

# Analyse en Composantes Principales

## Jeu de données decathlon

Ce jeu de données contient 41 individus et 13 variables, 1 variable qualitative est illustrative.

---

### 1. Observation d'individus extrêmes

L'analyse des graphes ne révèle aucun individu singulier.

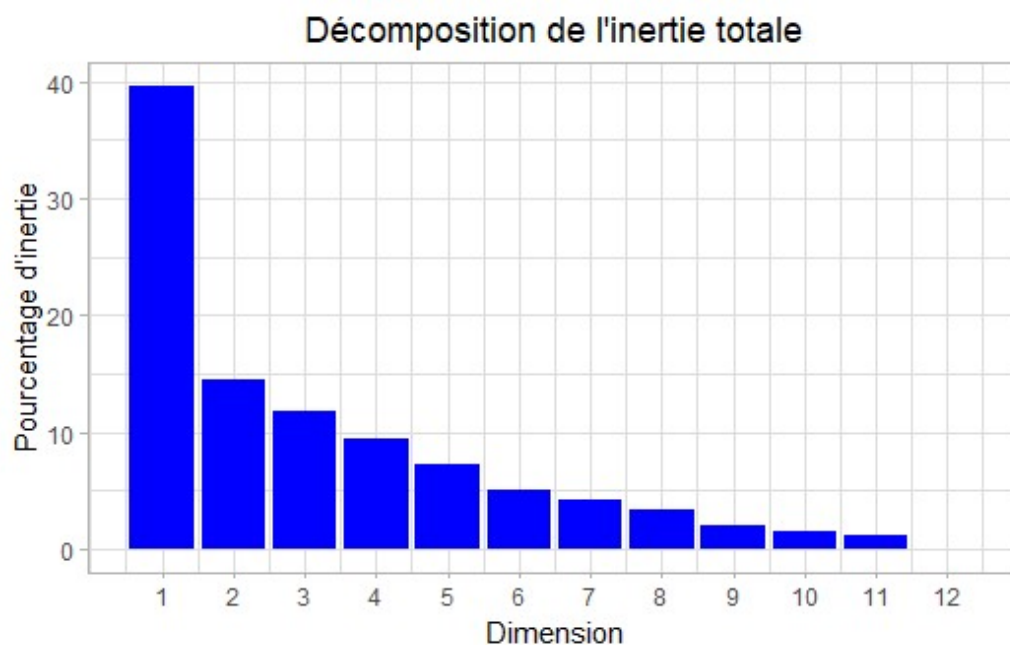
---

### 2. Distribution de l'inertie

L'inertie des axes factoriels indique d'une part si les variables sont structurées et suggère d'autre part le nombre judicieux de composantes principales à étudier.

Les 2 premiers axes de l'analyse expriment **54.16%** de l'inertie totale du jeu de données ; cela signifie que 54.16% de la variabilité totale du nuage des individus (ou des variables) est représentée dans ce plan. C'est un pourcentage assez important, et le premier plan représente donc convenablement la variabilité contenue dans une grande part du jeu de données actif. Cette valeur est nettement supérieure à la valeur référence de **33.89%**, la variabilité expliquée par ce plan est donc hautement significative (cette inertie de référence est le quantile 0.95-quantile de la distribution des pourcentages d'inertie obtenue en simulant 471 jeux de données aléatoires de dimensions comparables sur la base d'une distribution normale).

Du fait de ces observations, il serait tout de même probablement préférable de considérer également dans l'analyse les dimensions supérieures ou égales à la troisième.

