

Méthodes non supervisées

Table des matières

1 Introduction aux méthodes non supervisées	2
1.A Groupes homogènes de sujets	2
1.A.1 Analyse en composantes principales	2
1.A.1.1 Principe	2

1 Introduction aux méthodes non supervisées

- Objectif des méthodes non supervisées = mettre en évidence des groupes homogènes de **sujets** ou de **variables**.
- 2 options :
 - Réduction du nombre de dimensions pour rendre compatible avec nos yeux : **ACP**
 - Demander à l'ordinateur de se substituer à nous : nuées dynamiques, classification hiérarchiques...

1.A Groupes homogènes de sujets

1.A.1 Analyse en composantes principales

1.A.1.1 Principe

- But = réduire le nombre de dimensions (sujets ou variables) tout en conservant un maximum d'information.

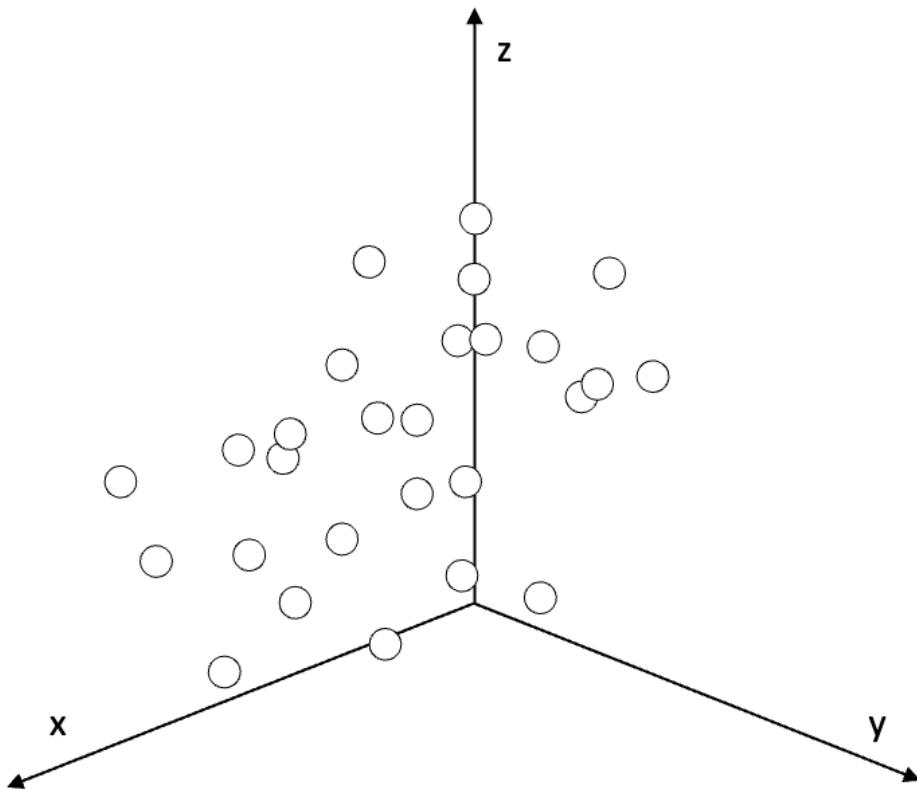


Figure 1: Représentation géométrique d'un tableau de données : les sujets sont symbolisés par des points (ici des « O »), les variables par les dimensions de l'espace.

La première composante principale est la droite dans le nuage de points qui **explique le maximum de variance**, c'est à dire la direction suivant laquelle le nuage de points s'étire au maximum.

Il s'agit donc d'une combinaison linéaire des variables initiales (ici x , y et z) qui maximise la variance des projections des points sur cette droite.

