

# Une introduction à l'Analyse Factorielle Multiple (AFM)

E. Morand

Semine-R

21 Février 2011

# Introduction

# De deux nuages de points à plusieurs nuages

## Comparaison de deux nuages de points

Exemple

Analyse Procustéenne

Coefficients RV

## Comparaison de plus de deux nuages de points

Analyse Procustéenne généralisée (APG)

Le problème

Quelques méthodes

# Introduction

Données

Résultats

- Cas des groupes

- Comparaison des groupes

- Modalités

Conclusion

# Un exemple

| 1° mois | 2°mois | 3°mois | ... | ieme mois | Age | Sexe | CSP | Etudes |
|---------|--------|--------|-----|-----------|-----|------|-----|--------|
| CDI     | CDD    | CDD    |     | CDI       |     |      |     |        |
| CDD     | CDD    | CDD    |     | CDI       |     |      |     |        |

| 1° mois | 2°mois | 3°mois | ... | ieme mois | Age | Sexe | CSP | Etudes |
|---------|--------|--------|-----|-----------|-----|------|-----|--------|
| CDI     | CDD    | CDD    |     | CDI       |     |      |     |        |
| CDD     | CDD    | CDD    |     | CDI       |     |      |     |        |

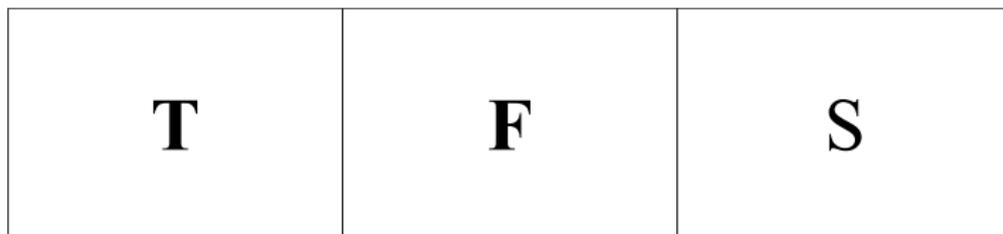
## Traitement des données

### Analyse des Correspondances Multiples

- ▶ Variables Actives : les trajectoires
- ▶ Variables Illustratives : signalétique (socio démographique)

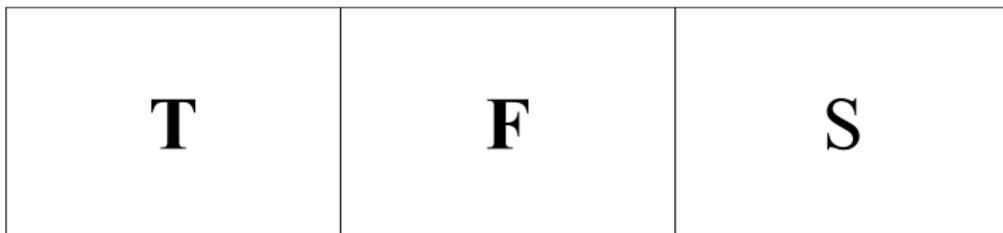
# Et si ...

On souhaite étudier une trajectoire familiale ?



## Et si ...

On souhaite étudier une trajectoire familiale ?

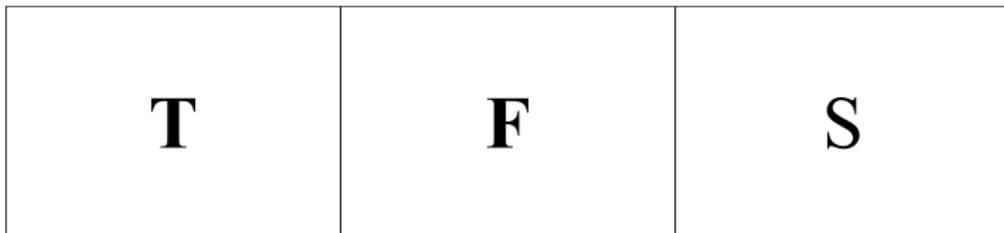


### Traitement des données

- ▶ On choisit de mettre les informations sur la trajectoire professionnelle et sur la trajectoire personnelle en tant que variables actives dans l'analyse.

## Et si ...

On souhaite étudier une trajectoire familiale ?



### Traitement des données

- ▶ On choisit de mettre les informations sur la trajectoire professionnelle et sur la trajectoire personnelle en tant que variables actives dans l'analyse.

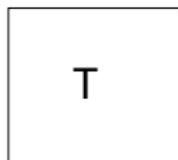
On ne tient alors pas compte de la différence entre les groupes de variables.

## Première partie I

### Deux nuages de points à plusieurs nuages...

# Comparaison de deux ACM

On réalise une ACM sur chacun des deux tableaux.



