European University of Britany

Agrocampus

Factorial analysis of qualitative and quantitative data both mixed and structured according to a hierarchy

Jérôme Pagès

Applied Mathematics dept.

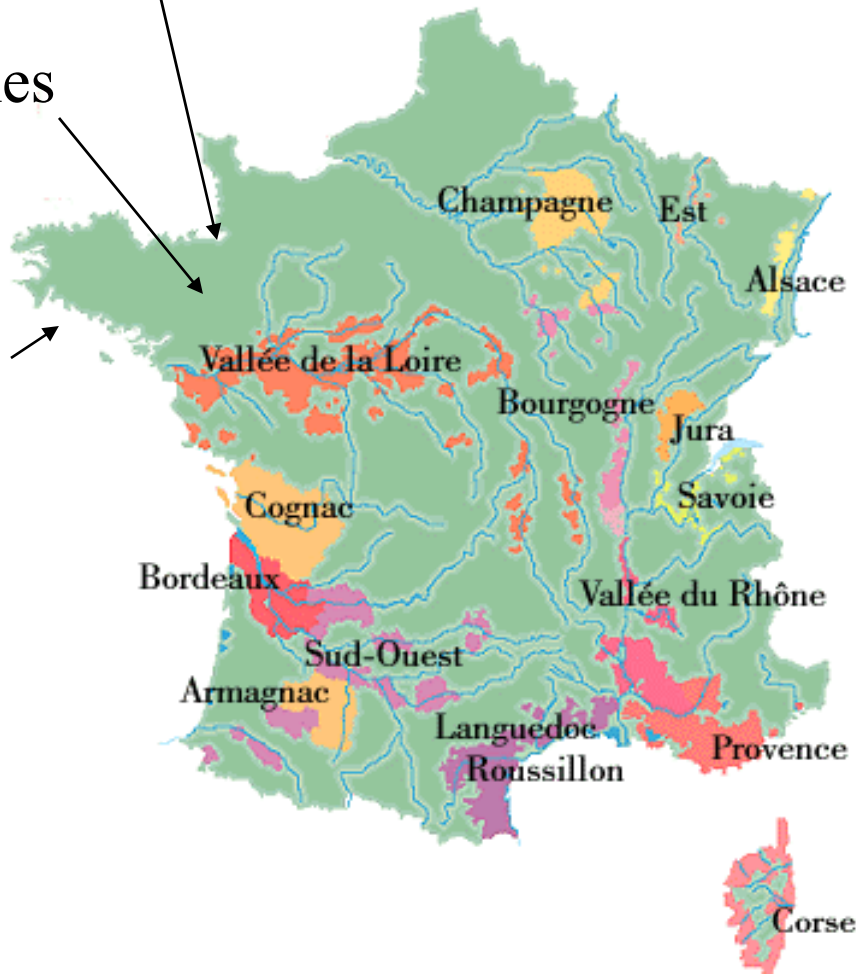
AGRO CAMPUS OUEST

LMA² Laboratoire de Mathématiques Appliquées Agrocampus

Mont Saint-Michel

Rennes

Britany



Context of factorial analysis :
One set of individuals described by several variables

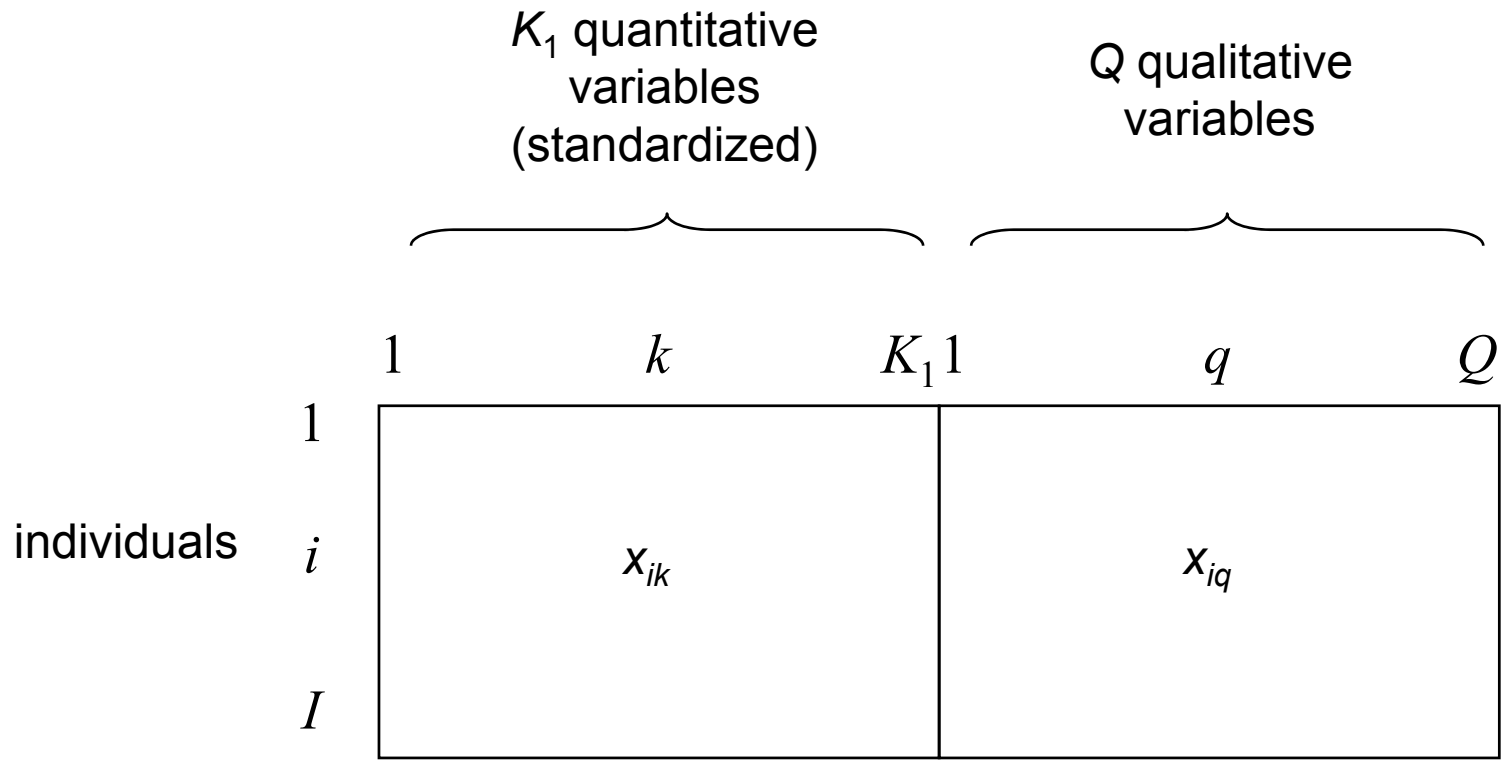
1 Taking into account both quantitative and qualitative variables
Factor Analysis for Mixed Data (FAMD)

2 Taking into account a partition of the variables
Multiple Factor Analysis (MFA)

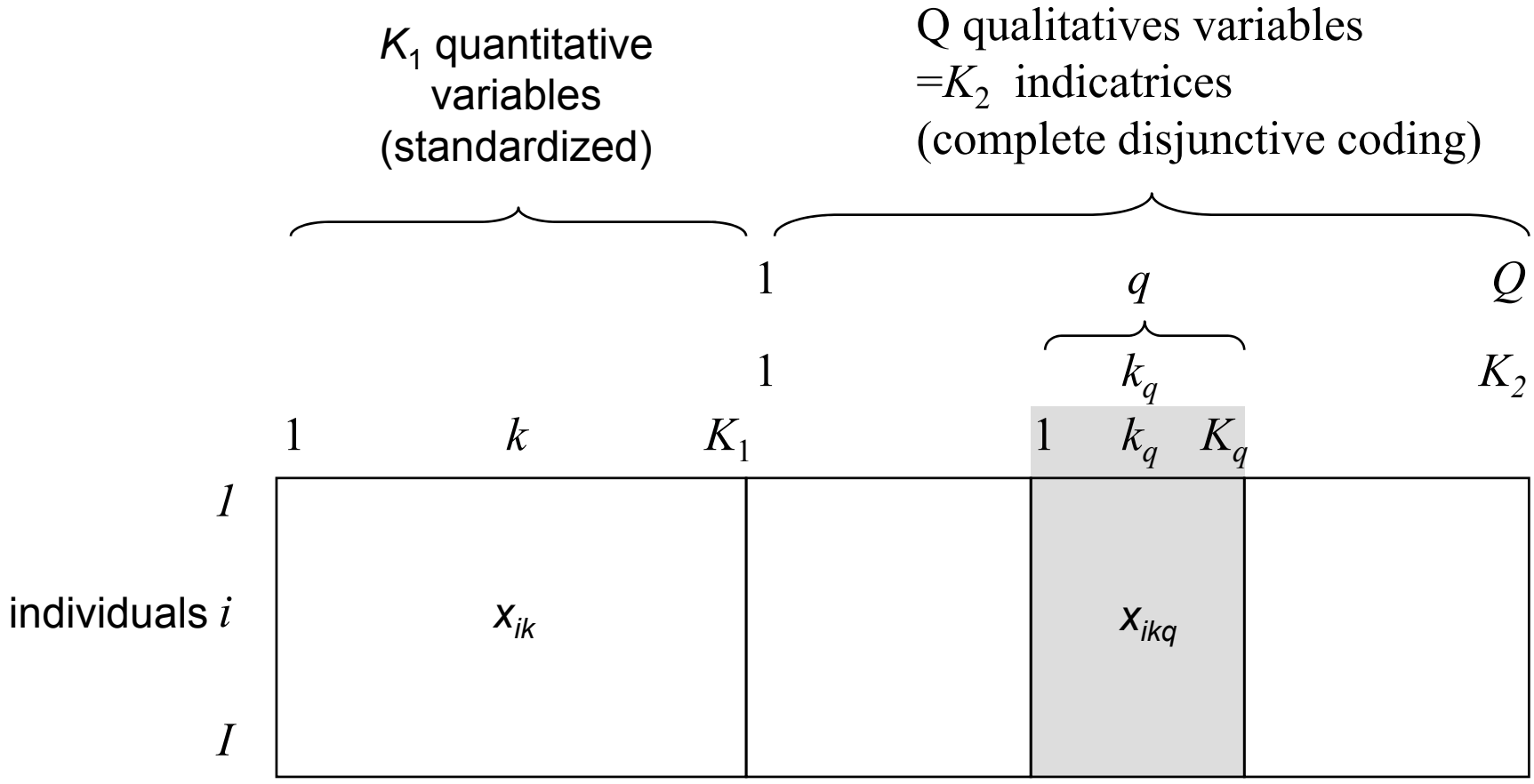
3 Taking into account a hierarchy defined on the variables
Hierarchical Multiple Factor Analysis (HMFA)

1 Taking into account both quantitative and qualitative variables Factor Analysis for Mixed Data (FAMD)

Data



Data


K_1 quantitative variables (standardized)
Q qualitatives variables =K_2 indicatrices (complete disjunctive coding)
1
q
Q
1
k_q
K_2
1
k
K_1
1
k_q
K_q
1
individuals i
I
x_{ik}
x_{ikq}

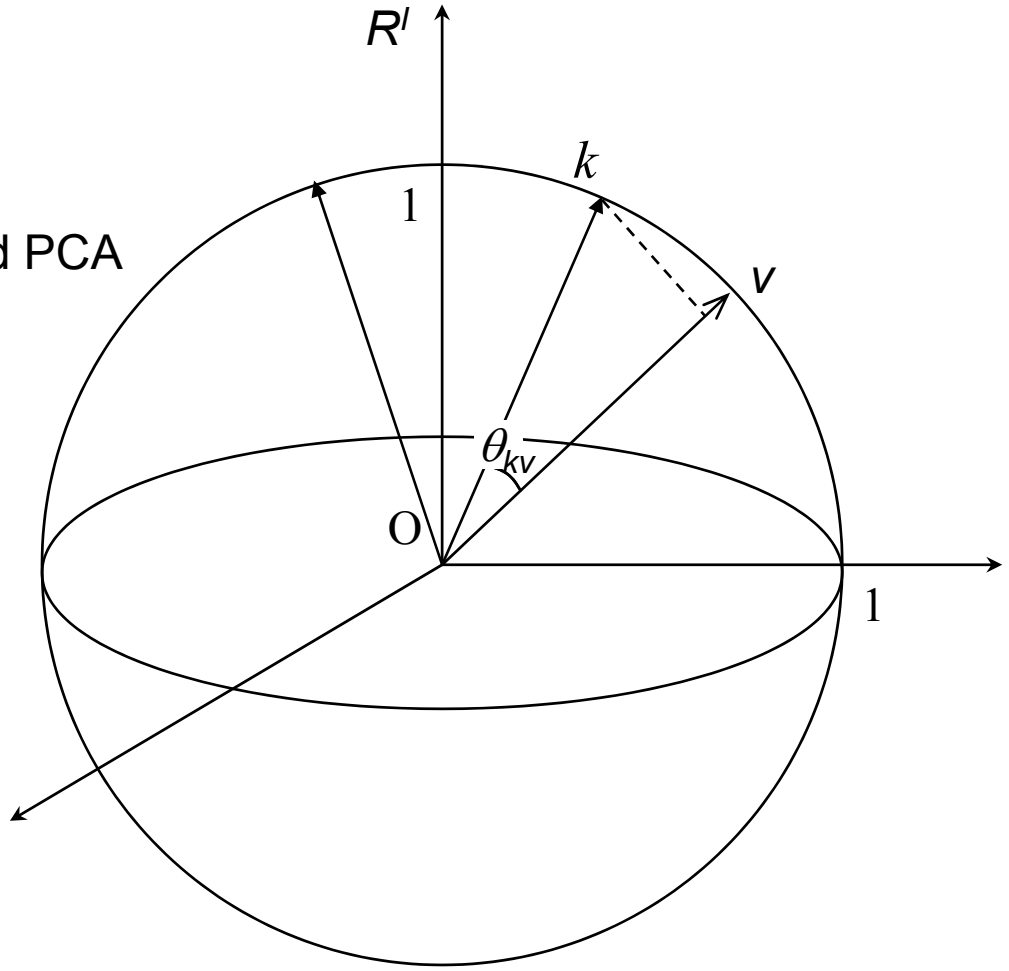
Principal components analysis
Representation of the variables and criterion

