

# ANALYSE FACTORIELLE MULTIPLE

## AFM

Principales caractéristiques de la méthode à partir d'un exemple

Jérôme Pagès

Mathématiques appliquées

Agrocampus Rennes  
France

1. Données ; notations ; exemple
2. Pondération des variables
3. L'AFM est basée sur une ACP de l'ensemble des variables actives
4. Représentation superposée des  $J$  nuages d'individus
5. Recherche de facteurs communs
6. Représentation globale des groupes de variables
7. Relations avec les facteurs des analyses séparées
8. Cas des variables qualitatives
9. Conclusion
10. Références (m.à j. mars 2008)

# 1. Données : cadre général

« Essai terroir » ; INRA Angers ; France (C. Asselin ; R. Morlat)

3 appellations de la vallée de la Loire : Saumur ; Chinon ; Bourgueil

## Objectif

Etudier la variabilité sensorielle des vins  
en relation avec le type de sol et l'appellation

20 parcelles

4 types de sol

Même : cépage, viticulture, vinification

21 vins : 1 vin est proposé deux fois

36 dégustateurs = juges (professionnels)

29 questions (= descripteurs = attributs)

## Tasting of red wines from the Loire Valley (INRA Angers)

1. Données

### Le questionnaire

<b>Olfaction at rest</b>		1	2	3	4	5
Olfactory intensity	nil → very high					
Olfactory quality	“					
Fruity	“					
Floral	“					
Spicy	“					

### **Vision**

Color intensity	light → dark					
Shade orange/purple	orange → purple					
Surface (size of tears)	nil → very high					

### **Olfaction after agitation**

Olfactory intensity	nil → very high					
Olfactory quality	“					
Fruity	“					
Floral	“					
Spicy	“					
Vegetal	“					
Phenolic	“					
Mouth aromas intensity	“					
Mouth aromas persistence	“					
Mouth aromas quality	nil → excellent					

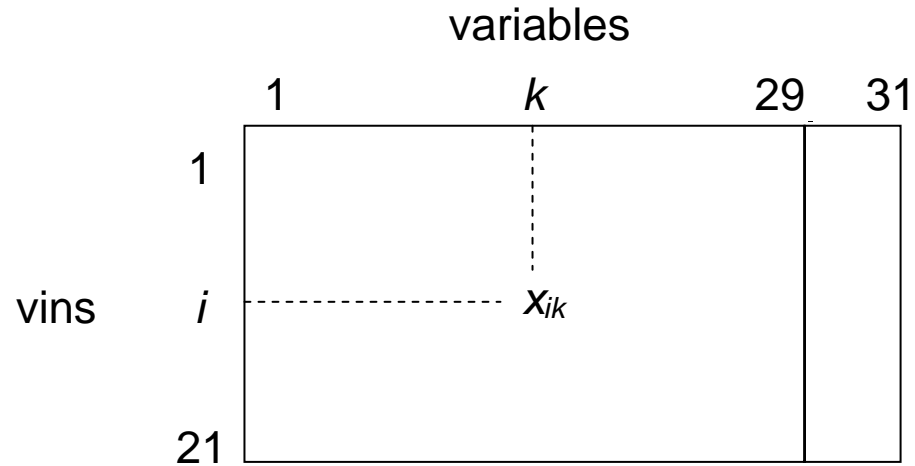
### **Gustation**

First mouth intensity	nil → very high					
Sour	“					
Astringency	“					
Alcohol	“					
Balance sour-astr.-alc.	“					
Velvety	“					
Bitter	“					
End mouth intensity	“					
Harmony	“					

### **Overall judgment**

Overall quality	nil → excellent					
Typicity	nil → very high					

Free comments

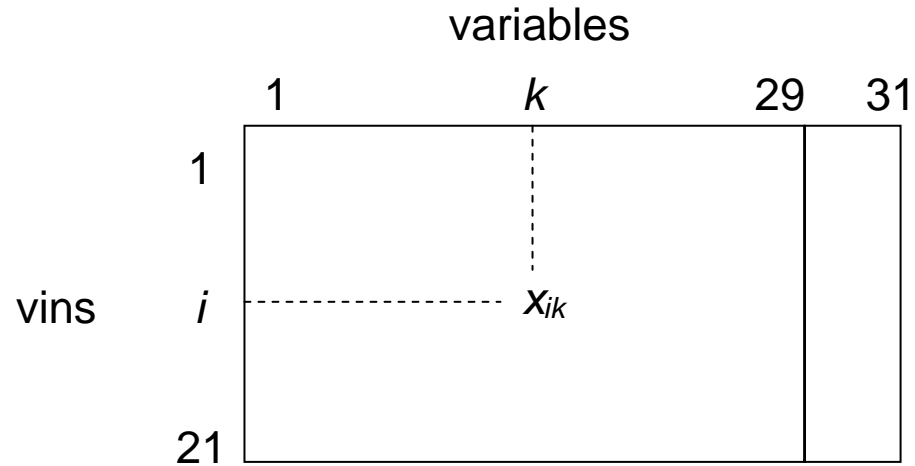


- $1 \leq k \leq 29$   $x_{ik}$  : moyenne des notes données par 36 juges  
au vin  $i$  considérant la variable  $k$
- $k=30,31$   $x_{ik}$  : modalité du vin  $i$  pour la variable  $k$ .

### Origine des vins

Variable 30 (appellation) : Saumur, Chinon, Bourgueil

Variable 31 (terroir) : séquence de référence, sol2, sol3, sol 4



### Objectif

Mettre en évidence les principales dimensions de la variabilité sensorielle des vins

Relier ces dimensions avec le type de terroir et l'appellation

### Méthodologie usuelle

Appliquer l'ACP à l'ensemble du tableau

les variables sensorielles (centrées-réduites) étant actives

les variables "origine" étant supplémentaires

groupes	1	2	3	4	5	6
variables	1					31
vins	1					
	21					

### 6 groupes de variables

- 1 Olfaction au repos (5 variables)
- 2 Vision (3 variables)
- 3 Olfaction après agitation (10 variables)
- 4 Gustation (9 variables)
- 5 Jugement d'ensemble (2 variables)
- 6 Origine (2 variables ; 7 modalités)

Premier problème : équilibrer l'influence des groupes au sein d'une analyse d'ensemble

## Notations générales

groupes	1	$j$	$J$		
variables	1	$K_1$	$K_j$	$1$	$K_J$
indiv.	1	$i$	$I$		

		$x_{ik}$	
--	--	----------	--

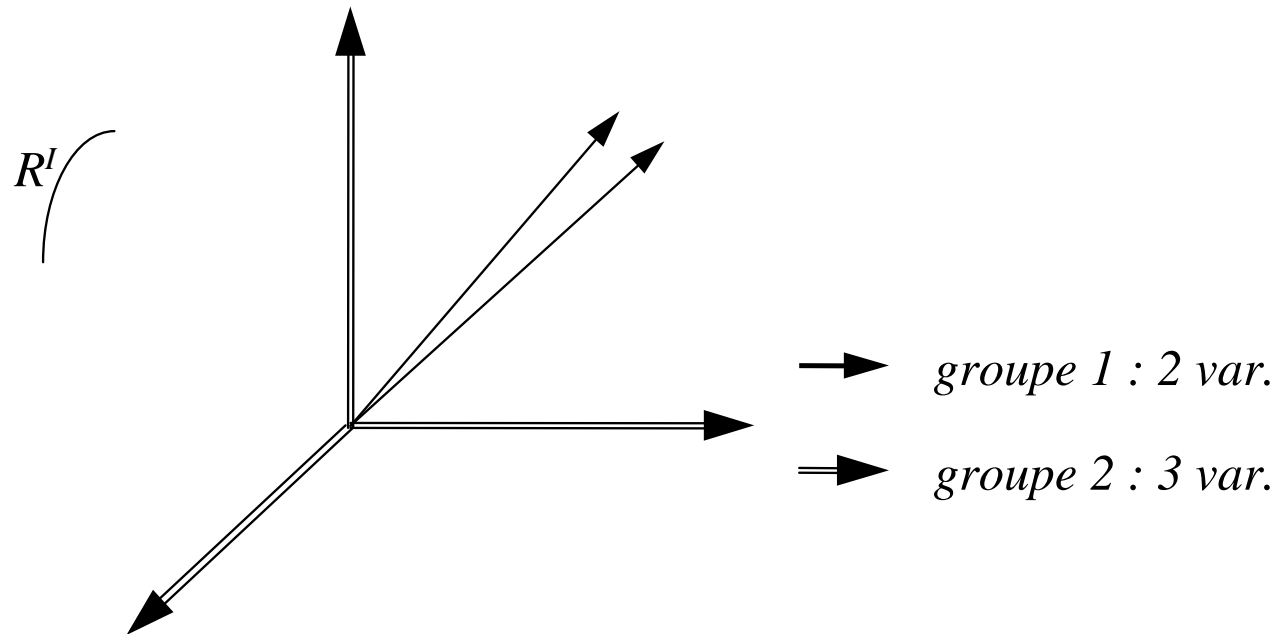
$x_{ik}$  : valeur de la variable  $k$  pour l'individu  $i$   
 variable quantitative : nombre  
 variable qualitative: n° ou libellé de modalité

Même notation  $I, J, K$  :  
 nombre d'éléments  
 ensemble des éléments



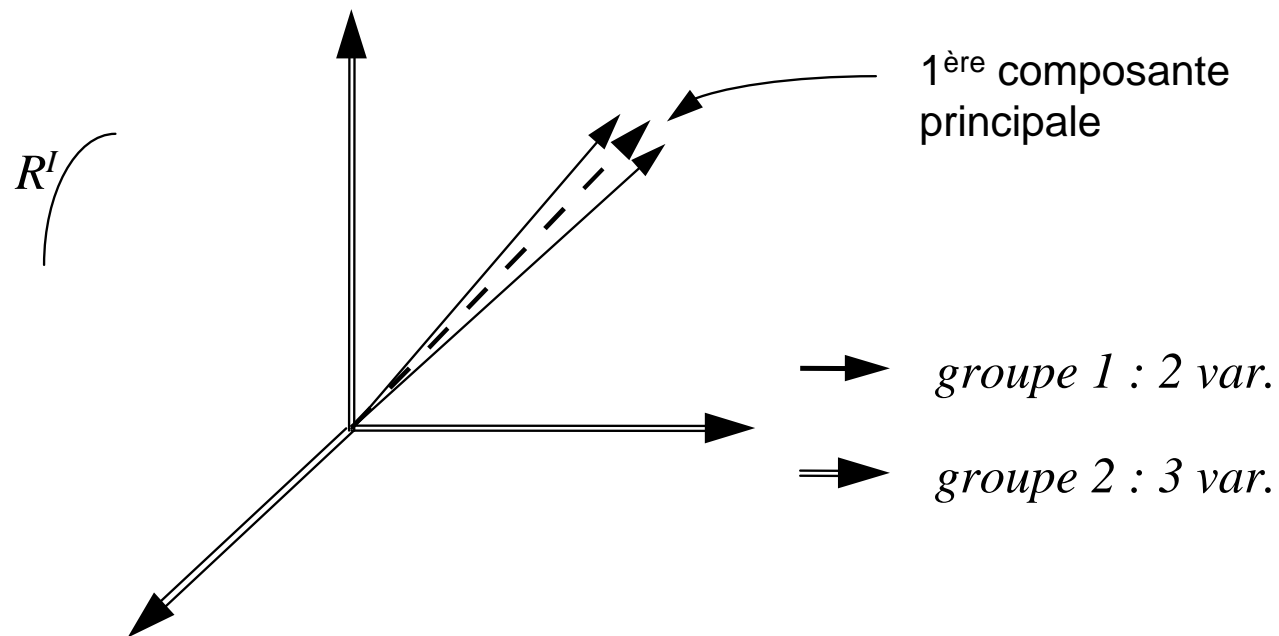
## 2. Pondération des variables

### Exemple de référence



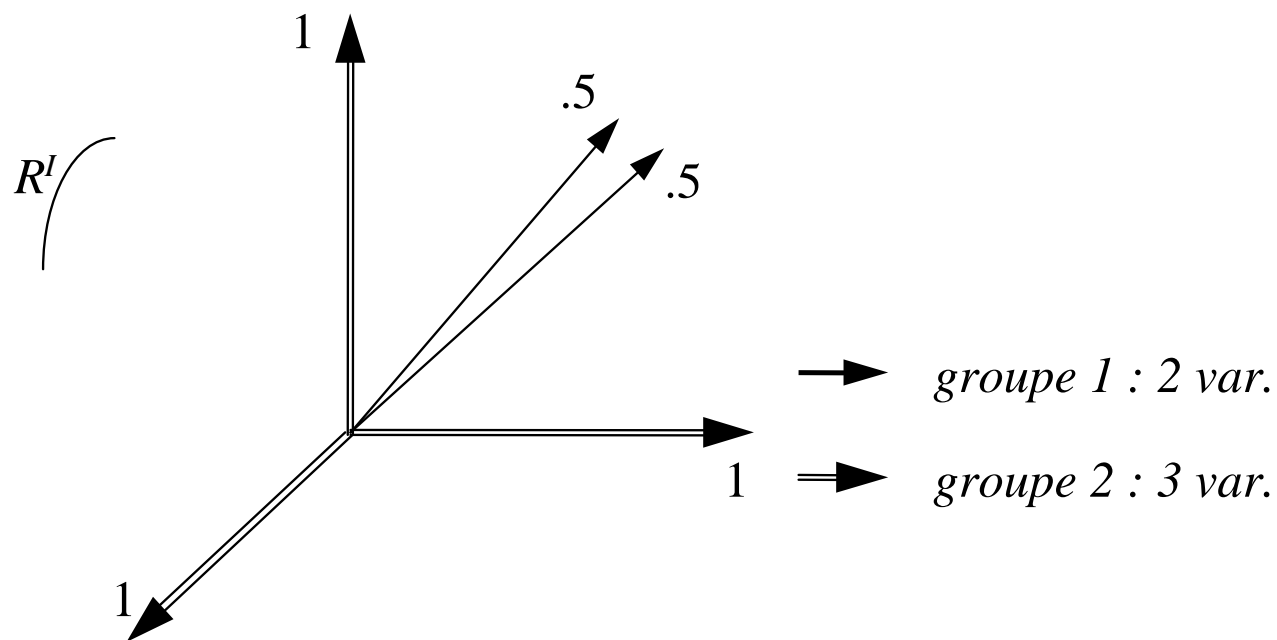
## Exemple de référence

ACP des 5 variables, sans considérer les groupes



## Exemple de référence

## Equilibre des groupes de variables en AFM



Chaque variable du groupe  $j$  est affectée du poids  $1/\lambda_1^j$   
 $\lambda_1^j$  : 1<sup>ère</sup> valeur propre de l'ACP séparée du groupe  $j$ .

