

# Chapitre 3 : Analyse en composante principale (ACP)

# I. Préambule

L'ACP propose, à partir d'un tableau de données relatives à  $p$  variables quantitatives portant sur  $n$  unités (individus), des représentations géométriques de ces unités et de ces variables.

Pour **les unités**, on cherche si l'on peut distinguer des groupes en regardant quelles sont les unités qui se ressemblent, celles qui se distinguent des autres...

Pour **les variables**, on cherche quelles sont celles qui sont très corrélées entre elles, celles qui, au contraire ne sont pas corrélées aux autres...

## II. Tableau de données

On dispose d'un tableau de données relatives à  $p$  variables quantitatives  $x_1, \dots, x_p$  portant sur  $n$  unités (individus).

Le tableau des données  $X$  a la forme suivante :

$$X = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1p} \\ \vdots & \vdots & & \vdots \\ x_{i1} & x_{i2} & \dots & x_{ip} \\ \vdots & \vdots & & \vdots \\ x_{n1} & x_{n2} & \dots & x_{np} \end{pmatrix} .$$

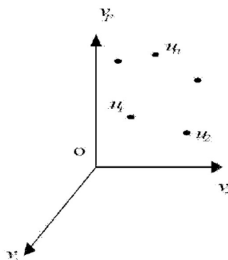
$x_{ij}$  est la valeur de la variable  $x_j$  pour l'unité  $i$ .

## II. Tableau de données (suite)

On peut représenter chaque unité par le vecteur de ses mesures sur les  $p$  variables :

$$u_j = \begin{pmatrix} x_{j1} \\ x_{j2} \\ \vdots \\ x_{jp} \end{pmatrix} \in \mathbb{R}^p.$$

L'ensemble des points qui représentent les unités est appelé "nuage des individus", qui est l'ensemble  $\{u_1, \dots, u_n\}$ .

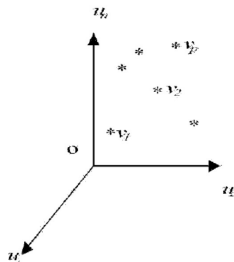


## II. Tableau de données (suite)

Aussi, on peut représenter chaque variable par un vecteur de  $\mathbb{R}^n$  dont les composantes sont les valeurs de la variable pour les  $n$  unités :

$$x_j = \begin{pmatrix} x_{1j} \\ x_{2j} \\ \vdots \\ x_{nj} \end{pmatrix} \in \mathbb{R}^n.$$

L'ensemble des points qui représentent les variables est appelé "nuage des variables", qui est l'ensemble  $\{x_1, \dots, x_p\}$ .



## Exemple (1)

On considère le tableau  $Y$  de notes sur 20 obtenues par 9 élèves en mathématiques, physique, français, et anglais. ( $n=9$  individus,  $p=4$  variables) :

	<i>mathématiques</i>	<i>physique</i>	<i>français</i>	<i>anglais</i>
<i>Jean</i>	6.0	6.0	5.0	5.5
<i>Aline</i>	8.0	8.0	8.0	8.0
<i>Annie</i>	6.0	7.0	11.0	9.5
<i>Monique</i>	14.5	14.5	15.5	15.0
<i>Didier</i>	14.0	14.0	12.0	12.5
<i>André</i>	11.0	10.0	5.5	7.0
<i>Pierre</i>	5.5	7.0	14.0	11.5
<i>Brigitte</i>	13.0	12.5	8.5	9.5
<i>Evelyne</i>	9.0	9.5	12.5	12.0

### III. Choix d'une distance

Pour faire une représentation géométrique, on doit choisir une distance entre deux points de l'espace.

La distance utilisée par l'ACP dans l'espace  $\mathbb{R}^p$  où sont représentées les unités, est la distance euclidienne classique.

La distance entre deux unités  $u_i$  et  $u_{i'}$  est

$$d^2(u_i, u_{i'}) = \sum_{j=1}^p (x_{ij} - x_{i'j})^2$$

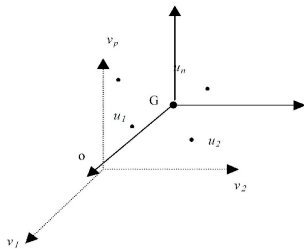
Avec cette distance, toutes les variables jouent le même rôle et les axes définis par les variables constituent une base orthogonale.

## IV. Choix de l'origine

Le point  $O$  correspondant au vecteur de coordonnées toutes nulles n'est pas une origine satisfaisante, car si les coordonnées des points du nuage des individus sont grandes, le nuage est éloigné de cette origine.

On choisit comme origine le point

$$G = \begin{pmatrix} \frac{1}{n} \sum_{i=1}^n x_{i1} \\ \vdots \\ \frac{1}{n} \sum_{i=1}^n x_{ij} \\ \vdots \\ \frac{1}{n} \sum_{i=1}^n x_{ip} \end{pmatrix} = \begin{pmatrix} \bar{x}_1 \\ \vdots \\ \bar{x}_j \\ \vdots \\ \bar{x}_p \end{pmatrix} .$$





## Exemple (Le centre de gravité)

Considérons les données de l'exemple (1)

On a

$$G = \begin{pmatrix} \frac{1}{9} \sum_{i=1}^9 x_{i1} \\ \frac{1}{9} \sum_{i=1}^9 x_{i2} \\ \frac{1}{9} \sum_{i=1}^9 x_{i3} \\ \frac{1}{9} \sum_{i=1}^9 x_{i4} \end{pmatrix} = \begin{pmatrix} 9.67 \\ 9.83 \\ 10.22 \\ 10.06 \end{pmatrix}.$$

Donc le centre de gravité est donné par le vecteur :  $G = \begin{pmatrix} 9.67 \\ 9.83 \\ 10.22 \\ 10.06 \end{pmatrix}$ .