One variable = one axis

Criterion of standardized PCA

$$\sum_k r^2(k,v)$$

$$\sum_k \cos^2 \theta_{kv}$$



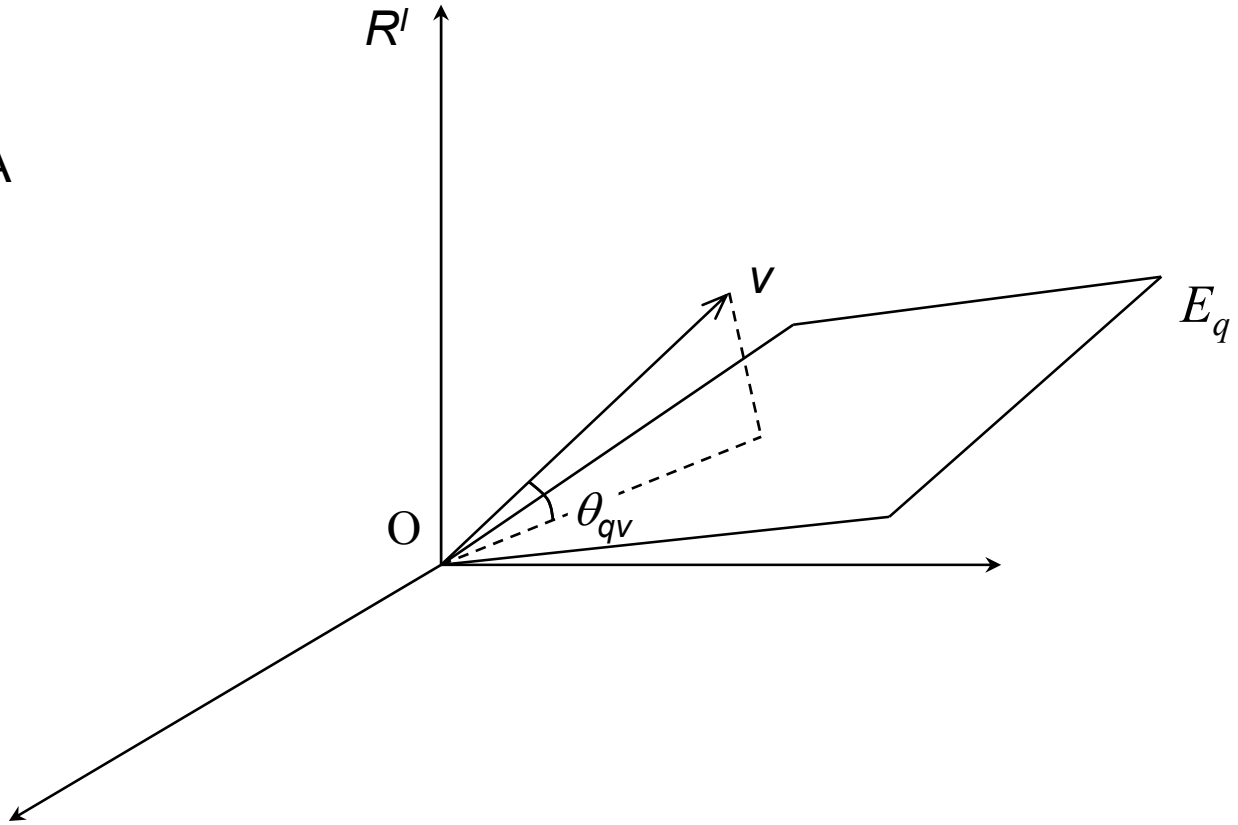
Multiple correspondence analysis
Representation of the variables and criterion

Variable q = sub-space E_q

Criterion of MCA

$$\sum_q \eta^2(q,v)$$

$$\sum_q \cos^2\theta_{qv}$$

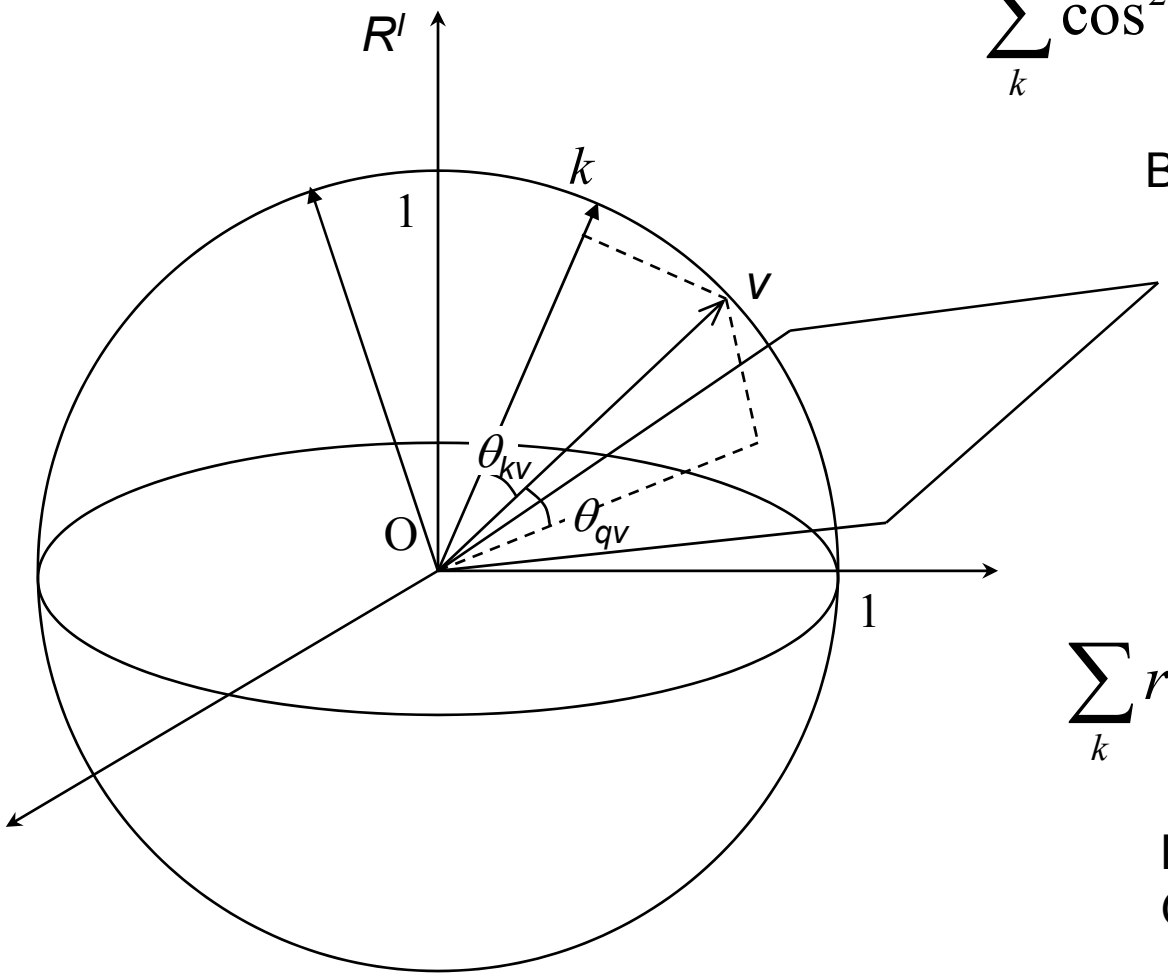


Factor analysis for mixed data (FAMD)

Criterion

$$\sum_k \cos^2 \theta_{kv} + \sum_q \cos^2 \theta_{qv}$$

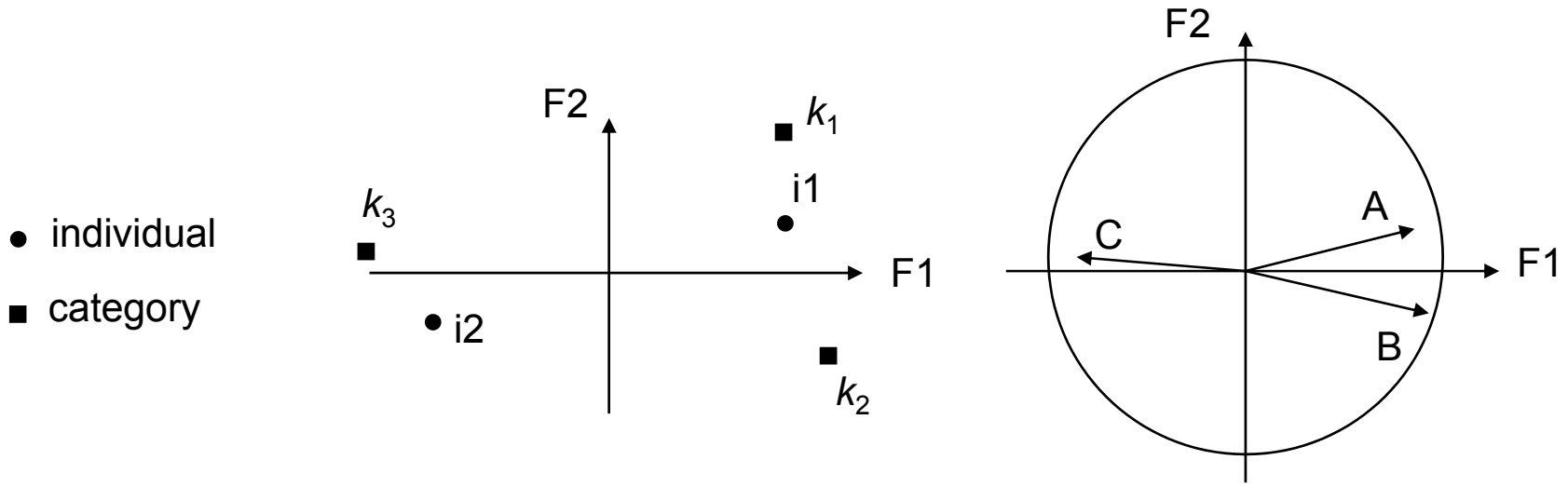
B. Escofier (1979)



