

INTERPRETATION GÉOMÉTRIQUE DE LA RÉGRESSION

A.F. IEMMA* et R. PALM†

RÉSUMÉ

Après un rappel de quelques éléments de calcul vectoriel, on présente principalement les aspects géométriques liés à la régression, en considérant successivement la régression linéaire simple sur données centrées et sur données non centrées, la régression multiple et la régression après décomposition par valeurs singulières de la matrice des variables explicatives.

SUMMARY

After a short review of vector spaces, we mainly examine the geometry of regression analysis. We successively consider simple linear regression on centered and non centered variables, multiple linear regression and regression after singular-value decomposition of the matrix of predictors.

1. INTRODUCTION

Depuis quelques décennies, on constate une utilisation croissante de la géométrie vectorielle comme outil didactique pour la présentation des notions de statistique.

Bien que les premiers travaux dans ce sens soient relativement anciens [SCHEFFE, 1959], la publication de MARGOLIS [1979] constitue le point de départ d'une série d'études dans ce domaine.

HERR [1980] a présenté le parallélisme entre l'approche algébrique et l'approche géométrique dans le cadre de l'estimation au sens des moindres carrés et a conclu à la supériorité de l'approche géométrique des modèles linéaires en passant en revue différents articles faisant référence à la géométrie vectorielle.

*Senior bio-statisticien du Département de Recherche et Développement de Glaxo-SmithKline Biologicals, Rixensart et Maître de conférences à la Faculté universitaire des Sciences agronomiques de Gembloux durant l'année académique 1990-1991.

†Chargé de cours associé à la Faculté universitaire des Sciences agronomiques de Gembloux.

Il cite notamment les articles de FISHER [1915] et de KRUSKAL [1975]. Il estime que la relative réticence à l'utilisation de la géométrie s'explique par une certaine inertie et par une résistance à l'abstraction.

MANDEL [1982] a utilisé la technique de la décomposition par valeurs singulières de la matrice du plan d'expérience pour éclaircir à partir de procédures graphiques le problème provoqué par la colinéarité en régression .

BRYANT [1984] a montré que géométrie, probabilité et statistique sont des variations sur thème commun.

SCHEY [1985] s'est intéressé plus particulièrement à la description géométrique des contrastes en analyse de la variance à un critère de classification et a donné une justification géométrique à la nullité de la somme des coefficients intervenant dans les contrastes orthogonaux.

SAVILLE et WOOD [1986] ont utilisé des concepts de géométrie élémentaire à n dimensions pour présenter, d'une manière simple mais rigoureuse, l'analyse de la variance et la régression.

IEMMA *et al.* [1993b] décrivent l'analyse de la variance comme une application du théorème de Pythagore et donnent une procédure très simple pour la construction des projecteurs orthogonaux associés aux différentes hypothèses dans les modèles avec échantillons déséquilibrés.

Cette note est une introduction aux aspects de la géométrie vectorielle dans le contexte de la régression. DRAPER et SMITH [1998] estiment que l'approche géométrique permet de mieux comprendre, par exemple, les difficultés liées à la singularité ou à la quasi-singularité de la matrice des variables explicatives ou les difficultés associées à l'interprétation du coefficient de détermination R^2 , même si la régression peut parfaitement être abordée sans la géométrie. Ils considèrent aussi que pour une compréhension plus approfondie des moindres carrés, la connaissance de la géométrie est essentielle. Une connaissance, au moins élémentaire, est utile également pour une meilleure compréhension de la bibliographie, dans la mesure où les aspects géométriques sont présents dans de nombreuses références.

Après une présentation des éléments de base du calcul vectoriel (paragraphe 2), nous considérons la représentation vectorielle de quelques statistiques élémentaires (paragraphe 3), puis nous examinons la représentation vectorielle de la régression (paragraphe 4). Enfin, nous donnons quelques informations complémentaires (paragraphe 5) avant de conclure (paragraphe 6).

2. QUELQUES ELEMENTS DE CALCUL VECTORIEL

2.1. Espace cartésien et vecteur

On appelle espace cartésien de dimension m , \mathfrak{R}^m , l'ensemble de tous les m -uplets ordonnés (x_1, \dots, x_m) , tels que $x_i \in \mathfrak{R}$ ($i = 1, \dots, m$), \mathfrak{R} étant l'ensemble des nombres réels et m étant un nombre entier positif.

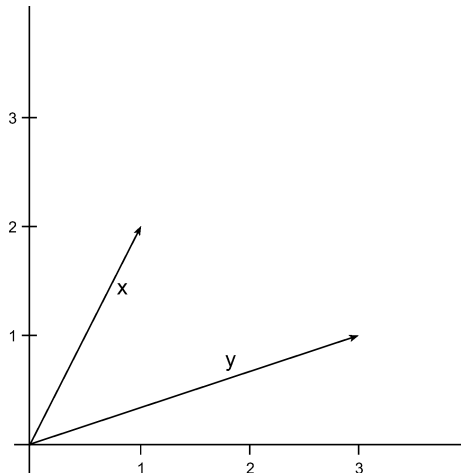


Figure 1. Vecteurs \mathbf{x} et \mathbf{y} dans l'espace vectoriel de dimension deux.

Un point donné de cet espace cartésien est donc défini par ses coordonnées (x_1, \dots, x_m) sur les m axes cartésiens et correspond à un vecteur \mathbf{x} :

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix} .$$

Géométriquement, ce vecteur \mathbf{x} est représenté par le segment de droite qui lie l'origine du système cartésien à ce point de l'espace.

A titre d'illustration, la figure 1 représente deux vecteurs, \mathbf{x} et \mathbf{y} , dans un espace cartésien de dimension deux, c'est-à-dire dans un plan :

$$\mathbf{x} = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \quad \text{et} \quad \mathbf{y} = \begin{bmatrix} 3 \\ 1 \end{bmatrix} .$$

Pour définir la longueur d'un vecteur \mathbf{x} , on utilise, dans cette note, la *norme euclidienne* ou simplement *norme*, qui est égale à :

$$\|\mathbf{x}\| = \left(\sum_{i=1}^m x_i^2 \right)^{1/2} .$$

La direction est donnée par la pente de la droite qui sous-tend le vecteur et le sens est l'orientation du vecteur sur la droite donnant la direction.

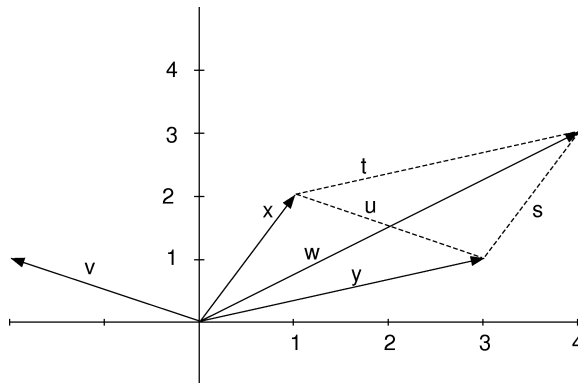


Figure 2. Addition et soustraction de deux vecteurs.

2.2. Addition et soustraction de vecteurs

Soit deux vecteurs quelconques, \mathbf{x} et \mathbf{y} , d'un même espace cartésien de dimension m . On définit l'addition et la soustraction par :

$$\mathbf{w} = \mathbf{x} \pm \mathbf{y} \quad \text{avec} \quad w_i = x_i \pm y_i \quad (i = 1, \dots, m).$$

L'addition de vecteurs présente les propriétés de commutativité :

$$\mathbf{x} + \mathbf{y} = \mathbf{y} + \mathbf{x}$$

et d'associativité :

$$(\mathbf{x} + \mathbf{y}) + \mathbf{z} = \mathbf{x} + (\mathbf{y} + \mathbf{z}) = (\mathbf{x} + \mathbf{z}) + \mathbf{y}.$$

Le vecteur nul est l'élément neutre :

$$\mathbf{x} + \mathbf{0} = \mathbf{0} + \mathbf{x} = \mathbf{x}$$

De plus, on a la propriété d'additivité inverse :

$$\mathbf{x} + (-\mathbf{x}) = (-\mathbf{x}) + \mathbf{x} = \mathbf{0}.$$

La figure 2 représente les vecteurs \mathbf{x} et \mathbf{y} examinés précédemment ainsi que les vecteurs :

$$\mathbf{v} = \mathbf{x} - \mathbf{y} = \begin{bmatrix} 1 \\ 2 \end{bmatrix} - \begin{bmatrix} 3 \\ 1 \end{bmatrix} = \begin{bmatrix} -2 \\ 1 \end{bmatrix}$$

et

$$\mathbf{w} = \mathbf{x} + \mathbf{y} = \begin{bmatrix} 1 \\ 2 \end{bmatrix} + \begin{bmatrix} 3 \\ 1 \end{bmatrix} = \begin{bmatrix} 4 \\ 3 \end{bmatrix}.$$

On remarque que l'addition correspond à la grande diagonale du parallélogramme construit en considérant que \mathbf{x} et \mathbf{y} sont deux cotés adjacents. La soustraction donne un vecteur parallèle à la petite diagonale de ce même parallélogramme.

Le côté \mathbf{t} du parallélogramme défini ci-dessus (figure 2) est obtenu par déplacement dans l'espace cartésien du vecteur \mathbf{y} parallèlement à lui-même. On dit que le vecteur \mathbf{t} est équivalent au vecteur \mathbf{y} , car ces deux vecteurs ont même longueur, même direction et même sens. De même, le vecteur \mathbf{s} est équivalent au vecteur \mathbf{x} et le vecteur \mathbf{u} est équivalent au vecteur \mathbf{v} .

On note aussi que :

$$\mathbf{x} - \mathbf{y} \neq \mathbf{y} - \mathbf{x} = -(\mathbf{x} - \mathbf{y}).$$

En effet :

$$\mathbf{y} - \mathbf{x} = \begin{bmatrix} 3 \\ 1 \end{bmatrix} - \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 2 \\ -1 \end{bmatrix}.$$

Le vecteur $\mathbf{y} - \mathbf{x}$ a donc même longueur, même direction mais le sens contraire du vecteur $\mathbf{x} - \mathbf{y}$.

2.3. Multiplication d'un vecteur par un scalaire

La multiplication d'un vecteur \mathbf{x} par un scalaire c est le vecteur obtenu par la multiplication de chaque composante de \mathbf{x} par ce scalaire :

$$\mathbf{w} = c \mathbf{x} \quad \text{avec} \quad w_i = c x_i \quad (i = 1, \dots, m).$$

Si c est positif, la multiplication d'un vecteur \mathbf{x} par c modifie uniquement la longueur du vecteur mais si c est négatif, la multiplication modifie la longueur mais aussi le sens du vecteur. La multiplication par un scalaire ne modifie donc jamais la direction d'un vecteur.

