

Analyse simultanée de variables quantitatives et qualitatives

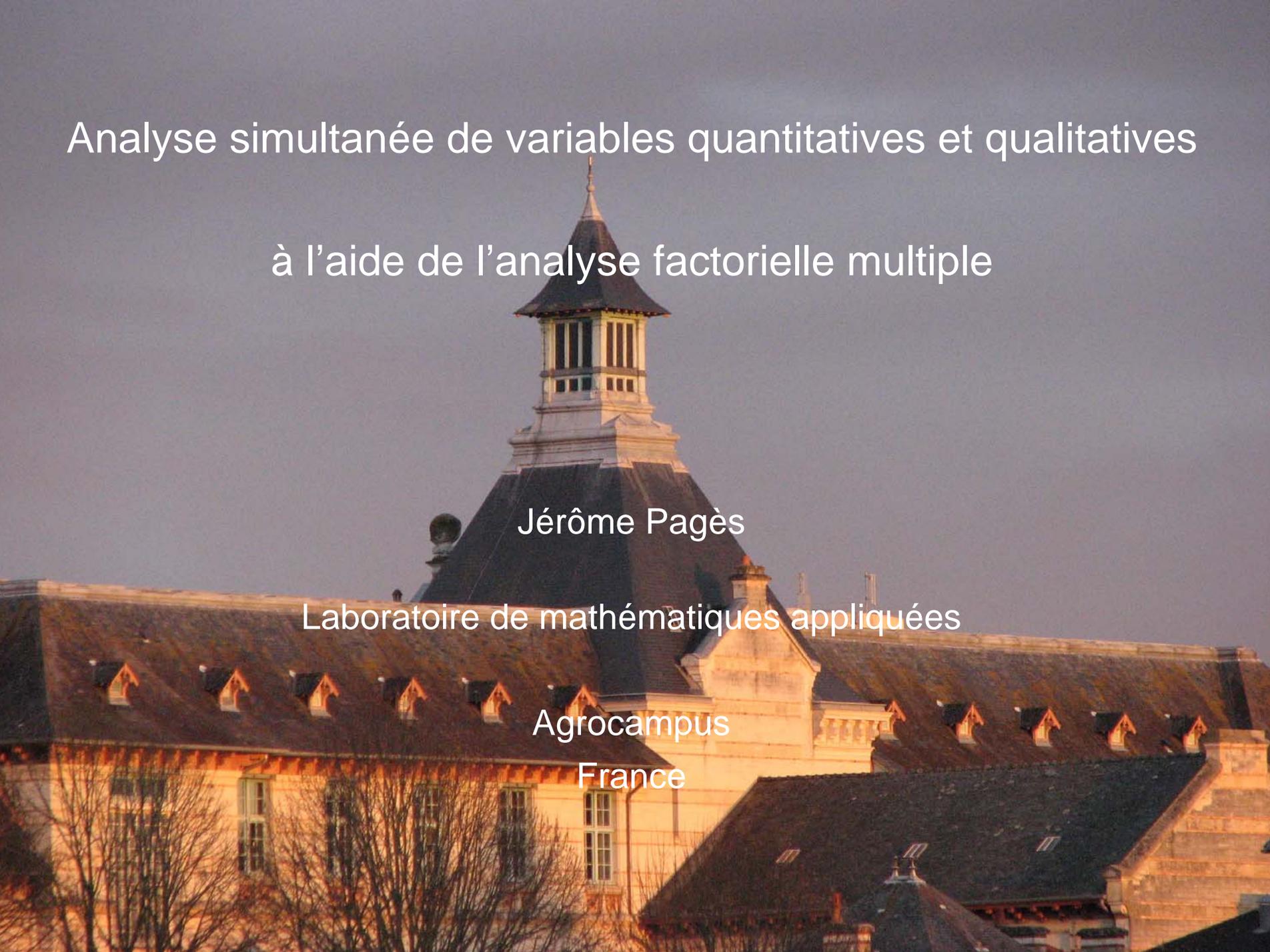
à l'aide de l'analyse factorielle multiple

Jérôme Pagès

Laboratoire de mathématiques appliquées

Agrocampus

France



Analyse Factorielle Multiple (AFM)

Un ensemble d'individus décrit par plusieurs groupes de variables

Exposé en deux étapes

- 1) Un groupe de variables
incluant à la fois des variables quantitatives et qualitatives
- 2) Plusieurs groupes de variables,
quantitatives, qualitatives ou des deux types (données mixtes)

Données

K_1 variables
quantitatives
(centrées réduites)

Q variables
qualitatives

The diagram shows a data matrix with two main sections. The first section, labeled ' K_1 variables quantitatives (centrées réduites)', is indicated by a bracket above the first K_1 columns. The second section, labeled ' Q variables qualitatives', is indicated by a bracket above the next Q columns. The matrix is labeled 'Individus' on the left. The rows are indexed from 1 to I , and the columns are indexed from 1 to K_1 and q to Q . The data point for individual i in the k -th quantitative column is x_{ik} , and the data point for individual i in the q -th qualitative column is x_{iq} .

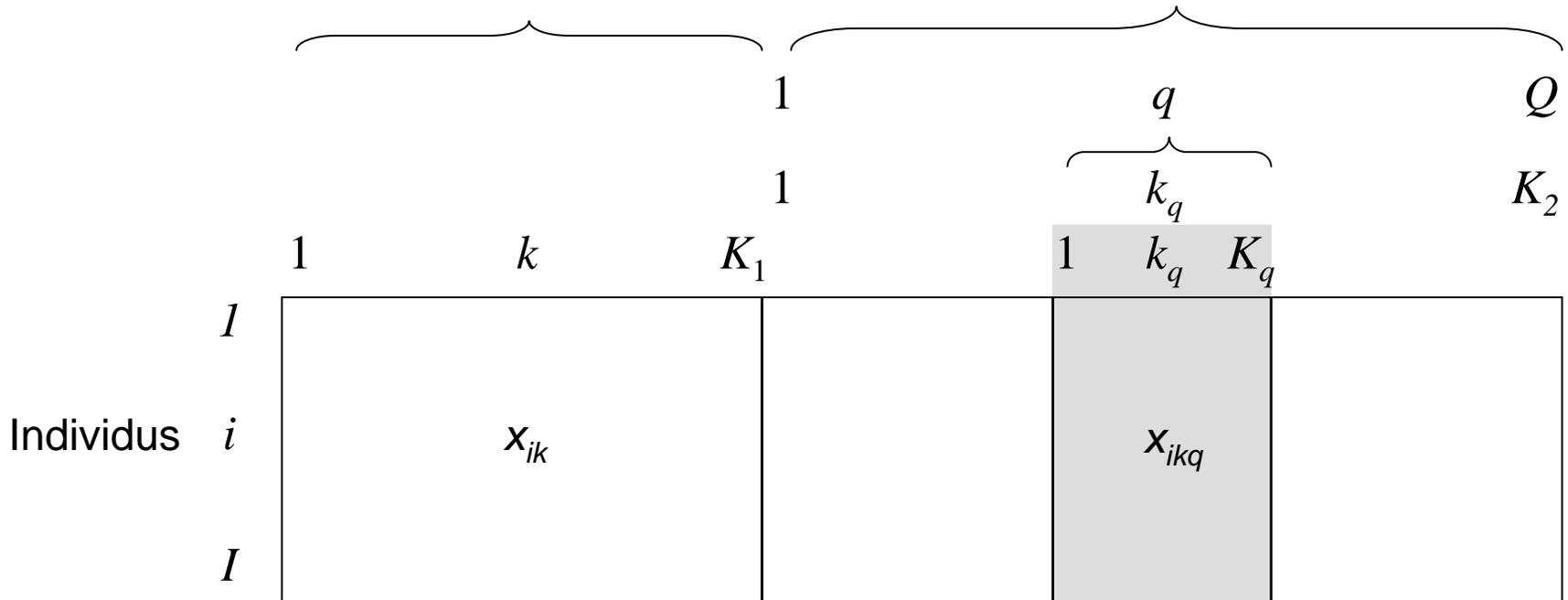
	1	k	K_1	q	Q
1					
i		x_{ik}		x_{iq}	
I					

Individus

Données

K_1 variables
quantitatives
(centrées réduites)

Q variables qualitatives
= K_2 indicatrices
(codage disjonctif complet)



Objectifs

Mettre en œuvre une analyse factorielle dans laquelle les deux types de variables jouent un rôle actif

Solution classique

Diviser l'intervalle de variation des variables quantitatives

Coder les variables quantitatives en qualitatives

Mettre en œuvre une Analyse des Correspondances Multiples (ACM)

Benzécri J.-P., Cazes P., Lebart L., etc.

Excellente solution en pratique

Deux cas « limite »

/ petit

$Q \ll K_1$

Autres solutions, moins connues

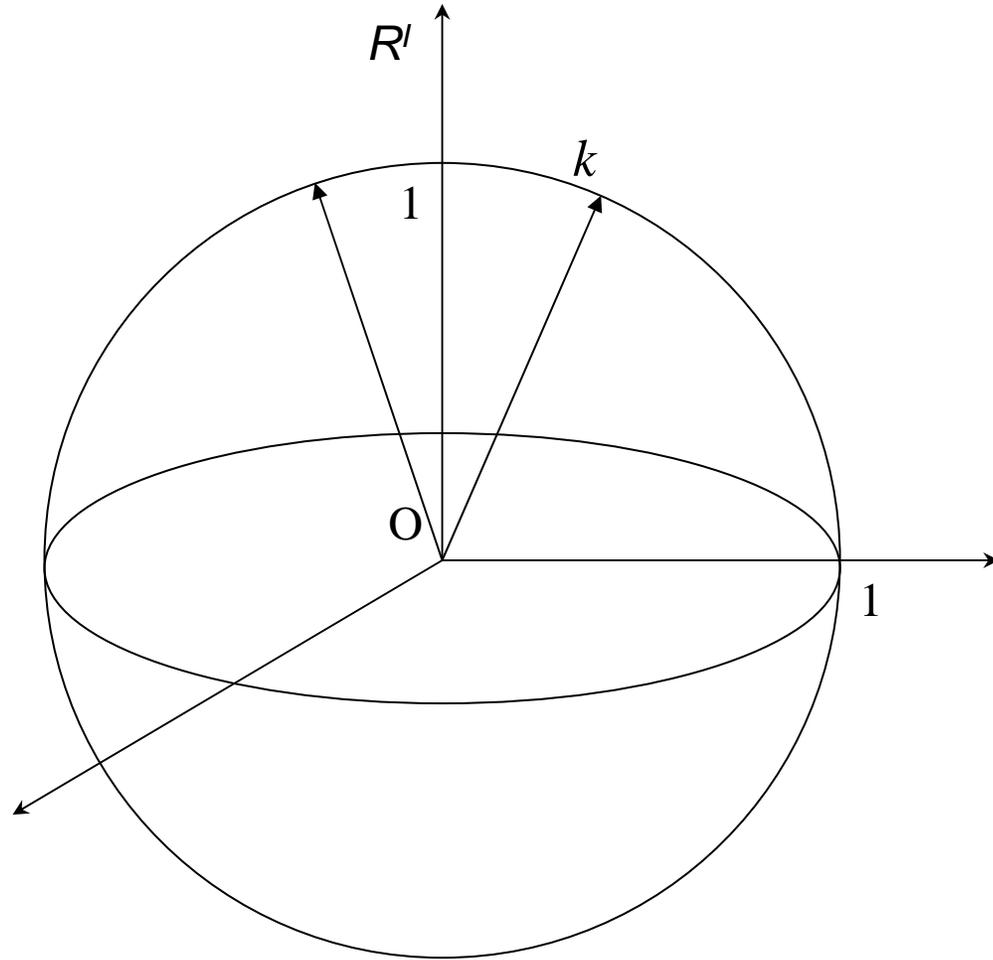
Introduction de variables quantitatives en ACM
Via un codage particulier
B. Escofier CAD 1979 (PUR 2003)

