Analyse de données

Analyse par Composantes Principales

Jamal Atif
jamal.atif@dauphine.fr

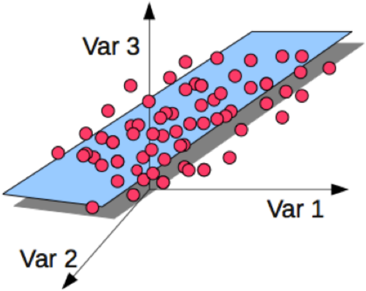
Université Paris-Dauphine, Licence MIDO

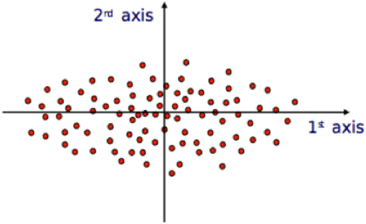
2015-2016

Copyright

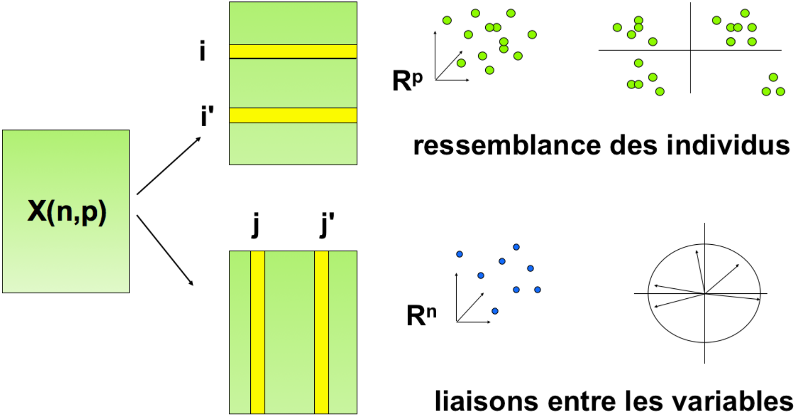
Ce cours est adapté librement des ressources des années précédentes. Remerciements particuliers à Patrice Bertrand.

L'ACP, principe

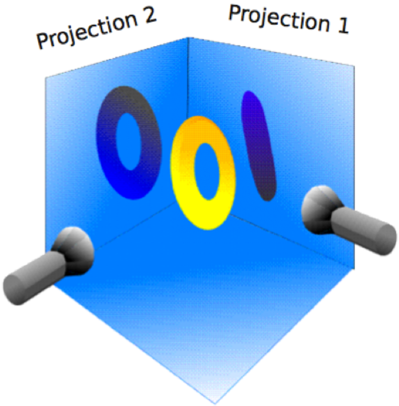




Deux espaces de représentation



Critères à optimiser

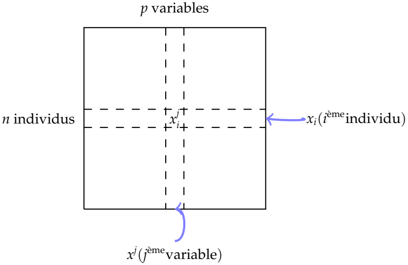


Nuage en projection le plus étalé possible.

$$\Updownarrow$$

Inertie par rapport au point G du nuage projeté maximum.

Données : nuages de points



Chaque individu i est muni d'un poids p_i

$$\forall i, p_i > 0 \text{ et } \sum_{i=1}^{n} p_i = 1 \,.$$

$$\boxed{\mathcal{N}_I = \{(x_i, p_i) \mid i = 1, \cdots, n\} \subset R^p}$$

Nuage des individus

Espace des individus

$$\boxed{\mathcal{N}_J = \{(x^j) \mid j = 1, \cdots, p\} \subset R^m}$$

Nuage des variables

Espace des variables

Exemple

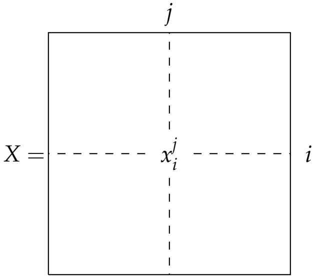
Tableau : notes de 9 élèves (Jean, Aline, Annie, Monique, Didier, André, Pierre, Brigitte, Evelyne) dans 5 matières (Math., Sciences, Français, Latin, Dessin-Musique)

