

Chap6. Analyse en composantes principales (ACP)

Pr. A. Fadil

EMSI Rabat

26 décembre 2025

- 1 Définition
- 2 Principe
- 3 Démarche
- 4 Exercice d'application (Résultats et interprétation)

Définition

Qu'est-ce que l'Analyse en Composantes Principales (ACP) ?

L'Analyse en Composantes Principales (ACP) est :

- Une méthode **descriptive** a pour objectif l'analyse des tableaux de données qui ne comportent pas de structure préalable (aucune distinction ni entre variable ni entre individu).
- Elle fait partie du groupe de méthodes descriptives multidimensionnelles appelées **méthodes factorielles**.
- Le but principal est de **résumer** l'information contenue dans un tableau composé d'un nombre élevé de lignes et de colonnes.
 - ⇒ Un outil statistique de **synthèse** de l'information.
 - ⇒ Un outil très important pour traiter les données **quantitatives**.
- Elle propose à partir d'un tableau rectangulaire de données comportant les valeurs des variables quantitatives pour des individus, des représentations géométriques de ces individus et de ces variables.

Principe

L'Analyse en Composantes Principales (ACP) permet d'étudier :

- La **variabilité entre les individus**, c'est-à-dire quelles sont les différences et les ressemblances entre individus ;
- Les **liaisons entre les variables** : y-a-t-il des groupes de variables très corrélées entre elles, qui peuvent être regroupées entre de nouvelles variables synthétiques ?

D'un point de vue géométrique :

- Le nuage de points représentant les données s'inscrit dans les données s'inscrit dans un espace de p dimensions, puisque chaque point représente un individu par rapport à x_1, x_2, \dots, x_n .
- ⇒ Il est difficile de visualiser les relations existant entre les variables dès que $p > 3$.

Principe

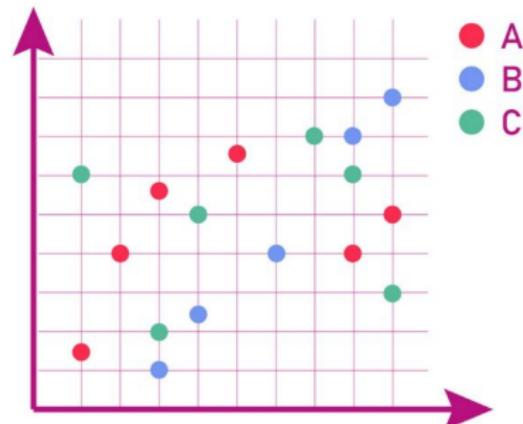


Figure – la dimension $p = 2$. Il est facile de présenter le nuage de points.

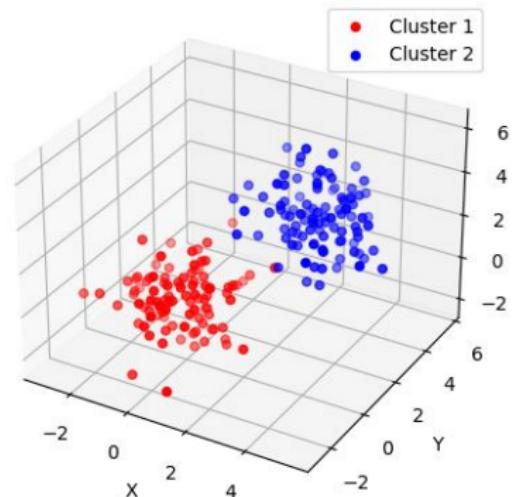


Figure – Si la dimension $p > 3$. Il est difficile de présenter le nuage de points.

