

# Multiple Factor Analysis (MFA)

Hervé Abdi<sup>1</sup> & Dominique Valentin

## 1 Overview

### 1.1 Origin and goal of the method

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. 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. 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.

---

<sup>1</sup>In: Neil Salkind (Ed.) (2007). *Encyclopedia of Measurement and Statistics*. Thousand Oaks (CA): Sage.

Address correspondence to: Hervé Abdi

Program in Cognition and Neurosciences, MS: Gr.4.1,

The University of Texas at Dallas,

Richardson, TX 75083-0688, USA

E-mail: [herve@utdallas.edu](mailto:herve@utdallas.edu) <http://www.utd.edu/~herve>

## 1.2 When to use it

MFA is used to analyze a set of observations described by several groups of variables. The number of variables in each group may differ and the nature of the variables (nominal or quantitative) can vary from one group to the other but the variables should be of the same nature in a given group. The analysis derives an integrated picture of the observations and of the relationships between the groups of variables.

## 1.3 The main idea

The goal of MFA is 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 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. After normalization, the data tables are concatenated into a data table which is submitted to PCA.

## 2 An example

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 results are presented in Table 1 (the same example is used in the entry for STATIS). 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.

### 3 Notations

The raw data consist in  $T$  data sets. Each data set is called a *study*. Each study is an  $I \times J_{[t]}$  rectangular data matrix denoted  $\mathbf{Y}_{[t]}$ , where  $I$  is the number of observations and  $J_{[t]}$  the number of variables of the  $t$ -th study. Each data matrix is, in general, preprocessed (*e.g.*, centered, normalized) and the preprocessed data matrices actually used in the analysis are denoted  $\mathbf{X}_{[t]}$ .

For our example, the data consist in  $T = 3$  studies. The data (from Table 1) were centered by column (*i.e.*, the mean of each column is zero) and normalized (*i.e.*, for each column, the sum of the squared elements is equal to 1). So, the starting point of the analysis consists in three matrices:

$$\mathbf{X}_{[1]} = \begin{bmatrix} -0.57 & 0.58 & 0.76 \\ 0.19 & -0.07 & -0.28 \\ 0.38 & -0.50 & -0.48 \\ 0.57 & -0.50 & -0.28 \\ -0.38 & 0.36 & 0.14 \\ -0.19 & 0.14 & 0.14 \end{bmatrix},$$

$$\mathbf{X}_{[2]} = \begin{bmatrix} -0.50 & 0.35 & 0.57 & 0.54 \\ 0.00 & 0.05 & 0.03 & -0.32 \\ 0.25 & -0.56 & -0.51 & -0.54 \\ 0.75 & -0.56 & -0.51 & -0.32 \\ -0.25 & 0.35 & 0.39 & 0.32 \\ -0.25 & 0.35 & 0.03 & 0.32 \end{bmatrix},$$

and

Table 1: Raw data for the wine example

wines	Oak-type	Expert 1			Expert 2			Expert 3			
		fruity	woody	coffee	red fruit	roasted	vanillin	woody	fruity	butter	woody
wine1	1	1	6	7	2	5	7	6	3	6	7
wine2	2	5	3	2	4	4	4	2	4	4	3
wine3	2	6	1	1	5	2	1	1	7	1	1
wine4	2	7	1	2	7	2	1	2	2	2	2
wine5	1	2	5	4	3	5	6	5	2	6	6
wine6	1	3	4	4	3	5	4	5	1	7	5

$$\mathbf{X}_{[3]} = \begin{bmatrix} -0.03 & 0.31 & 0.57 \\ 0.17 & -0.06 & -0.19 \\ 0.80 & -0.61 & -0.57 \\ -0.24 & -0.43 & -0.38 \\ -0.24 & 0.31 & 0.38 \\ -0.45 & 0.49 & 0.19 \end{bmatrix}. \quad (1)$$

Each observation is assigned a *mass* which reflects its importance. When all observations have the same importance, their masses are all equal to  $m_i = \frac{1}{I}$ . The set of the masses is stored in an  $I \times I$  diagonal matrix denoted  $\mathbf{M}$ .

