

# Analyse des Données

**Analyse Factorielle : Fondements 1**

*Notes du cours*

Données

variables  $P$

individus  $n$

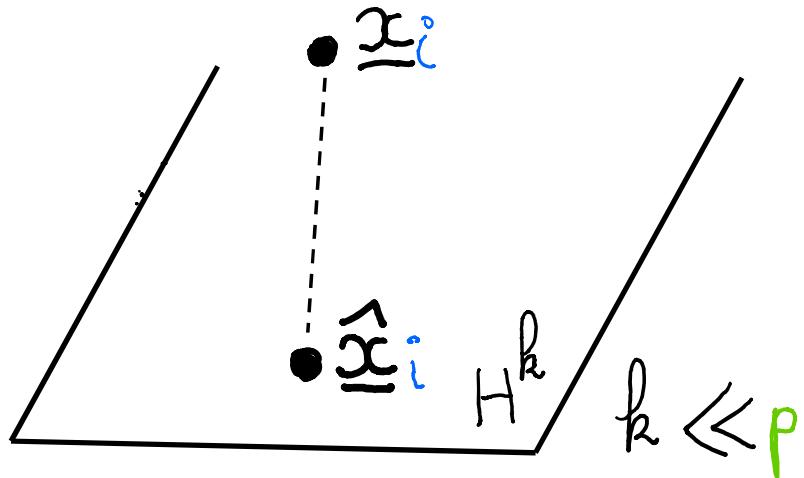
$\underline{x}_i = (x_i^1, \dots, x_i^P) \in \mathbb{R}^P$

The diagram shows a set of points labeled  $x_i^1, x_i^2, \dots, x_i^P$  enclosed in a circle, representing  $n$  individuals. Above the circle, the text "variables  $P$ " is written in green. To the left, the text "individus" and " $n$ " are written in blue.

Nuage des individus  $\mathcal{N}(I)$

The diagram shows a 3D coordinate system with axes labeled "var 1", "var 2", and "var P". A cluster of black dots represents the "Nuage des individus" ( $\mathcal{N}(I)$ ). One specific point in the cluster is highlighted and labeled  $\underline{x}_i$  in blue.

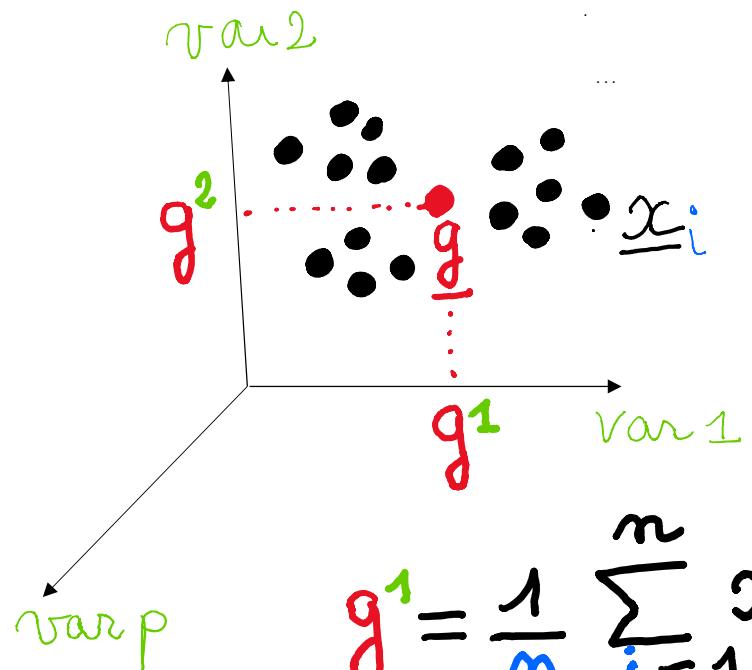
Problème : Représenter le nuage dans un espace de dimension faible  $k$  sans perdre trop d'information



Un problème d'optimisation

$$\underset{H^k}{\text{Min}} \quad M(H^k) = \frac{1}{n} \sum_{i=1}^n d^2(x_i, \hat{x}_i)$$

Les trois prochains cours sont consacrés à la résolution de ce problème !  
Il est nécessaire d'introduire différents ingrédients.