### 3. L'AFM est basée sur une ACP de l'ensemble des variables actives

Variables quantitatives : l'AFM est basée sur une ACP pondérée  
variables centrées-réduites  
variables centrées  
mélange (groupes homogènes)

Equivalence

Quand chaque groupe est réduit à 1 variable quantitative : AFM = ACP normée

L'AFM est basée sur une ACP de l'ensemble des variables actives

L'AFM fournit :

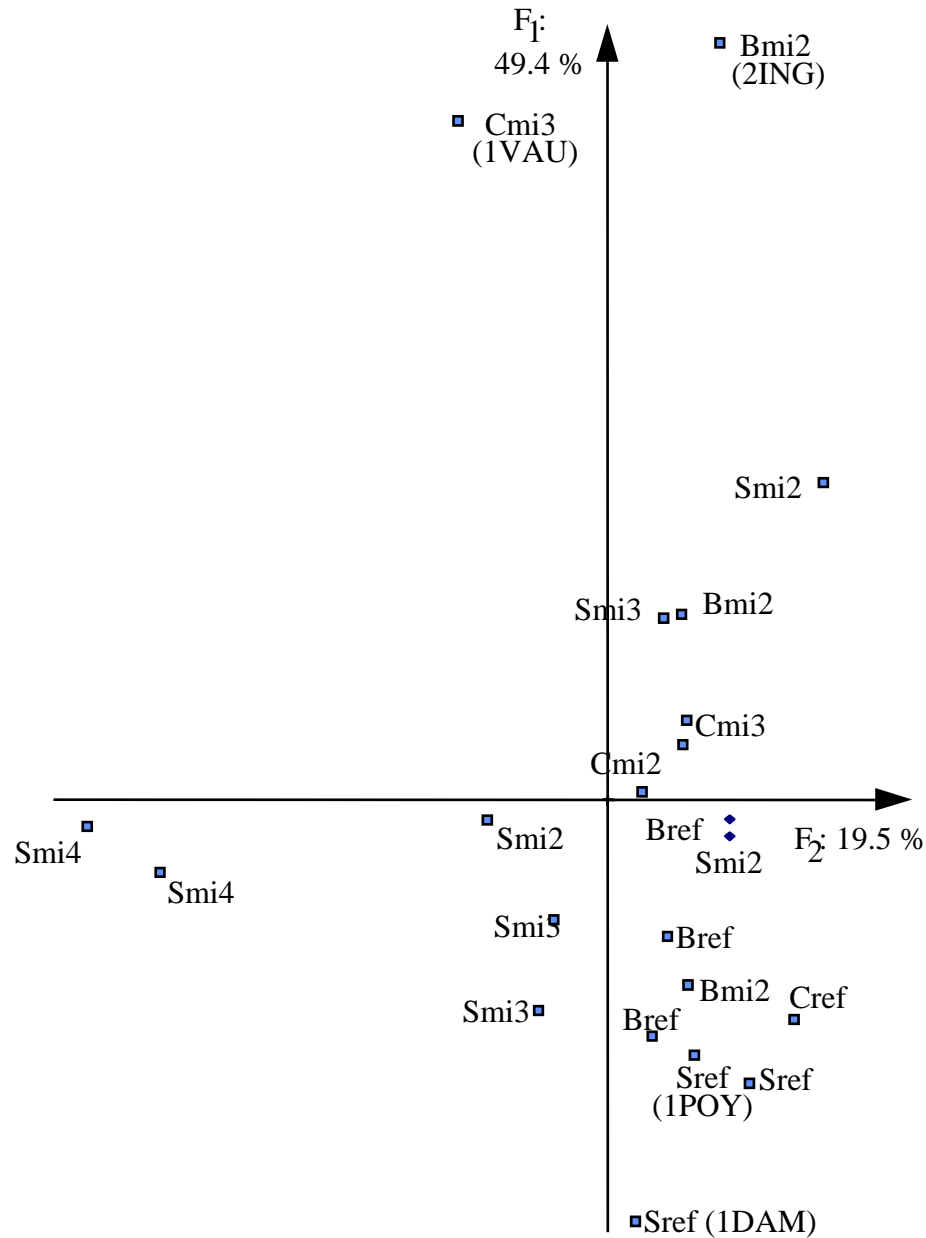
D'abord les résultats classiques de l'ACP

Pour chaque axe :

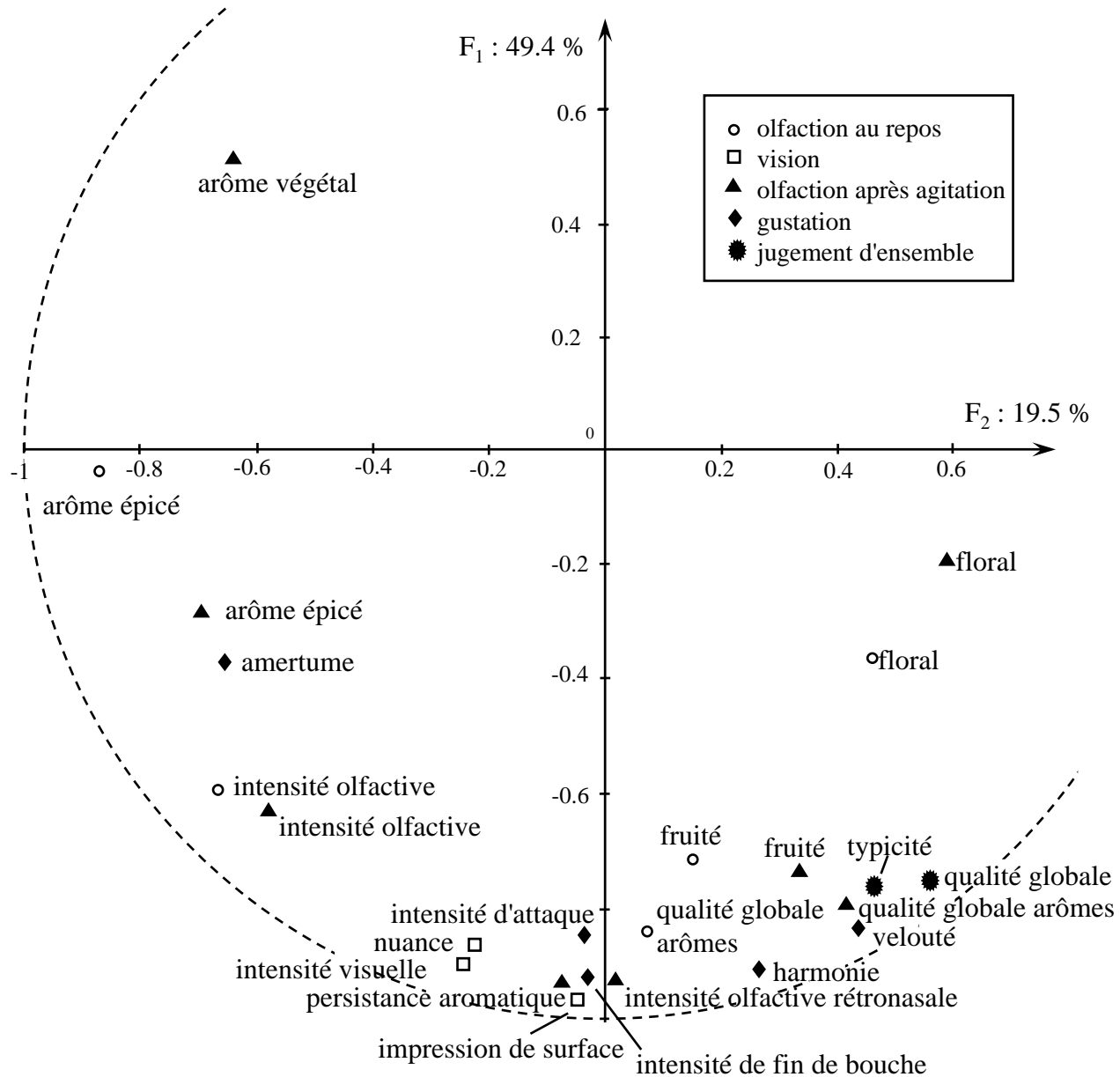
Coordonnées, contributions et cosinus carrés des individus

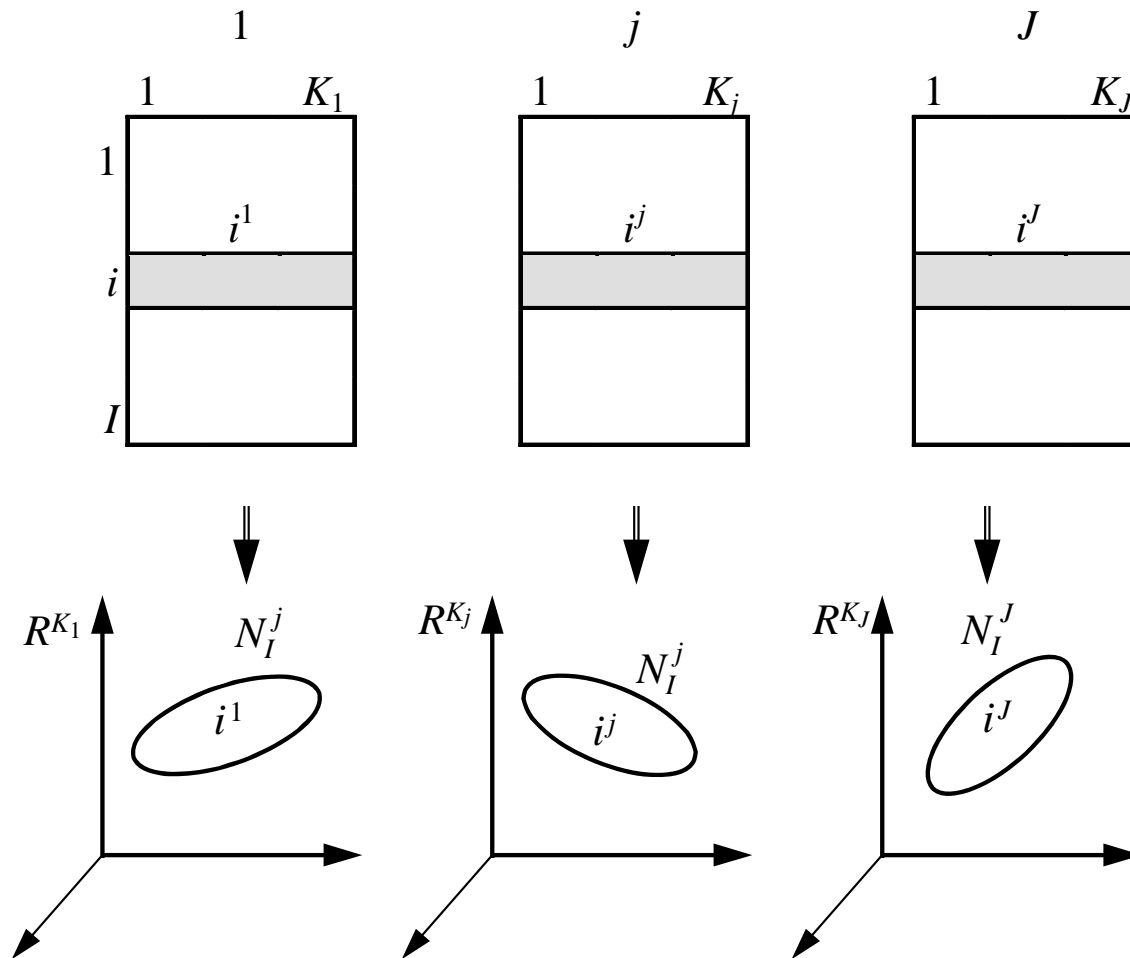
Coefficients de corrélation entre les facteurs et les variables quantitatives

## AFM : représentation des vins



## AFM : représentation des variables quantitatives

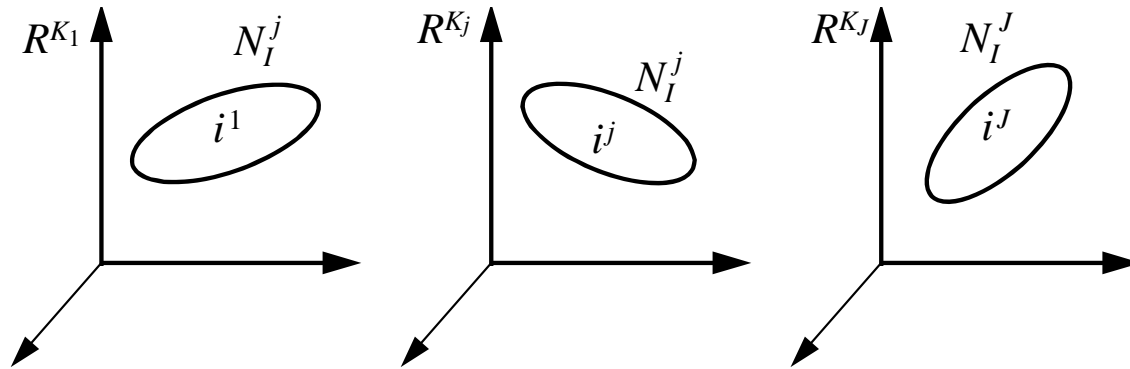


4. Représentation superposée des  $J$  nuages d'individus partiels en AFM

$N_j$  : nuage partiel (des individus vus par le groupe  $j$ )