## IV. Choix de l'origine (suite)

On travaille avec le tableau des données centrées :

$$X^c = \begin{pmatrix} x_{11} - \bar{x}_1 & \dots & x_{1p} - \bar{x}_p \\ \vdots & & \vdots \\ x_{i1} - \bar{x}_1 & \dots & x_{ip} - \bar{x}_p \\ \vdots & & \vdots \\ x_{n1} - \bar{x}_1 & \dots & x_{np} - \bar{x}_p \end{pmatrix}.$$

Le vecteur des coordonnées centrées de l'unité  $u_j$  et le vecteur des coordonnées centrées de la variable  $x_j$  sont respectivement :

$$u_j^c = \begin{pmatrix} x_{j1} - \bar{x}_1 \\ x_{j2} - \bar{x}_2 \\ \vdots \\ x_{jp} - \bar{x}_p \end{pmatrix} \in \mathbb{R}^p, \quad x_j^c = \begin{pmatrix} x_{1j} - \bar{x}_j \\ x_{2j} - \bar{x}_j \\ \vdots \\ x_{nj} - \bar{x}_j \end{pmatrix} \in \mathbb{R}^n.$$

## Exemple (Les données centrées)

Considérons les données de l'exemple (1)

$$X^c = \begin{pmatrix} 6.0 - 9.67 & 6.0 - 9.83 & 5.0 - 10.22 & 5.5 - 10.06 \\ 8.0 - 9.67 & 8.0 - 9.83 & 8.0 - 10.22 & 8.0 - 10.06 \\ 6.0 - 9.67 & 7.0 - 9.83 & 11.0 - 10.22 & 9.5 - 10.06 \\ 14.5 - 9.67 & 14.5 - 9.83 & 15.5 - 10.22 & 15.0 - 10.06 \\ 14.0 - 9.67 & 14.0 - 9.83 & 12.0 - 10.22 & 12.5 - 10.06 \\ 11.0 - 9.67 & 10.0 - 9.83 & 5.5 - 10.22 & 7.0 - 10.06 \\ 5.5 - 9.67 & 7.0 - 9.83 & 14.0 - 10.22 & 11.5 - 10.06 \\ 13.0 - 9.67 & 12.5 - 9.83 & 8.5 - 10.22 & 9.5 - 10.06 \\ 9.0 - 9.67 & 9.5 - 9.83 & 12.5 - 10.22 & 12.0 - 10.06 \end{pmatrix}$$

la première ligne de  $X^c$  est :  $X_1^c(-3.67 \quad -3.83 \quad -5.22 \quad -4.56)$

## Exemple (Les données centrées)

Considérons les données de l'exemple (1)

$$X^c = \begin{pmatrix} 6.0 - 9.67 & 6.0 - 9.83 & 5.0 - 10.22 & 5.5 - 10.06 \\ 8.0 - 9.67 & 8.0 - 9.83 & 8.0 - 10.22 & 8.0 - 10.06 \\ 6.0 - 9.67 & 7.0 - 9.83 & 11.0 - 10.22 & 9.5 - 10.06 \\ 14.5 - 9.67 & 14.5 - 9.83 & 15.5 - 10.22 & 15.0 - 10.06 \\ 14.0 - 9.67 & 14.0 - 9.83 & 12.0 - 10.22 & 12.5 - 10.06 \\ 11.0 - 9.67 & 10.0 - 9.83 & 5.5 - 10.22 & 7.0 - 10.06 \\ 5.5 - 9.67 & 7.0 - 9.83 & 14.0 - 10.22 & 11.5 - 10.06 \\ 13.0 - 9.67 & 12.5 - 9.83 & 8.5 - 10.22 & 9.5 - 10.06 \\ 9.0 - 9.67 & 9.5 - 9.83 & 12.5 - 10.22 & 12.0 - 10.06 \end{pmatrix}$$

la première ligne de  $X^c$  est :  $X_1^c(-3.67 \quad -3.83 \quad -5.22 \quad -4.56)$

## V. Correspondance entre statistique et géométrie

Les méthodes factorielles et leurs représentations géométriques utilisent la relation entre géométrie euclidienne et statistique empirique. Les statistiques élémentaires empiriques calculées sur  $n$  unités ont chacune leur correspondant géométrique dans un repère donné.

## V. Correspondance entre statistique et géométrie (suite)

Pour un ensemble quelconque de variables  $x_1, x_2, \dots, x_m$  :

- Variance et carré de la norme :  $Var(x_j) = \frac{1}{n} \sum_{i=1}^n (x_{ij} - \bar{x}_j)^2 = \frac{1}{n} \|\vec{Ox}_j\|^2$
- Covariance et produit scalaire :

$$cov(x_k, x_l) = \frac{1}{n} \sum_{i=1}^n (x_{ik} - \bar{x}_k)(x_{il} - \bar{x}_l) = \frac{1}{n} \langle \vec{Ox}_k, \vec{Ox}_l \rangle$$

- Coefficient de corrélation linéaire et cosinus d'angle :

$$\rho(x_k, x_l) = \frac{cov(x_k, x_l)}{\sigma_{x_k} \sigma_{x_l}} = \frac{\langle \vec{Ox}_k, \vec{Ox}_l \rangle}{\|\vec{Ox}_k\| \|\vec{Ox}_l\|} = \cos(\vec{Ox}_k, \vec{Ox}_l)$$

## Exemple

Considérons les données de l'exemple (1)

$$\overrightarrow{GX_1} = \begin{pmatrix} -3,666 \\ -1,666 \\ -3,666 \\ 4,833 \\ 4,333 \\ 1,333 \\ -4,166 \\ 3,333 \\ -0,666 \end{pmatrix} \quad \text{et} \quad \overrightarrow{GX_2} = \begin{pmatrix} -3,833 \\ -1,833 \\ -2,833 \\ 4,666 \\ 4,166 \\ 0,166 \\ -2,833 \\ 2,666 \\ -0,333 \end{pmatrix} .$$

$$\Rightarrow \frac{1}{9} \|\overrightarrow{GX_1}\|^2 = 11,388 = \text{Var}(X_1)$$

$$\frac{1}{9} \|\overrightarrow{GX_2}\|^2 = 8,944 = \text{Var}(X_2)$$

$$\frac{1}{9} \langle \overrightarrow{GX_1}, \overrightarrow{GX_2} \rangle = 9,916 = \text{Cov}(X_1, X_2)$$

### Définition

On appelle matrice de variance la matrice symétrique  $\Sigma$  contenant les variances  $\text{Var}(x_j)$  sur la diagonale et les covariances  $\text{cov}(x_k, x_l)$  en dehors de la diagonale (ligne  $k$  colonne  $l$  pour  $\text{cov}(x_k, x_l)$ ).

$$\Sigma = \begin{pmatrix} \text{Var}(x_1) & \text{cov}(x_1, x_2) & \dots & \text{cov}(x_1, x_n) \\ \text{cov}(x_2, x_1) & \text{Var}(x_2) & \dots & \text{cov}(x_2, x_n) \\ \vdots & \vdots & & \vdots \\ \text{cov}(x_n, x_1) & \text{cov}(x_n, x_2) & \dots & \text{Var}(x_n) \end{pmatrix}.$$

Cette matrice s'écrit :  $\Sigma = \frac{1}{n}(X^c)^t X^c$



## Exemple (Matrice de variance-covariances)

*Considérons les données de l'exemple (1)*

$$\Sigma = \begin{pmatrix} 11.389 & 9.917 & 2.657 & 4.824 \\ 9.917 & 8.944 & 4.120 & 5.481 \\ 2.657 & 4.120 & 12.062 & 9.293 \\ 4.824 & 5.481 & 9.293 & 7.914 \end{pmatrix}$$

## VI. Matrice de variance-covariances

De même, on définit le coefficient de corrélation linéaire entre les variables  $x_k$  et  $x_l$  par  $\rho(x_k, x_l) = \frac{\text{cov}(x_k, x_l)}{\sigma_{x_k} \sigma_{x_l}}$ .

