

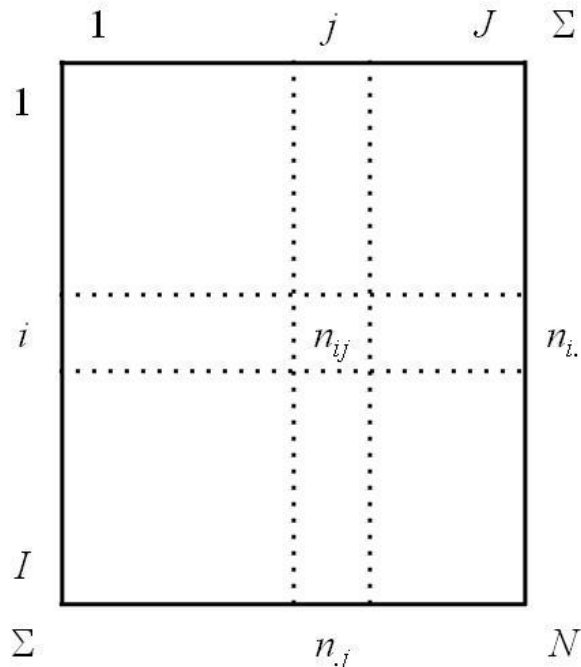
Analyse Factorielle des Correspondances Simples

Pr Roch Giorgi

 roch.giorgi@univ-amu.fr

Introduction (1)

- Étude simultanée de 2 variables qualitatives observées simultanément sur N individus
 - ✓ 2 variables qualitatives \Rightarrow tableau de contingence
 - ✓ Rôle symétrique des 2 variables
 - X a I modalités et Y a J modalités



Introduction (2)

- Liaison entre 2 variables qualitatives
 - ✓ X et Y ne sont pas liées si

$$\forall (i, j) \in \{1, \dots, I\} \times \{1, \dots, J\} : n_{ij} = \frac{n_{.j} \times n_{i.}}{N}$$

- ✓ Etude la liaison entre X et Y
 - ✓ Représentation Graphique des profils-lignes ou des profils-colonnes
 - ✓ Calcul du coefficient de Cramer,...
 - ✓ Test du Chi-deux

Objectifs

- Analyser la relation (correspondance) entre la typologie des lignes et la typologie des colonnes
 - ✓ Recherche des dimensions qui représentent le mieux l'écart à l'indépendance
- Décrire l'association entre les niveaux
- AFC est le résultat d'une double Analyse en Composantes Principales « pondérée »
 - ✓ ACP des profils-lignes
 - ✓ ACP des profils-colonnes
 - ✓ Relativement à la métrique du χ^2

Remarques

- AFC ignore les éventuelles structure d'ordre sur les modalités de X ou de Y
- Modalités mutuellement exclusives
- Chaque modalité doit avoir été observée au moins une fois (sinon, elle est supprimée)

Notations (1)