Définitions

- Un individu $e_i^\top = (x_{i1}, x_{i2}, \dots, x_{ip})$ est un vecteur de \mathbb{R}^p .
L'ensemble des vecteurs $e_i, i = 1, \dots, n$ constitue le nuage des individus.
- Une variable $X_k = (x_{1k}, x_{2k}, \dots, x_{nk})^\top$, de moyenne \bar{x}_k et de variance σ_k^2 , est un vecteur de \mathbb{R}^n .
L'ensemble des vecteurs $X_k, k = 1, \dots, p$ constitue le nuage des variables.
- On définit le vecteur individu moyen comme le centre de gravité du nuage des variables

$$g^\top = (\bar{x}_1, \bar{x}_2, \dots, \bar{x}_p)$$

Notion d'inertie

- La variabilité (dispersion) des données représente l'information (structure) ou l'inertie du nuage des individus par rapport à son centre de gravité mesurée par :

$$I_g = \frac{1}{I} \sum_{i=1}^n \sum_{k=1}^p (x_{ik} - \bar{x}_k)^2 = \sum_{k=1}^p \sigma_k^2$$

- Par centrage-réduction des variables X_k tel que $Z_k = \frac{X_k - \bar{X}_k}{\sigma_k}$.
Dans ce cas :

$$I_g = p \quad g^\top = (0, 0, \dots, 0)$$

On suppose, dans la suite, que les données ont été centrées et réduites.

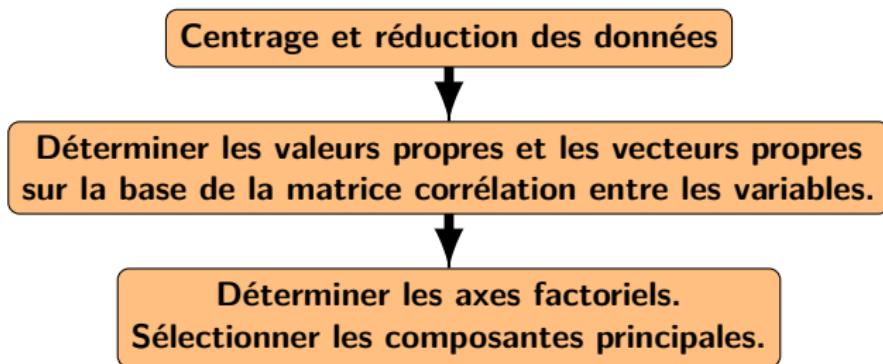
Démarche

Soit n individus (lignes) caractérisés par p variables métriques (Colonnes). Ces données sont présentées dans un tableau appelé la Matrice des données de dimension $n \times p$.

X_1	\dots	X_j	\dots	X_p
x_{11}	\dots	x_{1j}	\dots	x_{1p}
\vdots	\ddots	\vdots	\vdots	\vdots
x_{i1}	\dots	x_{ij}	\dots	x_{ip}
\vdots	\vdots	\vdots	\ddots	\vdots
x_{n1}	\dots	x_{nj}	\dots	x_{np}

Démarche

Les étapes pour déterminer la composante principale :



Centrage et réduction des données :

- Les p variables sont de nature différente, pour homogénéiser les unités, les p variables seront centrées et réduites.
- Les données sont centrées et réduites signifie que pour chaque variable la moyenne est nulle ($\bar{X} = 0$) et la variance égale à 1 ($\text{Var}(X) = 1$).

Démarche

Matrice Centrée Réduite :

La matrice centrée réduite (M_{cr}) est la matrice $M_{cr} = (m_{ij})_{1 \leq i \leq n, 1 \leq j \leq p}$ avec m_{ij} est obtenue par la formule suivante :

$$m_{ij} = \frac{x_{ij} - \bar{X}_j}{\sigma_i}.$$

La Matrice des Variances Covariances

La matrice des variances covariances permet de mesurer la liaison linéaire qui peut exister entre un couple de variables statistiques :

$$V = \begin{pmatrix} \mathbf{Var}(X_1) & \text{Cov}(X_1, X_2) & \text{Cov}(X_1, X_3) & \dots & \text{Cov}(X_1, X_p) \\ \text{Cov}(X_2, X_1) & \mathbf{Var}(X_2) & \text{Cov}(X_2, X_3) & \dots & \text{Cov}(X_2, X_p) \\ \text{Cov}(X_3, X_1) & \text{Cov}(X_3, X_2) & \mathbf{Var}(X_3) & \dots & \text{Cov}(X_3, X_p) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \text{Cov}(X_p, X_1) & \text{Cov}(X_p, X_2) & \text{Cov}(X_p, X_3) & \dots & \mathbf{Var}(X_p) \end{pmatrix}.$$