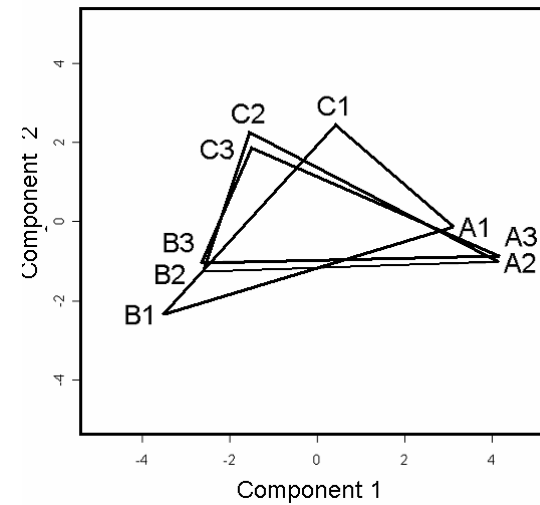
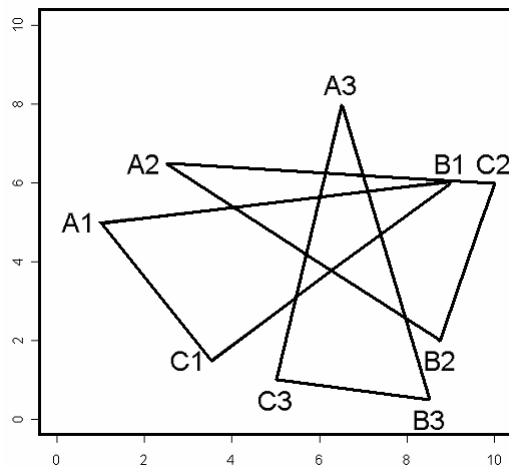
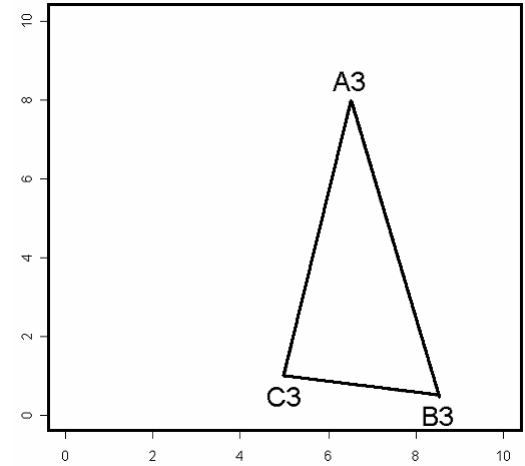
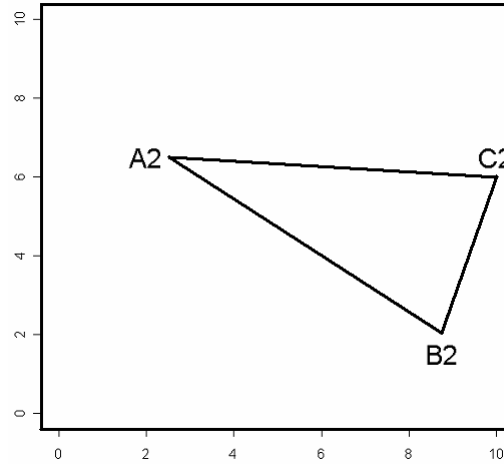
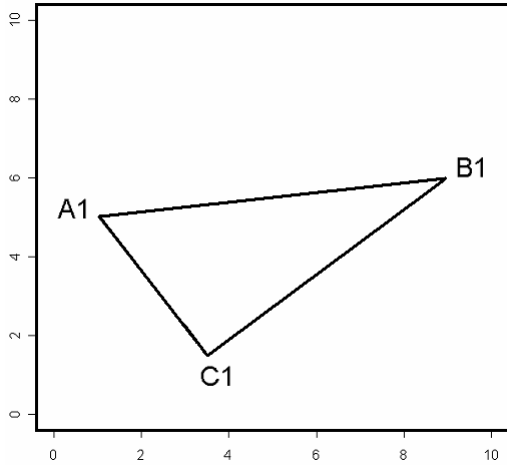
# Représentation superposée des $J$ nuages d'individus partiels en AFM



Comment comparer des nuages évoluant dans des espaces différents ?

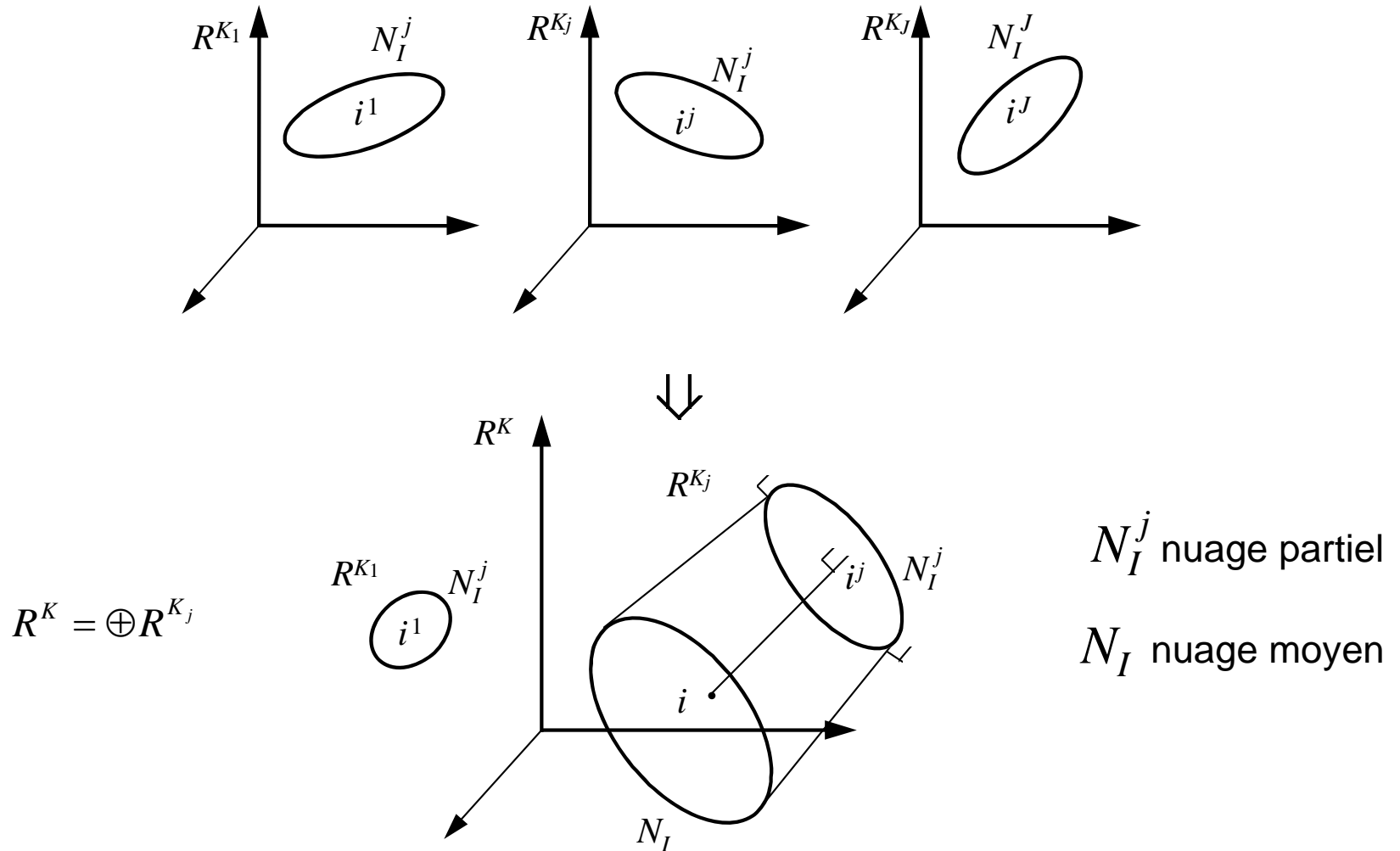
Méthode de référence : analyse procustéenne (Green, 1952; Gower, 1975)

## Analyse procustéenne : principe



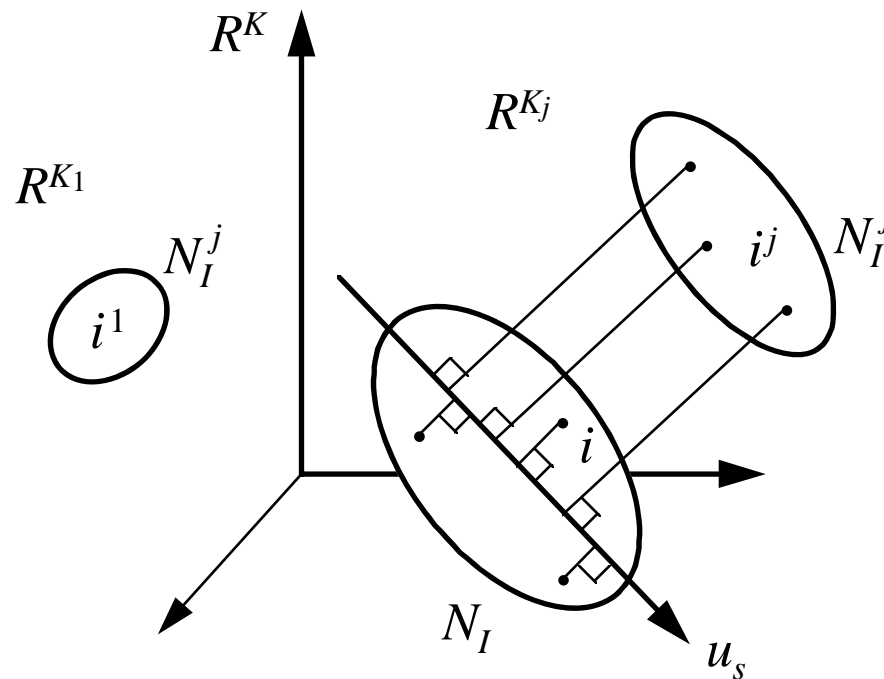
# Représentation superposée des $J$ nuages d'individus partiels en AFM

## Cadre géométrique



Représentation superposée des  $J$  nuages d'individus partiels en AFM

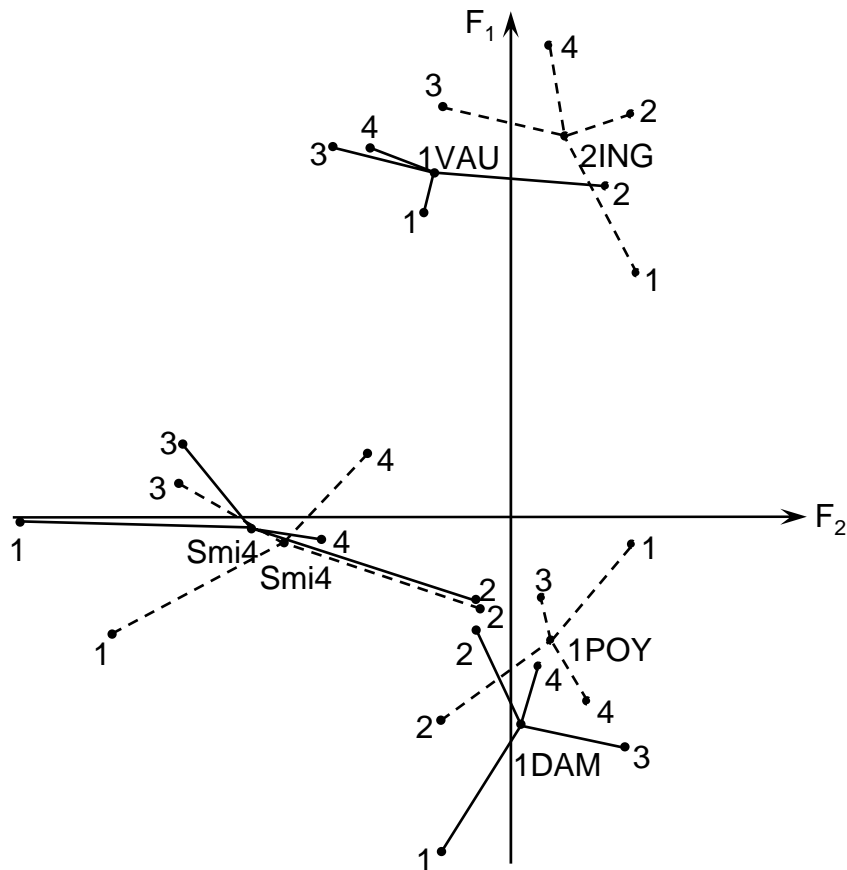
Principe



Les nuages partiels sont projetés sur les axes principaux d'inertie du nuage moyen

# Représentation superposée des $J$ nuages d'individus partiels en AFM

Exemple. Extraits (6 vins)



