ANALYSE EN COMPOSANTES PRINCIPALES (ACP)

PLAN

Notations du cours

Un vecteur est noté par une lettre latine en caractère gras. Il est toujours représenté en colonne.

$$\mathbf{x} = \begin{pmatrix} x_1 \\ \dots \\ x_i \\ \dots \\ x_n \end{pmatrix} \in \mathbb{R}^n.$$

Si l'on veut représenter un vecteur ligne, on écrit ${}^t\mathbf{x}$ (transposé de $\mathbf{x}$).

Une matrice est en général notée par une lettre capitale, par exemple X.

$$X = \begin{pmatrix} x_1^1 & \dots & x_1^j & \dots & x_1^p \\ \dots & \dots & \dots & \dots & \dots \\ x_i^1 & \dots & x_i^j & \dots & x_i^p \\ \dots & \dots & \dots & \dots & \dots \\ x_n^1 & \dots & x_n^j & \dots & x_n^p \end{pmatrix}$$

$\mathbf{x}_i^j$ valeur de la colonne j pour la ligne i.

Dimensions de X notées $n \times p$ soit $\underset{n\times p}{X}$

Un document extrait du rapport FPS01 "*Algèbre matricielle*" de J.Badia contient quelques rappels (Annexe 1). Ce document sera donné aux stagiaires. Il constitue un prérequis souhaitable. Toutefois, le cours rappellera toutes les notions algébriques nécessaires.

La définition de la multiplication matricielle sera présentée au tableau par le formateur.

- En analyse des données la colonne d'un tableau est associée à une variable, la ligne à un individu statistique.
- **Ne pas confondre** l'indice de colonne avec la puissance, par exemple, le carré de x_i^j est noté $(x_i^j)^2$.

On note $\mathbf{x}^j$ la j$\grave{e}me$ colonne de X : $\mathbf{x}^j = \begin{pmatrix} x_1^j \\ \dots \\ x_i^j \\ \dots \\ x_n^j \end{pmatrix}$, on peut écrire X par colonnes :

$$X = \left(\begin{array}{c|c|c|c|c} x_1^1 & \dots & x_1^j & \dots & x_1^p \\ \dots & \dots & \dots & \dots & \dots \\ x_i^1 & \dots & x_i^j & \dots & x_i^p \\ \dots & \dots & \dots & \dots & \dots \\ x_n^1 & \dots & x_n^j & \dots & x_n^p \end{array}\right) = \left(\begin{array}{c|c|c|c|c} \mathbf{x}^1 & \dots & \mathbf{x}^j & \dots & \mathbf{x}^p \end{array}\right)$$

En notant ${}^t\mathbf{x}_i$ la i$\grave{e}me$ ligne de X, on a :

$$X = \left(\begin{array}{ccccc} x_1^1 & \dots & x_1^j & \dots & x_1^p \\ \hline \dots & \dots & \dots & \dots & \dots \\ \hline x_i^1 & \dots & x_i^j & \dots & x_i^p \\ \hline \dots & \dots & \dots & \dots & \dots \\ \hline x_n^1 & \dots & x_n^j & \dots & x_n^p \end{array}\right) = \left(\begin{array}{c} {}^t\mathbf{x}_1 \\ \hline \dots \\ \hline {}^t\mathbf{x}_i \\ \hline \dots \\ \hline {}^t\mathbf{x}_n \end{array}\right)$$

Exemple : notes scolaires

$\underset{9\times5}{X} =$

nom	m	s	f	l	dm
JE	6.0	6.0	5.0	5.5	8.0
AL	8.0	8.0	8.0	8.0	9.0
AN	6.0	7.0	11.0	9.5	11.0
MO	14.5	14.5	15.5	15.0	8.0
DI	14.0	14.0	12.0	12.5	10.0
AD	11.0	10.0	5.5	7.0	13.0
PI	5.5	7.0	14.0	11.5	10.0
BR	13.0	12.5	8.5	9.5	12.0
EV	9.0	9.5	12.5	12.0	18.0

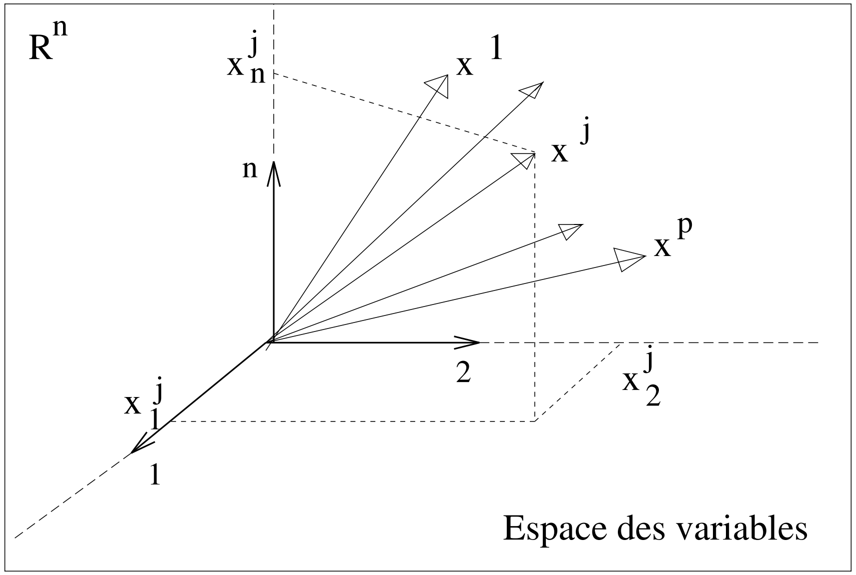
Les lignes du tableau sont associées aux élèves ou *individus*.
Les colonnes du tableau sont associées aux notes ou *variables*.

Jeu de données utilisé dans les applications numériques.

- Les individus (élèves) sont notés par un nom abrégé.
- Les variables sont les notes obtenues dans des matières scolaires,
  m : maths
  s : sciences
  f : français
  l : langue étrangère
  dm : dessin-musique
- JE $= \mathbf{x}_1 = \begin{pmatrix} 6 \\ 6 \\ 5 \\ 5.5 \\ 8 \end{pmatrix} \in \mathbb{R}^5, {}^t\mathbf{x}_1 = \begin{pmatrix} 6 & 6 & 5 & 5.5 & 8 \end{pmatrix}$
- $s = x^2$ deuxième colonne de X, vecteur de $\mathbb{R}^9$.
- Dans tout le cours, et dans la mesure du possible, un indice en exposant fait référence à un vecteur de $\mathbb{R}^n$, un indice bas fait référence à un vecteur de $\mathbb{R}^p$. Dans l'exemple $n = 9$ et $p = 5$.
- Bien noter qu'un vecteur est toujours écrit en colonne.

Nuage des colonnes

$$
\begin{array}{cc}
 & \begin{array}{ccccc} & & \mathbf{x}^j & & \end{array} \\
X = & \left(\begin{array}{c|c|c|c|c}
x_1^1 & \dots & x_1^j & \dots & x_1^p \\
\dots & \dots & \dots & \dots & \dots \\
x_i^1 & \dots & x_i^j & \dots & x_i^p \\
\dots & \dots & \dots & \dots & \dots \\
x_n^1 & \dots & x_n^j & \dots & x_n^p
\end{array}\right)
\end{array}
$$



- $\underset{n\times 1}{I^1} = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \underset{n\times 1}{I^2} = \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix}, \underset{n\times 1}{I^n} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix},$

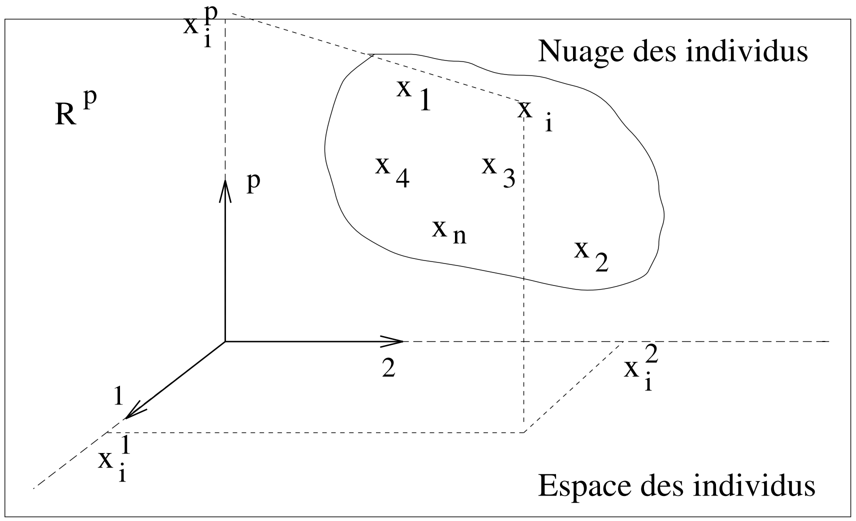
  sont les n vecteurs unitaires engendrant l'espace des variables $\mathbb{R}^n$. Ils forment la **base canonique** de $\mathbb{R}^n$. Sur le graphique un axe est associé à chacun d'eux.

  écrire au tableau la décomposition de x^j dans cette base :
  $\mathbf{x}^j = x_1^j I^1 + \ldots + x_i^j I^i + \ldots + x_n^j I^n$
  les coordonnées des points sont donc les valeurs prises par les variables pour les n individus. En cas de donnée manquante, le vecteur n'est pas défini.

- Par convention, une variable est représentée par un segment fléché.

Nuage des lignes

$$X = \begin{pmatrix} x_1^1 & \ldots & x_1^j & \ldots & x_1^p \\ \hline \ldots & \ldots & \ldots & \ldots & \ldots \\ \hline x_i^1 & \ldots & x_i^j & \ldots & x_i^p \\ \hline \ldots & \ldots & \ldots & \ldots & \ldots \\ \hline x_n^1 & \ldots & x_n^j & \ldots & x_n^p \end{pmatrix}$$



- $\underset{p\times 1}{J_1} = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \underset{p\times 1}{J_2} = \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix}, \underset{p\times 1}{J_p} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}$

forment la base canonique de $\mathbb{R}^p$.

Chaque ligne $\mathbf{x}_i$ s'écrit dans cette base :

$x_i = x_i^1 J_1 + \ldots + x_i^p J_p$

Chaque J_k définit un axe.

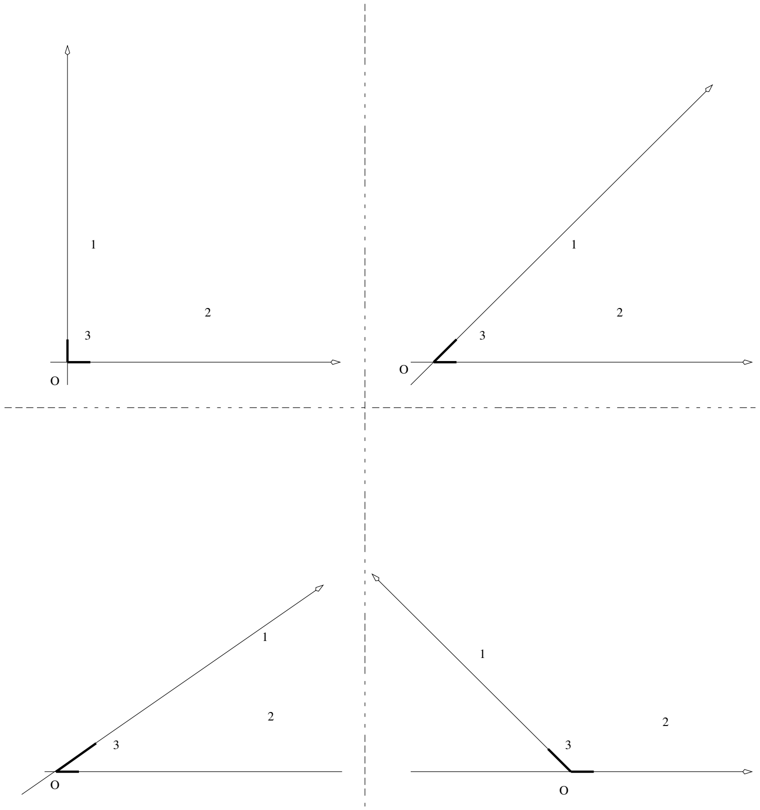
Les coordonnées des points sont les valeurs prises par les individus pour chaque variable. Par convention, un individu est représenté par un point.

Il existe bien sûr d'autres bases de $\mathbb{R}^n$ et $\mathbb{R}^p$.

Incidence du choix de la base

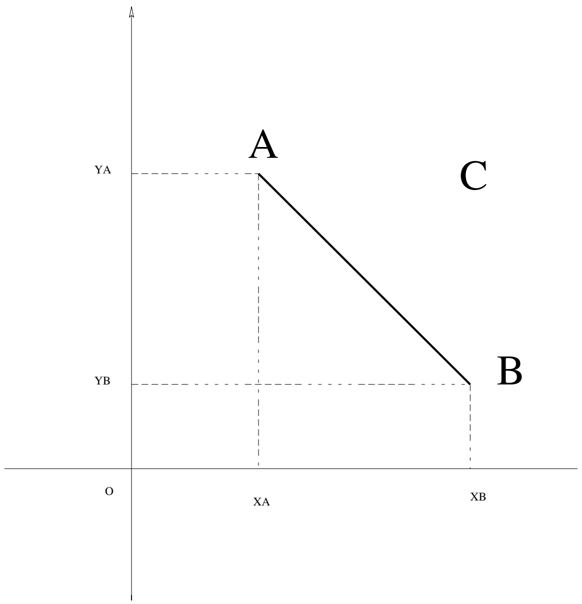
Soient les points :

$$1 = \begin{pmatrix} 1 \\ 5 \end{pmatrix}, 2 = \begin{pmatrix} 6 \\ 2 \end{pmatrix} \text{ et } 3 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$



Des bases différentes induisent des conclusions différentes. Le graphique doit être cohérent avec la perception visuelle. L'œil juge la distance entre deux points de manière équivalente quelle que soit la direction du bipoint dans l'espace. Le repère ou base qui correspond le mieux à notre perception visuelle est orthonormé. Il sera choisi dans les constructions graphiques des analyses multidimensionnelles.

Distance euclidienne usuelle dans $\mathbb{R}^2$



- $d(A, B) = \sqrt{(x_A - x_B)^2 + (y_A - y_B)^2}$

- Propriétés d'une distance :

  $d(A, B) > 0$ si $A \neq B$, $d(A, B) = 0$ si $A = B$

  $d(A, B) = d(B, A)$ symétrie

  $d(A, C) \leq d(A, B) + d(B, C)$ inégalité triangulaire

- La distance usuelle est appelée aussi distance de l'arpenteur.
- D'autres distances dans $\mathbb{R}^2$ vérifient ces propriétés.
- La distance usuelle à vol d'oiseau entre deux lieux A et B est une distance euclidienne. Sur une carte routière, le chemin le plus court pour aller de A en B (ou de B en A) n'est pas une distance euclidienne.

Distance euclidienne dans $\mathbb{R}^p$

- $\mathbf{x}_i$ et $\mathbf{x}_k$ deux points de $\mathbb{R}^p$
- $d_M(\mathbf{x}_i, \mathbf{x}_k) = \sqrt{\sum_{j=1}^{p} m_j (x_i^j - x_k^j)^2}$
- Les poids m_j sont positifs et définissent une métrique M dans l'espace des lignes de X.

Application au tableau des notes :

si $m_j = 1$ pour tout j :
$d(JE, EV) = 14.82, d(JE, AD) = 8.27, d(AD, EV) = 10.16$

si $m_1 = m_2 = m_3 = m_4 = 1$ et $m_5 = 3$:
$d(JE, EV) = 20.48, d(JE, AD) = 10.88, d(AD, EV) = 12.37$

On l'appelle aussi distance pondérée.

M est la matrice diagonale $(p \times p)$ d'éléments diagonaux m_j. M désigne à la fois la matrice et la métrique définie par la matrice.

Distance et métrique sont synonymes.

Distance euclidienne dans $\mathbb{R}^n$

- $\mathbf{x}^j$ et $\mathbf{x}^k$ deux points de $\mathbb{R}^n$ (2 variables)
- $d_D(\mathbf{x}^j, \mathbf{x}^k) = \sqrt{\sum_{i=1}^{n} p_i (x_i^j - x_i^k)^2}$
- Les poids p_i sont positifs **et** vérifient $\sum_{i=1}^{n} p_i = 1$
- d_D définit la distance de 2 colonnes de X.

C'est encore une distance pondérée, la somme des poids fait 1.

D est la matrice $(n \times n)$ d'éléments diagonaux p_i.