- Tableaux de données en AFC

	1	j	J	Σ
1				
i		n_{ij}		$n_{i.}$
I				
Σ		$n_{.j}$		N

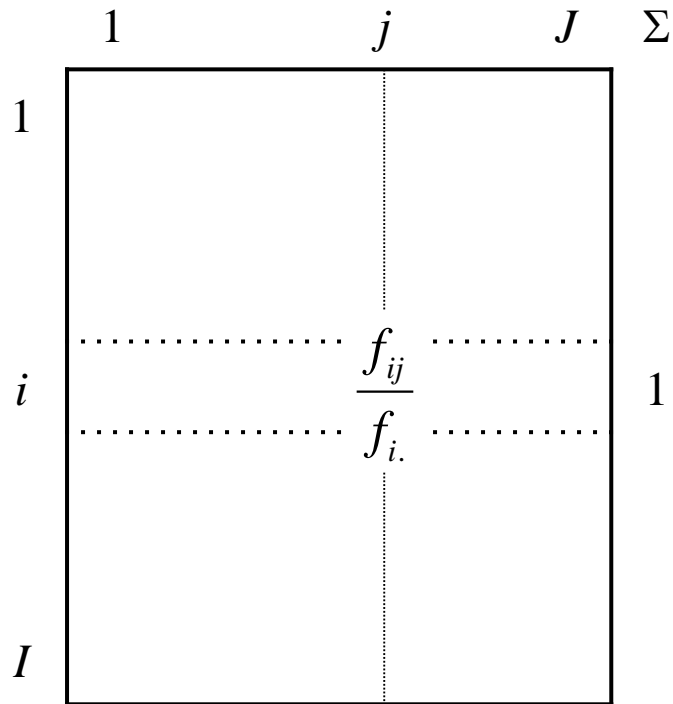
Effectifs

	1	j	J	Σ
1				
i		f_{ij}		$f_{i.}$
I				
Σ		$f_{.j}$		1

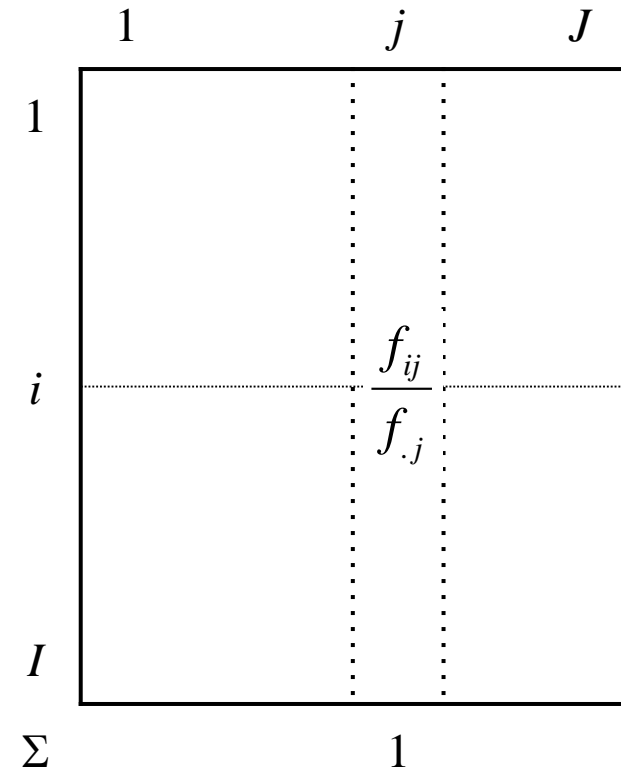
Fréquences relatives

Notations (2)

- Profils



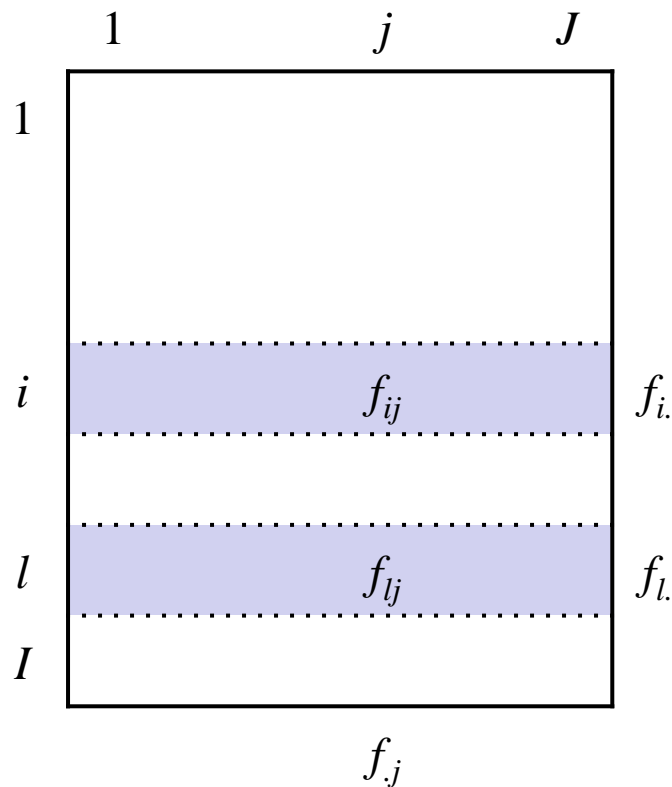
Lignes



Colonnes

Métriques du Chi-deux (1)