Comparaison de  
deux nuages de  
points

**Exemple**

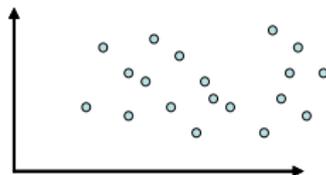
Analyse Procustéenne

Coefficients RV

Comparaison de  
plus de deux  
nuages de points

# Comparaison de deux ACM

On réalise une ACM sur chacun des deux tableaux.



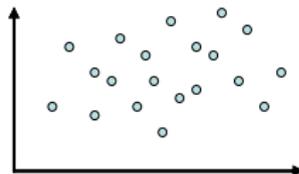
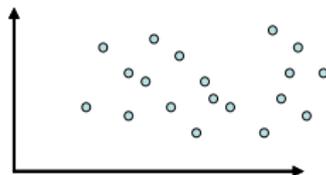
**Exemple**

Analyse Procustéenne

Coefficients RV

# Comparaison de deux ACM

On réalise une ACM sur chacun des deux tableaux.



Comparaison de  
deux nuages de  
points

**Exemple**

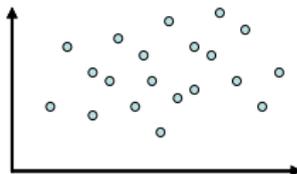
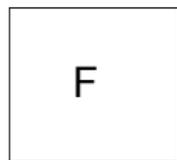
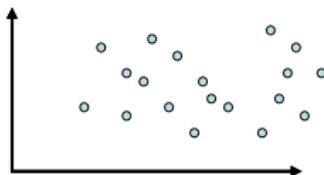
Analyse Procustéenne

Coefficients RV

Comparaison de  
plus de deux  
nuages de points

# Comparaison de deux ACM

On réalise une ACM sur chacun des deux tableaux.



Comment comparer les deux résultats ainsi obtenus ?

Comparaison de  
deux nuages de  
points

Exemple

Analyse Procustéenne

Coefficients RV

Comparaison de  
plus de deux  
nuages de points

# Analyse Procustéenne

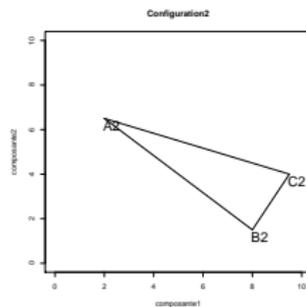
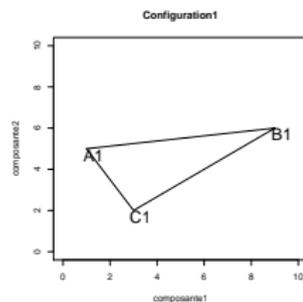
Comparaison de  
deux nuages de  
points

Exemple

**Analyse Procustéenne**

Coefficients RV

Comparaison de  
plus de deux  
nuages de points



Première idée : on superpose

# Analyse Procustéenne

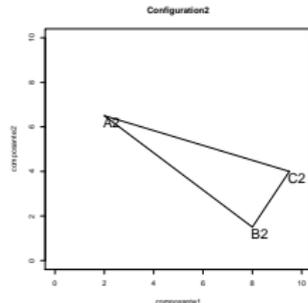
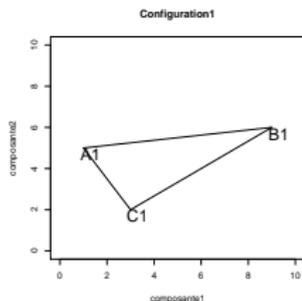
Comparaison de  
deux nuages de  
points

Exemple

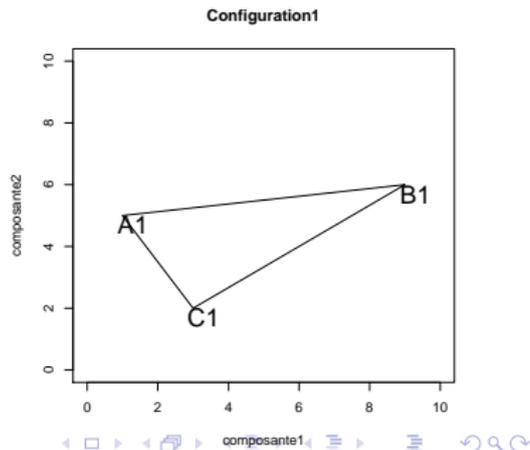
**Analyse Procustéenne**

Coefficients RV

Comparaison de  
plus de deux  
nuages de points



Première idée : on superpose



# Analyse Procustéenne

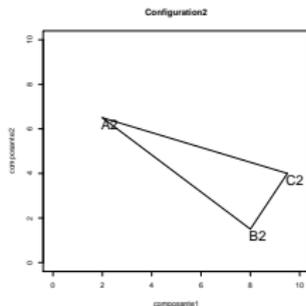
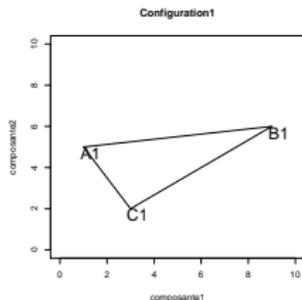
Comparaison de  
deux nuages de  
points

