TABLE DES MATIÈRES

CHAPITRE 1

Régression orthogonale dans $\mathbb{R}^2$

1.1 Notion d'espace vectoriel euclidien

1.1.1 Espace vectoriel $\mathbb{R}^n$

Soit n un entier strictement positif de et $\mathbb{R}$ le corps des nombres réels.
L'ensemble $\mathbb{R}^n$ des n-uples $(x_1, \cdots, x_n)$ de nombres réels est muni de sa structure usuelle d'**espace vectoriel réel**, définie par les opérations :

$$(x_1, \cdots, x_n) + (x'_1, \cdots, x'_n) = (x_1 + x'_1, \cdots, x_n + x'_n)$$
$$l(x_1, \cdots, x_n) = (lx_1, \cdots, lx_n), l \in \mathbb{R}.$$

Notations
On identifiera un élément $X = (x_1, \cdots, x_n)$ de $\mathbb{R}^n$ avec la matrice

$$X = \begin{pmatrix} x_1 \\ \vdots \\ x_i \\ \vdots \\ x_n \end{pmatrix}$$

à n lignes et 1 colonne.
La transposée de cette matrice est la matrice $X^t = (x_1, \cdots, x_n)$ à 1 ligne et n colonnes.
Les opérations dans $\mathbb{R}^n$ sont alors définies par des opérations sur les matrices :
Addition :

$$\begin{pmatrix} x_1 \\ \vdots \\ x_i \\ \vdots \\ x_n \end{pmatrix} + \begin{pmatrix} x'_1 \\ \vdots \\ x'_i \\ \vdots \\ x'_n \end{pmatrix} = \begin{pmatrix} x_1 + x'_1 \\ \vdots \\ x_i + x'_i \\ \vdots \\ x_n + x'_n \end{pmatrix}$$

$$(x_1, \cdots, x_n) + (x'_1, \cdots, x'_n) = (x_1 + x'_1, \cdots, x_n + x'_n)$$

Multiplication par un scalaire :

$$l \begin{pmatrix} x_1 \\ \vdots \\ x_i \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} lx_1 \\ \vdots \\ lx_i \\ \vdots \\ lx_n \end{pmatrix}$$

$$l(x_1, \cdots, x_n) = (lx_1, \cdots, lx_n)$$

Dans $\mathbb{R}^n$, les n éléments $e_i, i \in \{1, \cdots, n\}$, dont toutes les coordonnées sont nulles, sauf la i^{me} qui vaut 1, forment une base, appelée la **base canonique** de $\mathbb{R}^n$.
Tout élément $X = (x_1, \cdots, x_n)$ de $\mathbb{R}^n$ s'écrit de manière unique sous la forme

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

1.1.2 Produit scalaire dans $\mathbb{R}^n$

Soit F une application de $\mathbb{R}^n \times \mathbb{R}^n$ dans $\mathbb{R}$.
On notera aussi $\langle X|F|Y\rangle$ ou $\langle X|Y\rangle_F$, le nombre réel F(X,Y).

1.1.2.1 Définition

On appelle produit scalaire dans $\mathbb{R}^n$, toute application F de $\mathbb{R}^n \times \mathbb{R}^n$ dans $\mathbb{R}$ qui possède les propriétés suivantes :
a) Bilinéarité

- Linéarité par rapport à la première variable : $F(X + X', Y) = F(X, Y) + F(X', Y)$ et $F(lX, Y) = lF(X, Y)$, quels que soient l dans $\mathbb{R}$, $X, X', Y \in \mathbb{R}^n$; cette propriété s'écrit aussi
$$\langle X + X'|F|Y\rangle = \langle X|F|Y\rangle + \langle X'|F|Y\rangle$$
- Linéarité par rapport à la deuxième variable : $F(X, Y + Y') = F(X, Y) + F(X, Y')$ et $F(X, lY) = lF(X, Y)$, quels que soient l dans $\mathbb{R}$, $X, X', Y \in \mathbb{R}^n$; cette propriété s'écrit aussi
$$\langle X|F|Y + Y'\rangle = \langle X|F|Y\rangle + \langle X|F|Y'\rangle$$

b) Symétrie
F(X,Y)=F(Y,X), quels que soient X et Y dans $\mathbb{R}^n$: $\langle X|F|Y\rangle = \langle Y|F|X\rangle$
c) Positivité
F(X,X) est un nombre réel supérieur ou égal à 0, quel que soit X dans $\mathbb{R}^n$:

$$\langle X|F|X\rangle = 0$$

d) Non dégénérescence
F(X,X)=0 entraine X=$0_{\mathbb{R}^n}$:

$$\langle X|F|X\rangle = 0 \Rightarrow X = 0$$

Autrement dit, le vecteur $0 = (0, \cdots, 0, \cdots, 0)$ de $\mathbb{R}^n$ est l'unique solution de l'équation F(X,X)=0. On dit aussi qu'un produit scalaire sur $\mathbb{R}^n$ est une **forme bilinéaire symétrique positive non dégénérée**. Le mot "forme" fait simplement référence au fait que les valeurs sont des scalaires. Lorsqu'il est muni d'un produit scalaire, $\mathbb{R}^n$ est appelé un **espace vectoriel euclidien**.

1.1.2.2 Exemples

a) Produit scalaire canonique
L'application de $\mathbb{R}^n \times \mathbb{R}^n$ dans $\mathbb{R}$ définie par :

$$((x_1, \cdots, x_n) + (y_1, \cdots, y_n)) \mapsto \langle X|Y\rangle = X^tY = (x_1 \cdots, x_j, \cdots, x_n)\begin{pmatrix} y_1 \\ \vdots \\ y_i \\ \vdots \\ y_n \end{pmatrix} = \sum_{i=1}^{i=n} x_i y_i$$

est un produit scalaire sur $\mathbb{R}^n$ qu'on appelle le **produit scalaire canonique** de $\mathbb{R}^n$.
En effet, les propriétés de bilinéarité, de symétrie, de positivité et de non dégénérescence sont pratiquement évidentes à vérifier.
b) Produit scalaire défini par une matrice diagonale à éléments positifs
Considérons une matrice réelle M à n lignes et n colonnes dont tous les éléments en dehors de la diagonale principale sont nuls ($m_{ij} = 0$, quels que soient les entiers i et j dans $\{1, \cdots, n\}$ avec $i \leq j$)(on dit que M est une **matrice diagonale**) et dont tous les éléments de la diagonale principale sont des **nombres réels strictement positifs** ($m_{ij} = 0$, quels que soient les entiers i et j dans $\{1, \cdots, n\}$).
Alors l'application :

$$(X, Y) \mapsto \langle X|M|Y\rangle = X^tMY = (x_1 \cdots, x_j, \cdots, x_n)M\begin{pmatrix} y_1 \\ \vdots \\ y_i \\ \vdots \\ y_n \end{pmatrix} = m_{ij}x_jy_i = m_{ii}x_iy_i$$

est un produit scalaire sur $\mathbb{R}^n$. La matrice M est appelée la **matrice des poids** (les "'poids"' sont les éléments de la diagonale).

En effet, les propriétés de bilinéarité, de symétrie, de positivité et de non dégénérescence sont pratiquement évidentes à vérifier.

- Le produit scalaire canonique correspond au cas où la matrice M est la matrice unité I_n (tous les éléments de la diagonale sont égaux à 1 et les élements en dehors de la diagonale sont 0) : tous les poids sont égaux à 1.
- Autre exemple : $M = D_{\frac{1}{n}} = \frac{1}{n}I_n$. Tous les poids sont égaux à $\frac{1}{n}$ et la somme des poids vaut 1.

1.1.2.3 Propriétés

a) Matrice d'un produit scalaire

Pour tout produit scalaire F de $\mathbb{R}^n$, on peut écrire :

$$F(X,Y) = F(\sum_{i=1}^{i=n} x_i e_i, \sum_{j=1}^{j=n} y_j e_j) = \sum_{i,j=1}^{n} F(e_i,e_j)x_i y_j = (x_1,\cdots,x_i,\cdots,x_n)M_F \begin{pmatrix} y_1 \\ \vdots \\ y_j \\ \vdots \\ y_n \end{pmatrix}$$

La matrice $M_F = [F(e_i,e_j)]$ s'appelle la **matrice du produit scalaire** F dans la base canonique. Cette matrice est une matrice symétrique : $F(e_i,e_j) = F(e_j,e_i)$.
Les éléments de sa diagonale sont des nombres réels strictements positifs : $F(e_i,e_j) > 0$.
Remarquons que ces propriétés ne sont pas suffisantes : une matrice symétrique dont les éléments de la diagonale sont des nombres réels strictement positifs bilinéaire $((x_1,x_2),(y_1,y_2)) \mapsto (x_1,x_2)\begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}\begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$ qu'elle définit n'est pas un produit scalaire car le "produit scalaire" du vecteur propre (1,-1) pour la valeur propre négative, par lui même, est un nombre réel strictement négatif $(1,-1)\begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}\begin{pmatrix} 1 \\ -1 \end{pmatrix} = -2$
La matrice $\begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}$ n'est donc pas la matrice d'un produit scalaire sur $\mathbb{R}^2$, bien qu'elle soit symétrique et que les éléments de sa diagonale soient strictement positifs.
En réalité, pour qu'une matrice carrée symétrique réelle soit la matrice d'un produit scalaire, il faut et il suffit que toutes ses valauers propres, qui sont toujours des nombres réels, soient strictement positives. Mais ce résultat ne sera démontré, dans sa généralité, que plus tard, en algèbre.

b) Norme d'un vecteur

Si F est un produit scalaire sur $\mathbb{R}^n$, le nombre réel positif $\|X\|_F = \sqrt{\phi(X,X)}$ s'appelle la **F-norme** de X, ou **F-longueur** de X. Quand il n'y a pas de confusion à craindre, on parlera simplement de norme ou de longueur, qu'on notera $\|X\|$ au lieu de $\|X\|_F$.
On dit qu'un vecteur est **normé pour F** si sa F-longueur est 1.
Par exemple, dans $\mathbb{R}^2$ muni du produit scalaire canonique, la longueur de $X = (x_1,x_2)$ est $\|X\|_F = \sqrt{x_1^2 + x_2^2}$ et le vecteur (1,0) est normé.

c) Angle de deux vecteurs

Etant donnés deux vecteurs X et Y de $\mathbb{R}^n$ et un produit scalaire F sur $\mathbb{R}^n$, pour tout nombre réel l, on

a :
$F(X + lY, X + lY) = \|X + lY\|_F \geq 0$
$l^2 F(Y, Y) + l(F(Y, X) + F(X, Y)) + F(X, X) \geq 0$
$l^2 F(Y, Y) + 2lF(X, Y) + F(X, X) \geq 0$
$\|Y\|_F^2 l^2 + 2\langle X|Y\rangle_F l + \|X\|_F^2 \geq 0$
Comme cette relation est vraie pour tout nombre réel l, c'est que le déterminant de ce trinôme du second degré est négatif : $(\langle X|Y\rangle_F)^2 - \|X\|_F^2 \|Y\|_F^2 \leq 0$
$|\langle X|Y\rangle_F| \leq \|X\|_F \|Y\|_F$
Cette inégalité, valable pour tous vecteurs X et Y de $\mathbb{R}^n$ constitue l'**inégalité de Schwarz**.
Si les deux vecteurs X et Y sont différents de 0, leur longueur n'est pas nulle, le produit de leurs longueurs n'est pas nul, le rapport $\dfrac{\langle X|Y\rangle_\phi}{\|X\|_\phi \|Y\|_\phi}$ est compris entre -1 et 1, et il existe donc un angle compris entre 0 et p radians dont le cosinus est égal au rapport $\dfrac{\langle X|Y\rangle_\phi}{\|X\|_\phi \|Y\|_\phi}$.
Par définition, cet angle unique a compris entre 0 et p, vérifiant :

$$cos\ a = \frac{\langle X|Y\rangle_\phi}{\|X\|_\phi \|Y\|_\phi} = \frac{\langle X|\phi|Y\rangle}{\|X\|_\phi \|Y\|_\phi}$$

est appelé l'**angle des deux vecteurs non nuls** X et Y.

d) Orthogonalité

Etant donnés deux vecteurs X et Y de $\mathbb{R}^n$ et un produit scalaire F sur $\mathbb{R}^n$, on dit que X et Y sont **F-orthogonaux** (ou simplement "orthogonaux" s'il n'y a pas de confusion à craindre) si, et seulement si, leur produit scalaire est nul :

$$F(X, Y) = \langle X|Y\rangle_F = 0$$

Exemples :

- 0 est F-orthogonal à tout vecteur de $\mathbb{R}^n$.
- L'angle de deux vecteurs non nuls F-orthogonaux est $\dfrac{\pi}{2}$
- La base canonique de $\mathbb{R}^n$ muni du produit scalaire canonique est formée de vecteurs normés orthogonaux deux à deux : on parle alors de **base orthonormée**.

e) Projeté orthogonal

Soient X et Y deux vecteurs non nuls de $\mathbb{R}^n$ et F un produit scalaire sur $\mathbb{R}^n$.
Il existe un unique vecteur Z de $\mathbb{R}^n$, proportionnel à Y tel que X-Z soit orthogonal à Y.

Preuve

Pour tout vecteur Z, on peut écrire :

$$\langle X - Z|Y\rangle_F = \langle X|Y\rangle_F - \langle Z|Y\rangle_F$$

Si l'on prend un Z proportionnel à Y, on a Z=aY, donc :

$$\langle X - Z|Y\rangle_F = \langle X|Y\rangle_F - a\langle Y|Y\rangle_F = \langle X|Y\rangle_F - a\|Y\|_F^2$$

Pour que X-Z soit orthogonal à Y, soit $\langle X - Z|Y\rangle_F = 0$, il faut et il suffit que l'on prenne $a = \dfrac{\langle X|Y\rangle_\phi}{\|Y\|_\phi^2}$

f) Projeté orthogonal

Le projeté orthogonal Z_0 de X sur Y est le vecteur Z de $\mathbb{R}^n$ proportionnel à Y, qui minimise $\|X-Z\|_F^2$.