Introduction d'indicateurs dans une ACP
via une métrique particulière
G. Saporta *in* Revue Soc. ital. de stat. 1990

AFM de données mixtes
B. Escofier & J. Pagès (Dunod 1988 - 1997)
J. Pagès RSA 2002
idée : une variable = un groupe
Abascal E. *et al.* Carme 2003

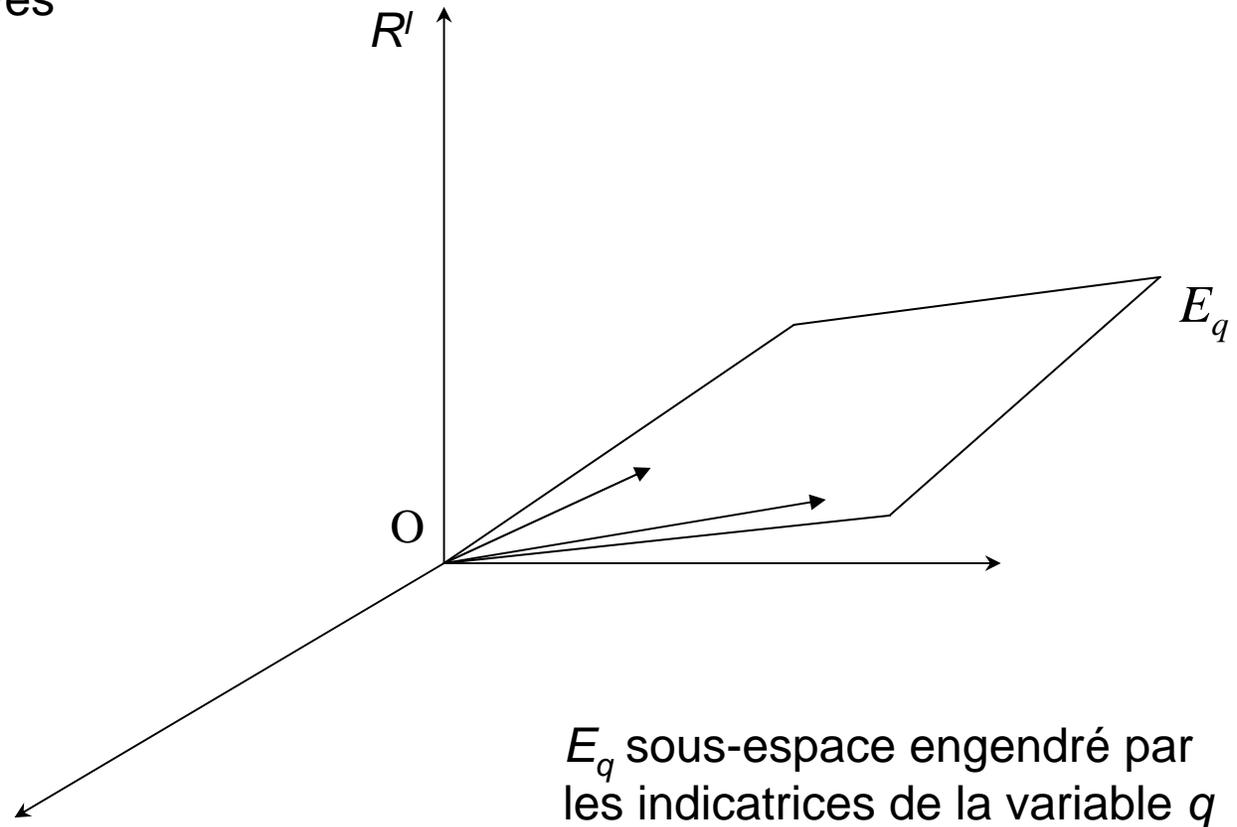
Représentation des variables dans R^l

Variables quantitatives
en ACP normée



Représentation des variables dans R^l

Variables qualitatives
en ACM



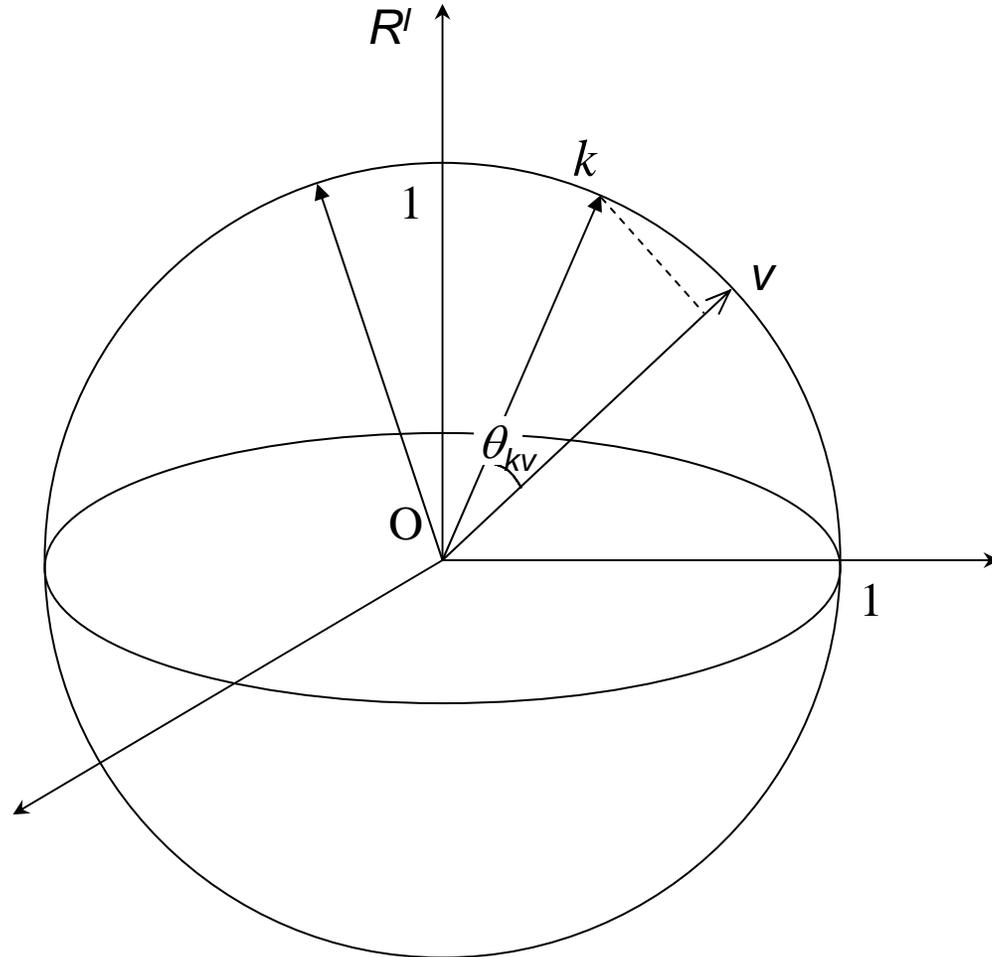
L'inertie des indicatrices de la variable q
est égale à 1 dans toutes les directions de E_q

