^{-1}$ is then used to project the studies onto the global space. For example, for the first set of variables we obtain

$$\mathbf{F}_{[1]} = T \times (\mathbf{X}_{[1]}\mathbf{X}_{[1]}^{T})\mathbf{P}$$

where T is the global number of variables

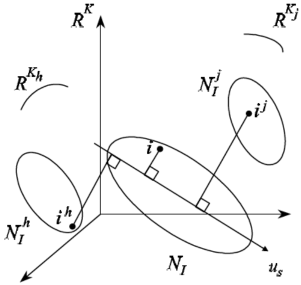


Figure 3: Principle of the superimposed representations provided by MFA
Each partial cloud N_I^j is projected onto main axes of the mean cloud N_I

- As in standard PCA, the variable loadings are the correlation between the original variables and the global factor scores
- MFA starts with a series of PCA's. Their relationship with the global analysis is explored by computing loadings (i.e., correlations) between the components of each studies and the components of the global analysis.
- These loadings relate the original PCA and the global analysis.

- The relationships between the studies and between the studies and the global solution are analyzed by computing the partial inertia of each study for each dimension of the global analysis.
- This is computed, for each study, as the sum of the squared projections of the variables on the right singular vectors of $\mathbf{X}$ multiplied by the corresponding eigenvalue.
- Because the singular vectors are normalized, the sum of the partial inertias for all the studies for a given dimension is equal to its eigenvalue.

Similarity between two groups

Measure of similarity between groups K_j and K_m:

$$\mathcal{L}_g(K_j, K_m) = \sum_{k \in K_j} \sum_{l \in K_m} cov^2 \left(\frac{x_{.k}}{\sqrt{\lambda_1^k}}, \frac{x_{.l}}{\sqrt{\lambda_1^l}} \right)$$

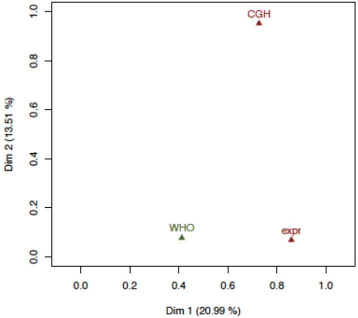
MFA = weighted PCA $\Rightarrow$ first principal component of MFA maximizes

$$\sum_{j=1}^{J} \mathcal{L}_g(v_1, K_j) = \sum_{j=1}^{J} \sum_{k \in K_j} cov^2 \left(\frac{x_{.k}}{\sqrt{\lambda_1^j}}, v_1 \right)$$

Inertia of K_j projected on v_1

Representation of the groups

Group j has the coordinates $(\mathcal{L}_g(v_1, K_j), \mathcal{L}_g(v_2, K_j))$



- 2 groups are all the more close that they induce the same structure
- The 1st dimension is common to all the groups
- 2nd dimension mainly due to CGH

$$0 \leq \mathcal{L}_g(v_1, K_j) = \frac{1}{\lambda_1^j} \underbrace{\sum_{k \in K_j} cov^2(x_{.k}, v_1)}_{\leq \lambda_1^j} \leq 1$$

- To illustrate MFA, we selected six wines, coming from the same harvest of Pinot Noir, aged in six different barrels made with one of two different types of oak.
- Wines 1, 5, and 6 were aged with the first type of oak, and wines 2, 3, and 4 with the second. Next, we asked each of three wine experts to choose from two to five variables to describe the six wines.
- For each wine, the expert rated the intensity of the variables on a 9-point scale.
- The goal of the analysis is twofold. First we want to obtain a typology of the wines and second we want to know if there is an agreement between the experts.