Preuve

Soit Z un vecteur proportionnel à Y

Soit $Z_0 = \dfrac{\langle X|Y\rangle_\phi}{\|Y\|_\phi^2} Y$ le projeté orthogonal de X sur Y.

$$\|X - Z\|_F^2 = \|X - Z_0 + Z_0 - Z\|_F^2$$

Comme Z est proportionnel à Y et que Z_0 est proportionnel à Y, la différence $Z_0 - Z$ est proportionnelle à Y. Or $X - Z_0$ est orthogonal à Y, donc $X - Z_0$ est orthogonal à $Z_0 - Z$ qui est proportionnel à Y. Il en résulte que l'on a :

$$\|X - Z\|_F^2 = \|X - Z_0 + Z_0 - Z\|_F^2 = \|X - Z_0\|_F^2 + \|Z_0 - Z\|_F^2 \geq \|X - Z_0\|_F^2$$

Et cette inégalité montre que $\|X - Z\|_F^2$ atteint son minimum lorsque $Z = Z_0$.

1.1.3 Approche euclidienne de ma regression

Considérons une variable statistique quantitative bidimensionnelle (X,Y) à valeurs dans $\mathbb{R}^2$, définie dans une population Ω de taille n. Elle est définie par l'ensembme des couples $\{(X(w), Y(w))\}_{w\in\Omega}$. $\mathbb{R}^2$ est l'**espace des variables**.

La variable statistique est représentée par un nuage de points dans $\mathbb{R}^2$ et chaque point du nuage statistique représente un individu dans la population Ω.

1.1.3.1 Espace des variables

Les n valeurs X(w) de X pour les n individus de la population peuvent être considérées comme les coordonnées d'un vecteur de $\mathbb{R}^n$. Ce vecteur est noté encore $X = \begin{pmatrix} x_1 \\ \vdots \\ x_i \\ \vdots \\ x_n \end{pmatrix}$

Les n valeurs Y(w) de Y pour les n individus de la population peuvent être considérées comme les coordonnées d'un vecteur de $\mathbb{R}^n$. Ce vecteur est noté encore $Y = \begin{pmatrix} y_1 \\ \vdots \\ y_i \\ \vdots \\ y_n \end{pmatrix}$

L'espace $E = \mathbb{R}^n$ apparait alors comme l'**espace des variables**.

Chaque élément de E peut être considéré comme les valeurs d'une variable statistique quantitative réelle définie sur Ω.

1.1.3.2 Produit scalaire

Dans cet espace des variables, la matrice $D_{\frac{1}{n}} = \frac{1}{n} I_n$, où I_n est la matrice unité à n lignes et n colonnes, définit un **produit scalaire** :

$$\langle X|Y\rangle_{D_{\frac{1}{n}}} = \langle X|D_{\frac{1}{n}}|Y\rangle = \sum_{i=1}^{i=n} \frac{1}{n} x_i y_i = \frac{1}{n} \sum_{i=1}^{i=n} x_i y_i = \frac{1}{n} \langle X|Y\rangle$$

en notant $\langle X|Y\rangle$ le produit scalaire canonique de $\mathbb{R}^n$. On note $1_n = \begin{pmatrix} 1 \\ \vdots \\ 1 \\ \vdots \\ 1 \end{pmatrix}$ le vecteur dont toutes les coordonnées sont égales à 1. On l'appelle le **vecteur unité** de $\mathbb{R}^n$.
On remarquera que ce vecteur unité est normé, sa longueur est $\|1_n\|_{D_{\frac{1}{n}}} = \frac{1}{n} \sum_{i=1}^{i=n} 1 \times 1 = \frac{1}{n} \times n = 1.$

1.1.3.3 Moyenne d'une variable statistique

la moyenne $\overline{X}$ de la variable statistique X est donné par :

$$\overline{X} = \frac{1}{n} \sum_{w} X(w) = \frac{1}{n} \sum_{i=1}^{i=n} x_i = \frac{1}{n} \sum_{i=1}^{i=n} x_i \times 1 = \langle X|D_{\frac{1}{n}}|1_n\rangle = \langle X|1_n\rangle D_{\frac{1}{n}}$$

La moyenne de X est le produit scalaire de X par le vecteur unité 1_n.
Notons X_0 la variable centrée correspondant à X : pour chaque individu w de la population, sa valeur est $X(w) - \overline{X}$:

$$X_0 = \begin{pmatrix} x_1 - \overline{X} \\ \vdots \\ x_i - \overline{X} \\ \vdots \\ x_n - \overline{X} \end{pmatrix} = \begin{pmatrix} x_1 \\ \vdots \\ x_i \\ \vdots \\ x_n \end{pmatrix} - \overline{X} \begin{pmatrix} 1 \\ \vdots \\ 1 \\ \vdots \\ 1 \end{pmatrix} = X - \overline{X} 1_n$$

$$X = X_0 + \overline{X} 1_n = X_0 + \langle X|1_n\rangle D_{\frac{1}{n}} 1_n$$

1.1.3.4 Variance d'une variable statistique

$$s^2(X) = \overline{X_0^2} = \frac{1}{n} \sum_{i=1}^{i=n} (x_i - \overline{X})^2 = \langle X_0|D_{\frac{1}{n}}|X_0\rangle = \|X_0\|^2$$
$$s^2(X) = \|X_0\|^2$$

La variance de X est le carré de la norme de la variable centrée.

1.1.3.5 Covariance

La covariance de deux variables quantitatives réelles X et Y définies sur Ω est la moyenne du produit des variables centrées :

$$Cov(X,Y) = \tfrac{1}{n}\sum_{i=1}^{i=n}(x_i - \overline{X})(y_i - \overline{Y}) = \langle X_0|D_{\frac{1}{n}}|Y_0\rangle = \langle X_0|Y_0\rangle D_{\frac{1}{n}}$$
$$Cov(X,Y) = \langle X_0|D_{\frac{1}{n}}|Y_0\rangle\langle X_0|Y_0\rangle D_{\frac{1}{n}}$$

La covariance est le produit scalare des variables centrées.

1.1.3.6 Coefficient de corrélation linéaire

$$r_{XY} = \frac{Cov(X,Y)}{s(X)s(Y)} = \frac{\langle X_0|D_{\frac{1}{n}}|Y_0\rangle}{\|X_0\|_\phi\|Y_0\|_\phi} = cos(X_0,Y_0)$$
$$r_{XY} = cos(X_0,Y_0)$$

Le coefficient de corrélation linéaire est le cosinus de l'angle des variables centrées.

1.1.3.7 Prédicteur linéaire

Soient Y la variable à expliquer, X la variable explicative, $X0$ et Y_0 les variables centrées.
Le prédicteur linéaire $D_{Y|X}$ est $y^* = a + bx$ ou $y^* - \overline{Y} = b(x - \overline{X})$, soit $y_0^* = bx_0$.
Il est représenté par la **droite de régression** de Y en X dans l'espace des individus.
Le coefficient b s'obtient par $b = \dfrac{Cov(X,Y)}{s^2(X)} = \dfrac{Cov(X,Y)}{s^2(X_0)} = \dfrac{\langle X_0|Y_0\rangle_{D_{\frac{1}{n}}}}{\|X_0\|^2_{D_{\frac{1}{n}}}}$.

D'après ce qui précède (I.1.2.3.e), $bX_0 = \dfrac{\langle X_0|Y_0\rangle_{D_{\frac{1}{n}}}}{\|X_0\|^2_{D_{\frac{1}{n}}}}X_0$ est le projeté orthogonal de Y_0 sur X_0, $Y_0 - bX_0$ est orthogonal à X_0 et b est la valeur qui minimise l'expression

$$S^2 = \frac{1}{n}\sum_{i=1}^{i=n}(Y_{0i} - bX_{0i})^2 = \|Y_0 - bX_0\|^2_{D_{\frac{1}{n}}} = s^2(Y - bX) = s^2(Y - a - bX) = s^2(Y - Y^*) = s^2(Y_0 - Y_0^*)$$

Le prédicteur linéaire de la variable centrée Y_0 est le projeté orthogonal de Y_0 sur X_0 dans $\mathbb{R}^n$. C'est la variable Y_0^* qui minimise la variance de $Y_0 - Y_0^*$.
Nous avons alors :

$$s^2(Y) = \|Y_0\|^2_{D_{\frac{1}{n}}} = \|Y_0 - bX_0 + bX_0\|^2_{D_{\frac{1}{n}}} = \|Y_0 - bX_0\|^2_{D_{\frac{1}{n}}} + \|bX_0\|^2_{D_{\frac{1}{n}}}$$

$$s^2(Y) = S^2_{min} + b^2\|X_0\|^2_{D_{\frac{1}{n}}} = S^2_{min} + (\frac{Cov(X,Y)}{s^2(X)})s^2(X) = S^2_{min} + (\frac{Cov(X,Y)}{s(X)s(Y)})s^2(Y)$$

$$s^2(Y) = S^2_{min} + r^2_{XY}s^2(X).$$

Nous retrouvons la variance résiduelle S^2_{min} et la variance expliquée par la régression $r^2_{XY}s^2(X)$.
De facon symétrique, si X est la variable explicative et Y la variable à expliquer, nous aurons une expression :

$$s^2(X) = S^2_{min} + r^2_{XY}s^2(X).$$

avec la variance résiduelle S'^2_{min} et la variance expliquée par la régression $r^2_{XY}s^2(X)$.

CHAPITRE 2

Analyse en Composantes Principales (ACP)

L'étude d'une population statistique de taille n passe le plus souvent par le recueil d'un nombre élevé p de données quantitatives par élément observé. L'analyse de ces données doit tenir compte de leur caractère multidimensionnel et révéler les liaisons existantes entre les comoposantes.
L'analyse en composantes principales (ACP), introduite en 1901 par K. Pearson et développée par H. Hotelling en 1933, est une méthode très puissante pour explorer la structure de telles données. Chaque donnée étant représentée dans un espase à p dimensions, l'ensemble des donées forment un «nuage de n points» dans $\mathbb{R}^p$. Le principe de l'ACP est d'obtenir un représentation approchée du nuage dans un sous-espace de dimension faible k par projections sur des axes bien choisis. Une métrique dans $\mathbb{R}^p$ étant choisie (een général normalisée par l'utilisation de variables centrées réduites), les k axes principaux sont ceux qui maximisent l'«inertie» du nuage projeté, c'est-à-dire la moyenne pondérée des carrés des distances des points projetés à leur centre de gravité. Les composantes principales sont les n vecteurs ayant pour coordonnées celles des projections orthogonales des n éléments du nuage sur les k axes principaux.
L'ACP construit ainsi de nouvelles variables, artificielles, et des représenatations graphiques permettant de visualiser les relations entre variables, ainsi que l'existence éventuelle de groupes d'éléments et de groupes de variables.

2.1 Regression orthogonales. Axe principal

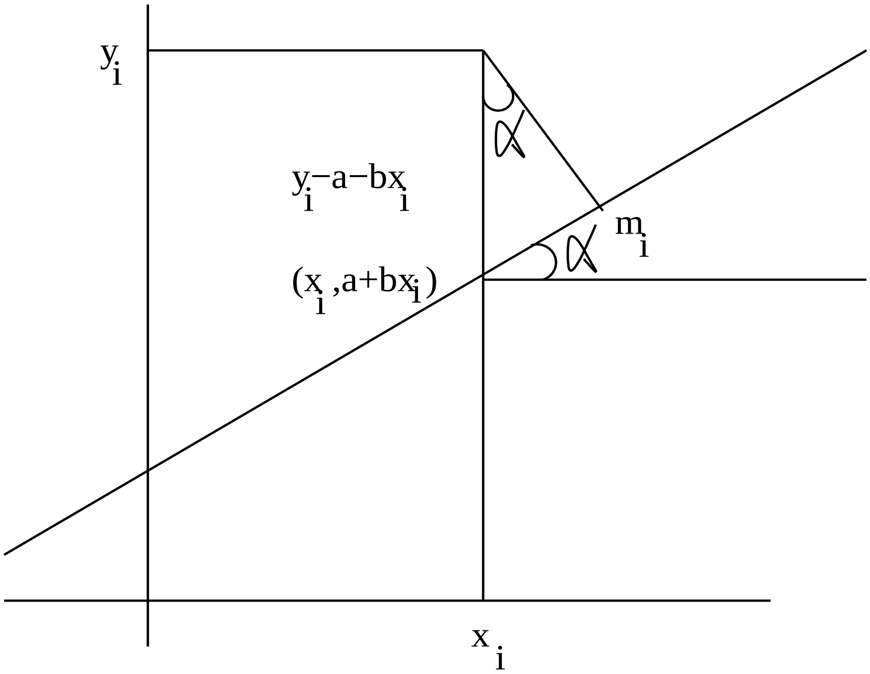
Soit $\mathbb{R}^2$ l'espace des individus, muni du produit scalaire canonique et de la base canonique $\{e_1, e_2\}$ qui, on l'a vu, est orthonormée pour ce produit scalaire.
Si aucune des variables statistiques, X ou Y ne peut s'interpréter par rapport à l'autre, il n'y a pas de raison de privilégier la régression linéaire de Y par rapport à X ou la régression linéaire de X par rapport à Y.
Nous sommes alors conduits à un autre point de vue, celui de la **réduction des données**.

2.1.1 Introduction

Nous cherchons alors dans $\mathbb{R}^2$ une droite (D) qui minimise la somme S^2 des carrés des distances des points du nuage de points à la droite.
La solution est donnée par la **droite de régression orthogonale**.

2.1.1.1 Calcul du terme constant a

L'équation de la droite de régression orthogonale est de la forme : $y = a + bx$.



b est la tangente de l'angle de la droite avec l'axe des abscisses : $b = \tan\alpha$.
$\|\overrightarrow{M_i m_i}\|^2 = \cos^2\alpha(y_i - a - bx_i)^2 = \dfrac{1}{1+b^2}(y_i - a - bx_i)^2$
En introduisant le point moyen $(\overline{X}, \overline{Y})$, on peut écrire :

$$\frac{1}{n}\sum_{i=1}^{n}\|M_i m_i\|^2 = \frac{1}{1+b^2}\times\frac{1}{n}\sum_{i=1}^{n}(y_i - \overline{Y} - b(x_i - \overline{X}) + (\overline{Y} - a - b\overline{X}))^2$$

$$= \frac{1}{1+b^2}\times\frac{1}{n}\sum_{i=1}^{n}(y_i - \overline{Y} - b(x_i - \overline{X}))^2 + \frac{1}{1+b^2}(\overline{Y} - a - b\overline{X})^2$$

$$+ 2\frac{1}{1+b^2}\times(\overline{Y} - a - b\overline{X})\frac{1}{n}\sum_{i=1}^{n}(y_i - \overline{Y} - b(x_i - \overline{X}))$$

Les relations $\overline{Y} = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} y_i$ et $\overline{X} = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} x_i$ entraînent que le dernier terme de la somme est nul.
Il reste :

$$\frac{1}{n}\sum_{i=1}^{n}\|\overrightarrow{M_i m_i}\|^2 = \frac{1}{1+b^2}\times\frac{1}{n}\sum_{i=1}^{n}(y_i-\overline{Y}-b(x_i-\overline{X}))^2+\frac{1}{1+b^2}(\overline{Y}-a-b\overline{X})^2$$

Quel que soit la valeur de b, cette somme sera la plus petite possible lorsque le deuxième terme est nul : $\overline{Y} = a + b\overline{X}$.