$$\sum_k r^2(k,v) + \sum_q \eta^2(q,v)$$

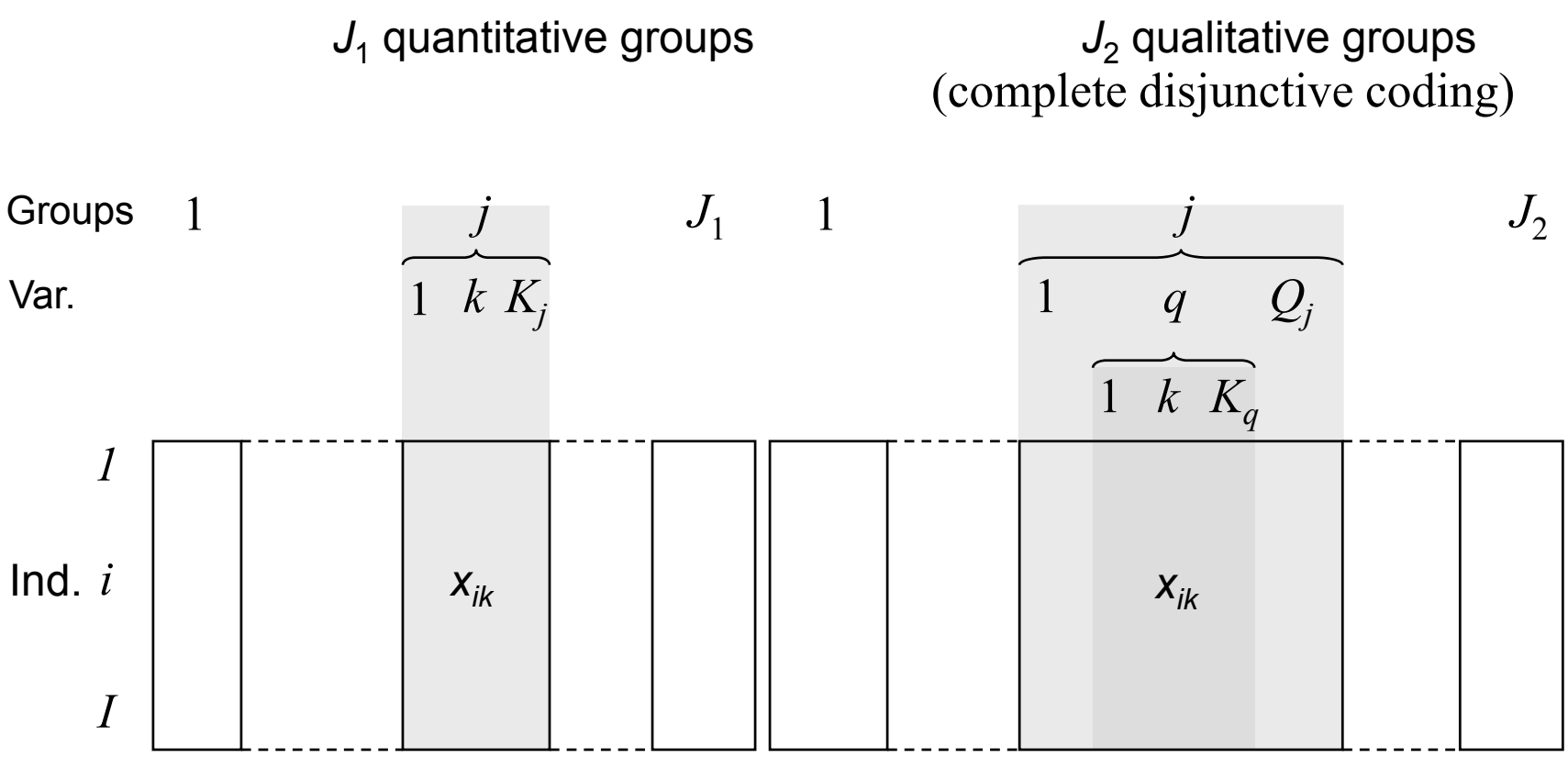
M. Tenenhaus (1985)
G. Saporta (1990)

Representations provided by FAMD



2 Taking into account a partition of the variables
Multiple Factor Analysis (MFA)

MFA applied to groups of variables : quantitative, qualitative or mixed



Weighting the groups of variables

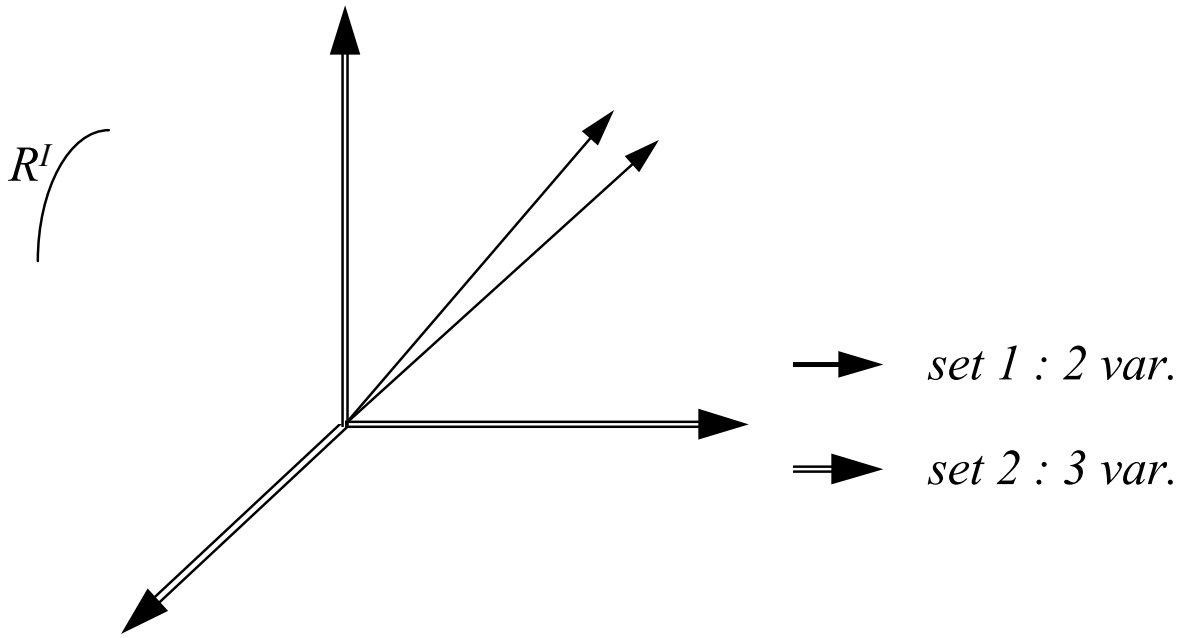
Several groups of active variables in a unique analysis

Question :

How to balance their influence ?

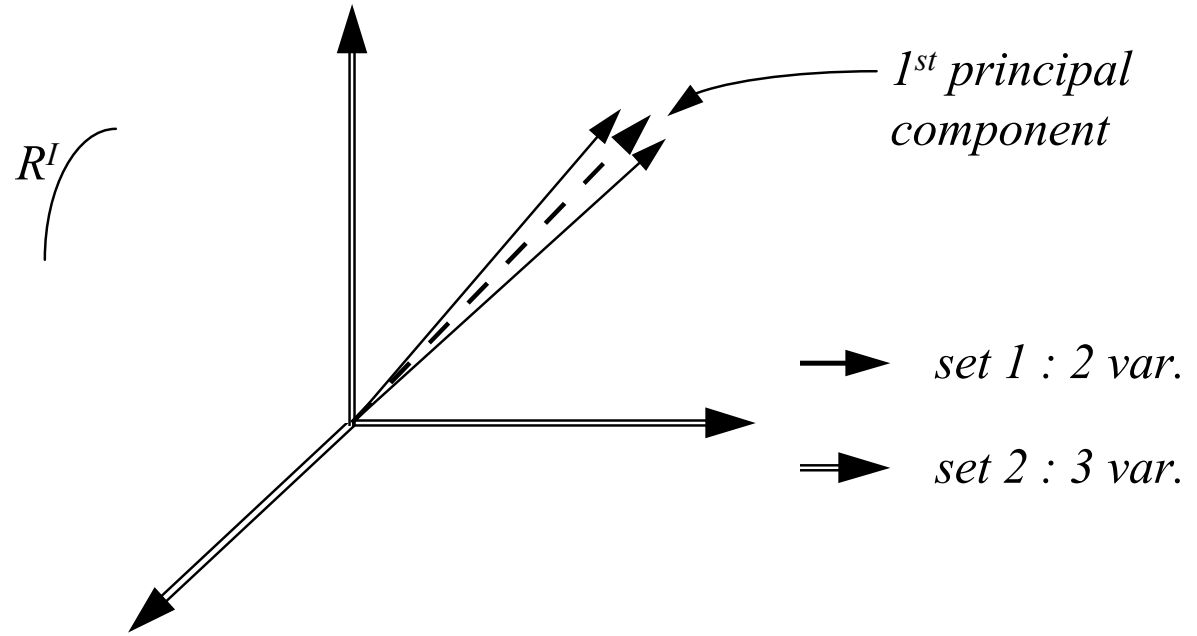
Weighting the variables in MFA

Reference example : two groups of quantitative variables



Reference example

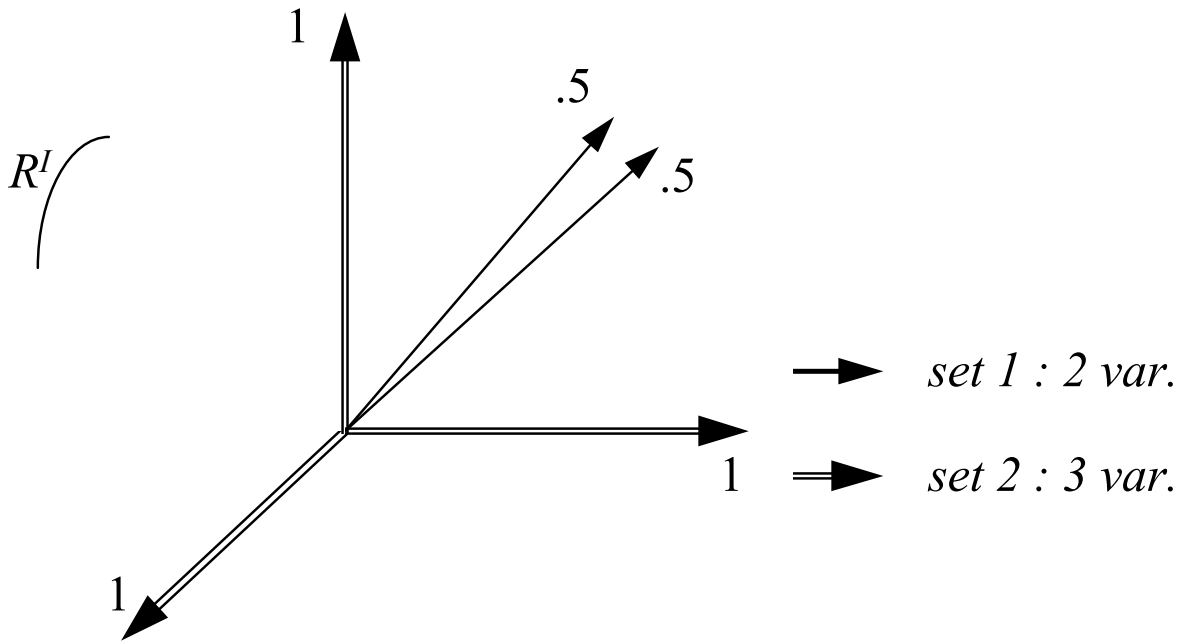
PCA of the 5 variables, without considering the sets



Reference example

Weighting the sets of variables in MFA
balancing the maximum axial inertia

Each variable of the set *j* is weighted by $1/\lambda_1^j$
λ_1^j : 1st eigenvalue of PCA applied to set *j*.



MFA is based on a factor analysis applied to all active sets of variables

The groups of variables can be

quantitative (standardized or not)

qualitative

mixed

Criterion (case of 2 groups : K_1 quantitative variables Q_2 qualitative variables)

$$\frac{1}{\lambda_1^1}\sum_{k\in K_1} r^2(k,v)+\frac{1}{\lambda_1^2 Q_2}\sum_{q\in Q_2}\eta^2(q,v)$$

Equivalences of MFA when each group contains a single variable

Quantitative variables	Standardized PCA
Qualitatives variables	MCA
Mixed data	FAMD

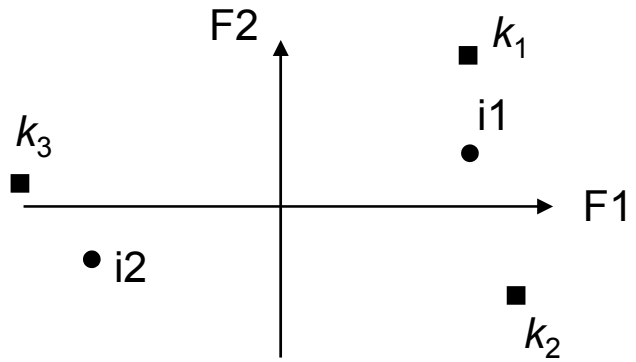
MFA is based on a factor analysis applied to all active sets of variables

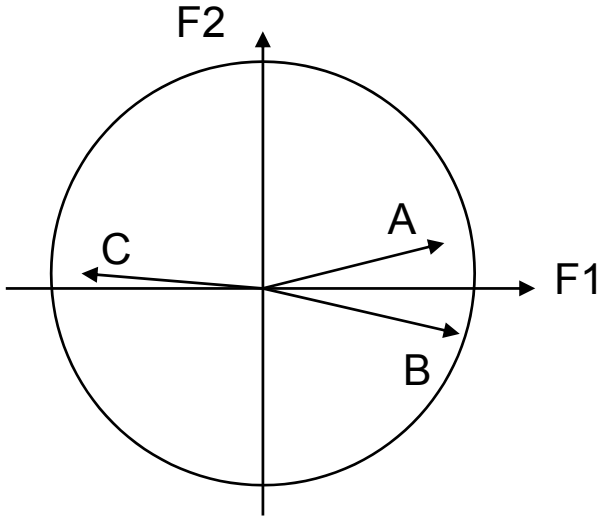
MFA provides :

Firstly : classical results of factor analysis

● individual

■ category





Specific representations (see HMFA)

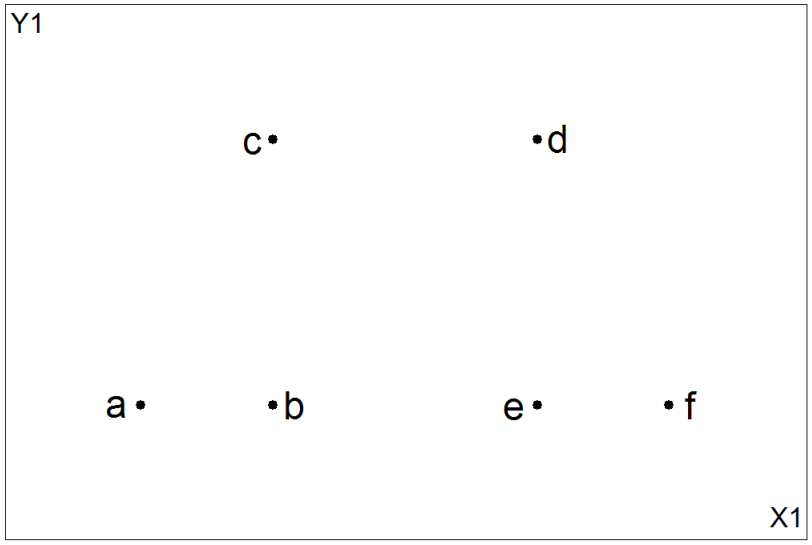
3 Taking into account a hierarchy defined on the variables
Hierarchical Multiple Factor Analysis (HMFA)

Sorted napping : an holistic approach in sensory evaluation

A set of products is given (a, b, c, d, e, f)

Task 1 : napping

Position the products on the tablecloth in such a way that :
two products are very near one another if they seem identical (for you) ,
two products are distant one another if they seem different (for you).
This must be done according to your own criteria.