Démarche

La Matrice des Variances Covariances

- Si $\text{Cov}(X_2, X_1) = 0$. Alors, les variables X_1 et X_2 sont indépendantes.
- Si $\text{Cov}(X_2, X_1) \neq 0$. Alors, les variables X_1 et X_2 sont dépendantes (existe une relation linéaire entre les variable).
- La matrice des variances covariances est obtenue par la formule suivante :

$$V = \frac{1}{n} M_c^T \times M_c$$

avec M_c : Matrice centrée $\iff M_c : a_{ij} = (x_{ij} - \bar{X}_j)$
et M_c^T : Matrice transposée de M_c .

Démarche

La matrice des corrélations entre variables

- La matrice des corrélations entre variables permet d'analyser les relations bilatérales entre les variables.
- Elle est obtenue par la formule suivante :

$$U = \frac{1}{n} M_{cr}^T \times M_{cr}$$

avec M_{cr} : Matrice centrée réduite
et M_{cr}^T : Matrice centrée réduite transposée.

Application (sans réponse)

Tableau des données :

	X_1	X_2	X_3
ind1	1	2	3
ind2	2	1	3
ind3	3	2	5
ind4	2	3	5

Exercice d'application (Résultats et interprétation)

Une étude consiste à déterminer les facteurs de la localisation internationale d'une marque. Soit le tableau des données suivant : (IDE=Investissements directs étrangers)

	IDE	Taux croissance économique (%)	Taux d'inflation (%)
Pays A	300	2	6
Pays B	450	2	4
Pays C	950	8	2
Pays D	700	7	5

Exercice d'application (Résultats et interprétation)

Travail à faire :

- (1) Calculer la moyenne et l'écart type des variables.
- (2) Déterminer la Matrice Centrée Réduite.
- (3) Déterminer la Matrice des variances covariance.
- (4) Déterminer la Matrice des corrélations entre variables.
- (5) Déterminer le polynôme caractéristique.
- (6) Calculer les valeurs propres.
- (7) Calculer et interpréter l'inertie expliquée des axes factoriels.
- (8) Déterminer les vecteurs propres orthogonaux associés aux valeurs propres.
- (9) Calculer et interpréter les composantes principales, la contribution des individus et la contribution des variables.
- (10) Calculer et interpréter la contribution (CONTR) des individus à la construction des axes.

Exercice d'application (Résultats et interprétation)

Réponses :

- (1) Calculer la moyenne et l'écart type des variables.

Définition 1.

La **moyenne** est un outil de calcul permet de résumer une liste de valeurs numériques en un seul nombre réel sans tenir compte de l'ordre de la liste.

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

avec x_i sont les valeurs de la variable.

Définition 2.

L'**écart type** est un outil de calcul permet de mesurer la **dispersion** des valeurs d'un échantillon. C'est la racine carrée de la variance : $\sigma = \sqrt{\text{Var}}$, avec la **variance**

est la moyenne des carrées des écarts à la moyenne :

$$\text{Var} = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{X})^2$$

Exercice d'application (Résultats et interprétation) : Réponses

(1) Calculons la moyenne et l'écart type des variables :

	IDE	Taux croissance économique (%)	Taux d'inflation (%)
Pays A	300	2	6
Pays B	450	2	4
Pays C	950	8	2
Pays D	700	7	5
Moyenne	600	4,75	4,25
Écart type	247,49	2,77	1,48

Calcul : Pour la variable $X_1 = \text{IDE}$:

$$\bar{X}_1 = \frac{300 + 450 + 950 + 700}{4} = 600.$$

$$\sigma_1 = \sqrt{\frac{(300 - 600)^2 + (450 - 600)^2 + (950 - 600)^2 + (700 - 600)^2}{4}} = 247,49.$$