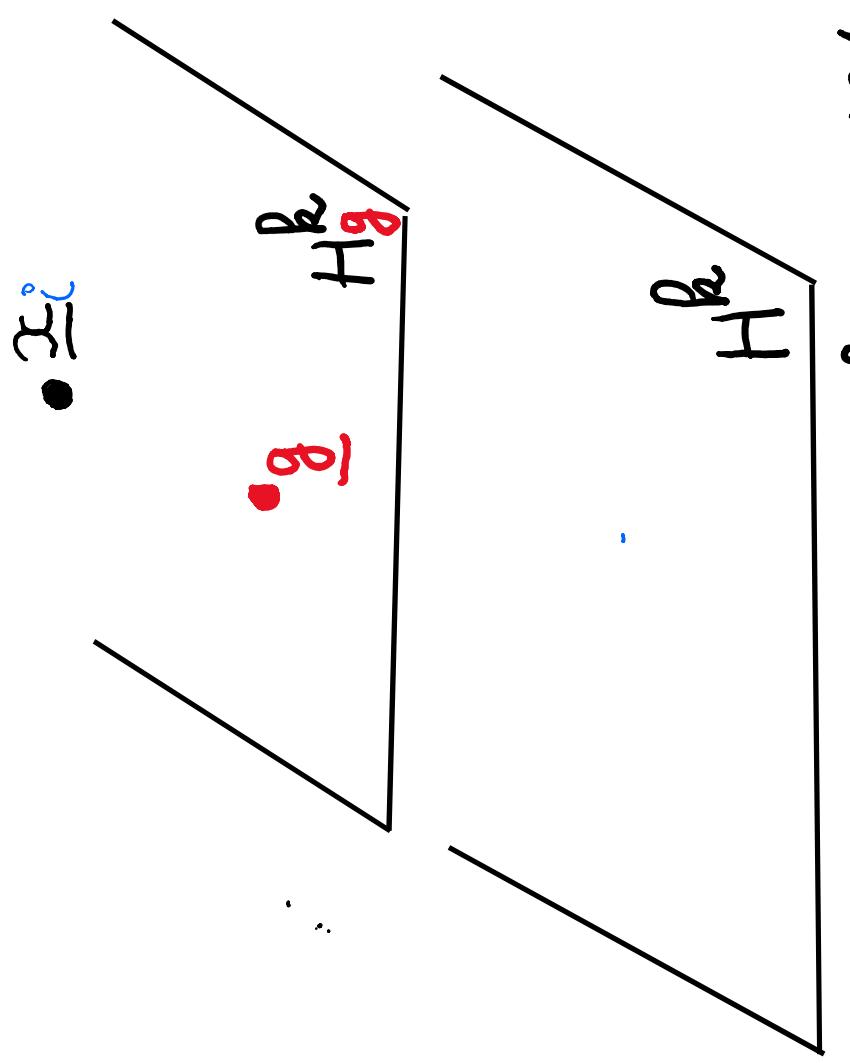
$$\underline{g} = (g^1, \dots, g^p) \in \mathbb{R}^p$$

$$g^1 = \frac{1}{n} \sum_{i=1}^n x_i^1$$

Généralisation de la moyenne  
au cas multi-dimensionnel

Barycentre du  
nuage  $\mathcal{N}(I)$

$H^k$  : sous-espace  
affine  $\parallel \bar{\alpha} H^k$   
passant par  $\bar{g}$



