Module 2 : L'analyse en composantes principales - Exercices préparatifs

L'analyse en composantes principales est notée ACP. Elle s'applique à tous les tableaux de données où les variables sont de type quantitatif. C'est la méthode de référence pour deux raisons :

- c'est la plus facile à exposer sur le plan mathématique,
- c'est une méthode qui peut servir de support à d'autres techniques statistiques comme par exemple la régression orthogonale, la construction d'indicateurs synthétiques, la prévision d'une chronique ou encore compléter une information manquante dans un tableau.

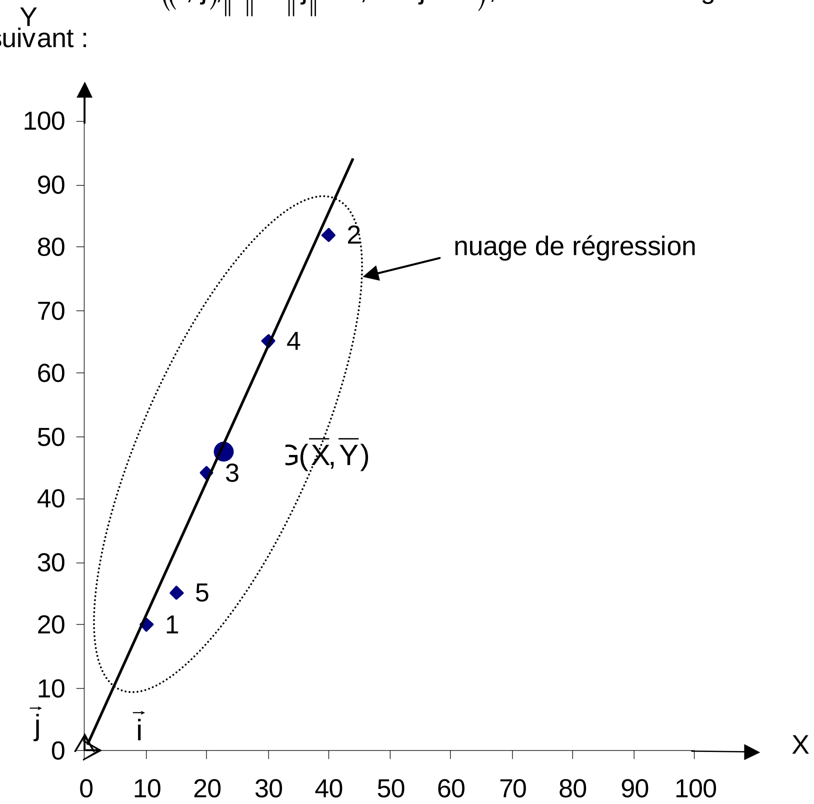
Avant de présenter formellement la méthode de l'ACP (Module suivant), on va essayer dans ce module, d'intuiter la démarche à travers deux exemples.

1er exemple

On considère le tableau suivant :

Individus\variables	Y	X
1	20	10
2	82	40
3	44	20
4	65	30
5	25	15
Somme	236	115

La représentation graphique des individus dans l'espace R² des deux variables, en utilisant une base orthonormée $\left((\vec{i}, \vec{j}), \|\vec{i}\| = \|\vec{j}\| = 1, \vec{i} * \vec{j} = 0\right)$, conduit au nuage des individus (nuage de régression) suivant :



Avec la régression, il est parfois possible de visualiser l'information contenue dans le nuage de régression (les proximités relatives des 5 points)

Le tableau des calculs permettant de trouver les éléments d'une régression sur données centrées (x,y) et non centrée (X,Y) est le suivant :.

Ind	Y	X	X^2	XY	Y^2	$y = Y - \overline{Y}$	$x = X - \overline{X}$	y^2	x^2	xy	$\hat{Y}$
1	20	10	100	200	400	-27,20	-13,00	739,84	169,00	353,60	18,99
2	82	40	1600	3280	6724	34,80	17,00	1211,04	289,00	591,60	84,09
3	44	20	400	880	1936	-3,20	-3,00	10,24	9,00	9,60	40,69
4	65	30	900	1950	4225	17,80	7,00	316,84	49,00	124,60	62,39
5	25	15	225	375	625	-22,20	-8,00	492,84	64,00	177,60	29,84
Somme	236	115	3225	6685	13910	0,00	0,00	2770,80	580,00	1257,00	236,00