Exercice d'application (Résultats et interprétation) : Réponses

(2) Déterminons la Matrice Centrée Réduite (M_{cr}) : On sait que

$$m_{ij} = \frac{x_{ij} - \bar{X}_j}{\sigma_i}$$

. La matrice M_{cr} est donnée par :

$$M_{cr} = \begin{pmatrix} -1,21 & -0,99 & 1,18 \\ -0,61 & -0,99 & -0,17 \\ 1,41 & 1,17 & -1,52 \\ 0,40 & 0,81 & 0,51 \end{pmatrix}.$$

Calcul : Pour la variable $X_1 = \text{IDE}$:

$$m_{11} = \frac{x_{11} - \bar{X}_1}{\sigma_1} = \frac{300 - 600}{247,49} = -1,21$$

$$m_{21} = \frac{x_{21} - \bar{X}_1}{\sigma_1} = \frac{450 - 600}{247,49} = -0,61$$

$$m_{31} = \frac{x_{31} - \bar{X}_1}{\sigma_1} = \frac{950 - 600}{247,49} = 1,41$$

$$m_{41} = \frac{x_{41} - \bar{X}_1}{\sigma_1} = \frac{700 - 600}{247,49} = 0,40$$

Exercice d'application (Résultats et interprétation) : Réponses

(3) Déterminons la matrice des variances covariance : $V = \frac{1}{n} M_c^T \times M_c$. La matrice centrée (M_c) est donnée par : $a_{ij} = x_{ij} - \bar{X}$. Donc,

$$M_c = \begin{pmatrix} -300 & -2,75 & 1,75 \\ -150 & -2,75 & -0,25 \\ 350 & 3,25 & -2,25 \\ 100 & 2,25 & 0,75 \end{pmatrix}.$$

Calcul : Pour la variable $X_1 = \text{IDE}$:

$$a_{11} = x_{11} - \bar{X}_1 = 300 - 600 = -300$$

$$a_{21} = x_{21} - \bar{X}_1 = 450 - 600 = -150$$

$$a_{31} = x_{31} - \bar{X}_1 = 950 - 600 = 350$$

$$a_{41} = x_{41} - \bar{X}_1 = 700 - 600 = 100$$

Exercice d'application (Résultats et interprétation) : Réponses

Donc,

$$V = \frac{1}{4} \begin{pmatrix} -300 & -150 & 350 & 100 \\ -2,75 & -2,75 & 3,25 & 2,25 \\ 1,75 & -0,25 & -2,25 & 0,75 \end{pmatrix} \times \begin{pmatrix} -300 & -2,75 & 1,75 \\ -150 & -2,75 & -0,25 \\ 350 & 3,25 & -2,25 \\ 100 & 2,25 & 0,75 \end{pmatrix}.$$

Ainsi,

$$V = \begin{pmatrix} 61250 & 650 & -300 \\ 650 & 7,6875 & -2,4375 \\ -300 & -2,4375 & 2,1875 \end{pmatrix}.$$

- (4) Déterminons la matrice des corrélations entre variables :

$$U = \frac{1}{n} M_{cr}^T \times M_{cr}$$

Exercice d'application (Résultats et interprétation) : Réponses

Alors,

$$U = \frac{1}{4} \begin{pmatrix} -1,21 & -0,61 & 1,41 & 0,40 \\ -0,99 & -0,99 & 1,17 & 0,81 \\ 1,18 & -0,16 & -1,52 & 0,51 \end{pmatrix} \times \begin{pmatrix} -1,21 & -0,99 & 1,18 \\ -0,61 & -0,99 & -0,16 \\ 1,41 & 1,17 & -1,52 \\ 0,40 & 0,81 & 0,51 \end{pmatrix}.$$

D'où,

$$U = \begin{pmatrix} 1 & 0,95 & -0,82 \\ 0,95 & 1 & -0,6 \\ -0,82 & -0,6 & 1 \end{pmatrix}.$$