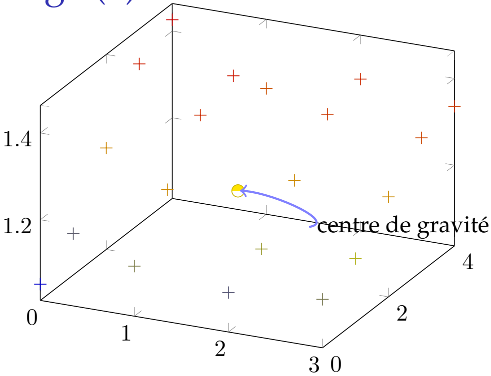
	M	S	F	L	DM
JE	6	6	5	5,5	8
AN	8	8	8	8	9
AN	6	7	11	9,5	11
MO	14,5	14,5	15,5	15	8
DI	14	14	12	12,5	10
AD	11	10	5,5	7	13
PI	5,5	7	14	11,5	10
BR	13	12,5	8,5	9,5	12
EV	9	9,5	12,5	12	18

$$\text{Jean} \hookrightarrow x_1 = \begin{pmatrix} 6 \\ 6 \\ 5 \\ 5{,}5 \\ 8 \end{pmatrix} \text{ ici} \qquad \text{Math} \hookrightarrow x_1 = \begin{pmatrix} 6 \\ 8 \\ \vdots \\ 9{,}5 \end{pmatrix}$$

$p_i = \frac{1}{9}$

Caractéristiques d'un nuage (1)



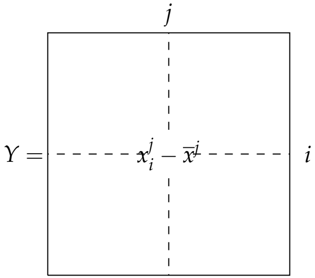


$$\mathcal{N} = \mathcal{N}_I = \{x_i \mid i = 1, \cdots, n\}, p_i \geq 0, \sum_i p_i = 1$$

Centre de gravité

$$g = \sum_{i=1}^{n} p_i x_i = \begin{pmatrix} \sum_i p_i x_i^1 \\ \vdots \\ \sum_i p_i x_i^p \end{pmatrix} = \begin{pmatrix} \overline{x}^1 \\ \vdots \\ \overline{x}^p \end{pmatrix}$$

Caractéristiques d'un nuage (2)



Exemple (2)

$$g = \frac{1}{9}\begin{pmatrix} 6 \\ 6 \\ 5 \\ 5,5 \\ 8 \end{pmatrix} + \frac{1}{9}\begin{pmatrix} 8 \\ 8 \\ 8 \\ 8 \\ 9 \end{pmatrix} + \cdots + \frac{1}{9}\begin{pmatrix} 9 \\ 9,5 \\ 12,5 \\ 12 \\ 18 \end{pmatrix}$$

$$g = \begin{pmatrix} 9,67 \\ 9,83 \\ 10,2 \\ 10,1 \\ 11,0 \end{pmatrix} = \begin{pmatrix} g^1 \\ g^2 \\ g^3 \\ g^4 \\ g^5 \end{pmatrix}$$

g correspondant à l'"élève moyen".

