

Analyse de données

TP 2

L3 MIASHS

Dans ce TP, on reprend les données du TP 1 : les performances réalisées par des athlètes lors de deux compétitions de décathlon, le Décastar de 2004 et les Jeux Olympiques.

1 Préparation des données

- a) On va se limiter aux données des Jeux Olympiques. Pour cela créer un tableau de données X contenant toutes les observations pour la compétition "JO" :
`> X = decathlon[decathlon$Competition == "JO",2 :11]`
- b) Centrer les données à l'aide de la fonction *scale*.
- c) Calculer la matrice de covariance. Quelles sont les variables qui ont une forte variance ? Certaines variables sont-elles fortement dépendantes ?

2 Calcul des axes principaux et des composantes principales

- a) Calculer les valeurs propres et les vecteurs propres de cette matrice (utiliser la fonction *eigen*).
- b) Représenter les valeurs propres graphiquement. Quel est le pourcentage d'inertie expliqué par le premier axe principal ? par le premier plan principal ? Combien d'axes faut-il retenir pour expliquer 90% de l'inertie ?
- c) Calculer les composantes principales. Calculer la matrice de covariance des composantes principales. Les résultats obtenus sont-ils conformes à la théorie ?

3 Qualité de l'ACP

- a) Calculer les contributions des individus aux différentes composantes (CTR). Existe-t-il des individus qui apportent une contribution importante à l'un des deux premiers axes principaux ?
- b) Calculer les quantités COS2 et discuter de la qualité de la représentation des individus sur le premier axe principal puis sur le premier plan principal.
- c) Tracer le cercle des corrélations puis, sur une autre figure, les individus projetés sur le premier plan principal (distingue-t-on des groupes d'individus ?). Interpréter les résultats obtenus.