Exemple

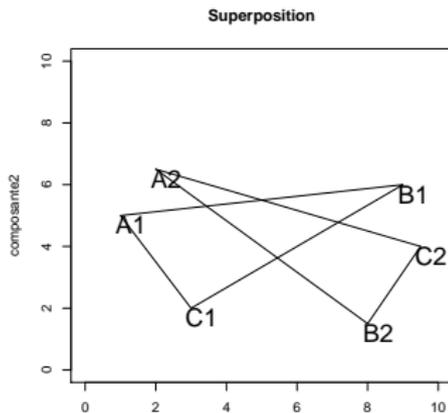
**Analyse Procustéenne**

Coefficients RV

Comparaison de  
plus de deux  
nuages de points



Première idée : on superpose



# Analyse Procustéenne

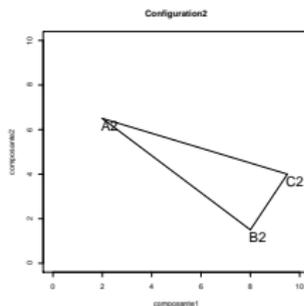
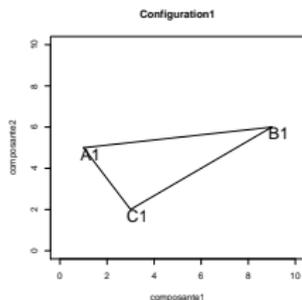
Comparaison de  
deux nuages de  
points

Exemple

**Analyse Procustéenne**

Coefficients RV

Comparaison de  
plus de deux  
nuages de points



Première idée : on superpose  
Deuxieme idée : on ajuste au  
mieux à l'aide de  
“transformations simples”

# Analyse Procustéenne

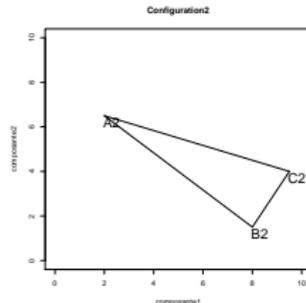
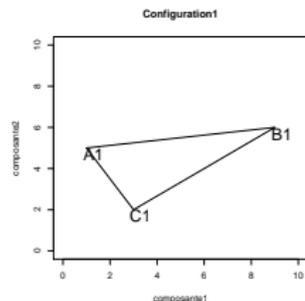
Comparaison de  
deux nuages de  
points

Exemple

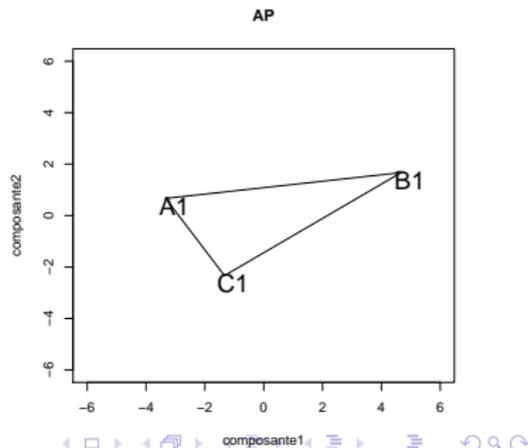
**Analyse Procustéenne**

Coefficients RV

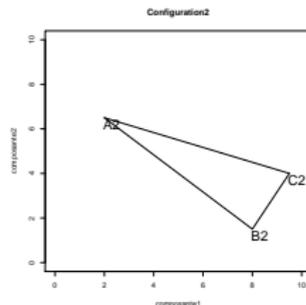
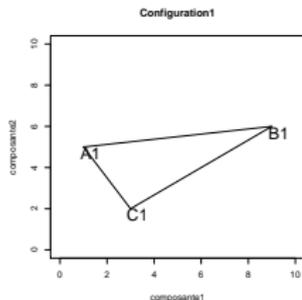
Comparaison de  
plus de deux  
nuages de points



