Analyse en Composantes Principales

Jeu de données decathlon

Ce jeu de données contient 41 individus et 10 variables.

### 1. Observation d’individus extrêmes

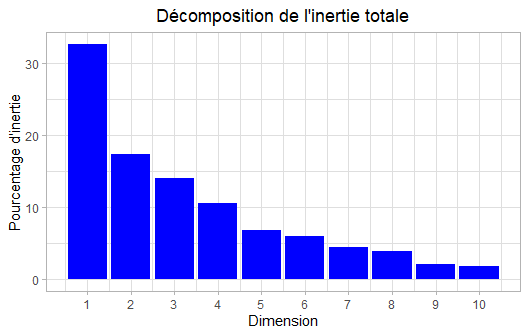
L’analyse des graphes ne révèle aucun individu singulier.

### 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 **50.09%** de l’inertie totale du jeu de données ; cela signifie que 50.09% 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 supérieure à la valeur référence de **37.79%**, la variabilité expliquée par ce plan est donc significative (cette intertie de référence est le quantile 0.95-quantile de la distribution des pourcentages d’inertie obtenue en simulant 6635 jeux de données aléatoires de dimensions comparables sur la base d’une distribution normale).

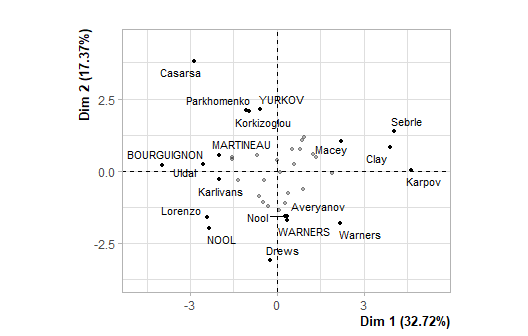
Du fait de ces observations, il serait tout de même probablement préférable de considérer également dans l’analyse les dimensions supérieures ou égales à la troisième.



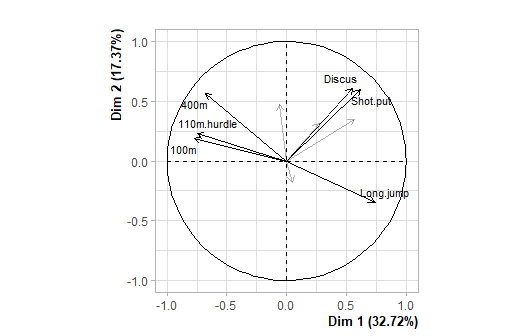
**Figure 2 - Décomposition de l’inertie totale**

Une estimation du nombre pertinent d’axes à interpréter suggère de restreindre l’analyse à la description des 3 premiers axes. Ces composantes révèlent un taux d’inertie supérieur à celle du quantile 0.95-quantile de distributions aléatoires (64.14% contre 51.61%). Cette observation suggère que seuls ces axes sont porteurs d’une véritable information. En conséquence, la description de l’analyse sera restreinte à ces seuls axes.

### 3. Description du plan 1:2



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



**Figure 3.2 - Graphe des variables (ACP)** *Les variables libellées sont celles les mieux représentées sur le plan.*

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

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

* de fortes valeurs pour les variables *High.jump*, *Discus*, *Shot.put* et *Long.jump* (de la plus extrême à la moins extrême).
* de faibles valeurs pour les variables *100m*, *110m.hurdle* et *400m* (de la plus extrême à la moins extrême).

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

* de fortes valeurs pour les variables *110m.hurdle* et *100m* (de la plus extrême à la moins extrême).
* de faibles valeurs pour les variables *Shot.put*, *High.jump* et *Discus* (de la plus extrême à la moins extrême).

La **dimension 2** oppose des individus tels que *Casarsa*, *YURKOV* et *Parkhomenko* (en haut du graphe, caractérisés par une coordonnées fortement positive sur l’axe) à des individus comme *Warners*, *Drews* et *WARNERS* (en bas du graphe, caractérisés par une coordonnées fortement négative sur l’axe).

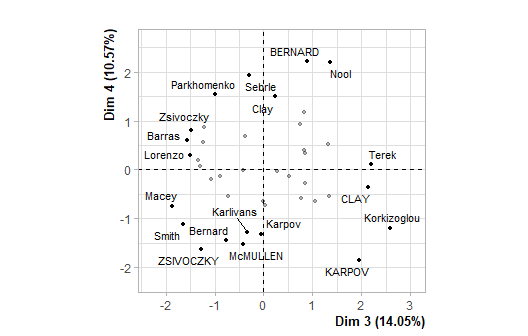
Le groupe auquel les individus *Casarsa*, *YURKOV* et *Parkhomenko* appartiennent (caractérisés par une coordonnée positive sur l’axe) partage :

* de fortes valeurs pour la variable *400m*.
* de faibles valeurs pour la variable *Long.jump*.