Distance de deux matrices

- $\underset{n\times p}{X}$ et $\underset{n\times p}{Y}$ deux matrices de même dimension.

- $d(X,Y)^2 = \sum_{j=1}^{p} m_j d_D(\mathbf{x}^j, \mathbf{y}^j)^2$
  (somme des carrés des distances entre les variables)

  $= \sum_{i=1}^{n} p_i d_M(\mathbf{x}_i, \mathbf{y}_i)^2$
  (somme des carrés des distances entre les individus)

  $= \sum_{i=1}^{n} \sum_{j=1}^{p} p_i m_j (x_i^j - y_i^j)^2$
  (somme des carrés des distances entre les éléments)

Notion de dissemblance entre matrices

Avec :

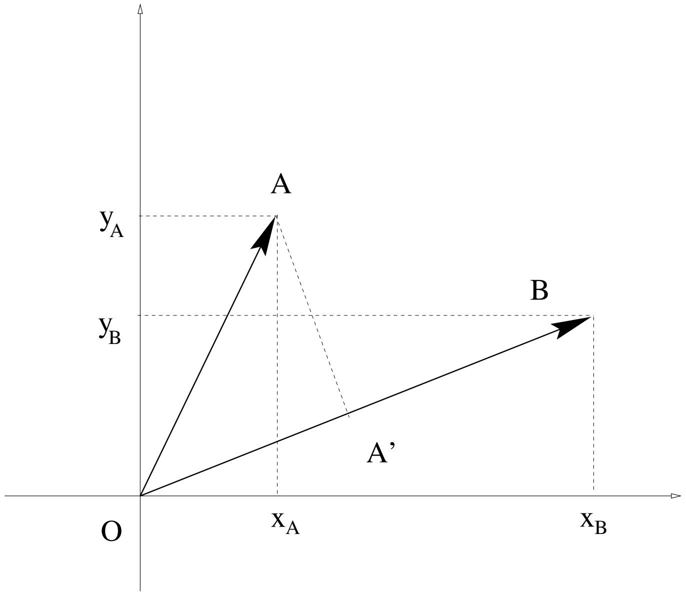
$$d_D(\mathbf{x}^j, \mathbf{y}^j) = \sqrt{\sum_{i=1}^{n} p_i (x_i^j - y_i^j)^2}$$

$$d_D(\mathbf{x}^j, \mathbf{y}^j)^2 = \sum_{i=1}^{n} p_i (x_i^j - y_i^j)^2$$

on trouve la $3ème$ formule :

$$d(X,Y)^2 = \sum_{j=1}^{p} m_j \sum_{i=1}^{n} p_i (x_i^j - y_i^j)^2$$

• **Produit scalaire des vecteurs OA et OB :** $< OA, OB >$



$< OA, OB > = x_A x_B + y_A y_B$

- **Norme de OA** : $\|OA\| = \sqrt{< OA, OA >}$
- $\cos\theta = \dfrac{\|OA'\|}{\|OA\|} = \dfrac{< OA, OB >}{\|OA\| \|OB\|}$
- De manière plus générale, $< OA, OB >_Q = q_1 x_A x_B + q_2 y_A y_B$
  Q, matrice diagonale 2×2 formée avec q_1 et q_2 **positifs**, est la métrique qui définit le produit scalaire.

Produit scalaire usuel

- Si $\theta \in \left[0, \dfrac{\pi}{2}\right]$, on a $< OA, OB >= \|OA'\| \times \|OB\|$
  Si $\theta \in \left[\dfrac{\pi}{2}, \pi\right]$, on a $< OA, OB >= -\|OA'\| \times \|OB\|$
- Les vecteurs étant non nuls, $< OA, OB >= 0$ si et seulement si OA et OB sont orthogonaux.

Produit scalaire dans $\mathbb{R}^p$, espace des individus

Pour tout $\mathbf{x} = \begin{pmatrix} x^1 \\ \vdots \\ x^p \end{pmatrix} \in \mathbb{R}^p$, et pour tout $\mathbf{y} = \begin{pmatrix} y^1 \\ \vdots \\ y^p \end{pmatrix} \in \mathbb{R}^p$,

on note $< \mathbf{x}, \mathbf{y} >_M = \sum_{j=1}^{p} m_j x^j y^j$

$\sqrt{< \mathbf{x}, \mathbf{x} >_M}$ est une norme dite **euclidienne**, notée $\|\mathbf{x}\|_M$:
$\|\mathbf{x}\|_M = \sqrt{\sum_{j=1}^{p} m_j (x^j)^2}$

- Écriture matricielle : $< \mathbf{x}, \mathbf{y} >_M = {}^t\mathbf{x} M \mathbf{y}$

- Propriétés : $\mathbf{x}$, $\mathbf{y}$, $\mathbf{z}$ vecteurs de $\mathbb{R}^p$, α réel

  $< \alpha\mathbf{x}, \mathbf{y} >_M = \alpha < \mathbf{x}, \mathbf{y} >_M$

  $< \mathbf{x}, \mathbf{y} >_M = < \mathbf{y}, \mathbf{x} >_M$ (symétrie)

  $< \mathbf{x} + \mathbf{y}, \mathbf{z} >_M = < \mathbf{x}, \mathbf{z} >_M + < \mathbf{y}, \mathbf{z} >_M$

- La norme d'un vecteur de $\mathbb{R}^p$ est définie avec la matrice M associée à la métrique de cet espace.
- Plus généralement, un produit scalaire peut être défini à l'aide d'une matrice Q symétrique et définie positive (i.e. ${}^t\mathbf{x}Q\mathbf{x} > 0$, pour tout $\mathbf{x}$ non nul).

  À une norme euclidienne est associée une **distance** euclidienne définie par :
  $$d_M(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|_M$$
  On retrouve alors la distance euclidienne de $\mathbb{R}^p$:
  $$d_M(\mathbf{x}, \mathbf{y}) = \sqrt{\sum_{j=1}^{p} m_j (x^j - y^j)^2}$$

Produit scalaire dans $\mathbb{R}^n$, espace des variables

$\mathbf{x}$, $\mathbf{y}$ vecteurs de $\mathbb{R}^n$.

Les n poids p_i positifs vérifiant $\sum_{i=1}^{n} p_i = 1$ définissent une matrice diagonale appelée métrique D.

$$< \mathbf{x}, \mathbf{y} >_D = \sum_{i=1}^{n} p_i x_i y_i = {}^t\mathbf{x} D \mathbf{y}$$

Norme associée : $\|\mathbf{x}\|_D = \sqrt{\sum_{i=1}^{n} p_i (x_i)^2}$

Distance associée : $d_D(\mathbf{x}, \mathbf{y}) = \sqrt{\sum_{i=1}^{n} p_i (x_i - y_i)^2}$

- Noter la différence entre poids sur ligne et colonne.
- Si l'on n'a pas d'*a priori*, on choisit $p_i = \dfrac{1}{n}$. Le poids caractérise l'importance relative de chaque individu dans l'échantillon.

Exemple : notes en Math et sciences et notes centrées :

nom	m	s
JE	6.0	6.0
AL	8.0	8.0
AN	6.0	7.0
MO	14.5	14.5
DI	14.0	14.0
AD	11.0	10.0
PI	5.5	7.0
BR	13.0	12.5
EV	9.0	9.5

nom	m.c	s.c
JE	-3.7	-3.8
AL	-1.7	-1.8
AN	-3.7	-2.8
MO	4.8	4.7
DI	4.3	4.2
AD	1.3	0.2
PI	-4.2	-2.8
BR	3.3	2.7
EV	-0.7	-0.3

Si $p_i = \frac{1}{9}$, i = 1, ..., 9

$$< m, s >_D = \frac{1}{9}(6 \times 6 + 8 \times 8 + 6 \times 7 + 14.5 \times 14.5 + 14 \times 14 + 11 \times 10 + 5.5 \times 7 + 13 \times 12.5 + 9 \times 9.5) = 104.97$$

$$< m.c, s.c >_D = \frac{1}{9}((-3.7) \times (-3.8) + (-1.7) \times (-1.8) + (-3.7) \times (-2.8) + (4.8) \times (4.7) + (4.3) \times (4.2) + (1.3) \times (0.2) + (-4.2) \times (-2.8) + (3.3) \times (2.7) + (-0.7) \times (-0.3)) = 9.92$$

C'est la covariance de ces deux variables (car les variables sont centrées).

- Écrire au tableau la matrice diagonale $D = diag(1/9)$.
- L'interprétation statistique du produit scalaire sera développée plus loin.

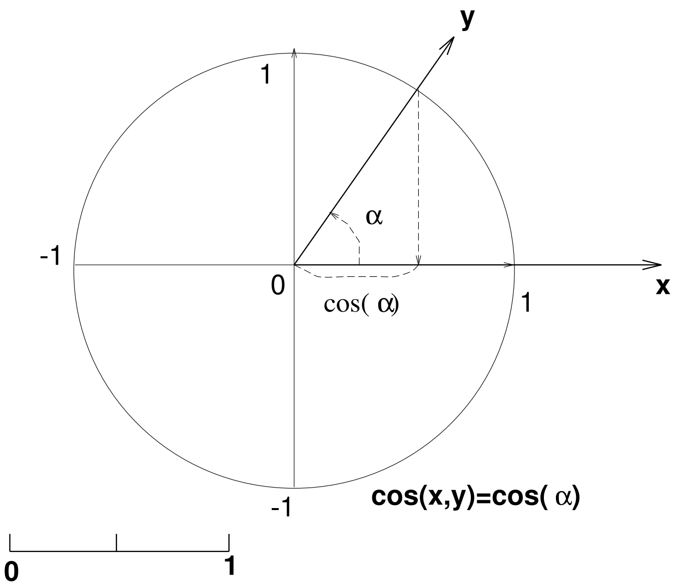
Le cosinus

On note $\cos(\mathbf{x}, \mathbf{y})$ la quantité

$$\cos(\mathbf{x}, \mathbf{y}) = \frac{< \mathbf{x}, \mathbf{y} >_D}{\|\mathbf{x}\|_D \|\mathbf{y}\|_D}$$

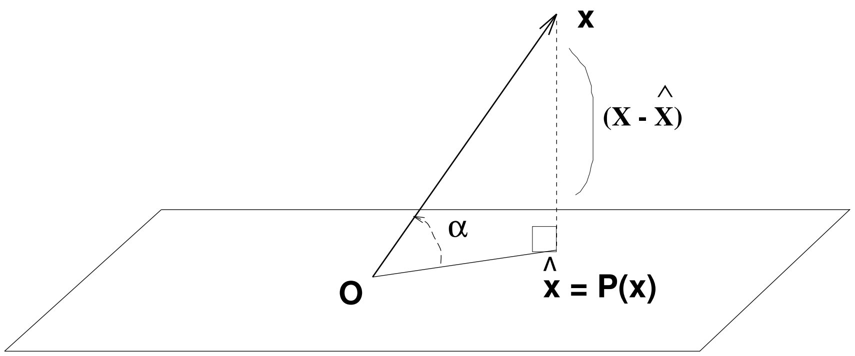
Propriétés

- $\cos(\mathbf{x}, \mathbf{y})$ varie entre -1 et +1.
- $\cos(\mathbf{x}, \mathbf{y}) = +1 \Leftrightarrow \mathbf{x}$ et $\mathbf{y}$ colinéaires et de même sens.
- $\cos(\mathbf{x}, \mathbf{y}) = -1 \Leftrightarrow \mathbf{x}$ et $\mathbf{y}$ colinéaires et de sens contraires.
- $\cos(\mathbf{x}, \mathbf{y}) = 0 \Leftrightarrow x \perp \mathbf{y}$.



- Si $\mathbf{x}$ et $\mathbf{y}$ sont normés à 1, leur produit scalaire est égal au cosinus de l'angle qu'ils forment.

Projection orthogonale



Pythagore :

$$\|\mathbf{x}\|^2 = \|\hat{\mathbf{x}}\|^2 + \|\mathbf{x} - \hat{\mathbf{x}}\|^2$$

$$\cos^2 \alpha = \frac{\|\hat{\mathbf{x}}\|^2}{\|\mathbf{x}\|^2}$$

représente le carré de la proportion de la longueur de $\mathbf{x}$ représenté par la projection.

$\Longrightarrow$ Coefficient de qualité de la représentation de $\mathbf{x}$ sur le plan.

- Le point $\hat{\mathbf{x}}$ du plan le plus proche de $\mathbf{x}$ au sens de la distance choisie, est la projection orthogonale de $\mathbf{x}$ sur ce plan. Il vérifie $< \mathbf{x} - \hat{\mathbf{x}}, \hat{\mathbf{x}} >= 0$.
- on a $\|\mathbf{x}\|^2 = \|\mathbf{x} - \hat{\mathbf{x}}\|^2 + \|\hat{\mathbf{x}}\|^2$ (Pythagore).
- cf. complément sur la projection dans l'annexe 1 page 6.
- Faire le bilan des connaissances : distance, norme, produit scalaire et projection.

Interprétation géométrique d'indices statistiques

- **moyenne pondérée** $\bar{x} = \sum_{i=1}^{n} p_i x_i$
- on note $\tilde{\mathbf{x}}$ le vecteur $\mathbf{x}$ centré avec sa moyenne $\bar{x}$
- **variance** $\sigma_x^2 = \sum_{i=1}^{n} p_i (x_i - \bar{x})^2 = \|\tilde{\mathbf{x}}\|_D^2$
- **covariance** $cov(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^{n} p_i (x_i - \bar{x})(y_i - \bar{y}) = < \tilde{\mathbf{x}}, \tilde{\mathbf{y}} >_D$
- **coefficient de corrélation linéaire** $\rho_{xy} = \frac{< \tilde{\mathbf{x}}, \tilde{\mathbf{y}} >_D}{\|\tilde{\mathbf{x}}\|_D \|\tilde{\mathbf{y}}\|_D}$

  c'est le cosinus de l'angle que font $\tilde{\mathbf{x}}$ et $\tilde{\mathbf{y}}$

- La variance d'une variable est égale au carré de la longueur de la variable centrée associée.
- Carré de la distance entre 2 colonnes centrées $\tilde{\mathbf{x}}$ et $\tilde{\mathbf{y}}$ de X :
  $d_D^2 = \|\tilde{\mathbf{x}} - \tilde{\mathbf{y}}\|_D^2 = < \tilde{x} - \tilde{y}, \tilde{x} - \tilde{y} >_D = \sum p_i (\tilde{x_i} - \tilde{y_i})^2 =$
  $\|\tilde{\mathbf{x}}\|_D^2 + \|\tilde{\mathbf{y}}\|_D^2 - 2 < \tilde{\mathbf{x}}, \tilde{\mathbf{y}} >_D = \sigma_x^2 + \sigma_y^2 - 2\rho_{xy}\sigma_x\sigma_y$
  La distance entre $\tilde{x}$ et $\tilde{y}$ est maximale si $\rho_{xy} = -1$ et minimale si $\rho_{xy} = +1$.
- Matrice de variance covariance des variables de X : ${}^t\tilde{X} D \tilde{X}$.

Décomposition aux valeurs singulières généralisées (DVSG)

Notion de rang

- Soit la matrice $\underset{n\times p}{X} = \left(\underset{n\times 1}{\mathbf{x}^1} \Big| \cdots \Big| \underset{n\times 1}{\mathbf{x}^p} \right)$
- Les combinaisons linéaires des p vecteurs $\mathbf{x}^j$ (colonnes de X) dans $\mathbb{R}^n$:
  $X\alpha = \sum_{j=1}^{p} \alpha_j \mathbf{x}^j$ avec $\alpha = \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_p \end{pmatrix}$
  sont les éléments de l'espace engendré par les colonnes de X.
- La dimension du sous-espace de $\mathbb{R}^n$ engendré par les p vecteurs colonnes de X est égale à celle du sous-espace de $\mathbb{R}^p$ engendré par ses n vecteurs lignes. Par définition, c'est le **rang** de X.

- Ne pas confondre dimension (nombre de lignes et colonnes) de X et rang.