Ce coefficient exprime le niveau de corrélation (linéaire) entre les variables  $x_k$  et  $x_l$  : plus il est proche de 1, plus les variables sont corrélées positivement, plus il est proche de -1, plus elles sont corrélées négativement. Un coefficient de corrélation nul indique l'absence de corrélation linéaire.

## VI. Matrice de variance-covariances

En divisant chaque colonne  $j$  du tableau centré  $X^c$  par l'écart-type  $\sigma_{x_j}$  de la variable  $x_j$ , on construit le tableau  $Z$  des données centrées réduites :

$z_{ij} = \frac{x_{ij} - \bar{x}_j}{\sigma_{x_j}}$ . La matrice  $Z$  s'exprime en fonction de  $X^c$  par  $Z = D_{1/\sigma} X^c$

où  $D_{1/\sigma}$  est la matrice diagonale contenant  $\frac{1}{\sigma_{x_1}}, \dots, \frac{1}{\sigma_{x_p}}$  sur sa diagonale.

Le terme réduit signifie que les variances des variables  $z_j$  sont égales à 1.

## VI. Matrice de variance-covariances

La matrice  $R = D_{1/\sigma} \Sigma D_{1/\sigma}$  est dite de corrélation. Regroupant les coefficients de corrélation linéaire entre les  $p$  variables prises deux à deux, elle résume la structure des dépendances linéaires entre les  $p$  variables. Elle est symétrique et sa diagonale est composée de 1.

# VII. Moments d'inertie

## 1. Inertie totale du nuage des individus

### Définition

*Le moment d'inertie du nuage des individus par rapport au centre de gravité  $G$  est :*

$$I_G = \frac{1}{n} \sum_{i=1}^n d^2(G, u_i) = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^p (x_{ij} - \bar{x}_j)^2$$

*C'est une mesure de la dispersion du nuage des individus par rapport à son centre de gravité.*

Si ce moment d'inertie est grand, cela signifie que le nuage est très dispersé ; s'il est petit, alors le nuage est très concentré autour de son centre de gravité.

## VII. Moments d'inertie (suite)

### Remarque

$I_G$  peut aussi s'écrire sous la forme suivante :

$$I_G = \sum_{j=1}^p \left[ \frac{1}{n} \sum_{i=1}^n (x_{ij} - \bar{x}_j)^2 \right] = \sum_{j=1}^p \text{var}(x_j),$$

où  $\text{Var}(x_j)$  est la variance de la variable  $x_j$ .

Sous cette forme, l'inertie totale est égale à la trace de la matrice de covariance  $\Sigma$  des  $p$  variables. ( $I_G = \text{trace}(\Sigma)$ )

### Exemple

Considérons les données de l'exemple (1)

$$I_G = \text{trace}(\Sigma) = 11.389 + 8.944 + 12.062 + 7.914 = 40.30$$

## 2. Inertie du nuage des individus par rapport à un axe passant par G

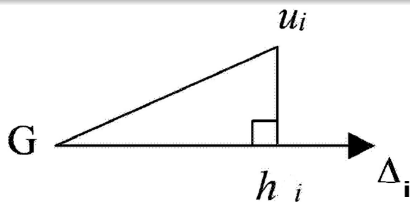
### Définition

L'inertie du nuage des individus par rapport à un axe  $\Delta$  passant par G est égale à :

$$I_{\Delta} = \frac{1}{n} \sum_{i=1}^n d^2(h_i, u_i)$$

où  $h_i$  est la projection orthogonale de  $u_i$  sur l'axe  $\Delta$ .

Cette inertie mesure la proximité à l'axe  $\Delta$  du nuage des individus.



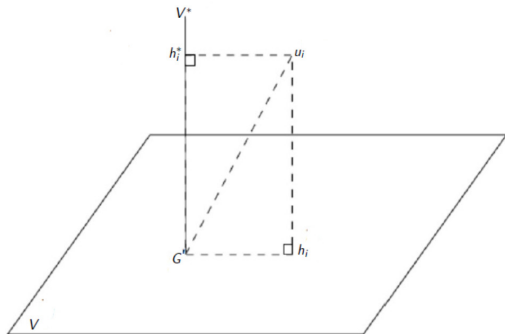
### 3. Inertie du nuage des individus par rapport à un sous-espace vectoriel $V$ passant par $G$

#### Définition

*Inertie du nuage des individus par rapport à un sous-espace vectoriel  $V$  passant par  $G$  est égale à :*

$$I_V = \frac{1}{n} \sum_{i=1}^n d^2(h_i, u_i),$$

*où  $h_i$  est la projection orthogonale de  $u_i$  sur le sous-espace  $V$ .*





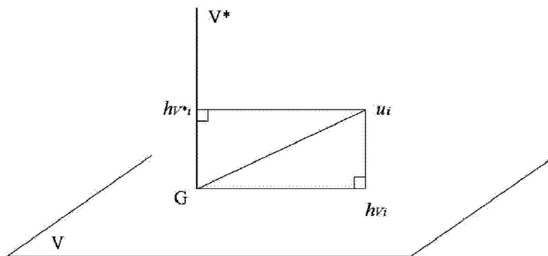
## VII. Moments d'inertie (suite)

### 4. Décomposition de l'inertie totale

Si on note  $V^*$  le complémentaire orthogonal d'un espace vectoriel  $V$  dans  $\mathbb{R}^p$  et  $h_i^*$  la projection orthogonale de  $u_i$  sur  $V^*$ , en appliquant le théorème de Pythagore, on peut écrire :

$$d^2(h_i, u_i) + d^2(h_i^*, u_i) = d^2(G, u_i) = d^2(G, h_i) + d^2(G, h_i^*),$$

où  $h_i$  est la projection orthogonale de  $u_i$  sur  $V$ , et  $h_i^*$  est la projection orthogonale de  $u_i$  sur  $V^*$ . Par exemple dans un espace de dimension 3 :



## VII. Moments d'inertie (suite)

On en déduit que :

$$I_G = I_V + I_{V^*},$$

c'est le théorème de Huygens.

