**Application : AFMD**

**Format des données**

Nous utiliserons le jeu de données wine dans le package FactoMineR:

**library**("FactoMineR")

data(wine)

df <- wine[,c(1,2, 16, 22, 29, 28, 30,31)]

Pour voir la structure des données, tapez ceci:

str(df)

Les données contiennent 21 lignes (individus) et 8 colonnes (variables):

* Les deux premières colonnes sont des variables catégorielles: label (Saumur, Bourgueil ou Chinon) et soil (Reference, Env1, Env2 or Env4).
* Les colonnes restantes sont numériques (variables continues).

L’objectif de cette étude est d’analyser les caractéristiques des vins.

**code R**

La fonction FAMD() [package FactoMineR] peut être utilisée pour le calcul. Un format simplifié est:

FAMD (base, ncp = 5, sup.var = NULL, ind = NULL, graph = TRUE)

* base: un data frame avec n lignes (individus) et p colonnes (variables).
* ncp: le nombre de dimensions conservées dans les résultats (par défaut 5)
* sup.var: un vecteur indiquant les positions des variables supplémentaires.
* ind: un vecteur indiquant les positions des individus supplémentaires.
* graphique: une valeur logique. Si TRUE le graphique est affiché.

Pour calculer l’AFDM, tapez ceci:

**library**(FactoMineR)

res.famd <- FAMD(df, graph = FALSE)

Le résultat de la fonction FAMD() est une liste comprenant:

print(res.famd)

**Visualisation et interprétation**

Nous utiliserons les fonctions suivantes du package factoextra:

* get\_eigenvalue(res.famd): Extraction des valeurs propres / variances des composantes principales.
* fviz\_eig(res.famd): Visualisation des valeurs propres.
* get\_famd\_ind(res.famd): Extraction des résultats pour les individus.
* get\_famd\_var(res.famd): Extraction des résultats pour les variables quantitatives et qualitatives.
* fviz\_famd\_ind (res.famd), fviz\_famd\_var(res.famd): Visualisation des résultats pour les individus et les variables, respectivement.

Dans les sections suivantes, nous allons illustrer chacune de ces fonctions.

Pour aider à l’interprétation de l’AFDM, nous recommandons fortement de lire l’interprétation de l’analyse en composantes principales et de l’analyse des correspondances multiples.

**Valeurs propres / Variances**

La proportion de variances expliquées par les différentes dimensions (axes) peut être extraite à l’aide de la fonction get\_eigenvalue() [factoextra] comme suit:

**library**("factoextra")

eig.val <- get\_eigenvalue(res.famd)

La fonction fviz\_eig() ou fviz\_screeplot() [factoextra] peut être utilisée pour visualiser les proportions de variances expliquées par les différents axes:

fviz\_screeplot(res.famd)

**Graphique des variables**

**Toutes les variables**

La fonction get\_mfa\_var() [factoextra] est utilisée pour extraire les résultats pour les variables. Par défaut, cette fonction renvoie une liste contenant les coordonnées, les cos2 (cosinus carré) et les contribution de toutes les variables:

var <- get\_famd\_var (res.famd)

var

## FAMD results for variables

## ===================================================

## Name Description

## 1 "$coord" "Coordinates"

## 2 "$cos2" "Cos2, quality of representation"

## 3 "$contrib" "Contributions"

Les différents composants peuvent être consultés comme suit:

# Coordonnées des variables

**head**(var$coord)

# Cos2: qualité de représentation

**head**(var$cos2)

# Contributions aux dimensions

**head**(var$contrib)

La figure suivante montre la corrélation entre les variables - variables quantitatives et qualitatives - et les axes principaux, ainsi que la contribution des variables aux axes 1 et 2. Les fonctions suivantes [dans le package factoextra] sont utilisées:

* fviz\_famd\_var() pour visualiser les variables quantitatives et qualitatives
* fviz\_contrib() pour visualiser la contribution des variables aux axes principaux

# Graphique des variables

fviz\_famd\_var (res.famd, repel = TRUE)

# Contribution à la première dimension

fviz\_contrib (res.famd, "var", axes = 1)

# Contribution à la deuxième dimension

fviz\_contrib (res.famd, "var", axes = 2)

La ligne en pointillé rouge sur le graphique ci-dessus indique la valeur moyenne attendue, si les contributions étaient uniformes. Lire plus dans le chapitre (Chapitre @ref(analyse-en-composantes-principales)).

