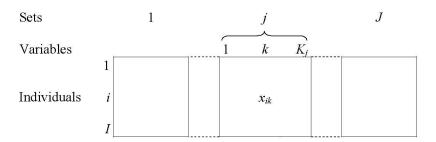
# Analyse Factorielle Multiple

Julie Josse Laboratoire de mathématique appliquées Agrocampus Ouest Rennes

## Analyse Factorielle Multiple

- 1 Données problématique
- 2 Une ACP globale
- 3 Etude des groupes
- 4 Nuages partiels
- 6 Analyses séparées
- 6 Compléments
- 7 Conclusion pour aller plus loin

### Tableaux multiples



#### Examples avec des variables continues et/ou qualitatives :

- genomique : ADN, protéines
- analyse sensorielle : sensorielles, physico-chimiques
- questionnaires : santé des étudiants (consommations de produits, conditions psychologiques, sommeil, signalétique)
- economie : indicateurs économiques chaque année

### Données vins - Experts

- 10 individus (lignes) : vins blancs du Val de Loire
- 30 variables (colonnes) :
  - 27 variables quantitatives : descripteurs sensoriels
  - 2 variables quantitatives : appréciation de l'odeur et générale
  - 1 variable qualitative : label (Vouvray Sauvignon)

	O.fruity	O.passion	O.citrus	 Sweetness	Acidity	Bitterness	Astringency	Aroma.intensity	Aroma.persistency	Visual.intensity	Odor.preferene	Overall.preference	Label
S Michaud	4.3	2.4	5.7	 3.5	5.9	4.1	1.4	7.1	6.7	5.0	6.0	5.0	Sauvignon
S Renaudie	4.4	3.1	5.3	 3.3	6.8	3.8	2.3	7.2	6.6	3.4	5.4	5.5	Sauvignon
S Trotignon	5.1	4.0	5.3	 3.0	6.1	4.1	2.4	6.1	6.1	3.0	5.0	5.5	Sauvignon
S Buisse Domaine	4.3	2.4	3.6	 3.9	5.6	2.5	3.0	4.9	5.1	4.1	5.3	4.6	Sauvignon
S Buisse Cristal	5.6	3.1	3.5	 3.4	6.6	5.0	3.1	6.1	5.1	3.6	6.1	5.0	Sauvignon
V Aub Silex	3.9	0.7	3.3	 7.9	4.4	3.0	2.4	5.9	5.6	4.0	5.0	5.5	Vouvray
V Aub Marigny	2.1	0.7	1.0	 3.5	6.4	5.0	4.0	6.3	6.7	6.0	5.1	4.1	Vouvray
V Font Domaine	5.1	0.5	2.5	 3.0	5.7	4.0	2.5	6.7	6.3	6.4	4.4	5.1	Vouvray
V Font Brûlés	5.1	0.8	3.8	 3.9	5.4	4.0	3.1	7.0	6.1	7.4	4.4	6.4	Vouvray
V Font Coteaux	4.1	0.9	2.7	 3.8	5.1	4.3	4.3	7.3	6.6	6.3	6.0	5.7	Vouvray

#### Données vins

#### Continuous variables

Categorical

	Expert (27)	Consu mer (15)	Student (15)	Preference (60)	Label (1)
wine 1					
wine 2					
wine 10					

- Comment est-ce que les vins sont décrits?
- Est-ce que les vins sont décrits de la même façon par les différents panels? Y-a t'il des spécificités par panel?

### **Objectifs**

- Equilibrer l'influence de chaque groupe (prendre en compte la structure des données)
- Etudier les ressemblances entre individus et les liaisons entre variables
- Etudier globalement les ressemblances et les différences entre groupes (voir les spécificités de chaque groupe)
- Etudier les ressemblances et les différences entre groupes du point de vue individuel
- Comparer les typologies issues des analyses séparées

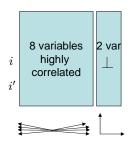
#### Equilibrer l'influence des groupes de variables

#### L'AFM est une ACP pondérée :

- calculer  $\lambda_1^j$  pour chaque groupe de variables
- réaliser l'ACP globale sur le tableau pondéré :

$$\left[\frac{X_1}{\sqrt{\lambda_1^1}}; \frac{X_2}{\sqrt{\lambda_1^2}}; ...; \frac{X_J}{\sqrt{\lambda_1^J}}\right]$$

 $\Rightarrow$  Même idée que pour l'ACP standardisée : les variables sont pondérées dans le calcul des distances entre les individus i et i'



