Master MLDS 17/18

TP3: AFD - LDA

On considère un jeu de données décrivant trois groupes d'insectes (notés A,B et C) sur lesquels six mesures anatomiques quantitatives ont été réalisées (notées X1,…,X6). On veut savoir si ces 6 mesures permettent de retrouver les groupes d'insectes. Pour cela on effectue avec le logiciel R une analyse factorielle discriminante (AFD) ainsi que des calculs de statistiques descriptives (moyenne, variance, corrélation linéaire).|

1. Combien d'axes discriminants peut-on construire avec les données dont on dispose ?

2. Préciser le nombre total d'insectes pris en compte dans l'étude, ainsi que le nombre d'insectes dans chaque groupe.

3. Quel est le centre de gravité g de l'ensemble des données ? Donner les centres de gravité gA, gB et gC de chacun des 3 groupes.

4. En considérant l'ensemble des données, quelle est la variable la plus dispersée ? la moins dispersée ?

5. Faire une ACP des données et visualiser les insectes et les variables sur le premier plan factoriel. Commentez. L'ACP vous semble-t-elle une bonne méthode pour discriminer

les trois groupes d'insectes ?

6. Faire maintenant un AFD. Au vu des pouvoirs discriminants, combien d'axes semblent nécessaires pour avoir une bonne discrimination des groupes ?

7. Visualisez les insectes sur le premier plan de l'AFD et les variables sur le cercle des corrélations correspondant. Commentez.

8. A votre avis, les données ont-elles été centrées avant de faire l'AFD ?

9. Calculer les coordonnées de la projection des centres de gravité gA, gB et gC (des données centrées) sur le premier axe discriminant. Vérifiez qu'il s'agit des moyennes de la première variable discriminante dans chaque groupe.

10. On considère que la première variable discriminante est un bon score. Calculez le score d'un nouvel insecte pour lesquels X = (193; 131; 55; 160; 16; 102) sur cette première variable discriminante.

11. Proposez des seuils pour construire une règle de décision. A votre avis, quel sera le taux d'erreur avec cette règle de décision sur les données ? A quelle classe appartient le nouvel insecte de la question précédente avec cette règle de décision ?

12. Construire le score et la règle de décision de l'AFD sur un échantillon aléatoire de 50 insectes (formant un échantillon d'apprentissage). Calculer les scores des 24 autres insectes (l'échantillon test) et classer ces 24 insectes avec la règle de décision construite sur l'ensemble d'apprentissage. Quel est le taux d'erreur ?