**A partir des graphiques ci-dessus, on constate t-on ?**

**Variables quantitatives**

Pour extraire les résultats pour les variables quantitatives, tapez ceci:

quanti.var <- get\_famd\_var(res.famd, "quanti.var")

quanti.var

## FAMD results for quantitative variables

## ===================================================

## Name Description

## 1 "$coord" "Coordinates"

## 2 "$cos2" "Cos2, quality of representation"

## 3 "$contrib" "Contributions"

Dans cette section, nous décrirons comment visualiser les variables quantitatives. De plus, nous allons montrer comment mettre en évidence les variables selon i) leurs qualités de représentation ou ii) leurs contributions aux dimensions.

Le code R ci-dessous présente les variables quantitatives. Nous utilisons repel = TRUE, pour éviter le chevauchement de texte.

fviz\_famd\_var(res.famd, "quanti.var", repel = TRUE,

col.var = "black")

En bref, le graphique des variables (cercle de corrélation) montre le lien entre les variables, la qualité de la représentation des variables, ainsi que la corrélation entre les variables et les dimensions.

Les variables quantitatives les plus contributives peuvent être mises en évidence sur le graphique en utilisant l’argument col.var = “contrib”. Cela produit un gradient de couleurs, qui peut être personnalisé à l’aide de l’argument gradient.cols.

fviz\_famd\_var(res.famd, "quanti.var", col.var = "contrib",

gradient.cols = c("#00AFBB", "#E7B800", "#FC4E07"),

repel = TRUE)

De même, vous pouvez mettre en évidence des variables quantitatives en utilisant leurs cos2 représentant la qualité de représentation. Si une variable est bien représentée par deux dimensions, la somme des cos2 est proche de 1. Pour certains des éléments, plus de 2 dimensions pourraient être nécessaires pour représenter parfaitement les données.

# Couleur par valeurs cos2: qualité sur le plan des facteurs

fviz\_famd\_var(res.famd, "quanti.var", col.var = "cos2",

gradient.cols = c("#00AFBB", "#E7B800", "#FC4E07"),

repel = TRUE)

**Graphique des variables qualitatives**

Comme les variables quantitatives, les résultats pour les variables qualitatives peuvent être extraits comme suit:

quali.var <- get\_famd\_var(res.famd, "quali.var")

quali.var

## FAMD results for qualitative variable categories

## ===================================================

## Name Description

## 1 "$coord" "Coordinates"

## 2 "$cos2" "Cos2, quality of representation"

## 3 "$contrib" "Contributions"

Pour visualiser les variables qualitatives, tapez ceci:

fviz\_famd\_var(res.famd, "quali.var", col.var = "contrib", gradient.cols = c("#00AFBB", "#E7B800", "#FC4E07"))

Le graphique ci-dessus montre les catégories des variables catégorielles.

**Graphique des individus**

Pour obtenir les résultats pour les individus, tapez ceci:

ind <- get\_famd\_ind(res.famd)

ind

## FAMD results for individuals

## ===================================================

## Name Description

## 1 "$coord" "Coordinates"

## 2 "$cos2" "Cos2, quality of representation"

## 3 "$contrib" "Contributions"

Pour visualiser les individus, utilisez la fonction fviz\_famd\_ind() [factoextra]. Par défaut, les individus sont colorés en bleu. Cependant, comme des variables, il est également possible de colorer les individus par leurs cos2 et leurs valeurs de contribution:

fviz\_famd\_ind(res.famd, col.ind = "cos2",

gradient.cols = c("#00AFBB", "#E7B800", "#FC4E07"),

repel = TRUE)

**Interpréter la sortie graphique**

Notez qu’il est possible de colorer les individus en utilisant l’une des variables qualitatives dans le tableau de données initial. Pour ce faire, l’argument habillage est utilisé dans la fonction fviz\_famd\_ind(). Par exemple, si vous souhaitez colorer les vins selon la variable qualitative supplémentaire “Label”, tapez ceci:

fviz\_mfa\_ind(res.famd,

habillage = "Label", # color by groups

palette = c("#00AFBB", "#E7B800", "#FC4E07"),

addEllipses = TRUE, ellipse.type = "confidence",

repel = TRUE # Avoid text overlapping

)

Si vous souhaitez colorer les individus à l’aide de plusieurs variables catégorielles en même temps, utilisez la fonction fviz\_ellipses() [factoextra] comme suit:

fviz\_ellipses(res.famd, c("Label", "Soil"), repel = TRUE)

Alternativement, vous pouvez spécifier les positions des variables catégorielles:

fviz\_ellipses(res.famd, 1:2, geom = "point")

**Résumé**

L’analyse factorielle des données mixtes (AFDM) permet d’analyser un jeu de données, dans lequel les individus sont décrits à la fois par des variables qualitatives et quantitatives.