Exemple : matrice 5×3 de rang 1

$$X = \begin{pmatrix} 1 & 2 & -2 \\ 4 & 8 & -8 \\ 8 & 16 & -16 \\ 5 & 10 & -10 \\ 4 & 8 & -6 \end{pmatrix}$$

Les colonnes de X sont les vecteurs $\mathbf{x}$, $2\mathbf{x}$ et $-2\mathbf{x}$ avec $\mathbf{x} = \begin{pmatrix} 1 \\ 4 \\ 8 \\ 5 \\ 3 \end{pmatrix}$

En effet, $\begin{pmatrix} 1 \\ 4 \\ 8 \\ 5 \\ 3 \end{pmatrix} \begin{pmatrix} 1 & 2 & -2 \end{pmatrix} = \begin{pmatrix} 1 & 2 & -2 \\ 4 & 8 & -8 \\ 8 & 16 & -16 \\ 5 & 10 & -10 \\ 4 & 8 & -6 \end{pmatrix} = X$

- Construction d'une matrice de rang 2

  Soient $\mathbf{u}^1$ et $\mathbf{u}^2$ deux vecteurs quelconques de $\mathbb{R}^n$
- $\mathbf{x}^j = v_1^j \mathbf{u}^1 + v_2^j \mathbf{u}^2 \qquad j = 1, \ldots, p$

  engendrés par $\mathbf{u}^1$ et $\mathbf{u}^2$
- $\underset{n \times p}{X} = \begin{pmatrix} u_1^1 & u_1^2 \\ \vdots & \vdots \\ u_n^1 & u_n^2 \end{pmatrix} \begin{pmatrix} v_1^1 & \ldots & v_1^p \\ v_2^1 & \ldots & v_2^p \end{pmatrix} = \mathbf{u}^1\ {}^t\mathbf{v}_1 + \mathbf{u}^2\ {}^t\mathbf{v}_2$

- Une matrice de rang 2 a été construite à partir de la somme de 2 matrices de rang 1.

  Plus généralement, une matrice de rang p sera construite à partir de la somme de p matrices de rang 1.
- La démarche consistant, à partir de X, à trouver les vecteurs u et v fait l'objet d'un théorème.
- Ici chaque variable j a deux composantes, par exemple :

  u_i^1 représenterait l'intelligence du sujet i,

  u_i^2 représenterait son aptitude manuelle

  et chaque note d'une matière j serait un certain mélange de ces deux qualités.

Décomposition aux valeurs singulières généralisées (DVSG) d'une matrice de rang 2 et généralisation

- Soit $\underset{n\times p}{X}$ de rang 2, on sait construire :

  – $\mathbf{u}^1$ et $\mathbf{u}^2$ de $\mathbb{R}^n$, **orthonormés** pour la métrique D

  – $\mathbf{v}_1$ et $\mathbf{v}_2$ de $\mathbb{R}^p$, **orthonormés** pour la métrique M

  – deux scalaires positifs $\lambda_1 \geq \lambda_2$

tels que

$$X = \begin{pmatrix} \mathbf{u}^1 & \mathbf{u}^2 \end{pmatrix} \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} \begin{pmatrix} {}^t\mathbf{v}_1 \\ {}^t\mathbf{v}_2 \end{pmatrix} = U\Lambda\, {}^tV = \sum_{s=1}^{2} \lambda_s \mathbf{u}^s\, {}^t\mathbf{v}_s$$

- $\mathbf{u}^1$ et $\mathbf{u}^2$ sont les **vecteurs singuliers à gauche**
- $\mathbf{v}_1$ et $\mathbf{v}_2$ sont les **vecteurs singuliers à droite**
- λ_1 et λ_2 sont les **valeurs singulières**

  – La DVSG générale d'une matrice de rang r s'écrit :
  $X = \sum_{s=1}^{r} \lambda_s \mathbf{u}^s\, {}^t\mathbf{v}_s$
  avec des propriétés analogues.

- Une matrice X peut se décomposer en une infinité de produits matriciels. Ainsi, si $X = \mathbf{u}^1\, {}^t\mathbf{v}_1 + \mathbf{u}^2\, {}^t\mathbf{v}_2$, on peut aussi écrire $X = (0.5\ \mathbf{u}^1)\ (2\ {}^t\mathbf{v}_1) + (0.5\ \mathbf{u}^2)\ (2\ {}^t\mathbf{v}_2)$ soit $X = \mathbf{w}^1\, {}^t\mathbf{z}_1 + \mathbf{w}^2\, {}^t\mathbf{z}_2$. On choisit une décomposition qui impose les contraintes suivantes :

  les vecteurs sont normés, $\|\mathbf{u}^1\|^2 = \|\mathbf{u}^2\|^2 = 1$
  $\|\mathbf{v}_1\|^2 = \|\mathbf{v}_2\|^2 = 1$
  ils sont orthogonaux, $\mathbf{u}^1 \perp \mathbf{u}^2$
  $\mathbf{v}_1 \perp \mathbf{v}_2$

  L'orthogonalité réduit le nombre de décompositions possibles. La première contrainte (norme égale à 1) impose l'ajout des coefficients λ_1 et λ_2 dans l'équation du transparent 22. On obtient alors la *décomposition aux valeurs singulières* (DVS) : $X = \lambda_1 \mathbf{u}^1\, {}^t\mathbf{v}_1 + \lambda_2 \mathbf{u}^2\, {}^t\mathbf{v}_2$

  Une extension de cette DVS (métrique identité) est obtenue en se donnant les métriques D et M :

  $\|\mathbf{u}^1\|^2_D = \sum_{i=1}^{n} p_i (u_i^1)^2 = 1 \quad \|\mathbf{u}^2\|^2_D = \sum_{i=1}^{n} p_i (u_i^2)^2 = 1$ et

  $< \mathbf{u}^1, \mathbf{u}^2 >_D = \sum_{i=1}^{n} p_i \mathbf{u}_i^1 \mathbf{u}_i^2 = 0$

  $\|\mathbf{v}_1\|^2_M = \sum_{j=1}^{p} m_j (v_1^j)^2 = 1 \quad \|\mathbf{v}_2\|^2_M = \sum_{j=1}^{p} m_j (v_2^j)^2 = 1$ et

  $< \mathbf{v}_1, \mathbf{v}_2 >_M = \sum_{j=1}^{p} m_j \mathbf{v}_1^j \mathbf{v}_2^j = 0$

  On obtient alors la *décomposition aux valeurs singulières généralisées* (DVSG).

- X étant de rang 2, on a construit deux matrices de rang 1 ($\mathbf{u}^1\, {}^t\mathbf{v}_1$ et $\mathbf{u}^2\, {}^t\mathbf{v}_2$ dont la somme pondérée est égale à X). Le théorème se généralise au cas d'une matrice de rang quelconque.

- Noter que :
  Toute colonne de X est combinaison linéaire de $\mathbf{u}^1$ et $\mathbf{u}^2$.
  Toute ligne de X est combinaison linéaire de $\mathbf{v}_1$ et $\mathbf{v}_2$.
  De façon générale, le rang de X est inférieur ou égal à $min(n, p)$.

Application du théorème d'Eckart-Young (1936) à l'approximation d'une matrice et interprétation géométrique

But : construction d'une matrice de faible rang (fixé) la plus proche d'une matrice X.

Soit X, de rang r, dont la DVSG s'écrit :

$$X = \sum_{s=1}^{r} \lambda_s \mathbf{u}^s \, {}^t\mathbf{v}_s$$

alors la matrice, notée $\hat{X}^k$, de rang k ($k \leq r$) la plus proche de X est la matrice

$$\widehat{X}^k = \sum_{s=1}^{k} \lambda_s \mathbf{u}^s \, {}^t\mathbf{v}_s$$

$$\widehat{X}^k = \begin{pmatrix} u_1^1 & \dots & u_1^k \\ \dots & \dots & \dots \\ u_i^1 & \dots & u_i^k \\ \dots & \dots & \dots \\ u_n^1 & \dots & u_n^k \end{pmatrix} \begin{pmatrix} \lambda_1 & \dots & 0 \\ \dots & \dots & \dots \\ 0 & \dots & \lambda_k \end{pmatrix} \begin{pmatrix} v_1^1 & \dots & v_1^j & \dots & v_1^p \\ \dots & \dots & \dots & \dots & \dots \\ v_k^1 & \dots & v_k^j & \dots & v_k^p \end{pmatrix}$$

Pour obtenir l'expression de l'approximation de rang k de X, il suffit de supprimer dans la DVSG de X les termes de rang supérieurs à k.

- Ce théorème constitue le fondement des techniques ACP et AFC présentées dans ce cours.
- Classement des valeurs singulières $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_r > 0$.
- On retrouve un schéma fréquent dans les méthodes statistiques :
  donnée = modèle + résidu
  soit en terme de tableaux :
  $X = \hat{X} + (X - \hat{X})$
  La structure d'espace vectoriel et les métriques définissent le modèle et par conséquent l'approximation.

Exemple “note” :

$$X = \begin{array}{cc} & \begin{array}{ccccc} m & s & f & l & dm \end{array} \\ \begin{array}{c} JE \\ AL \\ AN \\ MO \\ DI \\ AD \\ PI \\ BR \\ EV \end{array} & \begin{pmatrix} 6 & 6 & 5 & 5.5 & 8 \\ 8 & 8 & 8 & 8 & 9 \\ 6 & 7 & 11 & 9.5 & 11 \\ 14.5 & 14.5 & 15.5 & 15 & 8 \\ 14 & 14 & 12 & 12.5 & 10 \\ 11 & 10 & 5.5 & 7 & 13 \\ 5.5 & 7 & 14 & 11.5 & 10 \\ 13 & 12.5 & 8.5 & 9.5 & 12 \\ 9 & 9.5 & 12.5 & 12 & 18 \end{pmatrix} \end{array}$$

$$= \begin{array}{c} \begin{array}{ccccc} \mathbf{u}^1 & \mathbf{u}^2 & \mathbf{u}^3 & \mathbf{u}^4 & \mathbf{u}^5 \end{array} \\ \begin{pmatrix} -0.59 & -0.25 & 0.54 & 0.26 & -1.45 \\ -0.79 & -0.05 & 0.15 & 0.21 & 0.27 \\ -0.86 & -0.96 & -0.60 & -1.29 & 0.21 \\ -1.29 & 1.57 & -1.17 & 1.27 & 0.96 \\ -1.20 & 1.16 & -0.01 & -0.38 & -1.80 \\ -0.89 & -0.11 & 1.75 & 0.14 & 1.45 \\ -0.93 & -0.94 & -1.60 & -0.90 & 0.43 \\ -1.06 & 0.56 & 1.10 & -1.46 & 0.25 \\ -1.18 & -1.72 & 0.30 & 1.58 & -0.49 \end{pmatrix} \end{array} \begin{pmatrix} 23.26 & 0 & 0 & 0 & 0 \\ 0 & 3.61 & 0 & 0 & 0 \\ 0 & 0 & 3.41 & 0 & 0 \\ 0 & 0 & 0 & .456 & 0 \\ 0 & 0 & 0 & 0 & .1 \end{pmatrix}$$

$$\begin{array}{cc} \begin{pmatrix} -.429 & -.436 & -.453 & -.446 & -.471 \\ .518 & .414 & -.135 & .01 & -.736 \\ .365 & .196 & -.675 & -.37 & .486 \\ .318 & -.575 & -.371 & .656 & -.022 \\ .559 & -.519 & .428 & -.484 & .017 \end{pmatrix} & \begin{array}{c} {}^t\mathbf{v}_1 \\ {}^t\mathbf{v}_2 \\ {}^t\mathbf{v}_3 \\ {}^t\mathbf{v}_4 \\ {}^t\mathbf{v}_5 \end{array} \end{array}$$

$= U\Lambda {}^tV$

U est la matrice dont les colonnes sont $\mathbf{u}^1, \ldots, \mathbf{u}^5$ normés par D,
V est la matrice dont les colonnes sont $\mathbf{v}_1, \ldots, \mathbf{v}_5$ normés par M,
X est de rang 5.

Remarquer que dans ce produit matriciel, si l’on remplace les petites valeurs singulières par 0 (les 2 dernières), on “n’oublie” pas grand chose. Ainsi, le résultat est une matrice de “faible” rang (ici 3) assez facilement interprétable et peu différente de la matrice initiale.

On peut vérifier que :

${}^t\mathbf{u}^1\mathbf{u}^2 = 0$ (orthogonalité)

${}^t\mathbf{v}_1\mathbf{v}_1 = 1$ (vecteur normé) :

$(-.429)^2 + (-.436)^2 + (-.453)^2 + (-.446)^2 + (-.471)^2 = 1$

$1/9 \ {}^t\mathbf{u}^1\mathbf{u}^1 = 1$

$1/9[(-0.59)^2 + (-0.79)^2 + (-0.86)^2 + (-1.29)^2 + (1.20)^2 + (-0.89)^2 + (-0.93)^2 + (-1.06)^2 + (-1.18)^2] = 1$

$\hat{X}^k = \sum_{s=1}^{k} \lambda_s u^s \, {}^t v_s$

Approximation de rang 2 du tableau “note”

	m	s	f	l	dm
JE	5.4	5.6	6.3	6.1	7.1
AL	7.8	7.9	8.3	8.2	8.8
AN	6.8	7.3	9.5	8.9	12.0
MO	15.8	15.4	12.8	13.4	9.9
DI	14.1	13.9	12.0	12.4	10.0
AD	8.7	8.9	9.5	9.3	10.1
PI	7.5	8.0	10.2	9.6	12.7
BR	11.7	11.6	10.9	11.1	10.2
EV	8.6	9.4	13.3	12.2	17.5

Approximation de rang 3 du tableau “note”

	m	s	f	l	dm
JE	6.1	5.9	5.1	5.4	8
AL	8	8	8	8	9
AN	6.1	6.9	10.9	9.6	11
MO	14.4	14.7	15.5	14.9	8
DI	14.1	13.9	12.1	12.5	10
AD	10.9	10.1	5.4	7.1	13
PI	5.5	6.9	13.9	11.6	10
BR	13.1	12.4	8.4	9.7	12
EV	8.9	9.6	12.6	11.8	18

- Note technique pour le formateur :

  la fonction *svd* de Splus donne la décomposition aux valeurs singulières (DVS) d'un tableau avec des métriques identités.

  Pour obtenir la décomposition aux valeurs singulières **généralisée** (DVSG) de X avec des métriques D et M diagonales mais différentes de l'identité (cas présenté ici) , il suffit de faire la DVS de $D^{1/2}XM^{1/2}$. En effet, on a $X = U\Lambda \, {}^tV$ équivalent à $D^{1/2}XM^{1/2} = (D^{1/2}U)\Lambda \, {}^t(M^{1/2}V)$.

  $\tilde{U} = D^{1/2}U$ et $\tilde{V} = M^{1/2}V$ vérifient bien la normalisation de la DVS.

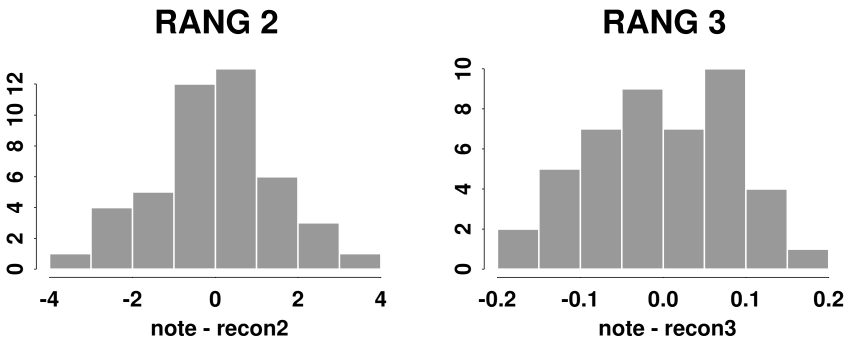
  Pour obtenir la DVSG de X avec Splus, il suffit donc de faire $svd(D^{**}.5\%^{*}\%X\%^{*}\%M^{**}.5)$ qui donnera $\tilde{U}$ et $\tilde{V}$.

  D'où $U = D^{-1/2}\tilde{U}$ et $V = M^{-1/2}\tilde{V}$.

  Les valeurs singulières de la DVSG de X sont données par le résultat de *svd*.

- $\hat{X}^k$ est calculée à l'aide des k premières valeurs singulières.
- Comparer $\hat{X}^2$ et $\hat{X}^3$ avec le tableau X.
- Une valeur singulière apporte d'autant plus d'information dans l'approximation qu'elle est grande.

Analyse des résidus $X - \hat{X}$



- On compare ici globalement les résidus sur l'ensemble des variables car on a la même unité. On constate que la gamme de variation des résidus diminue quand le rang de l'approximation augmente. L'examen des résidus élevés permettrait de détecter les individus et variables mal approximés.

  Attention : les échelles ne sont pas les mêmes.