Il permet de calculer les caractéristiques qui conduisent aux paramètres de la régression.

$$\overline{Y} = \frac{1}{n}\sum Y_i = \frac{236}{5} = 47.2$$

$$\overline{X} = \frac{1}{n}\sum X_i = \frac{115}{5} = 23$$

$$V[Y] = \frac{1}{n}\sum Y_i^2 - \overline{Y}^2 = \frac{13910}{5} - (47.2)^2 = 554.16$$

$$\sigma[Y] = \sqrt{V[Y]} = \sqrt{554.16} = 23.54$$

$$V[X] = \frac{1}{n}\sum X_i^2 - \overline{X}^2 = \frac{3225}{5} - (23)^2 = 116$$

$$\sigma[X] = \sqrt{V[X]} = \sqrt{116} = 10.77$$

$$r_{YX} = \frac{\operatorname{cov}[X,Y]}{\sigma[X]\sigma[Y]} = 0.9916 \quad r_{XY} = \frac{1}{n}\frac{\sum XY - n\overline{XY}}{\sigma_X\sigma_Y} = \frac{1}{5}\frac{(6685 - 5 * 47.2 * 23)}{10.77 * 23.54} = 0.9916$$

$$\boxed{r^2 = 0.98}$$

r^2, le coefficient de détermination, nous indique que 98% du nuage de régression est expliqué par la droite de régression $Y = aX + b$.

Il est donc possible d'utiliser cette droite pour résumer le nuage de régression.

La méthode de calcul des paramètres a et b de la droite de régression consiste à minimiser la somme des carrés des résidus entre les valeurs observées Y_i et le valeurs calculées $\hat{Y}_i$

La minimisation de la somme des écarts au carré porte le nom de méthode des MCO. Cela s'écrit :

$$\operatorname{Min}\sum\left(Y_i - \hat{Y}_i\right)^2$$

On démontre que :

$$\hat{a} = \frac{\operatorname{cov}(X,Y)}{V(X)} = r\frac{\sigma[Y]}{\sigma[X]} = 0.9916 * \frac{23.54}{10.77} = 2.17$$

$\hat{b} = \overline{Y} - \hat{a}\overline{X} \Rightarrow 47.2 - 2.17 * 23 = -2.71$ (la droite passe par le point $G(\overline{X},\overline{Y})$ qui est le centre de gravité du nuage des points des individus).

$$\boxed{\hat{Y} = 2.17X - 2.71}$$

Le nuage de régression permet de connaître l'information concernant les individus du tableau. Par exemple, on visualise le point 1 proche du point 5 et le point 1 loin du point 2. Il est possible aussi de quantifier cette information en calculant toutes les distances au carré (théorème de Pythagore) entre les paires de points et de les classer par ordre croissant.

Le graphe de régression montre que le nuage de point est inséré dans une ellipse aux bords aplatis , ce qui signifie que ce nuage peut être résumé au moyen d'une droite de régression. Cette observation est confirmée par le calcul du coefficient de corrélation $r = 0.99$, ce qui signifie qu'il existe une relation étroite et positive entre X et Y. Il est donc possible de substituer au nuage de régression, la droite $\hat{Y} = 2.17X - 2.71$ ou encore la droite sur variables centrées $\hat{y} = 2.17x$ qui a pour origine le point $G(\overline{X}, \overline{Y})$. (Cf le tableau précédent pour le détail des calculs)

On peut donc calculer les projections au sens des moindre carrés (parallèlement à l'axe des ordonnées) des 5 points sur la droite de régression.

Ces projections sont données pour les variables non centrées par les calculs $\hat{Y}_1, \cdots, \hat{Y}_5$. On constate alors que si on calcule la distance, par exemple, entre $\hat{Y}_1$ et $\hat{Y}_5$ au carré, on trouve environ celle du nuage de régression entre le point 1 et le point 5.