Ce résultat signifie que le **point moyen est sur la droite de régression orthogonale** et que, lorsque b est connu, le terme constant a est donné par :

$$a = \overline{Y} - b\overline{X}$$

Puisque le point moyen $G = (\overline{X}, \overline{Y})$ est sur la droite de régression orthogonale, nous le prendrons comme **origine** dans $\mathbb{R}^2$.

La droite de régression orthogonale a une équation de la forme :

$$y_0 = bx_0$$

avec $y_0 = y - \overline{Y}$ et $x_0 = x - \overline{X}$

2.1.1.2 Analyse en composantes principales (ACP)

En fait, la forme de la relation précédente fait disparaître la symétrie initiale entre les rôles de X et Y : ce n'est pas sous cette forme que nous exprimerons l'équation de la droite (D) de régression orthogonale.

Etant donnée une droite (D) passant par l'origine G, on considère plutôt le vecteur unitaire de $\mathbb{R}^2$ orthogonal à la droite (D) :

$$u_1 = \begin{pmatrix}\alpha\\ \beta\end{pmatrix} \text{avec } \alpha^2+\beta^2 = 1$$

Le vecteur unitaire u porté par la droite (D) est :

$$u = \begin{pmatrix}\beta\\ -\alpha\end{pmatrix}$$

La droite (D) est l'ensemble des points $M = (x, y)$ vérifiant $\langle u_1|\overrightarrow{GM}\rangle = 0$,
soit $\alpha x_0 + \beta y_0 = 0$.

Etant donné un point M_i du nuage de point et sa projection orthogonale m_i sur la droite (D), le vecteur $\overrightarrow{Gm_i}$ est le projeté orthogonal de $\overrightarrow{GM_i}$ sur le vecteur u :

$\overrightarrow{Gm_i} = \langle\overrightarrow{GM_i}|u\rangle u = (\beta x_{i0} - \alpha y_{i0})\begin{pmatrix}\beta\\ -\alpha\end{pmatrix}$

$\overrightarrow{m_iM_i} = \overrightarrow{GM_i} - \overrightarrow{Gm_i} = \begin{pmatrix}x_{i0}\\ y_{i0}\end{pmatrix} - (\beta x_{i0} - \alpha y_{i0})\begin{pmatrix}\beta\\ -\alpha\end{pmatrix} = \begin{pmatrix}(1-\beta^2)x_{i0}+\alpha\beta y_{i0}\\ \alpha\beta x_{i0}+(1-\alpha^2)y_{i0}\end{pmatrix} = \begin{pmatrix}\alpha^2 x_{i0}+\alpha\beta y_{i0}\\ \alpha\beta x_{i0}+\beta^2 y_{i0}\end{pmatrix} = (\alpha x_{i0} + \beta y_{i0})\begin{pmatrix}\alpha\\ \beta\end{pmatrix}$

$$\|\overrightarrow{m_iM_i}\|^2 = (\alpha x_{i0} + \beta y_{i0})^2 (\alpha \quad \beta) \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = (\alpha x_{i0} + \beta y_{i0})^2 (\alpha^2 + \beta^2) = (\alpha x_{i0} + \beta y_{i0})^2$$

$$\frac{1}{n}\sum_{i=1}^{n} \|\overrightarrow{m_iM_i}\|^2 = \frac{1}{n}\sum_{i=1}^{n} (\alpha x_{i0} + \beta y_{i0})^2 = \langle \alpha X_0 + \beta Y_0 | D_{\frac{1}{n}} | \alpha X_0 + \beta Y_0 \rangle = \|\alpha X_0 + \beta Y_0\|^2_{D_{\frac{1}{n}}}$$

La recherche de la droite de régression orthogonale se ramène donc à une question que l'on peut envoisager d'un double point de vue :

- soit rechercher, dans l'espace des individus $\mathbb{R}^2$, un vecteur unitaire $u_1 = \begin{pmatrix} \alpha \\ \beta \end{pmatrix}$ avec $\alpha^2 + \beta^2 = 1$ qui minimise la somme :

$$S^2 = \frac{1}{n}\sum_{i=1}^{n} \|\overrightarrow{m_iM_i}\|^2 = \frac{1}{n}\sum_{i=1}^{n} (\alpha x_{i0} + \beta y_{i0})^2,$$

- soit rechercher, dans l'espace des variables $\mathbb{R}^n$, un vecteur $\alpha X_0 + \beta Y_0$, combinaison linéaire fictive des deux variables centrées X_0 et Y_0, avec $\alpha^2 + \beta^2 = 1$, qui minimise $\|\alpha X_0 + \beta Y_0\|^2_{D_{\frac{1}{n}}}$, c'est-à-dire un vecteur de l'hyperplan défini par X_0 et Y_0, de norme minimum pour le produit scalaire défini par la matrice diagonale $D_{\frac{1}{n}}$, sous la contrainte $\alpha^2 + \beta^2 = 1$.

Sous la deuxième forme, la résolution du problème est appelée l'**analyse en composantes principales**.

2.2 Définitions

Appelons Z la matrice $\begin{pmatrix} x_{10} & y_{10} \\ \vdots & \vdots \\ x_{n0} & y_{n0} \end{pmatrix}$ des variables centrées

2.2.1 Inertie totale

On appelle inertie totale du nuage de points de $\mathbb{R}^2$ par rapport à l'origine G des axes, la quantité :

$$I_T = \frac{1}{n}\sum_{i=1}^{n} \|\overrightarrow{GM_i}\|^2 = \frac{1}{n}\sum_{i=1}^{n} (x_{i0}^2 + y_{i0}^2) = s^2(X) + s^2(Y).$$

2.2.2 Inertie statistique

On appelle inertie statistique du nuage de points de $\mathbb{R}^2$ par rapport à une direction Δ de $\mathbb{R}^2$ définie par un vecteur unitaire u, la quantité :

$$I_S(u) = \frac{1}{n}\sum_{i=1}^{n} \|\overrightarrow{Gm_i}\|^2$$

où $\overrightarrow{Gm_i}$ est le projeté orthogonal de $\overrightarrow{GM_i}$ sur u.

Le rapport $\dfrac{I_S(u)}{I_T}$ est le **taux d'inertie totale expliquée par la direction u**.

Par exemple, l'inertie statistique du nuage de points par rapport à l'axe des x est la variance de X et l'inertie statistique du nuage de points par rapport à l'axe des y est la variance de Y.

2.2.3 Inertie mécanique

On appelle inertie mécanique du nuage de points de $\mathbb{R}^2$ par rapport à une direction $Delta$ définie par un vecteur unitaire u, la quantité :

$$I_M(u) = \frac{1}{n}\sum_{i=1}^{n} \|\overrightarrow{M_i m_i}\|^2$$

où $\overrightarrow{Gm_i}$ est le projeté orthogonal de $\overrightarrow{GM_i}$ sur u.
Par exemple, l'inertie mécanique du nuage de points par rapport à l'axe des x est la variance de Y et l'inertie mécanique du nuage de points par rapport à l'axe des y est la variance de X.
Le théorème de Pythagore $\|\overrightarrow{GM_i}\|^2 = \|\overrightarrow{Gm_i}\|^2 + \|\overrightarrow{M_i m_i}\|^2$ entraîne :

$$I_M(u) = I_T - I_S(u)$$

2.2.4 Axes principaux, ou factoriels

On appelle **premer axe factoriel** du nuage de points de $\mathbb{R}^2$, l'axe dont la direction définie par un vecteur unitaire u maximise l'inertie statistique $I_S(u)$.
La direction définie par le vecteur u est appelée la **direction principale** ou **direction factorielle**.
On remarquera que, comme le premier axe factoriel maximise $I_S(u)$, il minimise $I_M(u)$: il donne donc la solution de notre problème, c'est-à-dire la droite de régression orthogonale.

2.2.5 Matrice des variances-covariances

Pour u=$\begin{pmatrix} \beta \\ -\alpha \end{pmatrix}$, l'inertie statistique $I_S(u) = \frac{1}{n}\sum_{i=1}^{n} \|\overrightarrow{Gm_i}\|^2$
s'écrit, avec $\overrightarrow{Gm_i} = \langle \overrightarrow{GM_i} | u \rangle u = (\beta x_{i0} - \alpha y_{i0}) \begin{pmatrix} \beta \\ -\alpha \end{pmatrix}$, sous la forme :
$I_S(u) = \frac{1}{n}\sum_{i=1}^{n}(\beta x_{i0} - \alpha y_{i0})^2 = \beta^2 \times \frac{1}{n}\sum_{i=1}^{n} x_{i0}^2 + \alpha^2 \times \frac{1}{n}\sum_{i=1}^{n} y_{i0}^2 - 2\alpha\beta \times \frac{1}{n}\sum_{i=1}^{n} x_{i0} y_{i0}$
Et comme on sait que :

$\frac{1}{n}\sum_{i=1}^{n} x_{i0}^2 = s^2(X), \frac{1}{n}\sum_{i=1}^{n} y_{i0}^2 = s^2(Y), \frac{1}{n}\sum_{i=1}^{n} x_{i0} y_{i0} = Cov(X,Y),$
l'inertie statistique devient :

$I_S(u) = \beta^2 s^2(X) + \alpha^2 s^2(Y) - 2\alpha\beta Cov(X,Y) = (\beta \quad -\alpha) \begin{pmatrix} s^2(X) & Cov(X,Y) \\ Cov(X,Y) & s^2(Y) \end{pmatrix} \begin{pmatrix} \beta \\ -\alpha \end{pmatrix} =$
$u^t A u$

La matrice A=$\begin{pmatrix} s^2(X) & Cov(X,Y) \\ Cov(X,Y) & s^2(Y) \end{pmatrix} = \frac{1}{n}\begin{pmatrix} x_{10} & \cdots & x_{n0} \\ y_{10} & \cdots & y_{n0} \end{pmatrix}\begin{pmatrix} x_{10} & y_{10} \\ \vdots & \vdots \\ x_{n0} & y_{n0} \end{pmatrix}$ s'appelle la **matrice des variances covariances**.

En introduisant la matrice Z=$\begin{pmatrix} x_{10} & y_{10} \\ \vdots & \vdots \\ x_{n0} & y_{n0} \end{pmatrix}$ des variables centrées, la matrice des variances covariances s'écrit sous les formes :

$$A = \begin{pmatrix} s^2(X) & Cov(X,Y) \\ Cov(X,Y) & s^2(Y) \end{pmatrix} = \frac{1}{n}\begin{pmatrix} x_{10} & \cdots & x_{n0} \\ y_{10} & \cdots & y_{n0} \end{pmatrix}\begin{pmatrix} x_{10} & y_{10} \\ \vdots & \vdots \\ x_{n0} & y_{n0} \end{pmatrix} = \frac{1}{n}Z^tZ = Z^tD_{\frac{1}{n}}Z$$

et l'inertie totale est la trace de cette matrice, somme des éléments diagonaux $s^2(X)$ et $s^2(Y)$:

$$I_T = Tr(A)$$

1re remarque : **valeurs propres**
La matrice des variances-covariances A est, comme on le voit, symétrique réelle.
Une valeur propre de A est un nombre réel λ tel qu'il existe un vecteur v non nul vérifiant $Av = \lambda v$.
Les valeurs propres de A sont donc les nombres réels λ tels que le noyau de l'endomorphisme (application linéaire de $\mathbb{R}^2$ dans $\mathbb{R}^2$) défini par la matrice $A - \lambda I_2$ ne soit pas réduit à 0.
Dire que le noyau n'est pas réduit à 0, c'est-à-dire que l'application lin'éaire n'est pas injective, donc qu'elle n'est pas bijective (puisque dans $\mathbb{R}^2$, injective=bijective) : pour cela, il faut et il suffit que son déterminant soit nul.
Les valeurs propres sont donc les solutions de l'équation :

$$Det(A - \lambda I_2) = \lambda^2 - (s^2(X) + s^2(Y))\lambda + s^2(X)s^2(Y) - (Cov(X,Y))^2 = 0$$

Le discriminant de cette équation du second degré est :

$$(s^2(X) + s^2(Y))^2 - 4(s^2(X)s^2(Y) - (Cov(X,Y))^2) = (s^2(X) - s^2(Y))^2 + 4(Cov(X,Y))^2 \geq 0$$

La matrice A possède donc, ainsi qu'on l'avait déjà dit pour toute matrice symétrique réelle, deux valeurs propres réelles λ_1 et λ_2 :
- la somme de ces valeurs propres est la **trace** de la matrice, somme des éléments diagonaux de la première diagonale :

$$\lambda_1 + \lambda_2 = s^2(X) + s^2(Y) \geq 0$$

-le produit de ces valeurs propres est le **déterminant** de la matrice :

$$\lambda_1\lambda_2 = s^2(X)s^2(Y) - (Cov(X,Y))^2 \geq 0 \quad \text{(d'après l'inégalité de Schwarz)}$$

Les deux valeurs propres de la matrice des variances-covariances sont donc des nombres réels positifs : il est très improbable que l'une soit nulle (il faudrait, pour cela, que le coefficient de corrélation

linéaire soit rigoureusement égal à 1, en valeur absolue, ce qui ne saurait se produire que si X et Y sont déduits l'un de l'autre par une relation linéaire, ou si X et Y sont constantes. Il est très improbabable aussi aussi que les deux valeurs propres soient égales : il faudrait pour cela que la covariance de X et Y soit strictement égale à 0 et que les variances de X et Y soient strictement égales, ce qui ne se produit jamais en pratique.)
Dans le cas général, on peut donc appeler λ_1 et λ_2 les **les valeurs propres de la matrice des variances-covariances**, rangés par ordre décroissant.

$$\lambda_1 > \lambda_2 > 0$$
$$\lambda_1 = \frac{1}{2}\left(s^2(X) + s^2(Y) + \sqrt{(s^2(X) - s^2(Y))^2 + 4(Cov(X,Y))^2}\right)$$
$$\lambda_2 = \frac{1}{2}\left(s^2(X) + s^2(Y) - \sqrt{(s^2(X) - s^2(Y))^2 + 4(Cov(X,Y))^2}\right)$$