- Distance entre les x_i et x_l de X

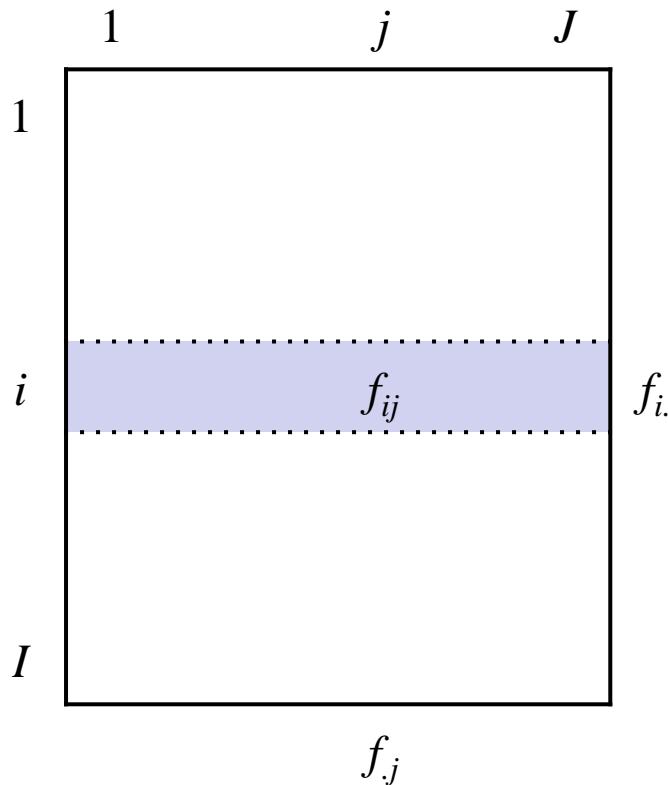


$$d^2(i, l) = \sum_{j=1}^J \frac{1}{f_{.j}} \left(\frac{f_{ij}}{f_{i.}} - \frac{f_{lj}}{f_{l.}} \right)^2$$

↑
pondération

Métriques du Chi-deux (2)

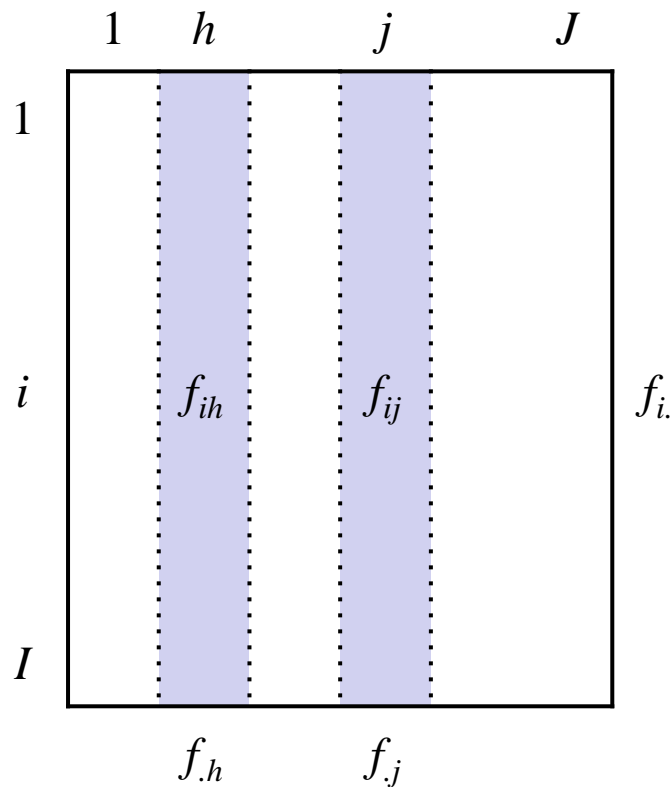
- Distance entre les x_i de X et le centre de gravité G



$$d^2(i, G) = \sum_{j=1}^J \frac{1}{f_{.j}} \left(\frac{f_{ij}}{f_{i.}} - f_{.j} \right)^2$$

Métriques du Chi-deux (3)

- Distance entre les y_h et y_j de Y

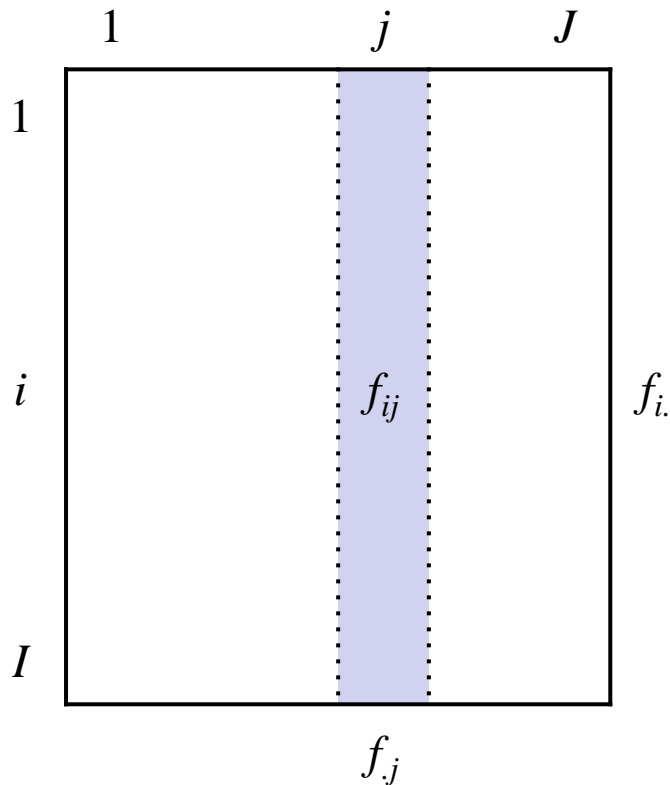


$$d^2(h, j) = \sum_{i=1}^I \frac{1}{f_{i.}} \left(\frac{f_{ij}}{f_{.j}} - \frac{f_{ih}}{f_{.h}} \right)^2$$