Métrique

$$d^2(x_1, x_2) = (x_1^1 - x_2^1)^2 + (x_1^2 - x_2^2)^2 + \cdots$$

physique : dimensions de même nature

$$d^2(x_1, x_2) = a_1(x_1^1 - x_2^1)^2 + a_2(x_1^2 - x_2^2)^2 + \cdots$$

statistique : donner à chaque attribut une importance différente.

$$\begin{aligned} d^2(x_1, x_2) &= (x_1 - x_2)^t M (x_1 - x_2) \\ &= \sum_{i=1}^{p} \sum_{j=1}^{p} m_{ij}(x_1^i - x_1^i)(x_1^j - x_1^j) \\ &= ||x_1 - x_2||_M^2 \end{aligned}$$

où M est symétrique, définie positive (cas d'axes obliques, i.e. non perpendiculaires).

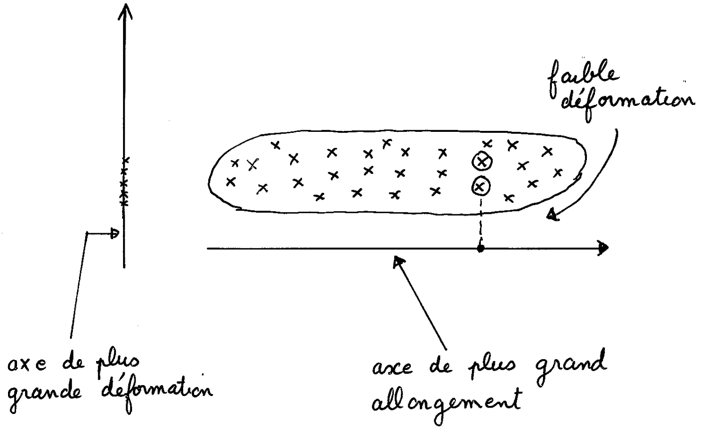
En pratique :

$$M = I, \quad M = D_{1/s^2} = \begin{pmatrix} 1/s_1^2 & 0 & \cdots & 0 \\ 0 & 1/s_2^2 & \cdots & 0 \\ \vdots & \cdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1/s_p^2 \end{pmatrix}$$

Formulation du critère

On recherche l'espace vectoriel V de $\mathbb{R}^p$ de faible dimension tel que par projection, le nuage $\mathcal{N}_I$ soit le moins déformé possible.

Définition intuitive des différents types d'axes



Déformation par projection

- P_V : projecteur orthogonal sur V. On note $P_V(x_i) = \tilde{x}_i$
- Déformation par projection orthogonale sur V

$$D_V = \sum_{i \neq j} p_i p_j [d^2(x_i, x_j) - d^2(\tilde{x}_i - \tilde{x}_j)]$$

$$\Updownarrow$$

$$D_V = \sum_i p_i \left(d^2(x_i, g) - d^2(\tilde{x}_i, \tilde{g}) \right)$$

- Sous-espace principal de dimension k, V_k : sous-espace de $\mathbb{R}^p$ de dimension k qui rend la déformation D_V minimum
- (Premier) axe principal : sous-espace principal de dimension 1
- Plan principal : sous-espace principal de dimension 2

Formulation du critère à l'aide de l'inertie

Propriété

$$D_V = I_g(\mathcal{N}_I) - I_{\tilde{g}}(\widetilde{\mathcal{N}_I})$$

où $I_g(\mathcal{N}_I) = \sum_i p_i d^2(x_i, g)$ est l'inertie du nuage $\mathcal{N}_I$ par rapport à g et $I_{\tilde{g}}(\widetilde{\mathcal{N}_I})$ et l'inertie du nuage projeté par rapport au projeté de g.

