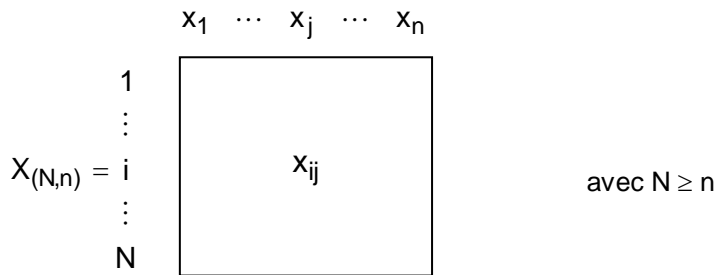


Module 3 : Formalisation mathématique de l'ACP

Après avoir introduit à l'aide de deux exemples l'ACP, nous présentons ici la formalisation mathématique de la méthode en généralisant ce que nous avons vu sur un espace réduit à un espace à dimensions quelconques.

A) Généralisation à des espaces de dimensions quelconques

Le tableau de départ qui sera soumis à une ACP se présente de la façon suivante :



Il s'agit d'un tableau de données quantitatives, avec les variables en colonnes, les individus en lignes et les observations à l'intérieur du tableau.

L'objectif de l'ACP est d'analyser l'information contenue dans le tableau, c'est-à-dire la structure du nuage des individus dans l'espace R^n et des variables dans l'espace R^N .

Pour des raisons mathématiques de simplification, mais aussi parce que les variables dans ces tableaux peuvent être de natures différentes, on transforme la matrice X en une matrice Z de variables centrées réduites qui conserve la même information :

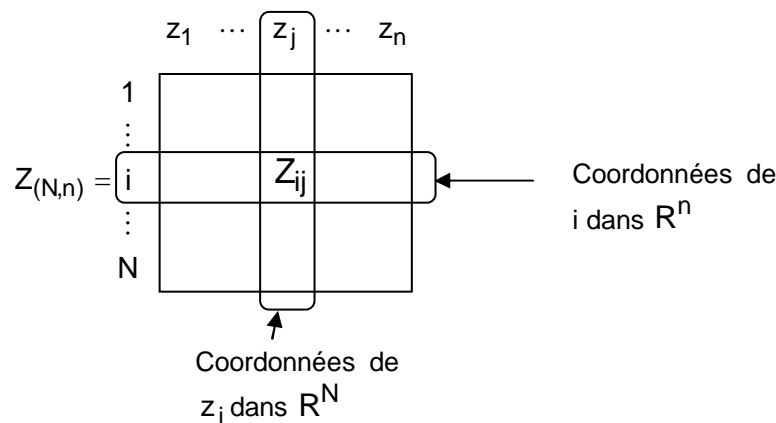
$$z_{ij} = \frac{x_{ij} - \bar{x}_j}{\sigma(x_j)} \text{ variables centrées réduites.}$$

- le centrage n'a pas d'influence sur la ressemblance entre individus
- la réduction supprime l'arbitrage des unités et toutes les variables ont la même influence dans le calcul des distances entre individus

Particularités de ces nouvelles variables :

- les moyennes sont toutes nulles
- les écart types sont égaux à 1
- $\text{COV}(z_j, z_k) = r_{z_j z_k} = \frac{1}{N} \sum z_{ij} z_{ik}$ avec $j \neq k$

La matrice Z des variables centrées réduites s'écrit alors :

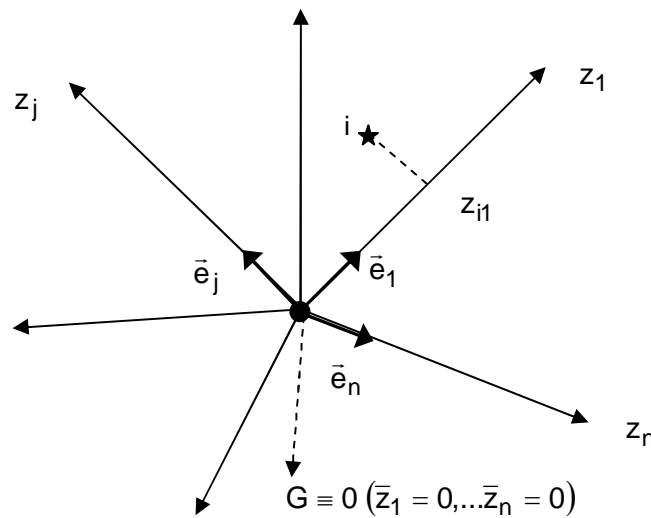


L'information contenue dans cette matrice est donnée par le nuage de points des individus dans l'espace \mathbb{R}^n et par le nuage de points des variables dans l'espace \mathbb{R}^N .

Plaçons nous dans l'espace \mathbb{R}^n des variables qui contient le nuage des N points individus. Le système des n axes est orthonormée ou encore la base de ce système $(\vec{e}_1, \vec{e}_2, \dots, \vec{e}_n)$ est orthonormée, c'est-à-dire :

$$\begin{cases} \|\vec{e}_1\| = \dots = \|\vec{e}_n\| = 1 \\ \vec{e}_l \cdot \vec{e}_j = 0 \quad \forall l \neq j \end{cases}$$

Un point i quelconque a pour coordonnées $(z_{i1}, z_{i2}, \dots, z_{in})$. La distance entre deux points est calculée par la distance euclidienne (théorème de Pythagores). On peut schématiser cet espace de la façon suivante :



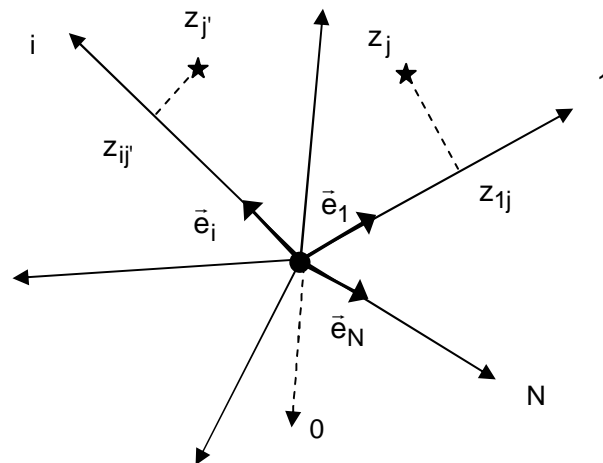
La distance au carré entre i et i' est égale à : $d^2(i, i') = \sum_{j=1}^n (z_{ij} - z_{i'j})^2$.

Les projections orthogonales (les coordonnées) des N points sur un axe quelconque z_j sont de moyennes nulles et de variance égale à un par construction. Le centre de gravité G est donc l'origine des axes. La variance totale du nuage multidimensionnel est égale à $1 \cdot n = n$. Chaque axe porte donc $\frac{1}{n} \cdot 100$ de la variance totale.