↑
pondération

Métriques du Chi-deux (4)

- Distance entre les y_j et le centre de gravité G



$$d^2(j, G) = \sum_{i=1}^I \frac{1}{f_{i.}} \left(\frac{f_{ij}}{f_{.j}} - f_{i.} \right)^2$$

Inertie Totale (1)

- Nuage des lignes

$$\begin{aligned}\text{Inertie totale} &= \sum_{i=1}^I f_{i.} \times d^2(i - G) \\ &= \frac{\chi^2}{n} = \phi^2\end{aligned}$$

ϕ^2 mesure l'intensité de la liaison

- ✓ Les coordonnées des lignes sont celles de l'analyse de l'inertie du nuage des I distributions conditionnelles par ligne, pondéré et centré par la distribution marginales des lignes utilisant la métrique des inverses des poids des colonnes

Inertie Totale (2)

- Nuage des colonnes

$$\begin{aligned}\text{Inertie totale} &= \sum_{j=1}^J f_{.j} \times d^2(j - G) \\ &= \frac{\chi^2}{n} = \phi^2\end{aligned}$$

ϕ^2 mesure l'intensité de la liaison

- ✓ Les coordonnées des colonnes sont celles de l'analyse de l'inertie du nuage des J distributions conditionnelles par colonne, pondéré et centré par la distribution marginales des colonnes utilisant la métrique des inverses des poids des lignes

ACP des Profils

- Lignes

- ✓ Les « individus » sont les modalités de X ou profils-lignes
- ✓ Pondérés par les fréquences marginales des lignes
- ✓ En utilisant la métrique du χ^2

- Colonnes

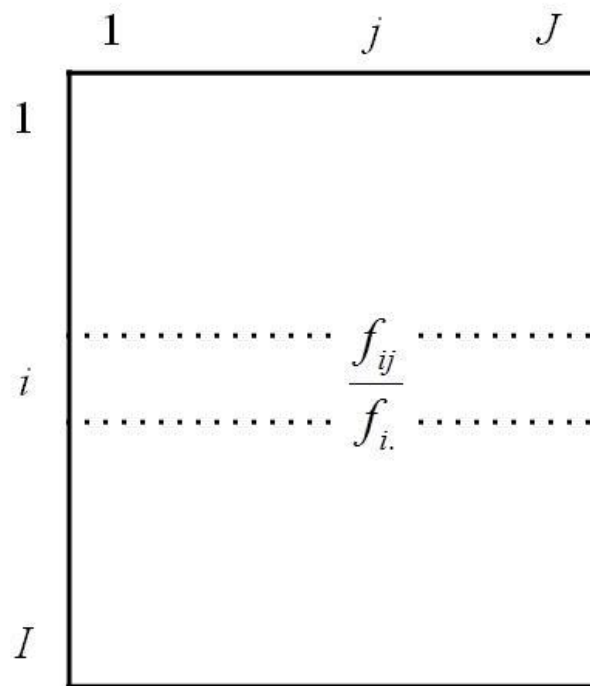
- ✓ Les « individus » sont les modalités de Y ou profils-colonnes
- ✓ Pondérés par les fréquences marginales des colonnes
- ✓ En utilisant la métrique du χ^2

Remarque

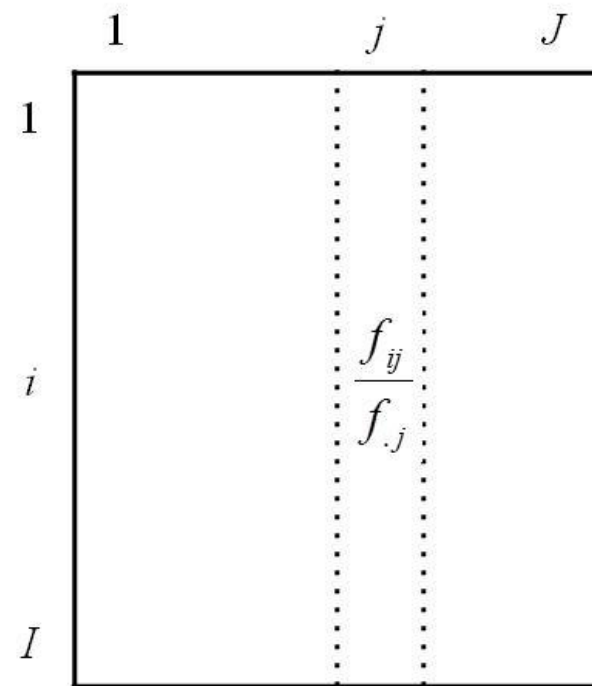
- Analyse effectuée sur une matrice de distance
- Equivalent à l'utilisation de l'analyse par positionnement multidimensionnel (MDS)