Critère

Cas de l'ACP normée

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

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

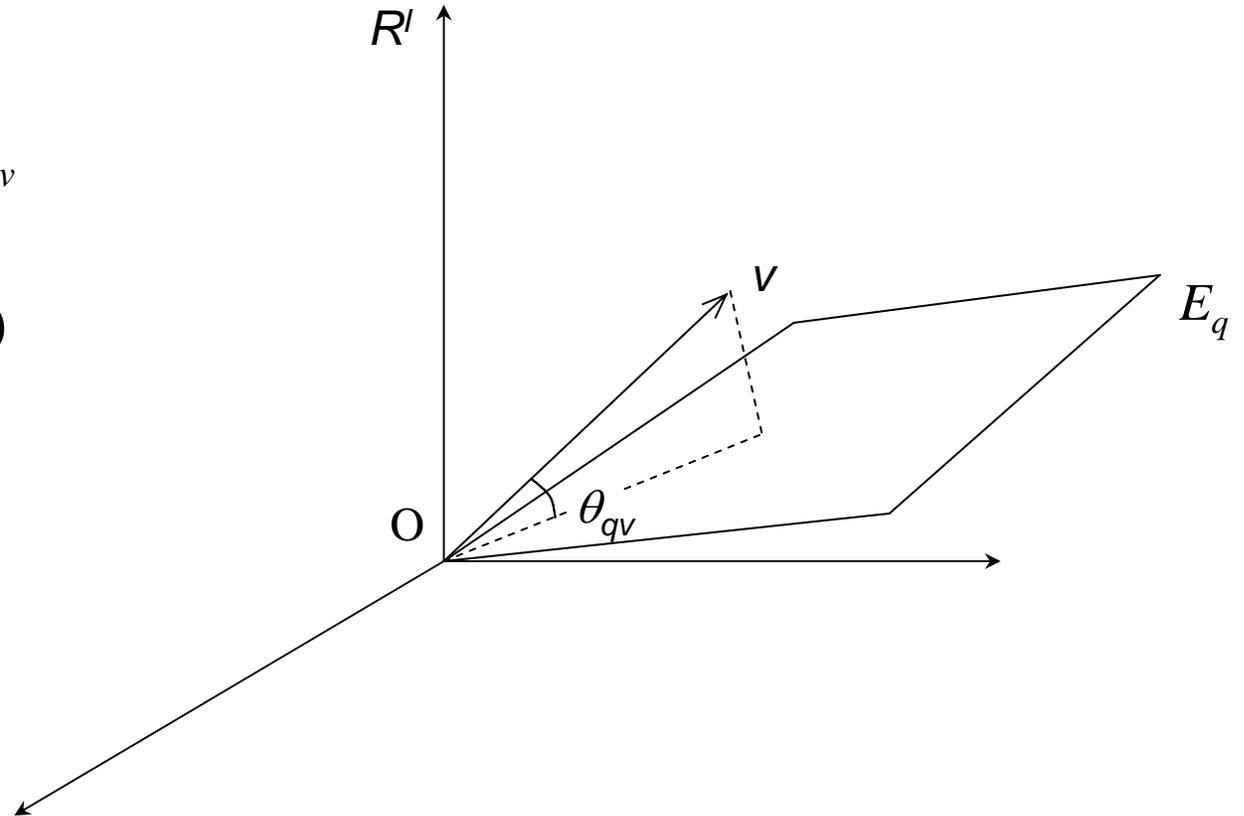


Critère

Cas de l'ACM

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

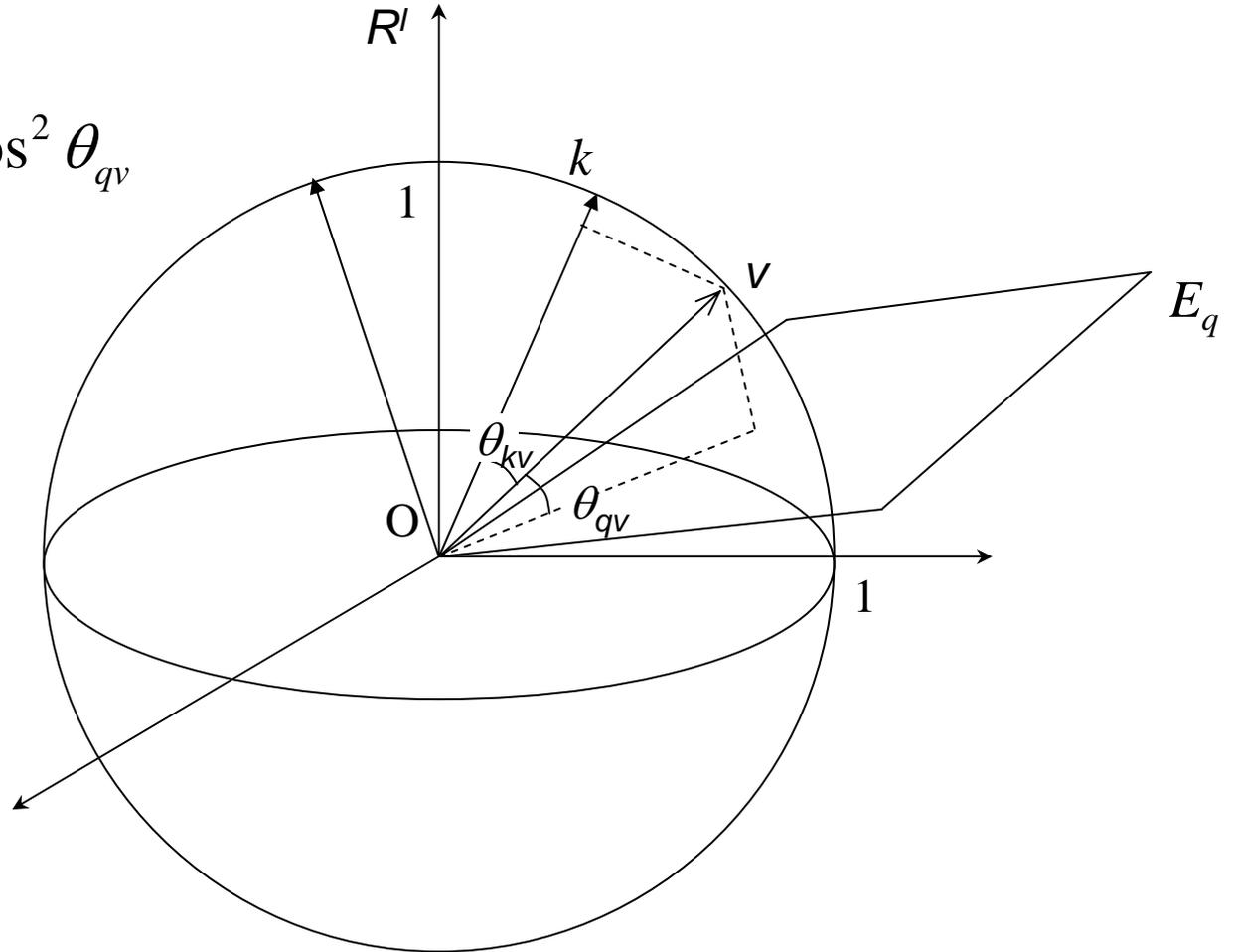
$$\sum_q \eta^2(q, v)$$



Méthode proposée par B. Escofier (1979)

Critère

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



Méthode proposée par G. Saporta (1990)

Critère

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

En pratique

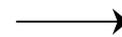
Les indicatrices sont juxtaposées aux variables quantitatives

L'ensemble est traité par une ACP avec une métrique particulière

Même méthode vue

dans le cadre de l'ACM (B. Escofier)

dans le cadre de l'ACP (G. Saporta)

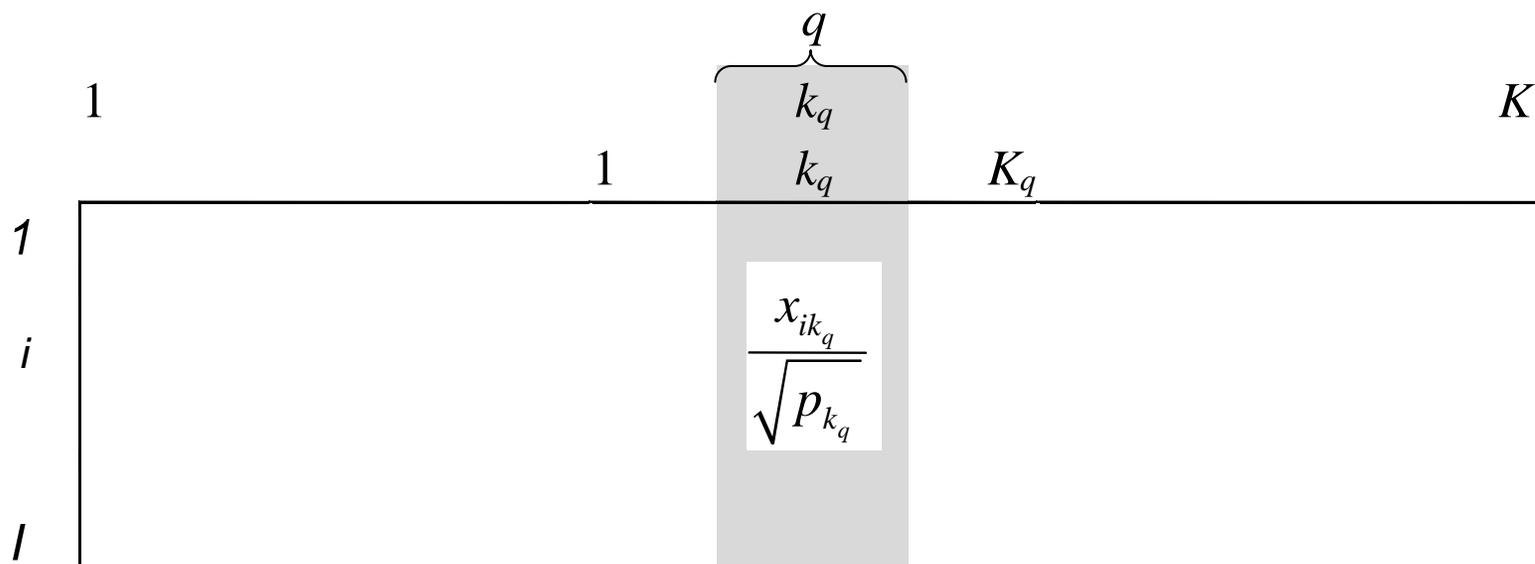


Analyse factorielle
de données mixtes
(AFDM)

AFDM à l'aide d'une ACP

Codage des indicatrices
pour garder, en ACP,
leurs propriétés en ACM

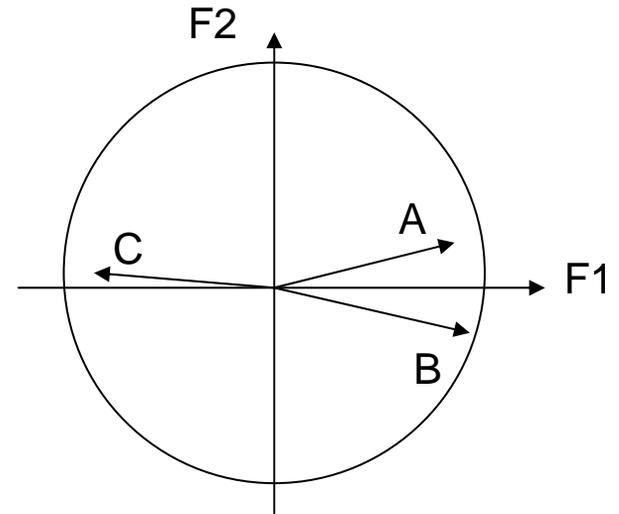
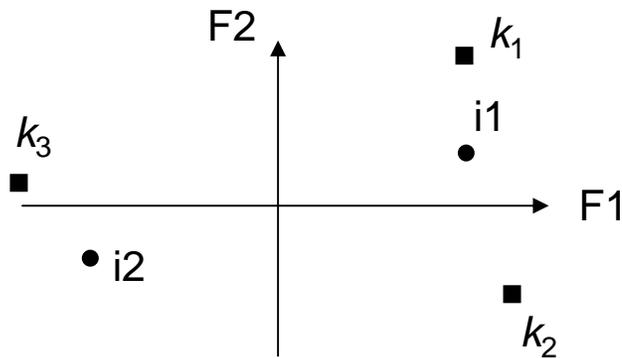
Q variables qualitatives = K_2 indicatrices