Conséquences du théorème d'Eckart-Young

Décomposition exacte	Approximation de rang k ($k \leq r$)
Elément par élément	
$x_i^j = \sum_{s=1}^{r} \lambda_s u_i^s v_s^j$	$\widehat{x_i^j}^k = \sum_{s=1}^{k} \lambda_s u_i^s v_s^j$
Par ligne (dans $\mathbb{R}^p$)	
$\mathbf{x}_i = \sum_{s=1}^{r} (\lambda_s u_i^s)\mathbf{v}_s$	$\widehat{\mathbf{x}_i}^k = \sum_{s=1}^{k} (\lambda_s u_i^s)\mathbf{v}_s$
Par colonne (dans $\mathbb{R}^n$)	
$\mathbf{x}^j = \sum_{s=1}^{r} (\lambda_s v_s^j)\mathbf{u}^s$	$\widehat{\mathbf{x}^j}^k = \sum_{s=1}^{k} (\lambda_s v_s^j)\mathbf{u}^s$

(1)	(2)
(3)	(4)
(5)	(6)

- Carré de la norme de l'écart entre une colonne j et son approximation de rang k :

$$\mathbf{x}^j - \hat{\mathbf{x}}^j = \sum_{s>k} \lambda_s v_s^j \mathbf{u}^s$$

$$\|\mathbf{x}^j - \hat{\mathbf{x}}^j\|_D^2 = \sum_{s>k} (\lambda_s v_s^j)^2 \|\mathbf{u}^s\|_D^2 = \sum_{s>k} (\lambda_s v_s^j)^2$$

1. L'élément x_i^j se reconstitue exactement par la formule (1).
2. $\widehat{x_i^j}^k$ est son approximation à l'ordre k.
3. La ligne $\mathbf{x}_i$ est combinaison linéaire des vecteurs $\mathbf{v}_s$ qui forment une base orthonormée.
4. $\widehat{\mathbf{x}_i}^k$ est obtenue en négligeant les derniers vecteurs de base.
5. La colonne $\mathbf{x}^j$ est combinaison linéaire des vecteurs $\mathbf{u}_s$ qui forment une base orthonormée.
6. $\widehat{\mathbf{x}^j}^k$ est obtenue en négligeant les derniers vecteurs de base.

Interprétation géométrique

Soit une base orthogonale $(\mathbf{u}^s)$, $s = 1, \ldots, r$ de $\mathbb{R}^n$.

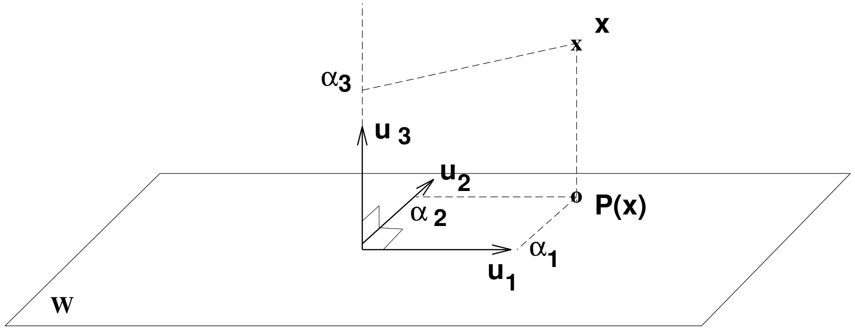
Projeter sur le sous-espace W engendré par certains vecteurs de la base $(\mathbf{u}^s)$, $s = 1, \ldots, k$ $(k \leq r)$, revient à annuler les coordonnées des vecteurs qui n'engendrent pas W.

Exemple avec $r = 3$:

$\mathbf{x} = \alpha_1 \mathbf{u}^1 + \alpha_2 \mathbf{u}^2 + \alpha_3 \mathbf{u}^3$.

P, projecteur orthogonal sur le plan engendré par $(\mathbf{u}^1, \mathbf{u}^2)$.

$P(\mathbf{x}) = \alpha_1 \mathbf{u}^1 + \alpha_2 \mathbf{u}^2$



On voit que l'opération qui consiste à passer des vecteurs lignes ou colonnes à leurs approximations est une **projection orthogonale**, sur le sous-espace engendré respectivement par les k premiers vecteurs singuliers à droite (lignes) ou à gauche (colonnes).

Ainsi, calculer la meilleure approximation de rang k d'une matrice (ou de ses lignes, de ses colonnes, de ses éléments) revient à remplacer :

- le nuage de ses lignes par sa projection sur le sous-espace de base orthonormée $(\mathbf{v}_1, \ldots, \mathbf{v}_k)$.
- le nuage de ses colonnes par sa projection sur le sous-espace de base orthonormée $(\mathbf{u}^1, \ldots, \mathbf{u}^k)$.

Formules de transition

De la DVSG de X,

$$X = \sum_{s=1}^{r} \lambda_s \mathbf{u}^s \, {}^t\mathbf{v}_s$$

on déduit :

$$XM\mathbf{v}_{s'} = (\sum_{s=1}^{r} \lambda_s \mathbf{u}^s \, {}^t\mathbf{v}_s) M\mathbf{v}_{s'} = \sum_{s=1}^{r} \lambda_s \mathbf{u}^s \, ({}^t\mathbf{v}_s M \mathbf{v}_{s'})$$

De ${}^t\mathbf{v}_s M\mathbf{v}_{s'} = 0$ si $s \neq s'$ et ${}^t\mathbf{v}_s M\mathbf{v}_{s'} = 1$ si $s = s'$ on déduit :

$$XM\mathbf{v}_s = \lambda_s \mathbf{u}^s \Longrightarrow \mathbf{u}^s = XM\mathbf{v}_s / \lambda_s$$

De même, l'expression de $\mathbf{v}_s$ en fonction de $\mathbf{u}^s$ est :

$$\mathbf{v}_s = {}^tXD\mathbf{u}^s / \lambda_s$$

- $\mathbf{u}^s$ est vecteur propre (cf. Module A) de la matrice $XM\,{}^tXD$ associé à la **valeur propre** $(\lambda)_s^2$.
- $\mathbf{v}_s$ est vecteur propre de la matrice ${}^tXDXM$ associé à la **valeur propre** $(\lambda)_s^2$.

Analyse générale d'un tableau X

- Construction d'indices de qualité de l'approximation du tableau.
- Représentation graphique de l'approximation de faible dimension :
  - du nuage de ses lignes
  - du nuage de ses colonnes

L'inertie totale de la matrice X

Définie comme le carré de la norme de X, elle mesure globalement l'importance des éléments de la matrice :

$$\begin{aligned}
\mathrm{In(X)} &= \|X\|^2 \\
&= \sum_i p_i \|\mathbf{x}_i\|_M^2 \\
&\quad \text{(somme pondérée des carrés des longueurs des lignes)} \\
&= \sum_j m_j \|\mathbf{x}^j\|_D^2 \\
&\quad \text{(somme pondérée des carrés des longueurs des colonnes)} \\
&= \sum_i \sum_j p_i m_j (x_i^j)^2 \\
&\quad \text{(somme pondérée des carrés des éléments de la matrice)}
\end{aligned}$$

$\dfrac{p_i \|\mathbf{x}_i\|_M^2}{\|X\|^2}$ = **contribution** d'un individu à l'inertie totale.

$\dfrac{m_j \|\mathbf{x}^j\|_D^2}{\|X\|^2}$ = **contribution** d'une variable à l'inertie totale.

- **L'inertie** est une somme de quantités associées aux lignes ou aux colonnes de X. L'**unité de mesure** et la **variabilité** des observations ont donc une influence sur le niveau de l'inertie du tableau. Ainsi une variable a une contribution à l'inertie totale qui peut varier selon son unité de mesure ou sa variabilité. Pour éliminer ses effets il faut effectuer une transformation sur les colonnes du tableau X ; la plus classique consistant à diviser chaque colonne par son écart type. Les colonnes ainsi obtenues seront sans dimension (i.e. pas d'unité) et auront une variance égale à 1. D'autres mesures de dispersion pourraient être envisagées (écart inter-quartile, référence, ...). Pour l'ACP on utilisera fréquemment la standardisation avec l'écart type.
- En mécanique physique, l'inertie d'un ensemble d'individus munis de masses p_i est calculée par rapport à un point (faire un dessin au tableau). On retrouve bien la même idée dans l'analyse générale d'un tableau, l'inertie est calculée par rapport à l'origine O. Choisir un autre point A de $\mathbb{R}^p$ consisterait à centrer les colonnes du tableau à l'aide des coordonnées de A et effectuer ensuite la DVSG sur ce tableau. Ce problème sera encore détaillé dans l'exposé de l'ACP.

Approximation globale de X

$$\|X\|^2 = \sum_{i=1}^{n} p_i \|\mathbf{x}_i\|_M^2 = \sum_{j=1}^{p} m_j \|\mathbf{x}^j\|_D^2 = \sum_{s=1}^{r} (\lambda_s)^2$$

$$\|\hat{X}\|^2 = \sum_{i=1}^{n} p_i \|\hat{\mathbf{x}}_i\|_M^2 = \sum_{j=1}^{p} m_j \|\hat{\mathbf{x}}^j\|_D^2 = \sum_{s=1}^{k} (\lambda_s)^2$$

$$\frac{\sum_{s=1}^{k} (\lambda_s)^2}{\sum_{s=1}^{r} (\lambda_s)^2} \times 100 = \text{ pourcentage d'inertie expliquée}$$

par les k premières dimensions.

On écrit le vecteur $\mathbf{x}_i$ dans la base $(\mathbf{v}_s,\ s = 1, \ldots, r)$ de vecteurs orthonormés de $\mathbb{R}^p$ ($\|v_s\|_M = 1$) :

$$\mathbf{x}_i = \sum_{s=1}^{r} (\lambda_s u_i^s) \mathbf{v}_s,$$

$\hat{\mathbf{x}}_i$ étant l'approximation calculée à partir des k premières valeurs singulières (on a supprimé l'indice k pour plus de clarté), on a :

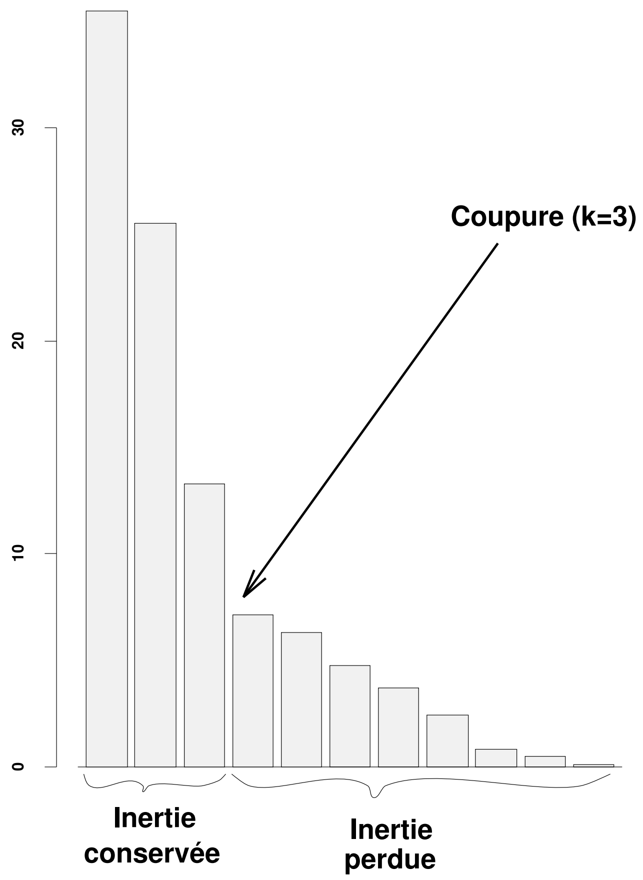
$$\|\hat{\mathbf{x}}_i\|_M^2 = \sum_{s=1}^{k} (\lambda_s u_i^s)^2$$

d'où

$$\sum_{i=1}^{n} p_i \|\hat{\mathbf{x}}_i\|_M^2 = \sum_{i=1}^{n} p_i \left(\sum_{s=1}^{k} (\lambda_s u_i^s)^2 \right) = \sum_{s=1}^{k} \lambda_s^2 \sum_{i=1}^{n} p_i (u_i^s)^2 = \sum_{s=1}^{k} \lambda_s^2 \|\mathbf{u}^s\|_D^2 = \sum_{s=1}^{k} \lambda_s^2$$

Pour les colonnes $\mathbf{x}^j$ la démonstration est du même type.

Éboulis des valeurs propres : diagramme des $\{\ \lambda_s^2,\ s=1,\ldots,r\ \}$ (ordonnées par construction)



- $\hat{\mathbf{x}}_i$ étant l'approximation calculée à partir de la seule valeur singulière s :

$$\sum_{i=1}^{n} p_i \|\hat{\mathbf{x}}_i\|_M^2 = \sum_{i=1}^{n} p_i (\lambda_s u_i^s)^2 = \lambda_s^2 \|\mathbf{u}^s\|_D^2 = \lambda_s^2$$

L'inertie des projections des n lignes sur l'axe engendré par le vecteur $\mathbf{v}_s$ est le carré de la valeur singulière λ_s. D'où l'idée de représenter le diagramme de ces valeurs $(\lambda_s)^2$, appelé **éboulis des valeurs propres**. Rappelez-vous que $(\lambda_s)^2$ est la valeur propre d'une certaine matrice. La somme de longueurs des bâtons est égale à l'inertie totale.

Quand on cherche un résumé le plus simple possible (de rang minimum) et contenant le maximum d'information (d'inertie maximum), on retiendra les valeurs singulières conservant une part d'inertie conséquente. La règle empirique appelée “règle du coude” consistera à repérer dans l'éboulis une rupture de pente importante afin de sélectionner les valeurs singulières les plus fortes.

$$\begin{aligned}\|X\|^2 &= \|\hat{X}\|^2 + \|X-\hat{X}\|^2 \\ &= \sum_{s=1}^{k}(\lambda_s)^2 + \sum_{s=k+1}^{r}(\lambda_s)^2\end{aligned}$$

Décomposition selon les lignes et les colonnes

Exemple

note	m	s	f	l	dm	inertie des lignes
JE	6.0	6.0	5.0	5.5	8.0	21.3
AL	8.0	8.0	8.0	8.0	9.0	37.4
AN	6.0	7.0	11.0	9.5	11.0	46.4
MO	14.5	14.5	15.5	15.0	8.0	105.5
DI	14.0	14.0	12.0	12.5	10.0	88.0
AD	11.0	10.0	5.5	7.0	13.0	52.1
PI	5.5	7.0	14.0	11.5	10.0	56.4
BR	13.0	12.5	8.5	9.5	12.0	70.2
EV	9.0	9.5	12.5	12.0	18.0	88.4
inertie des colonnes	104.8	105.6	116.6	109.0	129.7	565.7

Contribution des lignes (en millièmes)

JE	AL	AN	MO	DI	AD	PI	BR	EV
38	66	82	187	156	92	100	124	156

Contribution des colonnes (en millièmes)

m	s	f	l	dm
185	187	206	193	229

- Poids sur les lignes : $p_i = \dfrac{1}{9}$
- Vérifier quelques calculs :

Inertie de la première ligne :

$$[(6.0)^2 + (6.0)^2 + (5.0)^2 + (5.5)^2 + (8.0)^2]/9 = 21.3$$

Remarquer que le dénominateur n'est pas 5, mais 9.

Inertie de la première colonne :

$$[(6.0)^2 + (8.0)^2 + \ldots + (13.0)^2 + (9.0)^2]/9 = 104.8$$

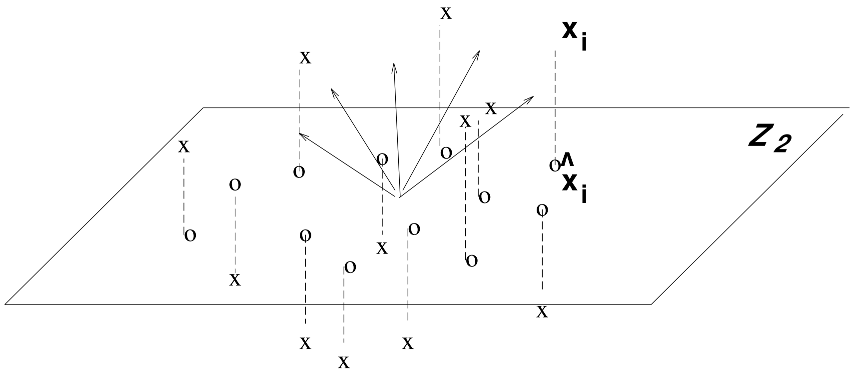
La contribution de JE est $21.3/565.7 = 0.038$.