Preuve

Nous avons

$$D_V = \sum_i p_i \left(d^2(x_i, g) - d^2(\tilde{x}_i, \tilde{g})\right)$$

Remarquons que : $\sum_i p_i(\tilde{x}_i - \tilde{g}) = \sum_i p_i(\widetilde{x_i - g}) = \sum_i \widetilde{p_i(x_i - g)} = \tilde{0} = 0$. Donc $\tilde{g}$ est le barycentre de $\widetilde{\mathcal{N}_I}$.
D'où :

$$I_{\tilde{g}}(\widetilde{\mathcal{N}_I}) = \sum_i p_i d^2(\tilde{x}_i, \tilde{g})$$

Donc :

$$I_g(\mathcal{N}_I) = I_{\tilde{g}}(\widetilde{\mathcal{N}_I}) + D_V$$

Nuage centré

- L'expression de D_V ne dépend que des vecteurs $x_i - g$
- On note $y_i = x_i - g$
- Nuage centré associé à $\mathcal{N}_I$: $\{(y_i, p_i) \mid i = 1, \cdots, n\}$
- Les expressions de $I(\mathcal{N}_I)$ et $I(\widetilde{\mathcal{N}_I})$ ne changent pas si l'on centre le nuage
- Par la suite on considère uniquement le nuage centré

Décomposition de l'inertie

Propriété

Si V est un s.e.v. de $\mathbb{R}^p$ qui est somme directe de V_1 et V_2, alors :

$$I(\widetilde{\mathcal{N}_I}^{V}) = I(\widetilde{\mathcal{N}_I}^{V_1}) + I(\widetilde{\mathcal{N}_I}^{V_2})$$

Deux conséquences

- $I(\mathcal{N}_I) = I(\widetilde{\mathcal{N}_I}^{V}) + I(\widetilde{\mathcal{N}_I}^{V^\perp})$
- Si $\{u_1, \cdots, u_k\}$ est une base orthogonale de V, alors :

$$I(\widetilde{\mathcal{N}_I}^{V}) = I(\widetilde{\mathcal{N}_I}^{u_1}) + \cdots + I(\widetilde{\mathcal{N}_I}^{u_k})$$

Recherche des sous-ensembles principaux

- Problème $\mathcal{P}_k$: déterminer un sous-espace principal V_k de dimension k
- Propriété : si k et l sont des enters tels que $0 < k < k + l \leq p$ et si V_k est solution de $\mathcal{P}_k$, alors il existe une solution V_{k+l} de $\mathcal{P}_{k+l}$ qui contient V_l.
- Conséquences :
  1. On peut choisir les sous-espaces principaux de telle sorte qu'ils soient emboîtés.
  2. Si V_{k-1} est solution de $\mathcal{P}_{k-1}$, on peut chercher V solution de $\mathcal{P}$ de la forme :

$$V = V_{k-1} \oplus \Delta u$$
$$I(\widetilde{\mathcal{N}}^V) = I(\widetilde{\mathcal{N}}^V_{k-1}) + I(\widetilde{\mathcal{N}}^{\Delta u})$$

  3. Le plan principal contient l'axe principal

Inertie du nuage le long d'une droite

$$I(\widetilde{\mathcal{N}}^u) = \sum_i p_i ||\widetilde{y}_i^u||^2 \text{ avec } \widetilde{y}_i^u = \frac{\langle y_i, u \rangle}{||u||^2} u$$

Donc :

$$I(\widetilde{\mathcal{N}}^u) = \sum_i p_i \frac{\langle y_i, u \rangle^2}{||u||^2}$$

Or $\langle y_i, u \rangle = y_i^t M u = u^t M y_i$, et on considère $||u||_M = 1$ $\Longrightarrow$

$$\begin{aligned} I(\widetilde{\mathcal{N}}^u) &= \sum_i p_i (u^t M y_i)(y_i^t M u) \\ &= \sum_i p_i (u^t M) y_i y_i^t (M u) \\ &= (u^t M) \underbrace{\left(\sum_i p_i y_i y_i^t\right)}_{\Sigma} (M u) \\ &= (u^t M) \Sigma (M u) \end{aligned}$$

$$\begin{aligned} \Sigma_{j,k} &= \sum_i p_i y_i^j y_i^k \\ &= \sum_i p_i (x_i^j - \bar{x}^j)(x_i^k - \bar{x}^k) \\ &= cov(x^j, x^k) \end{aligned}$$