## 4 Finding the global space

### 4.1 Computing the separate PCA's

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  ${}_1\varrho_1 = 2.86$  and a first singular value of  ${}_1\varphi_1 = \sqrt{{}_1\varrho_1} = 1.69$ . This gives the first normalized data matrix denoted  $\mathbf{Z}_{[1]}$ :

$$\mathbf{Z}_{[1]} = {}_1\varphi_1^{-1} \times \mathbf{X}_{[1]} = \begin{bmatrix} -0.33 & 0.34 & 0.45 \\ 0.11 & -0.04 & -0.16 \\ 0.22 & -0.30 & -0.28 \\ 0.33 & -0.30 & -0.16 \\ -0.22 & 0.21 & 0.08 \\ -0.11 & 0.08 & 0.08 \end{bmatrix}. \quad (2)$$

Matrices  $\mathbf{Z}_{[2]}$  and  $\mathbf{Z}_{[3]}$  are normalized with their first respective singular values of  ${}_2\varphi_1 = 1.91$  and  ${}_3\varphi_1 = 1.58$ . Normalized matrices have a first singular value equal to 1.

## 4.2 Building the global matrix

The normalized studies are concatenated into an  $I \times T$  matrix called the *global data matrix* denoted  $\mathbf{Z}$ . Here we obtain:

$$\mathbf{Z} = [ \mathbf{Z}_{[1]} \quad \mathbf{Z}_{[2]} \quad \mathbf{Z}_{[3]} ]$$

$$= \left[ \begin{array}{ccc|ccc|ccc} -0.33 & 0.34 & 0.45 & -0.26 & 0.18 & 0.30 & 0.28 & -0.02 & 0.19 & 0.36 \\ 0.11 & -0.04 & -0.16 & 0.00 & 0.03 & 0.02 & -0.17 & 0.11 & -0.04 & -0.12 \\ 0.22 & -0.30 & -0.28 & 0.13 & -0.29 & -0.27 & -0.28 & 0.51 & -0.39 & -0.36 \\ 0.33 & -0.30 & -0.16 & 0.39 & -0.29 & -0.27 & -0.17 & -0.15 & -0.27 & -0.24 \\ -0.22 & 0.21 & 0.08 & -0.13 & 0.18 & 0.20 & 0.17 & -0.15 & 0.19 & 0.24 \\ -0.11 & 0.08 & 0.08 & -0.13 & 0.18 & 0.02 & 0.17 & -0.29 & 0.31 & 0.12 \end{array} \right]. \quad (3)$$

## 4.3 Computing the global PCA

To analyze the global matrix, we use standard PCA. This amounts to computing the singular value decomposition of the global data matrix:

$$\mathbf{Z} = \mathbf{U}\mathbf{\Delta}\mathbf{V}^T \text{ with } \mathbf{U}^T\mathbf{U} = \mathbf{V}^T\mathbf{V} = \mathbf{I}, \quad (4)$$

(where  $\mathbf{U}$  and  $\mathbf{V}$  are the left and right singular vectors of  $\mathbf{Z}$  and  $\mathbf{\Delta}$  is the diagonal matrix of the singular values).

For our example we obtain:

$$\mathbf{U} = \left[ \begin{array}{ccccc} 0.53 & -0.35 & -0.58 & -0.04 & 0.31 \\ -0.13 & -0.13 & 0.49 & 0.51 & 0.54 \\ -0.56 & -0.57 & 0.01 & -0.36 & -0.25 \\ -0.44 & 0.62 & -0.48 & 0.15 & 0.03 \\ 0.34 & 0.04 & 0.16 & 0.39 & -0.73 \\ 0.27 & 0.40 & 0.40 & -0.65 & 0.11 \end{array} \right] \quad (5)$$

and

$$\text{diag}\{\mathbf{\Delta}\} = [ 1.68 \quad 0.60 \quad 0.34 \quad 0.18 \quad 0.11 ]$$

$$\text{and } \text{diag}\{\mathbf{\Lambda}\} = \text{diag}\{\mathbf{\Delta}^2\} = [ 2.83 \quad 0.36 \quad 0.11 \quad 0.03 \quad 0.01 ] \quad (6)$$