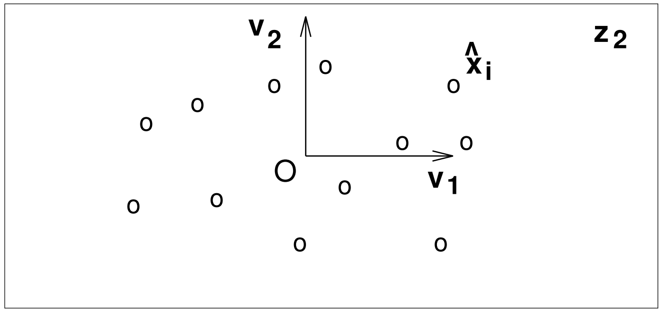
Décomposition par ligne :

$$\mathbf{x}_i = \sum_{s=1}^{r} (\lambda_s u_i^s) \mathbf{v}_s \qquad \widehat{x_i}^k = \sum_{s=1}^{k} (\lambda_s u_i^s) \mathbf{v}_s$$

$\lambda_s u_i^s$ représente les coordonnées de $\mathbf{x}_i$ dans la base $(\mathbf{v}_s)$.

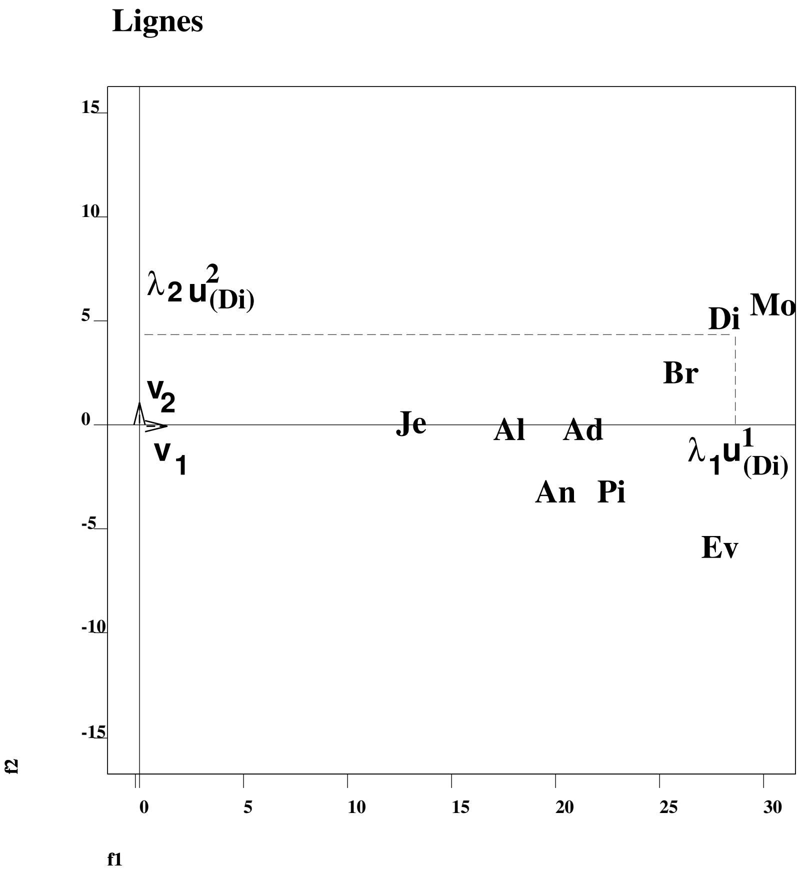
Nuage des lignes projeté sur le sous-espace $Z_2 = vect(\mathbf{v}_1, \mathbf{v}_2)$ de $\mathbb{R}^p$ (approximation de rang 2) :





- Z_2 est appelé le plan principal, les axes engendrés par $\mathbf{v}_1, \ldots, \mathbf{v}_s$ sont les axes principaux.
- La représentation graphique des lignes (resp. des colonnes) consiste à utiliser les coordonnées dans la base $(\mathbf{v}_s,\ s = 1, \ldots, r)$ (resp. dans la base $(\mathbf{u}^s,\ s = 1, \ldots, r)$).
- On appelle **facteurs** les nouvelles variables $\mathbf{u}^s$.
- $\|\hat{\mathbf{x}}_i\|_M^2 = (\lambda_1 u_i^1)^2 + (\lambda_2 u_i^2)^2$ est l'approximation de rang deux (ou d'ordre deux) de $\|\mathbf{x}_i\|_M^2 = \sum_{j=1}^{p} m_j (x_i^j)^2$.
- Si l'on perd peu d'inertie dans cette projection, la distance entre deux points (ou lignes du tableau) X est peu modifiée : on visualise alors sur le plan la **ressemblance** entre les lignes du tableau X.

Exemple
Nuage des lignes projeté sur Z_2



- Analyse générale du tableau de notes **non centré**.
- Représentation des lignes (élèves) sur les 2 premiers facteurs $\mathbf{v}_1$ et $\mathbf{v}_2$.
- Vérifier quelques coordonnées :

$$\lambda_1 = 23.26 \quad \lambda_2 = 3.61$$

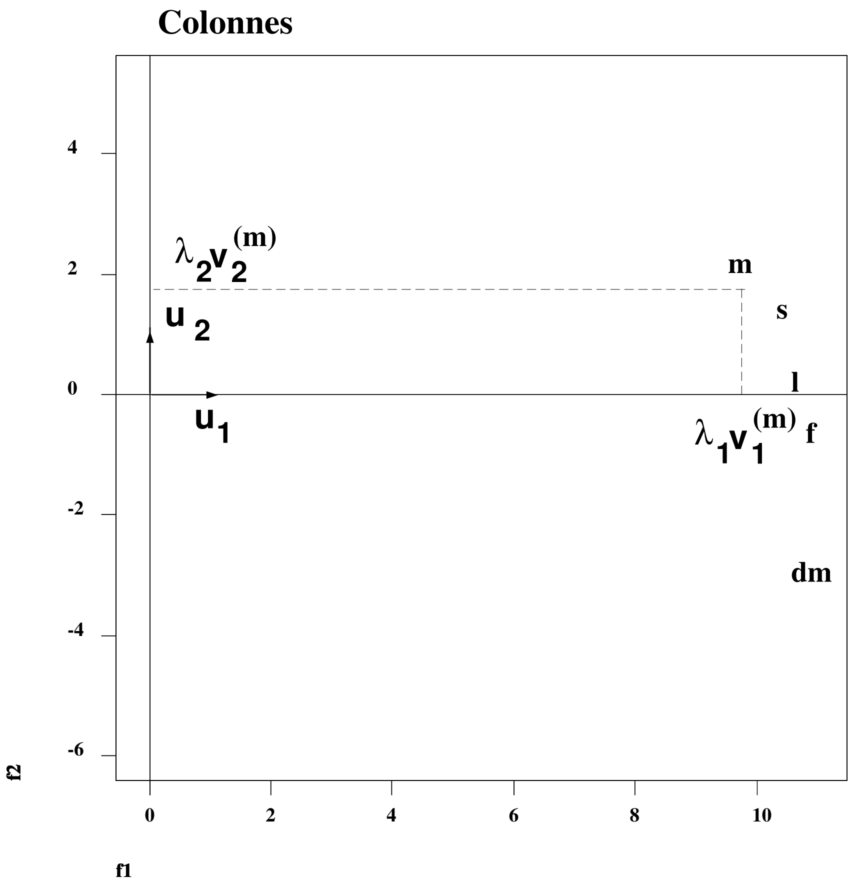
	$\mathbf{u}^1$	$\mathbf{u}^2$
JE	-0.59	-0.25
AL	-0.79	-0.05
AN	-0.86	-0.96
MO	-1.29	1.57
DI	-1.20	1.16
AD	-0.89	-0.11
PI	-0.93	-0.94
BR	-1.06	0.56
EV	-1.18	-1.72

$$MO \;:\; -1.29 \times 23.26 = -30 \text{ sur le facteur } \mathbf{v}_1$$
$$:\; 1.57 \times 3.61 = 5.67 \text{ sur le facteur } \mathbf{v}_2$$

$$DI \;:\; -1.20 \times 23.26 = -27.90 \text{ sur le facteur } \mathbf{v}_1$$
$$:\; 1.16 \times 3.61 = 4.19 \text{ sur le facteur } \mathbf{v}_2$$

Remarque : dans la représentation il y a un changement de signe sur **u** et **v** cependant les proximités entre points sont inchangées.

Nuage des colonnes projeté sur $W_2 = vect(\mathbf{u}^1, \mathbf{u}^2)$



- Représentation des colonnes (matières).
- nécessité de définir des indices de qualité de la représentation des lignes et des colonnes.

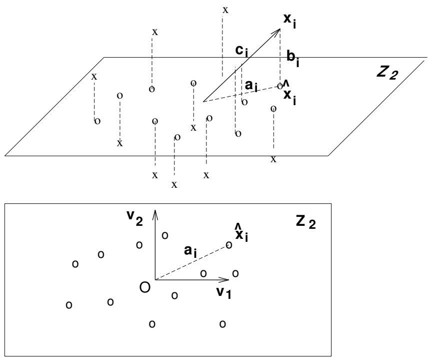
Critères équivalents (lignes de X)

$\sum_{i=1}^{n} p_i \|\widehat{\mathbf{x}}_i^k\|_M^2$ est l'inertie du nuage projeté des lignes de X sur $Z_k \subset \mathbb{R}^p$, $(Z_k = vect(\mathbf{v}_1, \ldots, \mathbf{v}_k))$.

On a les équivalences :

- la matrice approchée $\widehat{X}^k$ est la matrice de rang k la plus proche de X.
- Le sous-espace Z_k est le sous-espace de $\mathbb{R}^p$ qui, parmi les sous-espaces de dimension k, rend maximum l'inertie du nuage projeté des lignes de X ($\sum_i p_i \|\hat{\mathbf{x}}_i\|_M^2$ maximum).

Th. de Pythagore $\Rightarrow \sum_i p_i \|x_i\|_M^2 = \sum_i p_i \|\hat{\mathbf{x}}_i\|_M^2 + \sum_i p_i \|\mathbf{x}_i - \hat{\mathbf{x}}_i\|_M^2$



Inertie totale du nuage des lignes = Inertie du nuage projeté des lignes + somme pondérée des distances des lignes x_i au sous-espace Z_k.

- Maximiser l'inertie des projections $\sum_{i=1}^{n} p_i \|\hat{\mathbf{x}}_i\|_M^2$ revient à minimiser l'inertie des écarts $\sum_{i=1}^{n} p_i \|\mathbf{x}_i - \hat{\mathbf{x}}_i\|_M^2$.

  Le sous-espace principal Z_k est donc, encore parmi les sous-espaces de dimension k, celui qui passe le plus près possible du nuage des lignes.

Critères équivalents (colonnes de X)

De même $\sum_{j=1}^{p} m_j \|\widehat{\mathbf{x}^j}^k\|_D^2$ est l'inertie du nuage projeté des colonnes de X sur le sous-espace $W_k \subset \mathbb{R}^n$ ($W_k = vect(\mathbf{u}^1, \ldots, \mathbf{u}^k)$). On a les équivalences suivantes :

- la matrice approchée $\widehat{X}^k$ est la matrice de rang k la plus proche de X.
- W_k est le sous-espace de dimension k de $\mathbb{R}^n$ tel qu'en projection l'inertie du nuage des colonnes est maximum ($\sum_j m_j \|\widehat{\mathbf{x}^j}^k\|_D^2$ maximum).

Th. de Pythagore $\Rightarrow \sum_j m_j \|\mathbf{x}^j\|_D^2 = \sum_j m_j \|\hat{\mathbf{x}}^j\|_D^2 + \sum_j m_j \|\mathbf{x}^j - \hat{\mathbf{x}}^j\|_D^2$

L'inertie totale est la somme de

- l'inertie du nuage des projetés des colonnes
- la somme pondérée des distances des $\mathbf{x}^j$ au sous-espace W_k.

Maximiser l'inertie du nuage des projetés des colonnes équivaut à *minimiser* la distance du sous-espace au nuage.

Double décomposition de l'inertie (colonnes × facteurs)

Facteur	1	...	s	...	k	Sous-total	...	r	Total
Colonne									
1	...	...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...	...
j	...	...	$m_j(\lambda_s v_s^j)^2$	...	...	$m_j\|\widehat{\mathbf{x}^j}^k\|_D^2$	...	...	$m_j\|\mathbf{x}^j\|_D^2$
...	...	...	...	...	...	...	...	...	...
J	...	...	...	...	...	...	...	...	...
Total	λ_1^2	...	λ_s^2	...	λ_k^2	$\|\widehat{X}^k\|^2$	...	λ_r^2	$\|X\|^2$

Table 1 : la décomposition "colonne × facteur" de l'inertie.

- contribution du facteur s à l'inertie totale : $\dfrac{\lambda_s^2}{\|X\|^2}$
- contribution de la colonne j à l'inertie totale : $\dfrac{m_j\|\mathbf{x}^j\|_D^2}{\|X\|^2}$
- contribution du facteur s à la colonne j : $\dfrac{(\lambda_s v_s^j)^2}{\|\mathbf{x}^j\|_D^2}$
- contribution de la colonne j au facteur s : $m_j(v_s^j)^2$

• Décomposition de l'inertie totale :

1. $\sum_{s=1}^{r}(\lambda_s)^2 = \|X\|^2$
2. inertie totale calculée à l'aide des colonnes
   $\sum_{j=1}^{p} m_j\|\mathbf{x}^j\|_D^2 = \|X\|^2$
3. décomposition du carré de la norme de x^j
   $\|\mathbf{x}^j\|_D^2 = \sum_{s=1}^{r}(\lambda_s v_s^j)^2$
4. décomposition de $(\lambda_s)^2$
   $\sum_{j=1}^{p} m_j(\lambda_s v_s^j)^2 = (\lambda_s)^2$

De ces sommes on déduit les contributions présentées.

Double décomposition de l'inertie (lignes × facteurs)

Facteur	1	...	s	...	k	Sous-total	...	r	Total
Individu									
1	...	...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...	...
i	...	...	$p_i(\lambda_s u_i^s)^2$	...	...	$p_i\\|\widehat{\mathbf{x}_i}^k\\|_M^2$	...	...	$p_i\\|\mathbf{x}_i\\|_M^2$
...	...	...	...	...	...	...	...	...	...
n	...	...	...	...	...	...	...	...	...
Total	λ_1^2	...	λ_s^2	...	λ_k^2	$\\|\widehat{X}^k\\|^2$	...	λ_r^2	$\\|X\\|^2$

Table 2 : la décomposition "individu × facteur" de l'inertie.

- Faire construire les contributions analogues à celles des colonnes ;

contribution d'une ligne i à l'inertie totale :

$p_i\|\mathbf{x}_i\|_M^2/\|X\|^2$

contribution d'un facteur s à une ligne i :

$(\lambda_s u_i^s)^2/\|\mathbf{x}_i\|_M^2$

contribution d'une ligne i à l'inertie du facteur s :

$p_i(u_i^s)^2/(\lambda_s)^2$

Les mesures de qualité

1. Qualité globale de l'approximation d'ordre k :

$$\frac{\|\widehat{X}^k\|^2}{\|X\|^2} = \frac{\sum_{s=1}^{k} (\lambda_s)^2}{\sum_{s=1}^{r} (\lambda_s)^2}$$

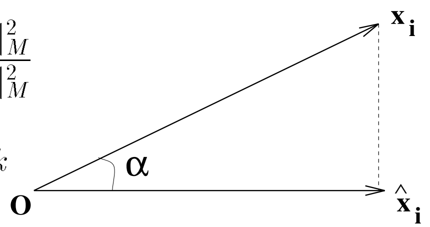
2. Qualité de l'approximation des lignes et des colonnes :

Approximation des lignes

$$\|\mathbf{x}_i\|_M^2 = \|\hat{\mathbf{x}}_i\|_M^2 + \|\mathbf{x}_i - \hat{\mathbf{x}}_i\|_M^2$$

$$\cos(\alpha)^2 = \cos(\mathbf{x}_i, \hat{\mathbf{x}}_i)^2 = \frac{\|\hat{x}_i\|_M^2}{\|x_i\|_M^2}$$

Carré de la proportion de la longueur de $\mathbf{x}_i$ expliquée en k dimensions.



Approximation des colonnes

$$\|\mathbf{x}^j\|_D^2 = \|\hat{\mathbf{x}}^j\|_D^2 + \|\mathbf{x}^j - \hat{\mathbf{x}}^j\|_D^2$$

$$\cos(\mathbf{x}^j, \hat{\mathbf{x}}^j)^2 = \frac{\|\hat{\mathbf{x}}^j\|_D^2}{\|\mathbf{x}^j\|_D^2}$$

Carré de la proportion de la longueur de $\mathbf{x}^j$ expliquée en k dimensions.