On dit qu'un ensemble de vecteurs réels de m composantes est un espace vectoriel s'il est fermé par rapport aux opérations d'addition de vecteurs et de multiplication par un scalaire, c'est-à-dire si, pour deux vecteurs quelconques et deux scalaires quelconques, le vecteur :

$$\mathbf{z} = c_1 \mathbf{x} + c_2 \mathbf{y}$$

appartient au même espace vectoriel.

L'ensemble \mathfrak{R}^m constitué de tous les vecteurs possibles de m composantes est un espace vectoriel. Un sous-ensemble S de \mathfrak{R}^m , fermé par rapport aux opérations d'addition de vecteurs et de multiplication par un scalaire est appelé sous-espace vectoriel de \mathfrak{R}^m . Ainsi, par exemple, dans un espace de dimension trois, les vecteurs :

$$\mathbf{x} = \begin{bmatrix} 6 \\ 4 \\ 2 \end{bmatrix} \quad \text{et} \quad \mathbf{y} = \begin{bmatrix} 4 \\ 2 \\ 3 \end{bmatrix}$$

définissent un sous-espace de dimension deux car, comme nous le préciserons au paragraphe 2.6, ils sont linéairement indépendants, c'est-à-dire que les composantes d'un vecteur ne sont pas, à une constante près, égales aux composantes de l'autre vecteur. Géométriquement, ce sous-espace est le plan contenant ces vecteurs. Le vecteur \mathbf{y} définit un sous-espace de dimension un, correspondant à la droite contenant ce vecteur.

2.4. Norme euclidienne d'un vecteur

Nous avons déjà signalé que la norme euclidienne d'un vecteur \mathbf{x} de l'espace cartésien \mathfrak{R}^m et défini par rapport à l'origine, est le nombre réel non négatif donné par (paragraphe 2.1) :

$$\|\mathbf{x}\| = \left(\sum_{i=1}^m x_i^2 \right)^{1/2} .$$

Géométriquement, la norme d'un vecteur est la mesure de sa longueur. Elle présente les propriétés suivantes :

$$\|\mathbf{x}\| > 0 \quad \text{si} \quad \mathbf{x} \neq \mathbf{0} \quad (\text{positivité})$$

$$\|\mathbf{x}\| = 0 \quad \text{si} \quad \mathbf{x} = \mathbf{0} \quad (\text{nullité})$$

$$\|c\mathbf{x}\| = c \|\mathbf{x}\| \quad (\text{homogénéité})$$

$$\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\| \quad (\text{inégalité triangulaire})$$

Considérons maintenant deux vecteurs \mathbf{x} et \mathbf{y} et leur somme $\mathbf{w} = \mathbf{x} + \mathbf{y}$. Par les relations dans les triangles quelconques, on a (figure 2) :

$$\|\mathbf{w}\|^2 = \|\mathbf{x} + \mathbf{y}\|^2 = \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2 + 2\|\mathbf{x}\|\|\mathbf{y}\|\cos\theta ,$$

θ étant l'angle formé par \mathbf{x} et \mathbf{y} et compris entre 0 et 180°.

Pour l'exemple illustré à la figure 2, l'angle θ est de 45° et les normes des trois vecteurs sont égales à :

$$\|\mathbf{x}\| = \sqrt{5}, \quad \|\mathbf{y}\| = \sqrt{10} \quad \text{et} \quad \|\mathbf{w}\| = 5 .$$

On vérifie bien la relation ci-dessus :

$$5 + 10 + 2\sqrt{5}\sqrt{10}\cos(45^\circ) = 25 .$$

Dans le cas particulier où $\theta = 90^\circ$, la relation ci-dessus se simplifie et on retrouve le théorème de Pythagore :

$$\|\mathbf{x} + \mathbf{y}\|^2 = \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2 .$$

Les égalités ci-dessus, justifiées géométriquement dans le cas d'un espace vectoriel de dimension deux, restent valables pour un espace de dimension quelconque.

2.5. Produit scalaire de deux vecteurs

On appelle produit scalaire (ou produit interne) de deux vecteurs \mathbf{x} et \mathbf{y} du même espace cartésien, la quantité :

$$\langle \mathbf{x}, \mathbf{y} \rangle = \sum_{i=1}^m x_i y_i = \|\mathbf{x}\| \|\mathbf{y}\| \cos \theta.$$

Il en résulte que, pour des vecteurs non nuls :

$$\text{si } \langle \mathbf{x}, \mathbf{y} \rangle > 0 \quad \text{alors} \quad \cos \theta > 0 \quad \text{et} \quad 0^\circ \leq \theta \leq 90^\circ$$

$$\text{si } \langle \mathbf{x}, \mathbf{y} \rangle = 0 \quad \text{alors} \quad \cos \theta = 0 \quad \text{et} \quad \theta = 90^\circ$$

$$\text{si } \langle \mathbf{x}, \mathbf{y} \rangle < 0 \quad \text{alors} \quad \cos \theta < 0 \quad \text{et} \quad 90^\circ < \theta \leq 180^\circ.$$

L'angle θ entre les deux vecteurs \mathbf{x} et \mathbf{y} est donc donné par :

$$\theta = \arccos \left(\frac{\langle \mathbf{x}, \mathbf{y} \rangle}{\|\mathbf{x}\| \|\mathbf{y}\|} \right) \quad (0 \leq \theta \leq 180^\circ)$$

et la condition nécessaire et suffisante pour que deux vecteurs d'un même espace cartésien soient orthogonaux est que leur produit scalaire soit nul.

Pour les vecteurs \mathbf{x} et \mathbf{y} de la figure 1, on a :

$$\langle \mathbf{x}, \mathbf{y} \rangle = 5 \quad \text{et} \quad \theta = \arccos \left(\frac{5}{\sqrt{5} \sqrt{10}} \right) = 45^\circ.$$

2.6. Dépendance linéaire de vecteurs

Un ensemble de vecteurs $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$, de l'espace cartésien \mathfrak{R}^m est linéairement dépendant si, pour un ensemble de scalaires $\alpha_1, \dots, \alpha_n$, il existe au moins un α_j ($j = 1, \dots, n$) non nul tel que :

$$\sum_{j=1}^n \alpha_j \mathbf{v}_j = \mathbf{0}$$

Dans le cas contraire, l'ensemble est dit linéairement indépendant.

Les vecteurs \mathbf{x} et \mathbf{y} de la figure 2 sont deux vecteurs linéairement indépendants. Il en est de même des vecteurs \mathbf{x} et \mathbf{v} ou \mathbf{y} et \mathbf{w} . Par contre, les trois vecteurs \mathbf{x} , \mathbf{y} et \mathbf{w} ne sont pas linéairement indépendants car \mathbf{w} est la somme des deux premiers.

Puisque les vecteurs \mathbf{x} et \mathbf{y} sont linéairement indépendants, n'importe quel vecteur de l'espace cartésien de dimension deux peut être obtenu par une combinaison linéaire de \mathbf{x} et \mathbf{y} .

La relation définissant la condition de dépendance linéaire des vecteurs $\mathbf{v}_1, \dots, \mathbf{v}_n$:

$$\sum_{j=1}^n \alpha_j \mathbf{v}_j = \mathbf{0}$$

peut encore s'écrire :

$$\mathbf{A} \boldsymbol{\alpha} = \mathbf{0},$$

\mathbf{A} étant la matrice de dimensions $m \times n$ obtenue par la juxtaposition des n vecteurs \mathbf{v}_i et $\boldsymbol{\alpha}$ est le vecteur-colonne contenant les n valeurs α_j .

Ainsi, pour les vecteurs \mathbf{x} , \mathbf{y} et \mathbf{w} de la figure 2, on peut écrire :

$$\begin{bmatrix} 1 & 3 & 4 \\ 2 & 1 & 3 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

La relation ci-dessus correspond à un système homogène d'équations linéaires. Si la matrice \mathbf{A} est de rang-colonne complet, c'est-à-dire de rang égal à n , un tel système possède comme unique solution la solution triviale $\alpha_j = 0$. Dans ce cas, les vecteurs \mathbf{v}_j sont linéairement indépendants. Par contre, le système d'équations est indéterminé si le rang de \mathbf{A} est inférieur à n et les n vecteurs sont linéairement dépendants [IEMMA et PALM, 1993].

D'une manière plus générale, si une matrice \mathbf{A} , formée de la juxtaposition de n vecteurs, est de rang k , alors l'ensemble des vecteurs contient k vecteurs linéairement indépendants.

2.7. Espace-ligne et espace-colonne d'une matrice

Soit \mathbf{A} une matrice de dimensions $m \times n$ et de rang k . Les n vecteurs-colonnes génèrent un sous-espace S , de dimension k ($k \leq n$). Ce même espace S peut être généré par k vecteurs linéairement indépendants : on dit que les k vecteurs sont une base pour le sous-espace S . Ce sous-espace est appelé l'espace-colonne de \mathbf{A} et est noté $C(\mathbf{A})$. De même, le sous-espace engendré par les m lignes de \mathbf{A} est défini comme l'espace-ligne de \mathbf{A} et est noté $L(\mathbf{A})$. Sa dimension est inférieure ou égale à m .

Comme le rang-colonne et le rang-ligne d'une matrice sont identiques, la dimension du sous-espace-colonne est identique à la dimension du sous-espace-ligne. Cette dimension est égale au rang de la matrice \mathbf{A} .

Considérons, à titre d'exemple, la matrice \mathbf{A} :

$$\mathbf{A} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{bmatrix}.$$

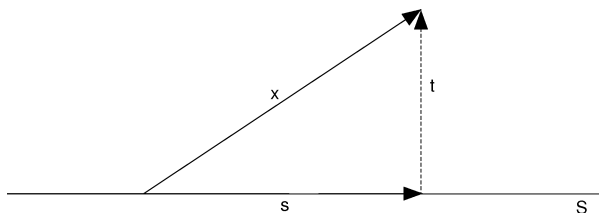


Figure 3. Projection orthogonale d'un vecteur dans un sous-espace.

