

TP: Analyse Factorielle Discriminante

Analyse des données

Master ISEFAR - M1

1 Packages R utiles

```
library("tidyverse")
library("FactoMineR") #pour effectuer l'ACP
library("factoextra")
library("ade4") # pour effectuer l'AFD
```

2 Script pour réaliser l'AFD sur le jeu de données insectes

```
# réalise l'AFD
X <- insectes %>% dplyr::select_if(is.numeric)
y <- insectes$type
insectes.dis = discrimin(dudi.pca(X, scan = FALSE),y, scan = FALSE)

# Pour les deux premiers axes
# Coordonnées des individus projetés
insectes.dis$li

# Vecteurs propres
insectes.dis$fa

# Valeurs propres
insectes.dis$eig

# Corrélation entre variables discriminantes et variables d'origines
insectes.dis$va

# Graphe des individus
coord <- insectes.dis$li %>% mutate(type=y)
ggplot(coord,aes(x=DS1,y=DS2,label =type))+
  geom_text(aes(color=type))+
  geom_hline(yintercept=0)+geom_vline(xintercept = 0)

# Graphe des variables
s.corcircle(insectes.dis$va)

# Coordonnées des centres de gravité projetés
insectes.dis$gc
```

3 Insectes

On reprend ici l'exemple du cours sur les insectes. Pour importer le jeu de données, faire

```
load("insectes.rda")
head(insectes)
```

Réaliser quelques analyses descriptives, une ACP et une AFD.

4 Les iris

On reprend le jeu de données classique des iris (issu du package **MASS**) et vu dans le TP sur l'analyse descriptive univariée et bivariée. On rappelle que dans ce jeu de données, on a 150 iris de 3 espèces différentes (variable **Species**) sur lesquels on mesure la longueur du sépale (variable **Sepal.Length**), la largeur du sépale (variable **Sepal.Width**), la longueur du pétale (variable **Sepal.Length**) et la largeur du pétale (variable **Petal.Width**).

Pour importer le jeu de données, faire

```
data(iris)
```

Réaliser quelques analyses descriptives, une ACP et une AFD.

5 Banque

On cherche à étudier les comportements bancaires de 500 clients en fonction de leur catégorie socio-professionnelle. Le jeu de données est **Banque.csv**. Les variables sont

- solde : Solde moyen du compte courant sur les 12 derniers mois (en euros)
- mdecouv : Montant cumulé des découverts sur le compte courant durant les 12 derniers mois (en euros)
- ncompte : Nombre de comptes utilisés en plus du compte courant (par exemple les livrets ...)
- emprunt : Montant total des emprunts effectués sur les trois dernières années (en euros)
- mdepot : Montant total des versements effectués sur le livret d'épargne lors des 5 dernières années (en euro)
- mretrait : montant total des retraits effectués sur le livret d'épargne sur les 12 derniers mois (en euros)
- nbenf : nombre d'enfants de moins de 18 ans
- age : age du client enquêté
- csp : catégorie socio-professionnelle du client: artisan-commerçant; cadre; employé; ouvrier; retraité; autre
- code : codification de la catégorie socio-professionnelle du client

Pour importer le jeu de données, faire

```
banque = read.table("Banque.csv", header = TRUE, dec = ",", sep=";")
```

Réaliser quelques analyses descriptives, une ACP et une AFD.