Exemple (4)

$$\Sigma = \begin{pmatrix} 11,4 & 9,92 & 2,66 & 4,82 & 0,111 \\ & 8,94 & 4,12 & 5,48 & 0,056 \\ & & 12,1 & 9,29 & 0,389 \\ & & & 7,91 & 0,667 \\ & & & & 8,67 \end{pmatrix}$$

Si $v_1 = \begin{pmatrix} 0,515 \\ 0,507 \\ 0,492 \\ 0,485 \\ 0,030 \end{pmatrix}$ et $M = Id$, alors

$$I(\widetilde{\mathcal{N}}^{v_1}) = v_1^t \Sigma v_1 = 28,5$$

Ecriture matricielle

$$\Sigma = \sum_i p_i y_i y_i^t = Y^t D_p Y, \text{ avec } D_p = \text{diag}(p_i) = \begin{pmatrix} p_1 & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & p_n \end{pmatrix}$$

$$\begin{aligned}
I(\mathcal{N}_I) &= \sum_i p_i ||y_i||_M^2 \\
&= \sum_i p_i y_i^t M y_i \\
&= trace\left(\sum_i p_i y_i^t M y_i\right) \text{ puisque } trace(a) = a \\
&= \sum_i p_i trace\left(y_i^t M y_i\right) \text{ puisque } trace(\sum aX_i) = a \sum trace(X_i) \\
&= \sum_i p_i trace\left(y_i y_i^t M\right) \text{ puisque } trace(XY) = trace(YX) \\
&= trace\left(\sum_i p_i y_i y_i^t M\right) \\
&= trace(\Sigma M)
\end{aligned}$$

Ecriture matricielle

$$I(\mathcal{N}_I) = trace(\Sigma M)$$

D'où :

$$trace(\Sigma M) = u^t MVMu + I(\widetilde{\mathcal{N}_I}^{u^\perp})$$

$\widetilde{\mathcal{N}_I}^{u^\perp}$ s'interprète comme l'inertie par rapport à Δu. On note $I_{\Delta_u}(\mathcal{N}_I) = I(\widetilde{\mathcal{N}_I}^{u^\perp})$.

$$\underbrace{I_{\Delta_u}(\mathcal{N}_I)}_{\text{“déformation } D_V\text{”}} = \underbrace{trace(\Sigma M)}_{I_T\text{: inertie totale}} - u^t MVMu$$

Minimiser D_V revient à maximiser $u^t MVMu$.

1er axe factoriel

- $I_{\Delta u} = trace(\Sigma M) - u^t M \Sigma M u$

Problème

$$\hat{u} = \arg\max_u u^t M \Sigma M u$$
$$\text{s.c. } u^t M u = 1$$

Solution

Ecriture sous forme de lagrangien :

$$L(u) = u^t M \Sigma M u - \lambda(u^t M u - 1)$$

où λ est un multiplicateur de Lagrange.
Le maximum est atteint lorsque la dérivée s'annule

$$\frac{\partial L(u)}{\partial u} = 2M\Sigma M u - 2\lambda M u = 0$$
$$\implies \Sigma M u = \lambda u$$

Le maximum est donc atteint si λ est la plus grande valeur propre de VM.

k^e axe factoriel

Pour tout $k \leq p$

- Le k^e axe factoriel est orienté par le vecteur propre normé u_k de ΣM associé à la k^e plus grande valeur propre λ_k.
- $I_{\Delta u_k^{\perp}} = \lambda_k = I(\widetilde{\mathcal{N}_I}^{\Delta u_k})$
- $I_T = \sum_{k=1}^{p} \lambda_k = trace(\Sigma M)$
- La projection du nuage sur $E_k = \Delta u_1 \oplus \cdots \Delta u_k$ est optimale : c'est celle qui donne la meilleure idée du nuage dans un espace de dimension k.
- Pourcentage d'inertie expliquée :

$$\tau_k = \frac{\lambda_1 + \cdots + \lambda_k}{I_T}$$

Exemple (3)