Dans l'espace \mathbb{R}^n des individus, on dispose du nuage des n variables. Le système des N axes est orthonormé : la base de ce système s'écrit : $(\vec{e}_1, \vec{e}_2, \dots, \vec{e}_N)$ avec :

$$\begin{cases} \|\vec{e}_1\| = \dots = \|\vec{e}_N\| = 1 \\ \vec{e}_i \cdot \vec{e}_l = 0 \quad \forall i \neq l \end{cases}$$

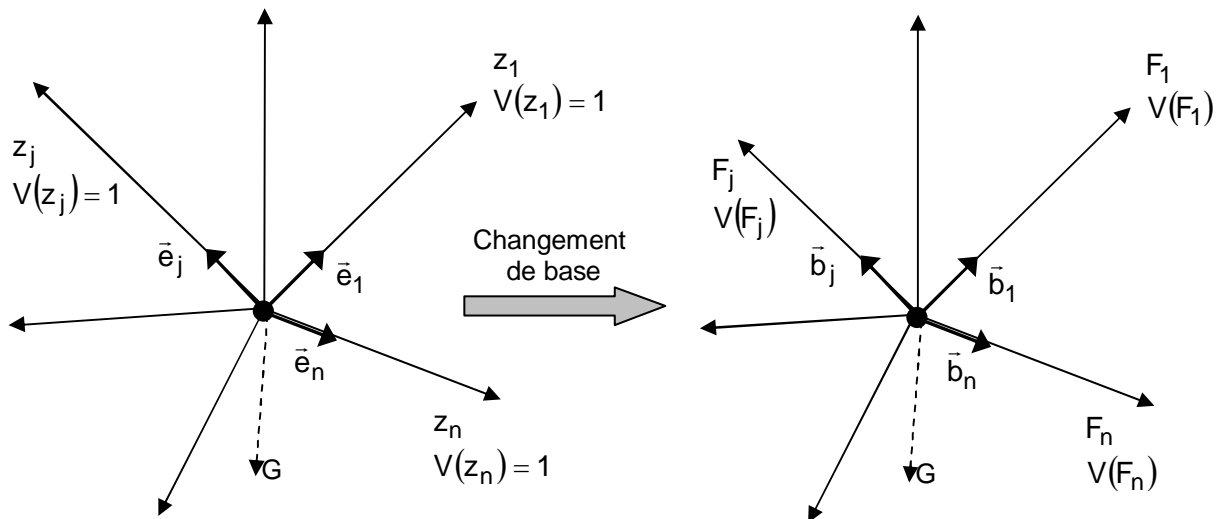
Un point z_i quelconque a pour coordonnées $(z_{1j}, z_{2j}, \dots, z_{Nj})$. La distance entre deux points est calculée par la distance euclidienne. On peut schématiser cet espace de la façon suivante :



La distance au carré entre z_j et z_j' est égale à $d^2(z_j, z_j') = \sum_{i=1}^N (z_{ij} - z_{ij'})^2$

L'information contenue dans ces espaces est illisible du fait du nombre d'axes. L'ACP a pour but de substituer à ces espaces, des espaces de même dimension mais de telle sorte qu'une grande part de l'information soit lisible à partir d'un, deux ou au maximum trois axes.

Le principe de l'ACP consiste donc à effectuer dans R^N et dans R^N un changement de base de telle sorte (lorsque cela est possible) que les variances des projections orthogonales (les coordonnées) sur les nouveaux axes (appelés axes principaux) rassemblent une part significative de la variance totale à partir des deux ou trois premiers axes. On peut schématiser ce principe de la façon suivante dans R^n .



La variance des coordonnées des N points individus sur z_j (quelquesoit j), explique $\frac{1}{n} * 100$ de la variance totale.

Système orthonormé d'axes principaux \bar{F}_j (quelquesoit j), de même origine, de base orthonormée : $(\bar{b}_1, \dots, \bar{b}_n)$ et tel que la variance des coordonnées de N points individus sur \bar{F}_1, \bar{F}_2 et au maximum \bar{F}_3 représente, par exemple, 70% à 80% de la variance totale n .

Le vecteur \vec{F}_j , quelquesoit j, dont les éléments sont les projections orthogonales (les coordonnées) des N points sur l'axe F_j porte le nom de composante principale. Il s'écrit :

$$\vec{F}_j = \begin{bmatrix} F_{1j} \\ \dots \\ F_{ij} \\ \dots \\ F_{Nj} \end{bmatrix}$$

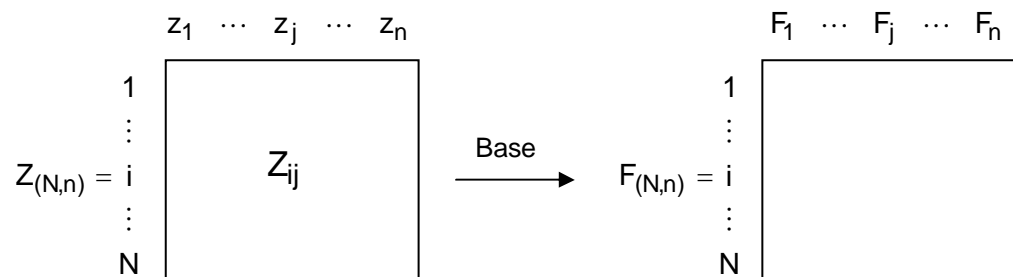
(N,1)

B) Ecriture des composantes principales dans R^n

1 Le produit scalaire

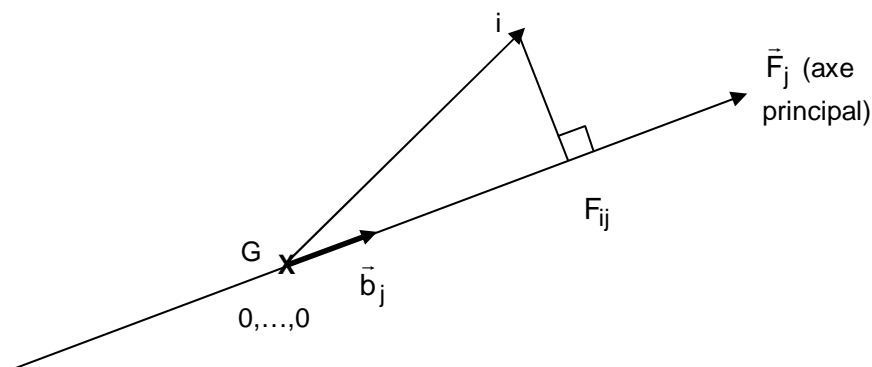
L'ACP vise donc à projeter dans R^n orthogonalement les N points individus sur n nouveaux axes appelés axes principaux, sachant que l'origine de ces nouveaux axes reste identique à celui de l'espace de départ. Ce changement d'axes a pour but (lorsque cela est possible), de lire l'information concernant les individus en utilisant un espace restreint à 2 ou 3 dimensions (au maximum).