p_{k_q} fréquence de la modalité k de la variable q

Représentations fournies par l'AFDM

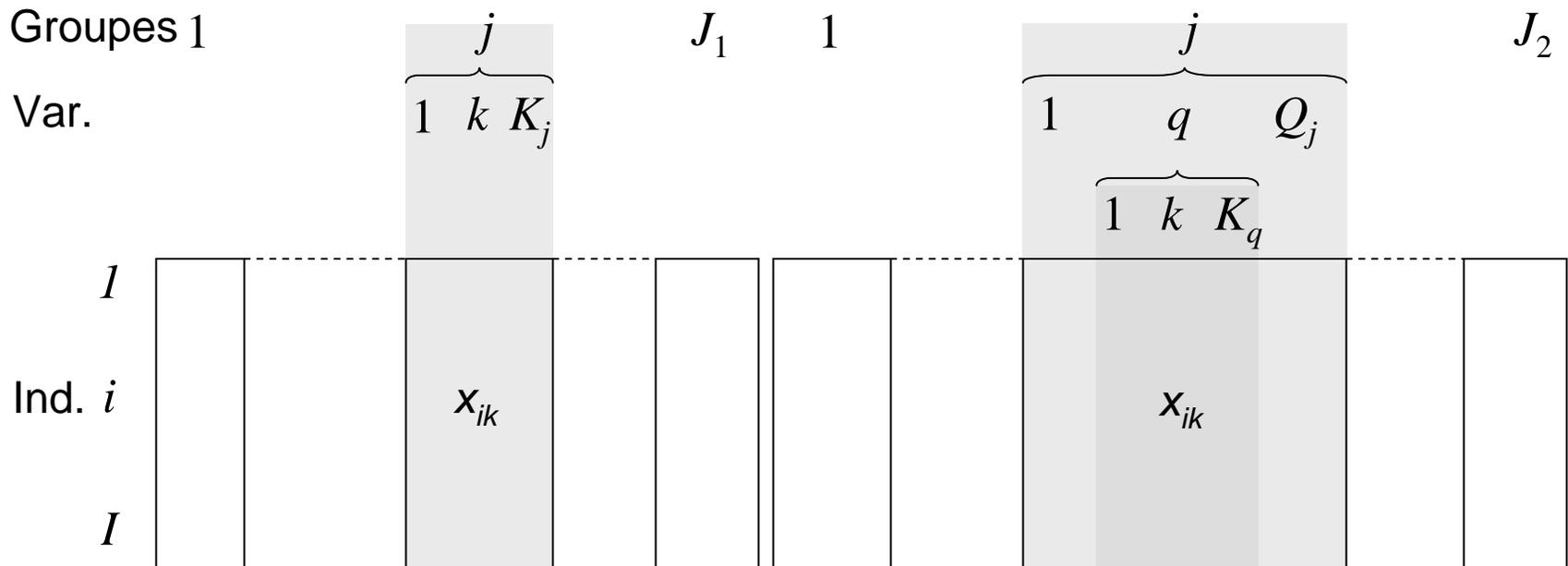
- individu
- modalité



AFM appliquée à plusieurs groupes de variables :
quantitatifs, qualitatifs ou mixtes

J_1 groupes quantitatifs

J_2 groupes qualitatifs
(codage disjonctif complet)



Données : exemple

Perception de 5 outils pédagogiques par des étudiants

Cours en ligne :

- 1) texte
- 2) animations
- 3) film sur logiciels

Livres

- 4) cours
- 5) exercices

Chaque outil est noté par chaque étudiant sur une échelle en 5 points

de 1 = *inutile*

à 5 = *très utile*

Tableau : 25 étudiants x 5 outils avec x_{ik} la note donnée par l'étudiant i à l'outil k

Analyse exploratoire

Les variables peuvent être considérées comme quantitatives (ACP) ou qualitatives (ACM)

Double objectif

Comparaison ACP - ACM

Introduction des deux points de vue dans une analyse unique

Données : exemple

5 outils
= variables quantitatives
(centrées-réduites)

5 outils
= 25 indicatrices
(codage disjonctif complet)

		5 outils = variables quantitatives (centrées-réduites)					5 outils = 25 indicatrices (codage disjonctif complet)				
		1					q				
		1					k				
		5					1 k 5				
I											
Individus i		x_{ik}					x_{ik}				
I											

Pondération des groupes de variables

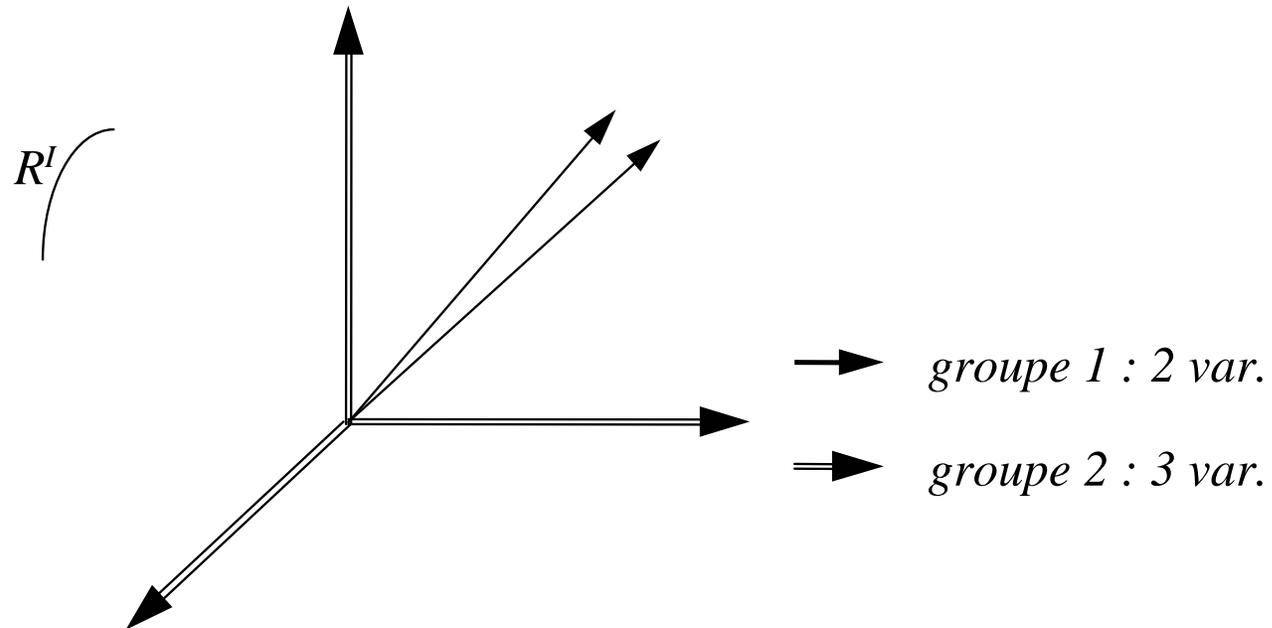
Plusieurs groupes de variables actives dans une analyse unique

Question

Comment équilibrer leur influence ?

