

## TP5 Traitement numérique de données : CAH Ward

1. Chargez le jeu de données : Papillons.
2. Visualisez les données Papillons à l'aide d'une ACP puis d'une AFC, en utilisant respectivement les fonctions « *PCA()* » et « *CA()* » disponibles dans le package *FactoMineR*.
3. Appliquez une CAH en utilisant le critère de Ward sur :
  - Les données Papillons originales.
  - Les résultats de l'ACP obtenus dans « 2 ».
  - Les résultats de l'AFC de l'étape « 2 ».
4. En utilisant la fonction « *cutree* » et en indiquant le nombre de classes adéquat, récupérez et visualisez les classes obtenues avec les différentes configurations dans « 3 ».
5. Confrontez les classes obtenues dans « 4 » avec les vraies classes suivantes : Espèce 1 : {1, 5, 11, 15, 20, 23}, Espèce 2 : {2, 4, 8, 9, 14, 16, 22}, Espèce 3 : {3, 7, 10, 12, 13, 17, 18, 19, 21}, Espèce 4 : {6}.
6. Interprétez et comparez les résultats obtenus dans « 5 ».
7. Appliquez une CAH en utilisant le critère de Ward sur les données Papillons mais cette fois-ci en utilisant la distance du Chi2 donnée par la formule suivante :

$$d(x_i, x_{i'}) = \sum_{j=1}^p \frac{1}{x_{.j}} \left( \frac{x_{ij}}{x_i} - \frac{x_{i'j}}{x_{i'}} \right)^2 = \sum_{j=1}^p \left( \frac{x_{ij}}{\sqrt{x_{.j}x_i}} - \frac{x_{i'j}}{\sqrt{x_{.j}x_{i'}}} \right)^2$$

- Proposez une partition en 4 classes et comparez le résultat avec ceux de l'étape « 5 ».
- Qu'en déduisez-vous ?