Par conséquent, l'information concernant les 5 points sur l'axe $\hat{Y}$ est conservée par rapport à celle du nuage de régression. On peut donc dire que l'analyse de données a eu lieu puisque l'**information est pratiquement identique sur l'axe que dans le plan.**

On peut aussi résumer l'information contenue dans le nuage de points en utilisant non pas les projections sur la droite de régression des points au sens des MCO, mais leurs projections orthogonales sur cette même droite, en conservant pour origine de l'axe, le point G et en construisant un vecteur unitaire dont on connaît les coordonnées dans l'espace R² ; les projections orthogonales des 5 points sur cette droite dont données par le **produit scalaire** entre le vecteur unitaire et un vecteur qui a pour origine le point G et pour extrémité le point à projeter. On pourrait constater que, dans ce cas aussi, la distance au carré par exemple entre le point 1 et le point 5 projetés est approximativement identique à celle du plan entre les mêmes points. L'analyse de données est donc encore réalisable en procédant de la sorte.

Remarque importante :

lorsque l'on travaille sur les variables centrées, on a les coordonnées suivantes des vecteurs $\vec{x}$ et $\vec{y}$:

$$X - \overline{X} = \vec{x} = \begin{bmatrix} -13 \\ 17 \\ -3 \\ 7 \\ -8 \end{bmatrix} \quad Y - \overline{Y} = \vec{y} = \begin{bmatrix} -27.2 \\ 34.8 \\ -3.2 \\ 17.8 \\ -22.2 \end{bmatrix}$$

Le produit scalaire entre les vecteurs $\vec{x}$ et $\vec{y}$ s'écrit :

$$\vec{x} * \vec{y} = (-13) * (-27.2) + \cdots + (-8) * (-22.2) = 1257 = \sum_i x_i y_i$$

De ce fait : $\dfrac{\vec{x} * \vec{y}}{n} = \dfrac{\sum xy}{n} = \mathrm{cov}(x, y)$

D'où : $V(x) = \mathrm{cov}(x, x) = \dfrac{(\vec{x})^2}{n}$

et $\sigma(x) = \dfrac{\|\vec{x}\|}{\sqrt{n}}$

De plus : $r = \dfrac{\mathrm{cov}(x,y)}{\sigma(x)\,\sigma(y)} = \dfrac{\dfrac{\vec{x} * \vec{y}}{n}}{\dfrac{\|\vec{x}\|}{\sqrt{n}} * \dfrac{\|\vec{y}\|}{\sqrt{n}}} = \dfrac{\vec{x} * \vec{y}}{\|\vec{x}\|\,\|\vec{y}\|}$

Par ailleurs on sait que :

$\vec{x} * \vec{y} = \|\vec{x}\| * \|\vec{y}\| * \cos\alpha$ avec α l'angle formé par les deux vecteurs.

D'où : $\boxed{r = \cos\alpha}$.

Ainsi, lorsque les variables sont centrées, le coefficient de corrélation entre les 2 variables est égal au cosinus de l'angle formé par les vecteurs représentant ces variables.

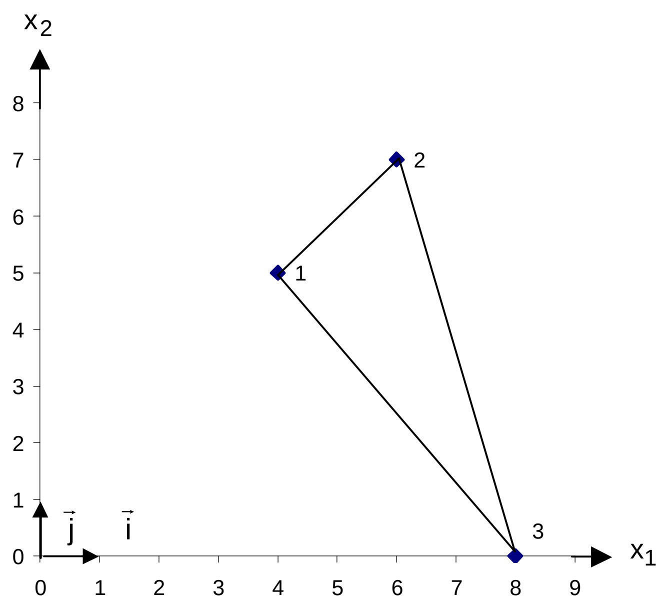
Quand on centre et on réduit des variables (par exemple $y_i = \dfrac{Y_i - \overline{Y}}{\sigma_Y}$), on forme des vecteurs qui ont tous la même dimension. ($V(y) = 1$). De ce fait, <u>la variance est la distance commune à tous les vecteurs</u> (ils se situent sur un cercle de rayon 1) et ils se positionnent les uns par rapport aux autres par le coefficient de corrélation linéaire que l'on déduit à partir de l'angle formé par les deux vecteurs.