AFC comme une Double ACP (1)

- Deux ACP pondérées sur les profiles-lignes et les profils-colonnes
- Présentation obtenue à partir des 2 profils



Lignes

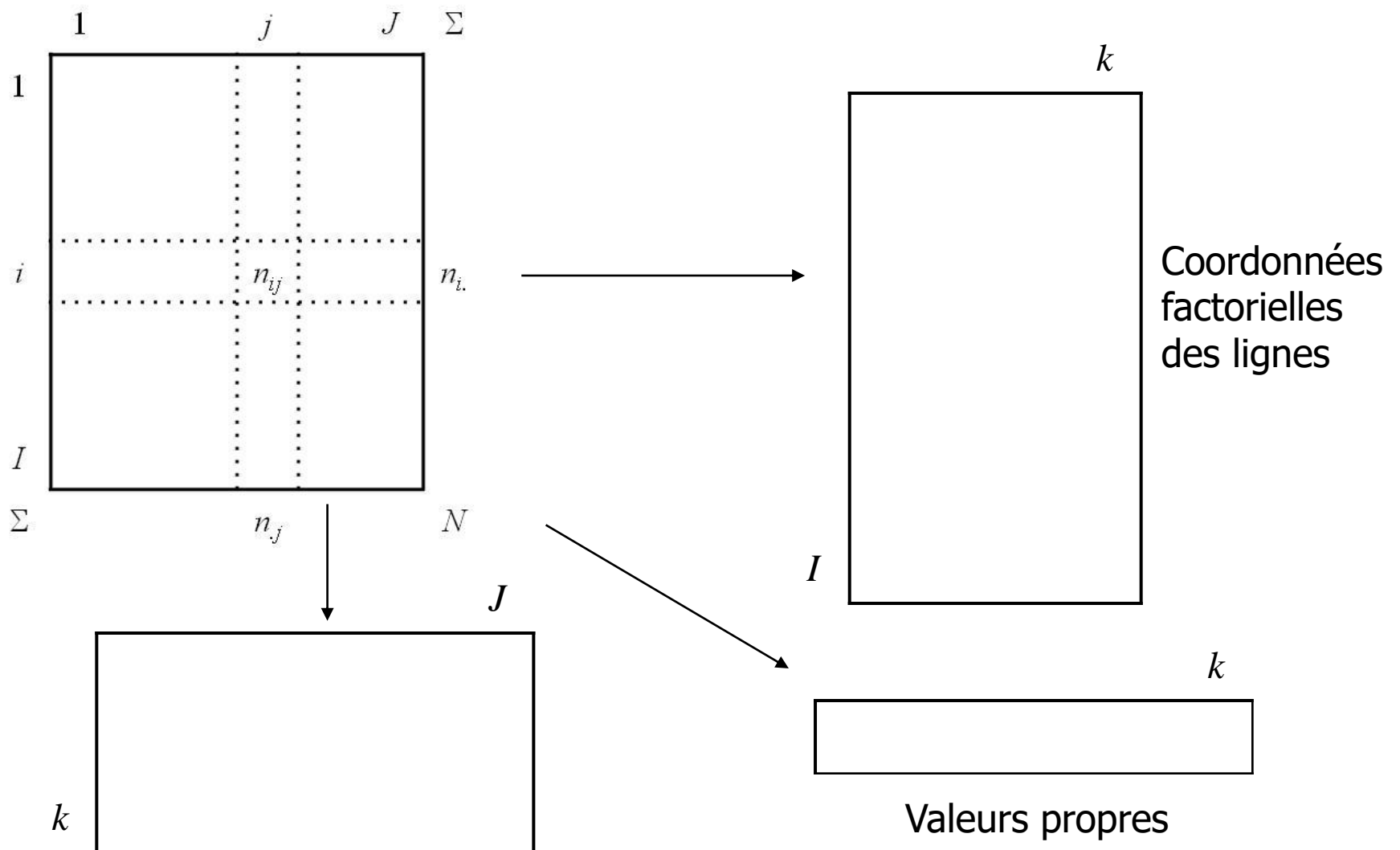


Colonnes

AFC comme une Double ACP (2)

- Formule de transition
 - ✓ La ligne i est au barycentre des colonnes pondérées
 - ✓ La colonne j est au barycentre des lignes pondérées
- Possibilité de représenter simultanément les 2 présentations

Principaux Résultats d'une AFC



Coordonnées factorielles des colonnes

Exemple (1)