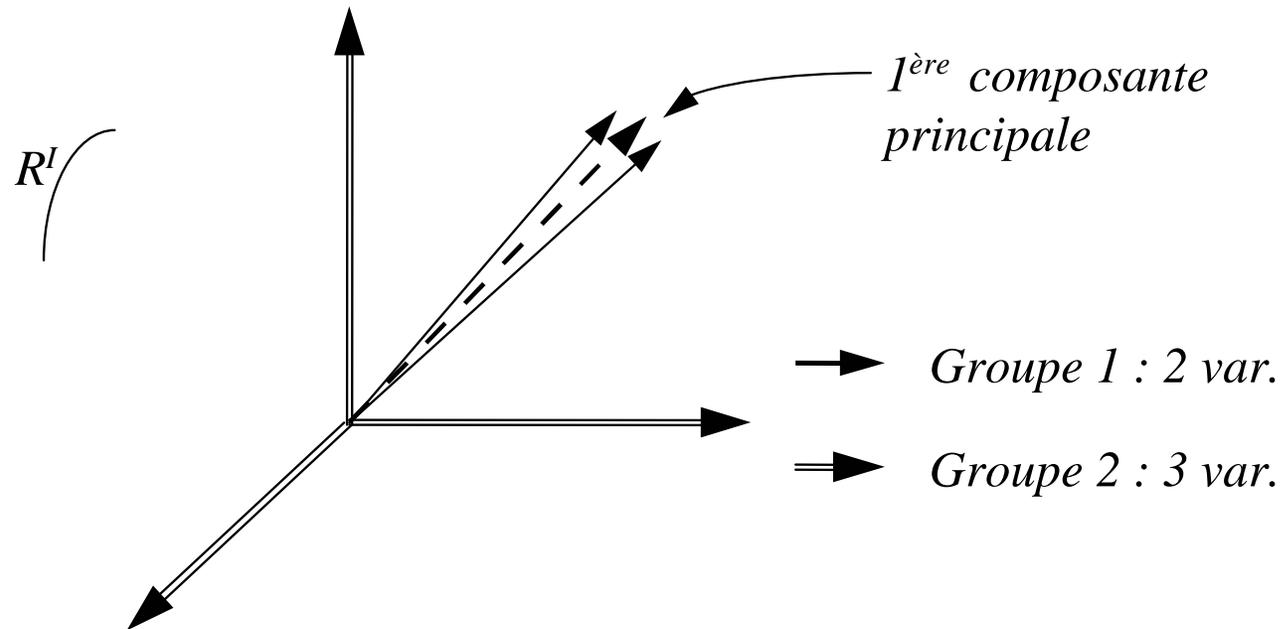
Pondération des groupes de variables

Exemple de référence : deux groupes de variables quantitatives



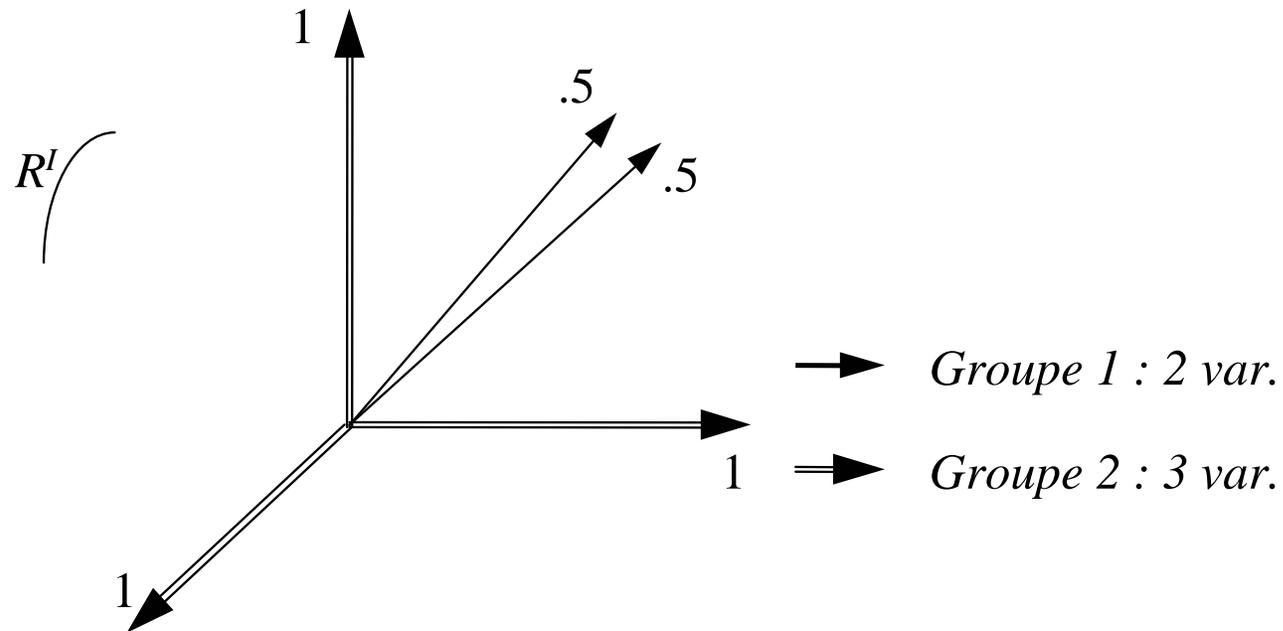
Pondération des groupes de variables

ACP des 5 variables, sans considérer les groupes



Exemple de référence

Pondération des groupes dans l'AFM
Equilibrer l'inertie axiale maximum



Chaque variable du groupe j est pondérée par $1/\lambda_1^j$
 λ_1^j : 1^{ère} valeur propre de l'ACP séparée du groupe j .

Le cœur de l'AFM est une analyse factorielle de tous les groupes actifs

Les groupes de variables peuvent être :

quantitatifs (centrés-réduits ou non)

qualitatifs

mixtes

Critère (cas de 2 groupes : K_1 variables quantitatives + Q_2 variables qualitatives)

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

Equivalences en AFM
quand chaque groupe
contient une seule variable

Variables quantitatives

ACP normée

Variables qualitatives

ACM

Données mixtes

AFDM

Le cœur de l'AFM est une analyse factorielle de tous les groupes actifs

L'AFM fournit :

1) : les résultats classiques de l'analyse factorielle

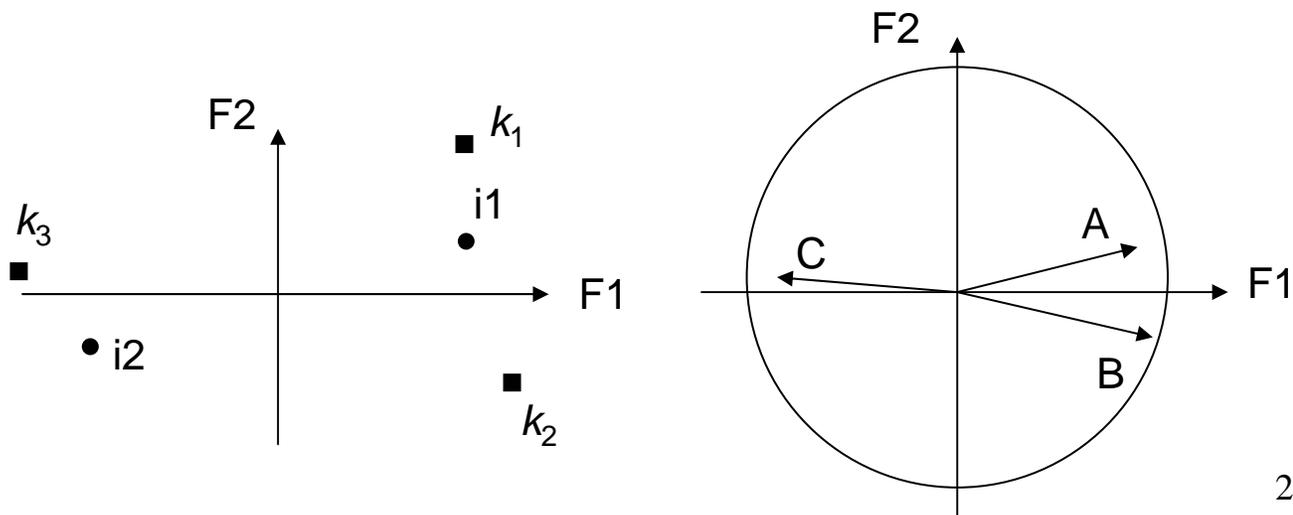
Soit, pour chaque axe :

Coordonnées, contributions et Cos^2 des individus

Coefficient de corrélation entre facteurs et variables continues

Coordonnées des modalités (au barycentre de leurs individus)

- individu
- modalité



Valeurs propres des analyses séparées

	F1	F2	F3	F4	F5
MCA	0,637	0,617	0,454	0,418	0,369
PCA	2,837	1,870	0,148	0,081	0,063

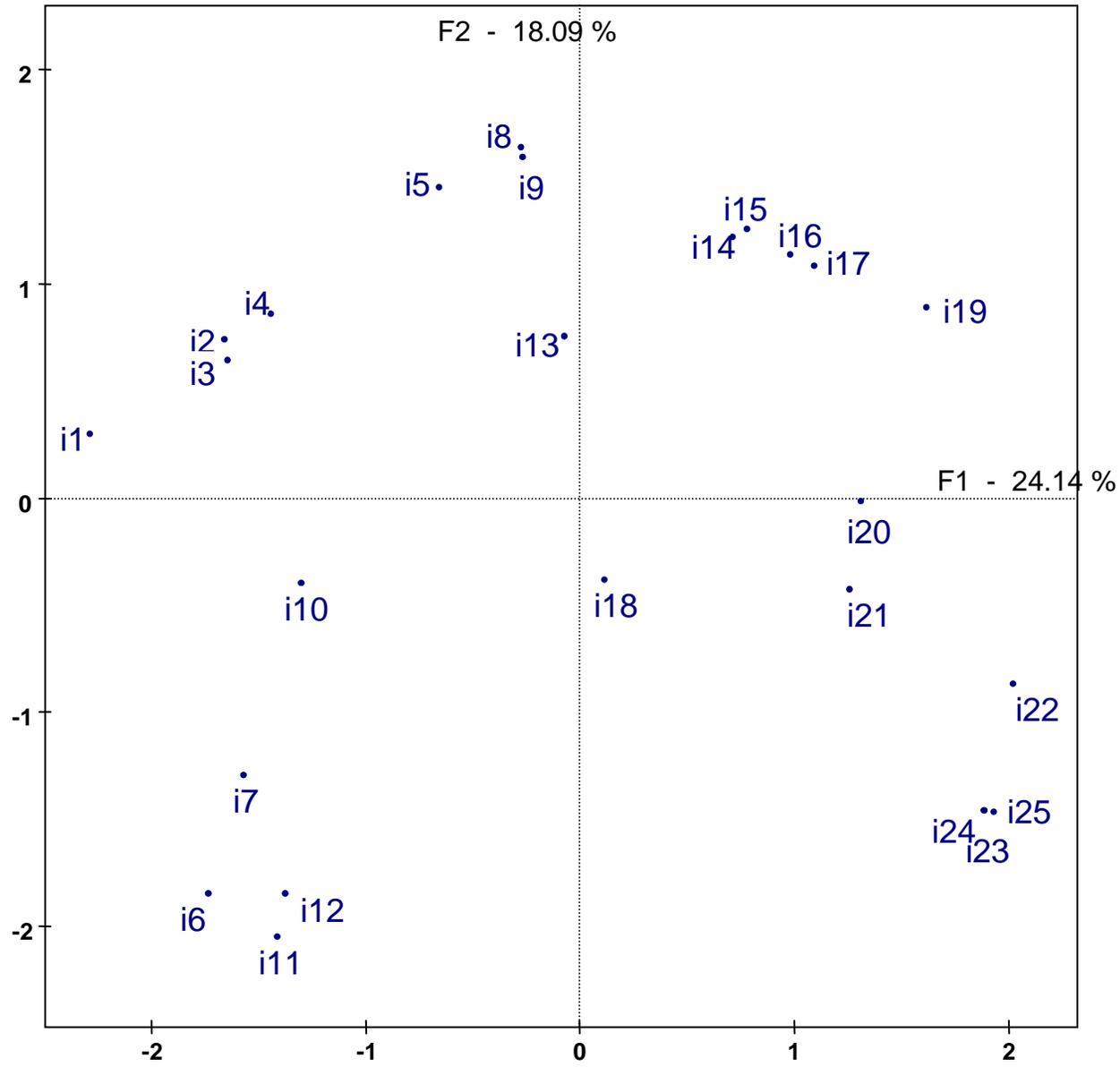
Valeurs propres des analyses séparées (% cumulés)