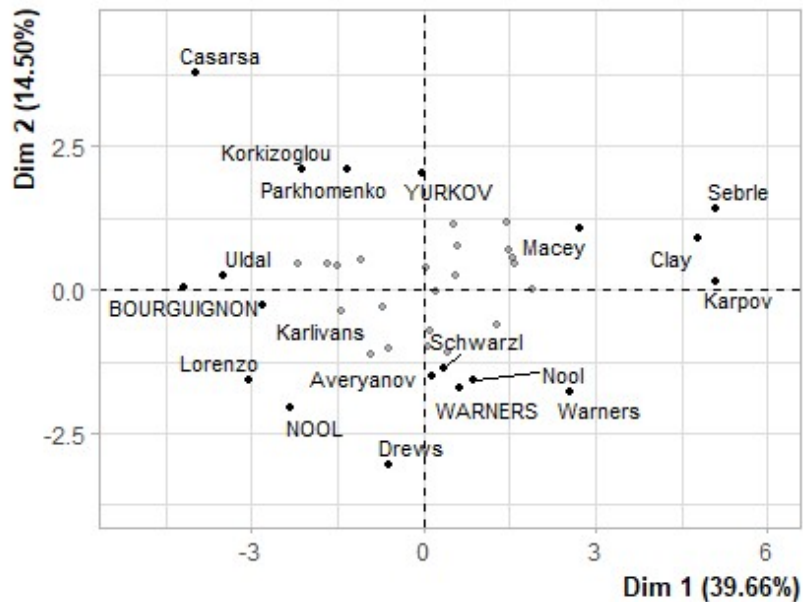


**Figure 2 - Décomposition de l'inertie totale**

Une estimation du nombre pertinent d'axes à interpréter suggère de restreindre l'analyse à la description des 1 premiers axes. Ces composantes révèlent un taux d'inertie supérieur à celle du quantile 0.95-quantile de distributions aléatoires (39.66% contre 19.37%). Cette observation suggère que seul cet axe est porteur d'une véritable information. En conséquence, la description de l'analyse sera restreinte à ces seuls axes.

---

### 3. Description de la dimension 1

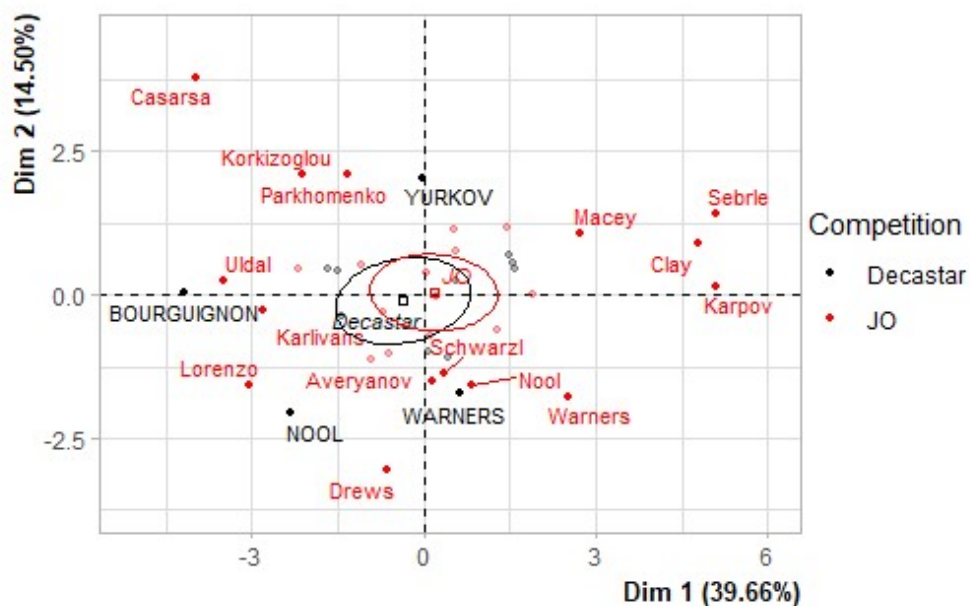


**Figure 3.1 - Graphe des individus (ACP)** Les individus libellés sont ceux ayant la plus grande contribution à la construction du plan.

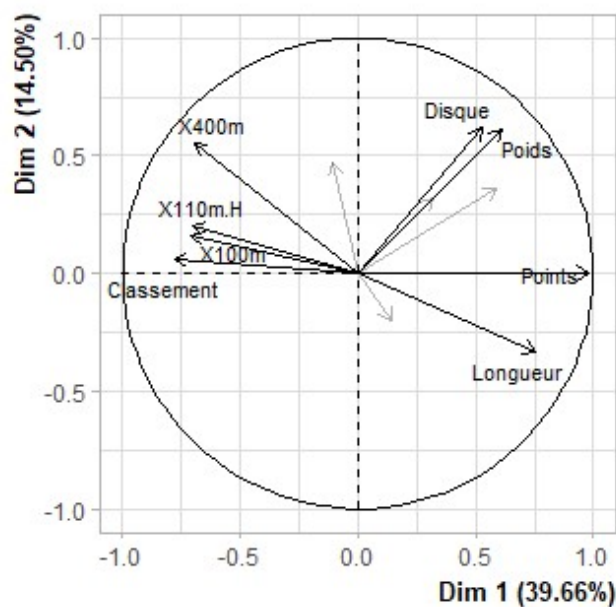
La probabilité critique du test de Wilks indique la variable dont les modalités sépare au mieux les individus sur le plan (i.e. qui explique au mieux les distances entre individus).

```
## Competition
## 0.7384367
```