**Cas général :**

### Théorème

*Si on décompose l'espace  $\mathbb{R}^p$  comme la somme de sous-espaces de dimension 1 et orthogonaux entre eux (i.e.  $\mathbb{R}^p = \Delta_1 \oplus \Delta_2 \oplus \dots \oplus \Delta_p$ ), alors*

$$I_G = I_{\Delta_1^*} + I_{\Delta_2^*} + \dots + I_{\Delta_p^*}.$$

## VIII. Dérivation matricielle (Rappel)

### Définition

Soient une forme linéaire  $f : \mathbb{R}^k \rightarrow \mathbb{R}$  et  $x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_k \end{pmatrix} \in \mathbb{R}^k$ .

On appelle dérivée de  $f$  en  $x$  et on note  $\frac{\partial f}{\partial x}$  ou  $\nabla f(x)$  ou encore  $f'(x)$  le vecteur colonne des dérivées partielles de  $f$  par rapport aux  $x_i$  :

$$\frac{\partial f}{\partial x} = \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \vdots \\ \frac{\partial f}{\partial x_k} \end{pmatrix}$$

## VIII. Dérivation matricielle (suite)

### Dérivation de formes linéaires

$$\text{Soit } f : \begin{cases} \mathbb{R}^k & \rightarrow \mathbb{R} \\ x = (x_1, x_2, \dots, x_k)^t & \mapsto \sum_{i=1}^k a_i x_i \end{cases}$$

#### Remarque

$$\sum_{i=1}^k a_i x_i = \langle x, a \rangle = \langle a, x \rangle = a^t x = x^t a$$

#### Proposition

$$\forall a, x \in \mathbb{R}^k \text{ on a } \frac{\partial a^t x}{\partial x} = \frac{\partial x^t a}{\partial x} = a$$

en effet On a  $\forall 1 \leq i \leq k$   $\frac{\partial f}{\partial x_i} = a_i$  donc  $\frac{\partial f}{\partial x} = a$

## VIII. Dérivation matricielle (suite)

### Dérivation d'une forme quadratique

#### Définition

Une forme quadratique est un polynôme homogène de degré 2 avec un nombre quelconque de variables :

$$f : \begin{cases} \mathbb{R}^k & \rightarrow \mathbb{R} \\ x = (x_1, x_2, \dots, x_k)^t & \mapsto \sum_{i=1}^k \sum_{j=1}^k a_{ij} x_i x_j \end{cases}$$

En notant  $A = (a_{ij}) \in \mathcal{M}_{k,k}(\mathbb{R})$ ,  $f$  s'écrit :

$$f : \begin{cases} \mathbb{R}^k & \rightarrow \mathbb{R} \\ x = (x_1, x_2, \dots, x_n)^t & \mapsto x^t A x \end{cases}$$

## VIII. Dérivation matricielle (suite)

### Remarque

$$x^t Ax = \langle x, Ax \rangle = \langle Ax, x \rangle$$

### Proposition

$$\forall x \in \mathbb{R}^k, A = (a_{ij}) \in \mathcal{M}_{k,k}(\mathbb{R}) \quad \frac{\partial x^t Ax}{\partial x} = (A + A^t)x$$

### Remarque

- Si  $A$  est symétrique, i.e : si  $A^t = A$ , on a :  $\frac{\partial x^t Ax}{\partial x} = 2Ax$ .
- En particulier, pour  $A = I_k$ , on a :  $\frac{\partial x^t x}{\partial x} = 2x$ .

# VI. Recherche des axes principaux

## 1. Recherche du premier axe principal $\Delta_1$ passant par $G$ d'inertie minimum

On cherche un axe  $\Delta_1$  passant par  $G$  d'inertie  $I_{\Delta_1}$  minimum car c'est l'axe le plus proche de l'ensemble des points du nuage des individus.

La droite  $\Delta_1$  s'appelle le premier axe principal.

Si on utilise la relation entre les inerties, donnée au paragraphe précédent, rechercher  $\Delta_1$  tel que  $I_{\Delta_1}$  est minimum, est équivalent à chercher  $\Delta_1$  tel que  $I_{\Delta_1^*}$  est maximum.

On définit l'axe  $\Delta_1$  par son vecteur directeur unitaire  $\overrightarrow{Ga_1}$ .

**Problème :**

Trouver  $\overrightarrow{Ga_1}$  tel que  $I_{\Delta_1^*}$  est maximum sous la contrainte que  $\|\overrightarrow{Ga_1}\|^2 = 1$ .

## IX. Recherche des axes principaux (suite)

Expressions algébriques de  $I_{\Delta_1^*}$  et de  $\|\overrightarrow{Ga_1}\|^2$

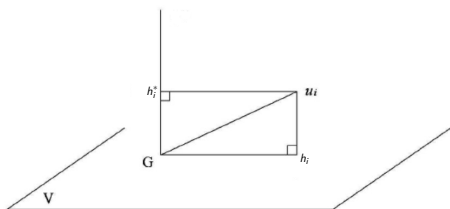
Proposition

$$I_{\Delta_1^*} = a_1^t \Sigma a_1 \quad \text{et} \quad \|\overrightarrow{Ga_1}\|^2 = a_1^t a_1.$$

Avec,  $\Sigma$  est la matrice de variance-covariance de  $x_1, \dots, x_p$ .

**Preuve :** On a

$$\langle \overrightarrow{Gu_i}, \overrightarrow{Ga_1} \rangle = \|\overrightarrow{Ga_1}\| \|\overrightarrow{Gh_i}\| = 1 \times \|\overrightarrow{Gh_i}\| = d(G, h_i) = d(u_i, h_i^*)$$





$$\implies d^2(u_i, h_i^*) = d^2(G, h_i) = \langle \overrightarrow{Gu_i}, \overrightarrow{Ga_1} \rangle^2 = a_1^t u_i^c (u_i^c)^t a_1,$$

par symétrie du produit scalaire.