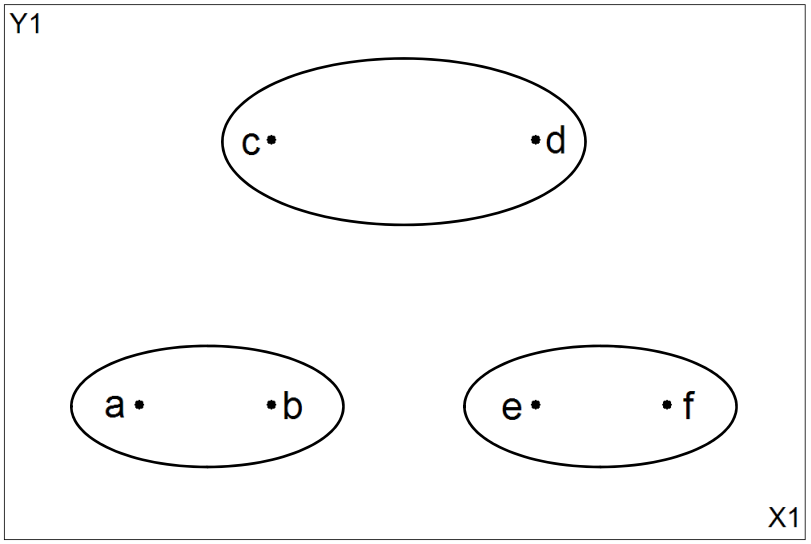


Sorted napping : an holistic approach in sensory evaluation

A set of product is given (a, b, c, d, e, f)

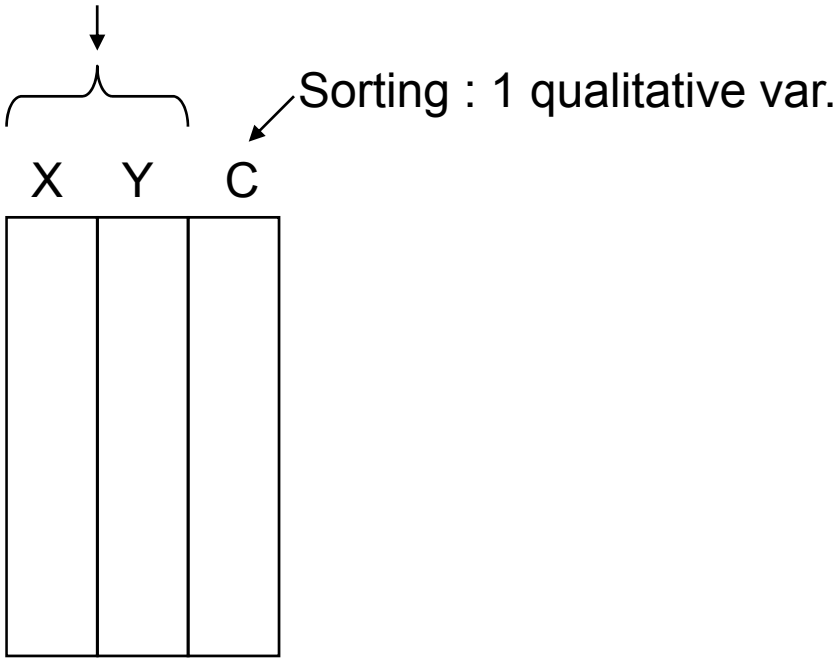
Task : sorted napping

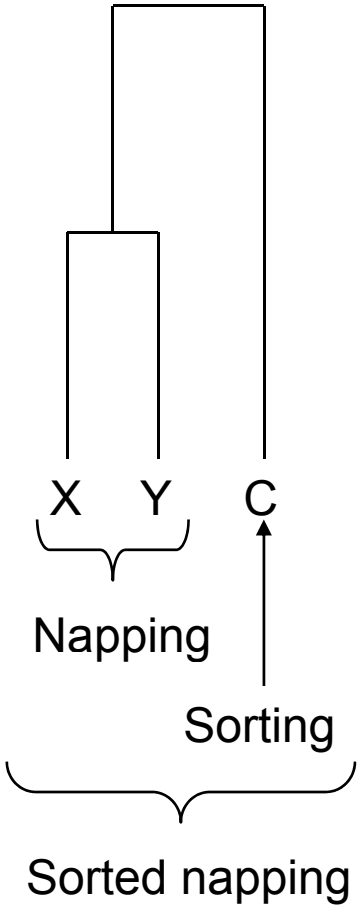
As the panellist forms his “tablecloth” (or ”nappe”),
he is asked to make groups of products,
i.e. to put in the same group the products that he perceives as similar.



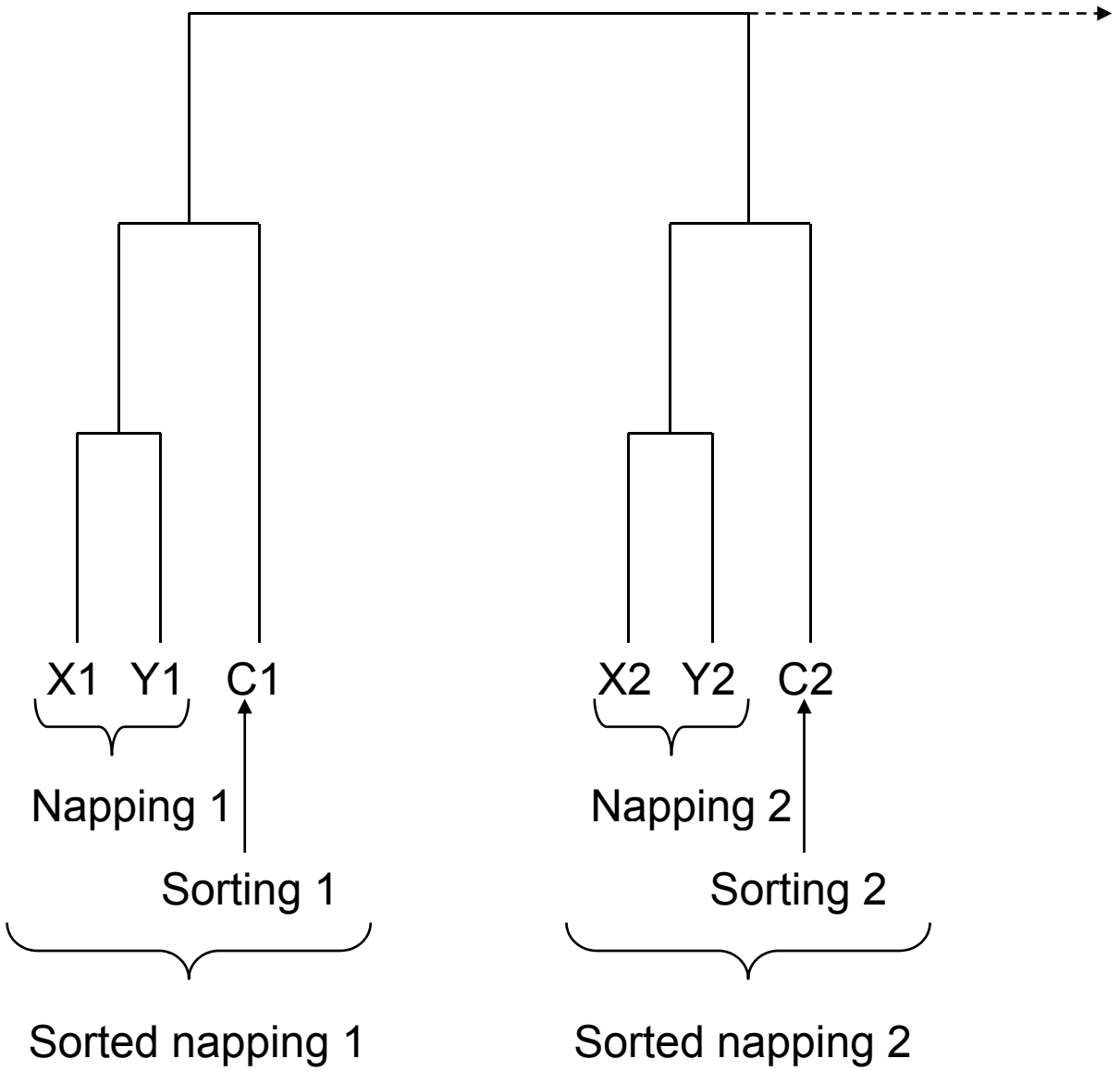
Data structure One sorted napping

Coordinates on the tableclothe
2 quantitative var. (not standardized)