2^{me} remarque : vecteurs propres

On démontre aussi, en algèbre, que $\mathbb{R}^2$ possède une **base propre orthonormée**, c'est-à-dire une base $\{u_1, u_2\}$, orthonormée pour le produit scalaire canonique, formée de vecteurs propres de la matrice A :

$$Au_1 = \lambda_1 u_1 \text{ et } Au_2 = \lambda_2 u_2,$$

avec

$$\|u_1\|^2 = 1, \|u_2\|^2 = 1, \langle u_1 | u_2 \rangle = 0$$

Ces vecteurs propres peuvent se calculer.
Soit λ une valeur propre. On a :
$\begin{pmatrix} s^2(X) - \lambda & Cov(X,Y) \\ Cov(X,Y) & s^2(Y) - \lambda \end{pmatrix} \begin{pmatrix} s^2(Y) - \lambda \\ -Cov(X,Y) \end{pmatrix} = \begin{pmatrix} (s^2(X) - \lambda)(s^2(Y) - \lambda) - (Cov(X,Y))^2 \\ (s^2(Y) - \lambda)Cov(X,Y) - (s^2(Y) - \lambda)Cov(X,Y) \end{pmatrix} = \begin{pmatrix} Det(A - \lambda I_2) \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} = 0$ donc le vecteur $\begin{pmatrix} s^2(Y) - \lambda \\ -Cov(X,Y) \end{pmatrix}$ est un **vecteur propre pour la valeur propre λ**.
le carré de la norme de ce vecteur pour le produit scalaire est donné par :

$$(s^2(Y) - \lambda - Cov(X,Y)) \begin{pmatrix} s^2(Y) - \lambda \\ -Cov(X,Y) \end{pmatrix} = (s^2(Y) - \lambda)^2 + (Cov(X,Y))^2$$

On peut donc prendre pour vecteur propre relatif à la valeur propre λ, le vecteur :

$$u = \frac{1}{\sqrt{(s^2(Y) - \lambda)^2 + (Cov(X,Y))^2}} \begin{pmatrix} s^2(Y) - \lambda \\ -Cov(X,Y) \end{pmatrix}$$

Le produit scalaire des deux vecteurs propres ainsi obtenu est nul, parce que la relation $\lambda_1 + \lambda_2 = s^2(X) + s^2(Y)$ entraîne :
$\begin{pmatrix} s^2(Y) - \lambda_1 & -Cov(X,Y) \end{pmatrix} \begin{pmatrix} s^2(Y) - \lambda_2 \\ -Cov(X,Y) \end{pmatrix} = \begin{pmatrix} \lambda_2 - s^2(X) & -Cov(X,Y) \end{pmatrix} \begin{pmatrix} s^2(Y) - \lambda_2 \\ -Cov(X,Y) \end{pmatrix} =$

$Det(A - \lambda_2 I_2) = 0$

Les deux vecteurs $\begin{pmatrix} s^2(Y)-\lambda_1 \\ -Cov(X,Y) \end{pmatrix}$ et $\begin{pmatrix} s^2(Y)-\lambda_2 \\ -Cov(X,Y) \end{pmatrix}$ forment une base de $\mathbb{R}^2$ parce que le déterminant de leurs coordonnées n'est pas nul :

$$-Cov(X,Y)\times(s^2(Y)-\lambda_1)+Cov(X,Y)\times(s^2(Y)-\lambda_2)=Cov(X,Y)\times(\lambda_1-\lambda_2)\neq 0$$

de sorte que les deux vecteurs ne sont pas proportionnels.
Les deux vecteurs :

$$u_1=\frac{1}{\sqrt{(s^2(Y)-\lambda_1)^2+(Cov(X,Y))^2}}\begin{pmatrix} s^2(Y)-\lambda_1 \\ -Cov(X,Y) \end{pmatrix}$$

$$u_2=\frac{1}{\sqrt{(s^2(Y)-\lambda_2)^2+(Cov(X,Y))^2}}\begin{pmatrix} s^2(Y)-\lambda_2 \\ -Cov(X,Y) \end{pmatrix}$$

forment donc une **base orthonormée** de $\mathbb{R}^2$.
Remarquons que, au lieu de prendre pour vecteur propre pour le valeur propre λ, le vecteur $\begin{pmatrix} s^2(Y)-\lambda \\ -Cov(X,Y) \end{pmatrix}$, on aurait pu prendre aussi le vecteur $\begin{pmatrix} -Cov(X,Y) \\ s^2(X)-\lambda \end{pmatrix}$ qui lui est proportionnel (le déterminant de la matrice de ces vecteurs est le déterminant de la matrice $A-\lambda I_2$).

2.3 Diagonalisation de la matrice des variances-covariances

Soit V= $\begin{pmatrix} \dfrac{s^2(Y)-\lambda_1}{\sqrt{(s^2(Y)-\lambda_1)^2+(Cov(X,Y))^2}} & \dfrac{s^2(Y)-\lambda_2}{\sqrt{(s^2(Y)-\lambda_2)^2+(Cov(X,Y))^2}} \\ \dfrac{-Cov(X,Y)}{\sqrt{(s^2(Y)-\lambda_1)^2+(Cov(X,Y))^2}} & \dfrac{-Cov(X,Y)}{\sqrt{(s^2(Y)-\lambda_2)^2+(Cov(X,Y))^2}} \end{pmatrix}$

la matrice des coordonnées des vecteurs propres u_1 et u_2.

$$Ve_1=u_1, Ve_2=u_2.$$

V donne, par produits, pour image d'une base orthonormée, une base orthonormée : c'est ce qu'on appelle une matrice $<< orthogonale >>$, ce qui veut dire que son inverse est égale à sa transposée :

$$V^{-1}=V^t$$

Pour le vérifier, remarquons que ouisque les bases $\{e_1,e_2\}$ et $\{u_1,u_2\}$ sont orthonormées, les coordonnées des vecteurs s'obtiennent par produits scalaires : :

$$u_1=\langle u_1|e_1\rangle e_1+\langle u_1|e_2\rangle e_2$$

$$u_2=\langle u_2|e_1\rangle e_1+\langle u_2|e_2\rangle e_2$$

de sorte que la matrice V, qui a, pour colonnes les vecteurs u_1 et u_2 dans la base $\{e_1, e_2\}$, est :

$$V = \begin{pmatrix} \langle u_1|e_1\rangle & \langle u_2|e_1\rangle \\ \langle u_1|e_2\rangle & \langle u_2|e_2\rangle \end{pmatrix}$$

et les relations inverses

$$e_1 = \langle e_1|u_1\rangle u_1 + \langle e_1|u_2\rangle u_2$$

$$e_2 = \langle e_2|u_1\rangle u_1 + \langle e_2|u_2\rangle u_2$$

montrent que la matrice inverse de V est la matrice :

$$V^{-1} = \begin{pmatrix} \langle e_1|u_1\rangle & \langle e_2|u_1\rangle \\ \langle e_1|u_2\rangle & \langle e_2|u_2\rangle \end{pmatrix}$$

qui, compte tenu de la symétrie du produit scalaire, est la transposée de V.

$$V^{-1} = \begin{pmatrix} \langle u_1|e_1\rangle & \langle u_1|e_2\rangle \\ \langle u_2|e_1\rangle & \langle u_2|e_2\rangle \end{pmatrix} = V^t$$

Il résulte alors des relations $Ve_1 = u_1$ et $Ve_2 = u_2$, que l'on a :

$$V^t u_1 = V^{-1}u_1 = e_1; \quad V^t u_2 = V^{-1}u_2 = e_2$$

Considérons maintenant la matrice $\Lambda = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}$, matrice diagonale des valeurs propres de A.
A est la matrice, dans la base canonique $\{e_1, e_2\}$, d'un endomorphisme f.
Cet endomorphisme f se réduit à deux homothéties, de rapport λ_1 selon le vecteur u_1, et de rapport λ_2 selon le vecteur u_2.
Λ est donc la matrice, dans la base propre $\{u_1, u_2\}$, de l'endomorphisme f.
La matrice de l'application identique de $\mathbb{R}^2$ muni de la base $\{u_1, u_2\}$ dans $\mathbb{R}^2$ muni de la base $\{e_1, e_2\}$ donne, par produits, pour image du vecteur $u_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ le vecteur

$$u_1 = \frac{1}{\sqrt{(s^2(Y) - \lambda_1)^2 + (Cov(X,Y))^2}} \begin{pmatrix} s^2(Y) - \lambda_1 \\ -Cov(X,Y) \end{pmatrix}$$

et pour image du vecteur $u_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ le vecteur

$$u_2 = \frac{1}{\sqrt{(s^2(Y) - \lambda_2)^2 + (Cov(X,Y))^2}} \begin{pmatrix} s^2(Y) - \lambda_2 \\ -Cov(X,Y) \end{pmatrix}$$

C'est donc la matrice V des vecteurs propres.

$$V = [Id_{\mathbb{R}^2}, \{u_1, u_2\}, \{e_1, e_2\}]$$

Réciproquement, la matrice de l'application identique de $\mathbb{R}^2$ muni de la base $\{e_1, e_2\}$ dans $\mathbb{R}^2$ muni de la base $\{u_1, u_2\}$ donne, par produits, pour image du vecteur $e_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ le vecteur $e_1 = \begin{pmatrix} \dfrac{s^2(Y) - \lambda_1}{\sqrt{(s^2(Y) - \lambda_1)^2 + (Cov(X,Y))^2}} \\ \dfrac{s^2(Y) - \lambda_2}{\sqrt{(s^2(Y) - \lambda_2)^2 + (Cov(X,Y))^2}} \end{pmatrix}$ et pour image du vecteur $e_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$

le vecteur $e_2 = \begin{pmatrix} \dfrac{-Cov(X,Y)}{\sqrt{(s^2(Y) - \lambda_1)^2 + (Cov(X,Y))^2}} \\ \dfrac{-Cov(X,Y)}{\sqrt{(s^2(Y) - \lambda_2)^2 + (Cov(X,Y))^2}} \end{pmatrix}$.

C'est donc la matrice V^t transposée et inverse de la matrice V des vecteurs propres.

$$V^t = [Id_{\mathbb{R}^2}, \{e_1, e_2\}, \{u_1, u_2\}]$$

Le diagramme commutatif suivant :

$$\begin{tikzcd}
R^2, \{e_1, e_2\} \arrow[r, "f", "A"'] & R^2, \{e_1, e_2\} \arrow[d, "Id"', "V"] \\
R^2, \{u_1, u_2\} \arrow[u, "Id", "V^t"'] \arrow[r, "f", "A"'] & R^2, \{u_1, u_2\}
\end{tikzcd}$$

met en évidence la relation $f = Id \circ f \circ Id$.
En termes de produit de matrices, cette relation s'écrit :

$$\Lambda = VA^tV$$

d'où l'on déduit aussitôt

$$V = V^t \Lambda V.$$

On dit qu'on a **diagonalisé le matrice** A.

2.4 Recherche des axes principaux

Pour un vecteur normé u, posons $v = Vu$. On a $v^t = u^t V^t$.

$$\|v\|^2 = v^t v = u^t V^t V u = u^t u = \|u\|^2 = 1$$

Le vecteur v est normé lui aussi. L'inertie statistique par rapport à u s'écrit :

$$I_S(u) = u^t A u = u^t V^t \Lambda V u = v^t \Lambda v$$

Dans $\mathbb{R}^2$ rapporté à la base $\{u_1, u_2\}$, notons $v = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}$

$$I_S(u) = v^t \Lambda v = \begin{pmatrix} v_1 & v_2 \end{pmatrix} \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \lambda_1 v_1^2 + \lambda_2 v_2^2, \text{avec } \ v_1^2 + v_2^2 = 1$$

Le problème de la recherche de la droite de régression orthogonale se ramène maintenant à la résolution du problème suivant :
$\mathbf{maximiser\ \lambda_1 v_1^2 + \lambda_2 v_2^2,\ sous\ la\ contrainte\ v_1^2 + v_2^2 = 1,\ avec\ \lambda_1 > \lambda_2 > 0}$
C'est maintenant un problème facile à résoudre :

$$I_S(u) = \lambda_1 v_1^2 + \lambda_2 v_2^2 = \lambda_1(1 - v_2^2) + \lambda_2 v_2^2 = \lambda_1 - (\lambda_1 - \lambda_2) v_2^2$$

La quantité $\lambda_1 - (\lambda_1 - \lambda_2) v_2^2,$ avec $\lambda_1 > \lambda_2$ atteint sa valeur maximum λ_1 lorsqu'on prend $v_2 = 0,$ donc $|v_1| = 1.$

La direction du premier axe factoriel est donc définie par le vecteur v de coordonnées $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ dans la base $\{u_1, u_2\}$: $v = u_1.$

$$I_S(u_1) = \lambda_1$$

D'où le résultat, qu'on peut énoncer sous forme de **thèorème**.

Théorème 2.1 *La direction du premier axe factoriel est définie par le vecteur propre associé à la plus grande valeur propre de la matrice des variances-covariances.*

Le premier axe factoriel est la droite de régression orthogonale.
Comme corollaire, la direction perpendiculaire au premier axe factoriel définit le **deuxième axe factoriel** : elle est définie par le vecteur propre associé à la plus petite valeur propre de la matrice des variances-covariances.
Le deuxième axe factoriel minimise l'inertie statistique $I_S(u$:
$I_S(u) = \lambda_2$ lorsque $|v_2| = 1,$ donc $v_1 = 0$ et $v = \begin{pmatrix} 0 \\ 1 \end{pmatrix} = u_2$ par exemple (on pourrait prendre aussi, bien sûr, $v = -u_2$, la direction serait la même).

$$I_S(u_2) = \lambda_2$$

Le taux d'inertie totale expliquée par le premier axe factoriel est le rapport

$$\frac{I_S(u_1)}{I_T} = \frac{\lambda_1}{s^2(X) + s^2(Y)} = \frac{\lambda_1}{\lambda_1 + \lambda_2}$$

Le taux d'inertie totale expliquée par le deuxième axe factoriel est le rapport

$$\frac{I_S(u_2)}{I_T} = \frac{\lambda_2}{s^2(X) + s^2(Y)} = \frac{\lambda_2}{\lambda_1 + \lambda_2}$$

La relation $\lambda_1 + \lambda_2 = s^2(X) + s^2(Y)$ (la somme des valeurs propres est la trace de la matrice des variances-covariances) s'écrit :

$$I_S(u_1) + I_S(u_2) = I_T$$

La somme des inerties statistiques par rapport aux deux axes factoriels est l'inertie totale du nuage de points.
Chaque valeur propre de la matrice des vairances-covariances correspond à l'inertie expliquée par l'axe factoriel correspondant.