- Complément de l'éboulis des valeurs propres.
- Si $\mathbf{x}_i$ est proche de O, point de référence dans le calcul de l'inertie, $\|\mathbf{x}_i\|_M^2$ est presque nulle et $\|\hat{\mathbf{x}}_i\|_M^2$ de même. Dans ce cas l'indice $\cos^2$ est peu pertinent car il peut prendre des valeurs faibles alors que le point est proche du plan.
- $\cos^2 \frac{\pi}{4} = \frac{1}{2}$, $\cos^2 \frac{\pi}{6} = \frac{3}{4}$
- le **cosinus carré** d'un vecteur $\mathbf{x}^j$ avec un sous-espace W est égal à la somme des cosinus carrés du vecteur avec les vecteurs du repère orthonormé de W. C'est bien sûr la même chose pour un vecteur ligne $\mathbf{x}_i$. D'où l'édition des sommes cumulées des cosinus carrés.

Illustration numérique

Édition des contributions

Contribution des colonnes à l'inertie

colonne	facteur 1	facteur 2	facteur 3	facteur 4	facteur 5	total	vari
m	184	269	133	101	313	185	233
s	190	171	39	330	270	187	183
f	205	18	456	138	184	206	246
l	199	0	137	430	234	193	162
dm	222	542	236	0	0	229	177
total	1000	1000	1000	1000	1000	1000	1000

Contribution des lignes à l'inertie

ligne	facteur 1	facteur 2	facteur 3	facteur 4	facteur 5	total	Norme^2
JE	38	7	32	7	235	38	38
AL	69	0	2	5	8	66	66
AN	82	103	40	185	5	82	82
MO	185	276	153	178	103	187	187
DI	159	150	0	16	361	156	156
AD	89	1	342	2	234	92	92
PI	96	98	286	89	20	100	100
BR	126	35	135	238	7	124	124
EV	155	330	10	279	27	156	156
total	1000	1000	1000	1000	1000	1000	1000

Remarque : l'avant dernière colonne donne (en millième) la contribution de la ligne à l'inertie totale ($p_i\|\mathbf{x}_i\|^2_M/\|X\|^2$).
La dernière est égale à $\|\mathbf{x}_i\|^2_M/\sum\|\mathbf{x}_i\|^2_M$.

- Les contributions permettent de visualiser les variables ou les individus qui orientent l'analyse. Demander aux stagiaires de calculer, à partir des éléments de la matrice V du transparent 24, la contribution de la variable :

  math (m) au facteur 1 : $m_1(v_1^1)^2 = (-.429)^2 = .184$

  sciences (s) au facteur 1 : $m_2(v_1^2)^2 = (-.436)^2 = .190$

  français (f) au facteur 5 : $m_3(v_5^3)^2 = (.428)^2 = .184$

  On fera observer que le poids de chaque colonne est égal à 1 ($m_j = 1\ \forall j$).
- Dans les sorties, les contributions aux facteurs (cf. transparents 41 et 42) sont exprimées en millièmes. Par exemple, la colonne f contribue pour 205 millièmes à l'inertie du premier facteur. La ligne MO contribue pour 185 millièmes à l'inertie du premier facteur.
- Les 2 colonnes à droite du tableau sont identiques si et seulement si les poids sont égaux.
- Vari : valeurs proportionnelles aux variances des variables.

  Norme^2 : en fonction des versions cette colonne peut aussi être intitulée ORI. C'est la mesure de l'originalité de la ligne (carré de la distance d'une ligne au centre du nuage). ORI est égale à l'inertie totale si et seulement si les poids des individus sont égaux.

Edition des contributions

Contribution des facteurs aux colonnes

colonne	facteur 1	facteur 2	facteur 3	facteur 4	facteur 5	total
m	951.78	33.41	14.74	2.35e-02	2.98e-02	1000
s	974.53	21.12	4.24	7.63e-02	2.55e-02	1000
f	952.45	2.03	45.46	2.87e-02	1.57e-02	1000
l	985.23	0.01	14.63	9.62e-02	2.144e-02	1000
dm	924.36	54.47	21.16	8.95e-05	2.159e-05	1000
total	956.33	23.04	20.56	4.31e-02	1.76e-02	1000

Contribution des facteurs aux lignes

ligne	facteur 1	facteur 2	facteur 3	facteur 4	facteur 5	total
JE	977.82	4.38	1.76e+01	0.008	0.110	1000
AL	999.18	0.08	7.27e-01	0.003	0.002	1000
AN	960.92	29.00	9.96e+00	0.097	0.001	1000
MO	949.05	34.04	1.68e+01	0.041	0.009	1000
DI	977.73	22.21	5.46e-04	0.004	0.041	1000
AD	923.36	0.34	7.62e+01	0.001	0.044	1000
PI	918.34	22.60	5.90e+01	0.038	0.003	1000
BR	970.97	6.49	2.24e+01	0.082	0.0011	1000
EV	950.02	48.62	1.27e+00	0.076	0.003	1000
total	956.33	23.04	2.05e+01	0.043	0.017	1000

- Demander aux stagiaires de calculer, à partir des éléments des matrices X et V du transparent 24, la contribution du :

  facteur 1 à la variable *math* (m),

  $(\lambda_1 v_1^1)^2/\|\mathbf{x}^1\|_D^2 = [23.26\times(-.429)]^2/[(6^2+8^2+\ldots+13^2+9^2)/9] = 0.951$

  facteur 3 à la variable *langue étrangère* (l),

  $(\lambda_3 v_4^3)^2/\|\mathbf{x}^4\|_D^2 = [3.41\times(-.37)]^2/[(5.5^2+8^2+\ldots+9.5^2+12^2)/9] = 0.0146$

- Dans les sorties, les contributions sont en millièmes. Par exemple, le facteur 1 contribue pour 924 millièmes à la variable dm.

Qualité de la représentation des colonnes ($\cos^2$)

colonne	facteur 1	facteur 2	facteur 3	facteur 4	facteur 5
m	952	985	1000	1000	1000
s	975	996	1000	1000	1000
f	952	954	1000	1000	1000
l	985	985	1000	1000	1000
dm	924	979	1000	1000	1000
total	956	979	1000	1000	1000

Qualité de la représentation des lignes ($\cos^2$)

ligne	facteur 1	facteur 2	facteur 3	facteur 4	facteur 5
JE	978	982	1000	1000	1000
AL	999	999	1000	1000	1000
AN	961	990	1000	1000	1000
MO	949	983	1000	1000	1000
DI	978	1000	1000	1000	1000
AD	923	924	1000	1000	1000
PI	918	941	1000	1000	1000
BR	971	977	1000	1000	1000
EV	950	999	1000	1000	1000
total	956	979	1000	1000	1000

• Les tableaux contiennent les sommes cumulées des cosinus carrés, i.e. la colonne “facteur 1” contient les cosinus carrés des colonnes et des lignes avec le premier facteur, la colonne “facteur 2” contient les cosinus carrés avec le plan des facteurs 1 et 2.

Par exemple, le cosinus carré de m avec le premier facteur est de 952 millièmes et de 985 millièmes avec le plan 1×2, soit de 33 millièmes avec le facteur 2.

L'Analyse en Composantes Principales (ACP)

Particularité de l'ACP par rapport à l'analyse générale

C'est une analyse générale d'un tableau X individus $\times$ variables <u>centrées</u>.

Objectifs

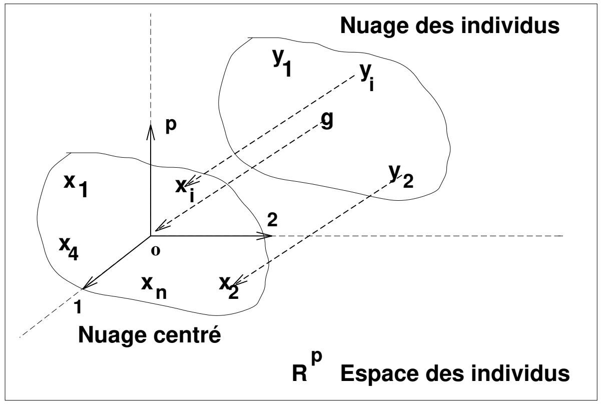
- Analyse des proximités des vecteurs individus.
- Analyse des corrélations entre les variables.
- Analyse simultanée : lorsque deux groupes d'individus se distinguent l'un de l'autre, pour quelles variables prennent-ils des valeurs différentes ?

Barycentre et nuage centré des individus

On note $Y = (y_i^j)$ (i individu, j variable) le tableau de données initial. Le nuage des individus est le nuage des vecteurs lignes $\mathbf{y}_i$ de Y dans $\mathbb{R}^p$.

Barycentre du nuage des individus

Le vecteur $\mathbf{g} = \sum_{i=1}^{n} p_i \mathbf{y}_i$ est le *barycentre* du nuage des individus.
Sa j*ième* coordonnée est $g_j = \sum_{i=1}^{n} p_i y_i^j$, moyenne pondérée de la variable $\mathbf{y}^j$.
$\mathbf{g}$ est encore l'*individu moyen* (les variables ont pour cet individu leurs valeurs moyennes).



Nuage des individus centrés

C'est le nuage des vecteurs $\mathbf{x}_i = \mathbf{y}_i - \mathbf{g}$, résultat de la translation du nuage initial par le vecteur $-\mathbf{g}$.

- $\mathbf{g}$ est un vecteur $\mathbb{R}^p$.
- Les poids sur les individus (lignes) sont positifs et leur somme est égale à 1. Ils peuvent caractériser l'importance des individus dans le n-échantillon. Si l'on n'a pas d'*a priori* particulier, on choisit $p_i = \frac{1}{n}$.
- La spécificité de l'ACP par rapport à l'analyse générale réside dans le fait de centrer les variables (colonnes) avec leur moyenne et donc d'analyser la dispersion des individus autour de l'individu moyen (être mythique s'il en est).

Le tableau centré

Tableau centré : tableau des écarts à la moyenne, ayant en position (i,j) $x_i^j = (y_i^j - \bar{y^j})$

$$Y = \begin{pmatrix} y_1^1 & \ldots & y_1^j & \ldots & y_1^p \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ y_i^1 & \ldots & y_i^j & \ldots & y_i^p \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ y_n^1 & \ldots & y_n^j & \ldots & y_n^p \end{pmatrix} \quad \text{Tableau non centré}$$

$$X = \begin{pmatrix} y_1^1 - \bar{y}^1 & \ldots & y_1^j - \bar{y}^j & \ldots & y_1^p - \bar{y}^p \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ y_i^1 - \bar{y}^1 & \ldots & y_i^j - \bar{y}^j & \ldots & y_i^p - \bar{y}^p \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ y_n^1 - \bar{y}^1 & \ldots & y_n^j - \bar{y}^j & \ldots & y_n^p - \bar{y}^p \end{pmatrix} \quad \text{Tableau centré}$$

- $\mathbf{g} = \begin{pmatrix} \bar{y}^1 \\ \vdots \\ \bar{y}^p \end{pmatrix}$
- remarquer que la somme pondérée des écarts à la moyenne est nulle : $\sum_{i=1}^{n} p_i(y_i^j - \bar{y}^j) = 0$.

Exemple

nom	m	s	f	l	dm
JE	6.0	6.0	5.0	5.5	8.0
AL	8.0	8.0	8.0	8.0	9.0
AN	6.0	7.0	11.0	9.5	11.0
MO	14.5	14.5	15.5	15.0	8.0
DI	14.0	14.0	12.0	12.5	10.0
AD	11.0	10.0	5.5	7.0	13.0
PI	5.5	7.0	14.0	11.5	10.0
BR	13.0	12.5	8.5	9.5	12.0
EV	9.0	9.5	12.5	12.0	18.0
g	9.7	9.8	10.2	10.1	11.0

tableau initial

← barycentre

nom	m	s	f	l	dm
JE	-3.7	-3.8	-5.2	-4.6	-3.0
AL	-1.7	-1.8	-2.2	-2.1	-2.0
AN	-3.7	-2.8	0.8	-0.6	0.0
MO	4.8	4.7	5.3	4.9	-3.0
DI	4.3	4.2	1.8	2.4	-1.0
AD	1.3	0.2	-4.7	-3.1	2.0
PI	-4.2	-2.8	3.8	1.4	-1.0
BR	3.3	2.7	-1.7	-0.6	1.0
EV	-0.7	-0.3	2.3	1.9	7.0

tableau <u>centré</u>

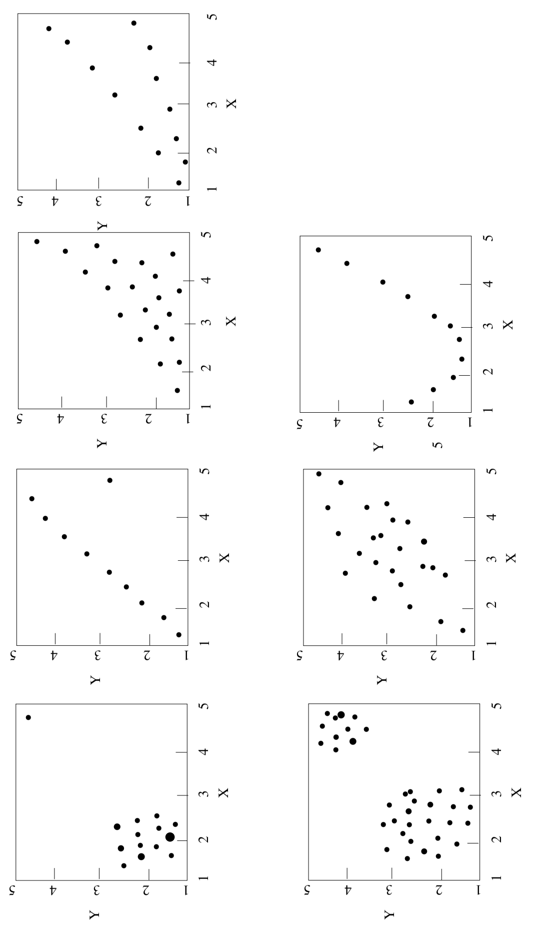
- $p_i = \frac{1}{9}$
- l'ACP agit donc sur ce tableau des écarts au barycentre.

Le centrage dans l'ACP (rappel)

- centrage avec la moyenne
- interprétation du nuage des variables (covariance, corrélation, variance)
- le centrage ne modifie pas la distance entre lignes du tableau initial

- Le fait de travailler avec des variables centrées avec leur moyenne permettra l'interprétation des normes et produit scalaire dans l'espace des variables muni de la métrique D des poids p_i en termes de variance, covariance ou corrélation linéaire.
  $var(\mathbf{y}^j) = \|\mathbf{x}^j\|_D^2$
  $cov(\mathbf{y}^j, \mathbf{y}^{j'}) = < \mathbf{x}^j, \mathbf{x}^{j'} >_D$
  $\rho(\mathbf{y}^j, \mathbf{y}^{j'}) = \dfrac{< \mathbf{x}^j, \mathbf{x}^{j'} >_D}{\|\mathbf{x}^j\|_D \|\mathbf{x}^{j'}\|_D} = \cos(\mathbf{x}^j, \mathbf{x}^{j'})$
- $\|\mathbf{y}_i - \mathbf{y}_{i'}\|_M = \|(\mathbf{y}_i - \mathbf{g}) - (\mathbf{y}_{i'} - \mathbf{g})\|_M$

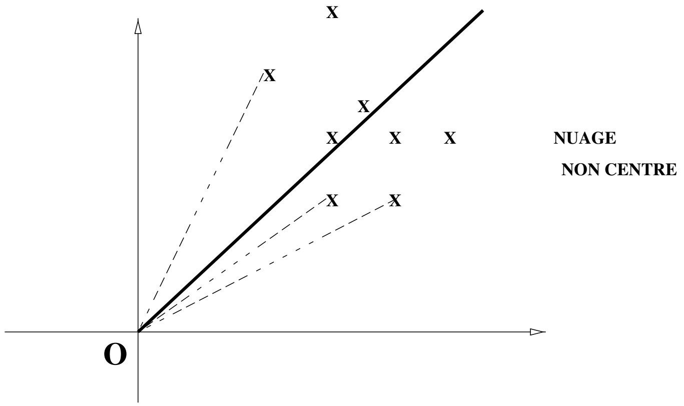
Nuage de points et valeur du coefficient de corrélation

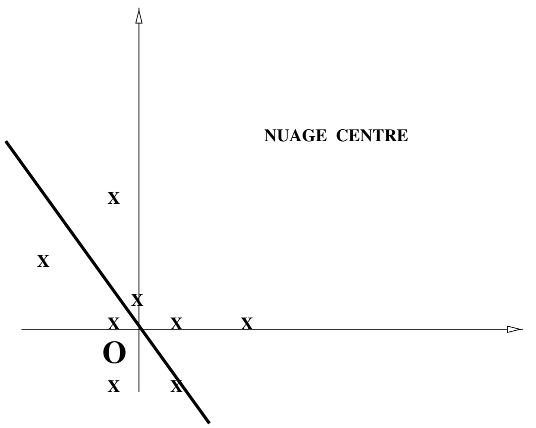


Corrélation linéaire de 0.70 dans chaque cas

- L'analyse du nuage des variables s'effectue à l'aide des corrélations (linéaires 2 à 2 des variables). Si les liaisons ne sont pas linéaires, même après transformation algébrique, la corrélation n'est pas pertinente pour analyser ce nuage. Montrer le transparent sur les formes de nuages associés à une valeur commune de la corrélation.