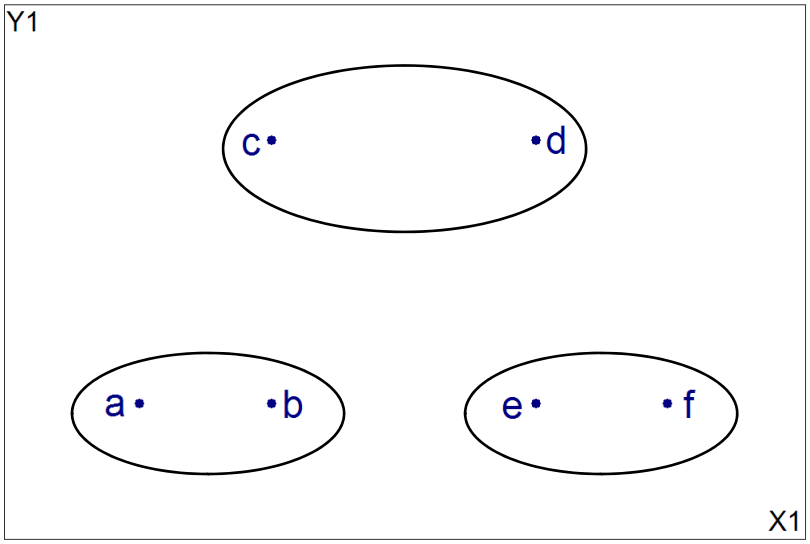


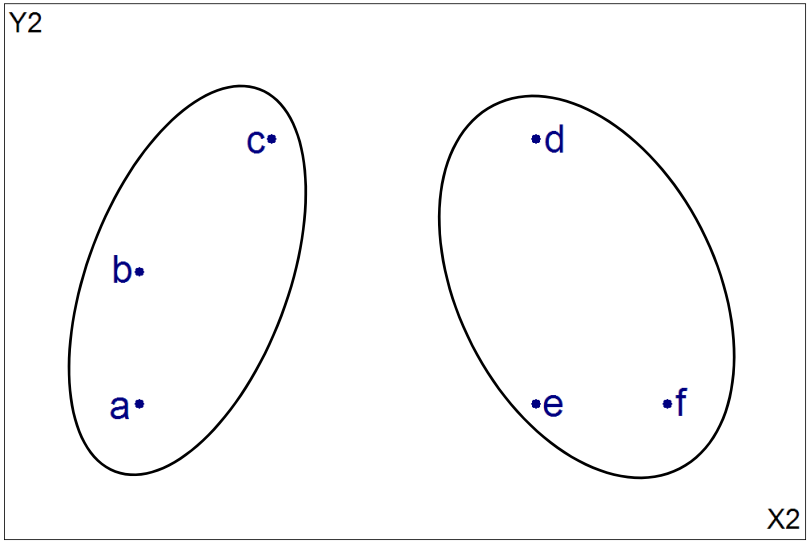


Case of several sorted nappings Data structure

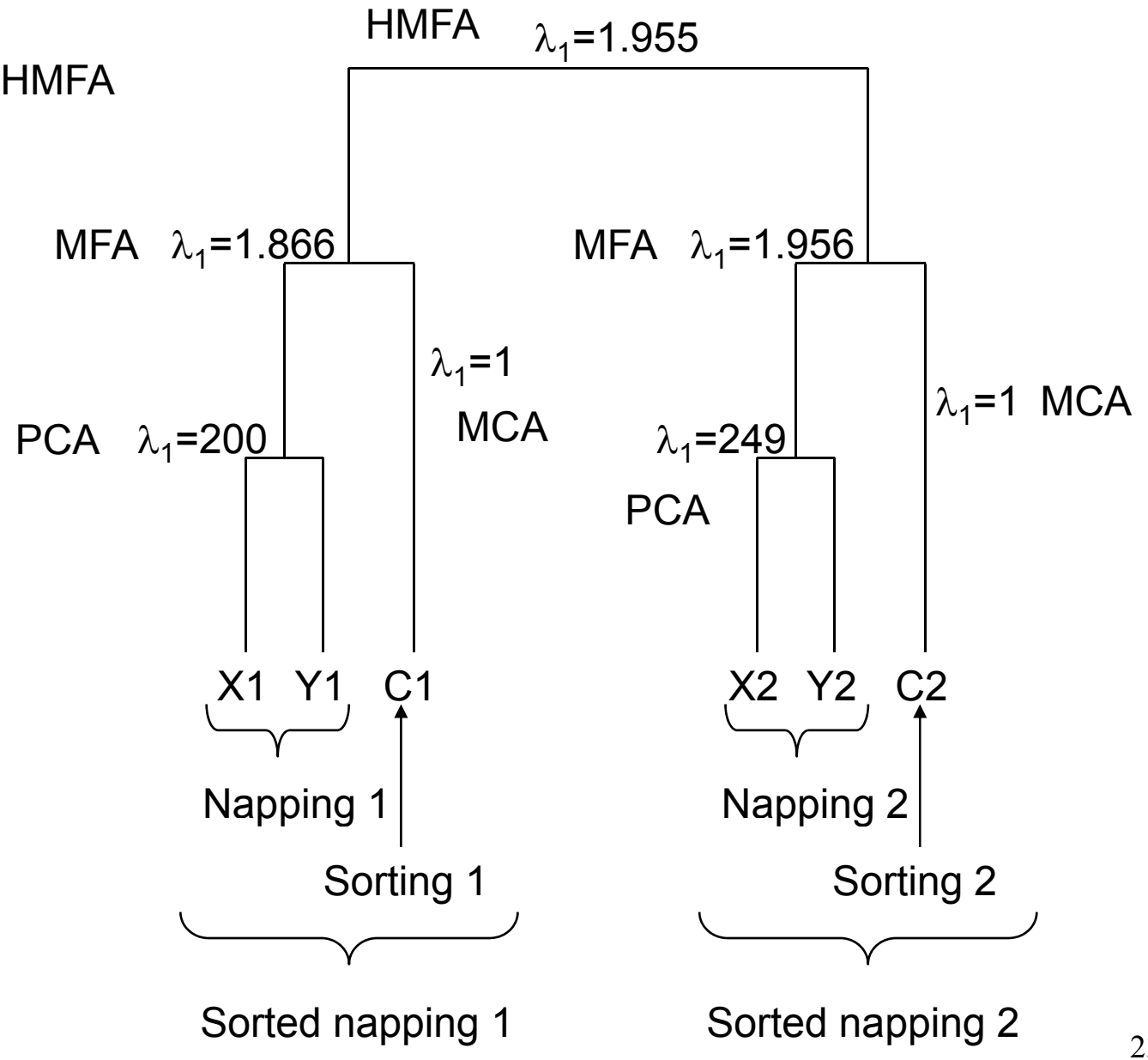


Exemple : two sorted nappings

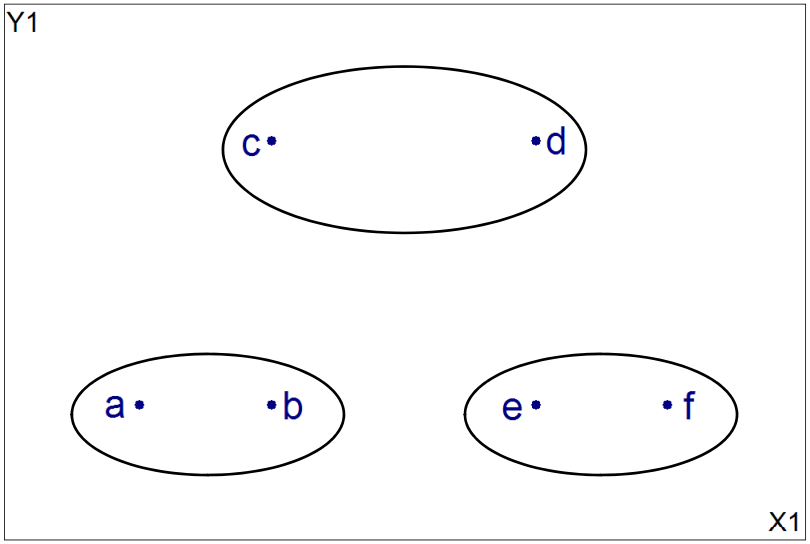


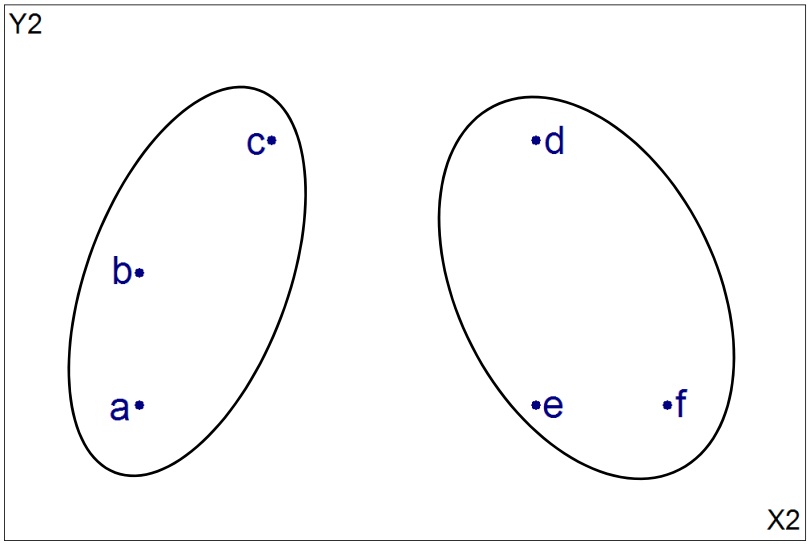


Balancing in HMFA

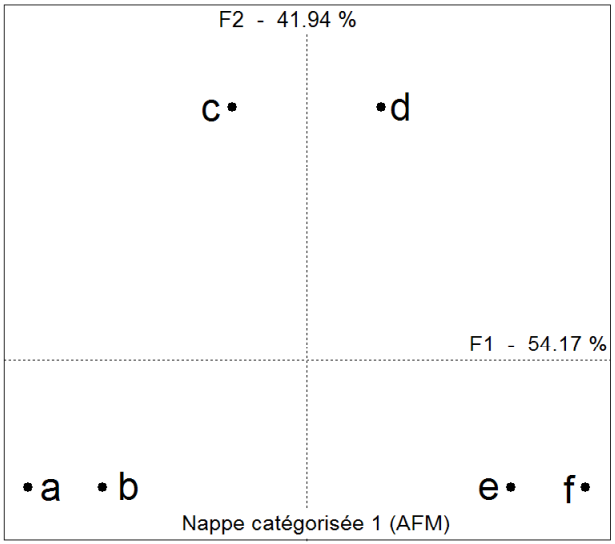


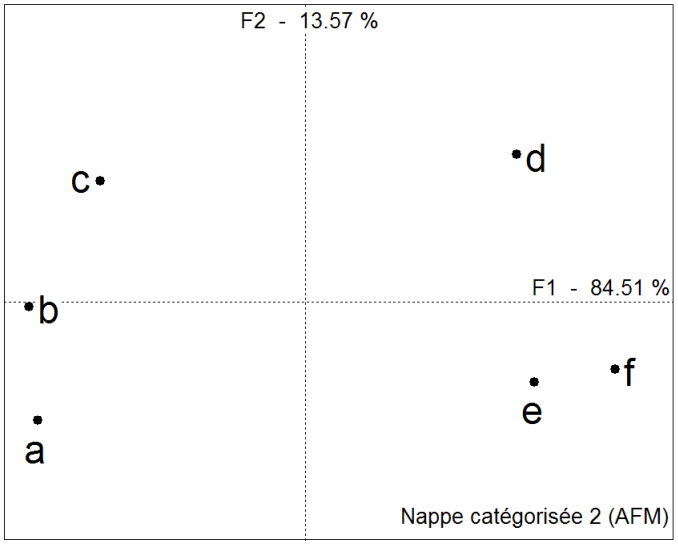
The raw sorted nappings



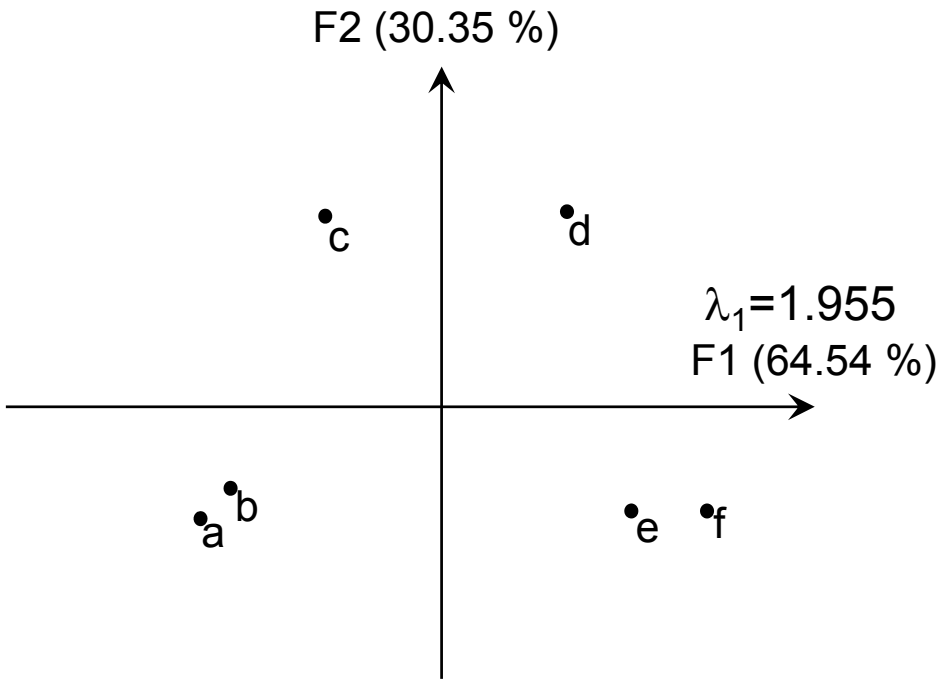


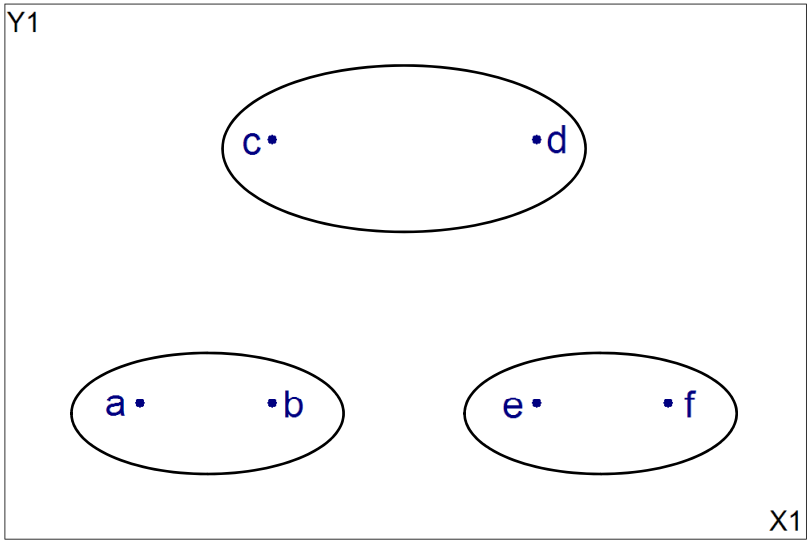
Sorted nappings seen through their MFA
First step of HMFA