Ainsi

$$l_{\Delta_1^*} = \frac{1}{n} \sum_{i=1}^n a_1^t u_i^c (u_i^c)^t a_1 = a_1^t \left[ \frac{1}{n} \sum_{i=1}^n u_i^c (u_i^c)^t \right] a_1$$

Par suite

$$l_{\Delta_1^*} = a_1^t \Sigma a_1$$

$$\text{et } \|\overrightarrow{Ga_1}\|^2 = a_1^t a_1.$$

## IX. Recherche des axes principaux (suite)

### Recherche du maximum

Le problème à résoudre : trouver  $a_1$  tel que  $a_1^t \Sigma a_1$  soit maximum avec la contrainte  $a_1^t a_1 = 1$ .

C'est un problème de recherche d'un optimum d'une fonction de plusieurs variables liées par une contrainte (les inconnues sont les composantes de  $a_1$ ).

La méthode des multiplicateurs de Lagrange peut alors être utilisée.

Dans le cas de la recherche de  $a_1$ , il faut calculer les dérivées partielles de :

$$g(a_1) = g(a_{11}, a_{12}, \dots, a_{1p}) = a_1^t \Sigma a_1 - \lambda_1 (a_1^t a_1 - 1).$$

En utilisant la dérivée matricielle, on obtient :

$$\frac{\partial g(a_1)}{\partial a_1} = 2\Sigma a_1 - 2\lambda_1 a_1 = 0.$$

## IX. Recherche des axes principaux (suite)

Le système à résoudre est :

$$\begin{cases} \Sigma a_1 - \lambda_1 a_1 = 0 & (1) \\ a_1^t a_1 - 1 = 0 & (2) \end{cases}$$

De l'équation matricielle (1) de ce système on déduit que  $a_1$  est vecteur propre de la matrice  $\Sigma$  associé à la valeur propre  $\lambda_1$ .

En multipliant à gauche par  $a_1^t$  les deux membres de l'équation (1), on obtient :

$$a_1^t \Sigma a_1 - \lambda_1 a_1^t a_1 = 0$$

et en utilisant l'équation (2) on trouve que :

$$a_1^t \Sigma a_1 = \lambda_1.$$

## IX. Recherche des axes principaux (suite) I

Le premier membre de l'équation précédente est égal à l'inertie  $I_{\Delta_1^*}$  qui doit être maximum. Cela signifie que la valeur propre  $\lambda_1$  est la plus grande valeur propre de la matrice de covariance  $\Sigma$ ; cette valeur propre est égale à l'inertie portée par l'axe  $\Delta_1^*$ .

### Proposition

*L'axe  $\Delta_1$  pour lequel le nuage des individus a l'inertie minimum a comme vecteur directeur unitaire le premier vecteur propre associé à la plus grande valeur propre de la matrice de covariance  $\Sigma$ .*

# IX. Recherche des axes principaux (suite)

## 2. Recherche du deuxième axe principal $\Delta_2$ passant par $G$ , orthogonal à $\Delta_1$ et d'inertie minimum

On recherche ensuite un deuxième axe  $\Delta_2$  orthogonal au premier axe principal  $\Delta_1$  et d'inertie minimum.

La droite  $\Delta_2$  s'appelle le deuxième axe principal.

On peut, comme dans le paragraphe précédent, définir l'axe  $\Delta_2$  passant par  $G$  par son vecteur directeur unitaire  $a_2$ . L'inertie du nuage des individus par rapport à son complémentaire orthogonal est égale à :

$$I_{\Delta_2^*} = a_2^t \Sigma a_2$$

elle doit être maximum avec les deux contraintes suivantes :

$$a_2^t a_2 = 1 \quad \text{et} \quad a_2^t a_1 = 0.$$

## IX. Recherche des axes principaux (suite)

En appliquant la méthode des multiplicateurs de Lagrange, cette fois avec deux contraintes, on trouve que  $a_2$  est le vecteur propre de  $\Sigma$  correspondant à la deuxième plus grande valeur propre. On peut montrer que le plan défini par les axes  $\Delta_1$  et  $\Delta_2$  est le sous-espace de dimension 2 qui génère un inertie minimum du nuage.

# IX. Recherche des axes principaux (suite)

## 3. Recherche des axes principaux suivants

On peut rechercher de nouveaux axes principaux en suivant la même procédure.

Les vecteurs des nouveaux axes principaux sont tous des vecteurs propres de  $\Sigma$  correspondant aux valeurs propres ordonnées. La matrice de covariance  $\Sigma$  étant une matrice symétrique réelle, elle possède  $p$  vecteurs propres réels, formant une base orthogonale de  $\mathbb{R}^p$ .

$$\left\{ \begin{array}{l} \Delta_1 \perp \Delta_2 \perp \dots \perp \Delta_p \\ a_1 \perp a_2 \perp \dots \perp a_p \\ \lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \\ I_{\Delta_1^*} \geq I_{\Delta_2^*} \geq \dots \geq I_{\Delta_p^*} \end{array} \right.$$

On passera de la base orthogonale initiale des variables centrées à la nouvelle base orthogonale des vecteurs propres de  $\Sigma$ .

## X. Contributions des axes à l'inertie totale

On utilise le théorème de Huygens pour décomposer l'inertie totale du nuage des individus

$$I_G = I_{\Delta_1^*} + I_{\Delta_2^*} + \dots + I_{\Delta_p^*} = \lambda_1 + \dots + \lambda_p.$$

### Définition

*La contribution absolue* de l'axe  $\Delta_k$  à l'inertie totale du nuage des individus est égale à :  $ca(\Delta_k/I_G) = \lambda_k$ .

*Sa contribution relative* est égale à :

$$cr(\Delta_k/I_G) = \frac{\lambda_k}{\lambda_1 + \dots + \lambda_p}$$

On emploie souvent l'expression "pourcentage d'inertie expliquée par  $\Delta_k$ ".



## X. Contributions des axes à l'inertie totale (suite)

On peut étendre ces définitions à tous les sous-espaces engendrés par les nouveaux axes. Ainsi, le pourcentage d'inertie expliqué par le plan engendré par les deux premiers axes  $\Delta_1$  et  $\Delta_2$  est égal à :

$$cr(\Delta_1 \oplus \Delta_2 / I_G) = \frac{\lambda_1 + \lambda_2}{\lambda_1 + \dots + \lambda_p}$$

Ces pourcentages d'inertie sont des indicateurs qui rendent compte de la part de variabilité du nuage des individus expliquée par ces sous-espaces. Si les dernières valeurs propres ont des valeurs faibles, on pourra négliger la variabilité qu'expliquent les axes correspondants.

## Exemple (Contributions des axes à l'inertie totale)

Considérons les données de l'exemple (1)

On a  $I_G = \text{tr}(V) = 40.30$ .