$$m(H^k) = m(H^g) + d^2(H^k, H^g)$$

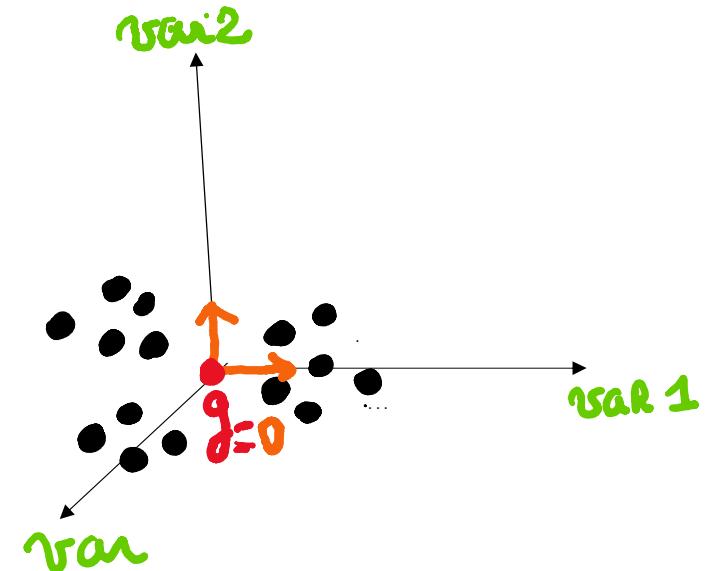
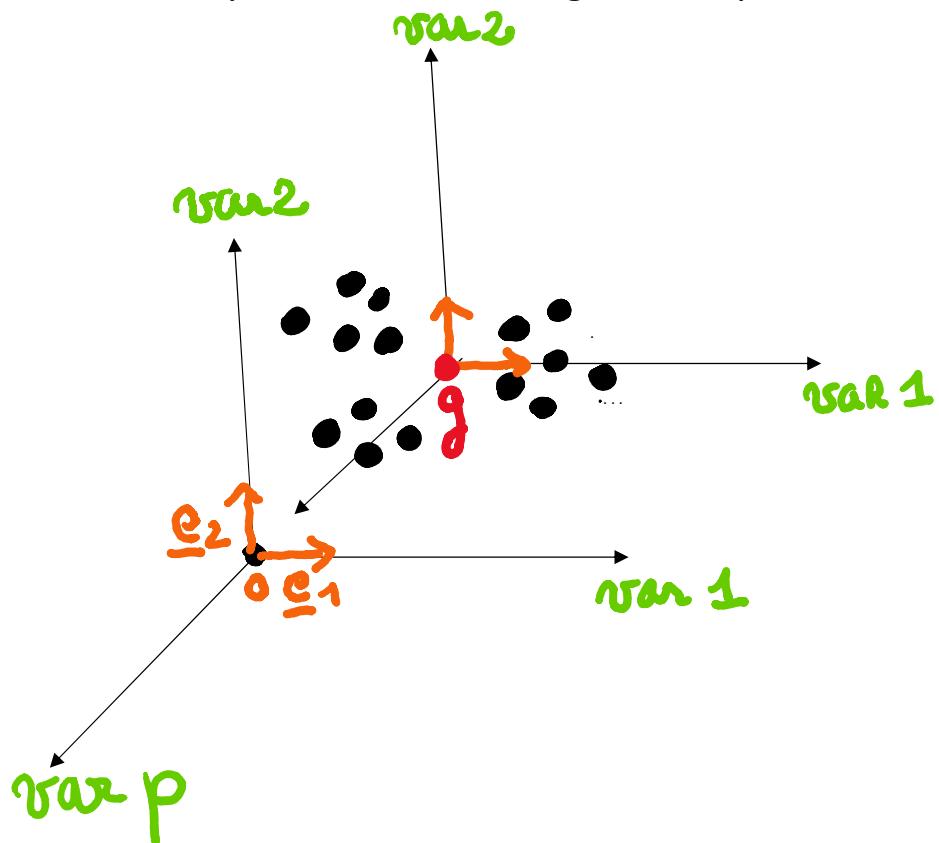
**Conséquence.** Se limiter dans la recherche aux sous espaces passant par le barycentre.

**Problème :**

Rechercher  $H^k$ ,  $k \ll p$ , qui passe  
par le barycentre  $g$  du nuage  $N(I)$  et qui  
minimise  $M(H^k)$

## Astuce

Le barycentre devient l'origine du repère

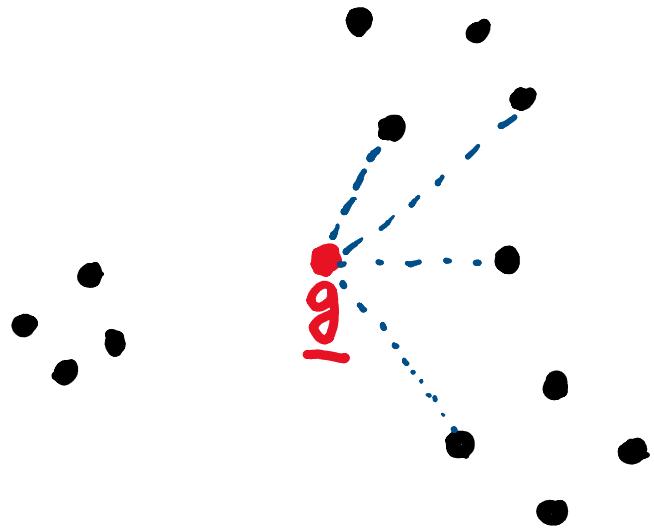


## Problème

Pour  $k < p$  fixé, chercher le sous-espace  $H_{g=0}^k$  passant par  $g$  (origine du repère) minimisant  $M(H_{g=0}^k)$

$$\min_{\substack{H_{g=0}^k}} M(H_{g=0}^k) = \frac{1}{m} \sum_{i=1} d^2(\underline{x}_i, \hat{\underline{x}}_i)$$