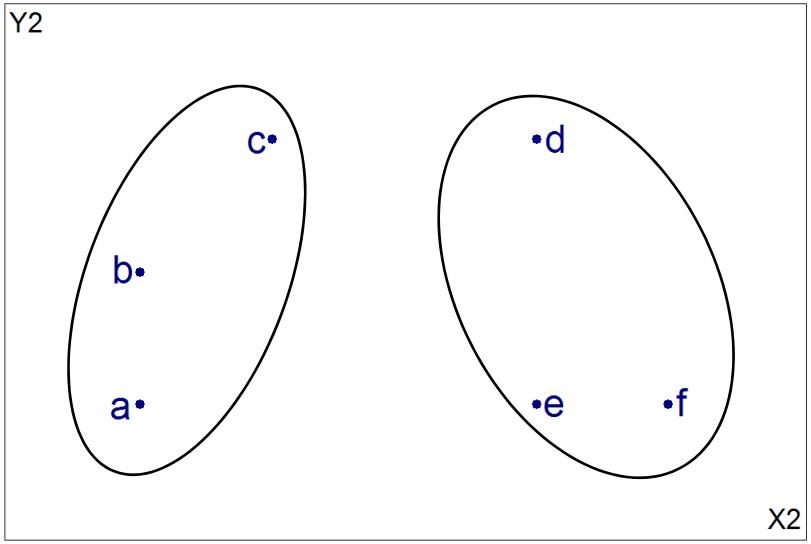




Individuals factor map

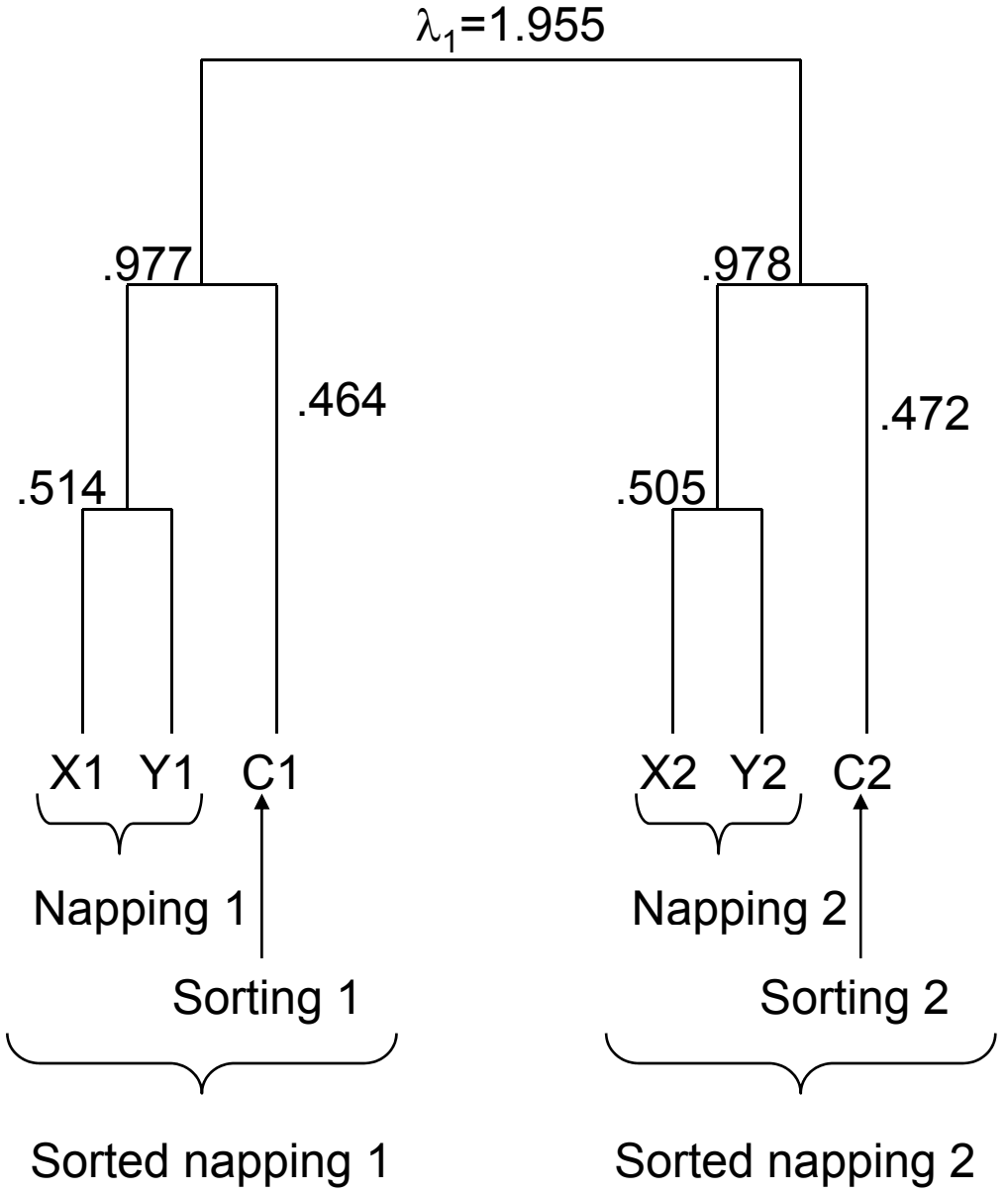






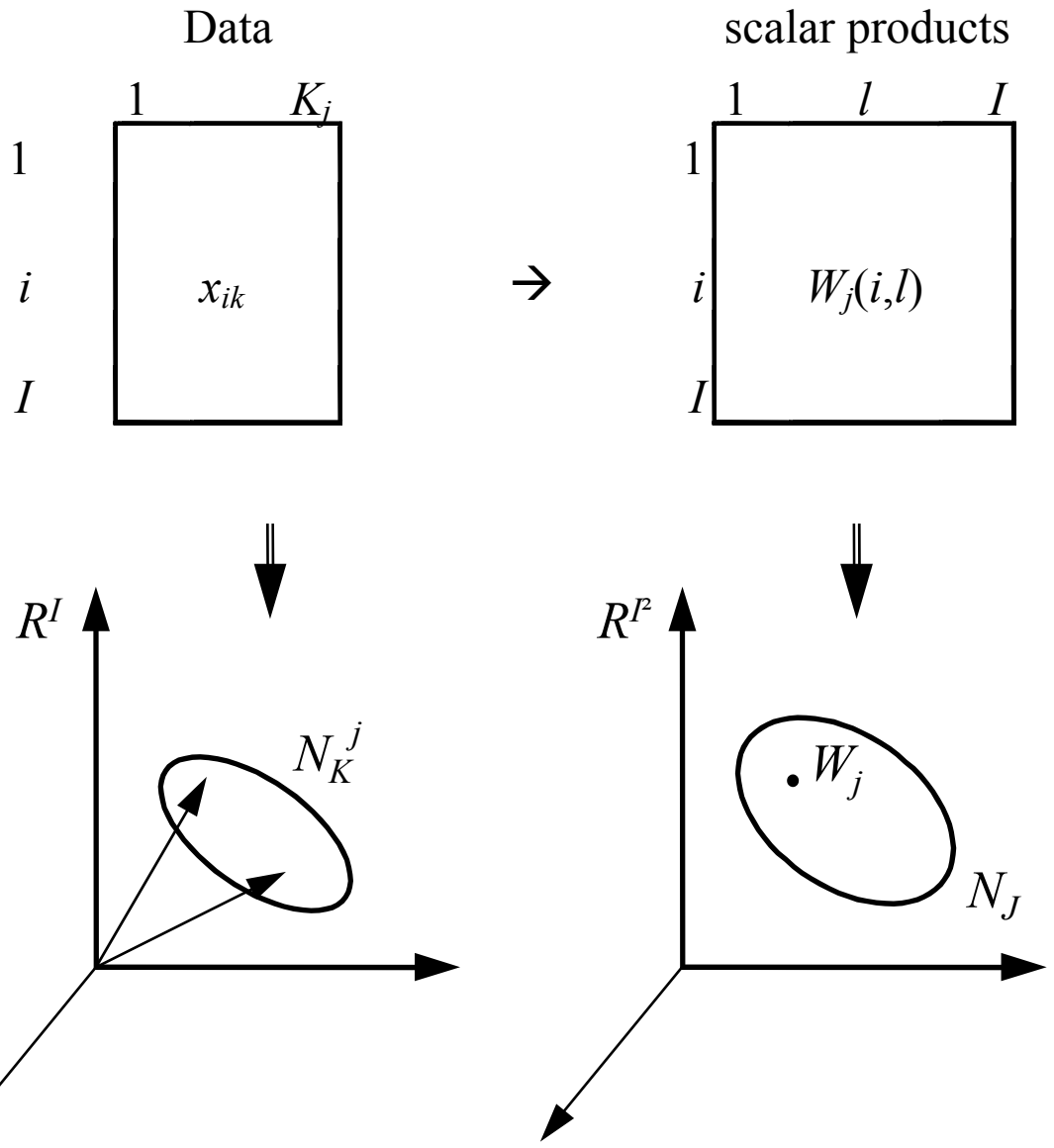
Efficiency of the balancing

Decomposition of the inertia for the first axis

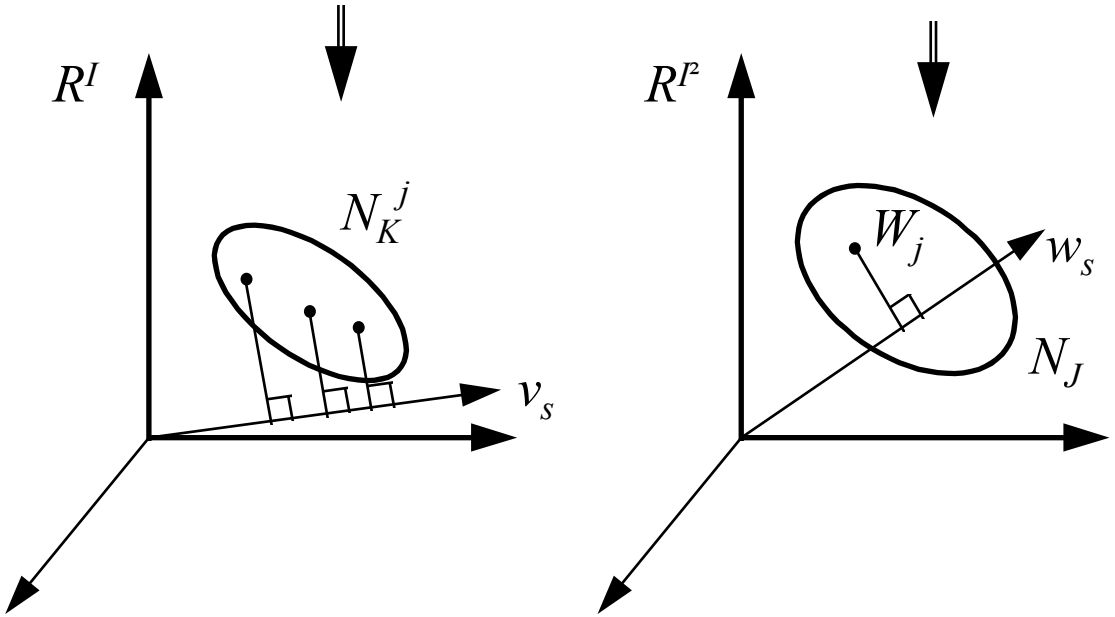


Representation of the nodes (=sets of variables) in HMFA

The set K_j of variables gathered in the node j



Representing N_J with HMFA



w_s : W associated to v_s

inertia of N_K^j projected upon v_s co-ordinate of W_j upon w_s

Projected inertia of the whole set of the variables K_j onto v_s

= a measure of relationship between

one variable v_s

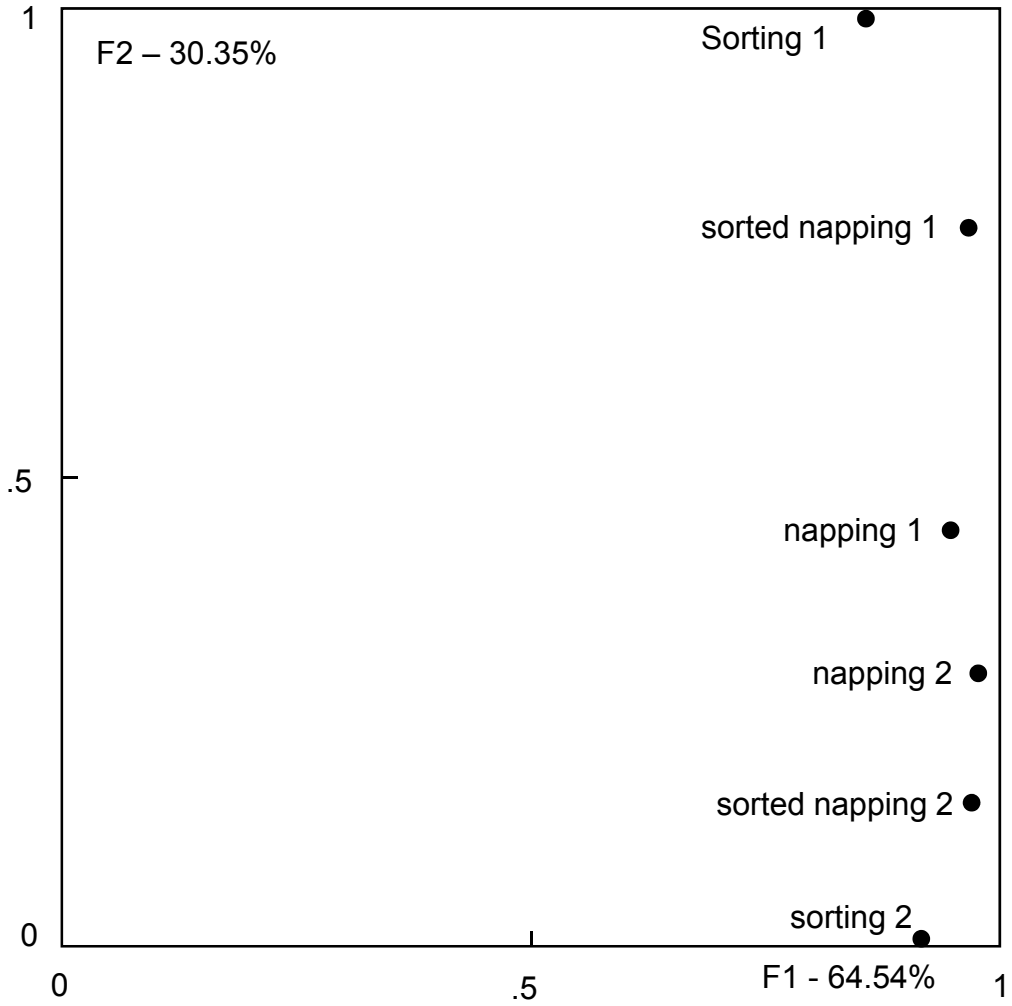
