

TP5 : Analyse des Correspondances Multiples avec R

1 Exemple d'une ACM avec le logiciel R

On considère le tableau de données brutes (contenu dans le fichier chiens.rda):

	taille	poids	velocite	intellig	affect	agress
beauceron	T++	P+	V++	I+	Af+	Ag+
basset	T-	P-	V-	I-	Af-	Ag+
ber_alle	T++	P+	V++	I++	Af+	Ag+
boxer	T+	P+	V+	I+	Af+	Ag+
bull-dog	T-	P-	V-	I+	Af+	Ag-
bull-mass	T++	P++	V-	I++	Af-	Ag+
caniche	T-	P-	V+	I++	Af+	Ag-
chihuahua	T-	P-	V-	I-	Af+	Ag-
cocker	T+	P-	V-	I+	Af+	Ag+
colley	T++	P+	V++	I+	Af+	Ag-
dalmatien	T+	P+	V+	I+	Af+	Ag-
dobermann	T++	P+	V++	I++	Af-	Ag+
dogue_all	T++	P++	V++	I-	Af-	Ag+
epagn_bre	T+	P+	V+	I++	Af+	Ag-
epagn_fra	T++	P+	V+	I+	Af-	Ag-
fox_hound	T++	P+	V++	I-	Af-	Ag+
fox_terri	T-	P-	V+	I+	Af+	Ag+
grand_ble	T++	P+	V+	I-	Af-	Ag+
labrador	T+	P+	V+	I+	Af+	Ag-
levrier	T++	P+	V++	I-	Af-	Ag-
mastiff	T++	P++	V-	I-	Af-	Ag+
pekinois	T-	P-	V-	I-	Af+	Ag-
pointer	T++	P+	V++	I++	Af-	Ag-
saint_ber	T++	P++	V-	I+	Af-	Ag+
setter	T++	P+	V++	I+	Af-	Ag-
teckel	T-	P-	V-	I+	Af+	Ag-
terre_neu	T++	P++	V-	I+	Af-	Ag-

On désire faire une ACM de ce petit jeu de données avec le logiciel R. Pour cela, on va utiliser la fonction MCA (“Multiple Correspondance Analysis”) du package **FactoMineR**. Les lignes de commandes suivantes permettent de réaliser cette ACM et d’obtenir les sorties numériques et graphiques nécessaires à l’interprétation.

```

#-----
# Chargement du jeu de donnees
#-----
load("../data/chiens.rda")
chiens

#-----
# Chargement du package FactoMineR
#-----
library(FactoMineR) # permet de charger le package "FactoMineR"
                    # afin de pouvoir l'utiliser par la suite
help(MCA)

#-----
# Mise en oeuvre de l'ACM
#-----
res <- MCA(chiens) # tous les calculs de l'ACM sont stockes dans l'objet "res"
                  # NB : par default les graphiques des plans factoriels 1-2
                  # sont affiches a l'ecran
res <- MCA(chiens, graph=FALSE) # idem sans les graphiques
res # permet de voir l'ensemble des sorties numeriques disponibles

#-----
# Choix du nombre d'axes à retenir
#-----
round(res$eig,digit=2) # permet d'afficher les valeurs propres et les pourcentages
                      # de variances expliquees par chaque axe
# Graphique de l'ebouli des valeurs propres
barplot(res$eig[,1],main="Eigenvalues", names.arg=1:nrow(res$eig))
abline(h=1/ncol(chiens),col=2,lwd=2)

#-----
# Graphiques des individus/modalités et des variables sur le plan factoriel 1-2
#-----
?plot.MCA # permet d'afficher la fenetre d'aide de la commande "plot.MCA"
plot(res,axes=c(1,2),choix="ind") # on retrouve ici le graphique de la représentation
                                  # simultanée des individus et des modalités (plan 1-2)
plot(res,axes=c(1,2),choix="ind", invisible = "var") # on retrouve ici le graphique des
                                                       # individus seuls
plot(res,axes=c(1,2),choix="ind", invisible = "ind") # on retrouve ici le graphique des
                                                       # modalités seules
plot(res,axes=c(1,2),choix="var") # on retrouve ici le graphique variables (plan 1-2)

#-----
# Sorties numeriques pour les individus, les modalités et les variables
#-----
res$ind # permet d'afficher l'ensemble des sorties numeriques associees aux individus :
        # coordonnees, contributions, cosinus carres
round(res$ind$cos2,digit=3) # uniquement les cosinus carres

res$var # permet d'afficher l'ensemble des sorties numeriques associees modalités
        # coordonnees, contributions, cosinus carres
round(res$var$cos2,digit=3) # uniquement les cosinus carres

```

```
res$var$eta2 # permet d'afficher les rapports de corrélations variables/CP
```

Travail à réaliser. Faire l'ACM en utilisant les commandes ci-dessus et commenter les différentes sorties graphiques et numériques.

2 Autre jeu de données pour faire une ACM puis du clustering

Voici ci-après le jeu de données `tea` avec lequel vous pouvez faire également une ACM. Ces données sont également disponibles par défaut dans le package `FactoMineR`. Elles concernent les réponses à un questionnaire sur le thé de $n = 300$ personnes. Seules les $p = 18$ premières variables nous intéressent ici (il s'agit de celles correspondant aux questions sur le comportement en terme de consommation de thé). Les autres variables peuvent éventuellement être mises en variables supplémentaires ou illustratives : une est quantitative (l'âge) et les autres sont qualitatives (questions signalétiques ou sur l'image du produit).

```
the <- read.table("../data/the.csv" , header=TRUE, sep=";")
```

Travail à réaliser. Faire l'ACM et commenter les différentes sorties graphiques et numériques. Faire ensuite un clustering des individus en choisissant une méthode de recodage des données en données quantitatives et une méthode de clustering. Pour rappel, vous pouvez recoder les données en prenant les composantes principales de l'ACM (toutes ou les q premières, q à choisir), ou encore en prenant les variables synthétiques d'une clustering de variables avec le package `ClustOfVar` (nombre de classes de variables à choisir). Enfin interprétez les classes des individus que vous aurez obtenu.