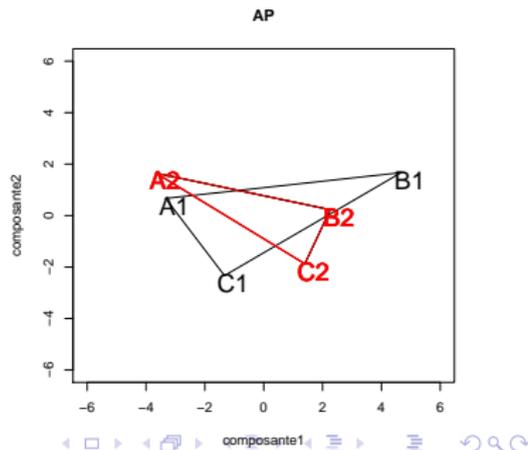
Première idée : on superpose  
Deuxième idée : on ajuste au  
mieux à l'aide de  
"transformations simples"



# Analyse Procustéenne



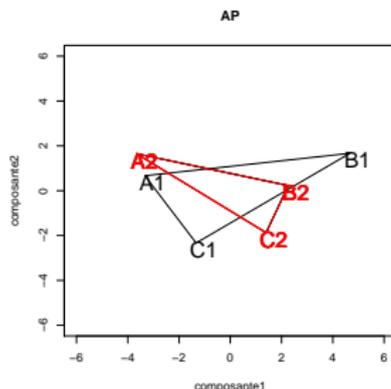
Première idée : on superpose  
Deuxième idée : on ajuste au  
mieux à l'aide de  
"transformations simples"



# Indice de similarité de Procrustes

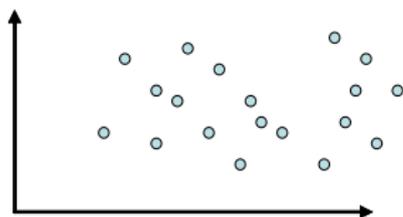
L'indice de similarité de procruste (Peron, 2000, Qannari et al., 1999) entre deux configurations  $\mathbf{X}_1$  et  $\mathbf{X}_2$ ,  $S(\mathbf{X}_1, \mathbf{X}_2)$ , vaut :

$$S(\mathbf{X}_1, \mathbf{X}_2) = \frac{\text{Tr } \mathbf{X}_1 \mathbf{H}_1 \mathbf{X}_2'}{\sqrt{\text{Tr } (\mathbf{X}_1 \mathbf{X}_1') \text{Tr } (\mathbf{X}_2 \mathbf{X}_2')}}}$$



Un indicateur classique pour comparer deux configurations entre elles est le coefficient RV (voir Escoufier, 1973 et Robert et Escoufier, 1976).

$$RV(\mathbf{X}_1; \mathbf{X}_2) = \frac{\text{Tr}(\mathbf{X}_1 \mathbf{X}'_1 \mathbf{X}_2 \mathbf{X}'_2)}{\sqrt{\text{Tr}(\mathbf{X}_1 \mathbf{X}'_1 \mathbf{X}_1 \mathbf{X}'_1) \text{Tr}(\mathbf{X}_2 \mathbf{X}'_2 \mathbf{X}_2 \mathbf{X}'_2)}}$$



# Analyse Procustéenne généralisée (APG)

Gower (1975)

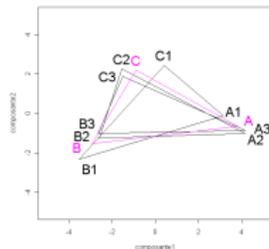
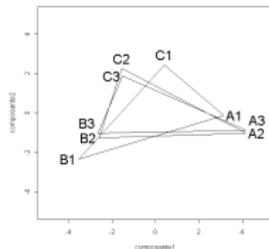
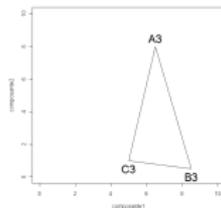
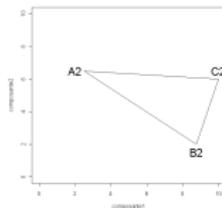
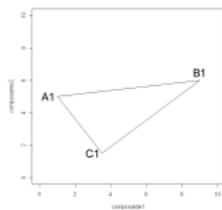
Comparaison de  
deux nuages de  
points

Comparaison de  
plus de deux  
nuages de points

Analyse Procustéenne  
généralisée (APG)

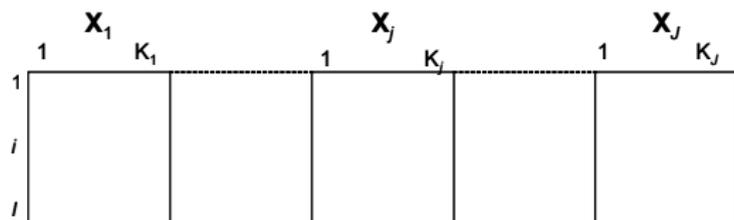
Le problème

Quelques méthodes



# Tableaux Multiples

Les données sont constituées d'un ensemble d'individus,  $\{i; 1, \dots, I\}$ , décrits par plusieurs groupes de variables. Ces données peuvent être regroupées sous la forme d'un tableau structuré en plusieurs sous-tableaux