On remplace donc Z par une nouvelle matrice F :



Pour cela, on effectue un changement de base dans l'espace R^n de départ.

Rappelons que si on connaît les coordonnées d'un vecteur quelconque \vec{b}_j dans la base R^n de départ, la projection orthogonale F_{ij} (la coordonnée) d'un point i du nuage des N points est donnée par le produit scalaire du vecteur \vec{b}_j par le vecteur $\vec{G\vec{i}}$ ou G est l'origine des axes (G est le centre de gravité du nuage des N points) :



$$\vec{F}_{ij} = \vec{G}_i * \vec{b}_j$$

$$= [z_{i1}, \dots, z_{ik}, \dots, z_{in}] \begin{bmatrix} b_{1j} \\ \dots \\ b_{kj} \\ \dots \\ b_{nj} \end{bmatrix}$$

= coordonnées du point i (ligne i de la matrice Z) x coordonnées de \vec{b}_j dans l'espace R^n de départ.

$$= [z_{i1} * b_{1j} + \dots + z_{ik} * b_{kj} + \dots + z_{in} * b_{nj}]$$

L'ensemble des projections orthogonales des N points du nuage sur l'axe principal \vec{F}_j constitue les éléments du vecteur \vec{F}_j que l'on appelle la composante principale. on a :

$$\boxed{\vec{F}_j = Z \vec{b}_j}$$

(N,1) (N,n) (n,1)

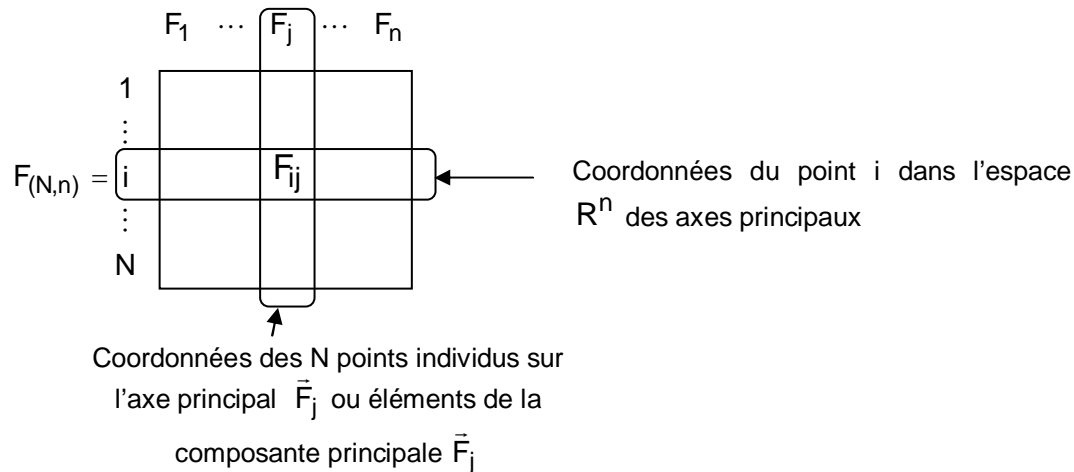
$$\begin{bmatrix} F_{1j} \\ \vdots \\ F_{ij} \\ \vdots \\ F_{Ni} \end{bmatrix} = \begin{bmatrix} z_{i1} & \dots & z_{ih} & \dots & z_{in} \end{bmatrix} \begin{bmatrix} b_{1j} \\ \vdots \\ b_{hj} \\ \vdots \\ b_{nj} \end{bmatrix} = \begin{bmatrix} b_{1j}z_{i1} + \dots + b_{nj}z_{in} \end{bmatrix}$$

Si dans l'expression encadrée de \vec{F}_j , on fait varier j de 1 à n on obtient par les produits scalaires l'ensemble des projections orthogonales (coordonnées) de tous les points N sur tous les axes principaux. Ces coordonnées sont les éléments de la matrice :

$$F = Z B$$

(N,n) (N,n) (n,n)

La matrice F s'écrit et s'interprète :



On vérifie avec le produit scalaire $F_j = Zb_j$ qui définit la $j^{\text{ème}}$ composante principale, qu'il s'agit d'une combinaison linéaire des variables de départ, les poids étant les éléments du nouveau vecteur \vec{b}_j du changement de base.

2 Détermination des composantes principales

Les composantes principales sont les coordonnées des N projections orthogonales sur les axes principaux. Ces composantes principales (vecteurs principaux) sont le résultat du produit scalaire

$$\underset{(N,n)}{F} = \underset{(N,n)}{Z} \underset{(n,n)}{B}$$

Pour déterminer le contenu de F il faut connaître celui de B. Pour cela, considérons la $j^{\text{ème}}$ composante principale $F_j = \underset{(N,1)}{Z} \underset{(n,1)}{b_j}$ et déterminons ses caractéristiques.

- Calculons la moyenne de la composante principale :

$$\begin{aligned} \bar{F}_j &= \frac{1}{N} \sum_i F_{ij} = \frac{1}{N} \sum_i Z_{i1} b_{1j} + \dots + \frac{1}{N} \sum_i Z_{in} b_{nj} \\ &= b_{1j} \underbrace{\left(\frac{1}{N} \sum_i Z_{i1} \right)}_{\bar{Z}_1=0} + \dots + b_{nj} \underbrace{\left(\frac{1}{N} \sum_i Z_{in} \right)}_{\bar{Z}_n=0} \end{aligned}$$

$$\bar{F}_j = 0$$

Toutes les composantes principales sont centrées. Les axes principaux ont toujours pour origine le point G, le centre de gravité du nuage des N points.