2.5 Coordonnées factorielles et composantes principales

Dans $\mathbb{R}^2$ rapporté à la base propre orthonormée $\{u_1, u_2\}$ les coordonnées des vecteurs $\overrightarrow{GM_i}$ s'appellent les **coordonnées factorielles**.
Comme la base $\{u_1, u_2\}$ est orthonormée, les coordonnées factorielles s'obtiennent par produit scalaire :

$$\overrightarrow{GM_i} = \langle\overrightarrow{GM_i}|u_1\rangle u_1 + \langle\overrightarrow{GM_i}|u_2\rangle u_2$$

Or la base canonique $\{e_1, e_2\}$ est, elle même, orthonormée et l'on a, par conséquent :

$$\overrightarrow{GM_i} = \langle\overrightarrow{GM_i}|e_1\rangle e_1 + \langle\overrightarrow{GM_i}|e_2\rangle e_2 = x_{i0}e_1 + y_{i0}e_2$$

d'où :

$$\langle\overrightarrow{GM_i}|u_1\rangle = x_{i0}\langle e_1|u_1\rangle + y_{i0}\langle e_2|u_1\rangle$$
$$\langle\overrightarrow{GM_i}|u_2\rangle = x_{i0}\langle e_1|u_2\rangle + y_{i0}\langle e_2|u_2\rangle$$

Les coordonnées factorielles s'obtiennent donc par la formule matricielle :

$$\begin{pmatrix} \langle\overrightarrow{GM_i}|u_1\rangle \\ \langle\overrightarrow{GM_i}|u_2\rangle \end{pmatrix} = \begin{pmatrix} \langle e_1|u_1\rangle & \langle e_2|u_1\rangle \\ \langle e_1|u_2\rangle & \langle e_2|u_2\rangle \end{pmatrix} \begin{pmatrix} x_{i0} \\ y_{i0} \end{pmatrix} = V^t \begin{pmatrix} \langle\overrightarrow{GM_i}|e_1\rangle \\ \langle\overrightarrow{GM_i}|e_2\rangle \end{pmatrix}$$

$$\begin{pmatrix} \langle\overrightarrow{GM_i}|u_1\rangle \\ \langle\overrightarrow{GM_i}|u_2\rangle \end{pmatrix} = V^t \begin{pmatrix} \langle\overrightarrow{GM_i}|e_1\rangle \\ \langle\overrightarrow{GM_i}|e_2\rangle \end{pmatrix} = V^t \begin{pmatrix} x_{i0} \\ y_{i0} \end{pmatrix}$$

La matrice V^t est ce qu'on appelle la **matrice du changement de base**.
Elle donne les nouvelles coordonnées (sur la base $\{u_1, u_2\}$) en fonction des anciennes (sur la base $\{e_1, e_2\}$).
Nous avons vu plus haut que cette matrice est la matrice de l'application identique, de $\mathbb{R}^2$ muni de la base $\{u_1, u_2\}$ dans $\mathbb{R}^2$ muni de la base $\{e_1, e_2\}$.
Les relations

$$\begin{pmatrix} \langle\overrightarrow{GM_i}|u_1\rangle & \langle\overrightarrow{GM_i}|u_2\rangle \end{pmatrix} = \begin{pmatrix} \langle\overrightarrow{GM_i}|u_1\rangle \\ \langle\overrightarrow{GM_i}|u_2\rangle \end{pmatrix}^t = \left(V^t \begin{pmatrix} x_{i0} \\ y_{i0} \end{pmatrix}\right)^t = \begin{pmatrix} x_{i0} & y_{i0} \end{pmatrix} V, \text{ pour } i \in \{1, \cdots, n\}$$

peuvent se condenser en une seule formule matricielle :

$$L = ZV$$

formule dans laquelle :

$$L = \begin{pmatrix} \langle \overrightarrow{GM_1} | u_1 \rangle & \langle \overrightarrow{GM_1} | u_2 \rangle \\ \vdots & \vdots \\ \langle \overrightarrow{GM_n} | u_1 \rangle & \langle \overrightarrow{GM_n} | u_2 \rangle \end{pmatrix}$$

est la matrice, à n ligne et 2 colonnes, dont les lignes sont les cordonnées factorielles du nuage de points dans $\mathbb{R}^2$ muni de la base $\{u_1, u_2\}$,

$$Z = \begin{pmatrix} x_{10} & y_{10} \\ \vdots & \vdots \\ x_{n0} & y_{n0} \end{pmatrix}$$

est la matrice, à n lignes et 2 colonnes, dont les colonnes sont les variables centrées
$X - \overline{X}$ et $Y - \overline{Y}$,

$$V = \begin{pmatrix} \frac{s^2(Y) - \lambda_1}{\sqrt{(s^2(Y) - \lambda_1)^2 + (Cov(X,Y))^2}} & \frac{s^2(Y) - \lambda_2}{\sqrt{(s^2(Y) - \lambda_2)^2 + (Cov(X,Y))^2}} \\ \frac{-Cov(X,Y)}{\sqrt{(s^2(Y) - \lambda_1)^2 + (Cov(X,Y))^2}} & \frac{-Cov(X,Y)}{\sqrt{(s^2(Y) - \lambda_2)^2 + (Cov(X,Y))^2}} \end{pmatrix}$$

est la matrice des coordonnées des vecteurs propres orthonormés $\{u_1, u_2\}$ de la matrice des variances-covariances, dans la base canonique $\{e_1, e_2\}$.
Les deux colonnes de la matrice L sont des éléments de l'espace des variables $\mathbb{R}^n$: on les appelle les **composantes principales** de la variable statistique (X,Y).
La première colonne de la matrice V est le vecteur propre u_1.
La première colonne de la matrice L=ZV est donc le vecteur propre $L_1 = Zu_1$.
De même, la deuxième colonne de la matrice L est le vecteur $L_2 = Zu_2$.
Les deux composantes principales L_1 et L_2 de la variable statistique (X,Y) s'obtiennent ainsi par les formules :

$$L_1 = \begin{pmatrix} x_{10} & y_{10} \\ \vdots & \vdots \\ x_{n0} & y_{n0} \end{pmatrix} u_1 = \frac{1}{\sqrt{(s^2(Y) - \lambda_1)^2 + (Cov(X,Y))^2}} \begin{pmatrix} x_{10} & y_{10} \\ \vdots & \vdots \\ x_{n0} & y_{n0} \end{pmatrix} \begin{pmatrix} s^2(Y) - \lambda_1 \\ -Cov(X,Y) \end{pmatrix}$$

$$L_2 = \begin{pmatrix} x_{10} & y_{10} \\ \vdots & \vdots \\ x_{n0} & y_{n0} \end{pmatrix} u_2 = \frac{1}{\sqrt{(s^2(Y) - \lambda_2)^2 + (Cov(X,Y))^2}} \begin{pmatrix} x_{10} & y_{10} \\ \vdots & \vdots \\ x_{n0} & y_{n0} \end{pmatrix} \begin{pmatrix} s^2(Y) - \lambda_2 \\ -Cov(X,Y) \end{pmatrix}$$

avec les valeurs propres λ_1 et λ_2 de la matrice

$$A = \frac{1}{n} \begin{pmatrix} x_{10} & \cdots & y_{10} \\ x_{n0} & \cdots & y_{n0} \end{pmatrix} \begin{pmatrix} x_{10} & y_{10} \\ \vdots & \vdots \\ x_{n0} & y_{n0} \end{pmatrix} = \frac{1}{n} Z^t Z = Z^t D_{\frac{1}{n}} Z = \begin{pmatrix} s^2(X) & Cov(X,Y) \\ Cov(X,Y) & s^2(Y) \end{pmatrix}$$

des variances-covariances :

$$\lambda_1 = \frac{1}{2}\left(s^2(X) + s^2(Y) + \sqrt{(s^2(X) - s^2(Y))^2 + 4(Cov(X,Y))^2}\right)$$

$$\lambda_2 = \frac{1}{2}\left(s^2(X) + s^2(Y) - \sqrt{(s^2(X) - s^2(Y))^2 + 4(Cov(X,Y))^2}\right)$$

2.6 Propriétés des composantes principales

2.6.1 Les composantes principales sont centrées.

$$\overline{L_1} = \langle L_1 | D_{\frac{1}{n}} | 1_n \rangle = \frac{1}{n}\langle Zu_1 | 1_n \rangle = \frac{1}{n}(Zu_1)^t 1_n = \frac{1}{n} u_1^t Z^t 1_n$$

$$Z^t 1_n = \begin{pmatrix} x_{10} & \cdots & y_{10} \\ x_{n0} & \cdots & y_{n0} \end{pmatrix} \begin{pmatrix} 1 \\ \vdots \\ 1 \\ \vdots \\ 1 \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^{n} x_{i0} \\ \sum_{i=1}^{n} y_{i0} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

puisque les variables X_0 et Y_0 sont centrées.
Il reste donc :

$$\overline{L_1} = \frac{1}{n} u_1^t \begin{pmatrix} 0 \\ 0 \end{pmatrix} = 0$$

De même

$$\overline{L_2} = \langle L_2 | D_{\frac{1}{n}} | 1_n \rangle = \frac{1}{n}\langle Zu_2 | 1_n \rangle = \frac{1}{n}(Zu_2)^t 1_n = \frac{1}{n} u_2^t Z^t 1_n = \frac{1}{n} u_2^t \begin{pmatrix} 0 \\ 0 \end{pmatrix} = 0$$

2.6.2 La variance d'une composante principale est la valeur propre correspondante

Comme les composantes sont centrées, leur variance est le carré de leur norme pour le produit scalaire défini par $D_{\frac{1}{n}}$:

$$s^2(L_1) = \|L_1\|^2_{D_{\frac{1}{n}}} = \langle L_1 | D_{\frac{1}{n}} | L_1 \rangle = \frac{1}{n} L_1^t L_1 = \frac{1}{n} u_1^t Z^t Z u_1$$

$$\frac{1}{n} Z^t Z = A$$

$$s^2(L_1) = u_1^t A u_1 = u_1^t \lambda_1 u_1 = \lambda_1 \|u_1\|^2 = \lambda_1$$

De même

$$s^2(L_2) = \|L_2\|^2_{D_{\frac{1}{n}}} = \langle L_2 | D_{\frac{1}{n}} | L_2 \rangle = \frac{1}{n} L_2^t L_2 = \frac{1}{n} u_2^t Z^t Z u_2 = u_2^t A u_2 = u_2^t \lambda_2 u_2 = \lambda_2 \|u_2\|^2 = \lambda_2$$

2.6.3 Les composantes principales sont non corrélées

$$Cov(L_1, L_2) = \langle L_1 | D_{\frac{1}{n}} | L_2 \rangle = \frac{1}{n} L_1^t L_2 = \frac{1}{n} u_1^t Z^t Z u_2 = \frac{1}{n} u_1^t A u_2 = \frac{\lambda_2}{n} \langle u_1 | u_2 \rangle = 0$$

puisque les vecteurs u_1 et u_2 sont orthogonaux pour le produit scalaire canonique.

2.6.4 Reconstruction des données

Les points du nuage centré sont définis par les vecteurs :

$$\overrightarrow{GM_i} = x_{i0} e_1 + y_{i0} e_2 = \langle \overrightarrow{GM_i} | u_1 \rangle u_1 + \langle \overrightarrow{GM_i} | u_2 \rangle u_2$$

Les projetés orthogonaux de ces vecteurs sur l'axe principal défini par u_1 sont les vecteurs :

$$\overrightarrow{Gm_i} = \langle \overrightarrow{GM_i} | u_1 \rangle u_1 (\langle u_1 | e_1 \rangle e_1 + \langle u_1 | e_2 \rangle e_2)$$

Les vecteurs $\overrightarrow{Om_i} = \overrightarrow{OG} + \overrightarrow{Gm_i}$ forment ce qu'on appelle l'**approximation de rang 1 du nuage de points** dans $\mathbb{R}^2$.

Les points m_i sont les projections orthogonales des points M_i sur la droite de régression orthogonale. L'équation de la droite de régression orthogonale, sur laquelle se situe l'approximation de rang 1 du nuage de points, peut prendre l'une des formes équivalents :

$$\langle \overrightarrow{GM_i} | u_2 \rangle = 0$$

$$(x - \overline{X})(s^2(Y) - \lambda_2) = (y - \overline{Y}) Cov(X, Y)$$

$$(x - \overline{X})(\lambda_1 - s^2(X)) = (y - \overline{Y}) Cov(X, Y)$$

$$(x - \overline{X}) Cov(X, Y) = (y - \overline{Y})(s^2(Y) - \lambda_1)$$

$$(x - \overline{X}) Cov(X, Y) = (y - \overline{Y})(\lambda_2 - s^2(X))$$

CHAPITRE 3

Analyse Factorielle des Correspondances (AFC)

L'analyse factorielle des correspondances (AFC), ou analyse des correspondances simples, est une méthode exploratoire d'analyse des tableaux de contingence. Elle a été développée essentiellement par J.-P. Benzecri durant la période 1970-1990.
L'AFC peut être considérée comme une ACP particulière dotée de la métrique du $\mathfrak{X}^2$ (Khi-2) qui ne dépend que du profil des colonnes du tableau. L'analyse permet, dans le plan des deux premiers axes factoriels, une représentation simultanée, souvent fort suggestive des ressemblances entre les colonnes ou les lignes du tableau et de la proximité entre lignes et colonnes.
Dans ce type d'Analyse Factorielle, nous allons étudier sur N individus les «liaisons» entre deux variables X et Y. Chaque variable détermine deux partitions de l'ensemble des individus selon les **modalités** déterminées choisies pour chacune d'elles. On note souvent I l'ensemble des modalités de la variable X et J celui des modalités de Y. Le cardinal de I est noté n et celui de J est noté m. Pour chercher les liaisons entre X et Y nous allons croiser les deux partitions pour obtenir un **tableau de contingence** indexé par $I \times J$ (on définit un ordre sur I et J, qui peut être éventuellement arbitraire, afin de pouvoir construire ce tableau). Dans la case associée à la i-ème ligne et à la j-ème colonne on écrit l'effectif des individus ayant la i-ème modalité pour la variable X et la j-ème modalité pour la variable Y, celui-ci est noté k_{ij}.

