

## Analyse Factorielle discriminante.

Travaux pratiques (statistiques exploratoires).

Les jeux de données étudiés sont disponibles sur <http://math.univ-lille1.fr/~marbaclo/>

### 1 Premiers pas en AFD

On dispose de 8 individus décrits par 3 variables quantitatives et répartis en deux groupes. On se propose d'effectuer une AFD pour représenter au mieux les classes. Les données sont regroupées dans le tableau 1.

numero	var1	var2	var3	groupe
1	2	3	2	gr1
2	3	2	2	gr1
3	2	3	3	gr1
4	3	2	3	gr1
5	-2	-3	-3	gr2
6	-2	-2	-2	gr2
7	-3	-3	-2	gr2
8	-3	-2	-3	gr2

TABLE 1 – Données exercice 1.

1. Déterminer le nuage pesant associé aux données  $\mathcal{N}$ .
2. Déterminer le centre de gravité  $g$  associé à  $\mathcal{N}$  et le centre de gravité des classes notés  $g_1$  et  $g_2$ .
3. Déterminer la matrice de covariance totale  $V_T$ , la matrice de covariance intra-classe  $V_W$  et la matrice de covariance inter-classe  $V_B$ .
4. Démontrer la relation suivante (vue en cours)

$$V_T = V_W + V_B.$$

5. Sachant que  $(V_T)^{-1}V_B \propto M$  avec  $M = \begin{pmatrix} 2 & 2 & 2 \\ 2 & 2 & 2 \\ 1 & 1 & 1 \end{pmatrix}$ . Déterminer les vecteurs propres et valeurs propres associés à  $M$ . En déduire les axes factoriels de l'AFD.
6. Déterminer les coordonnées des centres de gravité des classes et des individus notées  $c_i^k$ .
7. Représenter les centres de gravité des classes et les individus dans le premier plan de l'AFD.

### 2 AFD dans le cas de deux classes

On considère ici un tableau  $X$  décrivant pour  $n$  individus  $p$  variables quantitatives centrées. Les  $n$  individus sont répartis dans deux classes d'effectifs respectifs  $n_1$  et  $n_2$ . L'objectif de cet exercice est de réaliser l'analyse discriminante de ce tableau.

1. Rappeler brièvement l'objectif de l'analyse discriminante.
2. Combien d'axe peut-on retenir ?
3. Exprimer les centres de gravité  $g_1$  et  $g_2$  en fonction de  $x_i$ ,  $z_i$ ,  $n_1$ ,  $n_2$ .
4. Exprimer  $V_B$  en fonction de  $n_1$ ,  $n_2$ ,  $g_1$  et  $g_2$ .

5. Montrer que  $n_1 g_1 + n_2 g_2 = 0$ . En déduire une expression de  $g_2$  en fonction de  $g_1$ .
6. Montrer que  $V_B = \frac{n_1}{n_2} g_1 g_1^T$ .
7. On considère les individus suivants :

var1	var2	classe
-6	4	1
-2	1	1
2	-2	1
-2	2	2
2	-4	2
6	-1	1

TABLE 2 – Données exercice 2.

Représenter les centres de gravité et les individus dans le plan canonique puis sur l'axe issu de l'AFD.

### 3 AFD sous SAS

1. Générer en SAS un jeu de données de 300 individus provenant de trois classes gaussiennes trivariées de la façon suivante (utiliser la fonction `rand`). Chaque classe  $k$  ( $k = 1, \dots, 3$ ) comporte 100 individus issus de manière indépendante de la loi

$$\mathcal{N}_3 \left( \begin{pmatrix} 0 \\ 0 \\ k \end{pmatrix}, \begin{pmatrix} 100 & 0 & 0 \\ 0 & 100 & 0 \\ 0 & 0 & 0.01 \end{pmatrix} \right).$$

2. Représenter la partition dans le premier plan des variables initiale.
3. Faire l'ACP du nuage de points (poids identiques, métrique identité) et représenter la partition dans le premier plan factoriel (`proc princomp`).
4. Faire l'AFD du nuage de points (poids identiques) et représenter la partition dans le premier plan factoriel (`proc candisc`).
5. Commenter la différence entre les trois projections précédentes.

### 4 ACP et AFD sous R

On dispose de mesures anatomiques (au nombre de 6) sur trois espèces d'insectes (fichier `lubisch.txt`). On souhaite représenter les classes sur un plan factoriel.

1. Représenter les individus et les centres de gravité dans le premier plan ACP.
2. Représenter les individus et les centres de gravité dans le premier plan AFD en respectant les étapes suivantes :
  - (a) Programmer les fonctions `calcul_VT` et `calcul_VB` qui prennent un nuage pesant en paramètre d'entrée et qui retournent respectivement  $V_T$  et  $V_B$ .
  - (b) Programmer la fonction `Afd` qui prend un nuage pesant en paramètre d'entrée et qui retourne les coordonnées des individus et des centres de gravité sur les axes de l'AFD.
  - (c) Programmer la fonction `voir.afd` qui représente les individus et les centres de gravité sur les axes de l'AFD en utilisant les outils de visualisation (couleur, points,...).
3. Conclure.