Exemple 2

Soit le tableau de données suivant :

	Ind\var	x_1	x_2
	1	4	5
$X_{(3,2)} =$	2	6	7
	3	8	0

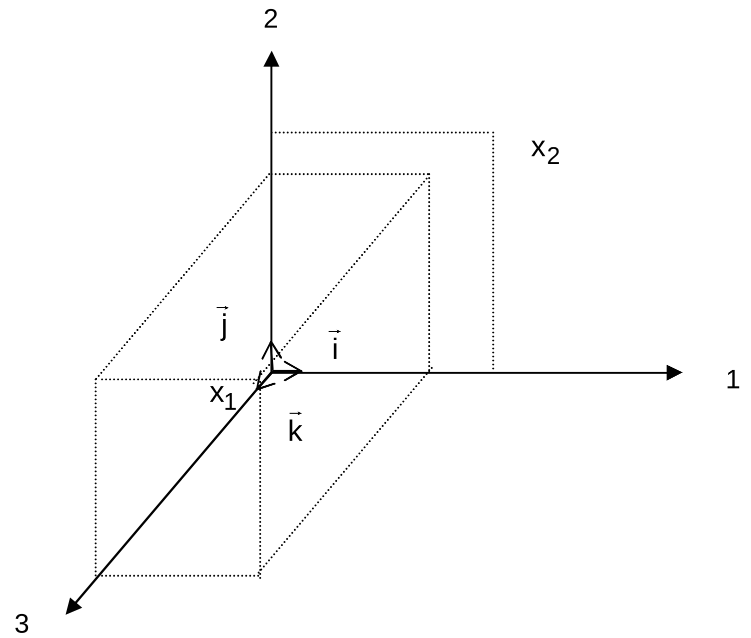
• <u>Représentation graphique</u> du nuage des 3 points individus dans l'espace R² des variables (x_1 en abscisse, x_2 en ordonnée). Le système d'axes est orthonormé : base $(\vec{i}, \vec{j})$ telle que $\|\vec{i}\| = \|\vec{j}\| = 1$, $\vec{i} * \vec{j} = 0$.



Les 3 points du nuage constituent l'information des lignes du tableau. Les positions relatives de ces 3 points peuvent être calculées en utilisant la distance euclidienne.

• <u>Représentation graphique</u> du nuage des 2 points variables dans l'espace R^3 des individus. Le système d'axes est orthonormé : base $\left(\vec{i}, \vec{j}, \vec{k}\right)$ telle que :

$$\|\vec{i}\| = \|\vec{j}\| = \|\vec{k}\| = 1, \ \vec{i} * \vec{j} = 0 \ \vec{i} * \vec{k} = 0 \ \vec{j} * \vec{k} = 0 .$$



Les points du nuage constituent l'information donnée par les colonnes du tableau. Ici aussi, on peut calculer la distance euclidienne entre les deux point.

• <u>Calcul des caractéristiques des colonnes du tableau</u>

Calcul de la moyenne et de l'écart type de x_1 et x_2 :

$$\bar{x}_1 = \frac{18}{3} = 6 \quad \bar{x}_2 = \frac{12}{3} = 4$$

$$V(x_1) = \frac{116}{3} - 6^2 = 2.67 \quad \sigma(x_1) = 1.633$$

$$V(x_2) = \frac{74}{3} - 4^2 = 8.67 \quad \sigma(x_2) = 2.944$$

Calcul de la moyenne et de l'écart type de 1, 2, 3 :

$$\bar{1} = \frac{9}{2} = 4.5 \quad \bar{2} = \frac{13}{2} = 6.5 \quad \bar{3} = \frac{8}{2} = 4$$