#### Equilibrer l'influence des groupes de variables

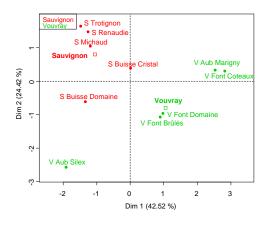
	Expert	Consommateur	Etudiant
$\overline{\lambda_1}$	11.74	7.17	7.89
$\lambda_2$	6.78	2.59	3.83
$\lambda_3$	2.74	1.63	1.70

- Même poids pour toutes les variables d'un même groupe : la structure du groupe est préservée
- Pour chaque groupe, la variance de la principale dimension de variabilité (première valeur propre) est égale à 1
- Aucun groupe ne peut générer à lui seul la première dimension
- Un groupe multidimensionnel contribue à plus de dimensions qu'un groupe uni-dimensionnel

### Equilibrer l'influence des groupes de variables

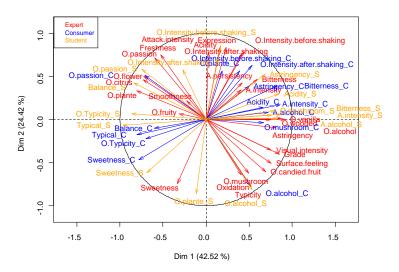
- ⇒ Le coeur de l'AFM est une ACP pondérée
- ⇒ Mêmes représentations qu'en ACP
  - Etudier les ressemblances entre individus du point de vue de l'ensemble des variables
  - Etudier les relations entre variables
  - Décrire les indidus à partir des variables
- ⇒ Mêmes sorties (coordonnées, cosinus, contributions)
- ⇒ Ajouter des individus et variables (quantitatives et qualitatives) en supplémentaire

#### Représentation des individus

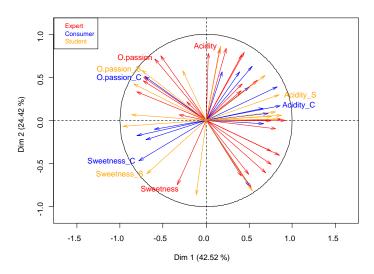


- Les deux labels sont bien séparés
- Les Vouvray sont plus différents du point de vue sensoriel
- Plusieurs groupes de vins, ...

#### Représentation des variables



#### Représentation des variables



### Etudes des groupes

⇒ Comparaison synthétique des groupes

⇒ Est-ce que la position relative des individus est similaire d'un groupe à l'autre? Est-ce que les nuages partiels sont similaires?

### Mesures de similarité entre groupes

 $X_{j_{(I \times K_i)}}$  et  $X_{m_{(I \times K_m)}}$  ne sont pas directement comparables  $W_{j_{(I\times I)}}=X_jX_j'$  et  $W_{m_{(I\times I)}}=X_mX_m'$  sont comparables Matrice de produits scalaires = position relative des individus

Covariance entre deux groupes :

$$< W_j, W_m > = \sum_{k \in K_j} \sum_{l \in K_m} cov^2(x_{.k}, x_{.l})$$

Corrélation entre deux groupes :

$$RV(K_j, K_m) = \frac{\langle W_j, W_m \rangle}{\|W_i\| \|W_m\|}$$
  $0 \le RV \le 1$ 

RV = 0: toutes les variables sont orthogonales RV = 1: les deux nuages de points sont homothétiques

### Mesures de similarité entre groupes

⇒ Le coefficient RV : corrélation entre matrices

 $\Rightarrow$  Le coefficient  $\mathcal{L}_{\varphi}$ :

$$\mathcal{L}_{g}(K_{j}, K_{m}) = \sum_{k \in K_{j}} \sum_{l \in K_{m}} cov^{2} \left( \frac{X_{.k}}{\sqrt{\lambda_{1}^{k}}}, \frac{X_{.l}}{\sqrt{\lambda_{1}^{l}}} \right)$$

 $\Rightarrow$  Le coefficient  $\mathcal{L}_{g}$  pour un groupe : un indice de dimensionalité

$$\mathcal{L}_{g}(K_{j}, K_{j}) = \frac{\sum_{k=1}^{K_{j}} (\lambda_{k}^{j})^{2}}{(\lambda_{1}^{j})^{2}} = 1 + \frac{\sum_{k=2}^{K_{j}} (\lambda_{k}^{j})^{2}}{(\lambda_{1}^{j})^{2}}$$