Tableau de contingence complété par ses marges

X \ Y	$\cdots$	j-ème colonne	$\cdots$	marge
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
i-ème ligne	$\cdots$	k_{ij}	$\cdots$	$k_{i.}$
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
marge	$\cdots$	$k_{.j}$	$\cdots$	N

On pose

$$k_{.j} = \sum_{i=1}^{n} k_{ij} \text{ et } k_{i.} = \sum_{j=1}^{m} k_{ij}$$

$k_{.j}$ est appelé l'**effectif marginal de la j-ème modalité** de Y,
$k_{i.}$ est appelé l'**effectif marginal de la i-ème modalité** de X.
Les éléments du tableau de contingence divisés par l'effectif total N constituent le **tableau des fréquences** où l'on note f_{ij} l'élément générique. Ce tableau permet de définir deux «marges» : une colonne indexée par i d'élément générique $f_{i.} = \sum_{j \in J} f_{ij}$ et une ligne indexée par j d'élément générique $f_{.j} = \sum_{i \in I} f_{ij}$, ce sont les **fréquences marginales**.
La fréquence $f_{i.}$ (respectivement $f_{.j}$) peut être interprétée comme le **poids de la i-ème modalité** de X, on peut noter celui-ci $p_{i.}$ (respectivement le poids $p_{.j}$ de la j-ème modalité de Y).
On obtient ainsi **deux nuages** $\mathfrak{N}_I$ et $\mathfrak{N}_J$ définis de la facon suivante :

Nuage $\mathfrak{N}_I$

Pour chaque $i \in I$ tel que $p_{i.} \neq 0$, on définit un point f^i de $\mathbb{R}^m$ de composantes :

$$f^i = \begin{pmatrix} \dfrac{f_{i1}}{p_{i.}} \\ \dfrac{f_{i2}}{p_{i.}} \\ \cdots \\ \dfrac{f_{ij}}{p_{i.}} \\ \cdots \\ \dfrac{f_{im}}{p_{i.}} \end{pmatrix} = \begin{pmatrix} f_1^i \\ f_2^i \\ \cdots \\ f_j^i \\ \cdots \\ f_m^i \end{pmatrix}$$

Ces composantes sont les **proportions (ou fréquences) conditionnelles** de la i-ème modalité de X individus qui ont la j-ème modalité pour Y.
Les points f^i du nuage $\mathfrak{N}_I$ sont appelés **profils-lignes**.
A chaque point f^i on associe le poids $p_{i.}$. (afin de prendre en compte l'importance de chaque classe).
On obtient ainsi les points pondérés $(f^i, p_{i.})$ du nuage $\mathfrak{N}_I$.

Nuage $\mathfrak{N}_J$

Par symétrie des rôles de I et J, pour tout $j \in J$ tel que $p_{.j} \neq 0$ on définit un nuage $\mathfrak{N}_J$ par les points f_j de $\mathbb{R}^n$ de composantes :

$$f_j = \begin{pmatrix} \dfrac{f_{1j}}{p_{.j}} \\ \dfrac{f_{2j}}{p_{.j}} \\ \cdots \\ \dfrac{f_{ij}}{p_{.j}} \\ \cdots \\ \dfrac{f_{nj}}{p_{.j}} \end{pmatrix} = \begin{pmatrix} f_j^1 \\ f_j^2 \\ \cdots \\ f_j^i \\ \cdots \\ f_j^n \end{pmatrix}$$

Les points f_j du nuage $\mathfrak{N}_J$ sont appelés **profils-colonnes** et chaque point f_j est affecté du poids $p_{.j}$.

Exemple

Prenons un exemple de résultats scolaire : on relève les notes mathématiques et d'anglais d'une classe de sixième, le tableau de données est le suivant :

Numéro	1	2	3	4	5	6	7	8	9	10	11	12	13
math.	9	13	11	10	12	16	18	12	15	18	13	9	17
angl.	9	7	8	10.5	11	12	16.5	9.5	13	16.5	12	3	17

14	15	16	17	18	19	20	21	22	23	24	25	26	27
13	10	10	15	10	12	10	16	11	10	8	11	14	14
12.5	7.5	6.5	13.5	7.5	9	12	17	10	14.5	7.5	12	8.5	11.5

Nous allons définir des «classes» en mathématiques et en anglais, l'écart type de l'anglais étant plus élevé que celui des maths nous allons définir, par exemple, 5 classes en anglais et 4 en maths :
Classes en maths : $[0; 5[$, $[5; 10[$, $[10; 15[$, $[15; 20]$,
classes en anglais : $[0; 4[$, $[4; 8[$, $[8; 12[$, $[12; 16[$, $[16; 20]$
Le **tableau complet de contingence** (ou tableau complet des effectifs), de terme général k_{ij} est le suivant :

Y : math / X : anglais	1 :CM1 [0 ;5[	2 :CM2 [5 ;10[	3 :CM3 [10 ;15[	4 :CM4 [15 ;20[	marge
1 :CA1 [0 ;4[	0	1	0	0	$k_{1.} = 1$
2 :CA2 [4 ;8[	0	1	4	0	$k_{2.} = 5$
3 :CA3 [8 ;12[	0	1	8	0	$k_{3.} = 9$
4 :CA4 [12 ;16[	0	0	5	3	$k_{4.} = 8$
5 :CA5 [16 ;20[	0	0	0	4	$k_{5.} = 4$
marge	$k_{.1} = 0$	$k_{.2} = 3$	$k_{.3} = 17$	$k_{.4} = 7$	effectif total=27

Remarque : La première classe de math contient 0 individus. Dans ce cas, généralement, on la supprime, nous la gardons ici pour ne pas changer les indices, mais nous ne l'utiliserons pas dans les calculs.
Ce tableau nous montre, par exemple, que 8 élèves ont leur note d'anglais entre 12 (compris) et 16 (non compris) puisque $K_{4.} = 8$.
Le **tableau complet des fréquences**, de terme général $f_{ij} = \dfrac{k_{ij}}{N}$, est :

Y : math / X : anglais	1	2	3	4	fréquence marginale
1	0	$\frac{1}{27}$	0	0	$f_{1.} = \frac{1}{27} = p_{1.}$
2	0	$\frac{1}{27}$	$\frac{4}{27}$	0	$f_{2.} = \frac{5}{27} = p_{2.}$
3	0	$\frac{1}{27}$	$\frac{8}{27}$	0	$f_{3.} = \frac{9}{27} = p_{3.}$
4	0	0	$\frac{5}{27}$	$\frac{3}{27}$	$f_{4.} = \frac{8}{27} = p_{4.}$
5	0	0	0	$\frac{4}{27}$	$f_{5.} = \frac{4}{27} = p_{5.}$
fréquence marginale	$f_{.1} = \frac{0}{27} = p_{.1}$	$f_{.2} = \frac{3}{27} = p_{.2}$	$f_{.3} = \frac{17}{27} = p_{.3}$	$f_{.4} = \frac{7}{27} = p_{.4}$	poids total=1

Dans la suite de ce chapitre nous notons **F** la matrice à 5 lignes et 3 colonnes (nous supprimons la colonne ne comportant que des zéros) représentant le tableau des fréquences.

Définition du nuage $\mathfrak{N}_I$ (profils-lignes)

Le cardinal de I est : $n = 5$ et celui de J est : $m = 3$.

Y : math / X : anglais	f_1^i	f_2^i	f_3^i	f_4^i	poids $p_{i.}$
f^1 (i=1)	•	1	0	0	$p_{1.} = \frac{1}{27}$
f^2 (i=2)	•	$\frac{1}{5}$	$\frac{4}{5}$	0	$p_{2.} = \frac{5}{27}$
f^3 (i=3)	•	$\frac{1}{9}$	$\frac{8}{9}$	0	$p_{3.} = \frac{9}{27}$
f^4 (i=4)	•	0	$\boldsymbol{\frac{5}{8}}$	$\frac{3}{8}$	$p_{4.} = \frac{8}{27}$
f^5 (i=5)	•	0	0	1	$p_{5.} = \frac{4}{27}$

exemple : Nous avons imprimé en gras le nombre $\dfrac{5}{8}$ dans la ligne f^4 : c'est la **proportion conditionnelle** de la 4-ème modalité de l'anglais des individus qui ont la 3-ème modalité en maths (5 élèves ont à la fois un note entre 12 compris et 16 non compris en anglais et une note entre 10 compris et 15 non compris et 15 non compris en math ; 8 élèves ont une note entre 12 compris et 16 non compris en anglais).

$$f_3^4 = \frac{f_{43}}{p_{4.}} = \frac{\frac{5}{27}}{\frac{8}{27}} = \frac{5}{8}, \text{ le poids du point } f^4 \text{ est } p_{4.} = \frac{8}{27}.$$

Remarque : La somme des composantes d'un point quelconque de $\mathfrak{N}_I$ (respectivement de $\mathfrak{N}_J$), étant égale à 1, tous les points de $\mathfrak{N}_I$ (respectivement de $\mathfrak{N}_J$) sont situés dans un hyperplan de $\mathbb{R}^m$ (respectivement de $\mathbb{R}^n$).
Dans notre exemple, où $\mathfrak{N}_I$ peut être représenté dans $\mathbb{R}^3$, les points seront donc dans un plan de $\mathbb{R}^3$.

Définition du nuage $\mathfrak{N}_J$ (profils-colonnes)

	f_1 (j=1)	f_2 (j=2)	f_3 (j=3)	f_4 (j=4)
f_j^1	•	$\frac{1}{1}$	0	0
f_j^2	•	$\frac{1}{3}$	$\frac{4}{17}$	0
f_j^3	•	$\frac{1}{3}$	$\frac{8}{17}$	0
f_j^4	•	0	$\frac{5}{17}$	$\frac{3}{17}$
f_j^5	•	0	0	$\frac{4}{17}$
poids $p_{.j}$	$p_{.1} = 0$	$p_{.2} = \frac{3}{27}$	$p_{.3} = \frac{17}{27}$	$p_{.4} = \frac{7}{27}$

3.1 Métriques et bases

Les éléments du nuage $\mathfrak{N}_I$ sont repérés dans la base canonique de $\mathbb{R}^m$ notée $(e_1, \cdots, e_j, \cdots, e_m)$. On définit une **métrique** sur l'espace vectoriel engendré par les éléments de $\mathfrak{N}_I$ à partir de l'**inverse de la matrice diagonale** D_m (tous les éléments de cette diagonale sont non nuls) de la facon suivante : si b est la forme bilinéaire définie par :

$$\begin{pmatrix} p_{.1} & 0 & 0 & \cdots & \cdots & 0 \\ 0 & p_{.2} & 0 & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & 0 & p_{.j} & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & 0 \\ 0 & \cdots & \cdots & 0 & 0 & p_{.m} \end{pmatrix}$$

$$b(e_j, e_k) = \begin{cases} 0 \text{ si } j \neq k \\ \frac{1}{p_{.j}} \text{ si } j = k \end{cases}$$

La distance de deux profils-lignes quelconques f^r et f^s est définie par
$d_m(f^r, f^s) = b(f^r - f^s, f^r - f^s)$ soit encore :
$d^2_m(f^r, f^s) = \sum_{j\in J} \frac{(f^r_j - f^s_j)^2}{p_{.j}} = \frac{(f^r_1 - f^s_1)^2}{p_{.1}} + \cdots + \frac{(f^r_j - f^s_j)^2}{p_{.j}} + \cdots + \frac{(f^r_m - f^s_m)^2}{p_{.m}}$
Remarque : ce sont donc les poids $p_{.j}$ du nuage $\mathfrak{N}_J$ qui définissent la métrique associée à $\mathfrak{N}_I$. De même, le nuage $\mathfrak{N}_J$ est rapporté à la base canonique de $\mathbb{R}^n$ et sa métrique est définie par les poids de $\mathfrak{N}_I$.
Les métriques ainsi définies sont appelées «métrique du $\mathfrak{X}^2$».
La matrice définissant la métrique n'est donc plus, comme c'est le cas généralement en ACP, le produit d'un scalaire par la matrice unité I_d (de dimension $m \times m$ dans le cas du nuage $\mathfrak{N}_I$).
Ainsi dans notre exemple la matrice des poids définissant la métrique sur $\mathfrak{N}_I$ est :

$$\begin{pmatrix} \frac{3}{27} & 0 & 0 \\ 0 & \frac{17}{27} & 0 \\ 0 & 0 & \frac{7}{27} \end{pmatrix} \text{ rappelons que } p_{.1} = 0$$

Le carré de la distance entre f^2 et f^3 est calculée par :
$d^2_m(f^2, f^3) = \sum_{j\in J} \frac{(f^2_j - f^3_j)^2}{p_{.j}} = \frac{(\frac{1}{5} - \frac{1}{9})^2}{\frac{3}{27}} + \frac{(\frac{4}{5} - \frac{8}{9})^2}{\frac{17}{27}} = 0.0825.$
Rappelons que la métrique de distance d_m, ici sur $\mathbb{R}^3$ car $m = 3$, est donnée par l'inverse de la marge ligne du tableau des fréquences.
Propriété importante de cette métrique : la distance entre deux points f^r et f^s de $\mathfrak{N}_I$ ne change pas si on remplace deux colonnes identiques du tableau des fréquences $\mathfrak{N}_I$ par une colonne égale à la somme des éléments de même ligne et le poids associé par la somme des poids des éléments de même ligne.
En effet, le carré de la distance entre f^r et f^s (pour $r \neq s$) est :
$d_m(f^r, f^s) = b(f^r - f^s, f^r - f^s)$ soit encore :
$d^2_m(f^r, f^s) = \sum_{j\in J} \frac{(f^r_j - f^s_j)^2}{p_{.j}} = \frac{(f^r_1 - f^s_1)^2}{p_{.1}} + \cdots + \frac{(f^r_h - f^s_h)^2}{p_{.h}} + \cdots + \frac{(f^r_k - f^s_k)^2}{p_{.k}} + \cdots + \frac{(f^r_m - f^s_m)^2}{p_{.m}}$ or par hypothèse, $\forall r, \forall s \quad f^r_h = f^s_h$ et
$p_{.h} = p_{.k}$ et $\frac{(f^r_h - f^s_h)^2}{p_{.h}} + \frac{(f^r_k - f^s_k)^2}{p_{.k}} = 2\frac{(f^r_h - f^s_h)^2}{p_{.h}}. \quad (a)$
Si on remplace les colonnes d'indice h et k du tableau des fréquence par une colonne d'élément générique égal à la somme des éléments de même ligne et les poids $p_{.h}$ et $p_{.k}$ par le poids $p_{.h} + p_{.k}$, l'expression (a) devient :

$$\frac{((f^r_h - f^s_h) - (f^r_k - f^s_k))^2}{p_{.h} + p_{.k}} = \frac{(2f^r_h - 2f^s_h)^2}{2p_{.h}} = \frac{2(f^r_h - f^s_h)^2}{p_{.h}},$$