- Calculons la variance :

$$\begin{aligned} V[F_j] &= \frac{1}{N} \sum_{i=1}^N (F_{ij} - \bar{F}_j)^2 = \frac{1}{N} \sum_{i=1}^N F_{ij}^2 = \frac{1}{N} F_j' F_j \\ &= \frac{1}{N} (Zb_j)' (Zb_j) = \frac{1}{N} (b_j' Z') (Zb_j) = b_j' \left(\frac{1}{N} Z' Z \right) b_j \end{aligned}$$

Et avec $\underset{(n,n)}{R} = \frac{1}{N} Z' Z$ (la matrice des coefficients de corrélation linéaire des variables z_j de Z) :

$$V[F_j] = b_j' R b_j \text{ qui est l'expression d'une forme quadratique.}$$

Pour atteindre l'objectif de l'ACP, il est nécessaire qu'avec un, deux ou au maximum trois axes principaux on puisse expliquer une part significative de la variance totale du nuage (70 à 80% par exemple) Cela revient donc à maximiser la variance d'une composante principale F_j quelconque. Le problème mathématique à résoudre est donc :

$\text{Max}(V[F_j])$ sachant que le nouveau vecteur de la base \vec{b}_j de l'axe principal \bar{F}_j est un vecteur normé, c'est-à-dire :

$$\|\vec{b}_j\| = 1 = b_{1j}^2 + \dots + b_{nj}^2$$

Il s'agit de la maximisation d'une forme quadratique sous contrainte :

$$\begin{cases} \text{Max}(V[F_j]) = b_j' R b_j \\ b_j' b_j = 1 \end{cases}$$

Pour résoudre ce problème on utilise le Lagrangien :

$L = b_j' R b_j - \lambda (b_j' b_j - 1)$ et on cherche la dérivée de L par rapport à l'inconnu b_j sachant que la dérivée de la forme quadratique est $\frac{\partial (b_j' R b_j)}{\partial b_j} = 2R b_j$

$$\frac{\partial L}{\partial b_j} = 2R b_j - 2\lambda b_j = 0$$

$$R b_j - \lambda b_j = 0 \Leftrightarrow R b_j = \lambda b_j \Leftrightarrow [R - \lambda I] b_j = 0$$

En définitive, maximiser la $V[F_j]$ sous la contrainte que b_j soit un vecteur unitaire revient à trouver la solution du système d'équation homogène $[R - \lambda I] b_j = 0$. Or on sait qu'en dehors de la solution triviale donnée par le déterminant $[R - \lambda I] \neq 0$, il existe une infinité de vecteurs b_j appelés vecteurs propres, associés aux différentes valeurs propres λ_j de la matrice R . On appelle diagonalisation de la matrice R l'ensemble de ces recherches de valeurs propres et de vecteurs propres.

Le vecteur propre b_j est un vecteur normé. De ce fait, parmi l'infinité des vecteurs b_j , on choisira un vecteur unitaire c'est-à-dire tel que $b_j' b_j = 1$.

On peut enfin constater que si dans l'expression de la variance $V[F_j]$, on remplace $R b_j$ par la solution que l'on vient de trouver ($\lambda_j b_j$), on aura :

$$\begin{aligned} V[F_j] &= b_j' \lambda_j b_j \\ &= \lambda_j \underbrace{b_j' b_j}_{=1} \end{aligned} \quad \boxed{V[F_j] = \lambda_j}$$

En définitive, la solution du problème est le calcul du vecteur propre normé qui correspond à la plus forte valeur propre de la matrice R . Comme R est de dimension n , on va faire l'hypothèse qu'il existe n valeurs propres distinctes pour cette matrice. On disposera donc de n valeurs propres et de n vecteurs propres normés.

En classant les valeurs propres par ordre décroissant et en affectant à chacune des valeurs propres classées son vecteur propre normé correspondant, on construit ainsi la matrice B du changement de base qui a ses coordonnées dans l'espace d'origine.

3 Caractéristiques des composantes principales

Considérons la matrice B obtenue précédemment :

$$\begin{array}{c}
 \lambda_1 \\
 \searrow \\
 \lambda_n \\
 \begin{array}{cccc}
 \bar{b}_1 & \cdots & \bar{b}_j & \cdots & \bar{b}_n \\
 \begin{array}{c} z_1 \\ \vdots \\ z_k \\ \vdots \\ z_n \end{array} & \boxed{\begin{array}{c} \\ \\ b_{kj} \\ \\ \end{array}} & & & \\
 \mathbf{B}_{(n,n)} = \mathbf{Z}_k & & & &
 \end{array}
 \end{array}$$

Les vecteurs colonnes de cette matrice sont normés et orthogonaux 2 à 2.

Les vecteurs b_j constituent donc une base orthonormée.

La matrice B qui contient cette nouvelle base est donc une matrice orthogonale. Elle vérifie la propriété suivante :

$$\mathbf{B}' = \mathbf{B}^{-1} \text{ ou encore } \mathbf{B}'\mathbf{B} = \mathbf{B}\mathbf{B}' = \mathbf{I}$$

Utilisons ces résultats dans l'expression précédente de la variance de F_j :

$$V[F_j] = \lambda_j = b_j' R b_j$$

En faisant varier j de 1 à n on a l'expression matricielle $\mathbf{B}'\mathbf{R}\mathbf{B} = \mathbf{\Lambda}$ avec :

$$\mathbf{\Lambda} = \begin{bmatrix} \lambda_1 & & 0 \\ & \lambda_j & \\ 0 & & \lambda_n \end{bmatrix} \text{ la matrice diagonale des valeurs propres rangées par ordre décroissant.}$$

Dans cette expression B est une matrice orthogonale. De ce fait, la trace de la matrice R est égale à la trace de la matrice $\mathbf{\Lambda}$, c'est-à-dire $1 + 1 + \dots + 1 = n = \sum_j \lambda_j$

Il est donc possible de calculer le pourcentage de variance totale $\frac{\lambda_j}{n} * 100 = \frac{\text{tr}[\mathbf{R}]}{n} * 100$ expliquée par les différentes composantes principales et de déterminer si avec un nombre restreint d'axes principaux on peut lire l'information donnée par le nuage des N points individus.

Conclusion : **résumé de la démarche dans \mathbf{R}^n**

- On dispose d'un tableau $X_{(N,n)}$.
- On transforme la matrice X en une matrice Z de variables centrées réduites. Les variables sont de moyenne nulle et de variance égale à 1. La covariance entre deux variables est égale au coefficient de corrélation linéaire entre les 2 variables.

On se place dans l'espace R^n avec un système orthonormé. On analyse le nuage des N points individus.

- On calcule $R = \frac{1}{N} Z'Z$ qui est la matrice des coefficients de corrélation linéaire entre les variables. C'est la matrice d'information des variables.

C'est une matrice carrée, symétrique qui contient sur la diagonale les variances des variables et de part et d'autre les coefficients de corrélation entre 2 variables quelconques.

- On diagonalise R , c'est-à-dire qu'on calcule les n valeurs propres de cette matrice.

On divise chacune des valeurs par n , ce qui donne le pourcentage de variance totale expliqué par une composante principale.

On ordonne par ordre décroissant ces pourcentages que l'on cumule et si avec 1, 2 ou au maximum 3 axes on explique 70% de la variance totale, alors l'ACP est réalisable, sinon on arrête les calculs.

- Si l'ACP est réalisable, on calcule alors les vecteurs propres normés associés aux valeurs propres de R et on place les coordonnées de ces vecteurs propres en colonne d'une matrice B dans l'ordre des valeurs propres décroissantes correspondantes.
- On effectue alors le produit scalaire $F = ZB$, la matrice F contient donc en colonne les composantes principales, c'est-à-dire les projections orthogonales du nuage des N points individus sur les nouveaux axes.
- On sélectionne alors les 2, au maximum les 3 premières colonnes de la matrice F et on réalise le graphique qui permet de lire l'information concernant les individus de la matrice Z .