Cette matrice est de rang égal à deux. Les trois vecteurs-colonnes, qui appartiennent à \mathbb{R}^4 , puisqu'ils ont quatre composantes, engendrent un sous-espace $C(\mathbf{A})$ de dimension deux. De même, les quatre vecteurs-lignes, qui appartiennent à \mathbb{R}^3 , puisqu'ils ont trois composantes, engendrent un sous-espace $L(\mathbf{A})$ de dimension deux.

2.8. Projection orthogonale

La projection d'un vecteur \mathbf{x} appartenant à un espace V sur un sous-espace S est un cas particulier d'une transformation linéaire.

La projection orthogonale de \mathbf{x} sur S , notée \mathbf{s} , est la projection sur S qui minimise la distance entre l'extrémité de \mathbf{s} et l'extrémité de \mathbf{x} . Géométriquement, le vecteur \mathbf{s} est obtenu en abaissant la perpendiculaire du point \mathbf{x} sur l'espace S . La figure 3 schématise la projection orthogonale d'un vecteur \mathbf{x} situé dans un espace de dimension quelconque mais supérieure à un sur le sous-espace S de dimension un, c'est-à-dire sur une droite.

On constate que le vecteur \mathbf{x} peut être décomposé en deux vecteurs :

$$\mathbf{x} = \mathbf{s} + \mathbf{t},$$

où \mathbf{s} est la projection orthogonale de \mathbf{x} sur S et où \mathbf{t} est un vecteur dont la direction est la direction de la projection. La particularité de \mathbf{s} est qu'il est le vecteur de l'espace S tel que la norme de :

$$\mathbf{t} = \mathbf{x} - \mathbf{s},$$

soit minimum. Ce vecteur est unique.

La décomposition de \mathbf{x} en deux vecteurs \mathbf{s} et \mathbf{t} peut se généraliser au cas où \mathbf{x} est un vecteur dans un espace de dimension m et où S est un espace de dimension k ($k < m$) engendré par un sous-ensemble de vecteurs orthonormaux, $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k$, c'est-à-dire des vecteurs orthogonaux et de norme unitaire.

Le vecteur \mathbf{s} est une combinaison linéaire des vecteurs \mathbf{v}_j ($j = 1, \dots, k$). En effet, on peut montrer que :

$$\mathbf{s} = \langle \mathbf{x}, \mathbf{v}_1 \rangle \mathbf{v}_1 + \dots + \langle \mathbf{x}, \mathbf{v}_k \rangle \mathbf{v}_k .$$

D'autre part, on peut aussi montrer que :

$$\mathbf{s} = (\mathbf{v}'_1 \mathbf{v}_1 + \dots + \mathbf{v}'_k \mathbf{v}_k) \mathbf{x} = \mathbf{P} \mathbf{x} ,$$

\mathbf{v}'_j étant le vecteur transposé de \mathbf{v}_j . La matrice \mathbf{P} est appelée matrice de projection ou encore projecteur orthogonal de \mathbf{x} sur S .

D'une manière plus générale, le projecteur \mathbf{P} d'un vecteur \mathbf{x} sur l'espace-colonne d'une matrice \mathbf{A} peut être obtenu de la façon suivante [IEMMA *et al.*, 1993b] :

$$\mathbf{P} = \mathbf{A}(\mathbf{A}' \mathbf{A})^G \mathbf{A}' = \mathbf{A} \mathbf{A}^+ = \mathbf{A} \mathbf{A}^l ,$$

où $(\mathbf{A}' \mathbf{A})^G$ est une inverse généralisée quelconque de \mathbf{A} , \mathbf{A}^+ est l'inverse généralisée de MOORE-PENROSE et \mathbf{A}^l est l'inverse généralisée des moindres carrés. Ces différences de matrices inverses généralisées ont fait l'objet d'une précédente note [IEMMA et PALM, 1992].

La matrice \mathbf{P} est symétrique, idempotente et ne dépend pas du choix de l'inverse généralisée.

De façon analogue, $(\mathbf{I} - \mathbf{P})$ est la matrice qui projette le vecteur \mathbf{x} sur le complément orthogonal de l'espace engendré par les colonnes de \mathbf{A} , $C^\perp(\mathbf{A})$:

$$\mathbf{t} = (\mathbf{I} - \mathbf{P}) \mathbf{x} \quad \mathbf{t} \in C^\perp(\mathbf{A}) ,$$

\mathbf{I} étant la matrice identité.

2.9. Décomposition par valeurs singulières

La décomposition par valeurs singulières d'une matrice \mathbf{X} , de dimensions $n \times p$ et de rang k , consiste en la factorisation suivante [ECKART et YOUNG, 1936] :

$$\mathbf{X} = \mathbf{U} \mathbf{S} \mathbf{V}' ,$$

où \mathbf{U} est une matrice de dimensions $n \times k$, \mathbf{S} une matrice carrée de dimensions $k \times k$ et \mathbf{V}' une matrice de dimensions $k \times p$, dont la transposée est \mathbf{V} . \mathbf{U} et \mathbf{V} sont des matrices orthonormées, c'est-à-dire telles que :

$$\mathbf{U}' \mathbf{U} = \mathbf{I} \quad \text{et} \quad \mathbf{V}' \mathbf{V} = \mathbf{I} .$$

De plus, \mathbf{U} est la matrice dont les colonnes sont les vecteurs propres normés à l'unité associés aux valeurs propres de $\mathbf{X} \mathbf{X}'$; \mathbf{V}' est la matrice dont les lignes sont les vecteurs propres normés à l'unité associés aux valeurs propres non nulles de $\mathbf{X}' \mathbf{X}$ et \mathbf{S} est une matrice diagonale dont les éléments diagonaux sont les

valeurs singulières de \mathbf{X} , c'est-à-dire les racines carrées des valeurs propres de $\mathbf{X} \mathbf{X}'$ ou de $\mathbf{X}' \mathbf{X}$.

On peut montrer que \mathbf{U} constitue une base pour l'espace des colonnes de \mathbf{X} , $C(\mathbf{X})$ et \mathbf{V} constitue une base pour l'espace des lignes de \mathbf{X}' , $L(\mathbf{X}')$ et on peut écrire :

$$\mathbf{v}_i = \frac{1}{s_i} \mathbf{X}' \mathbf{u}_i \quad \text{et} \quad \mathbf{u}_i = \frac{1}{s_i} \mathbf{X} \mathbf{v}_i \quad (i = 1, \dots, k).$$

3. REPRÉSENTATION VECTORIELLE DE QUELQUES STATISTIQUES ÉLÉMENTAIRES

3.1. Moyenne et variance

Considérons, par exemple, la série statistique suivante, relative à une caractéristique donnée :

$$y_1 = 4, \quad y_2 = 2 \quad \text{et} \quad y_3 = 3.$$

Cette série peut être représentée, dans l'espace de dimension trois, par le vecteur \mathbf{y} , dont les éléments sont les trois observations (paragraphe 2.1).

Soit $\mathbf{1}$ le vecteur dont les coordonnées sont toutes égales à l'unité. La projection orthogonale de \mathbf{y} sur le sous-espace, de dimension un, engendré par le vecteur $\mathbf{1}$, est donnée par (paragraphe 2.8) :

$$\mathbf{P} \mathbf{y} = \mathbf{1}(\mathbf{1}' \mathbf{1})^{-1} \mathbf{1}' \mathbf{y} = \begin{bmatrix} 1/3 & 1/3 & 1/3 \\ 1/3 & 1/3 & 1/3 \\ 1/3 & 1/3 & 1/3 \end{bmatrix} \begin{bmatrix} 4 \\ 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 3 \\ 3 \\ 3 \end{bmatrix},$$

car $\mathbf{1}' \mathbf{1}$ ne contient qu'un seul élément, égal à 3.

On constate que la projection orthogonale de \mathbf{y} sur $\mathbf{1}$ est un vecteur dont les composantes sont égales à la moyenne \bar{y} . Le vecteur $\mathbf{P} \mathbf{y}$ est donc égal à $\bar{y} \mathbf{1}$. Le vecteur orthogonal à $\mathbf{P} \mathbf{y}$ est égal à :

$$\mathbf{e} = (\mathbf{I} - \mathbf{P}) \mathbf{y} = \begin{bmatrix} 2/3 & -1/3 & -1/3 \\ -1/3 & 2/3 & -1/3 \\ -1/3 & -1/3 & 2/3 \end{bmatrix} \begin{bmatrix} 4 \\ 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}.$$

Les éléments de \mathbf{e} correspondent aux écarts par rapport à la moyenne pour chacune des observations et la décomposition vectorielle :

$$\mathbf{y} = \bar{y} \mathbf{1} + \mathbf{e}$$

permet de retrouver la décomposition habituelle de la somme des carrés des écarts :

$$\|\mathbf{y}\|^2 = \|\bar{y} \mathbf{1}\|^2 + \|\mathbf{e}\|^2 = \bar{y} \|\mathbf{1}\|^2 + \|\mathbf{e}\|^2$$

ou :

$$\sum_{i=1}^3 y_i^2 = n \bar{y}^2 + \sum_{i=1}^3 (y_i - \bar{y})^2,$$

soit aussi :

$$\sum_{i=1}^3 (y_i - \bar{y})^2 = \sum_{i=1}^3 y_i^2 - \frac{1}{3} \left(\sum_{i=1}^3 y_i \right)^2.$$

A une constante près, le carré de la norme du vecteur \mathbf{e} correspond donc à la variance. Cette constante, égale à $1/2$ pour l'exemple, est l'inverse de la dimension de l'espace complément orthogonal du vecteur $\mathbf{1}$, $C^\perp(\mathbf{1})$.

D'une manière plus générale, toute série statistique de n observations peut être représentée dans l'espace des individus de dimension n , par un vecteur \mathbf{y} . Ce vecteur \mathbf{y} peut être décomposé en deux vecteurs orthogonaux $\bar{y} \mathbf{1}$ et \mathbf{e} . Le vecteur $\bar{y} \mathbf{1}$ est la projection orthogonale du vecteur \mathbf{y} sur le sous-espace engendré par le vecteur $\mathbf{1}$, qui est un vecteur dont les n composantes sont égales à l'unité. La matrice de projection, de dimensions $n \times n$ est une matrice dont tous les éléments sont égaux à $1/n$. Les normes des vecteurs permettent d'obtenir la formule de calcul classique de la somme des carrés des écarts :

$$\|\mathbf{e}\|^2 = \|\mathbf{y}\|^2 - \|\bar{y} \mathbf{1}\|^2$$

ou :

$$\sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n y_i^2 - n \bar{y}^2 = \text{SCE}_y.$$

En divisant la somme des carrés des écarts par le nombre de degrés de liberté, égal à $n - 1$, on obtient la variance estimée. Ce nombre de degrés de liberté est la dimension du sous-espace orthogonal au sous-espace engendré par le vecteur $\mathbf{1}$.