- $\text{Cov}(X_1, X_2) = 0,95$: Forte corrélation positive entre IDE et Taux de Croissance. **Taux de Croissance augmente \implies IDE augmente.**
- $\text{Cov}(X_1, X_3) = -0,82$: Forte corrélation négative entre IDE et Taux d'inflation. **Taux d'inflation augmente \implies IDE diminue.**

Exercice d'application (Résultats et interprétation) : Réponses

(5) Déterminons le polynôme caractéristique. On a :

$$P(\lambda) = \det(U - \lambda I_3) = \begin{vmatrix} 1 - \lambda & 0,95 & -0,82 \\ 0,95 & 1 - \lambda & -0,6 \\ -0,82 & -0,6 & 1 - \lambda \end{vmatrix}.$$

Donc,

$$\begin{aligned} P(\lambda) &= (1 - \lambda) \begin{vmatrix} 1 - \lambda & -0,6 \\ -0,6 & 1 - \lambda \end{vmatrix} - 0,95 \begin{vmatrix} 0,95 & -0,82 \\ -0,6 & 1 - \lambda \end{vmatrix} - 0,82 \begin{vmatrix} 0,95 & -0,82 \\ 1 - \lambda & -0,6 \end{vmatrix} \\ &= (1 - \lambda) [(1 - \lambda)^2 - 0,6^2] - 0,95 [0,95(1 - \lambda) - 0,6 \times 0,82] \\ &\quad - 0,82 [-0,95 \times 0,6 + 0,82(1 - \lambda)]. \\ &= -\lambda (\lambda^2 - 3\lambda + 1,1). \end{aligned}$$

(6) Calculons les valeurs propres : On résout l'équation :

$$P(\lambda) = \det(U - \lambda I_3) = 0 \iff -\lambda (\lambda^2 - 3\lambda + 1,1) = 0.$$

Exercice d'application (Résultats et interprétation) : Réponses

Donc, $\lambda = 0$ ou $\lambda = 0.43$ ou $\lambda = 2.57$.

- Condition prendre les valeurs propres dans l'ordre décroissant.

On prend, $\lambda_1 = 2.57$, $\lambda_2 = 0.43$ et $\lambda_3 = 0$.

Axes principaux

On appelle axes principaux d'inertie les axes de direction les vecteurs propres de U normés à 1 .

Il y en a p.

Le premier axe est celui associé à la plus grande valeur propre. On le note u^1

Le deuxième axe est celui associé à la deuxième valeur propre. On le note u^2

...

(7) Calculons et interprétons l'inertie expliquée des axes principaux :

- L'inertie de l'axe principal : Le pourcentage (%) d'inertie exprimé par un axe principal permet d'évaluer la quantité d'information contenue dans cet axe.

$$\text{Inertie d'un axe} = \frac{\text{Valeur propre correspondante}}{\text{Somme des valeurs propres (Inertie totale)}}.$$

- Axe 1 : $\lambda_1 = 2.57$. Inertie axe 1 = $\frac{\lambda_1}{\lambda_1 + \lambda_2} = \frac{2.57}{2.57 + 0.43} = 0.86$.

Cet axe contient 86% des informations.

- Axe 2 : $\lambda_2 = 0.43$. Inertie axe 2 = $\frac{\lambda_2}{\lambda_1 + \lambda_2} = \frac{0.43}{2.57 + 0.43} = 0.14$.

Cet axe contient 14% des informations.