Effet du centrage





Inertie totale $= \sum_{i=1}^{n} p_i \|\mathbf{y}_i - \mathbf{g}\|^2 = \sum_{j=1}^{p} m_j \|\mathbf{x}^j\|^2$

- L'axe de plus grande inertie est tracé en gras.

Nuage non centré : les projections des points sur l'axe reproduisent au mieux les vecteurs initiaux.

Nuage centré : les projections des points sur l'axe reproduisent au mieux les <u>différences</u> entre les points individus (reproduction de la variabilité inter individus).
$\|\mathbf{x}_i\|^2 = \|\mathbf{y}_i - \mathbf{g}\|^2$ représente l'originalité de l'individu i (sa distance au carré à l'individu moyen).
$\|\mathbf{x}^j\|^2 = var(\mathbf{y}^j)$

Pratique de l'ACP : bilan des choix initiaux

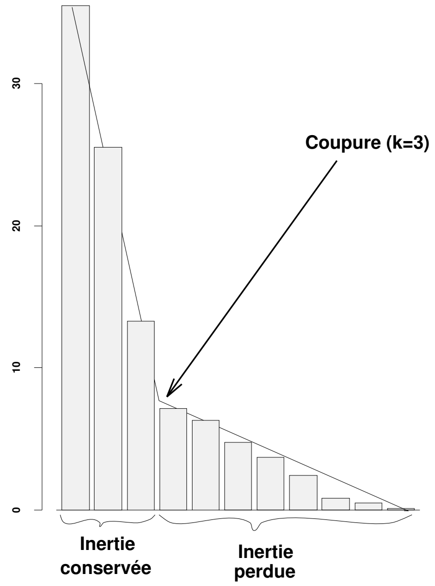
- X tableau de variables quantitatives ?
- unités identiques ?
- homogénéiser les variances ?
- définition de la métrique des variables
- définition de la métrique des individus
- choix les plus fréquents :
  ACP centrée ou **ACP centrée réduite**

- Si les unités des variables sont différentes, on peut éliminer l'effet de l'unité de mesure sur la distance entre individus avec la métrique M. Un choix classique consiste à diviser chaque variable par son écart-type σ_j. Cette opération revient à prendre $m_j = \dfrac{1}{(\sigma_j)^2}$ comme élément diagonal de la métrique M. Ou bien, on réduit les variables et on utilise la métrique $(m_j = 1)$. Ces deux façons de procéder sont équivalentes.
- On dira que l'on effectue une **ACP centrée réduite**. Les variances des variables seront égales à 1. La somme des valeurs propres sera égale au nombre de variables soit p.

Analyse des sorties de l'ACP

Choix du nombre de facteurs

But : obtenir le maximum d'inertie conservée avec le minimum de facteurs.
Règle du "coude" : on coupe l'éboulis des valeurs propres à l'endroit où celui-ci possède un "coude".



Le choix dépend des objectifs de l'analyse :
description des données $\longrightarrow$ en général au plus 4 ou 5 facteurs (difficultés d'interprétation).
Compression ou recodage $\longrightarrow$ ce nombre peut-être très grand.

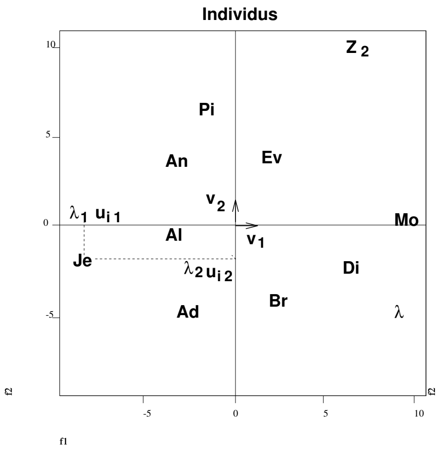
- On utilise la règle empirique du coude déjà vue dans l'analyse générale. Des tables établies à l'issue de simulations par Lebart, donnent, sous l'hypothèse d'indépendance des variables, la variabilité de la première et de la deuxième valeur propre selon le nombre de variables et d'individus. Le formateur pourra éventuellement s'y reporter (annexe 2).

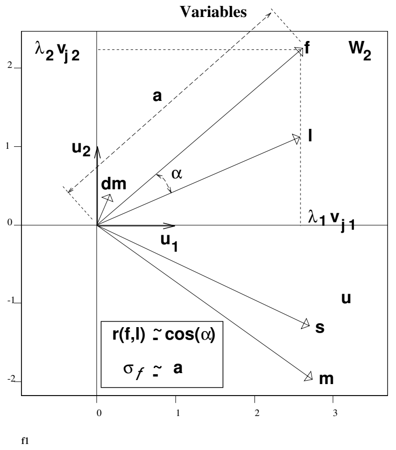
ACP non réduite

Les représentations graphiques

Représentation, la plus fidèle possible, du nuage centré des individus :

Représentation, la plus fidèle possible, du nuage des variables centrées :





- Un individu i a pour coordonnée $\lambda_1 u_i^1$ sur $\mathbf{v}_1$.
- Une variable j a pour coordonnée $\lambda_1 v_1^j$ sur $\mathbf{u}^1$.

On dit que ces représentations sont isométriques : les distances se lisent directement sur le graphique, sans distorsion entre les axes.

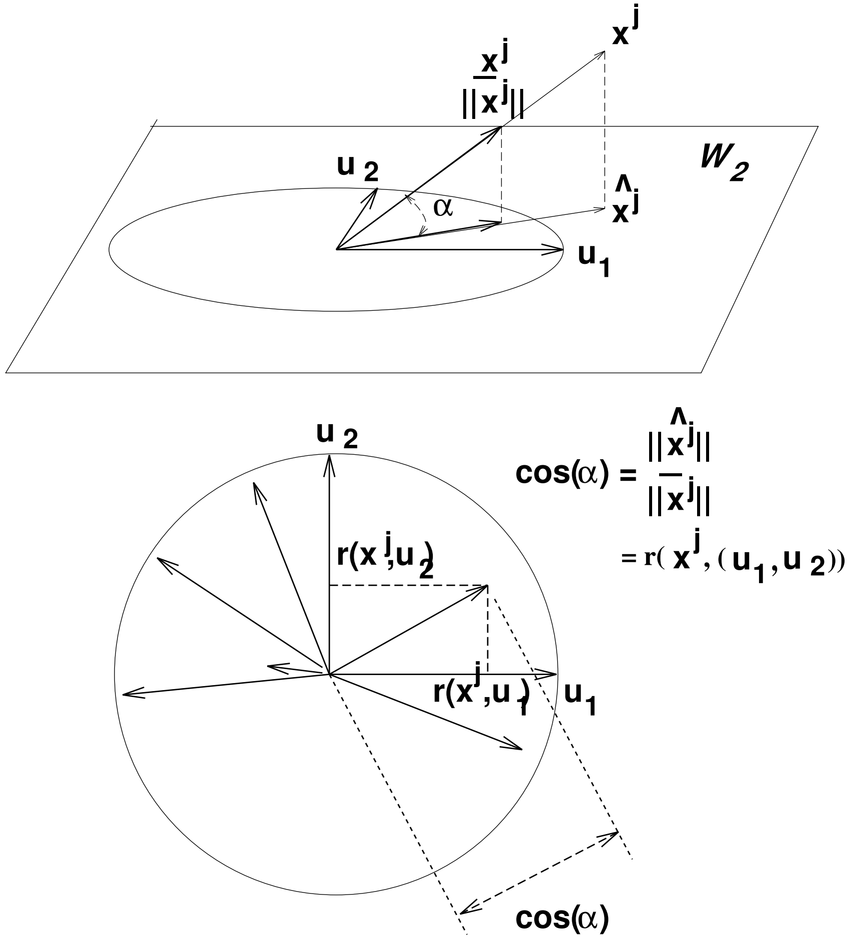
- Remarquer que l'origine des axes du nuage des individus est placée au barycentre du nuage. On a donc bien une image des écarts des points à ce barycentre.
- La longueur des vecteurs associés aux variables est proportionnelle à l'écart-type de la variable considérée. Le cosinus de l'angle α s'interprète comme une corrélation. Un angle droit entre 2 variables permet de penser que la corrélation est nulle et qu'il n'y a donc pas de liaison linéaire.

  Après projection, si la déformation du nuage n'est pas trop grande, on garde les mêmes interprétations.

  Z_2 est le plan principal (des individus) engendré par $\mathbf{v}_1$ et $\mathbf{v}_2$.

  W_2 est le plan principal (des variables) engendré par $\mathbf{u}_1$ et $\mathbf{u}_2$.

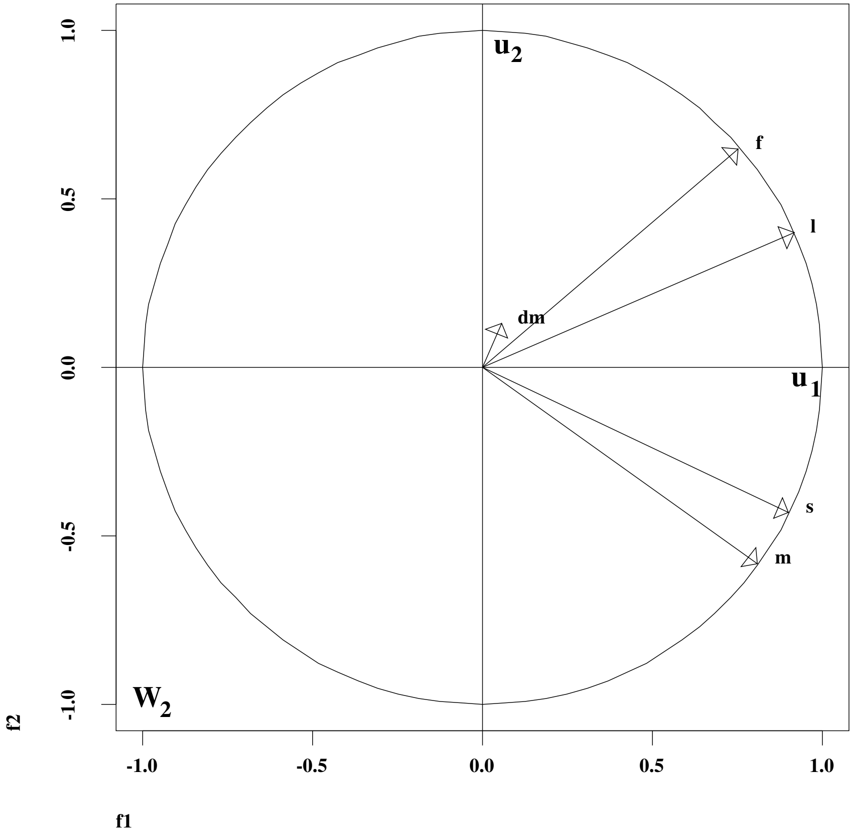
Le cercle des corrélations



- α angle entre $\mathbf{x}^j$ et sa projection $\hat{\mathbf{x}}^j$ sur le plan factoriel; $\cos(\alpha) = \dfrac{\|\hat{\mathbf{x}}^j\|_D}{\|\mathbf{x}^j\|_D}$. La projection de la variable x^j réduite, $\dfrac{\mathbf{x}^j}{\|\mathbf{x}^j\|_D}$, est égale à $\dfrac{\hat{\mathbf{x}}^j}{\|\mathbf{x}^j\|_D}$, la norme d'une variable étant l'écart-type. Par conséquent $\cos(\alpha)$ est aussi la norme, au signe près, de la projection du vecteur unitaire $\dfrac{\mathbf{x}^j}{\|\mathbf{x}^j\|_D}$.
- Les coordonnées du vecteur unitaire $\dfrac{\mathbf{x}^j}{\|\mathbf{x}^j\|_D}$ dans le plan sont données par les corrélations avec $\mathbf{u}^1$ et $\mathbf{u}^2$. Le **cercle des corrélations** visualise les projections des variables réduites dans le plan factoriel. La qualité de la représentation est lisible par la proximité du vecteur avec le cercle de rayon 1. Si le vecteur est proche du cercle unité la représentation est bonne. De plus les listes de l'indice $\cos^2$ sont fournies en sortie des programmes d'ACP, comme on l'a vu pour l'analyse générale.
- Cette représentation est la plus utilisée que l'on fasse l'ACP réduite ou non.

Exemple

Cercle des correlations



Ici, les variables f, l, m, s, sont dans W_2 (pas de déformation). Par contre dm est presque non corrélée avec le plan. On peut lire, pour les facteurs 1 et 2, les corrélations variables facteurs, les corrélations entre toutes les variables deux à deux, et la qualité des représentations.

Pratique de l'ACP : interprétation des facteurs

- Variables qui contribuent le plus au facteur.
- Fortes corrélations des variables au facteur.
- Fortes contributions des individus.
- Facteur **taille** : les corrélations de toutes les variables sont de même signe.
- Facteur **forme** : les variables avec des corrélations positives avec le facteur sont opposées à des variables avec des corrélations négatives avec le facteur.

Double décomposition de l'inertie (colonnes × facteurs)

Facteur	1	...	s	...	k	Sous-total	...	r	Total
Colonne									
1	...	...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...	...
j	...	...	$m_j(\lambda_s v_s^j)^2$	...	...	$m_j \Vert \widehat{\mathbf{x}^j}^k \Vert_D^2$	...	...	$m_j \Vert \mathbf{x}^j \Vert_D^2$
...	...	...	...	...	...	...	...	...	...
J	...	...	...	...	...	...	...	...	...
Total	λ_1^2	...	λ_s^2	...	λ_k^2	$\Vert \widehat{X}^k \Vert^2$	...	λ_r^2	$\Vert X \Vert^2$

Table 1 : la décomposition "colonne × facteur" de l'inertie.

Double décomposition de l'inertie (lignes × facteurs)

Facteur	1	...	s	...	k	Sous-total	...	r	Total
Individu									
1	...	...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...	...
i	...	...	$p_i(\lambda_s u_i^s)^2$	...	...	$p_i \Vert \widehat{x_i}^k \Vert_M^2$	...	...	$p_i \Vert x_i \Vert_M^2$
...	...	...	...	...	...	...	...	...	...
n	...	...	...	...	...	...	...	...	...
Total	λ_1^2	...	λ_s^2	...	λ_k^2	$\Vert \widehat{X}^k \Vert^2$	...	λ_r^2	$\Vert X \Vert^2$

Table 2 : la décomposition "individu × facteur" de l'inertie.

- L'analyse des contributions pour les variables fait référence au tableau du transparent 40, pour les individus au transparent 41. Transparents auxquels on se réfèrera souvent dans la pratique.
- L'examen des contributions a pour but de détecter notamment les individus ou variables qui contribuent beaucoup à l'inertie d'un facteur. Une fois repérés, on leur fait subir un traitement particulier qui fait l'objet d'un transparent suivant.

Pratique de l'ACP : individus et variables supplémentaires

- Représenter individus et variables non utilisés dans l'analyse.

  Exemple : individus mis à l'écart après une $1^{ère}$ analyse, point moyen d'un ensemble d'individus, variables illustratives, ...

Méthode :

- Individu supplémentaire (le vecteur associé a pour coordonnées les valeurs des p variables pour cet individu). On fait subir à ce vecteur les mêmes transformations que les vecteurs initiaux (centrage : $y^j_{sup} \rightarrow x^j_{sup} = y^j_{sup} - \bar{y^j}$, éventuellement réduction). Puis on projette la ligne transformée $\mathbf{x}_{sup}$ sur le sous-espace principal.