Table 1: Raw data for the wine example

		Expert 1			Expert 2				Expert 3		
wines	Oak-type	fruity	woody	coffee	red fruit	roasted	vanillin	woody	fruity	butter	woody
$wine_1$	1	1	6	7	2	5	7	6	3	6	7
$wine_2$	2	5	3	2	4	4	4	2	4	4	3
$wine_3$	2	6	1	1	5	2	1	1	7	1	1
$wine_4$	2	7	1	2	7	2	1	2	2	2	2
$wine_5$	1	2	5	4	3	5	6	5	2	6	6
$wine_6$	1	3	4	4	3	5	4	5	1	7	5

$$\mathbf{X}_{[1]} = \begin{bmatrix} -0.57 & 0.58 & 0.76 \\ 0.19 & -0.07 & -0.28 \\ 0.38 & -0.50 & -0.48 \\ 0.57 & -0.50 & -0.28 \\ -0.38 & 0.36 & 0.14 \\ -0.19 & 0.14 & 0.14 \end{bmatrix} \quad \mathbf{X}_{[2]} = \begin{bmatrix} -0.50 & 0.35 & 0.57 & 0.54 \\ 0.00 & 0.05 & 0.03 & -0.32 \\ 0.25 & -0.56 & -0.51 & -0.54 \\ 0.75 & -0.56 & -0.51 & -0.32 \\ -0.25 & 0.35 & 0.39 & 0.32 \\ -0.25 & 0.35 & 0.03 & 0.32 \end{bmatrix} \quad \mathbf{X}_{[3]} = \begin{bmatrix} -0.03 & 0.31 & 0.57 \\ 0.17 & -0.06 & -0.19 \\ 0.80 & -0.61 & -0.57 \\ -0.24 & -0.43 & -0.38 \\ -0.24 & 0.31 & 0.38 \\ -0.45 & 0.49 & 0.19 \end{bmatrix}$$

The data were centered by column and normalized

- To normalize the studies, we first compute a PCA for each study. The first singular value (i.e., the square root of the first eigenvalue) is the normalizing factor used to divide the elements of the data table.
- For example, the PCA of the first group gives a first eigenvalue $\lambda_1^1 = 2.86$ and a first singular value of $\sqrt{\lambda_1^1} = 1.69$. This gives the first normalized data matrix:

$$\mathbf{X}_{[1]} = \begin{bmatrix} -0.33 & 0.34 & 0.45 \\ 0.11 & -0.04 & -0.16 \\ 0.22 & -0.30 & -0.28 \\ 0.33 & -0.30 & -0.16 \\ -0.22 & 0.21 & 0.08 \\ -0.11 & 0.08 & 0.08 \end{bmatrix}$$

- Matrices $\mathbf{X}_2$ and $\mathbf{X}_3$ are normalized with their first respective singular values of $\lambda_1^2 = 1.91$ and $\lambda_1^3 = 1.58$. Normalized matrices have a first singular value equal to 1.

- Global matrix **X**

$$\left[\begin{array}{ccc|cccc|ccc} -0.33 & 0.34 & 0.45 & -0.26 & 0.18 & 0.30 & 0.28 & -0.02 & 0.19 & 0.36 \\ 0.11 & -0.04 & -0.16 & 0.00 & 0.03 & 0.02 & -0.17 & 0.11 & -0.04 & -0.12 \\ 0.22 & -0.30 & -0.28 & 0.13 & -0.29 & -0.27 & -0.28 & 0.51 & -0.39 & -0.36 \\ 0.33 & -0.30 & -0.16 & 0.39 & -0.29 & -0.27 & -0.17 & -0.15 & -0.27 & -0.24 \\ -0.22 & 0.21 & 0.08 & -0.13 & 0.18 & 0.20 & 0.17 & -0.15 & 0.19 & 0.24 \\ -0.11 & 0.08 & 0.08 & -0.13 & 0.18 & 0.02 & 0.17 & -0.29 & 0.31 & 0.12 \end{array}\right]$$

- Computing the global PCA. For our example we obtain:

$$\mathbf{U} = \begin{bmatrix} 0.53 & -0.35 & -0.58 & -0.04 & 0.31 \\ -0.13 & -0.13 & 0.49 & 0.51 & 0.54 \\ -0.56 & -0.57 & 0.01 & -0.36 & -0.25 \\ -0.44 & 0.62 & -0.48 & 0.15 & 0.03 \\ 0.34 & 0.04 & 0.16 & 0.39 & -0.73 \\ 0.27 & 0.40 & 0.40 & -0.65 & 0.11 \end{bmatrix}$$

$$\operatorname{diag}\{\boldsymbol{\Delta}\} = \begin{bmatrix} 1.68 & 0.60 & 0.34 & 0.18 & 0.11 \end{bmatrix}$$