3.2. Covariance et corrélation

Une série statistique double $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ peut être représentée, dans l'espace des observations de dimension n , par deux vecteurs \mathbf{x} et \mathbf{y} .

Soit $\mathbf{P}\mathbf{x}$ et $\mathbf{P}\mathbf{y}$, les projections orthogonales de \mathbf{x} et \mathbf{y} sur le sous-espace engendré par le vecteur $\mathbf{1}$, dont les composantes sont toutes égales à l'unité :

$$\mathbf{P}\mathbf{x} = \bar{x} \mathbf{1} \quad \text{et} \quad \mathbf{P}\mathbf{y} = \bar{y} \mathbf{1}.$$

Soit $\mathbf{x} - \bar{x} \mathbf{1}$ et $\mathbf{y} - \bar{y} \mathbf{1}$, les vecteurs, orthogonaux à $\mathbf{P}\mathbf{x}$ et $\mathbf{P}\mathbf{y}$ dont les composantes sont, d'une part, les écarts $(x_i - \bar{x})$ et d'autre part les écarts $(y_i - \bar{y})$ (paragraphe 3.1). Le produit scalaire des vecteurs $\mathbf{x} - \bar{x} \mathbf{1}$ et $\mathbf{y} - \bar{y} \mathbf{1}$ correspond à la somme des produits des écarts (paragraphe 2.5) :

$$\langle \mathbf{x} - \bar{x} \mathbf{1}, \mathbf{y} - \bar{y} \mathbf{1} \rangle = \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = \text{SPE}.$$

La division de cette somme de produits d'écart par la dimension du sous-espace complément orthogonal de $\mathbf{1}$, c'est-à-dire $n - 1$, conduit à la covariance estimée.

D'autre part, on sait que (paragraphe 2.5) :

$$\langle \mathbf{x} - \bar{x}\mathbf{1}, \mathbf{y} - \bar{y}\mathbf{1} \rangle = \|\mathbf{x} - \bar{x}\mathbf{1}\| \|\mathbf{y} - \bar{y}\mathbf{1}\| \cos \theta,$$

θ étant l'angle formé par les deux vecteurs $\mathbf{x} - \bar{x}\mathbf{1}$ et $\mathbf{y} - \bar{y}\mathbf{1}$. De plus (paragraphe 3.1) :

$$\|\mathbf{x} - \bar{x}\mathbf{1}\|^2 = \text{SCE}_x \quad \text{et} \quad \|\mathbf{y} - \bar{y}\mathbf{1}\|^2 = \text{SCE}_y.$$

Il en résulte que :

$$\cos \theta = \frac{\text{SPE}}{\sqrt{\text{SCE}_x \text{SCE}_y}} = r_{xy}.$$

Le coefficient de corrélation est donc égal au cosinus de l'angle formé par le vecteur des écarts par rapport à la moyenne de x et par le vecteur des écarts par rapport à la moyenne de y .

Si la corrélation est nulle, ces deux vecteurs forment un angle droit : ils sont orthogonaux. Si la corrélation est égale à l'unité, ils forment un angle nul : ils sont colinéaires et sont de même sens. Si la corrélation est égale à -1 , ils sont colinéaires mais sont de sens opposé. Les deux vecteurs forment un angle compris entre 0 et 90° , si la corrélation est positive et un angle compris entre 90 et 180° , si la corrélation est négative.

A titre d'illustration, nous reprenons le vecteur \mathbf{y} défini au paragraphe 3.1 et nous considérons un deuxième vecteur \mathbf{x} , dont les éléments sont respectivement égaux à 6, 4 et 2 :

$$\mathbf{x} = \begin{bmatrix} 6 \\ 4 \\ 2 \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} 4 \\ 2 \\ 3 \end{bmatrix}.$$

Les projections orthogonales sur le vecteur unitaire $\mathbf{1}$ sont :

$$\bar{x}\mathbf{1} = \begin{bmatrix} 4 \\ 4 \\ 4 \end{bmatrix}, \quad \bar{y}\mathbf{1} = \begin{bmatrix} 3 \\ 3 \\ 3 \end{bmatrix}$$

et les projections sur le sous-espace complément orthogonal à $\mathbf{1}$ valent :

$$\mathbf{x} - \bar{x}\mathbf{1} = \begin{bmatrix} 2 \\ 0 \\ -2 \end{bmatrix}, \quad \mathbf{y} - \bar{y}\mathbf{1} = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}.$$

On a donc :

$$r_{xy} = \cos \theta = \frac{\langle \mathbf{x} - \bar{x}\mathbf{1}, \mathbf{y} - \bar{y}\mathbf{1} \rangle}{\|\mathbf{x} - \bar{x}\mathbf{1}\| \|\mathbf{y} - \bar{y}\mathbf{1}\|} = \frac{2}{\sqrt{(8)(2)}} = 0,5,$$

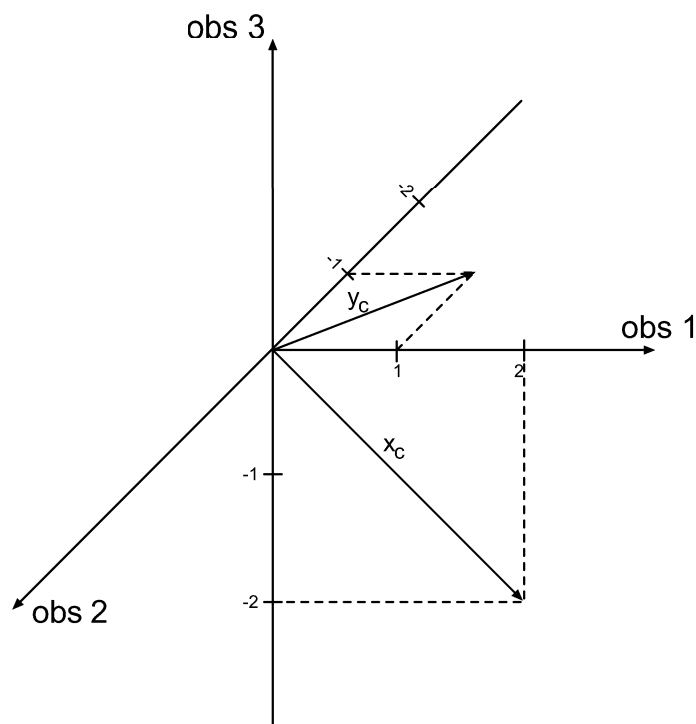


Figure 4. Représentation des vecteurs y_c et x_c dans l'espace de dimension trois.

et l'angle θ vaut 60° .

La figure 4 donne la représentation graphique, dans l'espace de dimension trois, des vecteurs $x - \bar{x} \mathbf{1}$ et $y - \bar{y} \mathbf{1}$, notés x_c et y_c .

4. REPRÉSENTATION VECTORIELLE DE LA RÉGRESSION

4.1. La régression linéaire simple sur données centrées

On sait que, pour une régression linéaire avec ordonnée à l'origine, le coefficient de régression peut être obtenu en calculant la régression passant par l'origine sur les données centrées $(x_i - \bar{x})$ et $(y_i - \bar{y})$. C'est l'approche que nous allons utiliser dans un premier temps.

Nous reprenons l'exemple du paragraphe 3.2. Nous avons donc, respectivement comme variable explicative et comme variable à expliquer :

$$x_c = x - \bar{x} \mathbf{1} = \begin{bmatrix} 2 \\ 0 \\ -2 \end{bmatrix} \quad \text{et} \quad y_c = y - \bar{y} \mathbf{1} = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}.$$

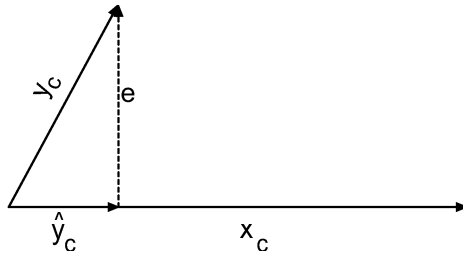


Figure 5. Projection de \mathbf{y}_c sur \mathbf{x}_c .

On considère le modèle ajusté suivant :

$$\mathbf{y}_c = \mathbf{x}_c \hat{\beta} + \mathbf{e},$$

qui correspond à la décomposition du vecteur \mathbf{y}_c en deux composantes. La première composante, $\mathbf{x}_c \hat{\beta}$, est la composante dans le sous-espace engendré par le vecteur \mathbf{x}_c , puisqu'il s'agit de la multiplication de \mathbf{x}_c par le scalaire $\hat{\beta}$, qui a été estimé. Elle correspond au vecteur des valeurs estimées $\hat{\mathbf{y}}_c$. La deuxième composante, \mathbf{e} , est le vecteur des résidus estimés.

Le principe de la méthode des moindres carrés est de déterminer la valeur de $\hat{\beta}$ de manière à rendre minimum la somme des carrés des résidus estimés. Géométriquement, le problème revient donc à décomposer le vecteur \mathbf{y}_c de manière à minimiser la norme de \mathbf{e} . Cela implique que $\mathbf{x}_c \hat{\beta}$ doit être la projection orthogonale de \mathbf{y}_c sur \mathbf{x}_c (paragraphe 2.8). On a donc :

$$\hat{\mathbf{y}}_c = \mathbf{x}_c \hat{\beta} = \mathbf{P} \mathbf{y}_c = \mathbf{x}_c (\mathbf{x}'_c \mathbf{x}_c)^{-1} \mathbf{x}'_c \mathbf{y}_c = \begin{bmatrix} 1/2 \\ 0 \\ -1/2 \end{bmatrix}$$

$$\hat{\beta} = (\mathbf{x}'_c \mathbf{x}_c)^{-1} \mathbf{x}'_c \mathbf{y}_c = 1/4$$

et

$$\mathbf{e} = (\mathbf{I} - \mathbf{P}) \mathbf{y}_c = \mathbf{y}_c - \mathbf{x}_c \hat{\beta} = \begin{bmatrix} 1/2 \\ -1 \\ 1/2 \end{bmatrix}.$$

La figure 5 reprend l'espace engendré par \mathbf{x}_c et \mathbf{y}_c et montre la projection de \mathbf{y}_c sur \mathbf{x}_c , notée $\hat{\mathbf{y}}_c$, et la projection de \mathbf{y}_c sur l'espace complément orthogonal de \mathbf{x}_c , notée \mathbf{e} .