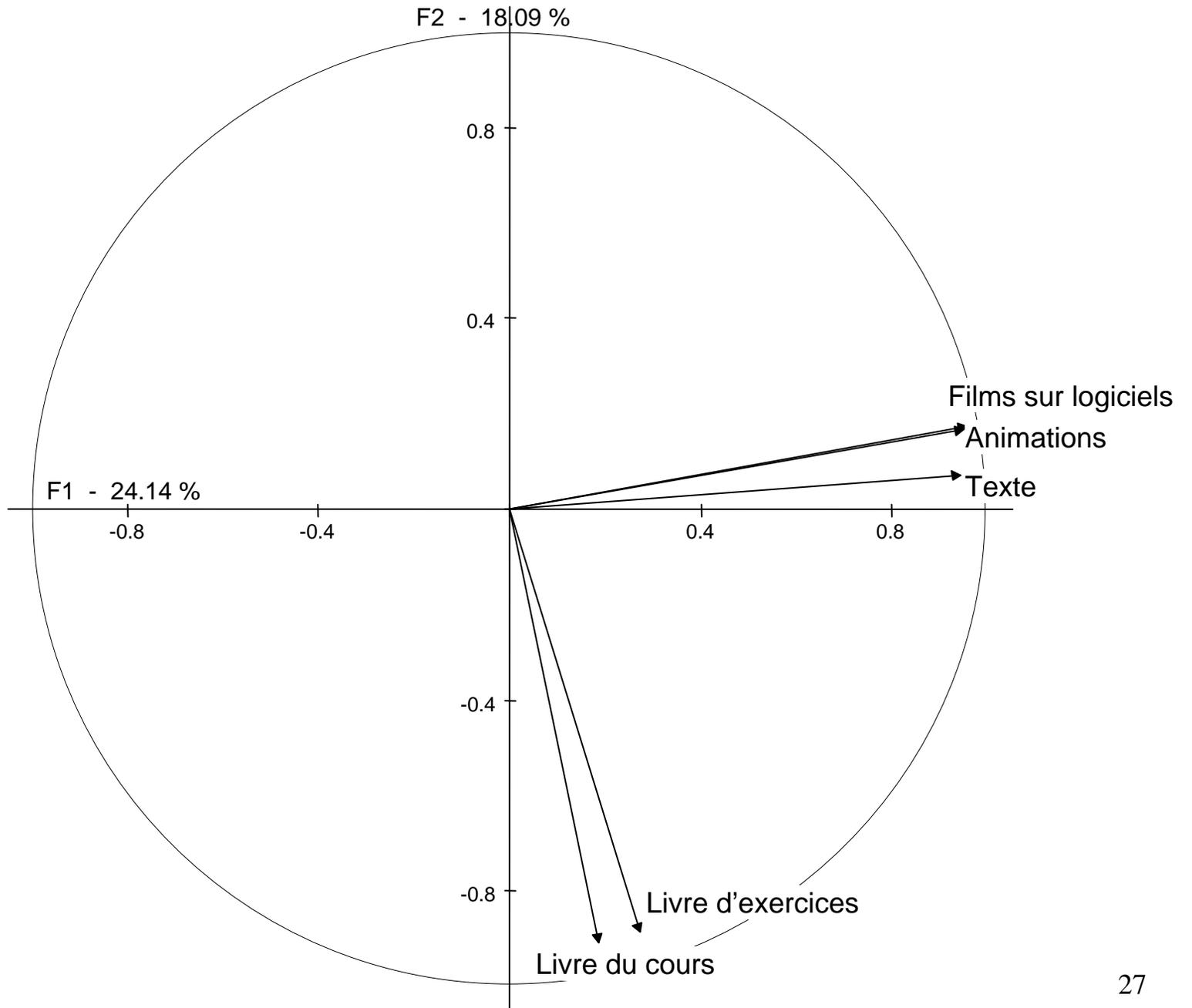
	F1	F2	F3	F4	F5
MCA	15,92	31,33	42,68	53,12	62,35
PCA	56,75	94,16	97,12	98,74	100,00

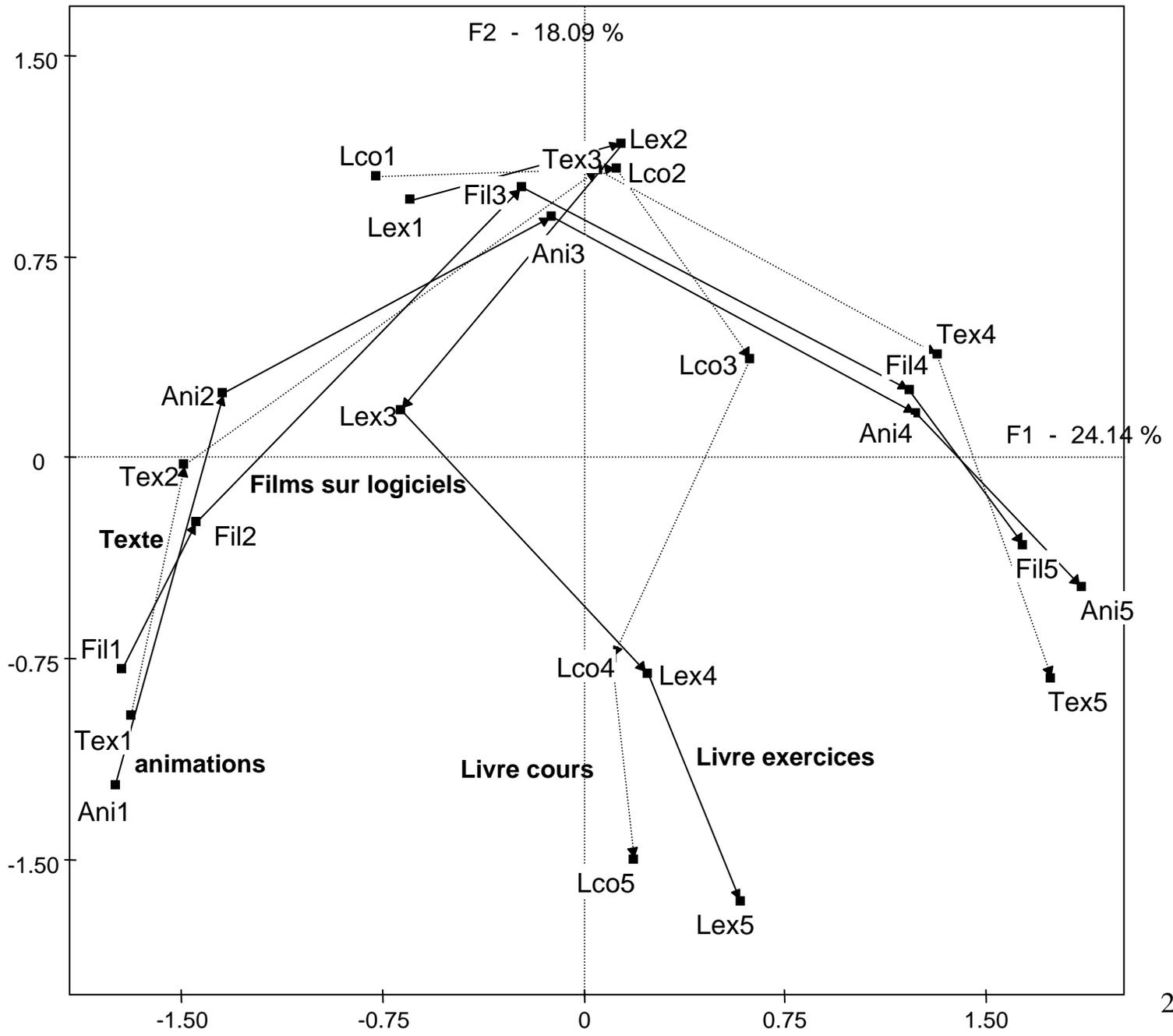
Décomposition de l'inertie dans l'AFM

	F1	F2	F3	F4	F5
MFA	1,942	1,455	0,826	0,694	0,612
	24,14%	18,09%	10,27%	8,63%	7,61%
MCA	0,950	0,864	0,781	0,689	0,597
PCA	0,992	0,591	0,045	0,005	0,015