( $\Lambda$  gives the eigenvalues of the PCA) and

$$\mathbf{V} = \begin{bmatrix} -0.34 & 0.22 & 0.03 & 0.14 & 0.55 \\ 0.35 & -0.14 & -0.03 & 0.30 & 0.02 \\ 0.32 & -0.06 & -0.65 & -0.24 & 0.60 \\ -0.28 & 0.34 & -0.32 & 0.31 & -0.18 \\ 0.30 & -0.00 & 0.43 & 0.11 & 0.19 \\ 0.30 & -0.18 & -0.00 & 0.67 & 0.11 \\ 0.30 & 0.09 & -0.22 & -0.36 & -0.38 \\ -0.22 & -0.86 & 0.01 & -0.12 & -0.00 \\ 0.36 & 0.20 & 0.45 & -0.30 & 0.19 \\ 0.37 & 0.01 & -0.21 & 0.18 & -0.28 \end{bmatrix} \quad (7)$$

The global factor scores for the wines are obtained as:

$$\mathbf{F} = \mathbf{M}^{-\frac{1}{2}} \mathbf{U} \mathbf{\Lambda} \quad (8)$$

$$= \begin{bmatrix} 2.18 & -0.51 & -0.48 & -0.02 & 0.08 \\ -0.56 & -0.20 & 0.41 & 0.23 & 0.15 \\ -2.32 & -0.83 & 0.01 & -0.16 & -0.07 \\ -1.83 & 0.90 & -0.40 & 0.07 & 0.01 \\ 1.40 & 0.05 & 0.13 & 0.17 & -0.20 \\ 1.13 & 0.58 & 0.34 & -0.29 & 0.03 \end{bmatrix} \cdot \quad (9)$$

In  $\mathbf{F}$ , each row represents an observation (*i.e.*, a wine) and each column a component. Figure 1 displays the wines in the space of the first two principal components. The first component has an eigenvalue equal to  $\lambda_1 = 2.83$ , which corresponds to 84% of the inertia ( $\frac{2.83}{2.83+0.36+0.11+0.03+0.01} = \frac{2.83}{3.35} \approx .84$ ). The second component, with an eigenvalue of .36, explains 11% of the inertia. The first component is interpreted as the opposition between the first (wines 1, 5, and 6) and the second oak type (wines 2, 3, and 4).

## 5 Partial analyses

The global analysis reveals the common structure of the wine space. In addition, we want to see how each expert “interprets” this space. This is achieved by projecting the data set of each expert onto the

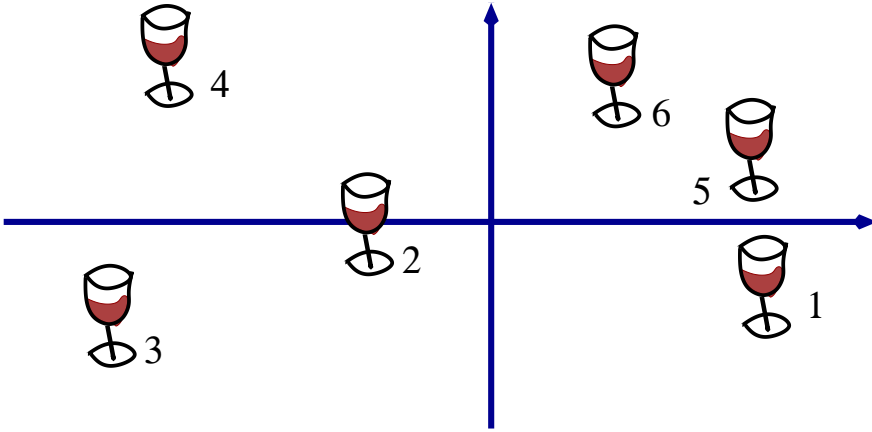


Figure 1: Global analysis: Plot of the wines on the first two principal components. First component:  $\lambda_1 = 2.83$ , explains 84% of the inertia, Second component:  $\lambda_2 = 2.83$ , explains 11% of the inertia.