En considérant les sommes des vecteurs, on peut écrire la relation suivante :

$$\|\mathbf{y}_c\|^2 = \|\mathbf{x}_c \hat{\beta}\|^2 + \|\mathbf{e}\|^2,$$

soit :

$$2 = 1/2 + 3/2.$$

Il s'agit en fait de la décomposition habituelle de la somme des carrés des écarts suivante :

$$SC_{y_c} = SC_{rég} + SC_{rés}.$$

Dans cette relation, SC_{y_c} est la somme des carrés de \mathbf{y}_c , c'est-à-dire la somme des carrés des écarts de \mathbf{y} , $SC_{rég}$ est la somme des carrés liée à la régression sur la variable \mathbf{x}_c et $SC_{rés}$ est la somme des carrés des résidus.

Le coefficient de détermination, r_{yx}^2 , est donné par :

$$r_{yx}^2 = \cos^2 \theta = \frac{(\langle \mathbf{y}_c, \hat{\mathbf{y}}_c \rangle)^2}{\|\mathbf{x}_c\|^2 \|\hat{\mathbf{y}}_c\|^2} = \frac{\|\hat{\mathbf{y}}_c\|^2}{\|\mathbf{y}_c\|^2} = \frac{SC_{rég}}{SC_{y_c}} = 0,25.$$

Le problème qui vient d'être examiné dans le cas de trois observations peut être étendu sans difficulté au cas où n est quelconque. Le vecteur \mathbf{y}_c , de l'espace de dimension $n - 1$, est divisé en deux composantes : la projection orthogonale $\mathbf{x}_c \hat{\beta}$ sur le sous-espace de dimension un, engendré par \mathbf{x}_c , et la projection \mathbf{e} sur le sous-espace orthogonal à \mathbf{x}_c , de dimension $n - 2$. Les carrés des normes de ces trois vecteurs sont respectivement la somme des carrés des écarts de \mathbf{y} , la somme des carrés liée à la régression sur \mathbf{x}_c et la somme des carrés des résidus. Les dimensions des espaces correspondent aux degrés de liberté associés aux sommes des carrés.

4.2. La régression linéaire simple sur données non centrées

Au paragraphe précédent, nous avons éliminé le problème de l'ordonnée à l'origine en centrant les variables y et x . Une autre solution consiste à prendre en compte simultanément l'ordonnée à l'origine et le coefficient de régression.

Il suffit, en effet, de considérer qu'on dispose de deux variables explicatives, notées x_0 et x_1 , x_0 étant une variable explicative pour laquelle toutes les observations sont égales à l'unité. Par la suite, le problème pourra très facilement être étendu au cas où on a plusieurs variables explicatives, avec ou sans ordonnée à l'origine.

Pour l'exemple traité aux paragraphes 3.2 et 4.1, on a :

$$\mathbf{y} = \begin{bmatrix} 4 \\ 2 \\ 3 \end{bmatrix}, \quad \mathbf{x}_0 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \quad \text{et} \quad \mathbf{x}_1 = \begin{bmatrix} 6 \\ 4 \\ 2 \end{bmatrix}.$$

Soit $\mathbf{X} = (\mathbf{x}_0 \mathbf{x}_1)$, la matrice formée des vecteurs \mathbf{x}_0 et \mathbf{x}_1 . On a le modèle :

$$\mathbf{y} = \mathbf{X} \hat{\beta} + \mathbf{e}.$$

Le vecteur des valeurs estimées $\hat{\mathbf{y}}$ est la projection de \mathbf{y} sur le sous-espace de dimension deux, engendré par les colonnes de \mathbf{X} , c'est-à-dire par le vecteur unitaire \mathbf{x}_0 et le vecteur \mathbf{x}_1 :

$$\hat{\mathbf{y}} = \mathbf{X} \hat{\beta} = \mathbf{P} \mathbf{y} = \mathbf{X} (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \mathbf{y} = \begin{bmatrix} 7/2 \\ 3 \\ 5/2 \end{bmatrix}.$$

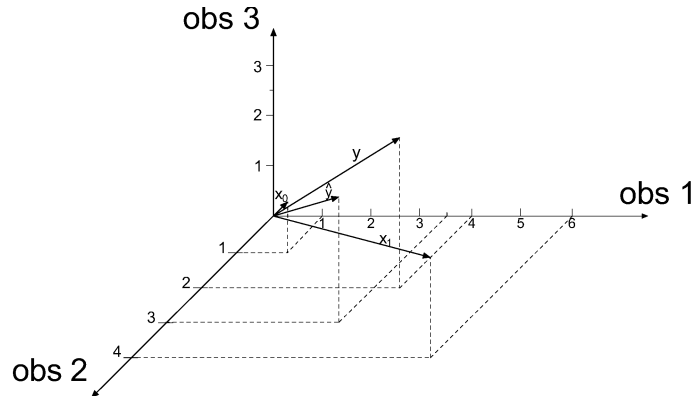


Figure 6. Représentation des vecteurs \mathbf{y} , \mathbf{x}_0 , \mathbf{x}_1 et $\hat{\mathbf{y}}$ dans l'espace de dimension trois.

La figure 6 donne la représentation des vecteurs \mathbf{y} , \mathbf{x}_0 , \mathbf{x}_1 et $\hat{\mathbf{y}}$ dans l'espace de dimension trois. On peut remarquer que les vecteurs \mathbf{x}_0 , \mathbf{x}_1 et $\hat{\mathbf{y}}$ sont situés dans un même plan, qui est le sous-espace engendré par les colonnes de \mathbf{X} . La représentation dans le plan de ces trois vecteurs fera l'objet des figures 7 et 8, qui seront examinées par la suite.

Le vecteur \mathbf{e} est le vecteur de la projection de \mathbf{y} sur l'espace complément orthogonal de \mathbf{X} , de dimension un. Il est identique au vecteur des résidus obtenu au paragraphe 4.1. Il n'est pas représenté dans la figure 6, afin de ne pas surcharger celle-ci.

Le vecteur $\hat{\boldsymbol{\beta}}$ est égal à :

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = \begin{bmatrix} 2 \\ 1/4 \end{bmatrix}.$$

En terme de sommes de carrés, on a :

$$\|\mathbf{y}\|^2 = \|\hat{\mathbf{y}}\|^2 + \|\mathbf{e}\|^2$$

soit : $29 = 27,5 + 1,5$.

Dans le sous-espace engendré par \mathbf{x}_0 et \mathbf{x}_1 , le vecteur $\hat{\mathbf{y}}$ peut se décomposer en deux parties : une première composante est la projection orthogonale de $\hat{\mathbf{y}}$ sur le sous-espace engendré par \mathbf{x}_0 et la deuxième composante est la projection de $\hat{\mathbf{y}}$ sur le complément orthogonal du sous-espace engendré par \mathbf{x}_0 . On a :

$$\hat{\mathbf{y}} = \hat{\mathbf{y}}(\mathbf{x}_0) + \hat{\mathbf{y}}(\mathbf{x}_1 | \mathbf{x}_0),$$

avec :

$$\hat{\mathbf{y}}(\mathbf{x}_0) = \mathbf{x}_0(\mathbf{x}_0'\mathbf{x}_0)^{-1}\mathbf{x}_0'\hat{\mathbf{y}} = \begin{bmatrix} 3 \\ 3 \\ 3 \end{bmatrix},$$

et :

$$\hat{\mathbf{y}}(\mathbf{x}_1 | \mathbf{x}_0) = \hat{\mathbf{y}} - \hat{\mathbf{y}}(\mathbf{x}_0) = \begin{bmatrix} 7/2 \\ 3 \\ 5/2 \end{bmatrix} - \begin{bmatrix} 3 \\ 3 \\ 3 \end{bmatrix} = \begin{bmatrix} 1/2 \\ 0 \\ -1/2 \end{bmatrix}.$$

On obtient ainsi :

$$\|\hat{\mathbf{y}}\|^2 = \|\hat{\mathbf{y}}(\mathbf{x}_0)\|^2 + \|\hat{\mathbf{y}}(\mathbf{x}_1 | \mathbf{x}_0)\|^2$$

soit : $27,5 = 27 + 0,5$.

La quantité $\|\hat{\mathbf{y}}(\mathbf{x}_0)\|^2$ correspond à la quantité $n\bar{y}^2$, c'est-à-dire au terme correctif utilisé lors du calcul de la somme des carrés des écarts de y et la quantité $\|\hat{\mathbf{y}}(\mathbf{x}_1 | \mathbf{x}_0)\|^2$ est la somme des carrés liée à la régression de y_c sur x_c , obtenue au paragraphe 4.1.

En combinant les décompositions successives, on obtient :

$$\|\mathbf{y}\|^2 = \|\hat{\mathbf{y}}(\mathbf{x}_0)\|^2 + \|\hat{\mathbf{y}}(\mathbf{x}_1 | \mathbf{x}_0)\|^2 + \|\mathbf{e}\|^2$$

soit : $29 = 27 + 0,5 + 1,5$.

La décomposition de $\hat{\mathbf{y}}$ peut être encore faite d'une autre manière : une première composante, $\hat{\mathbf{y}}(\mathbf{x}_1)$, est la projection orthogonale de $\hat{\mathbf{y}}$ sur le sous-espace engendré par \mathbf{x}_1 et la deuxième composante, $\hat{\mathbf{y}}(\mathbf{x}_0 | \mathbf{x}_1)$, est la projection \mathbf{y} sur le complément orthogonal du sous-espace engendré par \mathbf{x}_1 . On a alors :

$$\hat{\mathbf{y}} = \hat{\mathbf{y}}(\mathbf{x}_1) + \hat{\mathbf{y}}(\mathbf{x}_0 | \mathbf{x}_1),$$

avec :

$$\hat{\mathbf{y}}(\mathbf{x}_1) = \mathbf{x}_1(\mathbf{x}'_1\mathbf{x}_1)^{-1}\mathbf{x}'_1\hat{\mathbf{y}} = \begin{bmatrix} 4,07 \\ 2,71 \\ 1,36 \end{bmatrix},$$

et :

$$\hat{\mathbf{y}}(\mathbf{x}_0 | \mathbf{x}_1) = \hat{\mathbf{y}} - \hat{\mathbf{y}}(\mathbf{x}_1) = \begin{bmatrix} -0,57 \\ 0,29 \\ 1,14 \end{bmatrix}.$$

La décomposition de la somme des carrés s'écrit alors :

$$\|\hat{\mathbf{y}}\|^2 = \|\hat{\mathbf{y}}(\mathbf{x}_1)\|^2 + \|\hat{\mathbf{y}}(\mathbf{x}_0 | \mathbf{x}_1)\|^2,$$

soit : $27,5 = 25,79 + 1,71$.