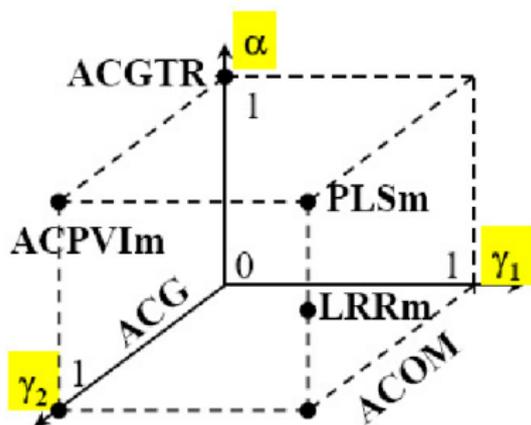


- ▶  $J$  le nombre de sous-tableaux ;
- ▶  $K_j$  le nombre de variables du groupe  $j$  ;
- ▶  $\mathbf{X}_j$  le tableau associé au groupe  $j$
- ▶  $\mathbf{X}$  le tableau complet  $\mathbf{X} = (\mathbf{X}_1 | \dots | \mathbf{X}_J)$  ;

- ▶ Les méthodes K tableaux : STATIS, AFM, Analyse Canonique Généralisée, Analyse de co-inertie
- ▶ Les méthodes K+1 tableaux

# K+1 tableaux

D'après Thèse de Stéphanie Bougeard 2007 :  
<http://tel.archives-ouvertes.fr/tel-00267595/en/>



## Interprétation des paramètres

- ▶  $\alpha$  : influence des variables  $Y$  dans la construction des composantes
- ▶  $\gamma_1$  : prise en compte de la multicolinéarité des variables  $X_k$
- ▶  $\gamma_2$  : prise en compte de la multicolinéarité des variables  $Y$

## Deuxième partie II

### AFM à partir d'un exemple

- ▶ Enquête 300 consommateurs de thé

- ▶ Enquête 300 consommateurs de thé
- ▶ 19 questions sur leur façon de consommer

- ▶ Enquête 300 consommateurs de thé
- ▶ 19 questions sur leur façon de consommer
- ▶ 12 questions sur l'image du produit

- ▶ Enquête 300 consommateurs de thé
- ▶ 19 questions sur leur façon de consommer
- ▶ 12 questions sur l'image du produit
- ▶ 4 questions de "signalétique"

# Les variables consommation (19)

- ▶ Moment de dégustation du thé en 6 questions binaires
- ▶ Lieu de consommation du thé (6 questions binaires)
- ▶ Fréquence de consommation
- ▶ Type de thé consommé
- ▶ Forme du thé
- ▶ “Comment consommez-vous du thé le plus souvent ?” :  
pur ; avec du citron ; avec du lait ; autre
- ▶ Lieu d'achat
- ▶ “Sucrez-vous votre thé ?”
- ▶ Variété de thé

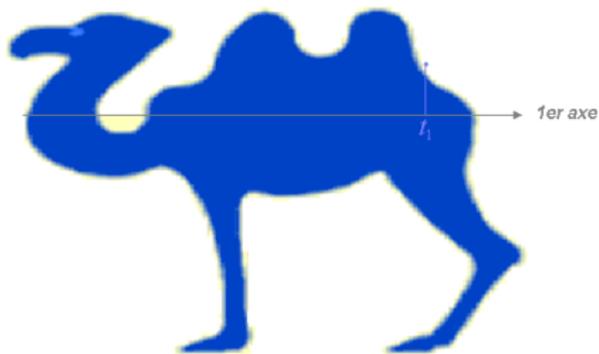
# Les variables sur l'image (12)

- ▶ “associez vous le thé à la spiritualité ?”
- ▶ “le thé est-il bon pour la santé ?”
- ▶ “Associez vous le thé à l'évasion ou l'exotisme ?”
- ▶ ...

- ▶ CSP
- ▶ Sexe
- ▶ Pratique régulier d'une activité sportive (oui/non)
- ▶ age

- ▶ Analyse Factorielle Multiple (AFM) (Escofier et Pagès,1998)
- ▶ 2 groupes actifs : Perception et consommation
- ▶ un groupe illustratif ( variables socio démographiques)

*On cherche les  
principales dimensions  
de variabilités*



- ▶ Etude des lignes (individus)
- ▶ Etude des colonnes (variables)
- ▶ Etude des groupes

# L'analyse factorielle multiple (AFM)

- ▶ L'analyse factorielle multiple (AFM ; Escofier & Pagès 1988-1998 ; Pagès 2002) s'utilise sur des données constituées d'individus décrits par 2 ou plusieurs groupes de variables
- ▶ Le coeur de l'analyse consiste à réaliser une ACP pondérée du tableau complet ( $X$ )
- ▶ Chaque variable du groupe  $j$  est pondérée par  $\lambda_{1j}$  (la première valeur propre de l'analyse du groupe  $j$ ).