Quelques valeurs des trois variables du groupe 1 les plus corrélées au premier facteur

	qualité globale des arômes	note fruitée	intensité olfactive
maximum	3.429	3.154	3.708
1DAM	3.429	3.154	3.607
1POY	3.107	2.731	3.071
moyenne	3.046	2.714	3.111

Quelques valeurs des 4 variables du groupe 4 les plus corrélées au premier facteur

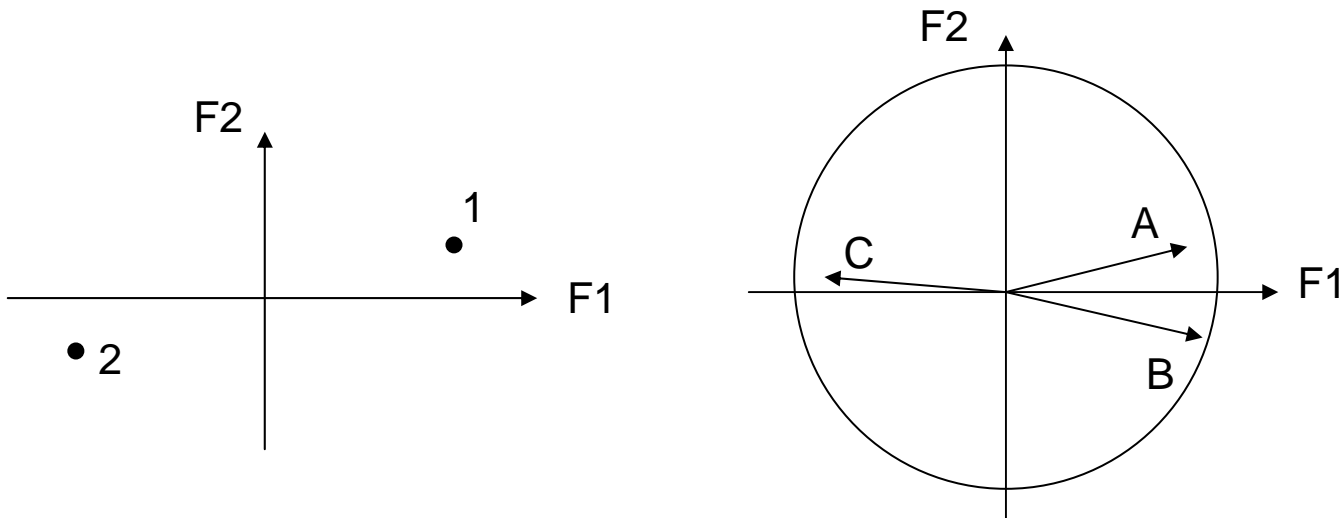
	velouté	intensité fin de bouche	harmonie	intensité d'attaque
maximum	3.286	3.676	3.786	3.519
1POY	3.231	3.667	3.786	3.519
1DAM	3.036	3.643	3.643	3.464
moyenne	2.674	3.166	3.148	3.156

## Une propriété importante de la représentation superposée

Relation de transition usuelle en ACP

$$F_s(i) = \frac{1}{\sqrt{\lambda_s}} \sum_{k \in K} x_{ik} G_s(k)$$

$F_s(i)$       coordonnée de  $i$  le long de l'axe  $s$   
 $G_s(k)$       coordonnée de la variable  $k$  le long de l'axe  $s$   
 $\lambda_s$           valeur propre associée à l'axe  $s$   
 $x_{ik}$           donnée (valeur de  $k$  pour  $i$ )



## Une propriété importante de la représentation superposée

Relation de transition usuelle en ACP

$$F_s(i) = \frac{1}{\sqrt{\lambda_s}} \sum_{k \in K} x_{ik} G_s(k)$$

Si la variable  $k$  a le poids  $m_k$

$$F_s(i) = \frac{1}{\sqrt{\lambda_s}} \sum_{k \in K} x_{ik} m_k G_s(k)$$

Relation de transition usuelle appliquée au nuage moyen en AFM

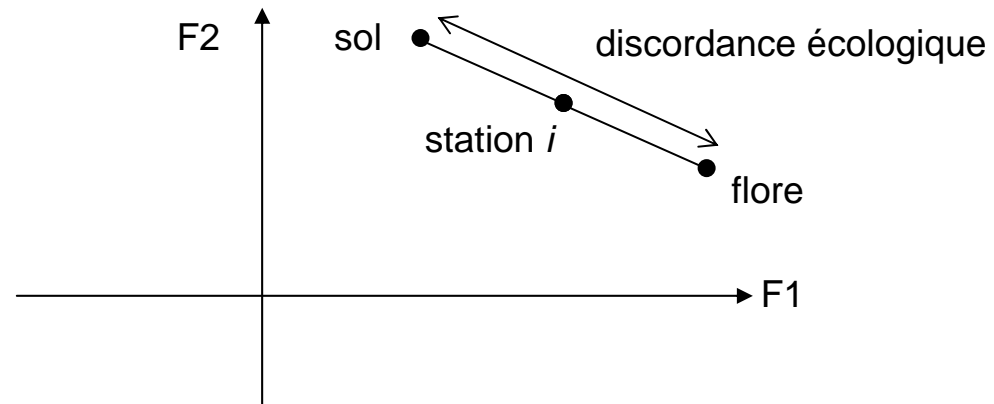
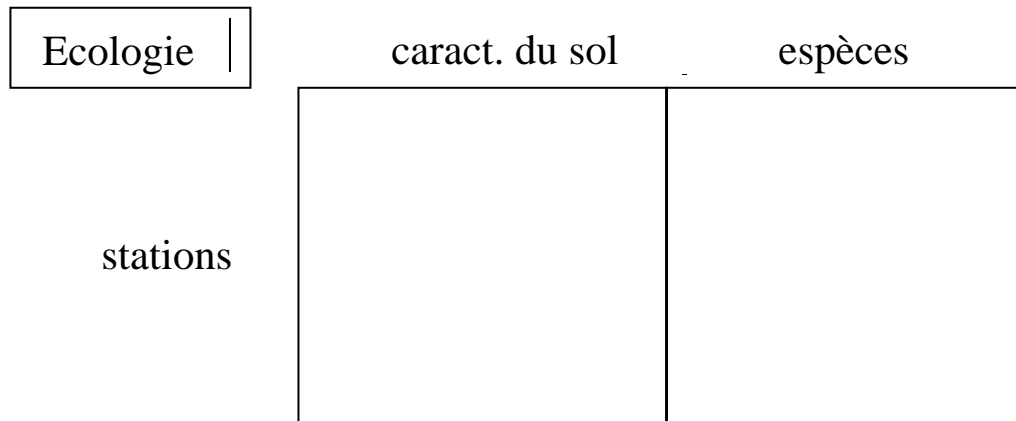
$$F_s(i) = \frac{1}{\sqrt{\lambda_s}} \sum_{j \in J} \frac{1}{\lambda_1^j} \sum_{k \in K_j} x_{ik} G_s(k)$$

Relation de transition partielle

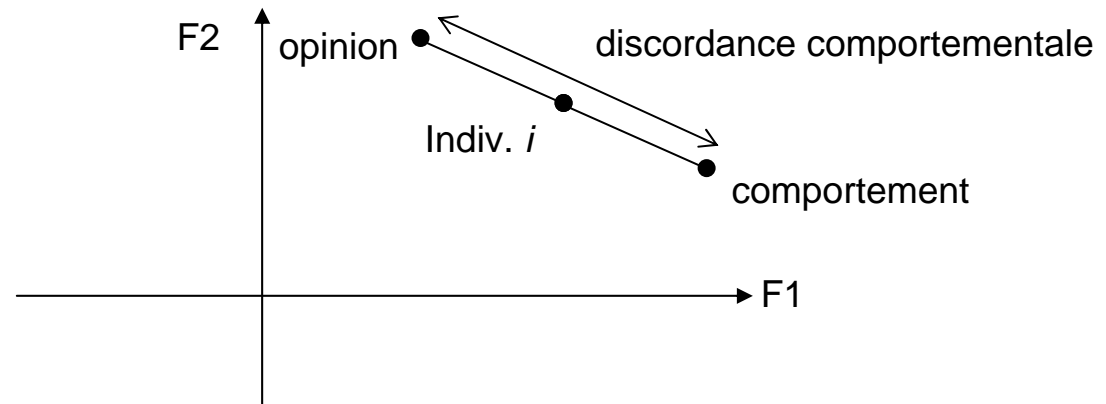
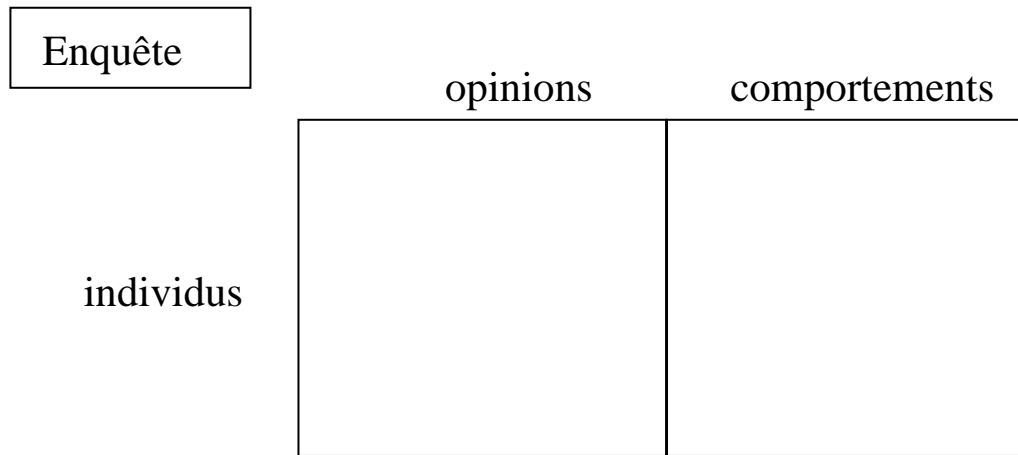
$$F_s(i^j) = \frac{1}{\sqrt{\lambda_s}} \frac{J}{\lambda_1^j} \sum_{k \in K_j} x_{ik} G_s(k) \qquad F_s(i) = \frac{1}{J} \sum_{j \in J} F_s(i^j)$$



## Autres exemples de représentation superposée

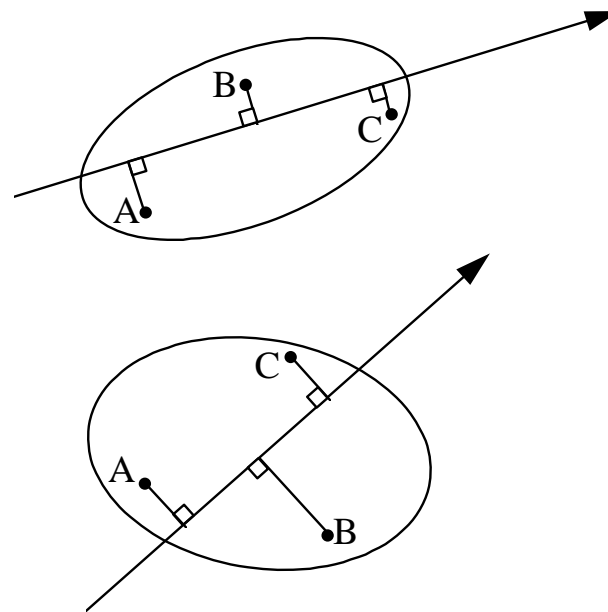


## Autres exemples de représentation superposée



## 5. Recherche de facteurs communs

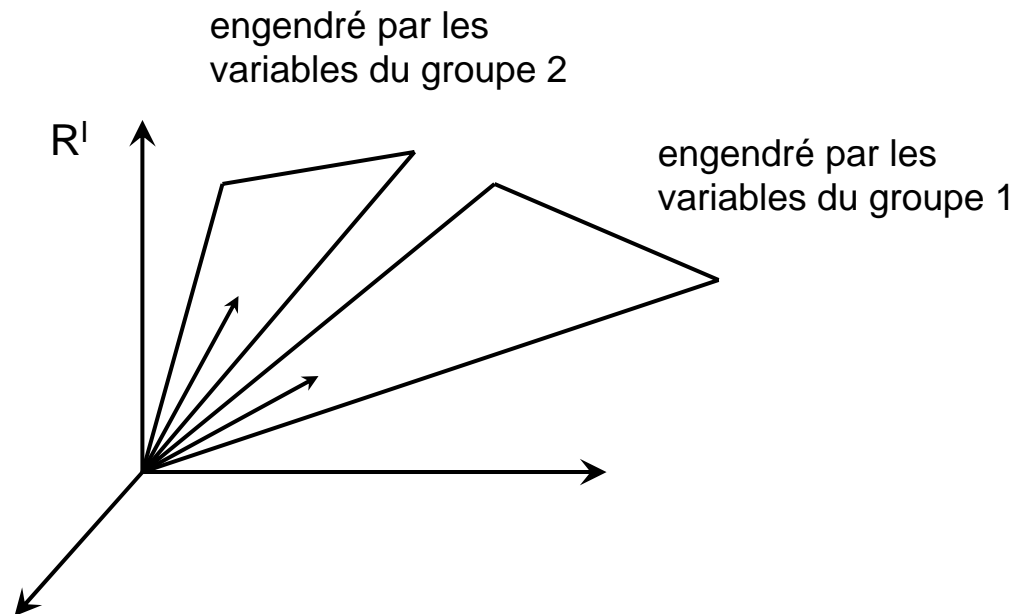
Un facteur commun aux deux nuages



Méthode de référence : analyse (multi)canonique

# Analyse canonique (Hotelling, 1936)

recherche simultanée de  
combinaisons linéaires des variables des groupes 1 et 2



