

## TP. 1

### Rappels ACP

---

#### Exercice 1

On étudie les performances des athlètes ayant participé en 2004 aux épreuves de décathlon des jeux Olympiques et du Décastar. Les données sont disponibles sur Moodle : decathlon2.dat. Les dix épreuves du décathlon sont :

- la course sur 100m `c100`
- le saut en longueur `long`
- le lancer de poids `poids`
- le saut en hauteur `haut`
- la course sur 400m `c400`
- la course de haies sur 110m `c110`
- le lancer de disques `disq`
- le saut à la perche `perch`
- le lancer de javelot `javel`
- la course sur 1500m `c1500`

Les performances de course sont mesurées en secondes, les autres en mètres. On dispose également des variables suivantes :

- rang de classement `RANG`
- nombre de points `POINTS`
- compétition `COMPET`, jeux olympiques `JO` ou décastar `DS`.

Les noms des participants sont en majuscule pour le décastar afin de différencier les participations d'un même athlète aux deux épreuves.

1. Télécharger les données. Afficher la matrice de corrélation des variables quantitatives. Quelles sont les couples de variables les plus corrélées, les moins corrélées ? Commenter les corrélations des différentes variables.
2. On effectue à présent l'ACP en mettant de côté les variables `RANG`, `POINTS` et `COMPET`. Calculer l'inertie associée à chacun des axes. À partir des critères vus en cours, combien d'axes doit-on garder pour l'analyse ? Dans la suite, on gardera uniquement les trois premiers axes (donner la part d'inertie totale représentée par ces trois premiers axes).
3. Tracer le cercle de corrélations entre les variables et les axes. On représentera le cercle de corrélations entre le premier et le deuxième axe, puis entre le deuxième et le troisième axe. Quelles sont les variables qui déterminent les trois premières composantes principales ?

4. Calculer les contributions de chaque individus aux axes. Quels individus contribuent le plus aux trois premiers axes ?
5. Calculer la qualité de représentation des individus par les espaces principaux. Quels sont les quatre individus les moins bien représentés par le sous espace composé des trois premiers axes ?
6. Calculer la corrélation entre les variables **RANG**, **POINTS** et les axes principaux. Commenter.
7. Reprendre les données et inverser le signe des variables de course. Interpréter de nouveau les trois axes principaux.