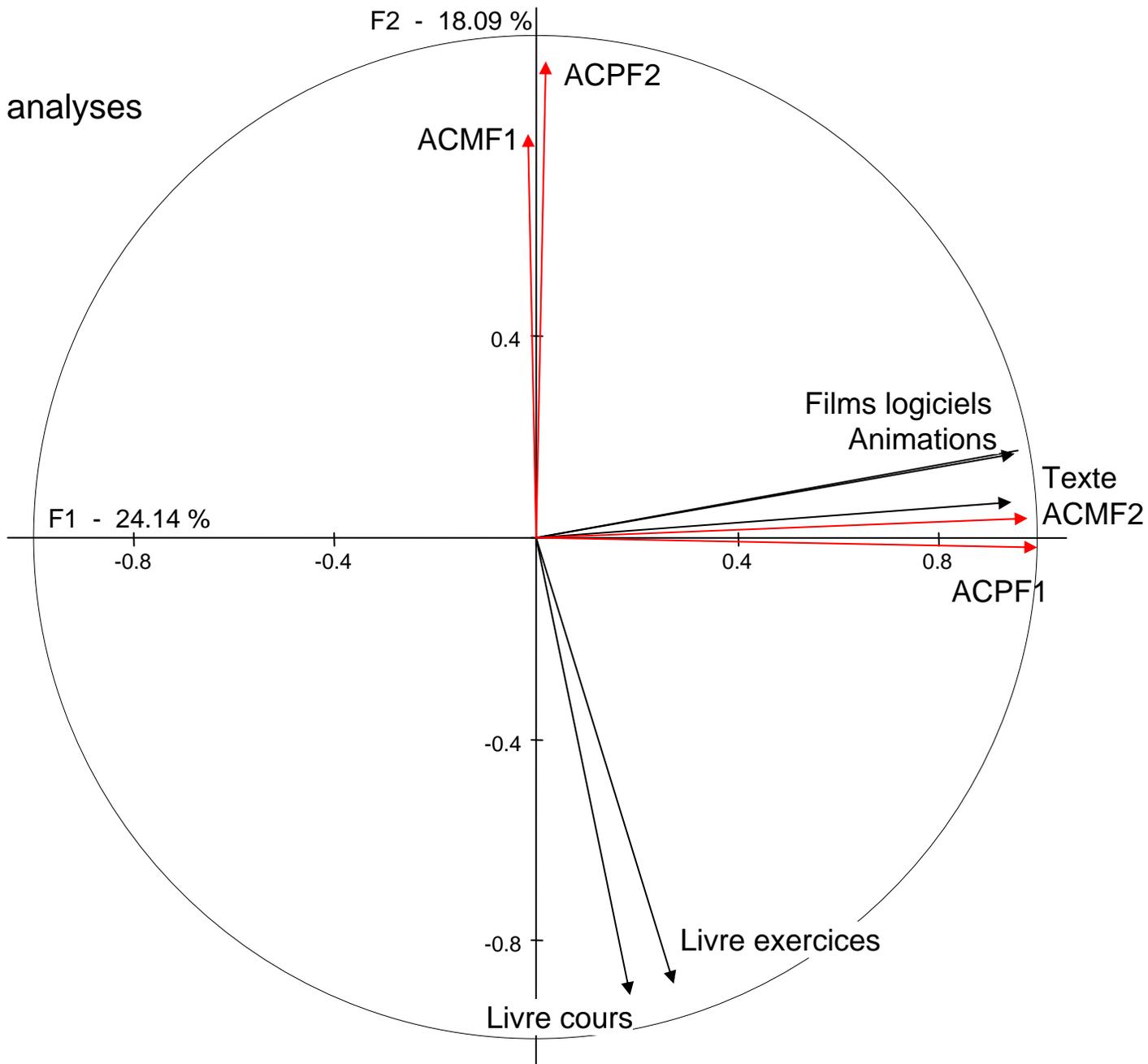
Les 25 étudiants



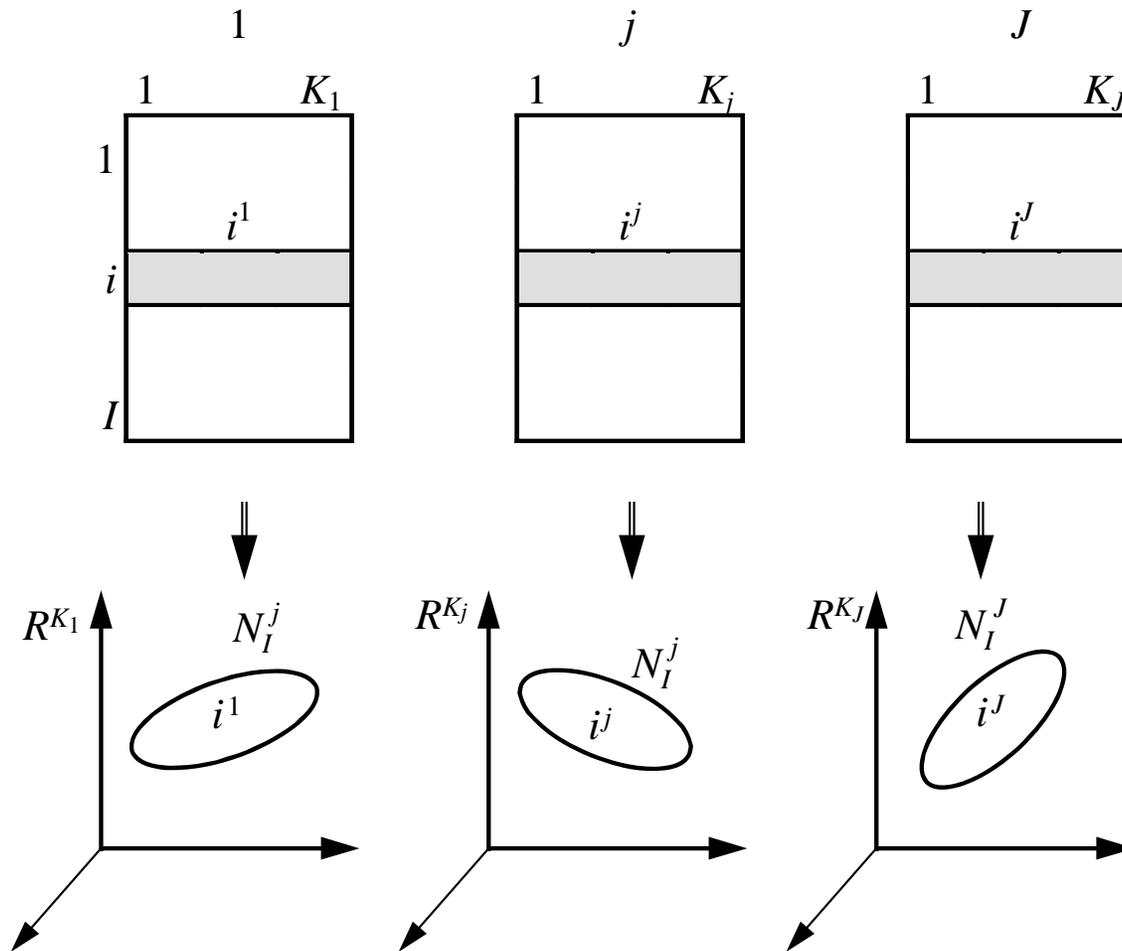




Axes des analyses
séparées

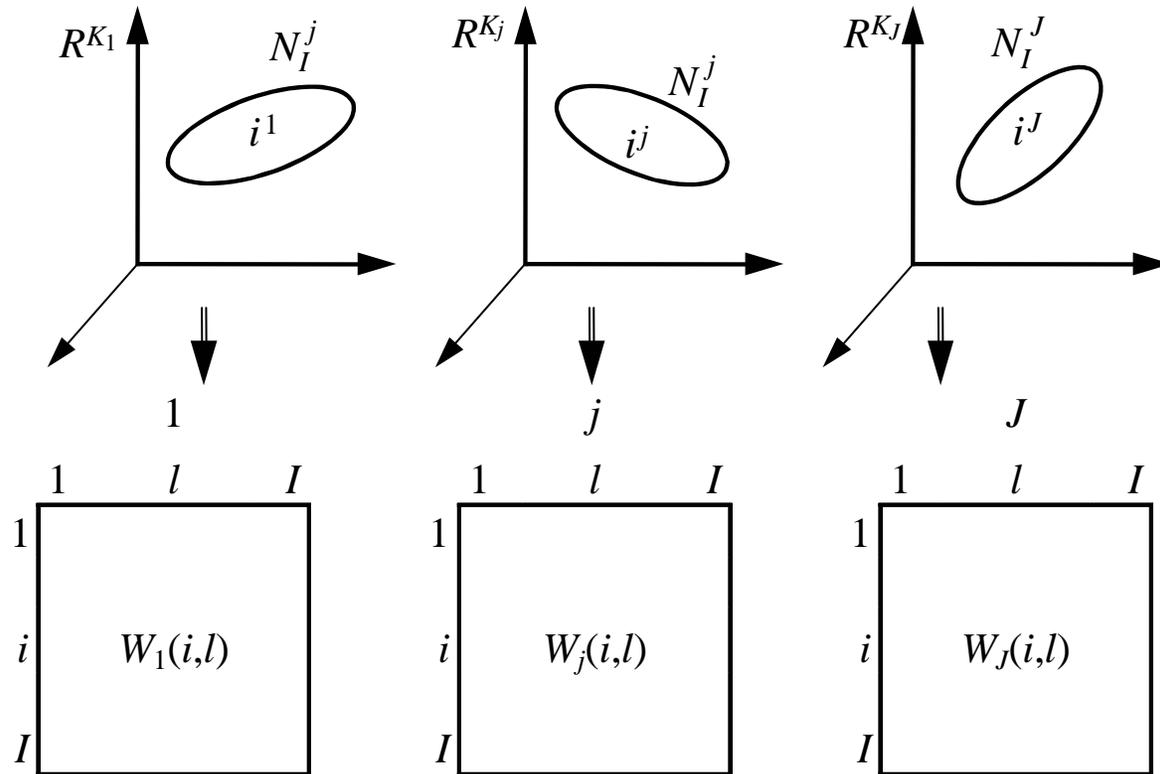


Représentation globale des groupes de variables



N_I^j : nuage partiel (des individus ; associé au groupe j)

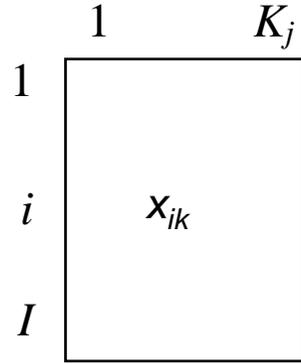
Etudier les ressemblances globales entre les N_i^j ?



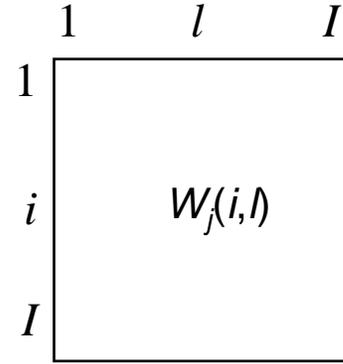
matrices des produits scalaires entre individus pour chaque groupe j

$$W_j = X_j X_j'$$

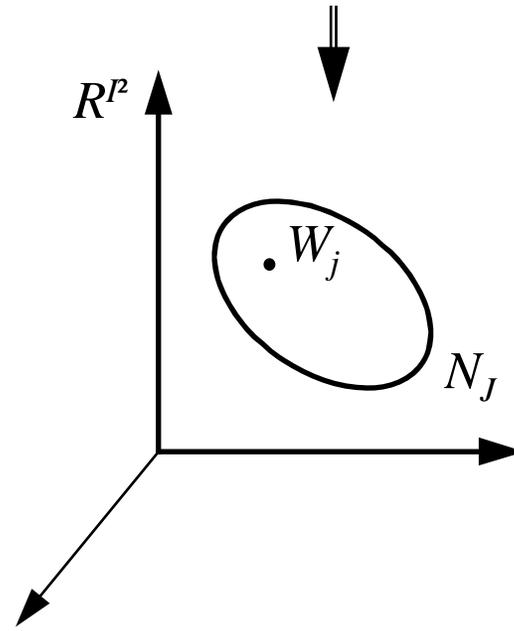
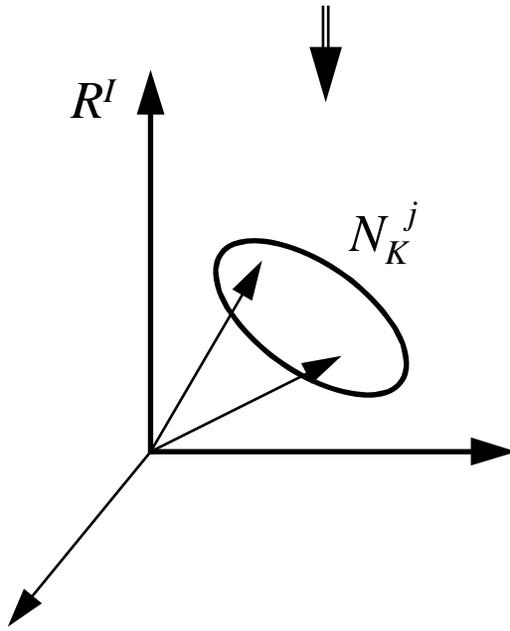
Données



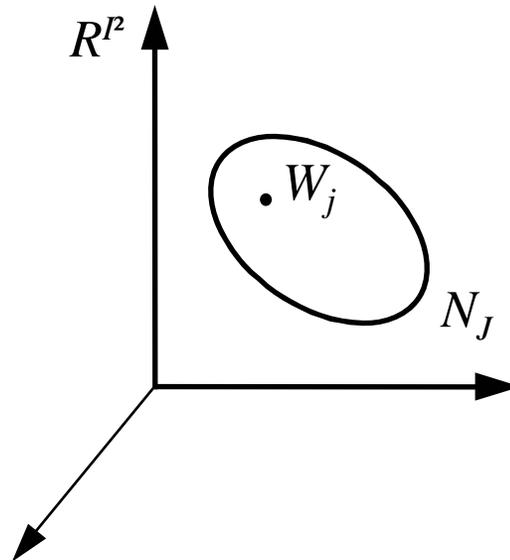
Produits scalaires



→



N_J : nuage des groupes



Etudier le nuage N_J

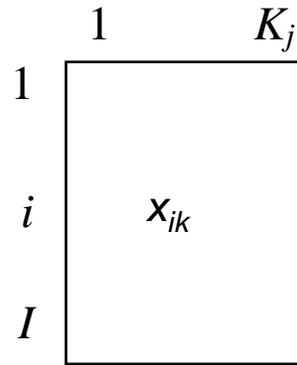
Méthode de référence : STATIS (Escoufier Y., Lavit C.)