# Réaliser l'analyse

```
> the<- read.table("donnees/the2.csv", header=TRUE,
+ sep=";", na.strings="NA", dec=".", strip.white=TRUE)
> the.MFA<-the[, c("ptt.dej", "gouter", "soiree",
+ "apres.dejeuner", "apres.diner", "tt.moment",
+ "maison", "travail", "salon.de.t", "amis", "resto",
+ "bar", "variete", "comment", "sucre", "forme", "lieuachat",
+ "type", "frequence", "evasion.exotisme", "spiritualite",
+ "bon.pr.la.sante", "diuretique", "convivialite",
+ "absorption.fer", "feminin", "raffine", "amaigrissant",
+ "excitant", "relaxant", "ss.effet.sante", "sexe", "CSP", "Sport")]
> # realisation d'une analyse factorielle multiple
> #avec deux groupes actifs
> #et un groupe de variable illustrative
>
>
> library(FactoMineR)
> res<-MFA(the.MFA, group=c(19, 12, 3),
+ type=c("n", "n", "n"), ncp=5,
+ name.group=c("consommation", "perception", "quali"),
+ num.group.sup=c(3),
+ graph=FALSE)
```

Données

Résultats

Cas des groupes

Comparaison des groupes

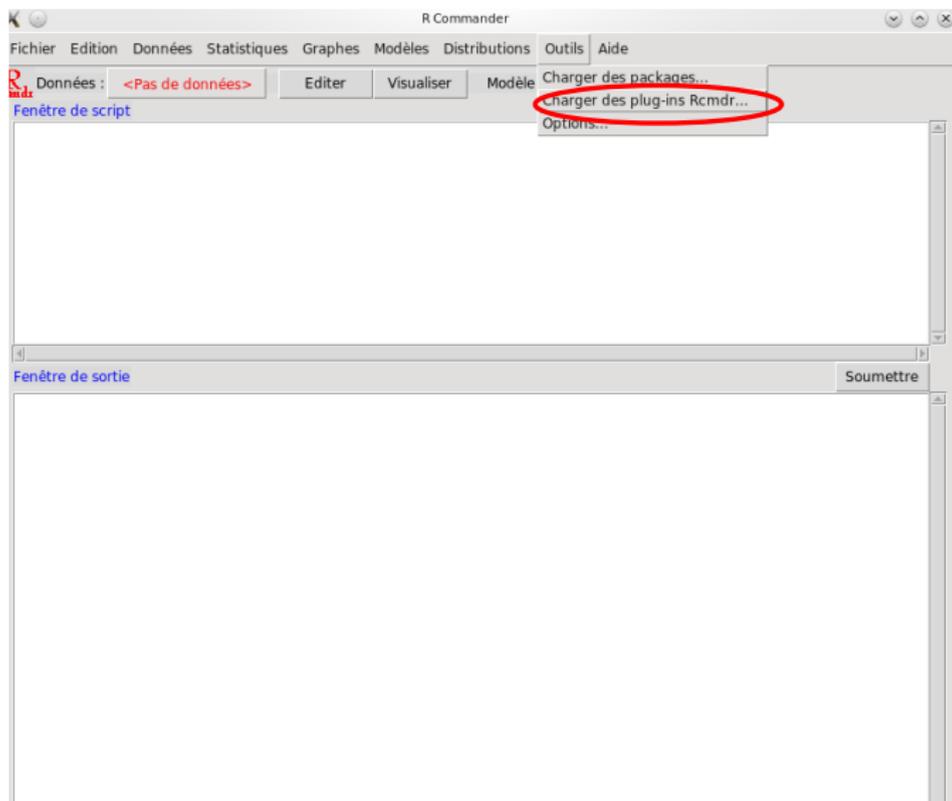
Modalités

Conclusion

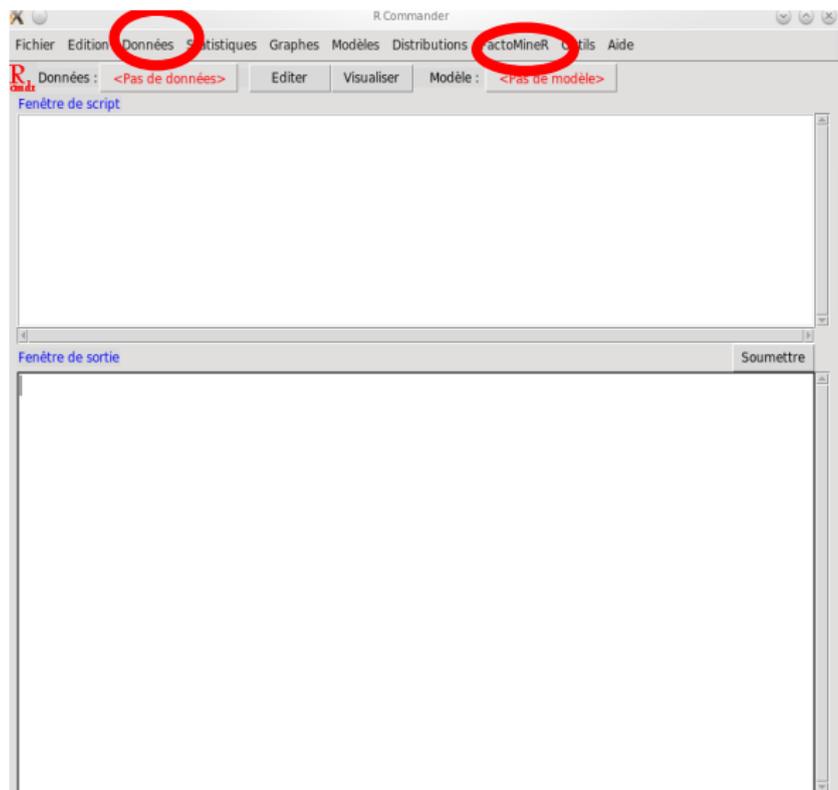
Références

# Réaliser l'analyse à l'aide de l'interface graphique

- > `install.packages("RcmdrPlugin.FactoMineR")`
- > `library(Rcmdr)`



# Réaliser l'analyse à l'aide de l'interface graphique(2)



Données

Résultats

- Cas des groupes
- Comparaison des groupes
- Modalités

Conclusion

Références

L'importation de données est réalisée à l'aide du menu

# Réaliser l'analyse à l'aide de l'interface graphique(3)

MFA

**Multiple Factor Analysis (MFA)**

**Qualitative groups**

Add quanti group !    Modify 1 group    Delete

**Qualitative groups**

Add quali group !    Modify 1 group    Delete

Select supplementary individuals    Graphical options

Outputs    Restart

**Main options**

Name of the result object:    res

Number of dimensions:    5

Select the dimensions for the graphs:    1    2

Perform Clustering after MFA

Apply