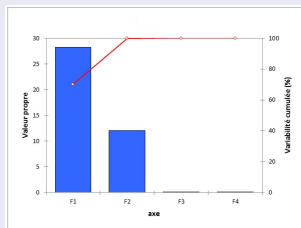
Les deux plus grandes valeurs propres de la matrice  $V$  des variances-covariances sont  $\lambda_1 = 28.253$ ,  $\lambda_2 = 12.03$ .

Donc :  $cr(I_{\Delta_1}/I_G) = \frac{\lambda_1}{I_G} = \frac{28.253}{40.30} = 0.7$  et

$cr(I_{\Delta_2}/I_G) = \frac{\lambda_2}{I_G} = \frac{12.03}{40.30} = 0.3$

$cr(I_{\Delta_1 \oplus \Delta_2}/I_G) = 0.7 + 0.3 = 1 = 100\%$

donc le nuage est pratiquement dans un espace de dimension 2



## Exemple (Contributions des axes à l'inertie totale)

Considérons les données de l'exemple (1)

On a  $I_G = \text{tr}(V) = 40.30$ .

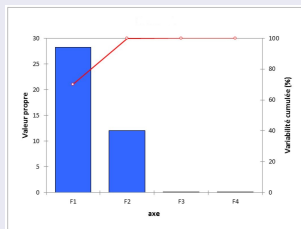
Les deux plus grandes valeurs propres de la matrice  $V$  des variances-covariances sont  $\lambda_1 = 28.253$ ,  $\lambda_2 = 12.03$ .

Donc :  $cr(I_{\Delta_1}/I_G) = \frac{\lambda_1}{I_G} = \frac{28.253}{40.30} = 0.7$  et

$cr(I_{\Delta_2}/I_G) = \frac{\lambda_2}{I_G} = \frac{12.03}{40.30} = 0.3$

$cr(I_{\Delta_1 \oplus \Delta_2}/I_G) = 0.7 + 0.3 = 1 = 100\%$

donc le nuage est pratiquement dans un espace de dimension 2



## Exemple (Contributions des axes à l'inertie totale)

Considérons les données de l'exemple (1)

On a  $I_G = \text{tr}(V) = 40.30$ .

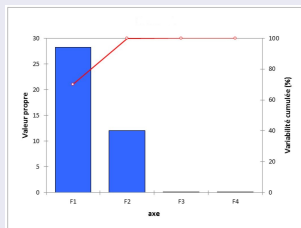
Les deux plus grandes valeurs propres de la matrice  $V$  des variances-covariances sont  $\lambda_1 = 28.253$ ,  $\lambda_2 = 12.03$ .

Donc :  $cr(I_{\Delta_1}/I_G) = \frac{\lambda_1}{I_G} = \frac{28.253}{40.30} = 0.7$  et

$cr(I_{\Delta_2}/I_G) = \frac{\lambda_2}{I_G} = \frac{12.03}{40.30} = 0.3$

$cr(I_{\Delta_1 \oplus \Delta_2}/I_G) = 0.7 + 0.3 = 1 = 100\%$

donc le nuage est pratiquement dans un espace de dimension 2



# VIII. Composantes principales

## 1. Expression des composantes principales

A chaque axe principal est associé une variable appelée "composante principale".

### Définition

*La composante principale  $C_k$  est le vecteur des coordonnées des projections des individus sur l'axe  $\Delta_k$ . Elle s'écrit sous forme de combinaison linéaire des variables initiales (centrée).*

$$C_k = X_c a_k \in \mathbb{R}^n$$

*Ce sont les coordonnées des  $n$  individus sur le  $k$ ème axe principal.*

## Proposition

*La kème composante principale est*

$$C_k = \begin{pmatrix} c_k^1 \\ c_k^2 \\ \vdots \\ c_k^n \end{pmatrix} = \begin{pmatrix} \sum_{j=1}^p x_{1j}^c a_j^k \\ \sum_{j=1}^p x_{2j}^c a_j^k \\ \vdots \\ \sum_{j=1}^p x_{nj}^c a_j^k \end{pmatrix} \in \mathbb{R}^n.$$

# Preuve :

en effet

$$\begin{aligned} C_k &= \begin{pmatrix} c_k^1 \\ c_k^2 \\ \vdots \\ c_k^n \end{pmatrix} = \begin{pmatrix} x_{11}^c & x_{12}^c & \cdots & x_{1p}^c \\ x_{21}^c & x_{22}^c & \cdots & x_{2p}^c \\ \vdots & \vdots & \vdots & \vdots \\ x_{n1}^c & x_{n2}^c & \cdots & x_{np}^c \end{pmatrix} \cdot \begin{pmatrix} a_1^k \\ a_2^k \\ \vdots \\ a_p^k \end{pmatrix} \\ &= \begin{pmatrix} \sum_{j=1}^p x_{1j}^c a_j^k \\ \sum_{j=1}^p x_{2j}^c a_j^k \\ \vdots \\ \sum_{j=1}^p x_{nj}^c a_j^k \end{pmatrix} \end{aligned}$$

# VIII. Composantes principales

## 2. Propriétés des composantes principales

### Proposition

- *Chaque composante principale  $C_k$  est une variable centrée.*
- *La variance d'une composante principale  $C_k$  est égale à l'inertie apportée par l'axe principal qui lui est associé.  
Ainsi,  $\text{var}(C_1) = \lambda_1$ ,  $\text{var}(C_2) = \lambda_2$ ,  $\dots$ ,  $\text{var}(C_p) = \lambda_p$ .*
- *Les composantes principales sont non corrélées deux à deux car les axes principaux associés sont orthogonaux.*



# Preuve :

$$\bar{C}_k = \frac{1}{n} \sum_{i=1}^n c_k^i = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^p x_{ij}^c a_j^k = \frac{1}{n} \sum_{j=1}^p a_j^k \sum_{i=1}^n x_{ij}^c = 0$$

$$\text{Var}(C_k) = \frac{1}{n} C_k^t C_k = \frac{1}{n} (X^c \cdot a_k)^t (X^c \cdot a_k) = \frac{1}{n} a_k^t (X^c)^t X^c \cdot a_k = \lambda_k$$

$$\begin{aligned} \text{cov}(C_k, C_l) &= \langle C_k, C_l \rangle = C_k^t C_l = (X^c \cdot a_k)^t \cdot (X^c \cdot a_l) \\ &= a_k^t \cdot (X^c)^t \cdot X^c \cdot a_l = n a_k^t \cdot \Sigma \cdot a_l = n a_k^t \cdot \lambda_l a_l \\ &= n \lambda_l a_k^t \cdot a_l = 0 \\ &\implies \rho(C_k, C_l) = 0 \\ &\implies C_k \perp C_l \end{aligned}$$

## 1. Coordonnées des individus

Dans la nouvelle base  $(a_1, \dots, a_p)$ , les coordonnées de l'individu  $i$  sont

$$\begin{pmatrix} c_1^i \\ c_2^i \\ \vdots \\ c_p^i \end{pmatrix} \in \mathbb{R}^p.$$

Ici,  $c_j^i$  est le  $i$ ème terme de la composante principale  $C_j$ .