La projection orthogonale de $\hat{\mathbf{y}}$ sur le sous-espace engendré par \mathbf{x}_1 , $\hat{\mathbf{y}}(\mathbf{x}_1)$, est donc différente de la projection de $\hat{\mathbf{y}}$ sur le sous-espace complément orthogonal du sous-espace \mathbf{x}_0 , $\hat{\mathbf{y}}(\mathbf{x}_1 | \mathbf{x}_0)$. Cela résulte de la non-orthogonalité de \mathbf{x}_0 et \mathbf{x}_1 : ces deux vecteurs forment en effet un angle qui n'est pas un angle droit.

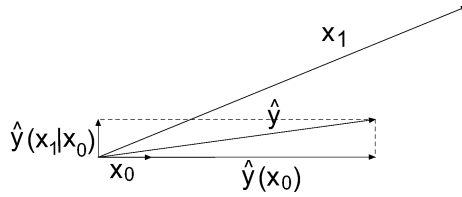


Figure 7. Décomposition de $\hat{\mathbf{y}}$ en $\hat{\mathbf{y}}(\mathbf{x}_0)$ et $\hat{\mathbf{y}}(\mathbf{x}_1 | \mathbf{x}_0)$.

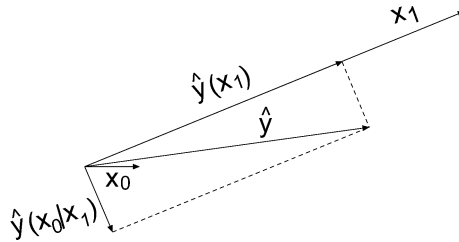


Figure 8. Décomposition de $\hat{\mathbf{y}}$ en $\hat{\mathbf{y}}(\mathbf{x}_1)$ et $\hat{\mathbf{y}}(\mathbf{x}_0 | \mathbf{x}_1)$.

Les figures 7 et 8 illustrent les deux décompositions du vecteur $\hat{\mathbf{y}}$. La première décomposition est la plus courante, dans la mesure où on s'intéresse surtout à la part de la somme des carrés des écarts de y qui est expliquée par la variable explicative x . On peut remarquer d'ailleurs que, pour cette décomposition, le vecteur $\hat{\mathbf{y}}(\mathbf{x}_1 | \mathbf{x}_0)$ est en fait la projection de $\hat{\mathbf{y}}$ sur le sous-espace engendré par $\mathbf{x} - \bar{x}\mathbf{1}$, dont il a été question à propos de la régression linéaire sur données centrées (paragraphe 4.1). Ce vecteur $\hat{\mathbf{y}}(\mathbf{x}_1 | \mathbf{x}_0)$ peut également être obtenu directement par la décomposition de \mathbf{y} , plutôt que par la décomposition de $\hat{\mathbf{y}}$. En effet, $\hat{\mathbf{y}}(\mathbf{x}_0)$ est la projection orthogonale de \mathbf{y} sur \mathbf{x}_0 :

$$\hat{\mathbf{y}}(\mathbf{x}_0) = \bar{y}\mathbf{1},$$

et $\hat{\mathbf{y}}(\mathbf{x}_1 | \mathbf{x}_0)$ est la projection de $\mathbf{y} - \bar{y}\mathbf{1}$ sur $\mathbf{x} - \bar{x}\mathbf{1}$. Cette projection est précisément celle qui est directement réalisée dans le cas de la régression linéaire sur les données centrées (paragraphe 4.1).

4.3. La régression multiple

Soit le modèle suivant :

$$\mathbf{y} = \mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{e} = \hat{\mathbf{y}} + \mathbf{e}$$

où \mathbf{X} est la matrice de variables explicatives. Nous ne traitons plus de manière séparée l'ordonnée à l'origine et les coefficients de régression. Si le modèle comporte un terme indépendant, il suffit d'ajouter à la matrice \mathbf{X} une colonne dont

tous les éléments sont égaux à l'unité. L'élément du vecteur $\hat{\beta}$ correspondant à cette colonne est alors l'ordonnée à l'origine.

Considérons que la matrice \mathbf{X} , de dimensions $n \times p$, est partitionnée en deux matrices, notées \mathbf{X}_1 et \mathbf{X}_2 respectivement de dimensions $n \times p_1$ et $n \times p_2$. Cela revient à dire que les variables explicatives sont réparties en deux groupes. Le vecteur $\hat{\beta}$ est partitionné de la même manière en $\hat{\beta}_1$ et $\hat{\beta}_2$. Le modèle peut alors s'écrire :

$$\mathbf{y} = \mathbf{X}_1 \hat{\beta}_1 + \mathbf{X}_2 \hat{\beta}_2 + \mathbf{e}.$$

La projection orthogonale de \mathbf{y} dans l'espace des colonnes de \mathbf{X} :

$$\hat{\mathbf{y}} = \mathbf{P}\mathbf{y} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y},$$

peut être décomposée en deux parties : la projection de $\hat{\mathbf{y}}$ dans le sous-espace des colonnes de \mathbf{X}_1 , notée $\hat{\mathbf{y}}(\mathbf{X}_1)$, et la projection de $\hat{\mathbf{y}}$ dans le sous-espace complément orthogonal à \mathbf{X}_1 , notée $\hat{\mathbf{y}}(\mathbf{X}_2 | \mathbf{X}_1)$. On a :

$$\hat{\mathbf{y}} = \hat{\mathbf{y}}(\mathbf{X}_1) + \hat{\mathbf{y}}(\mathbf{X}_2 | \mathbf{X}_1) = \mathbf{P}_1(\hat{\mathbf{y}}) + (\mathbf{I} - \mathbf{P}_1)\hat{\mathbf{y}}.$$

La projection de $\hat{\mathbf{y}}$ sur le sous-espace engendré par \mathbf{X}_1 est égale à la projection de \mathbf{y} sur ce même sous-espace. Il en résulte que :

$$(\mathbf{I} - \mathbf{P}_1)\hat{\mathbf{y}} = \mathbf{P}\mathbf{y} - \mathbf{P}_1\mathbf{y} = (\mathbf{P} - \mathbf{P}_1)\mathbf{y}.$$

La décomposition de $\hat{\mathbf{y}}$ peut donc s'écrire :

$$\hat{\mathbf{y}} = \mathbf{P}_1\mathbf{y} + (\mathbf{P} - \mathbf{P}_1)\mathbf{y},$$

et la décomposition de la somme des carrés des écarts correspondante s'écrit :

$$\|\hat{\mathbf{y}}\|^2 = \|\hat{\mathbf{y}}(\mathbf{X}_1)\|^2 + \|\hat{\mathbf{y}}(\mathbf{X}_2 | \mathbf{X}_1)\|^2;$$

$\|\hat{\mathbf{y}}\|^2$ est la somme des carrés de \mathbf{y} liée à la régression de \mathbf{y} sur \mathbf{X} ; $\|\hat{\mathbf{y}}(\mathbf{X}_1)\|^2$ est la somme des carrés de \mathbf{y} liée à la régression de \mathbf{y} sur \mathbf{X}_1 et $\|\hat{\mathbf{y}}(\mathbf{X}_2 | \mathbf{X}_1)\|^2$ est la somme des carrés de \mathbf{y} liée à la prise en compte de \mathbf{X}_2 quand on a déjà tenu compte de \mathbf{X}_1 . Les trois vecteurs en question se trouvent respectivement dans des espaces de dimension p , p_1 et p_2 . Ces dimensions sont les degrés de liberté associés aux sommes des carrés correspondantes.

La norme du vecteur \mathbf{e} est la somme des carrés des résidus. Ce vecteur se trouve dans un espace de dimension $n - p$. Le nombre de degrés de liberté associé à la somme des carrés des résidus est donc égal à $n - p$.

Disposant des sommes des carrés, on peut tester l'hypothèse de nullité des coefficients de régression partiels relatifs aux variables constituant \mathbf{X}_2 :

$$F_{obs} = \frac{\|\hat{\mathbf{y}}(\mathbf{X}_2 | \mathbf{X}_1)\|^2 / p_2}{\|\mathbf{e}\|^2 / (n - p)}.$$

La partition de la matrice \mathbf{X} en deux sous-matrices \mathbf{X}_1 et \mathbf{X}_2 peut évidemment se faire de différentes manières. Le choix d'une décomposition particulière dépend de l'hypothèse qu'on souhaite tester. Deux partitions sont particulièrement utiles en pratique. Dans le premier cas, la matrice \mathbf{X}_2 ne contient qu'une seule variable. La décomposition décrite ci-dessus permet alors de tester la signification du coefficient de régression partiel correspondant à la variable en question. Dans le deuxième cas, la matrice \mathbf{X}_2 contient les $p - 1$ variables explicatives, à l'exclusion de la colonne contenant les valeurs unitaires destinée à prendre en compte l'ordonnée à l'origine. La décomposition correspondante permet de tester globalement la nullité des $p - 1$ coefficients de régression. Elle permet aussi d'obtenir le coefficient de détermination multiple. Ce coefficient est, en effet, le carré de la corrélation entre \mathbf{y} et $\hat{\mathbf{y}}$ soit :

$$R^2 = \frac{(\langle \mathbf{y} - \bar{y}\mathbf{1}, \hat{\mathbf{y}} - \bar{y}\mathbf{1} \rangle)^2}{\|\mathbf{y} - \bar{y}\mathbf{1}\|^2 \|\hat{\mathbf{y}} - \bar{y}\mathbf{1}\|^2}$$

le vecteur $\mathbf{y} - \bar{y}\mathbf{1}$ est égal au vecteur $\mathbf{y} - \hat{\mathbf{y}}(\mathbf{X}_1)$ et le vecteur $\hat{\mathbf{y}} - \bar{y}\mathbf{1}$ est égal au vecteur $\hat{\mathbf{y}}(\mathbf{X}_2 | \mathbf{X}_1)$. La racine carrée du coefficient de détermination multiple est égale au cosinus de l'angle formé par les vecteurs $\mathbf{y} - \bar{y}\mathbf{1}$ et $\hat{\mathbf{y}}(\mathbf{X}_2 | \mathbf{X}_1)$.