C) Détermination des composantes principales dans R^N

Dans l'espace R^N se situe le nuage des n variables. Ici aussi le système choisi est orthonormé et une colonne du tableau fournit les coordonnées de la variable. Comme on travaille à partir de Z , les moyennes des lignes (c'est-à-dire des individus) n'ont aucune raison d'être nulles. Donc dans ce cas, l'origine des axes n'est pas au centre de gravité des variables.

Dans le deuxième exercice on a vu que si on conserve Z , la matrice d'information des individus est donnée par la matrice $V = ZZ'$ qui est une matrice de dispersion des individus. On pourrait donc appliquer le schéma précédent à cet espace, c'est-à-dire diagonaliser V , calculer les vecteurs propres normés correspondants, construire une matrice de changement de base et par l'opération produit scalaire calculer les projections orthogonales des n variables sur les axes principaux.

Concrètement, pour toutes les analyses de données, ces calculs sont inutiles. En effet, nous allons démontrer qu'il est possible de calculer les valeurs propres et les vecteurs propres de V sans utiliser la diagonalisation de cette matrice. Ce calcul s'opère à partir des formules de transition.

1 Démonstration des formules de transition

On va se placer dans R^n , où l'on connaît les valeurs propres et les vecteurs propres normés de R .

La diagonalisation de R s'écrit :

$$Rb_j = \lambda_j b_j \text{ avec } R = \frac{1}{N} Z'Z \text{ et } b_j' b_j = 1$$

$$\text{D'où } \frac{1}{N} Z'Z b_j = \lambda_j b_j$$

En multipliant les deux membres de cette équation par Z , on obtient :

$$\frac{1}{N} ZZ'Z b_j = \lambda_j Z b_j \text{ avec } V = ZZ' \text{ on a : } V(Zb_j) = N\lambda_j(Zb_j)$$

et en posant $Z \begin{matrix} b_j \\ \vdots \\ b_n \end{matrix} = \begin{matrix} u_i \\ \vdots \\ u_n \end{matrix}$ on a

$$Vu_i = (N\lambda_j)u_i$$

C'est l'écriture des vecteurs propres u_i associés aux valeurs propres $N\lambda_j$ de la matrice V .

Connaissant les valeurs propres λ_j de la matrice R , il suffit donc de les multiplier par N pour obtenir les valeurs propres de V . Or R est de dimension n , il n'y a donc que n valeurs propres pour R alors que V est de dimension (N,N) .

On peut démontrer que les $N-n$ autres valeurs propres sont nulles.

Le vecteur u_i qui nous intéresse doit être un vecteur propre normé. Il doit donc vérifier $u_i'u_i = 1$. Or

$$\begin{aligned} u_i'u_i &= (Zb_j)'(Zb_j) = b_j'Z'Zb_j \\ &= b_j'NRb_j = b_jN\lambda_j b_j \\ &= N\lambda_j b_j'b_j = N\lambda_j \neq 1 \end{aligned}$$

u_i n'est donc pas un vecteur propre normé. Pour le normer on lui impose que :

$$(ku_i)'(ku_i) = 1 \text{ avec } k \in \mathbb{R} ; \text{ soit :}$$

$$k^2 u_i'u_i = 1$$

$$k^2 (N\lambda_j) = 1$$

$$\text{D'où } k = \pm \frac{1}{\sqrt{N\lambda_j}}$$

Le vecteur propre normé noté c_i s'écrit alors : $c_i = ku_i = \frac{1}{\sqrt{N\lambda_j}} Zb_j$

Cette formule montre que connaissant b_j et λ_j , les vecteurs propres normés associés aux valeurs propres de R , il est possible de calculer le vecteur propre normé c_i associé à la valeur propre $N\lambda_j$ de la matrice V .

Cette formule qui permet de calculer les vecteurs propres normés de la nouvelle base de \mathbb{R}^N sans diagonaliser V porte le nom de formule de transition.

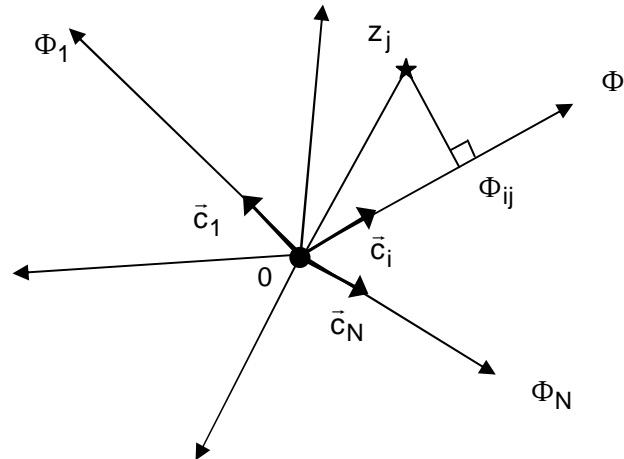
On peut démontrer que réciproquement, connaissant les vecteurs propres normés associés aux valeurs propres de V , il est possible de calculer les valeurs propres et les vecteurs propres normés b_j de la matrice R .

En définitive les formules de transition s'écrivent :

$$\begin{aligned} c_i &= ku_i = \frac{1}{\sqrt{N\lambda_j}} Zb_j \\ &\text{et} \\ b_j &= \frac{1}{\sqrt{N\lambda_j}} Z'c_i \end{aligned}$$

2 Les coordonnées des projections des variables

Le changement de base orthonormée $(\bar{c}_1, \dots, \bar{c}_i, \dots, \bar{c}_N)$ (où les c_i sont des vecteurs unitaires déterminés par leurs coordonnées dans l'ancienne base) permet de calculer, comme précédemment, les projections orthogonales (les coordonnées) des variables sur les i axes principaux : $\bar{\Phi}_i$ en utilisant le produit scalaire entre le vecteur unitaire \bar{c}_i et le vecteur $\bar{O}\bar{Z}_j$ où O est l'origine de N axes.



On appelle Φ_{ij} la coordonnée de z_j sur l'axe principal $\bar{\Phi}_i$ obtenu par le produit scalaire entre \bar{c}_i et $\bar{O}\bar{Z}_j$. De ce fait $\Phi_i = C_i' Z$ est la composante principale c'est-à-dire les coordonnées des

variables z_j ($j = 1 \dots n$) sur l'axe principal $\bar{\Phi}_i$.