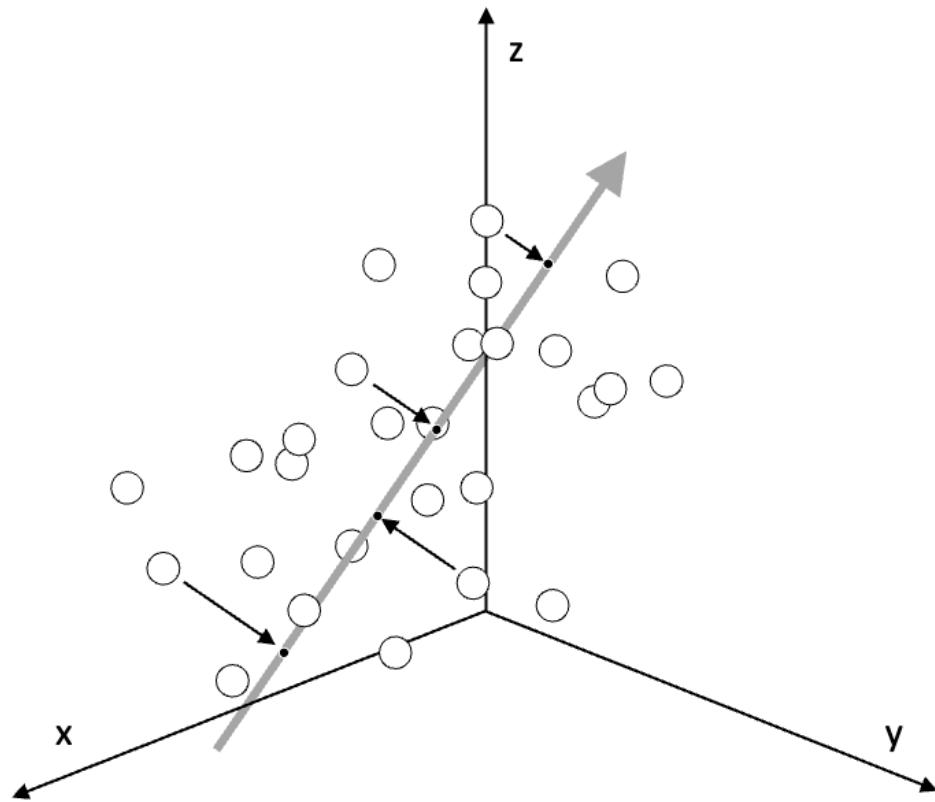


Figure 2: On recherche dans un premier temps la direction suivant laquelle le nuage de points s'étire au maximum

Représentation R de cette droite sur le plot précédent (avec les distances de chaque point à cette droite) :