Exercice d'application (Résultats et interprétation) : Réponses

(8) Déterminons les vecteurs propres orthogonaux associés aux valeurs propres :

- Axe 1 : $\lambda_1 = 2.57$. On cherche $u_1 \begin{pmatrix} x \\ y \\ z \end{pmatrix}$ tel que $Uu_1 = \lambda_1 u_1$.

$$\begin{pmatrix} 1 & 0,95 & -0,82 \\ 0,95 & 1 & -0,6 \\ -0,82 & -0,6 & 1 \end{pmatrix} \times \begin{pmatrix} x \\ y \\ z \end{pmatrix} = 2.57 \begin{pmatrix} x \\ y \\ z \end{pmatrix}.$$

$$\begin{cases} x + 0,95y - 0,82z = 2.57x \\ 0,95x + y - 0,6z = 2.57y \\ -0,82x - 0,6y + z = 2.57z \end{cases} \iff \begin{cases} -1.57x + 0,95y - 0,82z = 0 \\ 0,95x - 1.57y - 0,6z = 0 \\ -0,8x - 0,6y - 1.57z = 0 \end{cases}.$$

Donc,

$$\begin{cases} x = -0,62 \\ y = -0,57 \\ z = -0,54 \end{cases} \iff u_1 \begin{pmatrix} -0,62 \\ -0,57 \\ -0,54 \end{pmatrix}.$$

Exercice d'application (Résultats et interprétation) : Réponses

- Axe 2 : $\lambda_2 = 0,43$. On cherche $u_2 \begin{pmatrix} x \\ y \\ z \end{pmatrix}$ tel que $Uu_2 = \lambda_2 u_2$.

$$\begin{pmatrix} 1 & 0,95 & -0,82 \\ 0,95 & 1 & -0,6 \\ -0,82 & -0,6 & 1 \end{pmatrix} \times \begin{pmatrix} x \\ y \\ z \end{pmatrix} = 0,43 \begin{pmatrix} x \\ y \\ z \end{pmatrix}.$$

$$\begin{cases} x + 0,95y - 0,82z = 0,43x \\ 0,95x + y - 0,6z = 0,43y \\ -0,82x - 0,6y + z = 0,43z \end{cases} \iff \begin{cases} 0,57x + 0,95y - 0,82z = 0 \\ 0,95x + 0,57y - 0,6z = 0 \\ -0,82x - 0,6y + 0,57z = 0 \end{cases}.$$

Donc,

$$\begin{cases} x = 0,12 \\ y = 0,61 \\ z = 0,79 \end{cases} \iff u_2 \begin{pmatrix} 0,12 \\ 0,61 \\ 0,79 \end{pmatrix}.$$

Exercice d'application (Résultats et interprétation) : Réponses

Composantes principales

À chaque axe est associée une variable appelée composante principale.

- La composante \mathbf{C}_1 est le vecteur renfermant les coordonnées des projections des individus sur l'axe 1.
- La composante \mathbf{C}_2 est le vecteur renfermant les coordonnées des projections des individus sur l'axe 2.
-

Exercice d'application (Résultats et interprétation) : Réponses

(9) Calculer et interpréter les composantes principales, la contribution des individus et la contribution des variables. :

- Les composantes principales sont obtenues par la projection des individus sur les axes principaux.

$$M_{cr} \times u_1 = \text{Projection sur axe 1.}$$

$$M_{cr} \times u_2 = \text{Projection sur axe 2.}$$

- Projection sur axe 1 \Rightarrow
$$\begin{pmatrix} -1,21 & -0,99 & 1,18 \\ -0,61 & -0,99 & -0,17 \\ 1,41 & 1,17 & -1,52 \\ 0,40 & 0,81 & 0,51 \end{pmatrix} \times \begin{pmatrix} -0,62 \\ -0,57 \\ -0,54 \end{pmatrix}.$$

- Projection sur axe 2 \Rightarrow
$$\begin{pmatrix} -1,21 & -0,99 & 1,18 \\ -0,61 & -0,99 & -0,17 \\ 1,41 & 1,17 & -1,52 \\ 0,40 & 0,81 & 0,51 \end{pmatrix} \times \begin{pmatrix} 0,12 \\ 0,61 \\ 0,79 \end{pmatrix}.$$

Exercice d'application (Résultats et interprétation) : Réponses

Résultat de projection (les composantes principales) :

	C_1	C_2
Pays A	0.68	0.18
Pays B	1.03	-0.81
Pays C	-0.72	-0.32
Pays D	-0.99	0.94

- Ce tableau permet de calculer la contribution des axes.
- Les composantes principales sont centrées c-à-d. la moyenne de chaque composante est nulle.
- La variance de chaque composante est égale à la valeur propre associée à l'axe.
- La contribution des individus : Pour calculer la contribution de chaque individu à l'inertie, on utilise la formule suivante :
$$\text{CONTR} = \frac{1}{n} \frac{(m_{ij})^2}{I}$$
, avec : m_{ij} = valeur de M_{cr} et I = La somme des valeurs propres $\left(\sum \lambda_i\right)$.