global analysis. This is implemented by multiplication of a cross product matrix by a projection matrix. The projection matrix is obtained by rewriting Equation 8, to show that the global factor scores could be computed as

$$\mathbf{F} = \mathbf{M}^{-\frac{1}{2}} \mathbf{U} \mathbf{\Delta} = (\mathbf{Z} \mathbf{Z}^T) \times (\mathbf{M}^{-\frac{1}{2}} \mathbf{U} \mathbf{\Delta}^{-1}). \quad (10)$$

This shows that  $\mathbf{P} = \mathbf{M}^{-\frac{1}{2}} \mathbf{U} \mathbf{\Delta}^{-1}$  is a projection matrix which transforms the matrix  $\mathbf{Z} \mathbf{Z}^T$  into factor scores. Here, we obtain:

$$\mathbf{P} = \mathbf{M}^{-\frac{1}{2}} \mathbf{U} \mathbf{\Delta}^{-1} = \begin{bmatrix} 0.77 & -1.43 & -4.20 & -0.55 & 6.68 \\ -0.20 & -0.55 & 3.56 & 6.90 & 11.71 \\ -0.82 & -2.33 & 0.05 & -4.85 & -5.53 \\ -0.65 & 2.54 & -3.46 & 2.01 & 0.66 \\ 0.50 & 0.15 & 1.13 & 5.27 & -15.93 \\ 0.40 & 1.61 & 2.91 & -8.78 & 2.43 \end{bmatrix}, \quad (11)$$

The projection matrix is then used to project the studies onto the global space. For example, for the first expert we obtain

$$\mathbf{F}_{[1]} = T \times (\mathbf{Z}_{[1]} \mathbf{Z}_{[1]}^T) \mathbf{P} \quad (12)$$



$$= \begin{bmatrix} 2.76 & -1.10 & -2.29 & -0.39 & 0.67 \\ -0.77 & 0.30 & 0.81 & 0.31 & -0.27 \\ -1.99 & 0.81 & 1.48 & 0.08 & -0.39 \\ -1.98 & 0.93 & 0.92 & -0.02 & 0.59 \\ 1.29 & -0.62 & -0.49 & 0.10 & -0.51 \\ 0.69 & -0.30 & -0.43 & -0.07 & -0.08 \end{bmatrix}, \quad (13)$$

(multiplying by  $T$  is needed in order to scale one expert with all  $T = 3$  experts of the global solution). The same procedure gives for Experts 2 and 3:

$$\mathbf{F}_{[2]} = \begin{bmatrix} 2.21 & -0.86 & 0.74 & 0.27 & 0.06 \\ -0.28 & -0.13 & 0.35 & 0.55 & 0.52 \\ -2.11 & 0.50 & -0.77 & -0.49 & -0.01 \\ -2.39 & 1.23 & -1.57 & -0.20 & -0.68 \\ 1.49 & -0.49 & 0.62 & 0.40 & 0.13 \\ 1.08 & -0.24 & 0.63 & -0.53 & -0.03 \end{bmatrix}, \quad (14)$$

and

$$\mathbf{F}_{[3]} = \begin{bmatrix} 1.54 & 0.44 & 0.09 & 0.07 & -0.47 \\ -0.61 & -0.76 & 0.06 & -0.17 & 0.19 \\ -2.85 & -3.80 & -0.69 & -0.07 & 0.19 \\ -1.12 & 0.56 & -0.55 & 0.42 & 0.11 \\ 1.43 & 1.27 & 0.26 & 0.03 & -0.22 \\ 1.62 & 2.28 & 0.82 & -0.28 & 0.20 \end{bmatrix}. \quad (15)$$

Figure 2 shows the first two principal components of the global analysis along with the wine projections for the experts. Note that, the position of each wine in the global analysis is the barycenter (*i.e.*, centroid) of its positions for the experts. To facilitate the interpretation, we have drawn lines linking the expert wine projection to the global wine position. This picture shows that Expert 3 is at variance with the other experts in particular for Wines 3 and 6.