$$V(1) = \frac{41}{2} - (4.5)^2 = 0.25 \quad \sigma(1) = 0.5$$

$$V(2) = \frac{85}{2} - (6.5)^2 = 0.25 \quad \sigma(2) = 0.5$$

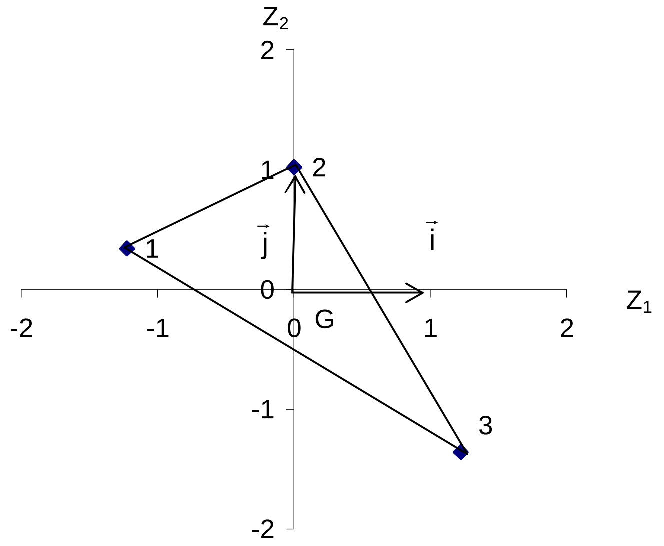
$$V(3) = \frac{64}{2} - (4)^2 = 16 \quad \sigma(3) = 4$$

• <u>Construction du tableau des variables centrées et réduites</u> :

		x_1	x_2	$x_1 - \bar{x}_1$	$x_2 - \bar{x}_2$	$z_1 = \dfrac{x_1 - \bar{x}_1}{\sigma(x_1)}$	$z_2 = \dfrac{x_2 - \bar{x}_2}{\sigma(x_2)}$
	1	4	5	-2	1	-1.225	0.34
$Z_{(3,2)} =$	2	6	7	0	3	0	1.02
	3	8	0	2	-4	1.225	-1.36
	Σ	18	12	0	0	0	0

on vérifie que : $\bar{z}_1 = \bar{z}_2 = 0$, $V(z_1) = V(z_2) = 1$ et $Cov(z_1, z_2) = r_{z_1, z_2}$

• <u>Représentation graphique</u> du nuage des 3 points individus dans l'espace R^2 des variables centrées réduites (z_1 en abscisse et z_2 en ordonnée). Le système d'axes est orthonormé : base $(\vec{i}, \vec{j})$ telle que $\|\vec{i}\| = \|\vec{j}\| = 1, \vec{i} * \vec{j} = 0$. Dans cet espace, l'origine des axes (point 0) est confondu avec le centre de gravité du triangle (Point $G(\bar{z}_1 = 0, \bar{z}_2 = 0)$)



Dans l'espace R^3 des individus, se situent les deux variables centrées réduites. Avec un système d'axes orthonormé on peut calculer :

En utilisant les variables centrées réduites dans l'espace à trois dimensions des individus avec un système orthonormé on peut calculer :

$$d^2(0, z_1) = \left(-\sqrt{\frac{3}{2}}\right)^2 + 0 + \left(\sqrt{\frac{3}{2}}\right)^2 = 3$$

D'où

$\frac{1}{3} d^2(0, z_1) = 1$ la variance de z_1

$$d^2(0, z_2) = \left(\sqrt{\frac{3}{26}}\right)^2 + \left(\frac{3\sqrt{2}}{\sqrt{26}}\right)^2 + \left(-\frac{4\sqrt{3}}{\sqrt{26}}\right)^2 = 3$$

$\frac{1}{3} d^2(0, z_2) = 1$ la variance de z_2

Dans cet espace, la distance au carré entre l'origine et une variable est, à N = 3 près, la variance de la variable. Quand les variables sont centrées et réduites, toutes les variables sont <u>équidistantes</u> de l'origine. Cette distance est, au nombre d'observations près, la variance des variables.

Récapitulatif : présentation des calculs :

$X_{(3,2)} =$

	x_1	x_2
1	4	5
2	6	7
3	8	0
$\bar{x}$	6	4
σx_j	$2\sqrt{\frac{2}{3}}$	$\sqrt{\frac{26}{3}}$

$$Z_{ij} = \frac{x_{ij} - \bar{x}_j}{\sigma x_i}$$

Tableau des variables centrées réduites :

$Z_{(3,2)} =$

	z_1	z_2		z_1	z_1
1	$-\sqrt{\frac{3}{2}}$	$\sqrt{\frac{3}{26}}$	=	-1,225	0,34
2	0	$\frac{3\sqrt{3}}{\sqrt{26}}$	=	0	1,02
3	$\sqrt{\frac{3}{2}}$	$-\frac{4\sqrt{3}}{\sqrt{26}}$	=	1,225	-1,36

= avec $z_{ij} = \frac{x_{ij} - \bar{x}_j}{\sigma_{x_i}}$.