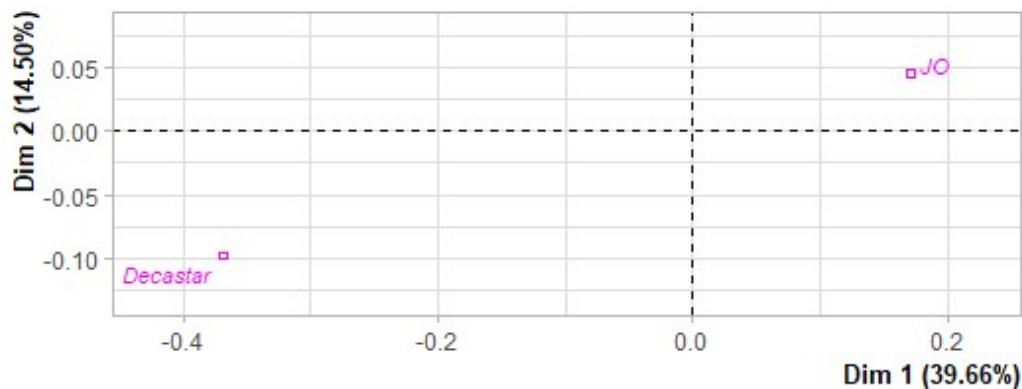
Il n'y a qu'une unique variable qualitative possible pour illustrer les distances entre individus : *Competition*.



**Figure 3.2 - Graphe des individus (ACP)** Les individus libellés sont ceux ayant la plus grande contribution à la construction du plan. Les individus sont colorés selon leur appartenance aux modalités de la variable Competition.



**Figure 3.3 - Graphe des variables (ACP)** Les variables libellées sont celles les mieux représentées sur le plan.



**Figure 3.4 - Graphe des modalités (ACP)** Les facteurs libellés sont ceux les mieux représentés sur le plan.

La **dimension 1** oppose des individus tels que *Sebrle, Karpov, Clay* et *Macey* (à droite du graphe, caractérisés par une coordonnée fortement positive sur l'axe) à des individus comme *Casarsa, Uldal, BOURGUIGNON, Lorenzo, Karlivans, NOOL, Parkhomenko* et *Korkizoglou* (à gauche du graphe, caractérisés par une coordonnée fortement négative sur l'axe).

Le groupe auquel les individus *Sebrle, Karpov, Clay* et *Macey* appartiennent (caractérisés par une coordonnée positive sur l'axe) partage :

- de fortes valeurs pour les variables *Hauteur, Points, Disque, Poids, Javelot* et *Longueur* (de la plus extrême à la moins extrême).
- de faibles valeurs pour les variables *Classement* et *X100m* (de la plus extrême à la moins extrême).

Le groupe auquel les individus *Uldal, BOURGUIGNON, Lorenzo, Karlivans* et *NOOL* appartiennent (caractérisés par une coordonnées négative sur l'axe) partage :

- de fortes valeurs pour les variables *X100m* et *X110m.H* (de la plus extrême à la moins extrême).
- de faibles valeurs pour les variables *Poids, Points, Hauteur* et *Disque* (de la plus extrême à la moins extrême).

Le groupe auquel les individus *Casarsa, Parkhomenko* et *Korkizoglou* appartiennent (caractérisés par une coordonnées négative sur l'axe) partage :

- de fortes valeurs pour les variables *X400m*, *X1500m* et *Classement* (de la plus extrême à la moins extrême).
- de faibles valeurs pour la variable *Longueur*.

Notons que la variable *Points* est extrêmement corrélée à cette dimension (corrélation de 0.96). Cette variable pourrait donc résumer à elle seule la dimension 1.

---

---

## Annexes