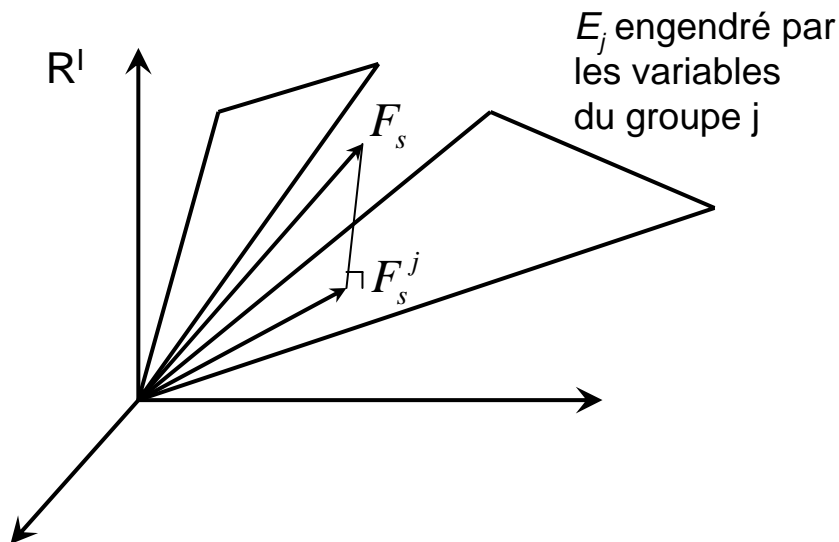
## Analyse multicanonique (Carroll, 1968)

A chaque pas  $s$  : recherche de

1) une variable générale (liée à tous les groupes de variables)  $F_s$

2) une variable canonique pour chaque groupe  $j$   $F_s^j$

(combinaison linéaire des variables du groupe  $j$  liée à la variable générale)



$R^2(z, K_j)$  : coefficient de détermination

$$F_s \text{ maximise } \sum_j R^2(z, K_j)$$

$F_s^j$  : projection de  $F_s$  sur  $E_j$

## L'AFM vue comme une analyse multicanonique

Une mesure de relation entre

une variable  $z$

un groupe de variables  $K_j = \{v_k; k = 1, K_j\}$

$Lg(z, K_j)$  = inertie projetée de l'ensemble des variables  $v_k$  sur  $z$

Cas de variables centrées réduite, pondérées en AFM

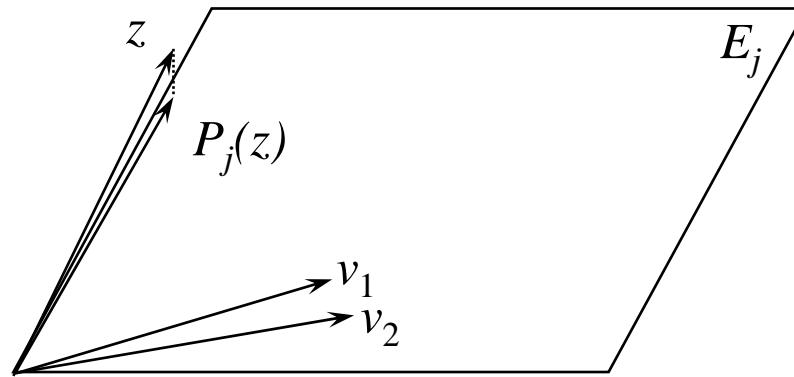
$$Lg(z, K_j) = \frac{1}{\lambda_1^j} \sum_{k \in K_j} r^2(z, v_k)$$

Du fait de la pondération de l'AFM

$$0 \leq Lg(z, K_j) \leq 1$$

# Coefficient de détermination et mesure $L_g$

Un cas particulier



## L'AFM vue comme une analyse multicanonique

La variable générale  $F_s$  doit être liée à tous les groupes de variables.

$F_s$  maximise

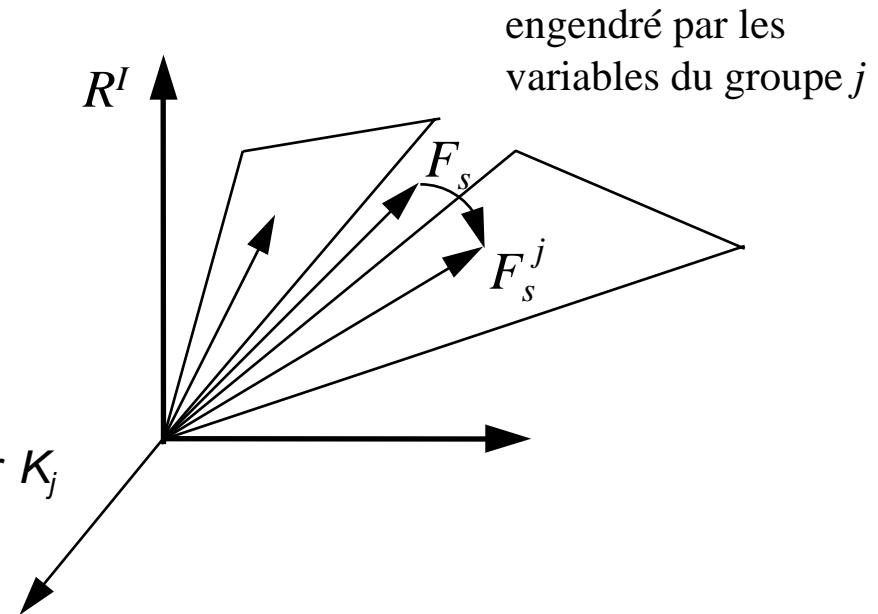
$$\sum_j \text{Lg}(z, K_j)$$

La variable canonique  $F_s^j$  doit être liée

à la variable générale  $F_s$   
aux variables du groupe  $j$

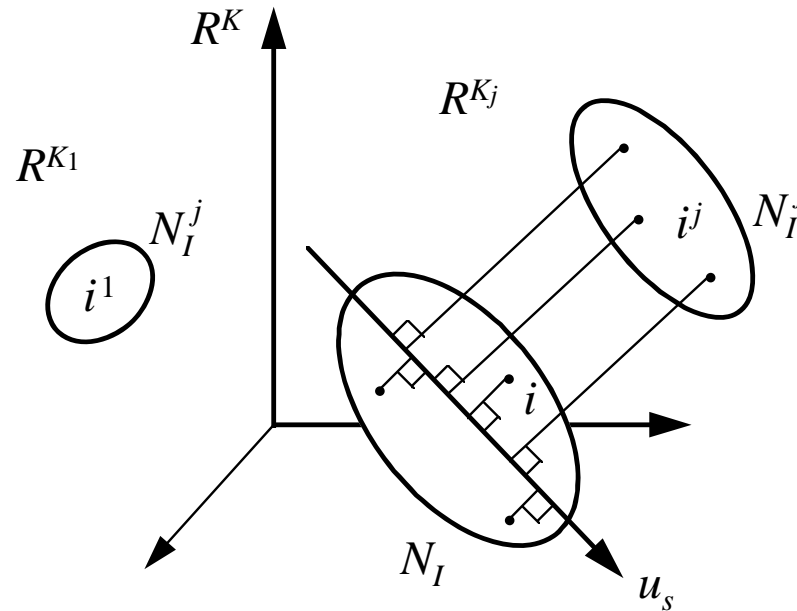
$$F_s^j = \frac{1}{\lambda_s} X_j X_j' D F_s = \frac{1}{\lambda_s} W_j D F_s$$

Lien avec la régression PLS de  $F_s$  sur  $K_j$

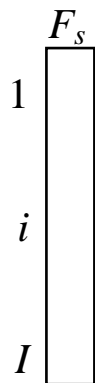




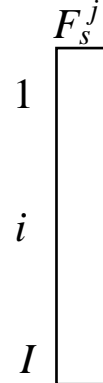
Lien avec la représentation superposée des  $N_I^j$



Coordonnées de  
 $N_I$  sur  $u_s$



Coordonnées de  
sur  $u_s$



Comment mettre en évidence un facteur commun ?

$r(F_s, F_s^j)$  = corrélation entre

le facteur “commun” :  $F_s$

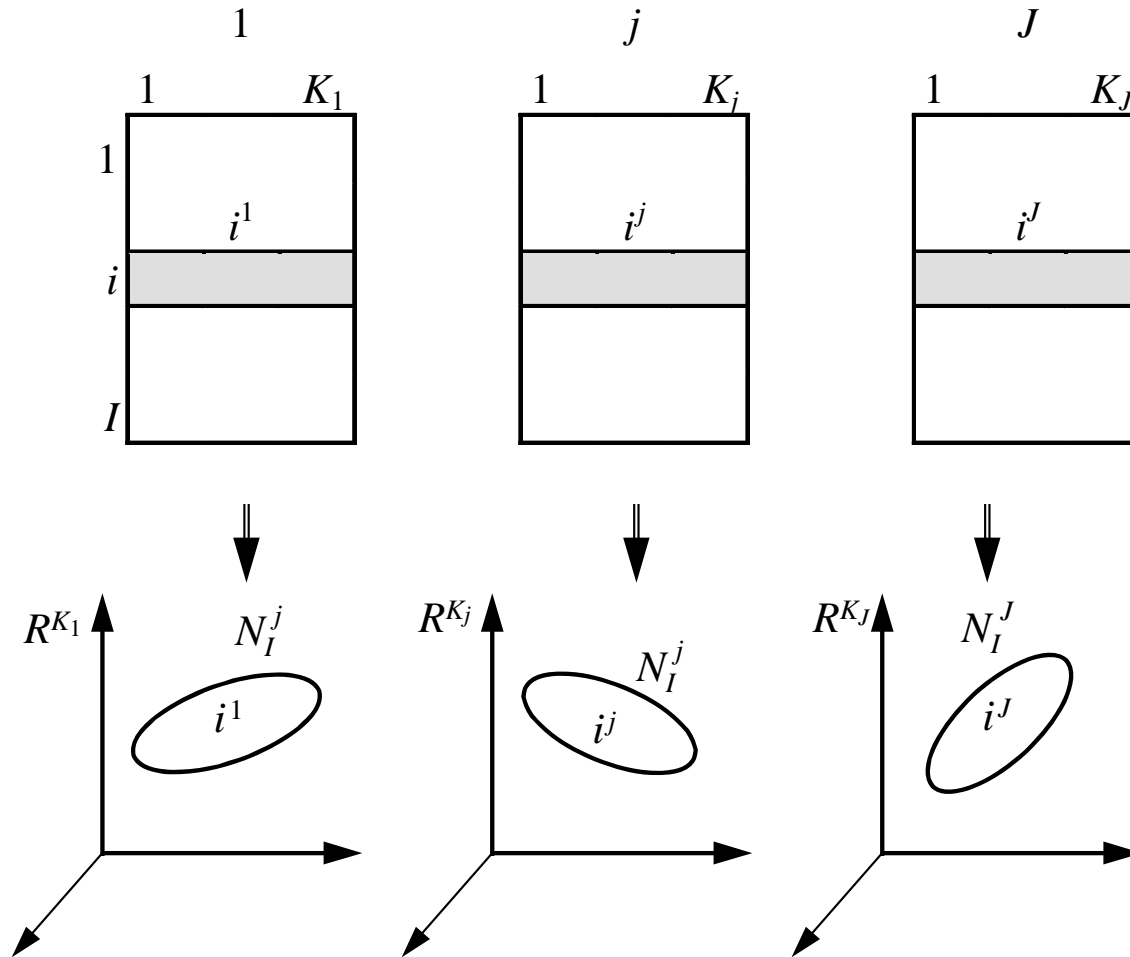
sa représentation dans le groupe  $j$  :  $F_s^j$

## Comment mettre en évidence un facteur commun ?

Pour chaque  $s$  et chaque  $j$ , corrélation entre  
 le facteur commun  $F_s$  (projection du nuage moyen  $N_l$ )  
 la variable canonique  $F_s^j$  (projection du nuage partiel  $N_l^j$ )

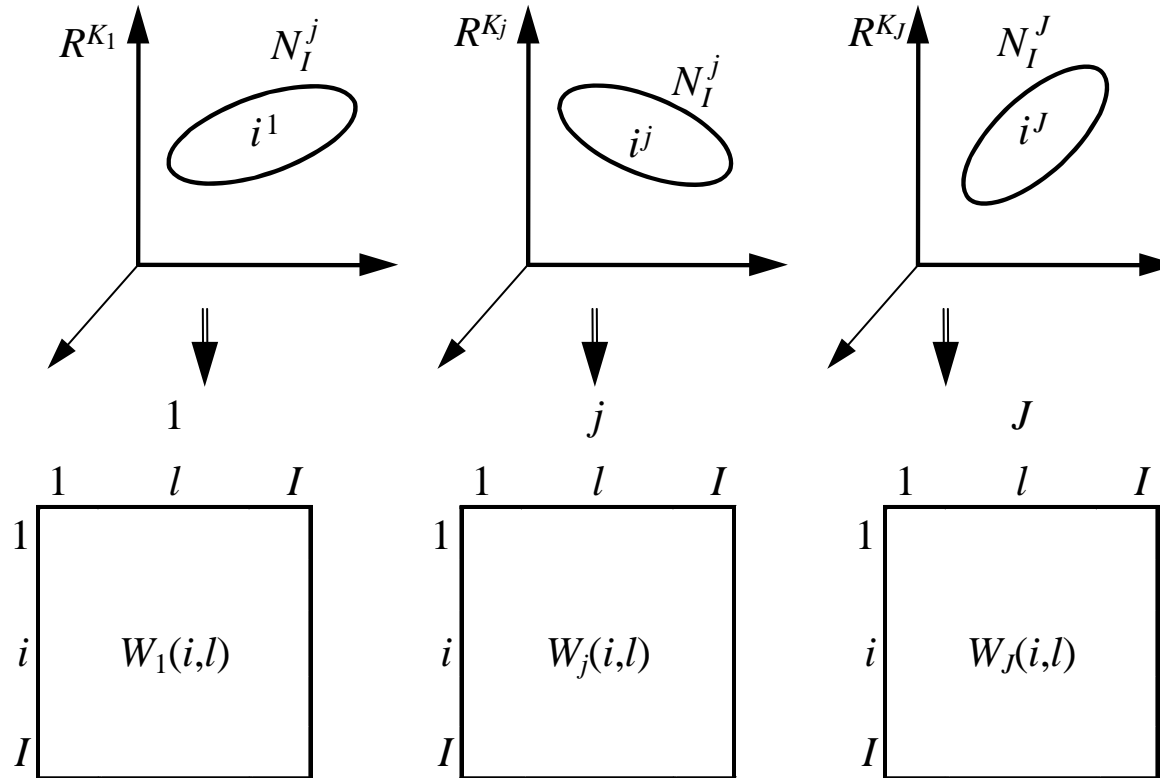
	axe 1	axe 2	axe 3	axe 4	axe 5	axe 6	axe 7
olfaction au repos	.89	.96	.89	.48	.42	.27	.42
vision	.93	.22	.16	.22	.17	.08	.21
olfaction après agitation	.97	.89	.90	.57	.66	.49	.46
gustation	.95	.87	.30	.25	.52	.56	.42