A titre d'illustration, nous considérons l'exemple proposé par DAGNELIE [1982] concernant l'étude de la relation entre le rendement du froment d'hiver (\mathbf{y} , en quintaux par hectare) et quatre variables météorologiques (\mathbf{x}_1 = précipitations des mois de novembre et décembre, en mm; \mathbf{x}_2 = température moyenne du mois de juillet, en degrés centigrades; \mathbf{x}_3 = précipitations du mois de juillet, en mm; \mathbf{x}_4 = radiations du mois de juillet, en ml d'alcool mesurés à l'actinomètre de Bellani). Les données concernent onze années successives, de l'année culturale 1920-1921 à l'année culturale 1930-1931. Elle sont présentées dans le tableau 1. Les mêmes données ont déjà été utilisées par PALM et IEMMA [1995], pour illustrer quelques alternatives à la régression classique dans le cas de la colinéarité.

La juxtaposition des quatre variables explicatives \mathbf{x}_1 , \mathbf{x}_2 , \mathbf{x}_3 et \mathbf{x}_4 et d'une colonne ne contenant que les éléments unitaires, \mathbf{x}_0 , donne la matrice \mathbf{X} .

Le modèle de régression s'écrit :

$$\mathbf{y} = \mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{e} \quad \text{avec} \quad \hat{\boldsymbol{\beta}}' = (\hat{\beta}_0 \hat{\beta}_1 \dots \hat{\beta}_4),$$

$\hat{\beta}_0$ étant l'ordonnée à l'origine.

Nous considérons les deux partitions suivantes :

$$\mathbf{X} = (\mathbf{X}_1 \quad \mathbf{x}_4) \quad \text{avec} \quad \mathbf{X}_1 = (\mathbf{x}_0 \quad \mathbf{x}_1 \quad \mathbf{x}_2 \quad \mathbf{x}_3)$$

et
$$\mathbf{X} = (\mathbf{x}_0 \quad \mathbf{X}_2) \quad \text{avec} \quad \mathbf{X}_2 = (\mathbf{x}_1 \quad \mathbf{x}_2 \quad \mathbf{x}_3 \quad \mathbf{x}_4).$$

La première partition permet de tester la signification du coefficient de régression partielle relatif à la variable x_4 et la deuxième partition permet de tester la signification globale de la régression.

Tableau 1. Variable à expliquer \mathbf{y} et variables explicatives $\mathbf{x}_0, \mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3$ et \mathbf{x}_4 (SC : somme des carrés).

i	\mathbf{y}	\mathbf{x}_0	\mathbf{x}_1	\mathbf{x}_2	\mathbf{x}_3	\mathbf{x}_4
1	28,37	1	87,9	19,6	1,0	1661
2	23,77	1	89,9	15,2	90,1	968
3	26,04	1	153,0	19,7	56,6	1353
4	25,74	1	132,1	17,0	91,0	1293
5	26,68	1	88,8	18,3	93,7	1153
6	24,29	1	220,9	17,8	106,9	1286
7	28,00	1	117,7	17,8	65,5	1104
8	28,37	1	109,0	18,3	41,8	1574
9	24,96	1	156,1	17,8	57,4	1222
10	21,66	1	181,5	16,8	140,6	902
11	24,37	1	181,4	17,0	74,3	1150
SC	7287,24	-	-	-	-	-

Tableau 2. Vecteur des valeurs estimées $\hat{\mathbf{y}}$, vecteur des résidus \mathbf{e} et décomposition du vecteur $\hat{\mathbf{y}}$ selon deux partitions des variables explicatives (SC = somme des carrés).

i	$\hat{\mathbf{y}}$	\mathbf{e}	$\hat{\mathbf{y}}(\mathbf{X}_1)$	$\hat{\mathbf{y}}(\mathbf{x}_4 \mathbf{X}_1)$	$\hat{\mathbf{y}}(\mathbf{x}_0)$	$\hat{\mathbf{y}}(\mathbf{X}_2 \mathbf{x}_0)$
1	29,41	-1,04	29,32	0,09	25,66	3,75
2	24,58	-0,81	24,64	-0,07	25,66	-1,08
3	26,73	-0,69	26,97	-0,24	25,66	1,07
4	25,39	0,35	24,95	0,44	25,66	-0,27
5	26,46	0,22	26,50	-0,05	25,66	0,80
6	23,77	0,52	23,44	0,33	25,66	-1,89
7	25,82	2,18	26,26	-0,44	25,66	0,16
8	27,72	0,65	27,24	0,48	25,66	2,07
9	25,44	-0,48	25,72	-0,27	25,66	-0,21
10	22,73	-1,07	22,83	-0,10	25,66	-2,93
11	24,20	0,17	24,38	-0,19	25,66	-1,46
SC	7278,01	9,23	7277,09	0,92	7242,28	35,73

Le tableau 2 reprend le vecteur $\hat{\mathbf{y}}$, le vecteur \mathbf{e} et les décompositions du vecteur $\hat{\mathbf{y}}$ résultant des deux partitions de \mathbf{X} définies ci-dessus. Il donne également les sommes des carrés des éléments de ces vecteurs. Pour la première partition, la décomposition de la somme des carrés s'écrit :

$$\|\mathbf{y}\|^2 = \|\hat{\mathbf{y}}(\mathbf{X}_1)\|^2 + \|\hat{\mathbf{y}}(\mathbf{x}_4 | \mathbf{X}_1)\|^2 + \|\mathbf{e}\|^2,$$

soit : $7287,24 = 7277,09 + 0,92 + 9,23$

La valeur F_{obs} relative au test de signification du coefficient de régression partielle, β_4 , est égale à :

$$F_{obs} = \frac{0,92}{9,23/(11-5)} = 0,60,$$

et possède 1 et 6 degrés de liberté.

Pour la deuxième partition, on a :

$$\|\mathbf{y}\|^2 = \|\hat{\mathbf{y}}(\mathbf{x}_0)\|^2 + \|\hat{\mathbf{y}}(\mathbf{X}_2 | \mathbf{x}_0)\|^2 + \|\mathbf{e}\|^2$$

soit : $7287,24 = 7242,28 + 35,73 + 9,23$.

La valeur F_{obs} relative à la signification globale de la régression est égale à :

$$F_{obs} = \frac{35,73/4}{9,23/(11-5)} = 5,81,$$

et possède 4 et 6 degrés de liberté.

Le coefficient de détermination multiple est donné par :

$$R^2 = \frac{\|\hat{\mathbf{y}}(\mathbf{X}_2 | \mathbf{x}_0)\|^2}{\|\mathbf{y} - \bar{y}\mathbf{1}\|^2} = \frac{35,73}{44,96} = 0,795.$$

L'angle formé par les vecteurs $\mathbf{y} - \bar{y}\mathbf{1}$ et $\hat{\mathbf{y}}(\mathbf{X}_2 | \mathbf{x}_0)$ est donc égal à $\arccos \sqrt{0,795}$, soit 27° environ.

4.4. Régression et décomposition par valeurs singulières

Nous avons vu, au paragraphe 2.9, que la décomposition par valeurs singulières d'une matrice consiste à factoriser cette matrice de la manière suivante :

$$\mathbf{X} = \mathbf{U} \mathbf{S} \mathbf{V}'$$

Cette décomposition peut être utilisée dans le cadre de la régression :

$$\mathbf{y} = \mathbf{U} \mathbf{S} \mathbf{V}' \hat{\boldsymbol{\beta}} + \mathbf{e}.$$

On peut montrer que, dans ces conditions, le vecteur des coefficients est donné par la relation :

$$\hat{\boldsymbol{\beta}} = \mathbf{V} \mathbf{S}^{-1} \mathbf{U}' \mathbf{y},$$

et les projecteurs permettant d'obtenir $\hat{\mathbf{y}}$ et \mathbf{e} sont :

$$\mathbf{P} = \mathbf{U} \mathbf{U}' \quad \text{et} \quad (\mathbf{I} - \mathbf{P}) = (\mathbf{I} - \mathbf{U} \mathbf{U}').$$

Cette décomposition par valeurs singulières de \mathbf{X} est à la base de la régression sur composantes principales, qui peut s'écrire :

$$\mathbf{y} = \mathbf{U}\hat{\boldsymbol{\alpha}} + \mathbf{e}$$

avec :

$$\hat{\boldsymbol{\alpha}} = \mathbf{S}\mathbf{V}'\hat{\boldsymbol{\beta}} = \mathbf{U}'\mathbf{y}.$$

On notera que les colonnes de \mathbf{U} ne sont pas les composantes principales habituelles de variance égale à la valeur propre mais bien les composantes standardisées de manière à ce que la somme des carrés des éléments soit égale à l'unité.

L'élément α_j du vecteur $\boldsymbol{\alpha}$ est simplement le produit scalaire du vecteur \mathbf{y} et du vecteur \mathbf{u}_j correspondant à la $j^{\text{ème}}$ colonne de \mathbf{U} . La matrice de variances et covariances du vecteur $\boldsymbol{\alpha}$ est égale à :

$$\mathbf{V}(\hat{\boldsymbol{\alpha}}) = \mathbf{I}\sigma^2,$$

ce qui signifie que les coefficients α_j ont même variance et sont non corrélés.

Le vecteur $\hat{\boldsymbol{\beta}}$ peut être exprimé en fonction de $\hat{\boldsymbol{\alpha}}$ par la relation :

$$\hat{\boldsymbol{\beta}} = (\mathbf{S}\mathbf{V}')^{-1}\hat{\boldsymbol{\alpha}} = \mathbf{V}'^{-1}\mathbf{S}^{-1}\hat{\boldsymbol{\alpha}} = \mathbf{V}\mathbf{S}^{-1}\hat{\boldsymbol{\alpha}},$$

la matrice \mathbf{V} étant orthogonale, et, pour un coefficient donné, on a :

$$\hat{\beta}_j = \sum_{k=1}^p v_{jk} \frac{\hat{\alpha}_k}{s_k}.$$

Il en résulte que la variance de $\hat{\beta}_j$ est égale à :

$$v(\hat{\beta}_j) = \sigma^2 \sum_{k=1}^p \frac{v_{jk}^2}{s_k^2},$$

puisque les $\hat{\alpha}_k$ sont de variance égale à σ^2 . Cette relation montre que si une valeur singulière, s_k , est petite par rapport aux autres, le coefficient $\hat{\alpha}_k$ contribue d'une façon importante à la variance de $\hat{\beta}_j$. Une telle situation se rencontre si les variables explicatives présentent des relations approximativement linéaires. On dit, dans ce cas, qu'il y a colinéarité approximative et que la matrice \mathbf{X} est mal conditionnée.

