

**TP-1 Analyse en Composantes Principales**

La base de données **Decathlon** se trouve dans le package **FactoMineR**. Les données contiennent les performances d'athlètes lors de deux compétitions. Le tableau de données contient 41 lignes et 13 colonnes. Les colonnes de 1 à 12 sont des variables quantitatives : les dix premières colonnes correspondent aux performances des athlètes pour les dix épreuves du décathlon et les colonnes 11 et 12 correspondent respectivement au rang et au nombre de points obtenus. La dernière colonne est une variable qualitative correspondant au nom de la compétition (Jeux Olympiques de 2004 ou Décastar 2004).

L'objectif de ce TP est de réaliser une première exploration et visualisation des données **Decathlon** puis de réaliser une ACP sur ces données.

### **I. Exploration et visualisation des données**

1. Importer et explorer le jeu de données.
2. Représenter des tableaux résumés des scores des athlètes selon la compétition : on utilisera la fonction *ddply* dans le package **plyr**.
3. Normaliser toutes les variables d'intérêt pour pouvoir les comparer.
4. Comparer les résultats des deux compétitions pour toutes les variables quantitatives.
5. Étudier les corrélations entre les différentes variables quantitatives. Penser à tracer un corrgram.

### **II. Programmer les étapes de L'ACP normée**

1. Centrer et réduire les variables d'intérêt.
2. Déterminer les valeurs propres et les vecteurs propres de l'ACP.
3. Calculer les coordonnées des variables et celles des individus sur les axes principaux.

### **III. Mise en oeuvre de l'ACP avec les packages FactoMineR, Factoextra, princomp et pca3d**

1. Exécuter l'ACP en utilisant **FactoMineR** et comparer les résultats obtenus à ceux de la partie précédente.
2. Exécuter l'ACP en utilisant le package **Factoextra** puis comparer les représentations graphiques des 2 packages.

### Pour aller plus loin

1. Effectuer une rotation varimax avec la fonction *principal* du package **psych** :  
`acp.varimax <- principal(r=cor.matrix, nfactors=4, rotate="varimax")`
2. Utiliser le package **Factoshiny** pour améliorer facilement et de façon interactive les graphiques : <http://factominer.free.fr/graphs/factoshiny-fr.html>
3. Aide à l'interprétation (à utiliser avec précaution...) : la carte de représentation simultanée individus-variables (biplot) obtenue en utilisant le package **ade4**.
4. Représenter les résultats de l'ACP en 3D en ayant recours au package **pca3d**