$\bar{Z}$	0	0
σz_j	1	1

La moyenne des variables centrées et réduites est égale à 0.

L'écart type des variables centrées et réduites est égal à 1.

De plus :

$$r = \frac{\text{cov}}{\sigma(x)\,\sigma(y)}$$

comme $\sigma(x) = 1$ et $\sigma(y) = 1$, le coefficient de corrélation linéaire r entre 2 variables est égal à la covariance.

Remarque : on peut aussi traiter l'information contenue dans le tableau de départ en utilisant le tableau des individus centrés réduits.

$X_{(3,2)} =$

	x_1	x_2	$\bar{x}_i$	$\sigma(x_i)$
1	4	5	4.5	0.5
2	6	7	6.5	0.5
3	8	0	4	4

$$Q_{(3,2)} = \begin{bmatrix} -1 & 1 \\ -1 & 1 \\ 1 & -1 \end{bmatrix} \qquad \text{avec } q_{ij} = \frac{x_{ij} - \overline{x}_i}{\sigma_{(x_i)}}$$

Il est possible de représenter l'information contenue dans ce nouveau tableau comme précédemment et d'en tirer des conclusions.

- **Calcul du produit matriciel** $\frac{1}{N} Z'Z$

$$\frac{1}{3}\begin{bmatrix} -\sqrt{\frac{3}{2}} & 0 & \sqrt{\frac{3}{2}} \\ \sqrt{\frac{3}{26}} & \frac{3\sqrt{3}}{\sqrt{26}} & \frac{-4\sqrt{3}}{\sqrt{26}} \end{bmatrix} \cdot \begin{bmatrix} -\sqrt{\frac{3}{2}} & \sqrt{\frac{3}{26}} \\ 0 & \frac{3\sqrt{3}}{\sqrt{26}} \\ \sqrt{\frac{3}{2}} & \frac{-4\sqrt{3}}{\sqrt{26}} \end{bmatrix} = \frac{1}{3}\begin{bmatrix} 3 & \frac{-15}{2\sqrt{13}} \\ \frac{-15}{2\sqrt{3}} & 3 \end{bmatrix} = \begin{bmatrix} 1 & \frac{-5}{2\sqrt{13}} \\ \frac{-5}{2\sqrt{3}} & 1 \end{bmatrix} = \begin{bmatrix} 1 & -0.69 \\ -0.69 & 1 \end{bmatrix}$$

Le résultat de ce calcul est une matrice carrée, de dimension (2,2), notée R, contenant les coefficients de corrélation linéaires des variables.

Cette matrice carrée R a pour dimension le nombre de variables. Elles possède les propriétés suivantes :

- Elle est symétrique.
- elle a des 1 sur la diagonale principale (les variances des variables)
- Elle a des valeurs inférieures ou égales à 1 en valeur absolue.

Dans cette matrice R, on a sur la diagonale les variances des variables, or dans l'exercice précédent on a vu que cette variance était, au nombre d'observations près, la distance de la variable à l'origine. Elle contient de part et d'autre de la diagonale le coefficient de corrélation linéaire entre les deux variables. Or dans l'exercice précédent, on a vu que ce coefficient de corrélation était le cosinus de l'angle formé par les deux variables. L'angle formé par les deux variables peut donc en être déduit.

Avec la matrice R, il est donc possible de représenter dans l'espace les positions relatives des variables entre elles. Cette matrice R nous donne donc l'information recherchée concernant les variables. C'est la raison pour laquelle elle porte le nom de matrice d'information des variables.