- 12 parfums décrits par 15 mots

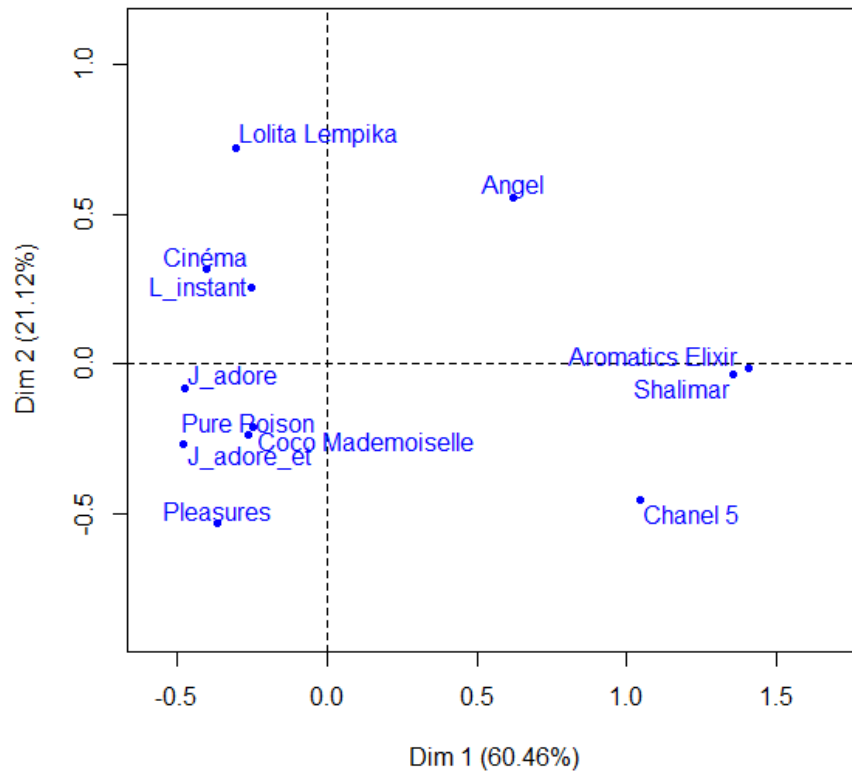
Parfums	floral	fruity	strong	soft	...
Angel	2	11	18	3	...
Aromatics Elixir	2	3	29	2	...
Chanel 5	5	0	19	3	...
Cinéma	14	14	3	12	...
...

- ✓ Image synthétique de ces 12 parfums de luxe ?
- ✓ Quels parfums sont sensoriellement proches ?
- ✓ Pour quelles raisons ces parfums s'opposent ?

Source : *Josse J. L'analyse des correspondances simples. 23 novembre 2010.*
FactoMineR

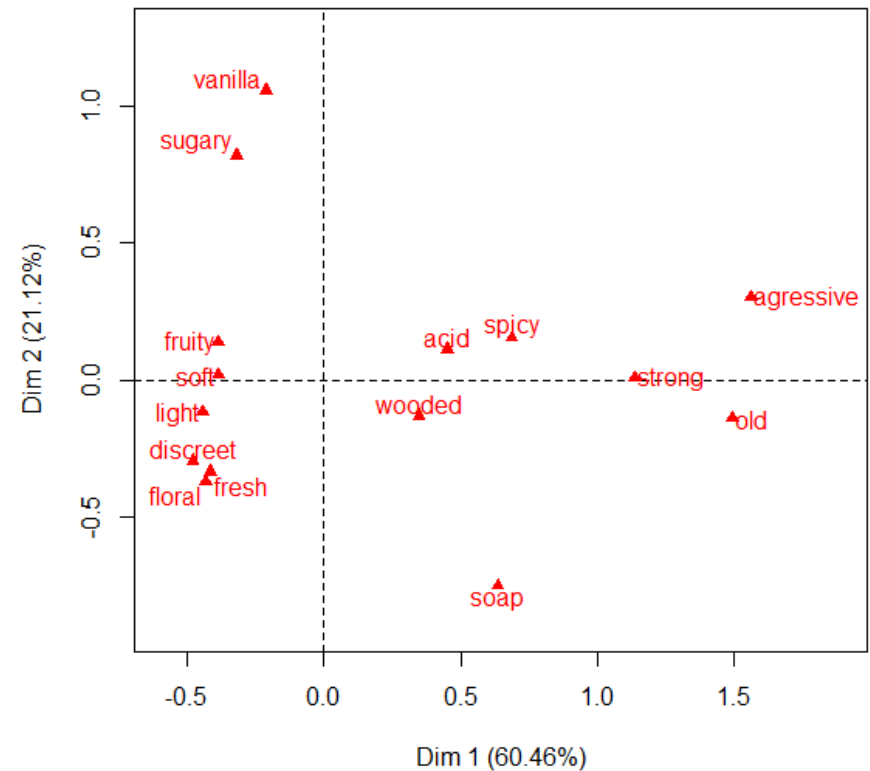
Exemple (2)