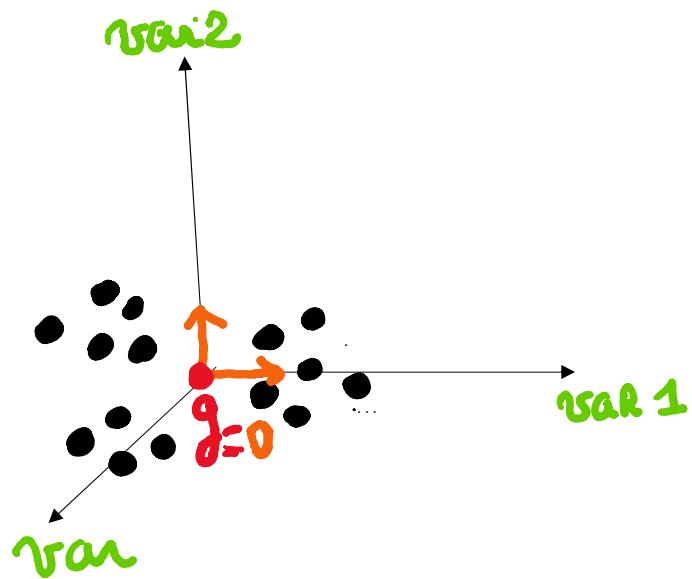
## Réécriture du problème



Inertie du nuage  $\mathcal{N}(I)$

$$I_g = \frac{1}{n} \sum_{i=1}^n d^2(\underline{x}_i, \underline{g})$$

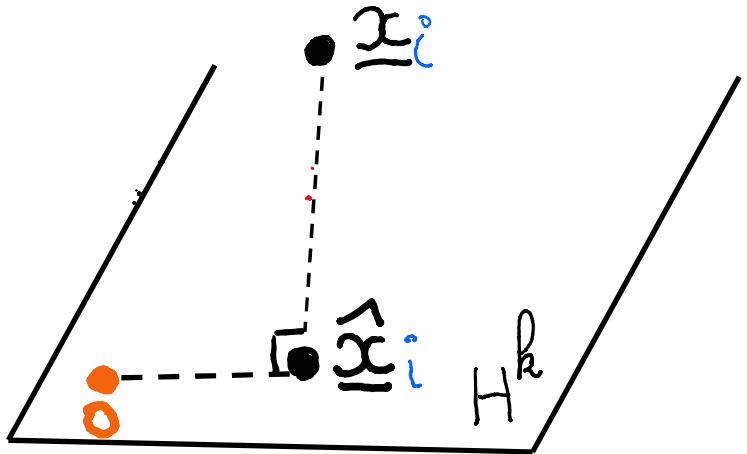
Généralisation de la variance au cas multidimensionnel



Réécriture de l'inertie  
en centrant le repère sur  $\underline{g}$

$$I_g = \frac{1}{n} \sum_{i=1}^n d^2(x_i, \underline{g})$$

$$\begin{aligned} I_{g=0} &= \frac{1}{n} \sum_{i=1}^n d^2(x_i, \underline{g}) \\ &= \frac{1}{n} \sum_{i=1}^n \|x_i\|^2 \end{aligned}$$



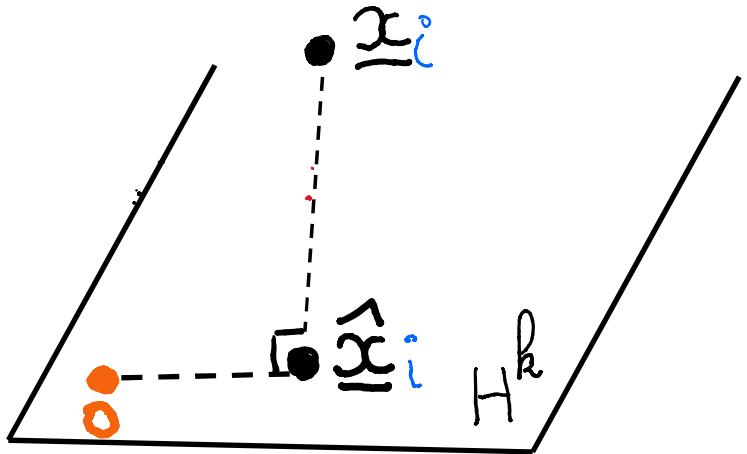
Pythagore

$$I_0 = \frac{1}{n} \sum_{i=1}^n \|\hat{x}_i\|^2 + \frac{1}{n} \sum_{i=1}^n \|x_i - \hat{x}_i\|^2$$

inertie totale

inertie expliquée

inertie résiduelle



$$I_o = \frac{1}{n} \sum_{i=1}^n \|\hat{x}_i\|^2 + \frac{1}{n} \sum_{i=1}^n \|x_i - \hat{x}_i\|^2$$

dépend du  
nuage et pas  
de la projection

Maximiser



Minimiser

## Réécriture du problème

$$I_0 = \frac{1}{m} \sum_{i=1}^n \|\hat{x}_i\|^2 + \frac{1}{m} \sum_{i=1}^n \|x_i - \hat{x}_i\|^2$$

Maximiser

Minimiser

Problème : Pour  $k$  fixé,  $k \ll p$ , chercher le sous-espace  $H^k$  passant par le barycentre tel que

$$\max_{\hat{x}_i \in H^k} \frac{1}{m} \sum_{i=1}^n \|\hat{x}_i\|^2$$