- **Calcul du produit matriciel ZZ' :**

$$ZZ' = \begin{bmatrix} -\sqrt{\frac{3}{2}} & \sqrt{\frac{3}{26}} \\ 0 & \frac{3\sqrt{3}}{\sqrt{26}} \\ \sqrt{\frac{3}{2}} & \frac{-4\sqrt{3}}{\sqrt{26}} \end{bmatrix} \begin{bmatrix} -\sqrt{\frac{3}{2}} & 0 & \sqrt{\frac{3}{2}} \\ \sqrt{\frac{3}{26}} & \frac{3\sqrt{3}}{\sqrt{26}} & \frac{-4\sqrt{3}}{\sqrt{26}} \end{bmatrix} \cdot = \begin{bmatrix} \frac{21}{13} & \frac{9}{26} & \frac{-51}{26} \\ \frac{9}{26} & \frac{27}{26} & \frac{-36}{26} \\ \frac{-51}{26} & \frac{-36}{26} & \frac{87}{26} \end{bmatrix} = \underset{(3,3)}{V}$$

Cette matrice V n'est pas une matrice de corrélation, mais elle y ressemble. On lui donne le nom de matrice d'information des individus. Elle est symétrique ; sa diagonale est la somme des carrés des individus lignes du tableau et de part et d'autre on trouve la somme des produits lignes deux à deux des individus

- **Caractéristiques de la matrice R**

Les caractéristiques d'une matrice sont données par les vecteurs propres associés aux valeurs propres de la matrice.

On appelle vecteur propre associé à la valeur propre λ de la matrice R la solution du système d'équation homogène $RX = \lambda X \Leftrightarrow [R-\lambda I]X = 0$.

On sait que si dans ce système d'équation le déterminant de la matrice $R - \lambda I \neq 0$, alors ce système possède une et une seule solution qui est X = 0 et que l'on appelle la solution triviale. C'est la raison pour laquelle pour que ce système ait des solutions autres que celle-ci, il faut que le déterminant $|R - \lambda I| = 0$. Or ce déterminant conduit à une équation (équation caractéristique) qui a pour variable λ et pour degré la dimension de la matrice R.

Les racines de cette équation donnent les différentes valeurs de λ et portent le nom de valeurs propres. Pour chacune des valeurs propres, on pourra calculer à partir du système de départ, une infinité de vecteurs X qu'on appelle les vecteurs propres. Parmi cette infinité de vecteurs propres, on recherche par la suite le vecteur propre de norme 1 (c'est-à-dire le vecteur unitaire).

Dans ce cas on a :

$$\underset{(2,2)}{R}\ \underset{(2,1)}{X} = \lambda \underset{(2,1)}{X} \quad \text{avec } \lambda \in R$$

$$[R - \lambda I]X = 0$$

$$R = \begin{bmatrix} 1 & -0.69 \\ -0.69 & 1 \end{bmatrix} - \lambda \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$R = \begin{bmatrix} 1-\lambda & -0.69 \\ -0.69 & 1-\lambda \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\begin{cases} (1-\lambda)x_1 - 0.69x_2 = 0 \\ -0.69x_1 + (1-\lambda)x_2 = 0 \end{cases}$$

Calcul du déterminant :

$$\begin{aligned} |R - \lambda I| = \begin{bmatrix} 1-\lambda & -0.69 \\ -0.69 & 1-\lambda \end{bmatrix} &= (1-\lambda)^2 - (-0.69)^2 \\ &= (1-\lambda-0.69)(1-\lambda+0.69) \\ &= (0.31-\lambda)(1.69-\lambda) \end{aligned}$$

$$\begin{matrix} \lambda_1 = 1.69 \\ \lambda_2 = 0.31 \end{matrix} \text{ deux valeurs propres de R.}$$

Si on additionne 1.69 + 0.31 = 2, on obtient la dimension de la matrice (le nombre de variables du tableau).

Calcul des vecteurs propres associés

- pour $\lambda_1 = 1.69$

$$[T - \lambda I]X = 0$$

$$\begin{bmatrix} 1 & -0.69 \\ -0.69 & 1 \end{bmatrix} - \begin{bmatrix} 1.69 & 0 \\ 0 & 1.69 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 0$$

$$\begin{bmatrix} -0.69 & -0.69 \\ -0.69 & -0.69 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 0$$

$$\begin{cases} -0.69\ x_1 - 0.69\ x_2 = 0 \\ -0.69\ x_1 - 0.69\ x_2 = 0 \end{cases}$$

$x_1 + x_2 = 0$

$\vec{V}_1 = \begin{pmatrix} k \\ -k \end{pmatrix} \quad k \in R$

On a une infinité de vecteurs propres portés par la seconde bissectrice du plan $(\vec{x}_1, \vec{x}_2)$.