a set of variables K_j

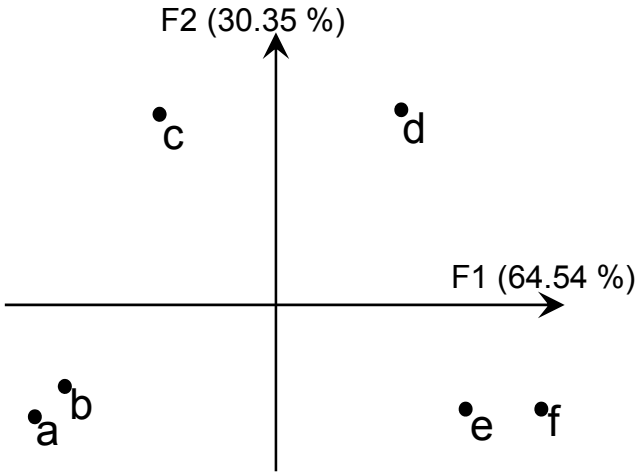
$$Lg(v_s, K_j)$$

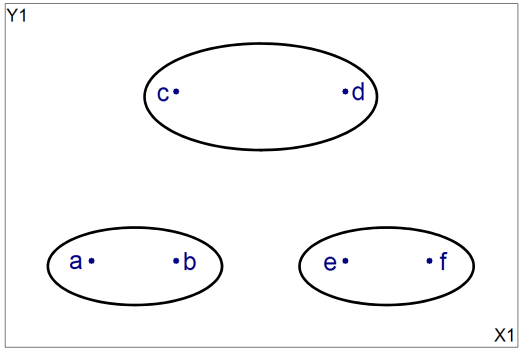
$$0 \le Lg(v_s, K_j) \le 1$$

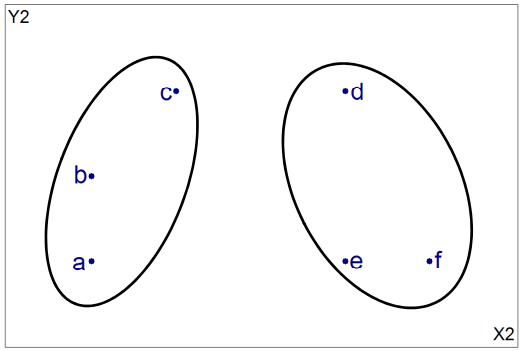
Can be applied to each node of the hierarchy

Lg square = relationship square

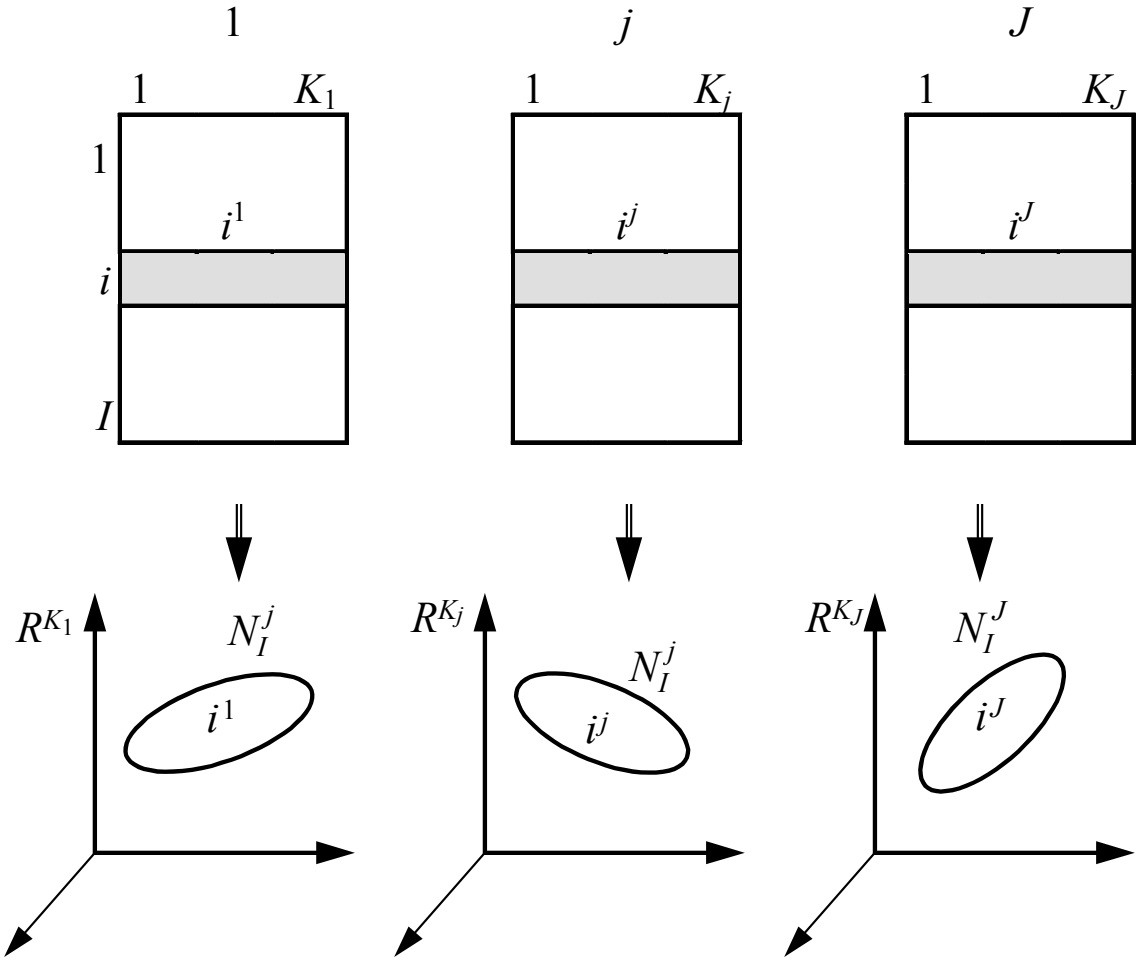








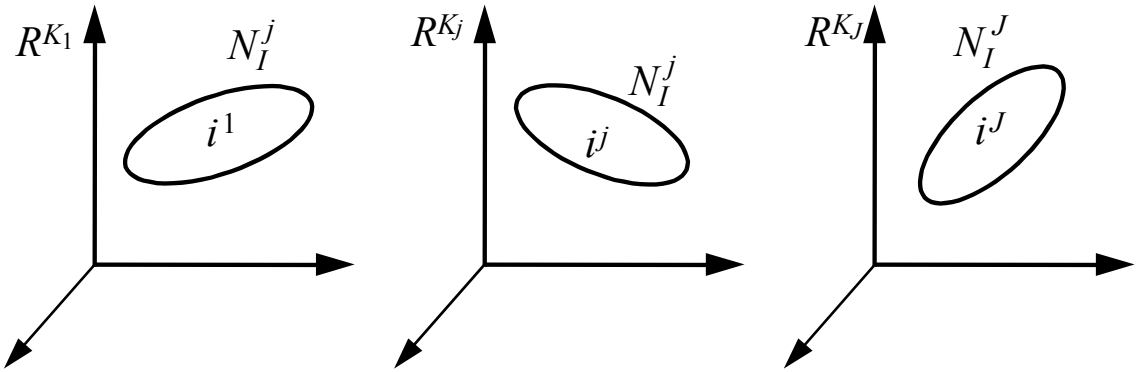
Superimposed representation of the partial clouds of individuals
(e.g. the *J* partial clouds associated to the highest partition in HMFA)

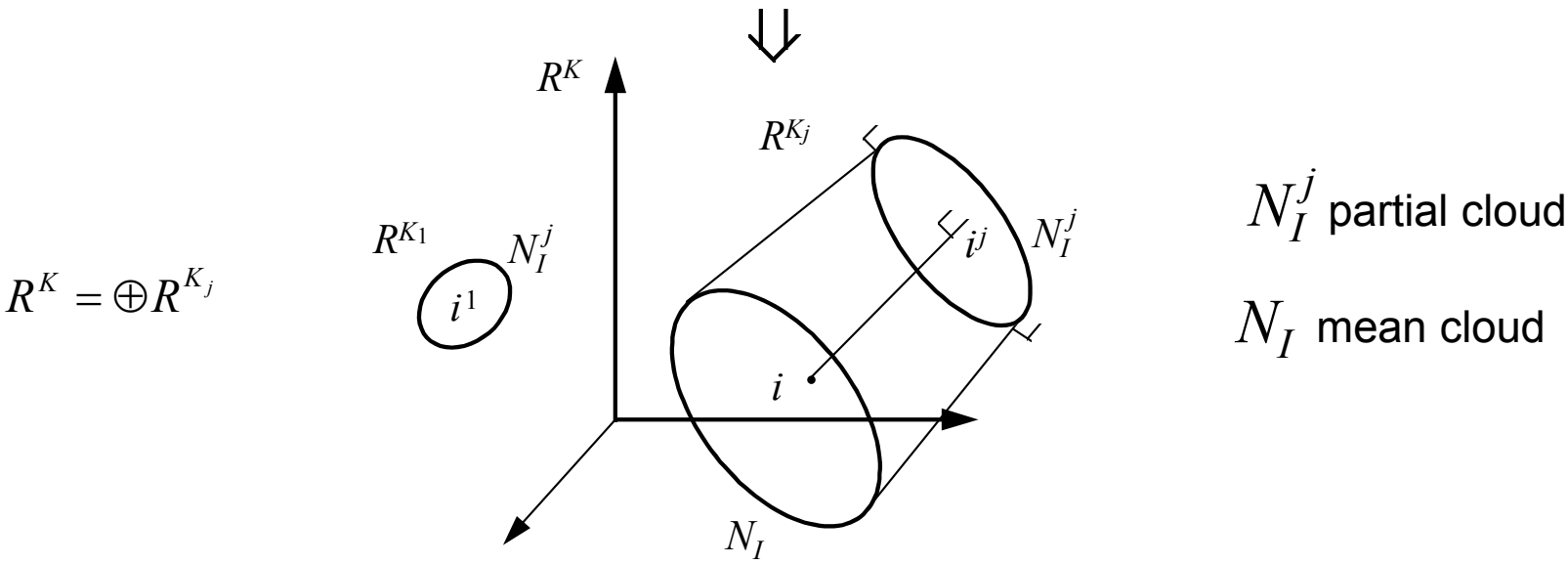


N_I^j : partial cloud (individuals seen through group *j*)

Superimposed representation of the *J* partial clouds of individuals (MFA)