Les autres éléments du calcul de la distance n'étant pas affectés par la transformations celle-ci reste

la même.
Exemple : soit le tableau de contingence suivant concernant deux variables X et Y mesurées sur 10 individus :

X \ Y	1	2	3	$k_{i.}$
1	1	1	0	2
2	2	2	0	4
3	1	1	1	3
4	0	0	1	1
$k_{.j}$	4	4	2	10

Il comporte deux colonnes égales (les colonnes 1 et 2) le tableau des fréquences correspondant est :

X \ Y	1	2	3	fréquence marginale $p_{i.}$
1	0.1	0.1	0	$p_{1.} = 0.2$
2	0.2	0.2	0	$p_{2.} = 0.4$
3	0.1	0.1	0.1	$p_{3.} = 0.3$
4	0	0	0.1	$p_{4.} = 0.1$
fréquence marginale $p_{.j}$	$p_{.1} = 0.4$	$p_{.2} = 0.4$	$p_{.3} = 0.2$	fréquence totale=1

Le tableau des profils-lignes du nuage $\mathfrak{N}_I$ est :

	f_1^i	f_2^i	f_3^i
f^1	$\frac{1}{2}$	$\frac{1}{2}$	0
f^2	$\frac{2}{4}$	$\frac{2}{4}$	0
f^3	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$
f^4	0	0	1
poids $p_{.j}$	$p_{.1} = \frac{4}{10}$	$p_{.2} = \frac{4}{10}$	$p_{.3} = \frac{2}{10}$

Calculons le carré de la distance entre les points f^3 et f^4 du nuage $\mathfrak{N}_I$, on a :

$$d_m^2(f^3, f^4) = \frac{(\frac{1}{3})^2}{\frac{4}{10}} + \frac{(\frac{1}{3})^2}{\frac{4}{10}} + \frac{(\frac{2}{3})^2}{\frac{2}{10}} = \frac{10}{36} + \frac{10}{36} + \frac{40}{18} = \frac{100}{36}.$$

Puisque les lignes 1 et 2 du tableau $\mathfrak{N}_I$ correspondantes aux 2 colonnes 1 et 2 du tableau des profils-lignes sont égales nous pouvons réduire le tableau $\mathfrak{N}_I$ en $\mathfrak{N}'_I$:

	f'^i_1	f'^i_2
f'^1	1	0
f'^2	1	0
f'^3	$\frac{2}{3}$	$\frac{1}{3}$
f'^4	0	1
poids $p'_{.j}$	$p'_{.1} = \frac{8}{10}$	$p'_{.2} = \frac{2}{10}$

On a maintenant :

$d^2_m(f'^3, f'^4) = \frac{(\frac{2}{3})^2}{\frac{8}{10}} + \frac{(\frac{2}{3})^2}{\frac{2}{10}} = \frac{40}{72} + \frac{40}{72} = \frac{100}{36}$.

Remarque : Le rôle des variables X et Y dans l'étude étant symétrique on pourra effectuer les mêmes simplifications sur le nuage $\mathfrak{N}_J$.

Métrique du $\mathfrak{X}^2$ et inertie

Rappelons que nous avons appelé la métrique associée à la distance d_m définie au paragraphe précédent «métrique du $\mathfrak{X}^2$», voyons quelle relations lie le calcul du test du $\mathfrak{X}^2$ (test d'indépendance de 2 caractères) à l'inertie des nuages $\mathfrak{N}_I$ et $\mathfrak{N}_J$ par rapport au barycentre G, calculée avec la métrique $\mathfrak{X}^2$. On définit le nombre $\mathfrak{X}^2$ par :

$$\mathfrak{X}^2 = \sum_{i \in I} \sum_{j \in J} \frac{(Nf_{ij} - Nf_{i.}f_{.j})^2}{Nf_{i.}f_{.j}}$$

où N est l'effectif total des individus sur lesquels sont étudiés les deux caractères ; Nf_{ij} est l'effectif réel des individus appartenant à la fois à la classe d'indice $i \in I$ et à la classe d'indice $j \in J$ et $Nf_{i.}f_{.j}$ est l'effectif théorique des individus dans l'hypothèse où les deux caractères étudiés seraient indépendants (la probabilité de l'événement d'indice ij serait alors le produit des probabilités des événements d'indices respectivement i et j).

Dans le cadre de l'AFC, calculons l'inertie d'un point f^i du nuage $\mathfrak{N}_I$ par rapport au barycentre G de vecteur associé g, pour cela calculons tout d'abord les coordonnées de G.

Calcul des coordonnées du barycentre G du nuage $\mathfrak{N}_I$ dont chacun des points d'indice i est affecté du poids $p_{i.}$.

G est un point de $\mathbb{R}^m$, sa j-ème coordonnée est $\sum_{i \in I} p_{i.} \frac{f_{ij}}{p_{i.}} = \sum_{i \in I} f_{ij} = p.j$

Donc les coordonnées du barycentre du nuage $\mathfrak{N}_I$ sont égales à la marge des poids de $\mathfrak{N}_I$.

Dans notre exemple les coordonnées de G sont donc $(\frac{3}{27}, \frac{17}{27}, \frac{7}{27})$.

Le carré de la distance de f^i à g pour la métrique d_m est :

$$d_m(f^i, g) = \sum_{j \in J} \frac{(\frac{f_{ij}}{p_{i.}} - p_{.j})^2}{p_{.j}} = \sum_{j \in J} \frac{(f_{ij} - p_{.j}p_{i.})^2}{p^2_{i.}p_{.j}}$$

L'inertie du point f^i affecté du poids $p_{i.}$ par rapport à G (ou g) est donc :

$$\sum_{j \in J} \frac{(f_{ij} - p_{.j}p_{i.})^2}{p_{i.}^2 p_{.j}} \times p_{i.} = \sum_{j \in J} \frac{(f_{ij} - p_{.j}p_{i.})^2}{p_{i.}p_{.j}}$$

or, $\forall i \in I, p_{i.} = f_{i.}$, et $\forall j \in J, p_{.j} = f_{.j}$ l'inertie totale du nuage $\mathfrak{N}_I$ par rapport à G est donc : $\sum_{j \in J} \frac{(f_{ij} - f_{.j}f_{i.})^2}{f_{i.}f_{.j}}$, ce qui montre la propriété :

$$\mathfrak{X}^2 = N \times Inertie_g \mathfrak{N}_I$$

et par un calcul analogue sur le nuage $\mathfrak{N}_J$:

$$\mathfrak{X}^2 = N \times Inertie_g \mathfrak{N}_J$$

Analyse factorielle des correspondances AFC), liens avec l'ACP

Nous allons effectuer sur le tableau du nuage $\mathfrak{N}_I$ le même travail qu'en ACP.
Nous recherchons les directions d'inertie maximale du nuage $\mathfrak{N}_I$; cells-ci sont déterminées par les points les plus éloignés (au sens de la métrique d_m) de G et munis des plus grands poids. Le poids attribué à chaque point étant égal à la fréquence de la classe correspondante, une classe de grand effectif intervient fortement dans la détermination des axes.
Reprenons par analogie avec l'ACP, les données relatives au nuage $\mathfrak{N}_I$. Nous disposons :
• d'un tableau de contingence relatif à deux caractères (ou variables) observés sur N individus dont le terme général est noté k_{ij} $\left(\sum_{(i,j) \in I \times J} k_{ij} = N \right)$;
• d'une matrice des fréquence, notée F, de dimmension $n \times m$, d'éléments $f_{ij} = \frac{k_{ij}}{N}$;
$\mathfrak{N}_I$.
• d'une matrice diagonale D_m, d'élément diagonale $f_{.j}$ non nul, c'est la matrice de dimension $m \times m$ définissant la métrique d_m sur l'espace vectoriel engendré $\mathfrak{N}_I$;
• d'une matrice diagonale D_n, d'élément diagonale générique $f_{i.}$ non nul, c'est la matrice de dimension $n \times n$ des poids des éléments de $\mathfrak{N}_I$.
Remarque importante :
Il faut bien prendre garde au fait que, pour définir les profils-lignes de $\mathfrak{N}_I$, nous avons divisé chaque ligne d'indice i du tableau F des fréquence par le poids $f_{i.}$ de la i-ème modalité de X : de même, pour définir les profils-colonnes de $\mathfrak{N}_J$, nous avons divisé chaque colonne d'indice j par le poids $f_{.j}$ de la j-ème modalité de Y.
Les matrices qui interviennent dans le calcul des axes principaux (de facon analogue à celui de l'ACP) sont donc D_n^{-1} et D_m^{-1} et non D_n et D_m, ainsi, la matrice à diagonaliser sera : $A = F^t D_n^{-1} F D_m^{-1}$.
Remarquons que les profils-lignes de $\mathfrak{N}_I$ ont pour coordonnées les n lignes de la matrice $D_n^{-1}F$.
Pratiquement il est difficile de diagonaliser A qui -contrairement aux matrices intervenant en ACP- n'est pas, en général symétrique.

On a donc recours à l'artifice de calcul suivant : nous remarquons tout d'abord que $F^t D_n^{-1} F$ est symétrique, nous allons décomposer la matrice D_m^{-1} en $D_m^{-\frac{1}{2}} D_m^{-\frac{1}{2}}$.
On a alors : $A = F^t D_n^{-1} F D_m^{-\frac{1}{2}} D_m^{-\frac{1}{2}}$.
Le but du calcul est d'obtenir les valeurs propres et les vecteurs propres de A. Soit λ une valeur propre de A et U la matrice colonne du vecteur propre associé à λ.
La relation $AU = \lambda U$ s'écrit $F^t D_n^{-1} F D_m^{-\frac{1}{2}} D_m^{-\frac{1}{2}} U = \lambda U$.
Multiplions cette expression à gauche par $D_m^{-\frac{1}{2}}$, il vient :

$$D_m^{-\frac{1}{2}} AU = \left(D_m^{-\frac{1}{2}} F^t D_n^{-1} F D_m^{-\frac{1}{2}} \right) D_m^{-\frac{1}{2}} U = D_m^{-\frac{1}{2}} \lambda U$$

ce qui montre que $D_m^{-\frac{1}{2}} U$ est la matrice colonne du vecteur propre de $D_m^{-\frac{1}{2}} F^t D_n^{-1} F D_m^{-\frac{1}{2}}$ relativement à la valeur propre λ (l'intérêt de ce calcul est que, D_n^{-1} et $D_m^{-\frac{1}{2}}$ étant diagonales, la matrice $D_m^{-\frac{1}{2}} F^t D_n^{-1} F D_m^{-\frac{1}{2}}$ est symétrique, ses valeurs propres et vecteurs propres se calculent plus rapidement).
Certains auteurs décomposent aussi D_n^{-1} en $D_n^{-\frac{1}{2}} D_n^{-\frac{1}{2}}$ et effectuent le produit de la matrice $D_n^{-\frac{1}{2}} F D_m^{-\frac{1}{2}}$ de terme général $\dfrac{f_{ij}}{\sqrt{f_{i.}}\sqrt{f_{.j}}}$ par sa tranposée ; on a alors :

$$D_m^{-\frac{1}{2}} AU = D_m^{-\frac{1}{2}} F^t D_n^{-\frac{1}{2}} D_n^{-\frac{1}{2}} F D_m^{-\frac{1}{2}} D_m^{-\frac{1}{2}} U.$$

On obtient ainsi, puisque $D_n^{-\frac{1}{2}}$ et $D_m^{-\frac{1}{2}}$ sont leurs transposées, une décomposition analogue à celle de l'ACP.
Pour calculer les valeurs propres et les vecteurs propres de A nous calculons donc ceux de $D_m^{-\frac{1}{2}} F^t D_n^{-1} F D_m^{-\frac{1}{2}} = D_m^{-\frac{1}{2}} A D_m^{-\frac{1}{2}}$ puis nous multiplions les vecteurs propres obetnus par $D_m^{\frac{1}{2}}$ (les valeurs propres sont les mêmes).
Les calculs effectués avec le logiciel de calculs mathématiques Math. Lab. donnent les résultats suivants en écriture décimale au dix-millième (la matrice étudiée étant de faible dimension, nous effectuons les calculs directment, sans utiliser la technique de symétrisation de A) :

$$F = \begin{pmatrix} 0.0370 & 0 & 0 \\ 0.0370 & 0.1481 & 0 \\ 0.0370 & 0.2963 & 0 \\ 0 & 0.1852 & 0.1111 \\ 0 & 0 & 0.1481 \end{pmatrix} ; \; D_n^{-1} = \begin{pmatrix} 27 & 0 & 0 & 0 & 0 \\ 0 & 5.4 & 0 & 0 & 0 \\ 0 & 0 & 3 & 0 & 0 \\ 0 & 0 & 0 & 3.375 & 0 \\ 0 & 0 & 0 & 0 & 6.75 \end{pmatrix} ;$$

$$D_m^{-1} = \begin{pmatrix} 9 & 0 & 0 \\ 0 & 1.5883 & 0 \\ 0 & 0 & 3.8565 \end{pmatrix} ; \; A = F^t D_n^{-1} F D_m^{-1} = \begin{pmatrix} 0.4370 & 0.0993 & 0 \\ 0.5530 & 0.7904 & 0.2679 \\ 0 & 0.1103 & 0.7321 \end{pmatrix}$$

Les valeurs propres de A sont : 0.3033, 0.6562, 1.