## 5.1 The original variables and the global analysis

As in standard PCA, the variable loadings are the correlation between the original variables and the global factor scores (*cf.* Table 2). These loadings are plotted in Figure 3 along with the “circles of correlation.” This figure shows that Expert 3 differs from the

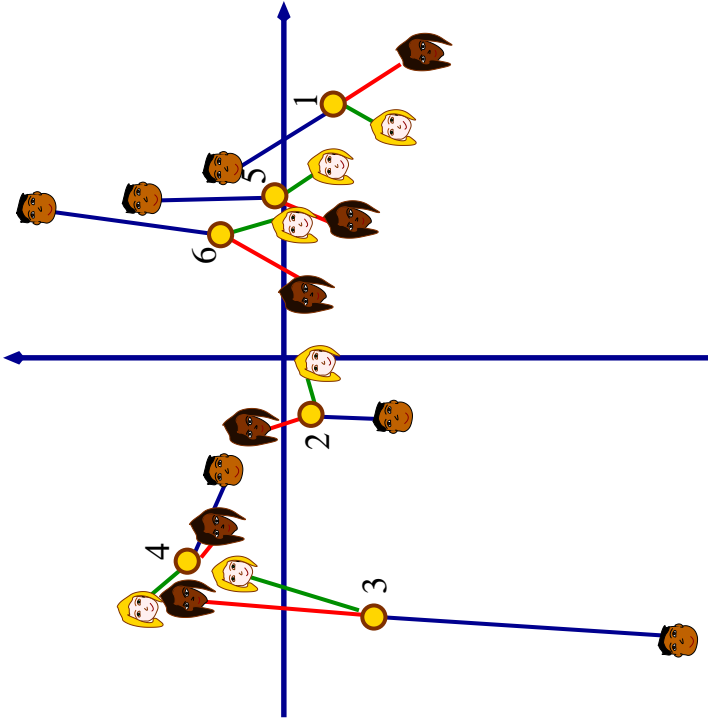


Figure 2: Projection of the experts onto the global analysis. Experts are represented by their faces. A line segment links the position of the wine for a given expert to its global position. First component:  $\lambda_1 = 2.83$ , explains 84% of the inertia, Second component:  $\lambda_2 = 2.83$ , explains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

Table 2: Loadings on the principal components of the global analysis of 1.) the original variables and 2.) the principal components of the study PCA's. Only the first three dimensions are kept.

Loadings with original variables												
		Expert 1			Expert 2			Expert 3				
Axis	$\lambda$	%	Fruity	Woody	Coffee	Fruit	Roasted	Vanillin	Woody	Fruity	Butter	Woody
1	2.83	85	-0.97	0.98	0.92	-0.89	0.96	0.95	0.97	-0.59	0.95	0.99
2	.36	11	0.23	-0.15	-0.06	0.38	-0.00	-0.20	0.10	-0.80	0.19	0.00
3	.12	3	0.02	-0.02	-0.37	-0.21	0.28	-0.00	-0.14	0.08	0.24	-0.11

Loadings with first 2 components from study PCA's									
		Expert 1		Expert 2		Expert 3			
Axis	$\lambda$	%	PC1	PC2	PC1	PC2	PC1	PC2	
1	2.83	85	.98	.08	.99	-.16	.94	-.35	
2	.36	11	-.15	-.28	-.13	-.76	.35	.94	
3	.12	3	-.14	.84	.09	.58	.05	-.01	