Geometrical framework



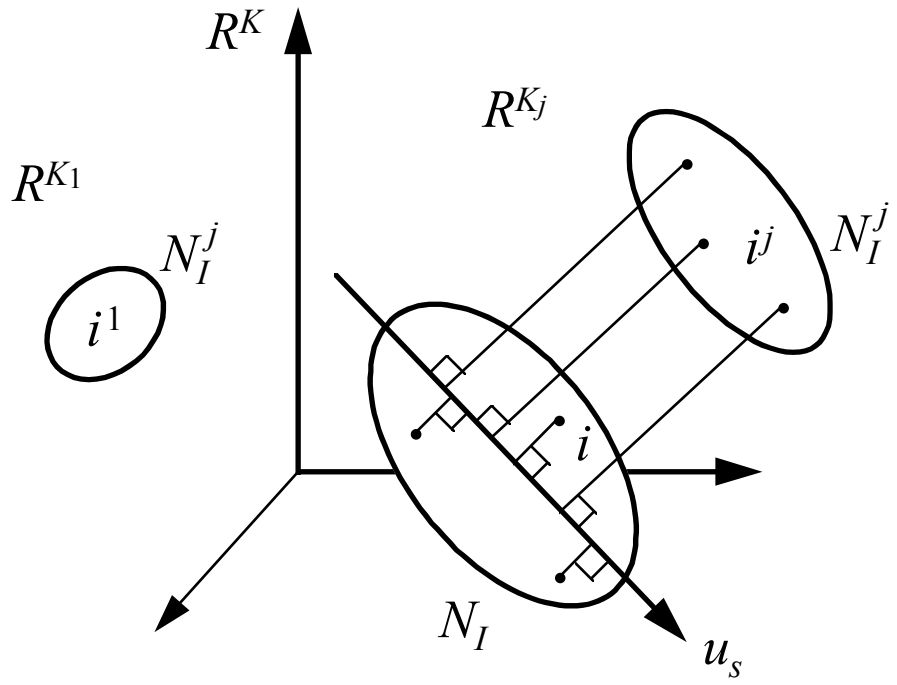


N_I^j partial cloud

N_I mean cloud

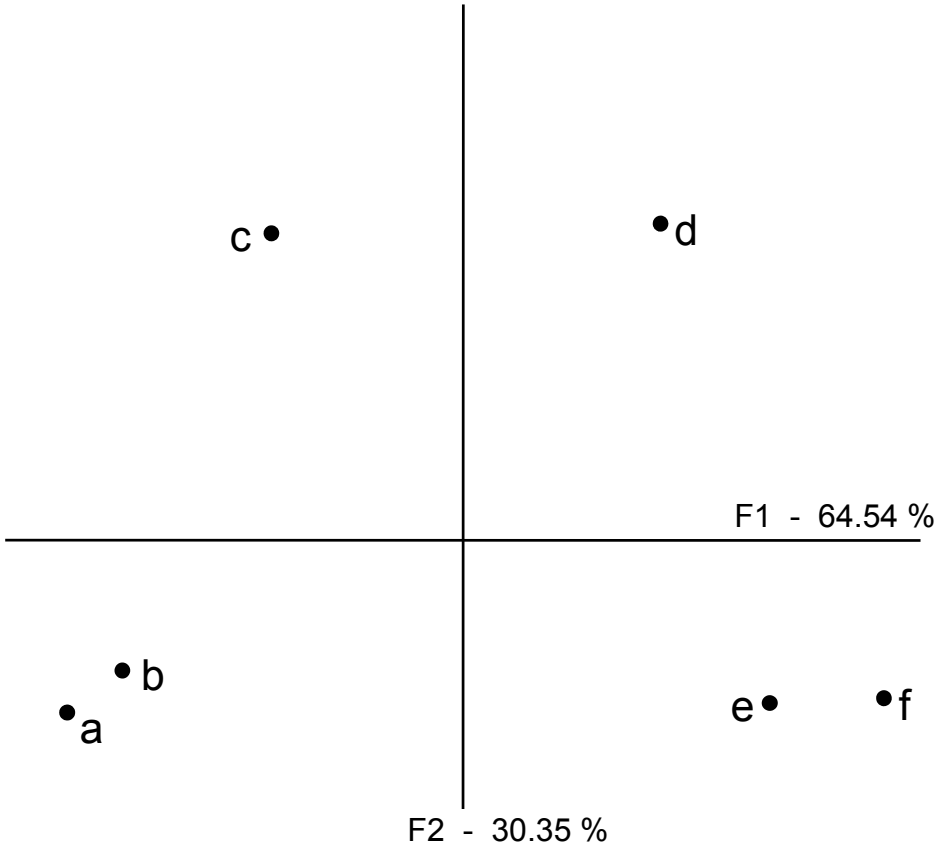
Superimposed representation of the *J* partial clouds of individuals (MFA)

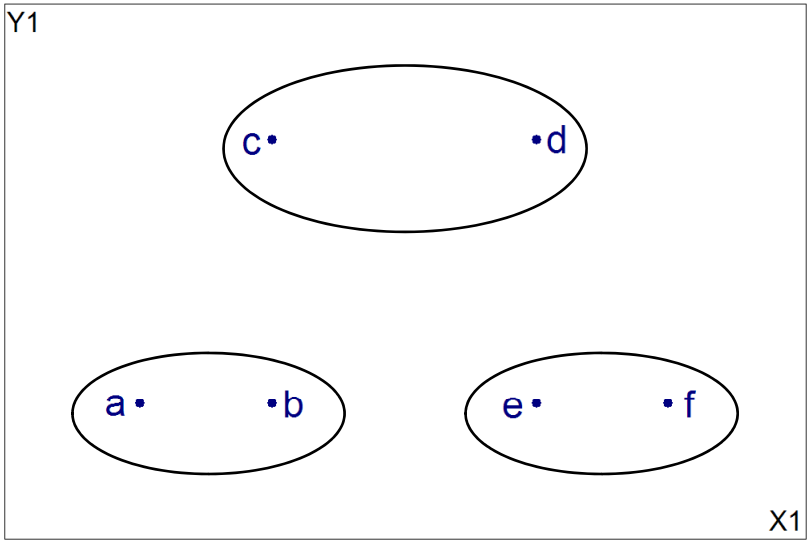
Principle

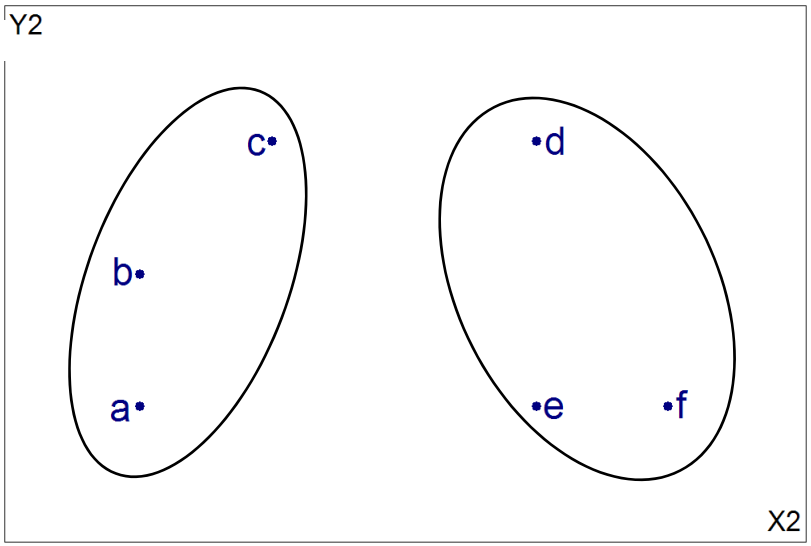


Partial clouds are projected onto principal axes of the mean cloud

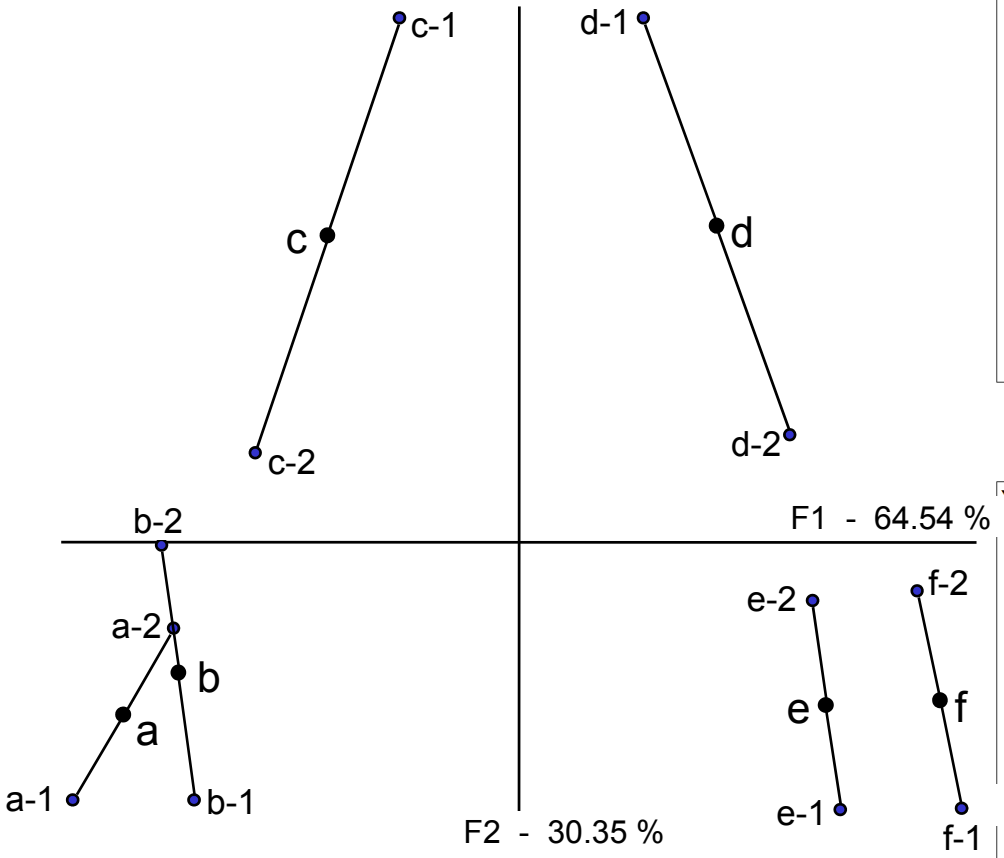
Representation of the partial clouds associated to the two highest nodes

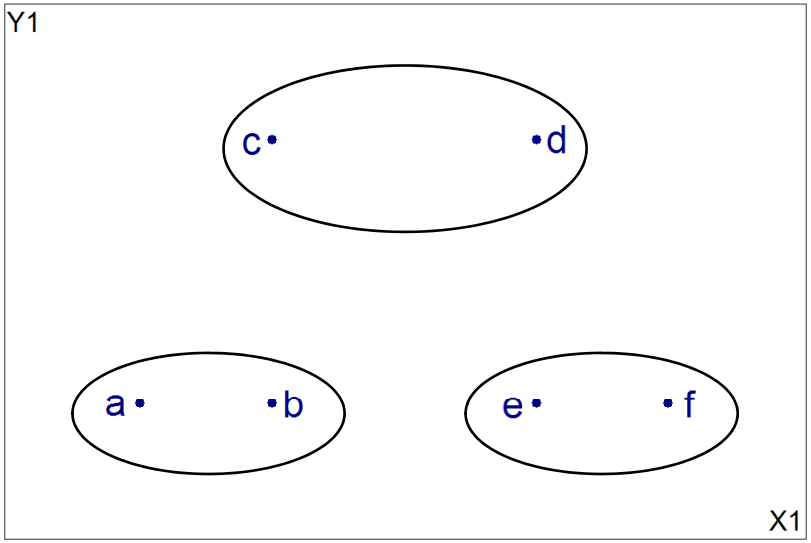


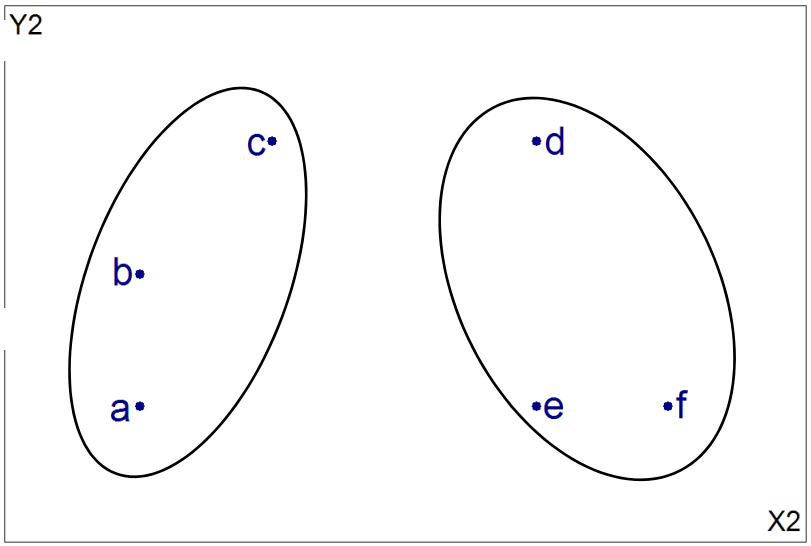




Representation of the partial clouds associated to the two highest nodes







Conclusion

HMFA is a factor analysis devoted to multiple tables in which
a set of individuals is described by
several sets of variables organized according to a hierarchy

The variables can be quantitative or categorical

The core of the method is a weighted factor analysis ; it works
- as a PCA for quantitative variables
- as MCA for categorical variables

It provides results
- usual in any factor analysis
  - representation of individuals, of variables, etc.
- specific to the hierarchy defined on the variables
  - representation of partials points, of nodes, etc.

The analyses were performed with

An R package dedicated to

Exploratory Analysis

See LMA² site