Choix utilisateur : ici 3 axes factoriels

	1	2	3
Val P.	28.253	12.075	8.6157
τ	57.7	24.7	17.6
τ cum.	57.7	82.4	99.94

	u_1	u_2	u_3
M	.515	-.567	-.0514
S	.507	-.372	-.0145
F	.492	.650	.108
L	.485	.323	-.0226
DM	.0306	.113	- .992

TABLE : Vecteurs axiaux factoriels

Composantes principales

- L'abscisse de la projection de l'individu i sur Δu_k est :

$$c_k^i = \langle y_i, u_k \rangle_M = y_i^t M u_k$$

- c_k^i représente la coordonnée de y_i sur le k^e axe factoriel.
- c_k : k^e composante principale

$$\begin{array}{lll} c_1 = \begin{pmatrix} c_1^1 \\ \vdots \\ c_1^n \end{pmatrix} & = YMu_1 & = Yb_1 \in \mathbb{R}^n \\ \vdots & \vdots & \vdots \\ c_p = \begin{pmatrix} c_p^1 \\ \vdots \\ c_p^n \end{pmatrix} & = YMu_p & = Yb_p \in \mathbb{R}^n \end{array}$$

Exemple (6)

	c_1	c_2	c_3
JE	-8.7	-1.7	2.5
AL	-3.9	-.7	1.8
AN	-3.2	3.4	.3
MO	9.7	.2	3.3
DI	6.3	-2.1	.9
AD	-2.9	-4.6	-2.6
PI	-1	6.2	1.6
BR	1.9	-4	-1.4
EV	1.7	3.4	-6.6

Exemple (7)

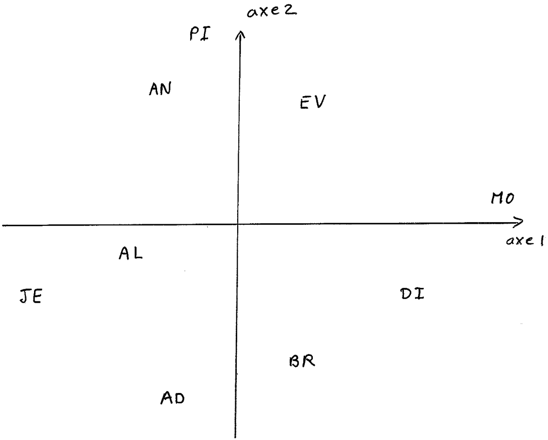


FIGURE : Plan factoriel 1×2

Exemple (8)

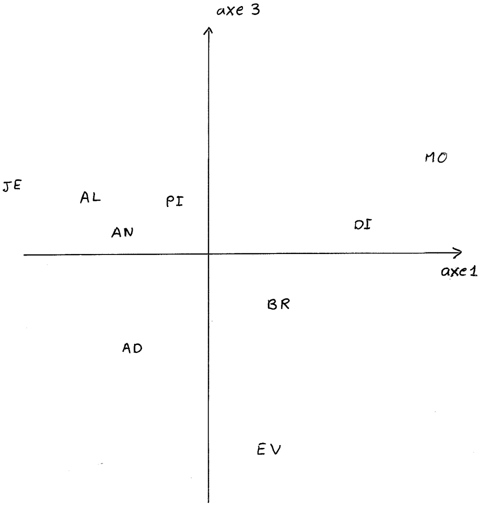


FIGURE : Plan factoriel 1×3

Représentation des variables

- Les r composantes principales, une fois normées, forment une base D_p-orthonormée du support du nuage formé par les p points $\{x^1, \cdots, x^p\} \subset \mathbb{R}^n$:

$$\{\frac{c_k}{\sqrt{\lambda_k} \mid k = 1, \cdots, n}\}$$

- La k^e coordonnée de la variable y^j dans cette base est :