Pour trouver un vecteur propre normé il faut que :

$$k^2 + k^2 = 1 \Leftrightarrow 2k^2 = 1 \Leftrightarrow k^2 = \frac{1}{2} \Leftrightarrow k = \pm\frac{\sqrt{2}}{2}$$

En retenant pour k la valeur positive, on définit :

$\vec{b}_1 = \begin{bmatrix} \frac{\sqrt{2}}{2} \\ -\frac{\sqrt{2}}{2} \end{bmatrix}$ = vecteur propre normé de R.

- Pour $\lambda_2 = 0.31$

$$\begin{bmatrix} 0.69 & -0.69 \\ -0.69 & 0.69 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 0$$

$$\begin{cases} 0.69\, x_1 - 0.69\, x_2 = 0 \\ -0.69\, x_1 + 0.69\, x_2 = 0 \end{cases}$$

$x_1 - x_2 = 0 \Rightarrow x_2 = x_1$

$\vec{V}_2 = \begin{pmatrix} k \\ k \end{pmatrix} \quad k \in R$

Pour trouver un vecteur propre normé il faut que :

$$k^2 + k^2 = 2k^2 = 1 \Leftrightarrow k = \pm\frac{\sqrt{2}}{2}$$

$\vec{b}_2 = \begin{bmatrix} \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} \end{bmatrix}$ = vecteur propre normé de R.

Ces vecteurs propres normés constituent une nouvelle base orthonormée dans laquelle la norme de chaque vecteur =1 et leur produit scalaire est nul :

$$\begin{matrix} \|\vec{b}_1\| = 1 \\ \|\vec{b}_2\| = 1 \end{matrix} \quad \text{et } \vec{b}_1 * \vec{b}_2 = \begin{pmatrix} \frac{\sqrt{2}}{2} \\ -\frac{\sqrt{2}}{2} \end{pmatrix} * \begin{pmatrix} \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} \end{pmatrix} = \left(\frac{\sqrt{2}}{2}\right)^2 - \left(\frac{\sqrt{2}}{2}\right)^2 = 0$$

On peut alors placer les coordonnées (dans l'ancienne base) de ces vecteurs dans une matrice $B_{(2,2)}$, dans l'ordres décroissant de leurs valeurs propres.

$B_{(2,2)} = \begin{bmatrix} \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \\ -\frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{bmatrix}$ coordonnés des vecteurs b_1 et b_2 dans l'ancien système d'axes.

Cette matrice est une matrice orthogonale et vérifie donc : $B^{-1} = B'$, soit $B'B = I$

- **Caractéristique de la matrice V**

Si on calcule comme précédemment les valeurs propres de la matrice V : $|V - \lambda I| = 0$ c'est-à-dire :

$$\begin{vmatrix} 1.62-\lambda & 0.35 & -1.96 \\ 0.55 & 1.04-\lambda & -1.38 \\ -1.96 & -1.38 & 3.35-\lambda \end{vmatrix} = 0$$

on trouve : $\begin{array}{l} \lambda_1 = 5.07 \\ \lambda_2 = 0.93 \\ \lambda_3 = 0 \end{array}$

Si on porte dans un tableau les valeurs propres de V et de R on a :

V	R
$\lambda_1 = 5.07$ $\lambda_2 = 0.93$ $\lambda_3 = 0$	$\lambda_1 = 1.69$ $\lambda_2 = 0.31$
$\sum \lambda_j = 6$	$\sum \lambda_j = 2 = n$

On voit que si on multiplie les valeurs propres de la matrice R par 3, on obtient les deux premières valeurs propres de la matrice V et que la dernière valeur propre de V est nulle.

On peut démonter que les valeurs propres de la matrice V sont égales aux valeurs propres de R multipliées par N et qu'il y a dans la matrice V, N-n valeurs propres nulles.

On peut aussi démontrer qu'il est possible de calculer les vecteurs propres de V connaissant ceux de R. Et donc, qu'en définitive, les caractéristiques de R permettent de calculer celles de V et réciproquement.