## XII. Représentation des individus

### 2. Qualité de la représentation des individus et l'étude de la proximité entre les individus

Si les projections des individus sont éloignés sur un axe (ou sur un plan), alors les points représentant ces individus sont éloignés dans l'espace.

Cependant, deux individus dont les projections sont proches sur un axe (ou sur un plan) peuvent ne pas être proches dans l'espace.

Pour interpréter correctement la proximité des projections de deux individus sur un plan, il faut s'assurer que ces individus sont bien représentés dans l'axe (ou dans le plan). Pour que l'individu  $i$  soit bien représenté sur un axe (ou sur un plan, ou un sous-espace), il faut que l'angle entre le vecteur  $\overrightarrow{Gu_i}$  et l'axe (ou le plan, ou le sous-espace) soit petit (le cosinus est proche de 1).

## XII. Représentation des individus

### Représentation par rapport à un axe

#### Proposition

Le carré du cosinus de l'angle  $\alpha_{ik}$  entre  $\overrightarrow{Gu_i}$  et un axe  $\Delta_k$  de vecteur directeur unitaire  $a_k$  est égal à :

$$\cos^2(\alpha_{ik}) = \frac{\langle \overrightarrow{Gu_i}, \overrightarrow{Ga_k} \rangle^2}{\|\overrightarrow{Gu_i}\|^2} = \frac{a_k^t u_i^c (u_i^c)^t a_k}{(u_i^c)^t u_i^c}.$$

## XII. Représentation des individus

### Représentation par rapport à un plan

En utilisant le théorème de Pythagore, on peut montrer que le carré du cosinus de l'angle d'un vecteur avec un plan engendré par deux vecteurs orthogonaux, est égal à la somme des carrés des cosinus des angles du vecteur avec chacun des deux vecteurs qui engendrent le plan. Ainsi, le carré du cosinus de l'angle  $\alpha_{ikk'}$  entre  $\overrightarrow{Gu_i}$  et le plan engendré par deux axes  $\Delta_k$  et  $\Delta_{k'}$  est

$$\cos^2 \alpha_{ikk'} = \cos^2 \alpha_{ik} + \cos^2 \alpha_{ik'}$$

### Remarque :

Cette propriété se généralise à l'angle d'un vecteur avec un sous-espace de dimension quelconque ( $\geq 2$ ).

## XII. Représentation des individus

### Interprétation de la Représentation

Si le carré du cosinus de l'angle entre  $\vec{Gu}_i$  et l'axe (ou le plan, ou le sous-espace) est proche de 1, alors l'individu  $i$  est bien représenté par sa projection sur l'axe (ou le plan, ou le sous-espace). Et si deux individus sont bien représentés en projection sur un axe (ou un plan, ou un sous-espace) et ont des projections proches, alors ils sont proches dans l'espace.

**Remarque :** Si un individu est très proche du centre de gravité dans l'espace, c'est-à-dire si  $\|\vec{Gu}_i\|^2$  est très petit, le point représentant cet individu sur un axe (ou un plan, ou un sous-espace) sera bien représenté.

### 3. Interprétation des axes principaux en fonction des individus

Un individu contribue à la confection d'un axe lorsque sa projection sur cet axe sera éloignée du centre de gravité du nuage. Inversement, un individu dont la projection sur un axe sera proche du centre de gravité contribue faiblement à l'inertie portée par cette axe.

## 1. Coordonnées des variables

Soient  $C_1, \dots, C_p$  les composantes principales. Il est intéressant de voir comment les anciennes variables  $x_j^c$  sont liées à ces composantes principales  $C_k$ ; on calcule alors les corrélations entre  $C_k$  et  $x_j^c$ .

La représentation des anciennes variables se fera en prenant comme coordonnées leurs coefficients de corrélation avec les composantes principales. On obtient alors ce que l'on appelle le "cercle des corrélations".



## Proposition

*On montre que le coefficient de corrélation entre  $C_k$  et  $x_j^c$  est*

$$\text{cor}(C_k, x_j^c) = \sqrt{\lambda_k} \frac{a_{kj}}{\sqrt{\text{var}(x_j)}},$$

*où  $a_{kj}$  est la  $j$ ème coordonnée du vecteur directeur unitaire  $a_k$  de  $\Delta_k$ .*

**Preuve :** On a  $x_j^c = x^c \cdot e_j$  avec  $e_j$  est le  $j$ ème vecteur de la base canonique de  $\mathbb{R}^n$ .

# XIII. Représentation des variables

Donc

$$\begin{aligned}\text{Cov}(C_k, x_j^c) &= \frac{1}{n} \langle C_k, x_j^c \rangle = \frac{1}{n} C_k^t \cdot x_j^c \\ &= \frac{1}{n} a_k^t (X^c)^t X^c e_j = a_k^t \Sigma e_j \\ &= \lambda_k a_k^t e_j = \lambda_k a_{kj}\end{aligned}$$

$$\begin{aligned}\implies \rho(C_k, x_j^c) &= \frac{\text{Cov}(C_k, x_j^c)}{\sqrt{\text{var}(C_k)} \sqrt{\text{var}(x_j)}} \\ &= \frac{\lambda_k a_{kj}}{\sqrt{\lambda_k} \sqrt{\text{var}(x_j)}} \\ &= \sqrt{\lambda_k} \frac{a_{kj}}{\sqrt{\text{var}(x_j)}}\end{aligned}$$

## 2. Qualité de la représentation des variables

Une variable sera d'autant mieux représentée sur un axe que sa corrélation avec la composante principale correspondante est en valeur absolue proche de 1.

Ainsi, une variable sera bien représentée sur un plan si elle est proche du bord du cercle des corrélations.

## 3. Interprétation des axes principaux en fonction des anciennes variables

Une variable  $x_j$  explique d'autant mieux un axe principal qu'elle est fortement corrélée avec la composante principale correspondant à cet axe.