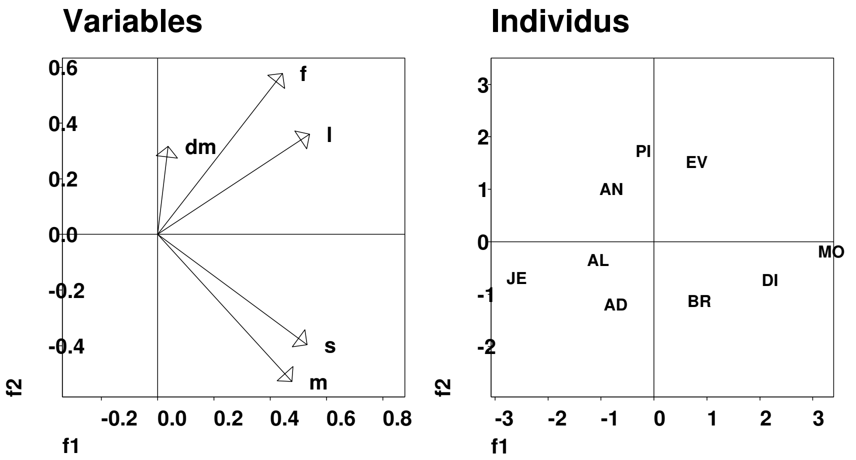
  Les coordonnées de $\mathbf{x}_{sup}$ sur $\mathbf{v}_s$ sont ${}^t\mathbf{v}_s M \mathbf{x}_{sup} = \sum_j m_j v^j_s x_{sup}$.

  ${}^t\mathbf{v}_s M$ est le projecteur orthogonal sur v_s.
- Variable continue supplémentaire : même démarche que pour un vecteur individu : centrage et éventuellement réduction, puis projection sur le sous-espace principal W_k.

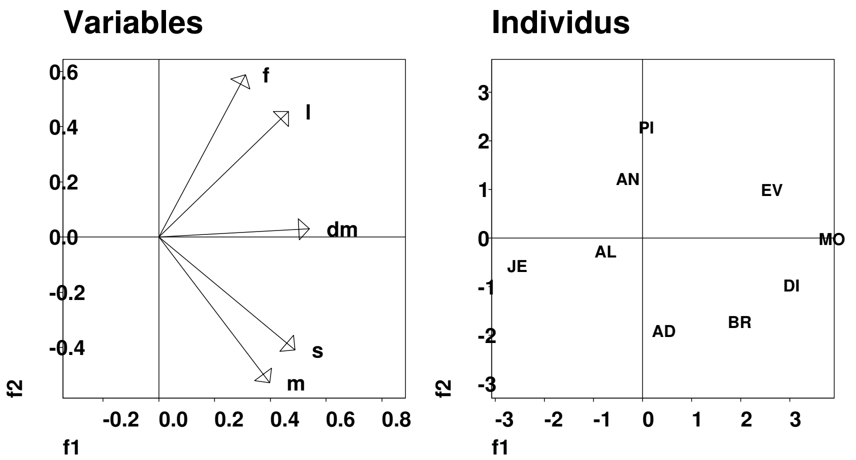
  La projection de $\mathbf{x}^{sup}$ sur W_k s'écrit $\sum_{s=1}^{k} ({}^t\mathbf{u}^s D \mathbf{x}^{sup})\mathbf{u}^s$.
- Variable catégorielle illustrative : elle définit des sous-populations (ensemble d'individus ayant la même modalité). On représente en individu supplémentaire les barycentres de ces groupes, ou encore les contours des individus d'une même sous-population.

- Bien noter que la moyenne utilisée dans le centrage des individus supplémentaires est calculée bien évidemment sans ces individus.
- La variance, en cas de réduction, est celle calculée sans les individus supplémentaires.
- Une variable qualitative ne peut être utilisée comme une variable quantitative. On ne peut analyser ses corrélations avec les autres variables.

ACP **centrée réduite** sur le tableau "note" **complet**



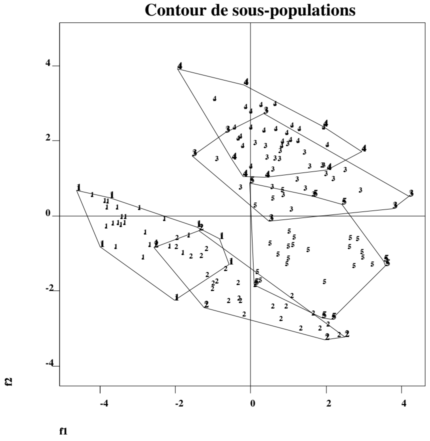
ACP centrée réduite sur le tableau "note" avec MO et DI individus supplémentaires

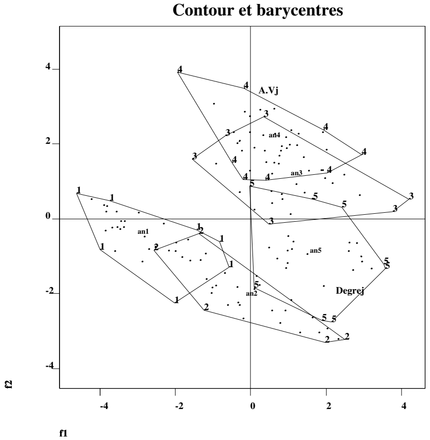


L'objectif du transparent est de montrer comment les graphiques **Variables** et **Individus** sont modifiés quand certains individus ne sont pas pris en compte dans l'analyse mais sont considérés comme des individus supplémentaires.

À gauche, représentation non isométrique (donc pas de cercle de corrélation) : les coordonnées de $\mathbf{x}^j$ sont (v_1^j, v_2^j).

À droite, représentation isométrique : les coordonnées de $\mathbf{x}_i$ sont $(\lambda_1 u_i^1, \lambda_2 u_i^2)$.





Représentation de variables illustratives

- On trace l'enveloppe convexe des individus associés à une certaine modalité de la variable illustrative qualitative.
- Le centre de gravité (point moyen calculé sur les coordonnées) est aussi dessiné dans chaque sous-ensemble.
- Nous allons chercher maintenant une **représentation simultanée des individus et variables.**

Le BIPLOT d'une matrice de rang 2

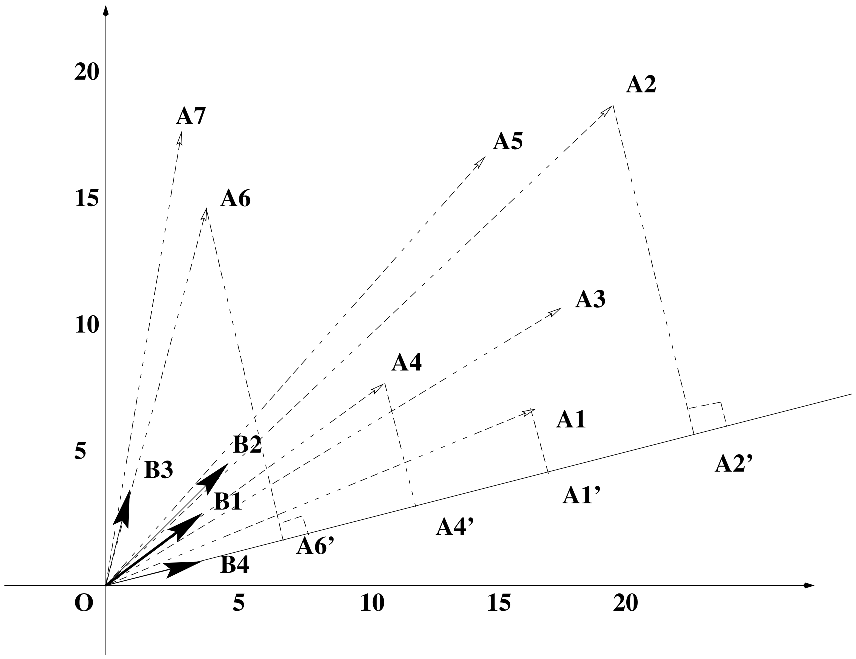
- Soit X de rang 2 telle que $X = \mathbf{u}^1\, {}^t\mathbf{v}_1 + \mathbf{u}^2\, {}^t\mathbf{v}_2 = U\, {}^tV$
- Par exemple :

$$X = \begin{pmatrix} 89 & 120 & 45 & 75 \\ 137 & 195 & 96 & 99 \\ 105 & 145 & 62 & 83 \\ 68 & 95 & 43 & 52 \\ 111 & 160 & 83 & 77 \\ 61 & 95 & 64 & 31 \\ 66 & 105 & 75 & 30 \\ 81 & 120 & 69 & 51 \\ 98 & 145 & 83 & 62 \\ 57 & 85 & 50 & 35 \end{pmatrix} = \begin{pmatrix} 17 & 7 \\ 20 & 19 \\ 18 & 11 \\ 11 & 8 \\ 15 & 17 \\ 4 & 15 \\ 3 & 18 \\ 9 & 15 \\ 11 & 18 \\ 6 & 11 \end{pmatrix} \begin{pmatrix} 4 & 5 & 1 & 4 \\ 3 & 5 & 4 & 1 \end{pmatrix}$$

$$x_3^2 = 145 = 18 \times 5 + 11 \times 5$$

Le BIPLOT d'une matrice de rang 2 (suite)

- A_i points ayant la ligne i de U pour coordonnées
- B_j points ayant la ligne j de V pour coordonnées



- si A_i' projection orthogonale de A_i sur l'axe OB_j

$$\begin{aligned} x_i^j &= < OA_i, OB_j > = \|OA_i\| \|OB_j\| \cos\alpha \\ &= \|OA_i'\| \|OB_j\| signe(\cos\alpha) \end{aligned}$$

- Lorsque l'angle (OA_i, OB_j) est aigu, la valeur x_i^j de la variable j pour l'individu i est positive.
  C'est le cas ici pour toutes les paires.
- Lorsque l'angle (OA_i, OB_j) est obtus, x_i^j est négatif.
- Plus un point A_i est éloigné dans la direction d'une flèche OB_j, plus la valeur de la variable j pour cet individu i est grande.
- Par exemple, projetons orthogonalement les points A_i sur l'axe engendré par B_4.
  Les projections A_i' s'ordonnent sur l'axe OB_4 selon les valeurs de x_i^4 que le formateur écrira sur le transparent au cours de l'exposé. Si le vecteur OB_4 était de norme 1, on aurait pu lire sur le graphique les valeurs de x_i^4.

Le BIPLOT : application à l'ACP

- soit l'approximation de rang 2 du tableau X
- 3 écritures de la décomposition :

$$\hat{x}_i^{j\,2} = \sum_{s=1}^{2} (\lambda_s u_s^i)(v_j^s) \quad (1)$$

$$= \sum_{s=1}^{2} (u_s^i)(\lambda_s v_j^s) \quad (2)$$

$$= \sum_{s=1}^{2} (\sqrt{\lambda_s} u_s^i)(\sqrt{\lambda_s} v_j^s) \quad (3)$$

(1) $A_i = \begin{pmatrix} \lambda_1 u_1^i \\ \lambda_2 u_2^i \end{pmatrix}$, $B_j = \begin{pmatrix} v_j^1 \\ v_j^2 \end{pmatrix}$, biplot **isométrique lignes**

(2) $A_i = \begin{pmatrix} u_1^i \\ u_2^i \end{pmatrix}$, $B_j = \begin{pmatrix} \lambda_1 v_j^1 \\ \lambda_2 v_j^2 \end{pmatrix}$, biplot **isométrique colonnes**

(3) $A_i = \begin{pmatrix} \sqrt{\lambda_1} u_1^i \\ \sqrt{\lambda_2} u_2^i \end{pmatrix}$, $B_j = \begin{pmatrix} \sqrt{\lambda_1} v_j^1 \\ \sqrt{\lambda_2} v_j^2 \end{pmatrix}$, biplot **symétrique**

- Quand le tableau X est de rang quelconque, le biplot sera utilisé pour représenter une approximation de rang 2 de X. Dans le cas de l'ACP, X est le tableau centré et réduit si nécessaire.
- On connait l'écriture de cette approximation donnée par la DVSG de X avec les métriques adéquates.
- La décomposition peut être écrite, notamment, de trois façons :

(1) nommée **biplot isométrique ligne**, c'est la représentation optimale des lignes au sens de la conservation des distances des individus. En effet les points A_i de coordonnées $(\lambda_1 u_i^1,\ \lambda_2 u_i^2)$ représentent au mieux les individus dans le premier plan factoriel. Ce n'est pas le cas pour les points B_j de coordonnées $(v_1^j,\ v_2^j)$ car il manque le coefficient en λ. Les distances sur l'axe 1 ne pourront pas être comparées à celles sur l'axe 2. Ceci dit, si λ_1 est proche de λ_2 la forme du nuage des variables est respectée à une homothétie près. Si les valeurs singulières ne sont pas trop différentes la déformation ne sera donc pas trop importante.

(2) nommée **biplot isométrique colonne**, c'est le cas symétrique du précédent. Cette fois on choisit de représenter de façon optimale les variables.

(3) nommée **biplot symétrique (ou à la hollandaise)**, compromis entre les deux représentations précédentes; pas d'optimalité pour les individus et les variables.

Exemple

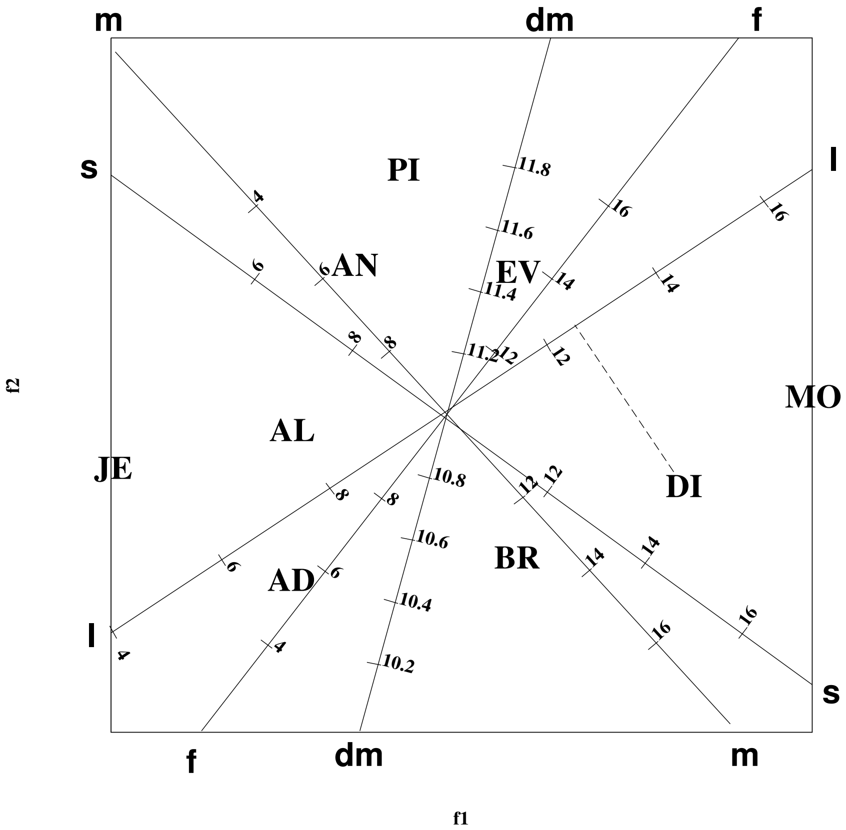
Notes centrées

nom	m	s	f	l	dm
JE	-3.7	-3.8	-5.2	-4.6	-3.0
AL	-1.7	-1.8	-2.2	-2.1	-2.0
AN	-3.7	-2.8	0.8	-0.6	0.0
MO	4.8	4.7	5.3	4.9	-3.0
DI	4.3	4.2	1.8	2.4	-1.0
AD	1.3	0.2	-4.7	-3.1	2.0
PI	-4.2	-2.8	3.8	1.4	-1.0
BR	3.3	2.7	-1.7	-0.6	1.0
EV	-0.7	-0.3	2.3	1.9	7.0

Approximation de rang 2

nom	m	s	f	l	dm
JE	-3.5	-3.8	-5.4	-4.8	-0.5
AL	-1.6	-1.7	-2.4	-2.1	-0.2
AN	-3.6	-2.9	0.7	-0.4	0.3
MO	4.9	4.9	4.9	4.8	0.3
DI	4.5	4.0	1.7	2.4	-0.1
AD	1.1	0.2	-4.5	-2.9	-0.6
PI	-4.1	-2.8	3.5	1.5	0.7
BR	3.3	2.5	-1.7	-0.4	-0.4
EV	-1.0	-0.4	3.1	2.0	0.4

Biplot gradué isométrique ligne



- Représentation optimale des individus.
- Des axes gradués avec l'unité de mesure originelle des variables permettent de lire, selon le principe déjà exposé, la valeur d'un individu sur les variables selon l'approximation d'ordre 2 choisie en le projetant sur les différents axes.
- Les valeurs lues sont d'autant plus proches des valeurs du tableau initial que le plan a une forte inertie.