Axes principaux de N_J

Interprétation de ces axes difficile

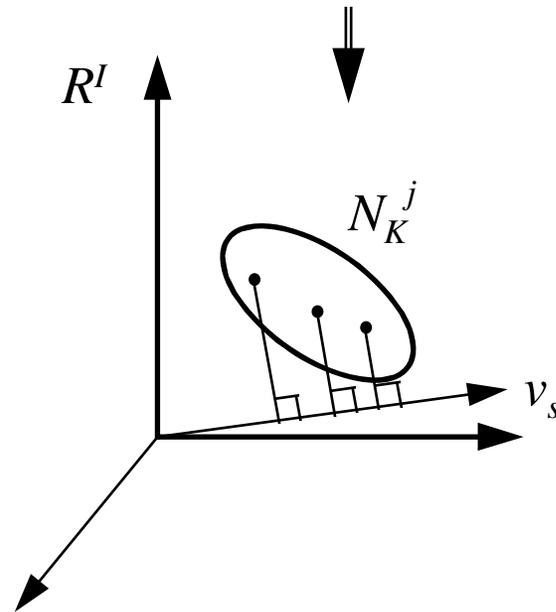
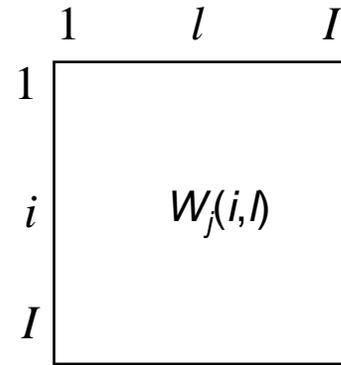
Etudier N_j avec l'AFM

Données

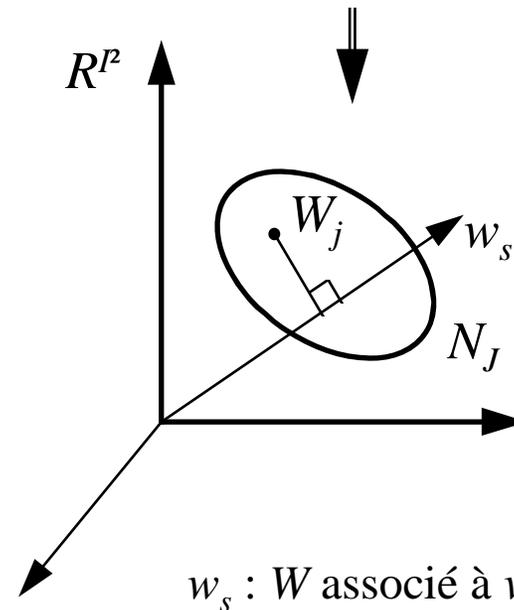


→

Produits scalaires

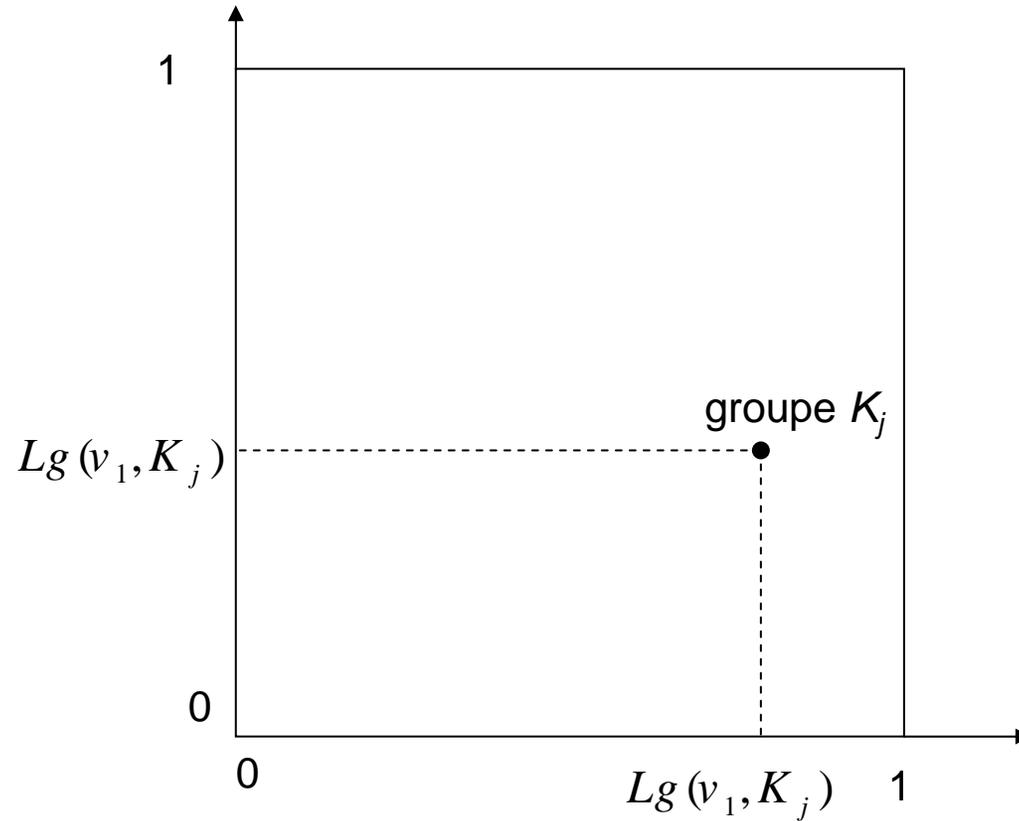


inertie de N_K^j projeté sur v_s



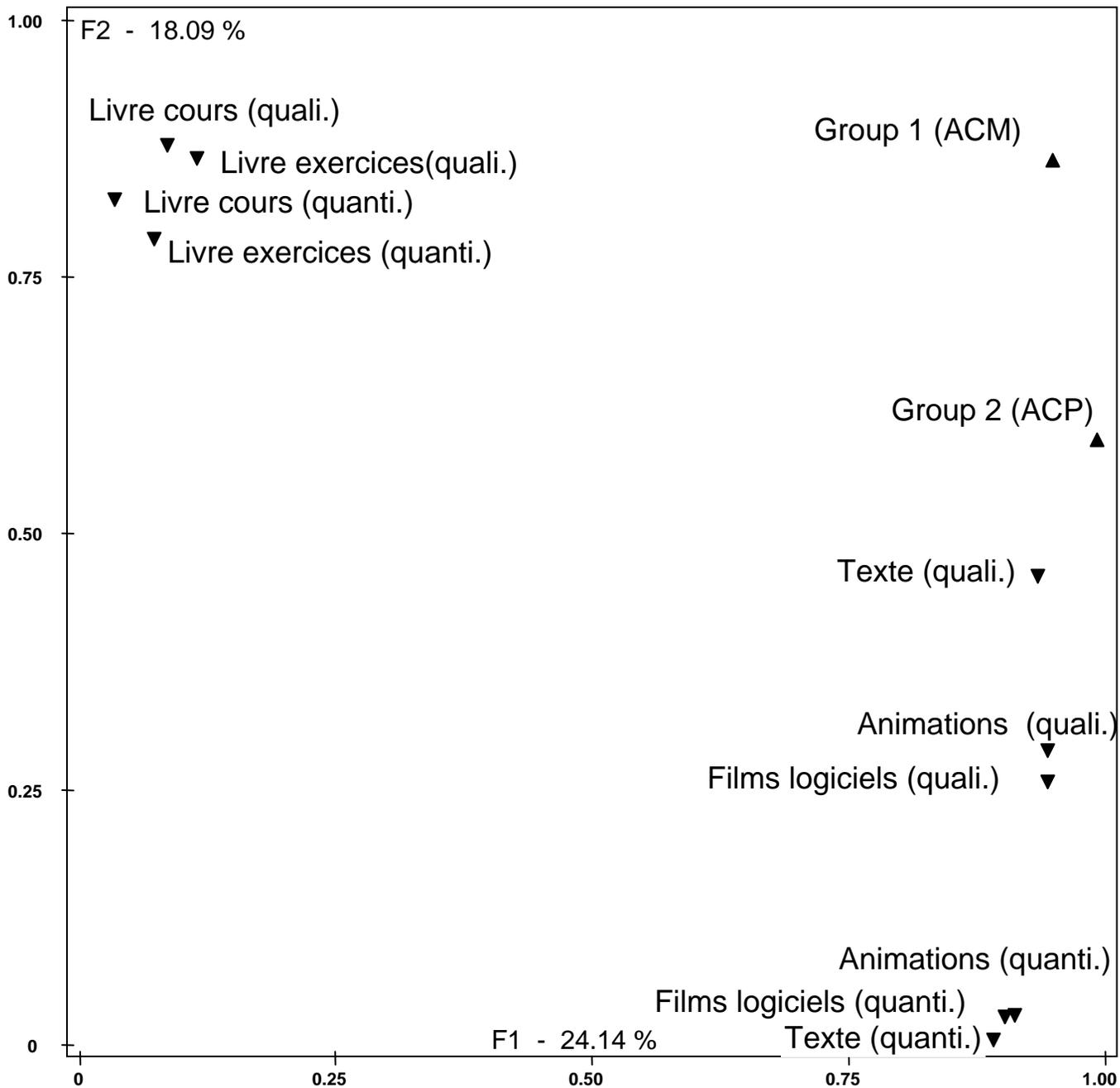
$w_s : W$ associé à v_s
 coordonnée de W_j sur w_s

Carré des liaisons



K_j quantitatif
$$Lg(v_s, K_j) = \frac{1}{\lambda_1^j} \sum_{k \in K_1} r^2(k, v_s)$$

K_j qualitatif
$$Lg(v_s, K_j) = \frac{1}{\lambda_1^j Q_2} \sum_{q \in Q_2} \eta^2(q, v_s)$$



Conclusion

L'AFM est une méthode factorielle dédiée aux tableaux multiples dans lesquels
Un groupe d'individus est décrit par plusieurs groupes de variables

Les groupes de variables peuvent être quantitatifs, qualitatifs ou mixtes

Le cœur de la méthode est une analyse factorielle pondérée ; elle fonctionne
comme une ACP pour les variables quantitatives
comme une ACM pour les variables qualitatives

Elle fournit des résultats

- classiques de l'analyse factorielle

 - représentation des individus, des variables, etc.

- spécifiques de la structure des variables en groupes

 - représentation des groupes, des axes des analyses séparées, etc.

Ces analyses sont disponibles dans

FACTOMINER

Un package R dédié à

l'analyse exploratoire des données

<http://www2.agrocampus-ouest.fr/math/>