CA factor map



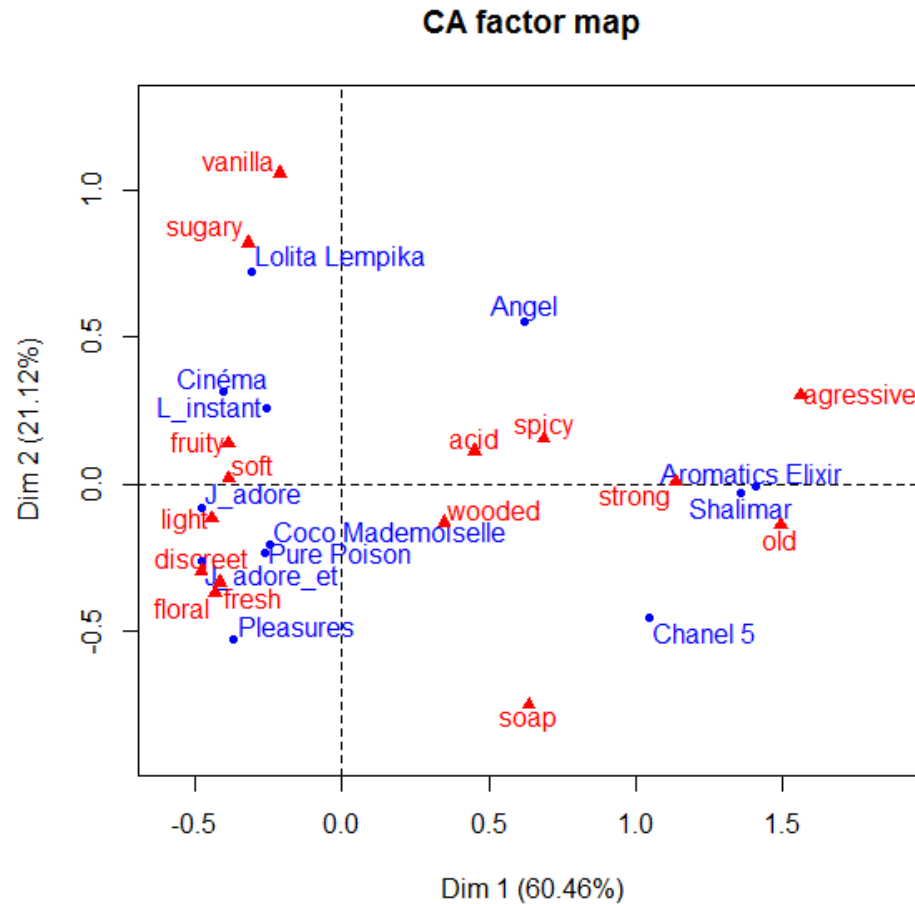
Lignes

CA factor map



Colonnes

Exemple (3)

