

## **TD3 Statistiques multivariées et Programmation R**

### **Méthodes exploratoires, synthétiques et classification**

- 1) Chargez les données des IRIS du Val-de-Marne dans R, puis identifiez les variables quantitatives.
- 2) Après avoir chargé la bibliothèque « FactoMineR », utilisez la fonction « PCA » pour réaliser une Analyse en Composantes Principales à 2 dimensions, puis affichez les résultats graphiques obtenus pour les variables (cercle des corrélations) et les individus.
- 3) Récupérez dans différentes variables les coordonnées des individus, les corrélations des variables avec les deux composantes principales et enfin les pourcentages de variance (d'inertie) expliqués par les différentes composantes principales. Affichez un « barplot » des pourcentages de variance (d'inertie) expliqués par les différentes composantes principales.
- 4) A l'aide de la fonction « HCPC », réalisez une classification ascendante hiérarchique des individus en 4 groupes sur les deux premières composantes principales. Affichez le résultat graphique en définissant l'argument « choice » égal à « map ».
- 5) Utilisez les groupes créés pour réaliser une cartographie des IRIS du Val-de-Marne. Etudiez les propriétés des classes créées.
- 6) Réalisez une Analyse en Composantes Principales à 4 dimensions. Affichez les individus sur un graphique reposant sur les troisième et quatrième composantes principales. Faites de même pour les variables et ainsi obtenir le cercle des corrélations correspondant.
- 7) A l'aide de la fonction « pairs », représentez l'ensemble des nuages de points des variables 12 à 17 des IRIS du Val-de-Marne.
- 8) A l'aide de la fonction « dist », calculez les dissimilarités entre les IRIS du Val-de-Marne pour les variables 13 à 17 par des distances euclidiennes.
- 9) Réalisez une Classification Ascendante Hiérarchique à partir des dissimilarités calculées plus haut en utilisant la fonction « hclust ». Pour cela, utilisez la méthode « ward.D2 ». Affichez le dendrogramme.
- 10) Calculez l'inertie expliquée par les différentes partitions en cumulant les valeurs d'inertie à l'aide de la fonction « cumsum ». Au préalable, trie par ordre décroissant les valeurs d'inertie à l'aide de la fonction « sort ». Affichez ces valeurs dans un « barplot ».
- 11) En utilisant la fonction « cutree », créez une partition en 4 groupes à partir de la CAH précédente, puis affichez l'inertie expliquée par cette partition.
- 12) Récupérez et affichez le nom de la 13<sup>ème</sup> variable (il est possible d'utiliser la fonction « colnames » pour cela), puis affichez le rapport entre les moyennes des groupes (en utilisant par exemple la fonction « tapply ») et la moyenne totale.
- 13) En vous fondant sur la question 12 et en utilisant une boucle « for » (décrite dans le support de cours), affichez les caractéristiques de chaque groupe de la CAH produite.
- 14) Réalisez une cartographie des IRIS du Val-de-Marne à partir de la CAH produite. On représentera le premier groupe en bleu, le deuxième groupe en rouge, le troisième groupe en orange et le quatrième groupe en jaune. Pour cela, on pourra créer un vecteur « couleurs » à l'aide de la fonction « vector ».