## 6. Représentation globale des groupes de variables



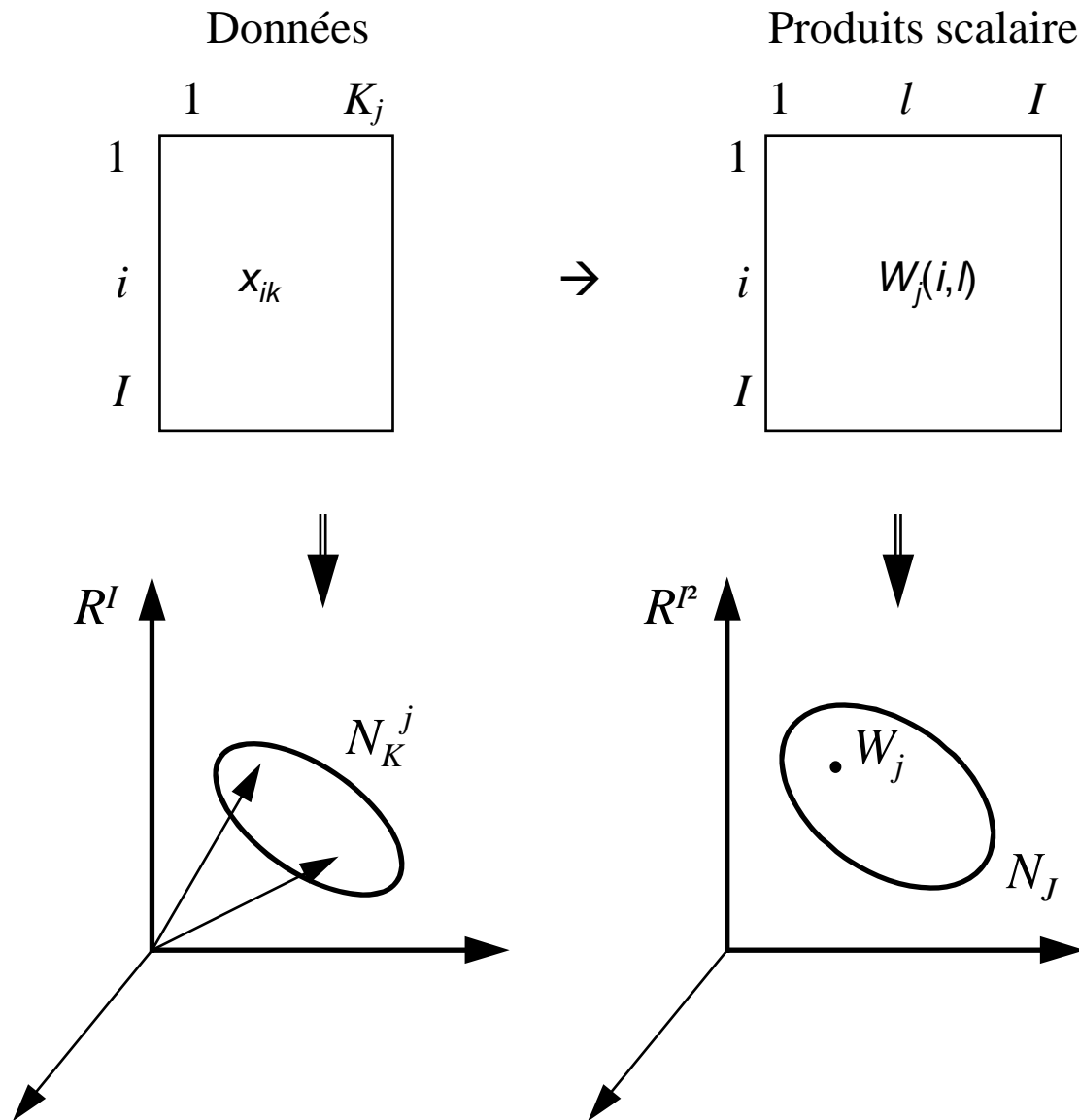
$N_I^j$  : nuage partiel (des individus ; associé au groupe  $j$ )

Etudier les ressemblances globales entre les  $N_i^j$

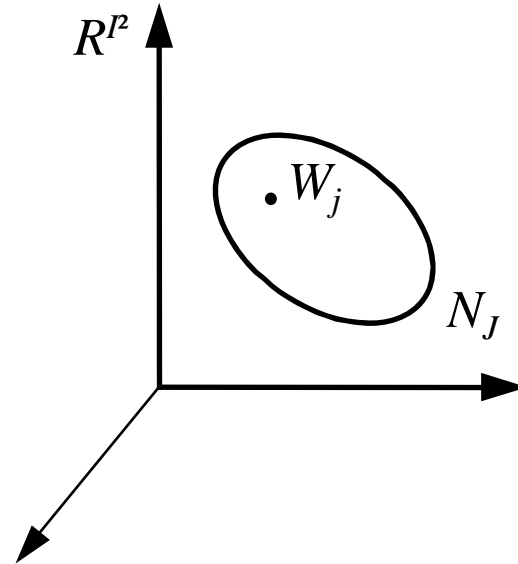


matrices des produits scalaires entre individus pour chaque groupe  $j$

$$W_j = X_j X_j'$$



$N_J$  : nuage des groupes



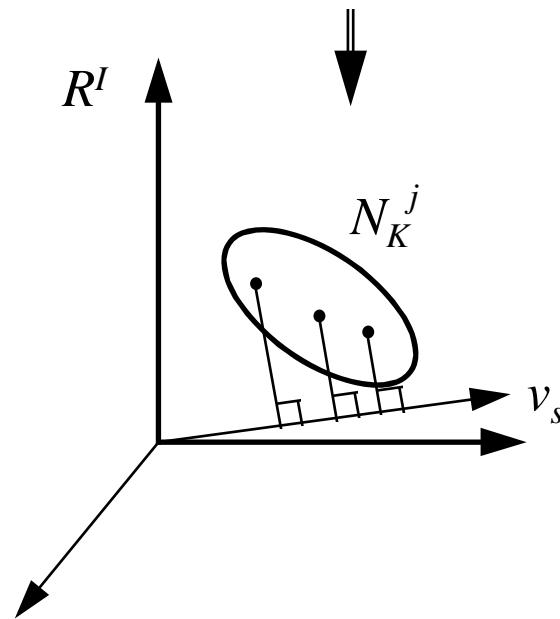
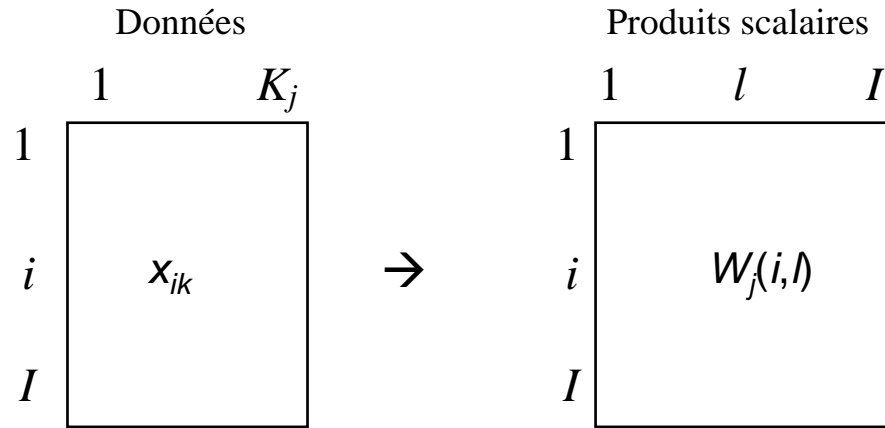
Etudier le nuage  $N_J$

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

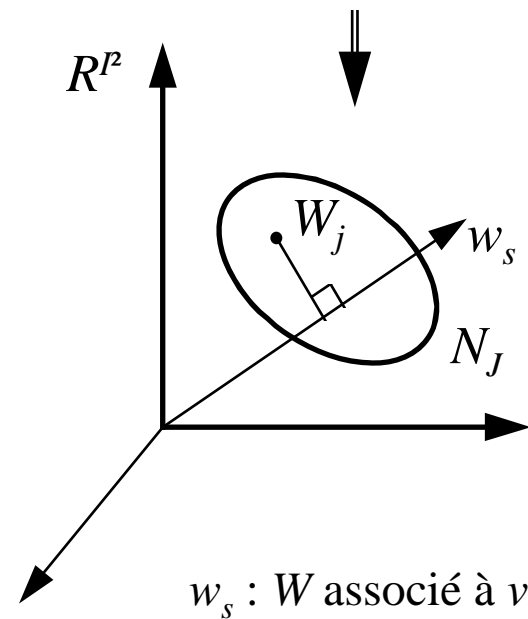
Axes principaux de  $N_J$

Interprétation de ces axes difficile

# Etudier $N_j$ avec l'AFM

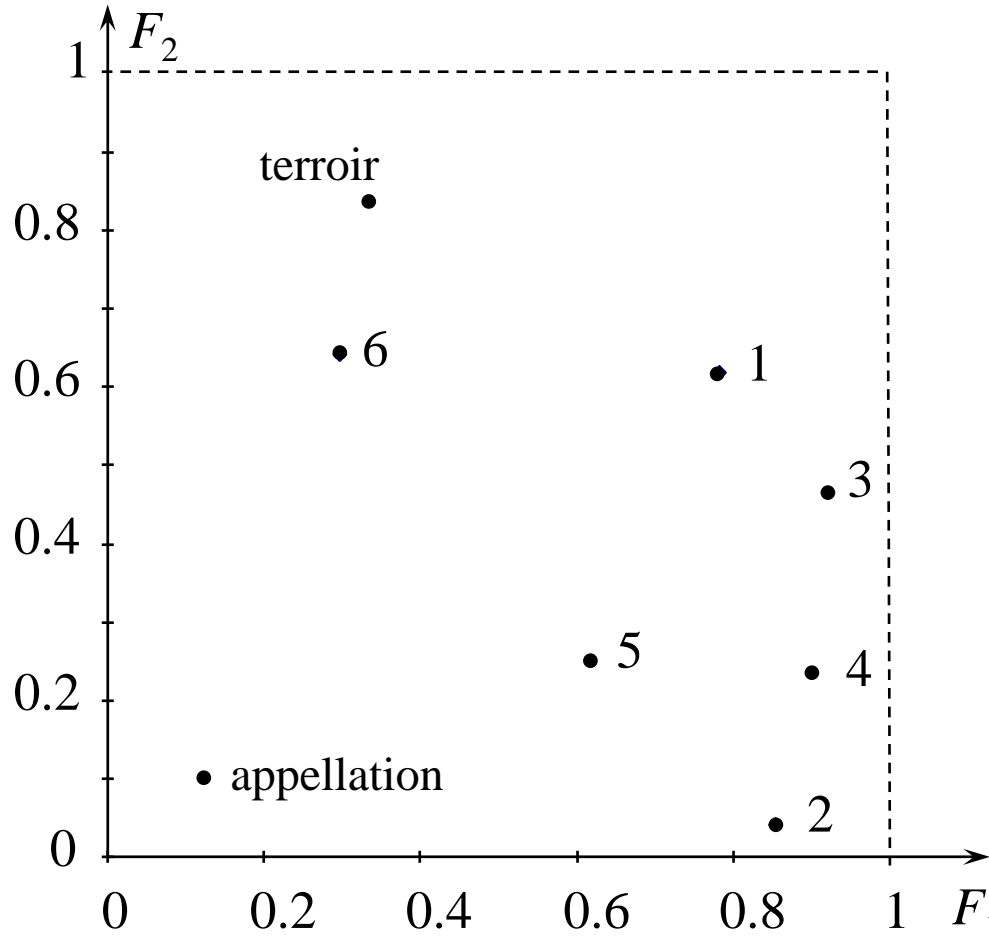


inertie de  $N_K^j$  projeté sur  $v_s$   
 $= \text{Lg}(v_s, K_j)$



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

## Représentation globale des groupes de variables



Carré des liaisons  
Coordonnée = Lg

Qualité de représentation des groupes  
(inertie projetée / inertie totale)