Pour généraliser ce résultat à une matrice Φ qui contient toutes les coordonnées il faut tenir compte du résultat (formules de transition) précédent qui indique qu'il y a dans cet espace $N-n$ valeurs propres de V qui sont nulles et donc $N-n$ axes principaux qui n'existent pas. La matrice Φ est donc de dimension (n,n) et non (N,n) . Elle s'écrit :

$$\Phi_{(N,n)} = \begin{matrix} & z_1 & z_j & z_n \\ \Phi_1 & \begin{bmatrix} \dots & \dots & \dots & \dots & \dots \end{bmatrix} \\ \dots & \dots & \dots & \dots & \dots \\ \Phi_i & \dots & \Phi_j & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ \Phi_n & \dots & \dots & \dots & \dots \end{matrix} = \Phi_{(n,n)}$$

$$\begin{matrix} \dots & \dots & \dots & \dots & \dots \\ \Phi_N & \dots & \dots & \dots & \dots \end{matrix}$$

De même pour la matrice C qui contient les coordonnées des vecteurs propres normés dans l'ancien système :

$$C_{(N,N)} = \begin{matrix} & c_1 & & c_i & & c_n & \dots & c_N \\ \begin{matrix} 1 \\ \vdots \\ i \\ \vdots \\ N \end{matrix} & \begin{bmatrix} \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & c_{ii} & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \end{bmatrix} & = C_{(N,n)} \end{matrix}$$

De ce fait on a : $\Phi = \begin{matrix} C' & Z \\ (n,n) & (n,N)(N,n) \end{matrix}$

Or on sait (formule de transition) que : $c_i = \frac{1}{\sqrt{N\lambda_j}} Zb_j$ ce qui s'écrit en faisant varier i et j :

$$C_{(N,n)} = \frac{1}{\sqrt{N}} \begin{matrix} Z & B & \Lambda^{-1/2} \\ (N,n) & (n,n) & (n,n) \end{matrix}$$

D'où

$$\Phi = \frac{1}{\sqrt{N}} \Lambda^{-1/2} B' Z' Z$$

soit avec $R = \frac{1}{N} Z' Z \Leftrightarrow Z' Z = NR$

$$\Phi = \frac{1}{\sqrt{N}} \Lambda^{-1/2} B' R = \sqrt{N} \Lambda^{-1/2} B' R$$

Or : $B' R b = \Lambda \Rightarrow B' R = \Lambda B^{-1} = \Lambda B'$

D'où

$$\Phi = \sqrt{N} \Lambda^{-1/2} \Lambda B'$$

$$\Phi = \begin{matrix} \sqrt{N} & \Lambda^{1/2} & B' \\ (n,n) & (n,n) & (n,n) \end{matrix}$$

3 Le choix des projections des variables

Dans le paragraphe précédent, nous avons vu que les projections orthogonales des n variables sur les axes principaux avaient pour coordonnées le contenu de la matrice $\phi = \begin{matrix} \sqrt{N} & \Lambda^{1/2} & B' \\ (n,n) & (n,n) & (n,n) \end{matrix}$.

Intéressons nous au produit $\Lambda^{1/2} B'$ dans cette formule, et pour cela rappelons quelles sont les caractéristiques des composantes principales F_j dans l'espace R^n .

$$\begin{cases} \bar{F}_j = 0 \\ V[F_j] = \lambda_j \\ Cov[F_j, F_k] = 0 \quad j \neq k \end{cases}$$

Cette dernière caractéristique peut être vérifiée de la façon suivante :

$$\text{Cov}(F_j, F_k) = \frac{1}{N} \mathbf{F}_k' \mathbf{F}_j \quad \text{Soit avec } F_j = \mathbf{Z}b_j \quad \forall j$$

$$\begin{aligned} \text{Cov}(F_j, F_k) &= \frac{1}{N} \mathbf{b}_k' \mathbf{Z}' \mathbf{Z} b_j \\ &= \mathbf{b}_k' \mathbf{R} b_j \\ &= \mathbf{b}_k' (\lambda_j b_j) \\ &= \lambda_j \mathbf{b}_k' b_j \end{aligned}$$

Comme \mathbf{b}_k et b_j sont des vecteurs propres normés, leur produit scalaire est nul et de fait :

$$\text{Cov}(F_j, F_k) = 0$$

Normons les composantes principales F_j

On sait que :

$$V[F_j] = \frac{1}{N} \sum_i F_{ij}^2 = \lambda_j$$

$$\text{Appelons alors } \tilde{F}_{ij} = \frac{F_{ij}}{\sqrt{\lambda_j}} = \lambda_j^{-1/2} F_{ij}$$

$$\text{De ce fait } V[\tilde{F}_{ij}] = \frac{1}{N} \sum_i \tilde{F}_{ij}^2 = \frac{1}{N} \sum_i \frac{F_{ij}^2}{(\sqrt{\lambda_j})^2} = \frac{1}{N} \sum_i F_{ij}^2 * \frac{1}{\lambda_j} = 1$$

Sous forme matricielle on aura :

$$\tilde{\mathbf{F}}_{(N,n)} = \mathbf{F} \mathbf{\Lambda}^{-1/2} \quad \text{avec } \mathbf{\Lambda}^{-1/2} = \begin{bmatrix} \frac{1}{\sqrt{\lambda_1}} & & \\ & \frac{1}{\sqrt{\lambda_j}} & \\ & & \frac{1}{\sqrt{\lambda_n}} \end{bmatrix}$$

Les nouvelles composantes principales ont alors pour caractéristiques :

$$\begin{cases} \tilde{\mathbf{F}}_j = 0 \\ V[\tilde{F}_j] = 1 \\ \text{Cov}[\tilde{F}_j, \tilde{F}_k] = 0 \quad j \neq k \end{cases}$$

On peut alors vérifier les résultats suivants :

$$\frac{1}{N} \tilde{\mathbf{F}}' \tilde{\mathbf{F}} = \mathbf{I}_n$$

$\frac{1}{N} \tilde{F}'Z = A_{(n,n)}$ qui contient les coefficients de corrélation linéaires entre les composantes principales \tilde{F}_j et les variables z_k . avec $j = 1, \dots, n$ et $k = 1, \dots, n$

