



---

## TP : Analyse multivariée

### Analyse en composantes principales

---

*Durée : 2h30*

---

*Ce TP a pour objectif de mettre en œuvre une analyse en composantes principales à l'aide d'un logiciel de statistique.*

#### Exercice 1

*Données : DecathlonData.xls*

Le tableau de données contient 41 lignes et 13 colonnes (visualiser les données).

- Les colonnes 1 à 12 sont des variables continues:
  - les dix premières colonnes correspondent aux performances des athlètes pour les dix épreuves du décathlon
  - 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).
- Les lignes désignent les athlètes.

Nous allons faire une ACP sur les colonnes de 1 à 10.

- 1) Quel pourcentage de l'inertie totale contiennent les deux premières composantes principales ? Combien faut-il choisir de composantes principales pour avoir plus de 70% de l'inertie totale ?
- 2) Etude des variables.
  - a) Pourquoi les variables « X100m », « X400m », « X110m.hurdle » et « X1500m » se trouvent-elles à gauche de l'axe des ordonnées ?
  - b) Comment interprétez-vous la corrélation entre les variables « X100m » et « long.jump » ?
  - c) Peut-on distinguer des groupes de variables ? Quelle est la corrélation entre ces groupes ? Comment l'interprétez-vous ?
  - d) Quelles variables contribuent majoritairement à la première composante principale, à la deuxième composante principale ? Comment pouvez-vous interpréter le plan défini par les deux premières composantes principales ?

### 3) Etude des individus

- a) Comment qualifieriez-vous l'athlète Lorenzo ? A votre avis, comment se fait-il qu'il ne soit pas dernier de sa compétition ?
- b) Comment qualifieriez-vous les athlètes suivants : Karpov, Sebrle, Casarsa ? Quel est leur classement ?
- c) Peut-on en conclure qu'il faut être rapide pour gagner le décathlon et que la puissance ne suffit pas ? Pour répondre à cette question, on ajoute les variables supplémentaires « Rank » et « Points ». Ces variables n'entrent pas en compte dans le calcul des composantes principales mais aident à une meilleure compréhension des axes. Que pouvez-vous en conclure ?
- d) Comparer la position de Karpov, Clay,... aux jeux olympiques et au décastar. Peut-on en conclure que le niveau des deux compétitions n'est pas le même ? Pour répondre à cette question, on ajoute la variable supplémentaire « Competition ». Cette variable est qualitative et est qualifiée de facteur. Deux nouveaux individus représentant un individu moyen pour chaque compétition sont ajoutés au graphique. Que pouvez-vous en conclure ?

## Exercice 2

*Données : EconomieEuroData.xls*

Le tableau de données contient 20 lignes et 8 colonnes (visualiser les données).

- Les lignes désignent des pays.
- La colonne 1 désigne le nom du pays
- Les colonnes 2 à 6 sont des données économiques des pays
- La colonne 7 est la population du pays
- La colonne 8 est une variable qualitative désignant l'appartenance du pays à la zone euro.

Nous allons faire une ACP sur les colonnes de 2 à 6. Les colonnes 7 et 8 seront ajoutées comme variable et facteur supplémentaires.

- 1) Quel pourcentage de l'inertie totale contiennent les deux premières composantes principales ? Combien faut-il choisir de composantes principales pour avoir environ 70% de l'inertie totale ?
- 2) Interpréter les axes, la liaison entre les variables et donner une cartographie.
- 3) Etudier des individus. Que pouvez-vous dire du Luxembourg, de la Grèce et de la Slovaquie ? Peut-on caractériser les pays de la zone euro ?

- 4) Que pouvez-vous conclure sur la contribution des individus (pays) à la construction des axes ? Refaite une ACP avec le Luxembourg (et la Slovaquie) en individu(s) supplémentaire(s). Que se passe-t-il ? Que pouvez-vous dire sur vos conclusions précédentes ?

### Exercice 3

*Données : PoissonsData.xls*

24 mulets ont été répartis dans trois aquariums, A1, A2 et A3, contaminés par radioactivité de façon identique mais, par contre, avec un temps d'exposition différent. Le poisson 17 est mort pendant l'expérience et n'apparaît pas dans le tableau.

On cherche à savoir s'il y a une influence du temps d'exposition à la radioactivité sur la taille des poissons. Nous disposons de mesures sur les 16 variables suivantes :

- 9 concernant la contamination à radioactivité, des yeux (ŒIL), des branchies (BRAN), des opercules (OPER), des nageoires (NAGE), du foie (FOIE), du tube digestif (TUDI), des reins (REINS), des écailles (ECAI), des muscles (MUSC).
- 7 concernant la morphologie du poisson : le poids (Poid), la longueur (Long), la longueur standard (Lng), la largeur de la tête (Tete), la largeur (Larg), la largeur du museau (Muse), le diamètre des yeux (Doeil)

Effectuer une analyse en composantes principales pour répondre à la question.