1 olfaction au repos : 0.617

2 vision : 0.729

3 olfaction après agitation : 0.784

4 gustation : 0.772

5 appréciation d'ensemble : 0.444

6 origine : 0.189

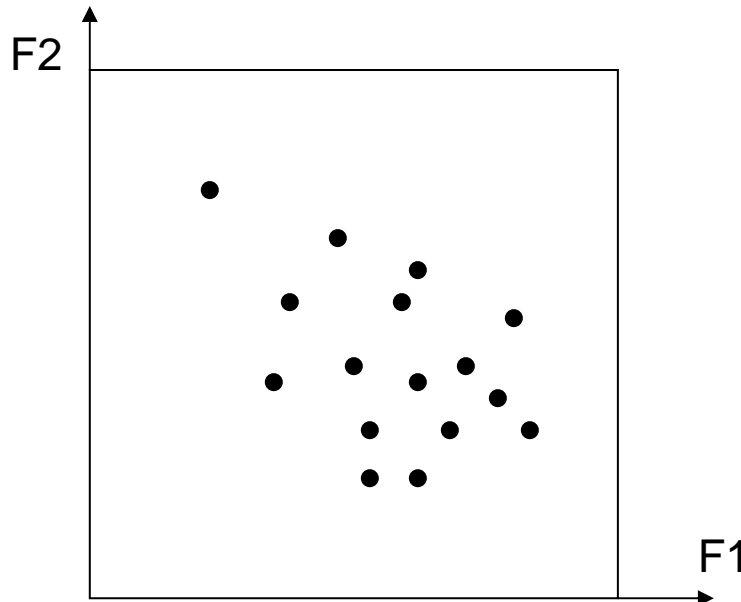


# Autre exemple de représentation des groupes

50 vins mousseux (incluant 26 Champagne) décrits par  
 25 amateurs  
 7 œnologues

Les données individuelles sont juxtaposées en ligne : un juge = un groupe

	25 amateurs			7 œnologues		
50 vins	Juge 1	...	Juge 25	Juge 26	...	Juge 32

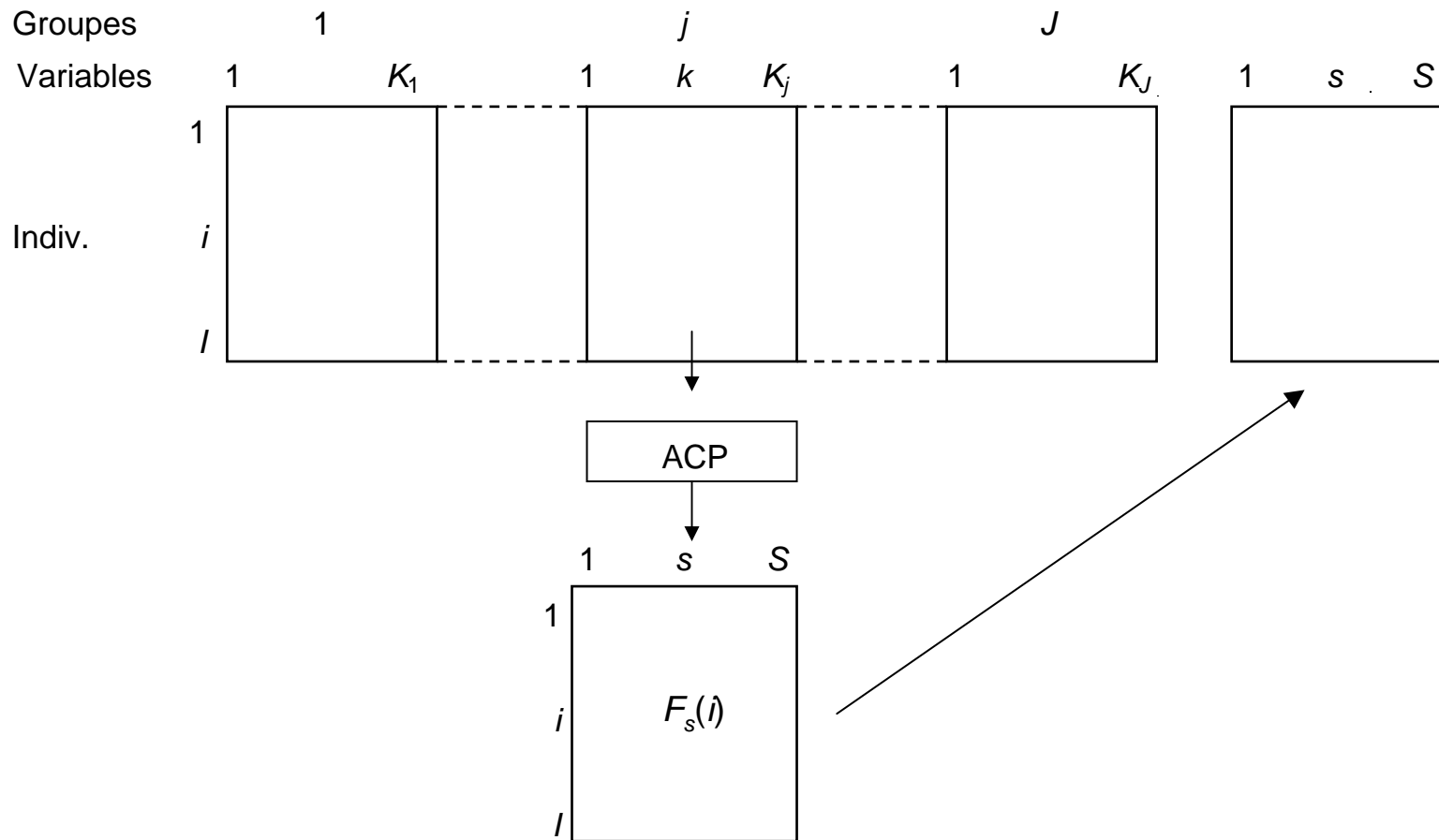


1 juge = 1 point

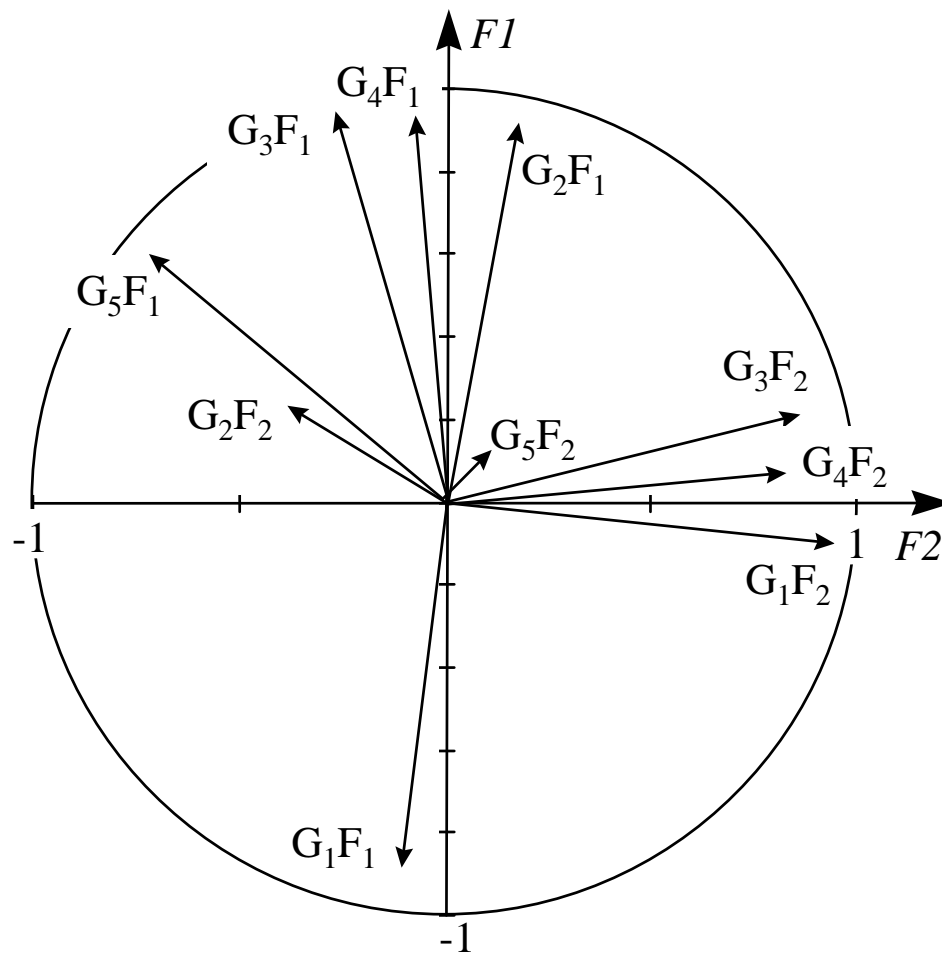
Les amateurs et les œnologues  
 sont-ils mélangés ?

## 7. Relations avec les facteurs des analyses séparées

Les facteurs des analyses séparées sont projetés comme des éléments supplémentaires



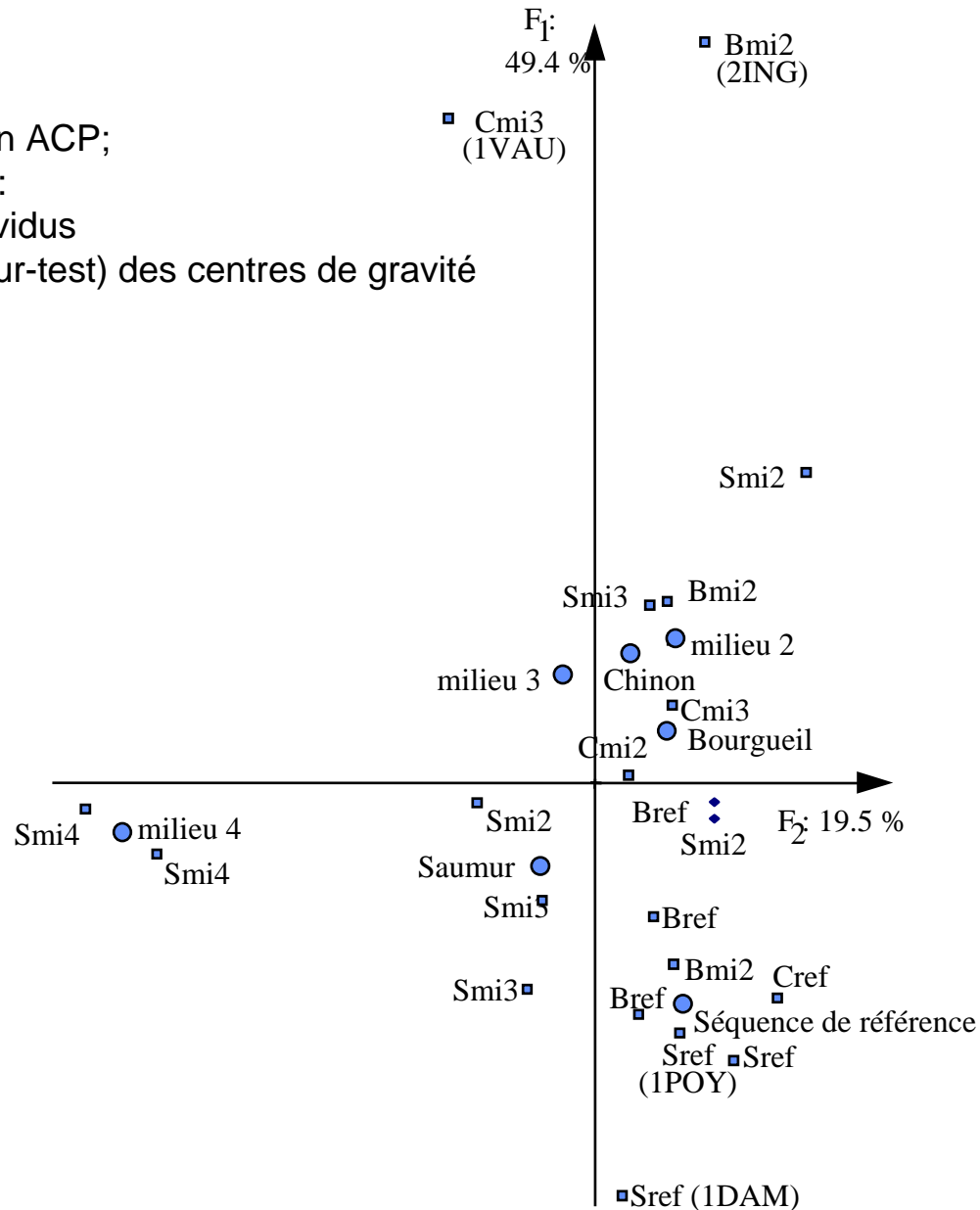
Les facteurs des analyses séparées sont projetées comme des éléments supplémentaires