Exercice d'application (Résultats et interprétation) : Réponses

	X_1	X_2	X_3	CONTR
Pays A	0,12	0,08	0,12	0,32
Pays B	0,03	0,08	0,003	0,113
Pays C	0,17	0,11	0,19	0,47
Pays D	0,1	0,05	0,02	0,17
	0,42	0,32	0,333	

Calcul : $\sum \lambda_i = 3$ | $(X_1; \text{Pays A}) : \text{CONTR} = \frac{1}{4} \times \frac{(1,21)^2}{3} = 0,12.$

$$(X_1; \text{Pays B}) : \text{CONTR} = \frac{1}{4} \times \frac{(0,61)^2}{3} = 0,03.$$

$$(X_1; \text{Pays C}) : \text{CONTR} = \frac{1}{4} \times \frac{(1,41)^2}{3} = 0,17.$$

$$(X_1; \text{Pays D}) : \text{CONTR} = \frac{1}{4} \times \frac{(0,4^2)}{3} = 0,10.$$

Exercice d'application (Résultats et interprétation) : Réponses

Le tableau permet de déterminer la contribution des individus dans l'analyse :

- Le pays C contribue de 0,47 (47%) pour expliquer le phénomène.
- Le pays A contribue de 0,32 (32%) pour expliquer le phénomène.

Le tableau permet de déterminer la contribution des variables dans l'analyse :

- La variable X_1 contribue de 42% dans l'analyse.
- La variable X_2 contribue de 32% dans l'analyse.
- La variable X_3 contribue de 33,3% dans l'analyse.

(10) La contribution CONTR des individus à la construction des axes est calculée par la formule suivante :

$$\text{Axe 1 : } \lambda_1 = 2.57 \implies \text{CONTR} = \frac{1}{n} \frac{(C_1^i)^2}{\lambda_1}.$$

$$\text{Axe 2 : } \lambda_2 = 0.43 \implies \text{CONTR} = \frac{1}{n} \frac{(C_2^i)^2}{\lambda_2}.$$

Exercice d'application (Résultats et interprétation) : Réponses

La contribution CONTR des individus à la construction des axes :

	C_1	C_2	CONTR Axe 1 (C_1)	CONTR Axe 2 (C_2)
Pays A	0.68	0.18	0,04	0,02
Pays B	1.03	-0.81	0,10	0,38
Pays C	-0.72	-0.32	0,05	0,06
Pays D	-0.99	0.94	0,09	0,51

Calcul :

$$\text{CONTR}(\text{Axe1}; \text{Pays A}) = \frac{1}{4} \times \frac{(0,68)^2}{2.57} = 0,04.$$

$$\text{CONTR}(\text{Axe1}; \text{Pays B}) = \frac{1}{4} \times \frac{(1,03)^2}{2.57} = 0,10.$$

$$\text{CONTR}(\text{Axe2}; \text{Pays A}) = \frac{1}{4} \times \frac{(0,18)^2}{0,43} = 0,02.$$

$$\text{CONTR}(\text{Axe2}; \text{Pays B}) = \frac{1}{4} \times \frac{(0,81)^2}{0,43} = 0,38.$$

Exercice d'application (Résultats et interprétation) : Réponses

Interprétation :

- Pour l'Axe 1, le point déterminant est 0,10 (10%) c-à-d. le pays B contribue de 10% à la construction de l'axe 1.
- Pour l'Axe 2, le point déterminant est 0,51 (51%) c-à-d. le pays D contribue de 51% à la construction de l'axe 2.