### Mesures de similarité entre groupes

> res.mfa\$gr	oup\$Lg					
	Expert	Consumer	Student	Preference	Label	MFA
Expert	1.45	0.94	1.17	1.01	0.89	1.33
Consumer	0.94	1.25	1.04	1.11	0.28	1.21
Student	1.17	1.04	1.29	1.03	0.62	1.31
Preference	1.01	1.11	1.03	1.47	0.37	1.18
Label	0.89	0.28	0.62	0.37	1.00	0.67
MF A	1.33	1.21	1.31	1.18	0.67	1.44
> res.mfa\$gr	oup\$RV					
	Expert	Consumer	Student	Preference	Label	MF A
Expert	1.00	0.70	0.85	0.69	0.74	0.92
Consumer	0.70	1.00	0.82	0.82	0.25	0.90
Student	0.85	0.82	1.00	0.75	0.55	0.96
Preference	0.69	0.82	0.75	1.00	0.31	0.81
Label	0.74	0.25	0.55	0.31	1.00	0.56
MF A	0.92	0.90	0.96	0.81	0.56	1.00

- Les Experts donnent une description plus riches ( $\mathcal{L}_g$  supérieur)
- Les groupes Student et Expert sont liés (RV = 0.85)

#### Première composante de l'AFM

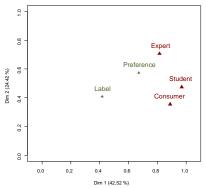
⇒ La première composante principale de l'AFM maximise :

$$\sum_{j=1}^{J} \sum_{k \in \mathcal{K}_j} cov^2 \left( \frac{x_{.k}}{\sqrt{\lambda_1^j}}, v_1 \right) = \sum_{j=1}^{J} \mathcal{L}_{g}(\mathcal{K}_j, v_1)$$

Inerties des  $K_i$  projectées sur  $v_1$ 

### Représentation des groupes

Le groupe j a une coordonnée  $(\mathcal{L}_g(v_1, K_i), \mathcal{L}_g(v_2, K_i))$ 



$$0 \leq \mathcal{L}_{g}(v_{1}, \mathcal{K}_{j}) = \frac{1}{\lambda_{1}^{j}} \sum_{k \in \mathcal{K}_{j}} cov^{2}(x_{.k}, v_{1}) \leq 1$$

- Première dimension commune à tous les groupes
- Deuxième dimension due au groupe Expert
- Les préférences sont liées aux descriptions sensorielles

$$\underbrace{K_{i}}_{\leq K_{i}}$$

#### Autre exemple de représentation des groupes

50 vins mousseux (dont 26 champagnes) décrits par 25 amateurs et 7 oenologues

Les données individuelles sont juxtaposées en ligne : tableau produits  $\times$  juges-descripteurs

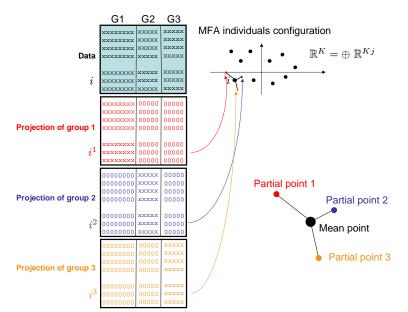
$$\Rightarrow$$
 1 juge = 1 groupe

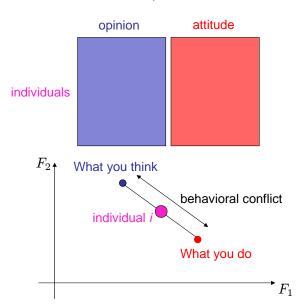
Les amateurs et les oenologues sont ils mélangés?

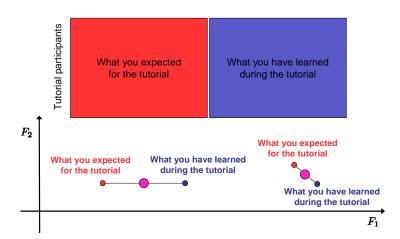
⇒ Comparaison des groupes au travers des individus

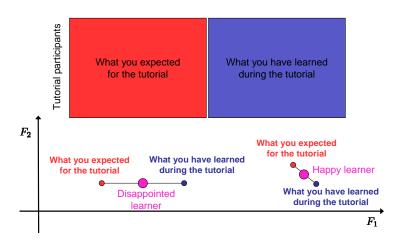
 $\Rightarrow$  Est-ce qu'il y a des individus très particuliers pour un certain groupe de variables ?