$G2F1$  : premier facteur de l'ACP du groupe 2

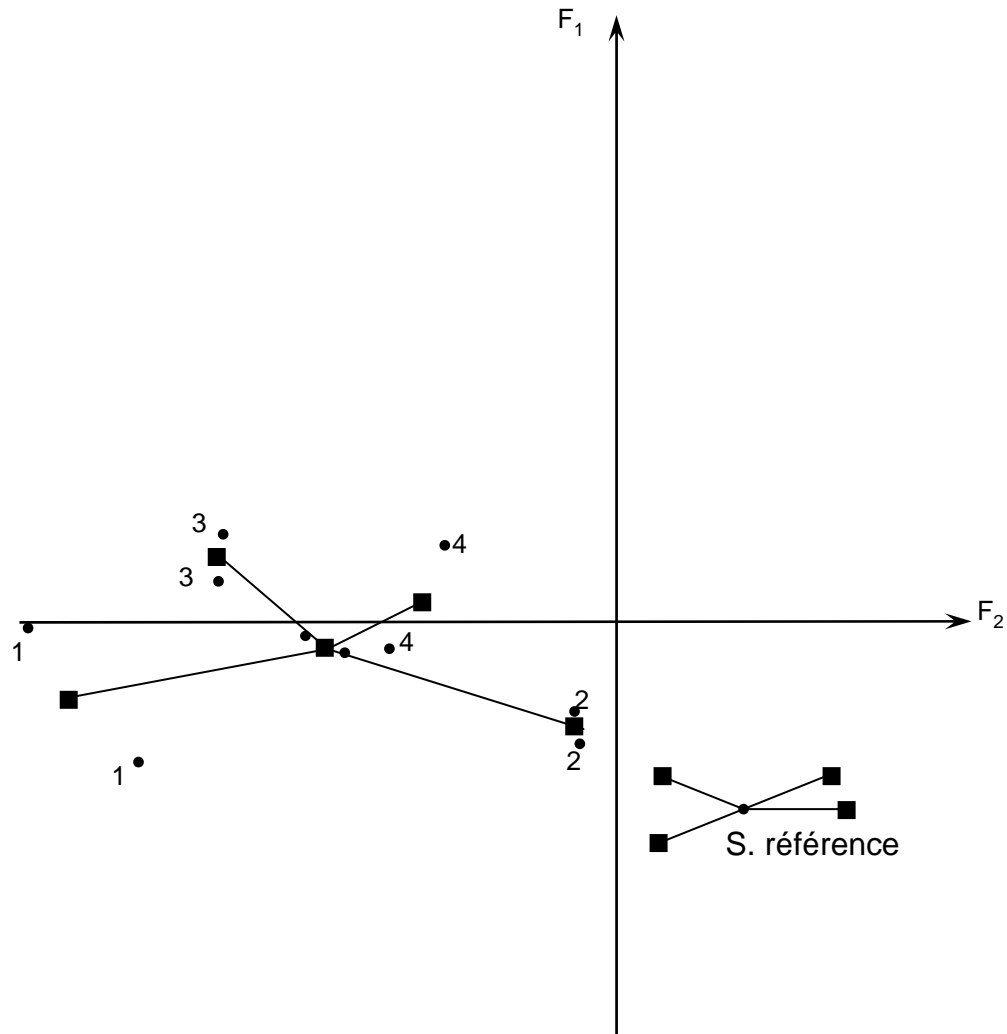
## 8. Cas des variables qualitatives

Illustratives : comme en ACP;  
 Pour chaque modalité :  
 identification des individus  
 coordonnées (et valeur-test) des centres de gravité



# Cas des variables qualitatives

Spécificité de l'AFM : points partiels associés aux centres de gravité



## Cas des variables qualitatives

Un seul groupe de variables : méthode de référence = analyse des correspondances multiples

Même problématique que pour les variables quantitatives

1. Equilibrer les groupes de variables dans une analyse globale
2. Représentation superposée des  $J$  nuages partiels
3. Recherche de facteurs communs
4. Représentation globale des groupes de variables
5. Relations entre l'analyse globale et les analyses séparées set

Même démarche en remplaçant ACP par ACM

Lorsque chaque groupe est réduit à une variable qualitative : AFM = ACM

## Données mixtes

Groupes composés de variables quantitatives

Groupes composés de variables qualitatives

La pondération de l'AFM permet d'analyser les deux types de variables ensemble

L'AFM fonctionne 'localement' comme

une ACP pour les variables quantitatives

une ACM pour les variables qualitatives

Lorsque chaque groupe est composé d'une seule variable, quantitative ou qualitative,  
AFM = Analyse Factorielle pour Données Mixtes (AFDM)

## 9. Conclusion

L'AFM est une méthode d'analyse de tableaux multiples dans lesquels les individus sont décrits par plusieurs groupes de variables

Les groupes peuvent être composés de variables quantitatives ou qualitatives

Le cœur de la méthode est une analyse factorielle pondérée qui fonctionne comme  
une ACP pour les groupes de variables quantitatives  
une ACM pour les groupes de variables qualitatives  
une AFDM pour les groupes mixtes

Elle fournit des résultats

classiques de l'analyse factorielle

représentation des individus, des variables, etc.

spécifiques de la structure en groupes de variables

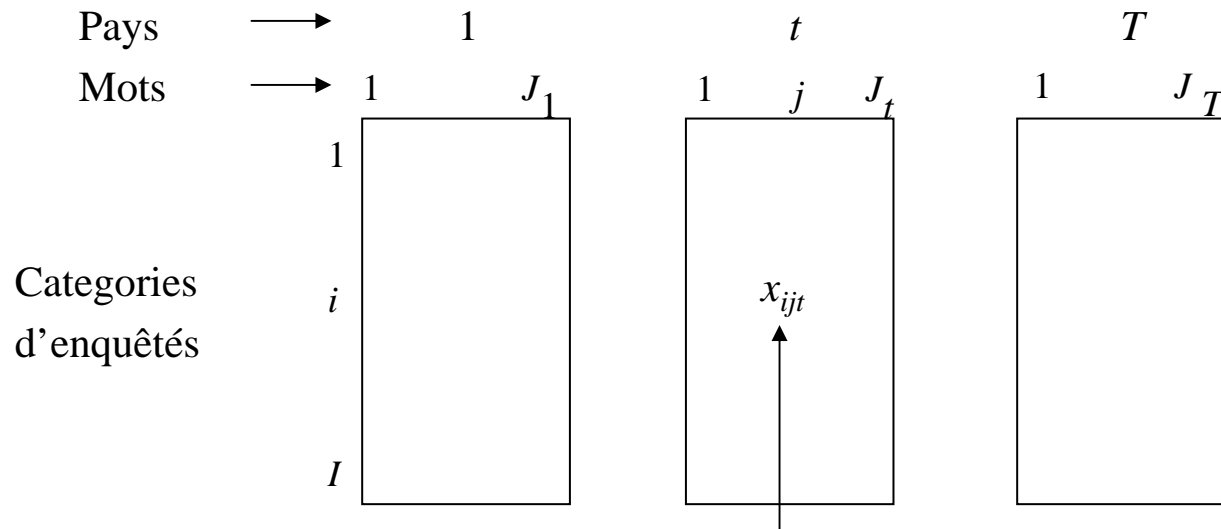
représentation des points partiels, des groupes, etc.



## Conclusion : extensions

L'AFM a été étendue aux tableaux de contingence multiples  
 Recherche menée avec Monica Becue (Barcelone, Espagne)  
 Les tableaux doivent avoir une dimension en commun

Application en analyse de données textuelles ; cas d'une même enquête  
 réalisée dans plusieurs pays



*fréquence du mot  $j$  pour les enquêtés appartenant à la catégorie  $i$  dans le pays  $t$*

## Conclusion : extensions

Ensemble d'individus décrits par plusieurs variables

Les variables sont structurées selon une hiérarchie  
Exemple : dans les questionnaires, les questions sont généralement structurées en thèmes et en sous-thèmes

Beaucoup d'applications en analyse sensorielle ; exemple de la comparaison entre jury et entre juges au sein des jurys

Jury 1			Jury 2			
Juge 1	Juge 2	Juge 3	Juge 4	Juge 5	Juge 6	Juge 7

Analyse Factorielle Multiple Hiérarchique (AFMH)

*Ces analyses sont incluses dans*

**FACTOMINER**

*Un package R dédié à*

*l'analyse exploratoire multidimensionnelle*

*développé par*

*le laboratoire de mathématiques appliquées*

*d'Agrocampus*

<http://factominer.free.fr>

## 10. Quelques références sur l'AFM

### Livre incluant une présentation de l'AFM

**Escofier B. & Pagès J.** (1998) Analyses factorielles simples et multiples ; objectifs, méthodes et interprétation. 3<sup>rd</sup> edition. 284 p, Dunod, Paris.

**Escofier B. & Pagès J.** (1992) Analisis factoriales simples y multiples ; objetivos, métodos e interpretación. Traducción de Elena Abascal Fdez., Karmele Fdez. Aguirre, Isabel Landaluce Calvo, José M. Piris Laespada, Amaya Zarraga Castro. Servicio editorial Universidad del país Vasco. (Information : etpfeagk@bs.ehu.es)

### Références méthodologiques

**Escofier B. & Pagès J.** (1994). Multiple Factor Analysis (AFMULT package). *Computational statistics & data analysis* 18 121-140

**Pagès J.** (1996). Eléments de comparaison entre l'Analyse Factorielle Multiple et la méthode STATIS. *Rev. Statistique appliquée* XLIV (4) 81-95

**Pagès J. & Ténenhaus M.** (2001). Multiple factor analysis combined with path modelling. Application to the analysis of relationships between physicochemical variables, sensory profiles and hedonic judgements. *Chemometrics and Intelligent Laboratory Systems* 58 261-273.

**Pagès J. & Ténenhaus M.** (2002) Analyse factorielle multiple et approche PLS. *Rev Statistique appliquée*, L (1) 5-33

# Quelques références sur l'AFM

## Références méthodologiques (suite)

**Pagès J.** (2002) Analyse factorielle multiple appliquée aux variables qualitatives et aux données mixtes. *Rev Statistique appliquée*, **L** (4), 5-37.

**Le Dien S., Pagès J.** (2003) Analyse factorielle multiple hiérarchique. *Rev Statistique appliquée*. **LI** (2) 47-73.

**Bécue M. & Pagès J.** (2003) A principal axes method for comparing contingency tables: MFACT. *Computational Statistics and Data Analysis*. (45) 3 481-503

**Pagès J.** (2004) Analyse factorielle de données mixtes. *Rev Statistique appliquée*. **LII** (4), 93-111.

**Pagès J.** (2005) Analyse factorielle multiple et analyse procustéenne. *Rev Statistique appliquée*. **LIII** (4), 61-86

**Pagès J., & Husson F.** (2005). Multiple factor analysis with confidence ellipses: a methodology to study the relationships between sensory and instrumental data. *J. Chemometrics*. 19. 1-7

**Husson, F. & Pagès, J.** (2006) Indscal Model: geometrical interpretation and methodology. *Computational Statistics and Data Analysis*. **50** (2). pp. 358-378

**Morand E. & Pagès J.** (2006). Procrustes multiple factor analysis to analyse the overall perception of food products. *Food quality and preference* 17 36-42.

## Références méthodologiques (suite)

**Husson, F. & Pagès, J.** (2006). Aspects méthodologiques du modèle Indscal. *Revue de Statistique Appliquée*. **LIV** (2) 83-100.

**Morand E. & Pagès J.** (2007). L'analyse factorielle multiple procustéenne. *Journal de la Société Française de Statistique*. 148 2 65-97

## Documentation SPAD

**Pagès J.** (1997). Procédure AFMUL : la méthode. In *SPAD TM : analyse des tableaux multiples*. CISIA Ed. 11-22.

**Pagès J.** (1997). Commentaire d'un exemple d'analyse factorielle multiple. In *SPAD TM : analyse des tableaux multiples*. CISIA Ed. 1-35.

# Quelques références sur l'AFM

## Applications

**Pagès J. & Escofier B. & Haury J.** (1991). Multiple factor analysis : a method to analyse several groups of variables measured on the same set of individuals. *Applied multivariate analysis in SAR and environmental studies*. Devillers & Karcher (Eds). Kluwer Academic Publishers. 33-83.

**Pagès J.** (1996). Quelques apports de l'AFM à l'analyse de données sensorielles. *J. int. Sci. Vigne Vin*. 30, n°4

**Pagès J. & Husson F.** (2001) Inter-laboratory comparison of sensory profiles. Methodology and results. *Food quality and preference* 12 297-309.

**Husson F., Le Dien, Pagès J.** (2001) Which value can be granted to sensory profiles given by consumers ? Methodology and results *Food quality and preference* 12 291-296

**Bécue M. & Pagès J.** (2001). Analyse simultanée de questions ouvertes et de questions fermées. Méthodologie, exemple. *Journal de la SFdS*. 142 (4), 91-104.

**Le Dien S., Pagès J.** (2003) Hierarchical Multiple Factor Analysis : application to the comparison of sensory profiles. *Food quality and preference*. 14 397-403.

**Pagès J.** (2003) Recueil direct de distances sensorielles : application à l'évaluation de dix vins blancs du Val de Loire. *Sciences des aliments*. (23) 679-688.

## Quelques références sur l'AFM

### Applications (suite)

**Pagès J., Deneulin P. & Asselin C.** (2004) Recherche méthodologique pour une caractérisation du cépage Chenin. Application au "rendez-vous du chenin ". *Revue française d'Oenologie*. 208, 35-40.

**Pagès, J. & Périnel E.** (2004) Panel performance and number of evaluations in a descriptive sensory study. *Journal of Sensory Studies*. **19** 273-291 .

**Pagès J.** (2005). Collection and analysis of perceived product inter-distances using multiple factor analysis; application to the study of ten white wines from the Loire Valley. *Food quality and preference* 16 642-649.

**Pagès J., Bertrand C., Ali R., Husson F. & Lê S.** (2007). Compared sensory analysis of eight biscuits by French and Pakistani panels. *Journal of Sensory Studies*. 22 665-686

**Perrin L., Symoneaux R., Maître I., Asselin C., Jourjon F. & J. Pagès** (2007). Comparison of Three Sensory Descriptive Methods for use with the Napping® procedure: Case of Ten Wines From Loire Valley. *Food quality and preference*. 19 (1) 1-11.

**Bécue-Bertault M., Alvarez-Esteban R & Pagès J.** (2008). Rating of products through scores and free text assertions : comparing and combining both. *Food quality and preference* 19 122-134.