OK    Annuler    Aide

> res

**\*\*Results of the Multiple Factor Analysis (MFA)\*\***

The analysis was performed on 300 individuals, described by

\*Results are available in the following objects :

|    | name                  | description                           |
|----|-----------------------|---------------------------------------|
| 1  | "\$eig"               | "eigenvalues"                         |
| 2  | "\$separate.analyses" | "separate analyses for each group of  |
| 3  | "\$group"             | "results for all the groups"          |
| 4  | "\$partial.axes"      | "results for the partial axes"        |
| 5  | "\$inertia.ratio"     | "inertia ratio"                       |
| 6  | "\$ind"               | "results for the individuals"         |
| 7  | "\$quali.var"         | "results for the categorical variable |
| 8  | "\$quali.var.sup"     | "results for the categorical suppleme |
| 9  | "\$summary.quali"     | "summary for the categorical variable |
| 10 | "\$global.pca"        | "results for the global PCA"          |

Données

Résultats

Cas des groupes

Comparaison des groupes

Modalités

Conclusion

Références

## Coefficient RV et Lg

Coefficient Lg :

Pour un groupe donnée  $(Lg(X_1, X_1))=$  la somme des carrés des valeurs propres du groupe (indicateur de multidimensionalité)

- ▶ Entre deux groupes : somme pondérée des carrés des covariances entre chaque colonne des deux groupes.
- ▶ Lg est d'autant plus élevé que les deux groupes comparés ont des directions d'inertie importantes communes.

|              | consommation | perception |
|--------------|--------------|------------|
| consommation | 4.89         | 0.37       |
| perception   | 0.37         | 4.64       |

Coefficient RV :

|              | consommation | perception |
|--------------|--------------|------------|
| consommation | 1            | 0.078      |
| perception   | 0.078        | 1          |