### Projection des points partiels



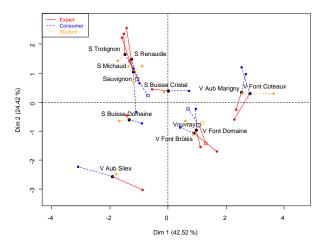






jectifs ACP globale Etude des groupes Points partiels Analyses séparées Compléments Conclusio

#### Representation des points partiels



- Un individu est au barycentre de ses points partiels
- Un individu est d'autant plus "homogène" que sont "étoile" est petite

### Optimalité

- Tous les nuages de points doivent être bien représentés ⇒ Maximiser l'inertie projetée globale
- Les J points representant le même individus doivent être proche les un des autres
- ⇒ Minimiser l'inertie projetée intra

Huygens theorem: inertie totale = inertie inter + inertia intra

$$\sum_{i=1}^{I} \sum_{j=1}^{J} (F_{ij}_{q})^{2} = \sum_{i=1}^{I} \sum_{j=1}^{J} (F_{iq})^{2} + \sum_{i=1}^{I} \sum_{j=1}^{J} (F_{ij}_{q} - F_{iq})^{2}$$

⇒ Maximiser l'inertie inter : réaliser une ACP globale sur le tableau de données concaténé

#### Ratio d'inertie

$$\sum_{i=1}^{I} \sum_{j=1}^{J} (F_{ij}_{q})^{2} = \sum_{i=1}^{I} \sum_{j=1}^{J} (F_{iq})^{2} + \sum_{i=1}^{I} \sum_{j=1}^{J} (F_{ij}_{q} - F_{iq})^{2}$$

inertie totale = inertie inter individus + inertia intra individu

```
> res.mfa$inertia.ratio
Dim.1 Dim.2 Dim.3 Dim.4 Dim.5
0.93 0.82 0.78 0.54 0.53
```

- Sur la première dimension, les coordonnées des points partiels sont proches (0.93 proche de 1)
- L'inertie intra mesure la ressemblance entre les nuages partiels dimension par dimension
- L'inertie intra peut être décomposée par individus

## Analyse procrustéenne / AFM

 $\Rightarrow$  Les relations de transitions s'appliquent toujours pour les points partiels

 $\Rightarrow$  La représentation superposée peut être interprétée dans un cadre unique

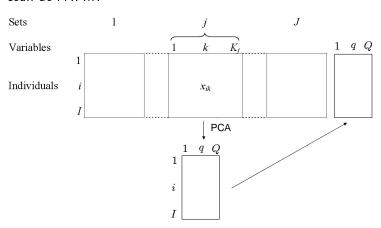
### Analyses séparées

 $\Rightarrow$  Comparison des typologies obtenues pour chaque groupe dans un espace commun

⇒ Comparaison des ACP séparées

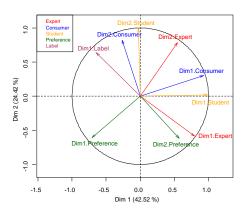
#### Représentation des composantes principales

Est-ce que les analyses séparées donnent les mêmes résultats que ceux de l'AFM?



jectifs ACP globale Etude des groupes Points partiels **Analyses séparées** Compléments Conclusio

#### Représentation des composantes principales

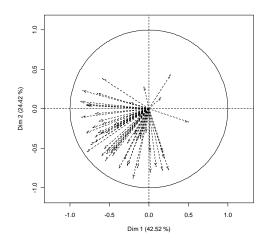


- Les deux premières dimensions de chaque groupe sont bien projetées
- Les étudiants ont les mêmes dimensions que l'AFM

⇒ Les composantes principales des analyses séparées sont projetées en supplémentaires

ojectifs ACP globale Etude des groupes Points partiels Analyses séparées **Compléments** Conclusio

#### Représentation des variables supplémentaires



Les préférences sont liées aux descriptions sensorielles Le vin préféré est le *Vouvray Aubussière Silex*  jectifs ACP globale Etude des groupes Points partiels Analyses séparées **Compléments** Conclusio

