

TP 4

L3 MIASHS - UBO

Dans ce TP, on reprend les données du TP3 sur les systèmes de santé de 185 pays à travers le monde. On considère que le traitement des données a été préalablement effectué (partie 1 “Préparation des données” du TP 3).

1. Calculer la matrice des distances entre les individus à l’aide de la fonction *dist*. Quelle est la distance utilisée par défaut par *R* ?
2. Tracer le dendrogramme. Pour cela, appliquer la fonction *hclust* à la matrice des distances calculée. Quel est la mesure de dissimilarité utilisée par défaut par *R* ?
3. Proposer un nombre *k* de classes à retenir. Pour cela, tracer la variation de l’inertie en fonction du nombre de classes à l’aide des lignes suivantes (où “res_cah” est le résultat de *hclust*) :

```
inertie=sort(res_cah$height, decreasing = TRUE)
plot(inertie[1:15], type = "s", xlab = "Nombre de classes", ylab = "Inertie",lwd=2);grid()
```

4. Représenter sur le dendrogramme obtenu à la question 2 les classes retenues à l’aide de la fonction *rect.hclust*.
5. On cherche maintenant à interpréter les différentes classes.
 - Utiliser la fonction *cutree* pour effectuer le regroupement en *k* classes. Afficher et interpréter le résultat.
 - Représenter les individus dans le premier plan principal de l’ACP en associant une couleur différente aux individus des différentes classes. Interpréter. Les classes sont-elles cohérentes ? Ont-elles du sens ?
6. Obtient-on les mêmes résultats si on réduit les données ?
7. Reprendre les questions 3 à 5 en changeant la mesure de dissimilarité utilisée (saut minimum, saut maximum et saut moyen). Observe-t-on des différences selon les méthodes utilisées ? Quelle est la meilleure classification ? Pourquoi ?
8. On veut maintenant comparer les résultats obtenus avec la classification hiérarchique ascendante (CAH) avec ceux obtenus avec la méthode des moyennes mobiles (fonction *kmeans* sous *R*).
 - Réaliser une classification avec la méthode des moyennes mobiles en initialisant le centre des classes avec les centres des classes obtenus par CAH. Comparer avec les classes identifiées par CAH.
 - Réaliser une classification avec la méthode des moyennes mobiles en initialisant le centre des classes de manière aléatoire. Comparer avec les résultats obtenus précédemment.