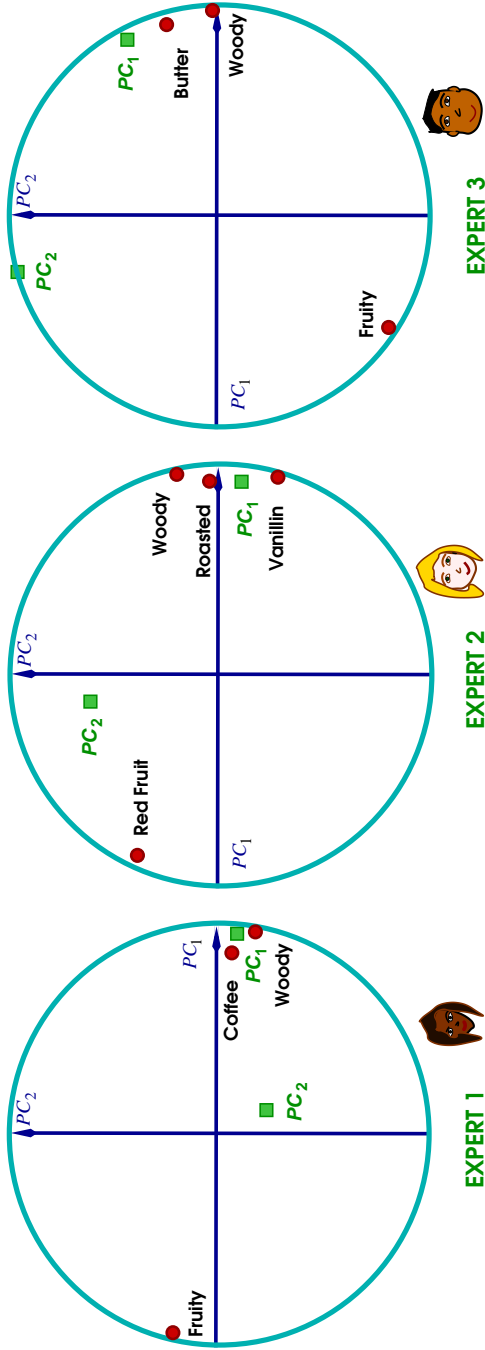


Figure 3: Circle of correlations for the original variables.

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
$\Sigma$	2.83	.36	.12	.03	.01
	$\lambda_1$	$\lambda_2$	$\lambda_2$	$\lambda_4$	$\lambda_5$

other experts, and is mostly responsible for the second component of the compromise.

## 5.2 The original PCA's and the global analysis

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, given in Table 2, are displayed in Figure 3. They relate the original PCA and the global analysis.

## 6 Analyzing the between study structure

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  $\mathbf{Z}$  (*cf.* Equation 7) 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. For example, for study one, and component one, the partial inertia is

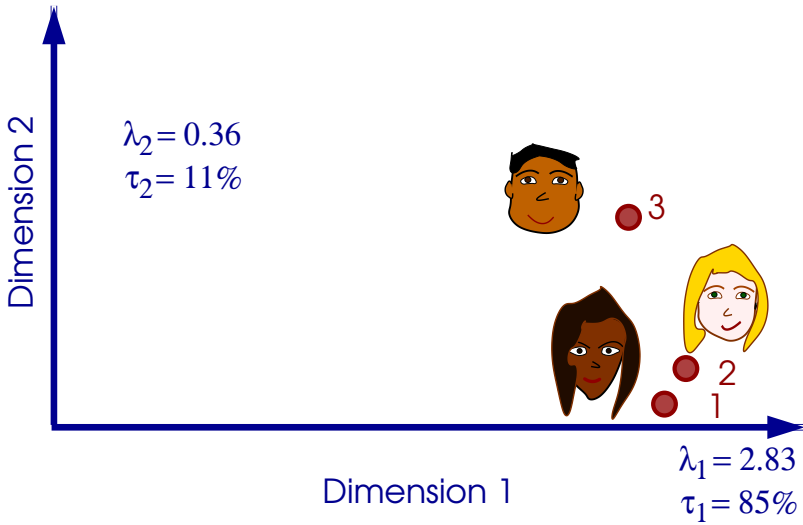


Figure 4: Partial Inertia: Plot of the experts on the first two components.

obtained as

$$\lambda_1 \times \sum_j q_{j,1}^2 = 2.83 \times [(-.34)^2 + (.35)^2 + (.32)^2] = 2.83 \times .34 = .96 . \quad (16)$$

Similar computations gives the values reported in Table 3. These values are used to plot the studies as shown in Figure 4. The plot confirms the originality of Expert 3, and its importance for Dimension 2.

## References

- [1] Abdi, H. (2003). Multivariate analysis. In M. Lewis-Beck, A. Bryman, & T. Futing (Eds): *Encyclopedia for research methods for the social sciences*. Thousand Oaks: Sage.
- [2] Escofier, B. and Pagès, J. (1990). *Analyses factorielles simples et multiples: objectifs, méthodes, interprétation*. Dunod, Paris.
- [3] Escofier, B. and Pagès, J. (1990). Multiple factor analysis. *Computational Statistics & Data Analysis*, **18**, 121–140.