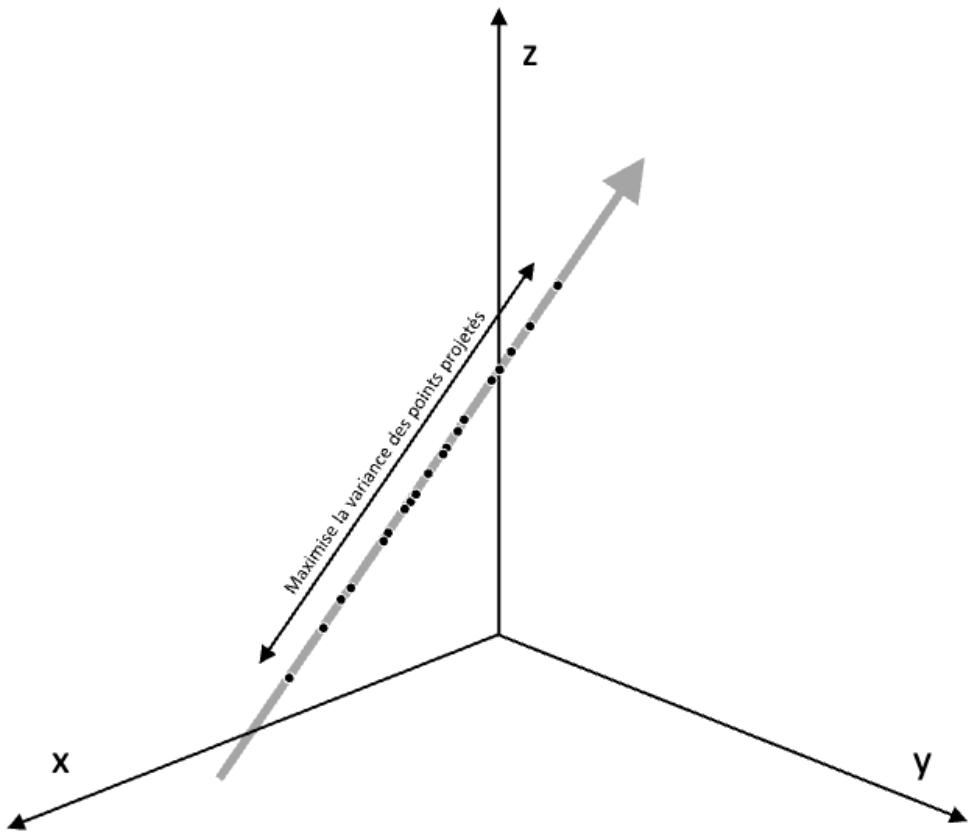


Figure 3: A cette direction correspond une variable dont la variance est maximale : c'est la première composante principale

Pour obtenir la seconde composante principale :

- Il faut d'abord projeter les points sujets sur un plan perpendiculaire à la droite première composante

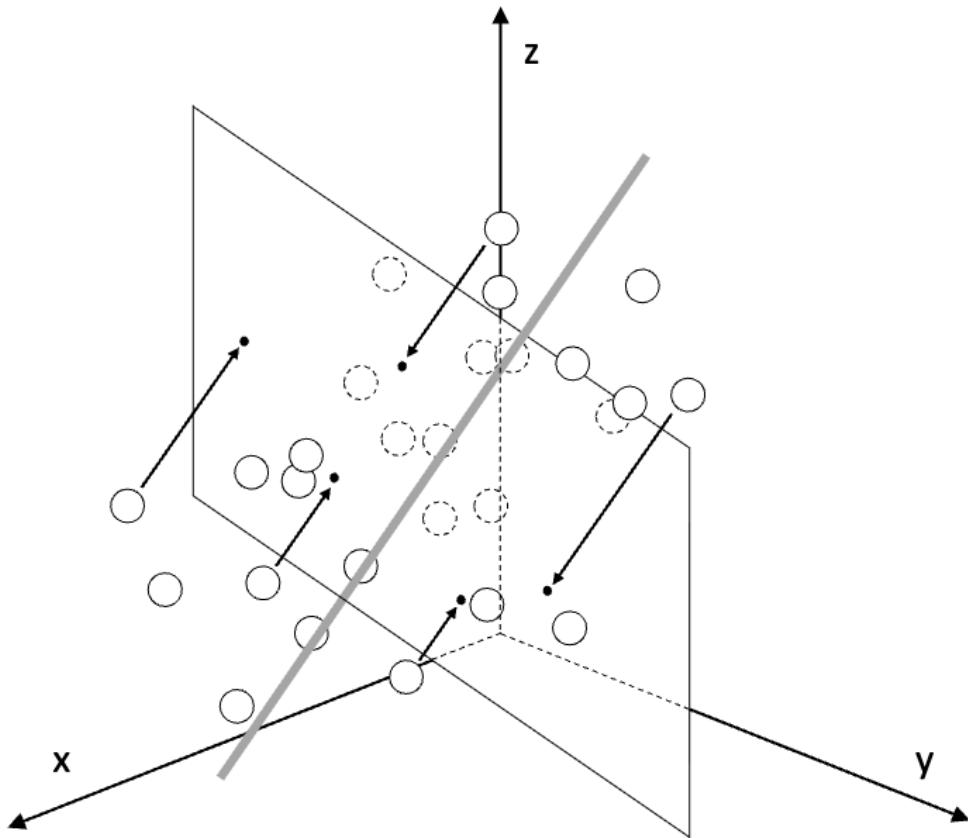


Figure 4: pour obtenir la seconde composante principale, il faut projeter les points sur un plan perpendiculaire à la première composante

Représentation R de ce plan perpendiculaire à la droite de la première composante principale et de la projection des points sur ce plan précédent :