## 4. Etude des liaisons entre les variables

Sur le graphique du cercle des corrélations, on peut aussi interpréter les positions des anciennes variables les unes par rapport aux autres en termes de corrélations.

- Deux points très proches entre elles et très proches du cercle des corrélations, donc bien représentées dans le plan, seront très corrélées positivement entre elles.
- Si elles sont proches du cercle, mais dans des positions symétriques par rapport à l'origine, elles seront très corrélées négativement.
- Deux variables proches du cercle des corrélations et dont les vecteurs qui les joignent à l'origine forment un angle droit, ne seront pas corrélées entre elles.

A partir des relations données précédemment, nous pouvons définir quelques règles pour l'interprétation :

- Il est naturel de commencer l'examen détaillé des graphiques par les variables parce qu'elles sont moins nombreuses et plus chargée de sens que les individus
- Interprétation axe par axe : Interpréter un axe factoriel consiste à donner un " sens " à l'axe en recensant les variables les plus liées à chaque axe. Deux situations typiques peuvent se produire :
  - ① Toutes les variables très liées au facteur sont situées d'un même coté de l'axe. Le facteur apparaît alors comme une synthèse entre ces variables.
  - ② les variables très liées au facteur présentent une coordonnée positive pour les unes et négative pour les autres. Il faut alors rechercher un dénominateur commun qui, à la fois, relie les variables situées de même côté et oppose les variables situées de part et d'autre de l'origine.

Par exemple, supposons que les variables soient des notes dans différentes matières : un facteur peut traduire l'opposition entre matières scientifiques et matières littéraire.

- Interprétation par plans : Le plans factoriel apport le pouvoir synthétique du graphique, et la prise en compte simultanée de deux dimensions qui donne une image plus fidèle des données et peut aussi suggérer d'interpréter d'autre directions que les axes factoriels. Il est utile de représenter en plus des points (variables) : " Le cercle de rayons 1, ou cercle de corrélations. " Les vecteurs joignant l'origine aux points variables afin de visualiser les angles qui mesurent la liaison entre variables.

# Interprétation (suite)

- Un individu sera du côté des variables pour lesquelles il a de fortes valeurs, inversement il sera du côté opposé des variables pour lesquelles il a de faibles valeurs.
- Plus les valeurs d'un individu sont fortes pour une variable plus il sera éloigné de l'origine suivant l'axe factoriel décrivant le mieux cette variable.
- Deux individus à une même extrémité d'un axe (i.e. éloignés de l'origine) sont proches (i.e. se ressemblent).
- Deux variables très corrélées positivement sont du même côté sur un axe.
- Il n'est pas possible d'interpréter la position d'un individu par rapport à une seule variable, et réciproquement, il n'est pas possible d'interpréter la position d'une variable par rapport à un seul individu. Les interprétations doivent se faire de manière globale.

Jusqu'à présent, on a étudié l'ACP simple, pour laquelle, tous les individus ont le même poids et toutes les variables sont traitées de façon symétrique (elles jouent le même rôle) et les axes principaux sont issus de la matrice de covariance des variables. Cela pose parfois certains problèmes :

- le premier problème : les anciennes variables sont hétérogènes, comme par exemple des poids, des tailles et des âges ;
- le deuxième problème : si on change d'unités sur ces variables, on peut changer complètement les résultats de l'ACP ;
- le dernier problème : une variable contribue d'autant plus à la confection des premiers axes que sa variance est forte.



## XIV. ACP normée (suite)

Pour échapper à tous ces problèmes, on travaille sur des variables centrées et réduites.

Cela revient à faire la même analyse que pour l'ACP simple, mais à choisir une autre distance euclidienne entre les individus que la distance euclidienne classique. La distance choisie est :

$$d^2(u_i, u_{i'}) = \sum_{j=1}^p \frac{1}{\sigma_j^2} (x_{ij} - x_{i'j})^2$$

## XIV. ACP normée (suite)

Si on reprend tous les calculs de l'ACP simple, mais en remplaçant les variables de départ par les variables centrées réduites, on voit que ce n'est plus la matrice de covariance, mais la matrice de corrélation  $R$  qui intervient pour la recherche des nouveaux axes.

Puisque la matrice de corrélation  $R$  n'a que des 1 sur sa diagonale principale, sa trace est égale à  $p$ . Par suite, l'inertie totale du nuage des individus dans  $\mathbb{R}^p$  est égale à  $p$ .

## XV. Individus et variables supplémentaires

Il arrive que l'on veuille faire apparaître dans les représentations graphiques certains individus sans qu'ils interviennent dans la détermination des axes. Cela peut être le cas de nouveaux individus que l'on veut simplement positionner par rapport aux autres sans que les positions de ceux-ci soient influencées par les nouveaux. On dit d'eux qu'ils sont des individus supplémentaires.

Il en est de même pour les variables. On peut, par exemple, vouloir représenter une variable qui dépend de façon synthétique des  $p$  variables choisies pour faire l'ACP, afin de mieux comprendre comment cette variable est liée aux anciennes, mais on ne souhaite pas qu'elle intervienne dans la confection des axes car ses liaisons avec les  $p$  variables de départ fausseraient la représentation si elle faisait partie intégrante de l'ACP. Elles sont appelées variables supplémentaires.

Pour représenter un individu supplémentaire, il suffit d'exprimer les coordonnées de cet individu dans la nouvelle base des axes principaux. Pour une variable supplémentaire, il suffit de calculer ses coefficients de corrélation avec les composantes principales.

## XVI. Récapitulatif : démarche d'une ACP

- Tableau de données ( $n$  unités/ $p$  variables quantitatives).
- Calcul de la matrice de covariance (pour une ACP simple) ou la matrice de corrélation (pour une ACP normée).
- Recherche des valeurs propres  $\lambda_1, \dots, \lambda_p$ .
- Recherche des vecteurs propres  $a_1 \in \mathbb{R}^n, \dots, a_p \in \mathbb{R}^n$ .
- Calcul du pourcentage d'inertie ou de contribution de l'axe à la variation totale (pourcentage expliqué par les axes principaux).
- Détermination des composantes principales  $C_1 \in \mathbb{R}^n, \dots, C_p \in \mathbb{R}^n$ .

## XV. Récapitulatif : démarche d'une ACP(suite)

- Etude du nuage des variables : Coordonnées des variables sur les axes principaux (corrélation entre les variables et les composantes principales) ; qualité de la représentation des variables, étude de liaison entre les variables.
- Etude du nuage des individus : Coordonnées des individus sur les axes principaux, qualité de la représentation des individus, étude de la proximité.