# Interet de la représentation simultanée

Raport inertie inter/inertie totale

```
> res$inertia.ratio
```

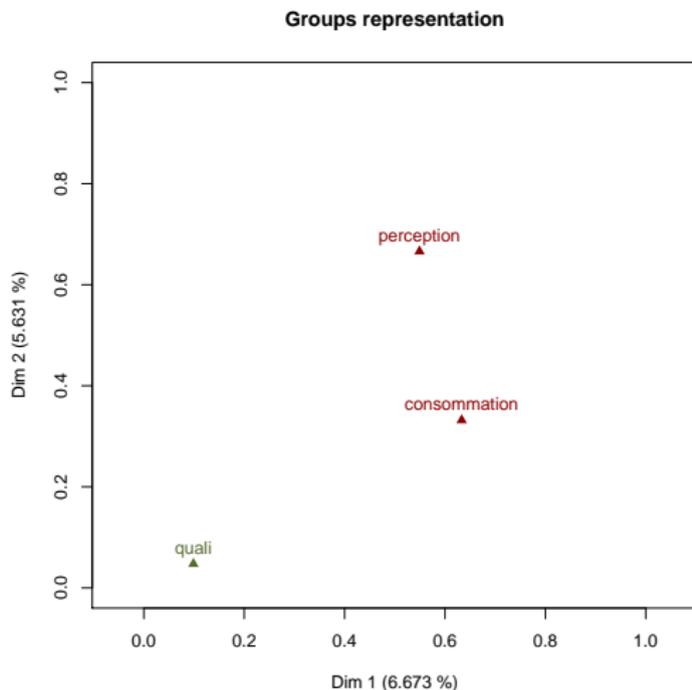
| res.inertia.ratio |      |
|-------------------|------|
| Dim.1             | 0.62 |
| Dim.2             | 0.57 |
| Dim.3             | 0.60 |
| Dim.4             | 0.60 |
| Dim.5             | 0.55 |

```
> res$group$correlation
```

|              | Dim.1 | Dim.2 | Dim.3 | Dim.4 | Dim.5 |
|--------------|-------|-------|-------|-------|-------|
| consommation | 0.63  | 0.33  | 0.26  | 0.54  | 0.30  |
| perception   | 0.55  | 0.67  | 0.68  | 0.30  | 0.53  |

# Graphique des groupes

```
> plot(res, choix = "group")
```



# Aide à l'interprétation sur les groupes

```
> res$group$coord
```

|              | Dim.1 | Dim.2 | Dim.3 | Dim.4 | Dim.5 |
|--------------|-------|-------|-------|-------|-------|
| consommation | 0.63  | 0.33  | 0.26  | 0.54  | 0.30  |
| perception   | 0.55  | 0.67  | 0.68  | 0.30  | 0.53  |

```
> res$group$contrib
```

|              | Dim.1 | Dim.2 | Dim.3 | Dim.4 | Dim.5 |
|--------------|-------|-------|-------|-------|-------|
| consommation | 53.55 | 33.26 | 27.93 | 63.96 | 36.59 |
| perception   | 46.45 | 66.74 | 72.07 | 36.04 | 63.41 |

- ▶ 0.63 (coordonnée du groupe consommation sur l'axe 1) est l'inertie de l'ensemble des variables du groupe consommation sur l'axe 1
- ▶ Contribution du groupe au facteur.

Données

Résultats

Cas des groupes

Comparaison des groupes

Modalités

Conclusion

Références





# Pour aller plus loin dans les graphiques

Le package `dyngraph` voir par exemple

`http://www.quantit.ihmc.ens.fr/IMG/pdf/Tutoriel\_FactomineR\_ACM.pdf`

# Pourquoi utiliser l'AFM

- ▶ Analyse de données à la française, lecture "facile" puisque identique à l'ACP, ACM par exemple
- ▶ Beaucoup d'aide à l'interprétation
- ▶ Mise en oeuvre adaptée pour les données qualitatives
- ▶ Calcul exact, pas d'algorithme

# Quelques extensions de l'AFM

- ▶ Analyse Factorielle Multiple Hierarchique  
Le Dien et Pagès (2003)



Escofier, B. et Pagès, J. (1998). *Analyses factorielles simples et multiples*. Dunod.

Escoufier, Y. (1973). Le traitement des variables vectorielles. *Biometrics*, 29, p. 751–760.

Gower, J. C. (1975). Generalized procrustes analysis. *Psychometrika*, 40(1), p. 33–51.

Le Dien, S. et Pagès, J. (2003). Hierarchical multiple factor analysis : application to the comparison of sensory profiles. *Food Quality and Preferences*, 18(6), p. 453–464.

Pagès, J. (2004). Analyse factorielle de données mixtes. *Revue de Statistique Appliquée*, 52(4), p. 93–111.

Peron, L. (2000). Statistical analysis of sensory profiling data : data reduction and generalised procustes analysis. *Food quality and preference*, 11, p. 155–157.

- Qannari, E., Macfie, H., et Courcoux, P. (1999).  
Performance indices and isotropic scaling factors in  
sensory profiling. *Food Quality and Preference*, 10, p.  
17–21.
- Robert, P. et Escoufier, Y. (1976). A Unifying Tool for  
Linear Multivariate Statistical Methods : the  
RV-Coefficient. *Applied Statistics*, 29(3), p. 257–265.