La matrice des vecteurs propres de A est : $\begin{pmatrix} -0.5840 & -0.2488 & -0.1610 \\ 0.7862 & -0.5491 & -0.9126 \\ -0.2022 & 0.7979 & -0.3758 \end{pmatrix}$

Nous remarquons qu'il apparait une valeur propre «paraiste» égale à1, qui correspond au dernier vecteur propre, colinéaire à g, nous reviendrons sur ce cas lorsque nous traiterons le cas des nuages $\mathfrak{N}'_I$ centrés par rapport à G.
Cette valeur propre ne sera pas prise en compte dans le calcul de l'inertie autour de G et nous obtenons dans notre exemple les résultats suivants :
L'inertie totale est : 0.3033+0.6562=0.9595 ;
La première valeur propre est 0.6562 ;
Le premier axe factoriel exprime donc 68% de l'inertie ($\frac{0.6562}{0.9595} \times 100 \approx 68$) ;

Le deuxième axe factoriel exprime donc 32% de l'inertie ($\frac{0.3033}{0.9595} \times 100 \approx 32$).
Le premier axe factoriel représente la totalité de l'intertie ce qui était prévisible puisque les profils-lignes de $\mathfrak{N}_I$ sont coplanaires dans $\mathbb{R}^3$ (dans le cas général ils appartiennet à un hyperplan de $\mathbb{R}^n$).
Si nous effectuons le même calcul sur la matrice $D_m^{-\frac{1}{2}} A D_m^{-\frac{1}{2}}$ les valeurs propres données par le logiciel de calcul numérique Math. Lab sont dans cet ordre : 0.3030, 1, 0.6562.
La matrice des vecteurs propres est : $\begin{pmatrix} -0.8540 & -0.3333 & -0.3995 \\ -0.4830 & -0.7935 & -0.3703 \\ 0.1936 & -0.5082 & 0.8386 \end{pmatrix}$
Si l'on effectue sans précaution le produit de cette matrice par $D_m^{\frac{1}{2}}$ on ne trouve pas les vecteurs propres de A.
En effet, il faut remarquer que les valeurs propres ne sont pas fournies dans le même ordre que précédemment. Nous avons, en effet, effectué les calculs avec un logiciel de mathématiques générales, un logiciel de statistiques nous aurait délivré les valeurs propres par ordre décroissant et ordonné les vecteurs propres en conséquence.
Il faut donc réordonner la matrice des vecteurs propres avant de la multiplier par $D_m^{\frac{1}{2}}$. Nous laissons au lectuer le soin d'effectuer la vérification (attention : les vecteurs propres ainsi obtenues ne sont pas obligatoirement égaux aux vecteurs propres de A obtenus par un autre procédé mails ils leur sont colinéaires).

AFC du nuage centré $\mathfrak{N}'_I$

Nous avons remarqué que L'AFC de notre exemple faisait apparaitre une valeur propre égale à 1 que nous n'avons pas pris en compte dans le calcul de l'inertie autour de G, cette valeur propre est inutile, car comme nous le verrons plus loin elle est associée au vecteur g défini par le barycentre G du nuage $\mathfrak{N}_I$, ce vecteur est orthogonal à l'hyperplan contenant les éléments de $\mathfrak{N}_I$, l'inertie par rapport à G, des points projetés sur l'axe défini par ce vecteur est donc nulle.
Pour éliminer ces éléments parasites et rendre l'interprétation plus facile nous allons centrer le nuage $\mathfrak{N}_I$ en prenant comme nouvelle origine le barycentre G.
Rappelons que la matrice colonne des coordonnées de G est égale à la marge des poids de $\mathfrak{N}_J$. Les nouvelles coordonnées f'^i d'un profil-ligne du nuage centre $\mathfrak{N}'_I$ sont donc :

$$f'^i = \left(\frac{f_{i1}}{p_{i.}} - p_{.1}, \frac{f_{i2}}{p_{i.}} - p_{.2}, \cdots, \frac{f_{im}}{p_{i.}} - p_{.m}\right)$$

Dans notre exemple la matrice des poids de $\mathfrak{N}_J$ est :

$$\begin{pmatrix} 0.1111 & 0 & 0 \\ 0 & 0.6296 & 0 \\ 0 & 0 & 0.2593 \end{pmatrix}$$

Le tableau associé au nuage centré $\mathfrak{N}'_I$ est donc :

	$f_1'^i$	$f_2'^i$	$f_3'^i$	$f_4'^i$	poids $p_{i.}$
f'^1 (i=1)	•	0.8889	-0.6296	-0.2593	$p_{1.}$ =0.0370
f'^2 (i=2)	•	0.0889	0.1704	-0.2593	$p_{2.}$=0.1852
f'^3 (i=3)	•	0	0.2593	-0.2593	$p_{3.}$=0.3333
f'^4 (i=4)	•	-0.1111	-0.0046	0.1157	$p_{4.}$=0.2963
f'^5 (i=5)	•	-0.1111	-0.6296	0.7407	$p_{5.}$=0.1481

Remarquons que dans le paragraphe précédent nous avons effectué nos calculs sur la matrice F des fréquences et non sur la matrice des points du nuage $\mathfrak{N}_I$, ceci pour bien faire apparaître le rôle des deux matrices D_m et D_n.
Centrer la matrice du nuage $\mathfrak{N}_I$ s'effectue en soustrayant les poids des profils-colonnes du nuage $\mathfrak{N}_J$, donc centrer la matrice F s'effectue en soustrayant aux éléments f_{ij} de F les produits $p_{i.}p_{.j}$.
Remarquons que le terme général de la matrice centrée notée F' est : $f_{ij} - p_{i.}p_{.j}$ qui n'est autre que l'écart utilisé dans un calcul de la distance du $\mathfrak{X}^2$ entre l'effectif réel f_{ij} du croisement des classes d'indices i et j et l'effectif théorique de croisement dans l'hypothèse ou les classe indexées par i celles indexées par j seraient indépendantes.
La matrice des fréquences F', centrée est :

$$F' = \begin{pmatrix} 0.0329 & -0.0233 & -0.0096 \\ 0.0165 & 0.0316 & -0.0480 \\ 0 & 0.0864 & -0.0864 \\ -0..329 & -0.0014 & 0.0343 \\ -0.0165 & -0.0933 & 0.1097 \end{pmatrix}$$

Calcul à l'ordinateur de la matrice $A' = F'^t D_n^{-1} F' D_m^{-1}$, de ses valeurs propres et de ses vecteurs propres

$$A' = \begin{pmatrix} 0.3257 & -0.0117 & -0.1112 \\ -0.0661 & 0.1605 & -0.3615 \\ -0.2595 & -0.1489 & 0.4726 \end{pmatrix}$$

Les valeurs propres de A' sont : 0.3028, 0 et 0.6560.
Une valeur propre est nulle, ce qui était prévisible puisque les profils-lignes du nuage $\mathfrak{N}_I$ sont tous dans le même plan, un des 3 axes représente donc une inertie nulle (c'est l'axe portant le vecteur propre normé colinéaire à g).
Les matrices colonnes des vecteurs propres associés sont :

$$U_0 = \begin{pmatrix} 0.1610 \\ 0.9126 \\ 0.3758 \end{pmatrix} \quad U_1 = \begin{pmatrix} -0.2492 \\ -0.5489 \\ 0.7979 \end{pmatrix} \quad U_2 = \begin{pmatrix} -0.5837 \\ 0.7863 \\ -0.2026 \end{pmatrix}$$

(Les indices 0,1 et 2 choisis de telle sorte que le vecteur u_0 corresponde à la valeur propre 0, u_1 à 0.6560 et u_2 à 0.3028.)
Le lecteur peut vérifier que u_0, (de matrice colonne U_0), est colinéaire au vecteur g défini par G.
Nous allons maintenant normer u_0, u_1 et u_2 pour la métrique d_m.

$$\|U_0\|^2 = U_0^t D_m^{-1} U_0 = 2.1009; \quad \frac{U_0}{\|U_0\|} = \begin{pmatrix} 0.1111 \\ 0.6296 \\ 0.2593 \end{pmatrix}$$

$$\|U_1\|^2 = U_1^t D_m^{-1} U_1 = 3.4930; \quad \frac{U_1}{\|U_1\|} = \begin{pmatrix} -0.1333 \\ -0.2937 \\ 0.4269 \end{pmatrix}$$

$$\|U_2\|^2 = U_2^t D_m^{-1} U_2 = 4.2066; \quad \frac{U_2}{\|U_2\|} = \begin{pmatrix} -0.2846 \\ 0.3834 \\ -0.0988 \end{pmatrix}$$

Faisons quelques remarques sur ces résultats :
• le vecteur u_0, après normalisation, est égal au vecteur g ;
• la matrice F' étant centrée les vecteurs colonnes qui la composent ont la somme de leurs composantes égale à 0, les deux vecteurs u_1 et u_2 appartiennent à l'espace engendré par ces vecteurs colonnes, la somme de leurs composantes est aussi égale à 0 : ils appartiennent à l'hyperplan d'équation x+y+z=0.
Pour obtenir la matrice des composantes principales CP_I il suffit de projeter les individus (c'est à dire les vecteurs dont les composantes sont les profils-lignes du nuage centré $\mathfrak{N}'_I$, ces vecteurs sont les n lignes de la matrice $D_n^{-1}F'$) sur les axes principaux définis par les vecteurs normés u_0, u_1, u_2.
Cette projection sera obtnue par produit scalaire pour la métrique d_m :

$$CP_I = D_n^{-1} F' D_m^{-1} VEC_I$$

où VEC_I est la matrice des vecteurs propres normés correspondants respectivement aux valeurs propres 0.3028, 0, 0.6560.

$$VEC_I = \begin{pmatrix} -0.2846 & 0.1111 & -0.1333 \\ 0.3834 & 0.6296 & -0.2937 \\ -0.0988 & 0.2593 & 0.4269 \end{pmatrix} = (U_2\, U_0\, U_1)$$

On a

$$CP_I = \begin{pmatrix} -2.5597 & 0.0000 & -1.1991 \\ -0.0256 & 0.0005 & -0.6132 \\ 0.2566 & 0.0000 & -0.5476 \\ 0.2374 & 0.0000 & 0.3260 \\ -0.3798 & 0.0000 & 1.6462 \end{pmatrix}$$

Nous remarquons que la deuxième composante principale est nulle, ce qui est normal puisque tous les profils-lignes de $\mathfrak{N}_I$ appartiennent à un plan orthogoanal à g (ce dernier est colinéaire àu_0), le

premier facteur correspond à la projection sur u_2 des coordonnées des classes d'aglaisla troisième à la projection sur u_1.
Afin donc de donner les résultats de facon bien claire nous allons réordonner la matrice des coordonnées des cinq classes d'anglais :

$$CP_I \text{ réduite et réordonnée : } \begin{pmatrix} -1.1991 & -2.5596 \\ -0.6134 & -0.0254 \\ -0.5478 & 0.2468 \\ 0.3260 & 0.2373 \\ 1.6463 & -0.3800 \end{pmatrix}$$

Ces résultats sont égaux au signe des vecteurs colonnes et au centième près : le sens des axes factoriels est indéfini car les vecteurs propres sont seulement normés.

AFC du nuage $\mathfrak{N}'_J$ centré

Dans le cas du nuage $\mathfrak{N}'_J$ la matrice à diagonaliser devient $F'D_m^{-1}F'^tD_n^{-1}$. Un vecteur propre w de matrice colonne W de cette matrice vérifie :

$$(F'D_m^{-1}F'^tD_n^{-1})W = \lambda W$$

Rappelons que d'après notre étude en ACP les valeurs propres non nulles sont les mêmes que celles de $F'^tD_m^{-1}F'^tD_n^{-1}$.
Considérons de nouveau la matrice $F'^tD_n^{-1}F'D_m^{-1}$ de valeur propre λ et de vecteur propre correspondant u, de matrice colonne U, on a :

$$(F'^tD_n^{-1}F'D_m^{-1})U = \lambda U$$

. Multiplions cette expressions à gauche par $F'^tD_m^{-1}$, il vient :

$$F'D_m^{-1}F'D_n^{-1}(F'D_m^{-1}U) = \lambda(F'D_m^{-1}U).$$

On voit que le vecteur $F'D_m^{-1}U$ est vecteur propre de $F'D_m^{-1}F'^tD_n^{-1}$ pour la valeur propre λ.
Nous chercherons donc un vecteur w non nul de matrice colonne W, colinéaire à $F'D_m^{-1}U$, de même sens et normé pour la métrique d_n définie sur $\mathfrak{N}'_J$ ce qui s'écrit :
Il existe $k \in \mathbb{R}_+^*$ tel que $W = kF'D_m^{-1}U$ et $W^tD_n^{-1}W = 1 \quad (a)$
Or u a été normé dans l'AFC de $\mathfrak{N}'_I$ donc vérifie : $U^tD_m^{-1}U = 1$.
Remplacons W par sa valeur dans l'expression (a) :
$k^2U^tD_m^{-1}F'^tD_n^{-1}F'D_m^{-1}U = 1$; en remarquant que $D_m^{-1t} = D_m^{-1}$ on a :
$k^2U^tD_m^{-1}(F'^tD_n^{-1}F'D_m^{-1}U) = 1$, or $F'^tD_n^{-1}F'D_m^{-1}U = \lambda U$, on a donc $k^2U^tD_m^{-1}\lambda U = 1$ et puisque $U^tD_m^{-1}U = 1$, on a $k^2\lambda = 1$ donc $k = \dfrac{1}{\sqrt{\lambda}}$ pour $\lambda > 0$.

Pour obtenir chacun des vecteurs propres normés relatifs au nuage $\mathfrak{N}'_J$ il suffit de multiplier chacun des vecteurs propres normés relatifs à $\mathfrak{N}'_I$ par l'expression $\imath\sqrt{\lambda}F'D_m^{-1}$ où λ est la valeur propre

strictement positive associée à chacun de ces vecteurs.
Dans le cas de notre exemple nopus obtenons donc pour vecteurs propres normés associés au nuage $\mathfrak{N}'_J$ centré :

$$W_1 = \frac{1}{\sqrt{0.6560}} F' D_m^{-1} = \begin{pmatrix} 0.1333 \\ -0.2937 \\ 0.4269 \end{pmatrix} = \begin{pmatrix} -0.0548 \\ -0.1402 \\ -0.2254 \\ 0.1193 \\ 0.3011 \end{pmatrix}$$

$$W_2 = \frac{1}{\sqrt{0.3028}} F' D_m^{-1} = \begin{pmatrix} -0.2846 \\ 0.3834 \\ -0.0988 \end{pmatrix} = \begin{pmatrix} -0.1723 \\ -0.0086 \\ 0.1554 \\ 0.1278 \\ -0.1023 \end{pmatrix}$$

Il nous reste, pour obtenir les coordonnées des classes relatives au nuage cantre $\mathfrak{N}\prime_J$ par rapport aux axes principaux, à effectuer le produit scalaire représenté matriciellement par :

$$D_m^{-1} F' D_n^{-1} (W_1 W_2) = \begin{pmatrix} -0.9718 & -1.4094 \\ -0.3778 & 0.3351 \\ 1.3337 & -0.2097 \end{pmatrix}$$

Chaque ligne de cette dernière matrice représente les coordonnées sur les deux axes principaux des 3 classes d'effectifs non nuls mathématiques.