En utilisant ces deux résultats on a :

$$\tilde{F} = F\Lambda^{-1/2} = ZB\Lambda^{-1/2}$$

On en déduit que :

$$Z = \tilde{F}\Lambda^{1/2}B' \text{ (puisque } B' = B^{-1}\text{)}$$

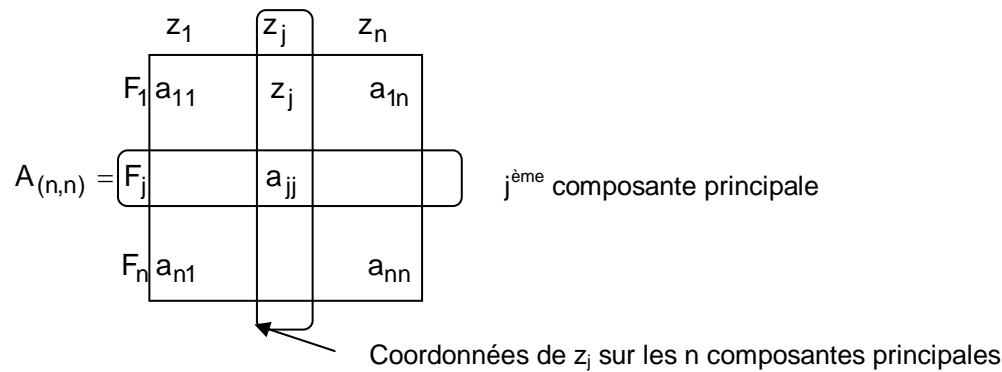
D'où

$$A = \frac{1}{N} \tilde{F}'Z \text{ s'écrit : } A = \frac{1}{N} \tilde{F}'\tilde{F}\Lambda^{1/2}B'$$

Soit :

$$A = \Lambda^{1/2}B'$$

En définitive on va choisir pour coordonnées des variables dans la nouvelle base, le contenu de la matrice A (au lieu de celui de Φ) qui est constitué de coefficients de corrélation linéaire, c'est-à-dire de chiffres en valeurs absolues inférieurs ou égaux à 1. Cette matrice s'écrit :



Propriétés de la matrice A :

En utilisant l'expression $R = \frac{1}{N} Z'Z$ dans la formule $Z = \tilde{F}\Lambda^{1/2}B' = \tilde{F}A$ on a :

$$\begin{aligned} R &= \frac{1}{N} (\tilde{F}A)' (\tilde{F}A) \\ &= \frac{1}{N} A' \tilde{F}' \tilde{F} A \\ &= A' \left(\frac{1}{N} \tilde{F}' \tilde{F} \right) A \end{aligned}$$

$$R = A'A$$

Quand on norme les composantes principales, la matrice $R = A'A$.

Développons la diagonale de cette égalité :

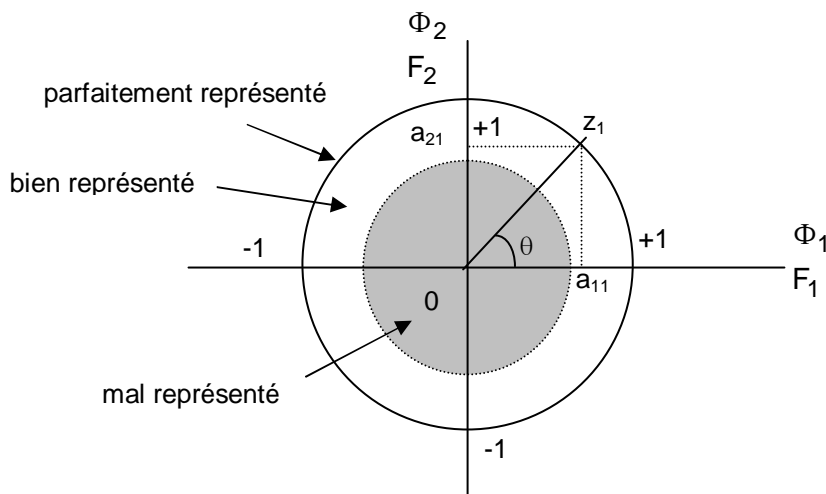
diag R = diag A'A

1 ^{er} élément	$a_{11}^2 + \dots$	+	$a_{j1}^2 + \dots$	+	a_{n1}^2	= 1 = V(z ₁)
p ^{ème} élément	$a_{1p}^2 + \dots$	+	$a_{jp}^2 + \dots$	+	a_{np}^2	= 1 = V(z _p)
n ^{ème} élément	$a_{1n}^2 + \dots$	+	$a_{jn}^2 + \dots$	+	a_{nn}^2	= 1 = V(z _n)
Σ	λ_1		λ_j		λ_n	Trace R = VT = n

D - L'interprétation d'une ACP

1) la projection des variables

En retenant pour projection des variables le contenu de A, les coordonnées de ces projections sont toutes en valeur absolue inférieures ou égales à 1, c'est-à-dire sont toutes insérées dans un cercle de rayon unité appelé cercle des corrélations. Considérons, par exemple, le cercle de corrélation dans le plan constitué par deux axes principaux et une variable z_1 sur ce cercle. Elle a pour coordonnées a_{11} et a_{21} (Cf. matrice A).



D'après le théorème de pythagore, on peut écrire que :

$$a_{11}^2 + a_{21}^2 = 0z_1^2 = 1$$

En utilisant la propriété précédente de la matrice A (1^{er} élément du développement de la diagonale), on a :

$$a_{11}^2 + a_{21}^2 + a_{31}^2 + \dots + a_{n1}^2 = V[z_1] = 1$$

Comme dans cette expression $a_{11}^2 + a_{21}^2 = 1$ (z_1 sur le cercle), on a :

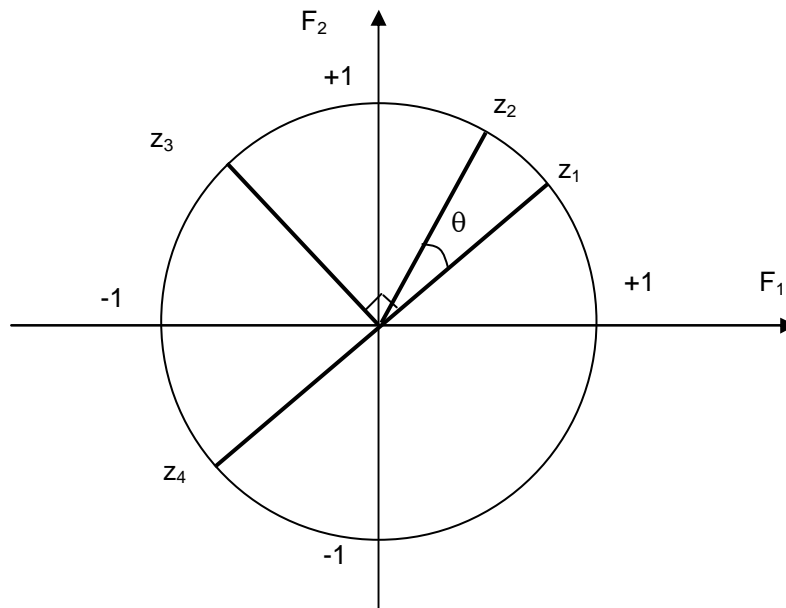
$$a_{31}^2 + \dots + a_{n1}^2 = 0$$

Ce qui signifie que z_1 a ses coordonnées nulles sur les autres axes principaux. En d'autres termes, z_1 est parfaitement représenté sur le plan F_1, F_2 .