$$d_j^k = (y^j)^t D_p \frac{c_k}{\sqrt{\lambda_k}} = cov(y^j, \frac{c_k}{\sqrt{\lambda_k}})$$

$$d^k = Y^t D_p \frac{c_k}{\sqrt{\lambda_k}} = \underbrace{\underbrace{Y^t D_p Y}_{\Sigma} M u_k}_{\Sigma M u_k = \lambda_k u_k} / \sqrt{\lambda_k}$$

$$\implies d^k = \sqrt{\lambda_k} u_k$$

Exemple (9)

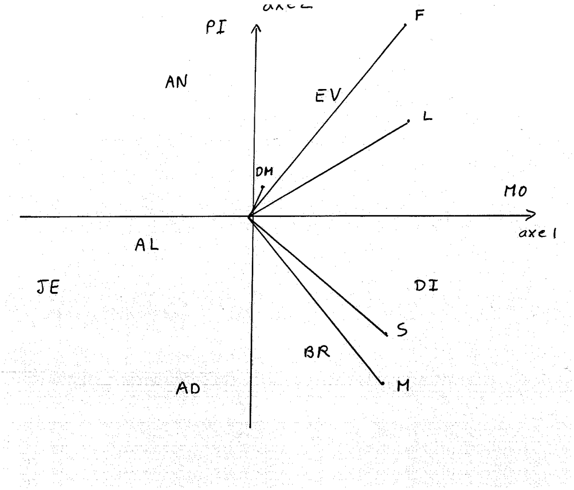


FIGURE : Plan principal 1 × 2, représentation simultanée (axe1 x axe2)

Exemple (10)

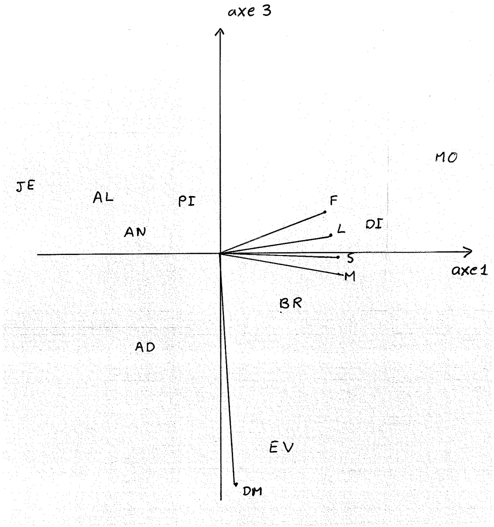


FIGURE : Plan principal 1×3, représentation simultanée (axe1 x axe3)

Aide à l'interprétation

Représentation des individus

- Qualité de la représentation par E_k :

$$\frac{\lambda_1 + \cdots + \lambda_k}{I_T}$$

- Qualité de la représentation de x_i sur Δu_k

$$COR_k(i) = \frac{(c_k^i)^2}{||y_i||_M^2} = cos^2(y_i, u_k)$$

- Contribution de x_i à l'inertie de l'axe Δu_k

$$CTR_k(i) = \frac{p_i (c_k^i)^2}{\lambda_k}$$

- Contribution de x_i à l'inertie totale

$$INR(i) = \frac{p_i ||y_i||_M^2}{I_T}$$

Aide à l'interprétation

Représentation des variables

- Qualité de la représentation de x sur $\Delta(\frac{c_k}{\sqrt{\lambda_k}}$

$$COR_k(j) = \frac{(d_j^k)^2}{||x^j||^2} = cos^2(x^j, c_k) = \rho_{j,k}^2$$

- Contribution de x^j à l'inertie de l'axe k^e axe

$$CTR_k(j) = \frac{m_j(d_j^k)^2}{\lambda_k} = m_j(u_k^j)^2$$

- Contribution de x_i à l'inertie totale

$$INR(J) = \frac{m_j||x^j||^2}{I_T}$$