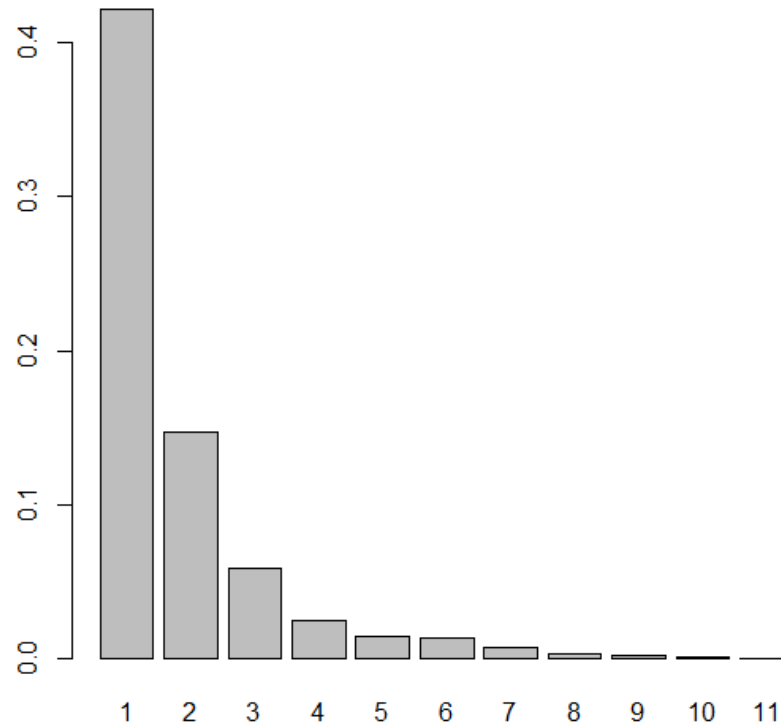


Présentation superposée

Pourcentage d'Inertie

- Pourcentage de variance de l'axe s $\frac{\lambda_s}{\sum_k \lambda_k}$

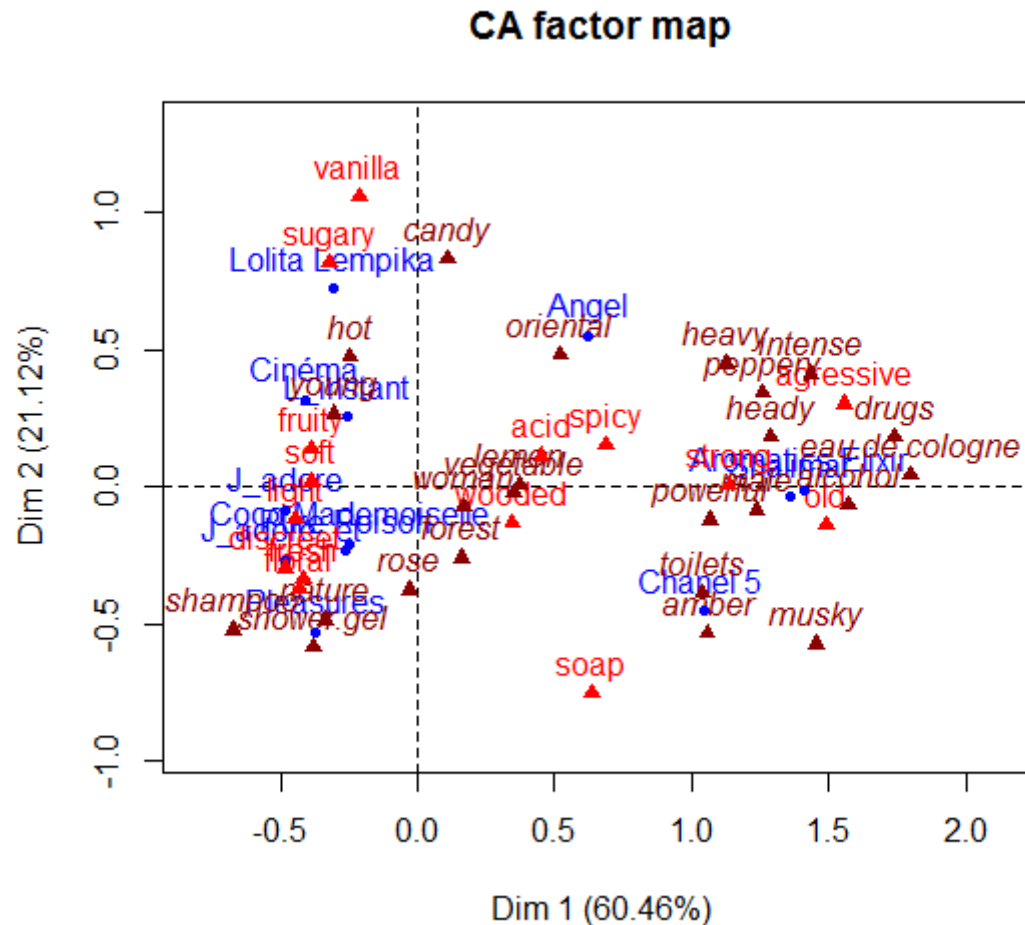
Valeurs propres



Aides à l'Interprétation

- Mesure de la qualité globale
 - ✓ Rapport entre la somme des q première valeurs propres et leur somme complète
- Mesure de la qualité de chaque modalité
 - ✓ Inertie projetée d'un point sur l'axe s divisée par l'inertie total du point : cosinus carré de l'angle formé
- Contribution à l'inertie totale
 - ✓ Inertie projetée d'un point sur l'axe s divisée par l'inertie de l'axe s
 - ✓ Les points extrêmes ne sont pas ceux qui contribuent le plus à la construction des axes
- Superposition de variables supplémentaires

Superposition Variables Supplémentaires



Remarques

- En général on n'interprète pas les axes d'une AFC
- L'interprétation s'appuie sur la position relative des différentes modalités repérées comme les plus importantes
- Equivalence distributionnelle
 - ✓ Considérer 2 points-lignes (resp. colonnes) comme un seul point affecté de la somme de leurs masses ne modifie pas les distances
 - ✓ Regroupement possible de mots/libellés équivalents

Sources

- Besse P., Baccini A. Statistique descriptive multidimensionnelle. 1999.
- Josse J. L'analyse des correspondances simples. 23 novembre 2010.
- FactoMineR : <http://factominer.free.fr/>