### Description des dimensions

#### Par des variables quantitatives :

- corrélation entre chaque variable et la composante principale de rang q est calculée
- les coefficients de corrélations sont triés et les significatifs sont donnés

```
> dimdesc(res.pca)
            $Dim.1$quanti
                                                         $Dim.2$quanti
                 corr p.value
                                                              corr p.value
O. candied. fruit
                 0.93 9.5e-05
                               Odor.Intensity.before.shaking 0.97 3.1e-06
Grade
                 0.93 1.2e-04
                               Odor. Intensity.after. shaking
                                                              0.95 3.6e-05
Surface feeling 0.89 5.5e-04
                                                              0.85 1.7e-03
                               Attack.intensity
Typicity
                 0.86 1.4e-03
                               Expression
                                                              0.84 2.2e-03
O.mushroom
                 0.84 2.3e-03
                               Aroma.persistency
                                                              0.75 1.3e-02
Visual.intensity 0.83 3.1e-03
                               Bitterness
                                                              0.71 2.3e-02
                                Aroma.intensity
                                                              0.66 4.0e-02
   . . .
0.plante
                -0.87 1.0e-03
0.flower
                -0.894.9e-04
O.passion
                -0.90 4.5e-04
Freshness
                -0.91 2.9e-04 Sweetness
                                                             -0.78 8.0e-03
```

jectifs ACP globale Etude des groupes Points partiels Analyses séparées **Compléments** Conclusio

#### Description des dimensions

#### Par des variables qualitatives :

- réalise une analyse de variance avec les coordonées des individus (F<sub>.q</sub>) expliquées par la variable qualitative
  - un F-test par variable
  - pour chaque categorie, un test de Student t-test

```
> dimdesc(res.pca)
Dim. 1$quali
              R.2
                      p.value
                     7.30e-05
           0.874
Label
Dim. 1$category
           Estimate
                           p.value
               3.203
                          7.30e-05
Vouvray
Sauvignon
              -3.203
                          7.30e-05
```

#### Données qualitatives

Un seul groupe de variables : méthode de référence = ACM

- ⇒ Même problématique que pour les variables quantitatives :
  - Equilibrer les groupes de variables dans une analyse globale
  - Représentation superposée des J nuages partiels
  - Représentation globale des groupes de variables
  - Relations entre l'analyse globale et les analyses séparées
- ⇒ 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 et 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)

#### Conclusion

- AFM : une méthode multi-tableaux pour les variables quantitatives et qualitatives
- Le coeur de l'AFM est une ACP pondérée
- Equivalence AFM-ACP ou AFM-ACM quand chaque groupe est composé d'une seule variable
- Sorties classiques (individus, variables)
- Sorties spécifiques (groupes, points partiels)
- AFM est disponible dans de nombreux logiciels: R
   (FactoMineR), SPAD, Statgraphics, XLSTAT (Microsoft Excel), SAS (macro B. Gelein)

#### Menu déroulant de FactoMineR

source("http://factominer.free.fr/install-facto.r")



#### Mise en œuvre d'une AFM

- 1 Definir les groupes actifs et les éléments supplémentaires
- 2 Réduire ou non les variables?
- 3 Réaliser une AFM
- 4 Choisir le nombre de dimensions à interpréter
- 6 Interprétation simultanée du graphe des individus et des variables
- 6 Etudes des groupes
- Analyses partielles
- 8 Utilisation d'indicateurs pour enrichir l'interprétation

jectifs ACP globale Etude des groupes Points partiels Analyses séparées Compléments Conclusion

### Pour aller plus loin

- AFM utilisée dans un but méthodologique :
  - comparaison de codage (quantitative ou qualitative?)
  - comparaison entre preprocessing (ACP standardisée et ACP non standardisée)
  - comparaison de résultats de différentes analyses
- Analyse Factorielle Multiple Hierarchique ⇒ prendre en compte une hierarchie sur les variables : les variables sont structurées en groupes et en sous-groupes
  - questionnaires avec des thèmes et des sous-thèmes
  - beaucoup d'applications en analyse sensorielle : exemple de la comparaison entre jury et entre juges au sein des jurys
- Gestion des données manquantes

### Bibliographie

- Escofier, B. & Pagès, J. (2008). Analyses Factorielles Simples et Multiples: Objectifs, Méthodes et Interprétation. Dunod, 4th edn.
- Husson, F., Lê, S. & Pages, J. (2009). Analyse de données aves R. Presse Universitaires de Rennes.
- Pagès, J. (1996). Quelques apports de l'AFM à l'analyse de données sensorielles. J. int. Sci. Vigne Vin. 30, (4).
- Pagès, J. & Husson, F. (2001) Inter-laboratory comparison of sensory profiles. Methodology and results. Food quality and preference (12) 297-309.
- 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.