$$\text{and } \operatorname{diag}\{\boldsymbol{\Lambda}\} = \operatorname{diag}\{\boldsymbol{\Delta}^2\} = \begin{bmatrix} 2.83 & 0.36 & 0.11 & 0.03 & 0.01 \end{bmatrix}$$

$$\mathbf{F} = \mathbf{M}^{-\frac{1}{2}}\mathbf{U}\boldsymbol{\Delta}$$

$$= \begin{bmatrix} 2.18 & -0.51 & -0.48 & -0.02 & 0.08 \\ -0.56 & -0.20 & 0.41 & 0.23 & 0.15 \\ -2.32 & -0.83 & 0.01 & -0.16 & -0.07 \\ -1.83 & 0.90 & -0.40 & 0.07 & 0.01 \\ 1.40 & 0.05 & 0.13 & 0.17 & -0.20 \\ 1.13 & 0.58 & 0.34 & -0.29 & 0.03 \end{bmatrix}$$

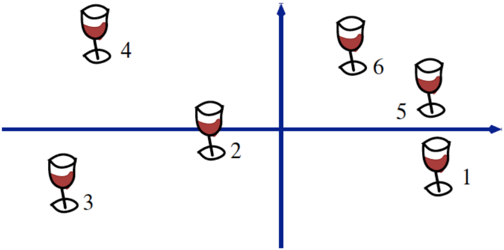


Figure 1: Global analysis: Plot of the wines on the first two principal components. First component: $\lambda_1 = 2.83$, explains 84% of the inertia, Second component: $\lambda_2 = 2.83$, explains 11% of the inertia.

Partial analyses

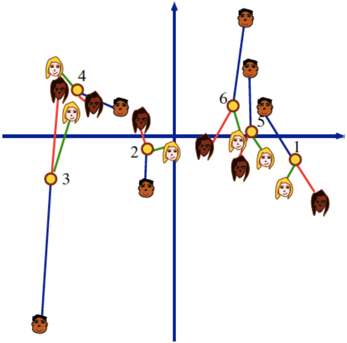


Figure 2: Projection of the experts onto the global analysis. Experts are represented by their faces. A line segment links the position of the wine for a given expert to its global position. First component: $\lambda_1 = 2.83$, explains 84% of the inertia, Second component: $\lambda_2 = 2.83$, explains 11% of the inertia.

The original variables and the global analysis

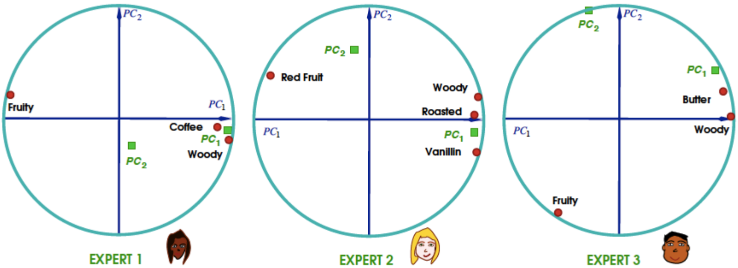


Figure 3: Circle of correlations for the original variables.

Analyzing the between study structure

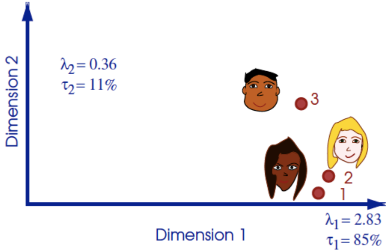


Table 3: Partial inertias for the first three dimensions.

	Axis 1	Axis 2	Axis 3	Axis 4	Axis 5
Expert 1	0.96	0.03	0.05	0.01	0.01
Expert 2	0.98	0.06	0.04	0.02	0.00
Expert 3	0.90	0.28	0.03	0.00	0.00
Σ	2.83	.36	.12	.03	.01
	λ_1	λ_2	λ_2	λ_4	λ_5

For example, for study one, and component one, the partial inertia is obtained as

$$\lambda_1 \times \sum_{j}^{J_k} q_{j,1}^2 = 2.83 \times \left[(-.34)^2 + (.35)^2 + (.32)^2\right] = 2.83 \times .34 = .96$$