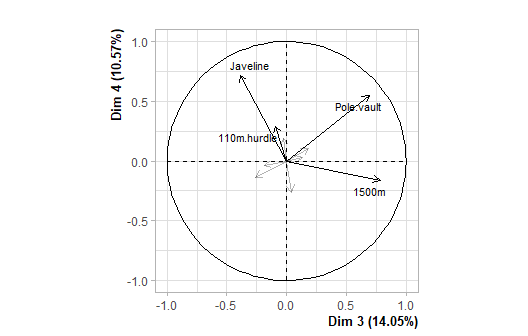
Le groupe auquel les individus *Warners*, *Drews* et *WARNERS* appartiennent (caractérisés par une coordonnées négative sur l’axe) partage :

* de fortes valeurs pour la variable *Pole.vault*.
* de faibles valeurs pour la variable *110m.hurdle*.

### 4. Description de la dimension 3



**Figure 4.1 - Graphe des individus (ACP)** *Les individus libellés sont ceux ayant la plus grande contribution à la construction du plan.*



**Figure 4.2 - Graphe des variables (ACP)** *Les variables libellées sont celles les mieux représentées sur le plan.*

La **dimension 3** oppose des individus tels que *KARPOV*, *Korkizoglou*, *Terek* et *CLAY* (à droite du graphe, caractérisés par une coordonnée fortement positive sur l’axe) à des individus comme *ZSIVOCZKY*, *Barras*, *Zsivoczky*, *McMULLEN*, *Macey*, *Bernard* et *Smith* (à gauche du graphe, caractérisés par une coordonnée fortement négative sur l’axe).

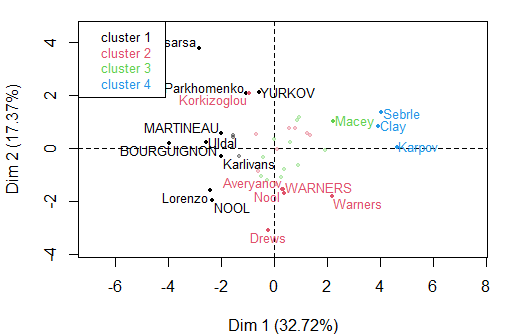
Le groupe auquel les individus *KARPOV*, *Korkizoglou*, *Terek* et *CLAY* appartiennent (caractérisés par une coordonnée positive sur l’axe) partage :

* de fortes valeurs pour la variable *1500m*.
* de faibles valeurs pour la variable *Javeline*.

Le groupe auquel les individus *ZSIVOCZKY*, *Barras*, *Zsivoczky*, *McMULLEN*, *Macey*, *Bernard* et *Smith* appartiennent (caractérisés par une coordonnées négative sur l’axe) partage :

* de faibles valeurs pour les variables *Pole.vault* et *1500m* (de la plus extrême à la moins extrême).

### 5. Classification



**Figure 5 - Classification Ascendante Hiérachique des individus.** *La classification réalisée sur les individus fait apparaître 4 classes.*

La **classe 1** est composée d’individus tels que *YURKOV*, *MARTINEAU*, *NOOL*, *BOURGUIGNON*, *Parkhomenko*, *Lorenzo*, *Karlivans*, *Uldal* et *Casarsa*. Ce groupe est caractérisé par :

* de fortes valeurs pour les variables *100m*, *110m.hurdle* et *400m* (de la plus extrême à la moins extrême).
* de faibles valeurs pour les variables *Long.jump* et *Shot.put* (de la plus extrême à la moins extrême).

La **classe 2** est composée d’individus tels que *WARNERS*, *Warners*, *Nool*, *Averyanov*, *Drews* et *Korkizoglou*. Ce groupe est caractérisé par :

* de fortes valeurs pour les variables *Pole.vault* et *1500m* (de la plus extrême à la moins extrême).
* de faibles valeurs pour les variables *Javeline* et *100m* (de la plus extrême à la moins extrême).

La **classe 3** est composée d’individus tels que *Macey*. Ce groupe est caractérisé par :

* de faibles valeurs pour les variables *Pole.vault* et *1500m* (de la plus extrême à la moins extrême).

La **classe 4** est composée d’individus tels que *Sebrle*, *Clay* et *Karpov*. Ce groupe est caractérisé par :

* de fortes valeurs pour les variables *Long.jump*, *Discus*, *Shot.put*, *Javeline* et *High.jump* (de la plus extrême à la moins extrême).
* de faibles valeurs pour les variables *100m*, *400m* et *110m.hurdle* (de la plus extrême à la moins extrême).

## Annexes