Plus généralement :

- une variable qui se trouve en projection sur le cercle de corrélation est une variable parfaitement déterminée sur le plan.
- Une variable proche du cercle de corrélation est une variable dite « bien représentée » dans le plan.
- Une variable proche de l'origine du cercle (zone grisée du graphique) est une variable mal représentée.

Considérons à présent le plan (F_1, F_2) avec les quatre variables z_1, z_2, z_3 et z_4 situées sur le cercle de corrélation de la façon suivante :



On sait que dans un plan, 2 vecteurs forment un angle θ dont le $\cos = r$

Pour des variables parfaitement représentées (ou bien représentées à la limite) :

- si l'angle θ tend vers 0 : $r_{z_1, z_2} \longrightarrow 1$

- si l'angle θ tend vers $90^\circ = \left(\frac{\pi}{2}\right)$: $r_{z_1, z_2} \longrightarrow 0$

- si l'angle θ tend vers $180 = \pi$: $r_{z_1, z_2} \longrightarrow -1$

En définitive, pour interpréter le graphique des projections des variables :

- on sélectionne les variables parfaitement ou bien représentées
- on analyse les proximités des variables sélectionnées en terme de corrélation

2) Le graphique des individus

On porte sur les plans factoriels les points individus dont les coordonnées sont données par la matrice F.

3) les aides à l'interprétation : les contributions absolues et relatives

La contribution absolue (CTA) d'un individu (ou d'une variable) à une composante principale est la part de sa variance expliquée par cet individu ou cette variable.

La contribution relative (CTR) d'un individu (ou d'une variable) à une composante principale indique la position de cet individu (ou variable) par rapport à l'axe principal qui porte cette composante principale.

En ACP, on utilise ces aides à l'interprétation que pour les individus. En effet, de part la représentation des variables dans le cercle de corrélation, ces contributions sont inutiles.

a) Les contributions absolues des individus (CTA)

La contribution absolue d'un individu i sur la composante principale F_j est notée : CTA_{i/F_j}

$$\text{Comme } V[F_j] = \frac{1}{N} \sum_i F_{ij}^2 = \frac{1}{N} (F_{1j}^2 + \dots + F_{ij}^2 + \dots + F_{nj}^2)$$

Par définition :

$$CTA_{i/F_j} = \frac{F_{ij}^2}{\sum_i F_{ij}^2} = \frac{F_{ij}^2}{N \cdot V[F_j]} = \frac{F_{ij}^2}{N \lambda_j}$$

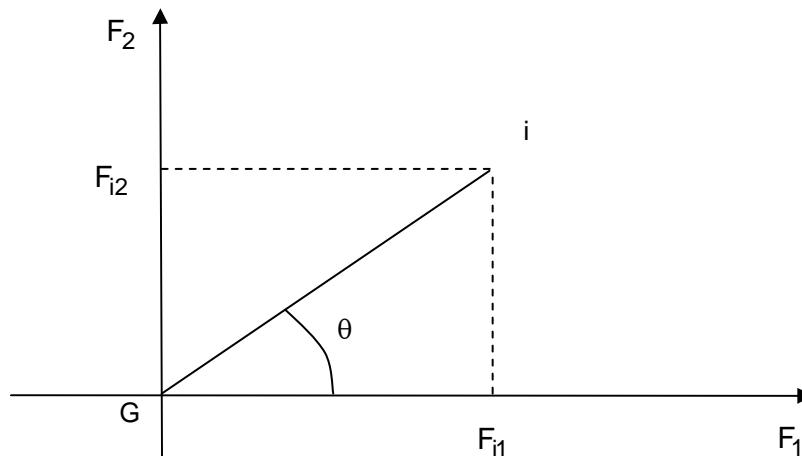
Par construction $0 \leq CTA \leq 1$ et $\sum_{i=1}^n CTA_{i/F_j} = 1$.

Les CTA s'interprètent comme des pourcentages et permettent de sélectionner les individus qui contribuent le plus à l'apparition d'un axe principal.

b) La contribution relative (CTR)

La contribution relative d'un individu i sur la composante principale F_j est notée : CTR_{i/F_j} .

Considérons un plan factoriel (F_1, F_2) avec un point i quelconque de coordonnées F_{i1} et F_{i2} (Cf. la matrice F) :



L'angle θ indique par sa valeur la plus ou moins grande proximité de i par rapport à l'axe F_1 . On ainsi :

$$\cos\theta = \frac{\text{coté adjacent}}{\text{hypoténuse}} = \frac{F_{i1}}{d(G,i)} \text{ soit}$$

$$\cos^2\theta = \frac{F_{ij}^2}{d^2(G,i)}$$

et si θ tend vers 0, $\cos^2\theta$ tend vers 1 et i est proche de F_1 . Au contraire, si θ tend vers $\frac{\pi}{2}$, $\cos^2\theta$ tend vers 0 et i s'éloigne de F_1 pour se rapprocher de F_2 .

On note la quantité $\frac{F_{i1}^2}{d^2(G,i)}$ la CTR_{i/F_1}

En réitérant la procédure pour l'individu i par rapport à F_2 on aura : $\frac{F_{i2}^2}{d^2(G,i)}$ la CTR_{i/F_2}

On constate alors que :

$$0 \leq CTR \leq 1$$

$$\text{et } CTR_{i/F_1} + CTR_{i/F_2} = \frac{F_{i1}^2 + F_{i2}^2}{d^2(G,i)} = 1.$$

En généralisant ce résultat pour un axe principal quelconque, on aura :

$$CTR_{i/F_j} = \frac{F_{ij}^2}{d^2(G,i)} = \frac{F_{ij}^2}{\sum_{j=1}^n F_{ij}^2}$$

avec $0 \leq CTR \leq 1$ et $\sum_{j=1}^n CTR_{i/F_j} = 1$

Les CTR permettent de vérifier si les individus sélectionnés par les CTA sont proches ou au contraire éloignés des axes principaux.

Schéma de calcul des CTA et CTR :

		F_1	...	F_j	...	F_n	\sum^2	CTR_1	CTR_j	CTR_n	Σ
$F =$ (N,n)	1	...					Y				1
	...										
	i	F_{1i}	...	$F_{ij}=A$	$\frac{A^2}{Y}$			1			
									
	N							...			

		Z		
		CTA_1	CTA_j	CTA_n
\sum^2	1	...		
	...			
	i	...	$\frac{A^2}{Z}$	
		
	N			
Σ	1		1	

Avec les CTR on sait si les individus sont proches ou loin de l'axe. (Si CTR proche de 1 alors individu proche de l'axe)

Les CTA permettent de répondre à la question : « Qui fait l'axe ? ». Les plus fortes contributions sont les individus qui sont responsables de l'apparition de l'axe