L'effet de la colinéarité approximative, dans le cas de l'exemple du paragraphe 4.3, est illustré dans PALM et IEMMA [1995].

S'il existe, entre les colonnes de \mathbf{X} , q relations linéaires exactes, le nombre de valeurs singulières non nulles sera égal à $r = p - q$. Les relations données ci-dessus pour $\hat{\boldsymbol{\alpha}}$ et $\hat{\boldsymbol{\beta}}$ restent cependant valables. Le vecteur $\hat{\boldsymbol{\alpha}}$ contiendra r éléments et le vecteur $\hat{\boldsymbol{\beta}}$ contiendra p éléments. La solution obtenue par l'intermédiaire des composantes principales traduit alors ce phénomène de colinéarité.

Géométriquement, l'existence d'une relation linéaire exacte entre les variables dans le cas de deux variables explicatives, \mathbf{x}_1 et \mathbf{x}_2 , signifie que ces deux variables correspondent à deux vecteurs formant un angle nul. Ces deux variables n'engendrent donc pas un espace de dimension deux, c'est-à-dire un plan, mais un espace de dimension un, c'est-à-dire une droite. La projection orthogonale du vecteur \mathbf{y} sur cette droite est unique, mais cette projection peut s'exprimer comme une fonction de \mathbf{x}_1 et \mathbf{x}_2 d'une infinité de manières : le vecteur des coefficients de régression est indéterminé.

La généralisation au cas de p variables est immédiate : s'il existe q relations linéaires entre les variables, l'espace engendré par ces p variables est le sous-espace de dimension $r = p - q$.

5. QUELQUES INFORMATIONS COMPLÉMENTAIRES

Dans cette note, nous nous sommes principalement intéressés à quelques aspects géométriques liés à la régression. Des informations plus détaillées peuvent être trouvées notamment dans les articles de MANDEL [1982] et de NELDER [1985], ainsi que dans l'ouvrage de SAVILLE et WOOD [1993].

La régression linéaire faisant partie du modèle linéaire, il est naturel que divers aspects qui ont été examinés se retrouvent dans le cadre de l'analyse de la variance. Dans ce cas, la matrice des variables explicatives \mathbf{X} , appelée matrice du plan expérimental, n'est en général pas de rang égal au nombre de colonnes et les phénomènes de colinéarité sont plus souvent la règle que l'exception. Nous n'approfondissons pas ici ce problème qui a été illustré dans une autre note [IEMMA et PALM, 1992]. Des informations complémentaires relatives à l'utilisation de la géométrie dans le contexte de l'analyse de la variance sont données par EUBANK et WEBSTER [1985], HERR [1980], IEMMA *et al.* [1993 a,b], LANCASTER [1965], LARA et IEMMA [1998 a,b], LIMA FILHO [1981], MARGOLIS [1979], SAVILLE et WOOD [1993], SCHEFFE [1959] et SCHEY [1985].

L'approche géométrique est aussi très largement utilisée, par de nombreux auteurs français notamment, pour la présentation de l'analyse des données. Dans ce contexte, on peut citer, par exemple, les ouvrages de BENZÉCRI et BENZÉCRI [1980], BERTIER et BOUROCHE [1975], BOUROCHE et SAPORTA [1980], CAILLIEZ et PAGES [1976], LEBART *et al.* [1995], ainsi que l'article de synthèse de PAGES *et al.* [1979].

Un autre domaine qui n'a pas été abordé est l'approche géométrique de la notion de variable aléatoire, qui permet de définir, par exemple, l'espérance mathématique d'un produit de deux variables comme un produit scalaire, le coefficient de corrélation de deux variables comme le cosinus d'un angle, l'espérance mathématique conditionnelle comme une projection orthogonale. Des informations à ce sujet sont données par BRYANT [1984], GRAYBILL [1993], LARA et IEMMA [1998 a,b], LOWERE [1982], SAPORTA [1978] et SAVILLE et WOOD [1986; 1993], notamment.

Enfin, parmi les applications récentes de la géométrie, on peut encore mentionner l'utilisation des projections orthogonales pour l'identification de groupes

d'individus [PENA et PRIETO, 2001], et l'utilisation de l'analyse "AMMI" pour l'interprétation des interactions génotype-milieu [DUARTE et WENCOVSKY, 1999].

6. CONCLUSIONS

Dans cette note, nous avons d'abord présenté un bref rappel de quelques notions de calcul vectoriel. Ces notions ont ensuite été utilisées pour la présentation des aspects géométriques liés à la régression.

Comme nous l'avons déjà signalé dans l'introduction, l'étude de la régression est possible sans cette approche géométrique. Certains ouvrages entièrement consacrés à la régression ne font d'ailleurs pas allusion à cet aspect géométrique [CHATTERJEE et PRICE, 1991; RYAN, 1997; WEISBERG, 1985] ou bien y consacrent un ou deux chapitres, qui peuvent être omis en première lecture [DRAPER et SMITH, 1998].

Cette note donne à l'utilisateur une vue un peu moins habituelle d'une méthode statistique couramment utilisée dans la pratique. L'approche géométrique permet en effet de mieux comprendre certains problèmes de la régression, comme, par exemple, les phénomènes de colinéarité et, surtout, elle ouvre l'accès à tout un secteur de la bibliographie susceptible d'intéresser les utilisateurs souhaitant approfondir leur maîtrise de la régression.

BIBLIOGRAPHIE

- BENZÉCRI J.P., BENZÉCRI F. [1980]. *Pratique de l'analyse des données. Vol. 1 : Analyse des correspondances, exposé élémentaire*. Paris, Dunod, 424 p.
- BERTIER P., BOUROCHE J.M. [1975]. *Analyse des données multidimensionnelles*. Paris, Presses Universitaires de France, 270 p.
- BOUROCHE J.M., SAPORTA G. [1980]. *L'analyse des données*. Paris, Presses Universitaires de France, 127 p.
- BRYANT P. [1984]. Geometry, statistics, probability: variations on a common theme. *Amer. Stat.* 38, 1, 38-48.
- CAILLIEZ F., PAGES J.P. [1976]. *Introduction à l'analyse des données*. Paris, SMASH, 616 p.
- CHATTERJEE S., PRICE B. [1991]. *Regression analysis by example*. New York, Wiley, 298 p.
- DAGNELIE P. [1982]. *Analyse statistique à plusieurs variables*. Gembloux, Presses agronomiques, 362 p.
- DRAPER S., SMITH H. [1998]. *Applied regression analysis*. New York, Wiley, 706 p.
- DUARTE J.B., WENCOVSKY R. [1999]. *Interação genótipos x ambientes: uma introdução à análise "AMMI"*. Soc. Brasileira de Genética - Série Monografias. Ribeirão Preto, 60 p.

- ECKART C., YOUNG G. [1936]. The approximation of a matrix by another of lower rank. *Psychometrika* 1, 211-218.
- EUBANK R.L., WEBSTER J.B. [1985]. The singular-value decomposition as a tool for solving estimability problems. *Amer. Stat.* 38, 2, 94-105.
- FISHER R.A. [1915]. Frequency distribution of the value of the correlation coefficient in samples from an indefinitely large population. *Biometrika* 10, 507-521.
- GRAYBILL F.A. [1983]. *Matrices with applications in statistics*. Belmont, Wadsworth, 372 p.
- HERR D.G. [1980]. On the history of the use of geometry in the general linear model. *Amer. Stat.* 34, 1, 43-47.
- IEMMA A.F., PALM R. [1992]. Les matrices inverses généralisées et leur utilisation dans le modèle linéaire. *Notes Stat. Inform.* (Gembloux) 92/1, 25 p.
- IEMMA A.F., PALM R. [1993]. Sobre sistemas lineares homogêneos, matrizes de Helmert e raízes características de matrizes uniformes. *Seminarios em estatística e experimentação agrônoma da ESALQ/USP*. DME, 22 p.
- IEMMA A.F., PALM R., ALVEZ M.I.F. [1993a]. Orthogonal projectors and statistical hypothesis about unbalanced data. *Biom. Bull.* 10, 1, p. 18 (abstracts).
- IEMMA A.F., PALM R., CLAUSTRIAUX J.J. [1993b]. Sobre a construção de projetores ortogonais. *Revista Matemática Estatística II*, 133-142.
- KRUSKAL W.H. [1975]. The geometry of generalized inverses. *J. R. Stat. Soc., Ser. B*, 37, 272-283.
- LANCASTER H.O. [1965]. The Helmert matrices. *Amer. Math. Monthly* 72, 1, 4-12.
- LARA I.A.R., IEMMA A.F. [1998a]. Geometria : um recurso didático para ensinar estatística. Anais de resumos - 2º CONAPE. São Carlos, p. 286.
- LARA I.A.R., IEMMA A.F. [1998b]. Tópicos de geometria com aplicações em estatística 13º SINAPE. São Paulo, p. 95.
- LEBART L., MORINEAU A., PIRON M. [1995]. *Statistique exploratoire multidimensionnelle*. Paris, Dunod, 439 p.
- LIMA FILHO E.C. [1981]. *Planejamento de observações adicionais em modelos lineares : decomposição em valores singulares da matriz de planejamento e sua caracterização*. (Thèse). USP/São Carlos, 38 p.
- LOWERE J.M. [1982]. An introduction to modern matrix methods and statistics. *Amer. Stat.* 36, 2, 113-115.
- MANDEL J. [1982]. Use of the singular value decomposition in regression analysis. *Amer. Stat.* 36, 1, 15-24.
- MARGOLIS M.S. [1979]. Perpendicular projections and elementary statistics. *Amer. Stat.* 33, 3, 131-135.
- NELDER J.A. [1985]. An alternative interpretation of the singular-value decomposition in regression. *Amer. Stat.* 39, 1, 63-64.

- PAGES J.P., CAILLIEZ F., ESCOUFIER Y. [1979]. Analyse factorielle : un peu d'histoire et de géométrie. *Rev. Stat. Appl.* 27 (1), 5-28.
- PALM R., IEMMA A.F. [1995]. Quelques alternatives à la régression classique dans le cas de la colinéarité. *Rev. Stat. Appl.* 43, 2, 5-33.
- PENA D., PRIETO F.J. [2001]. Cluster identification using projections. *J. Amer. Stat. Assoc.* 96, 1433-1445.
- RYAN T.P. [1997]. *Modern regression methods*. New York, Wiley, 515 p.
- SAPORTA G. [1978]. *Théories et méthodes de la statistique*. Paris, Technip, 386 p.
- SAVILLE D.J., WOOD G.R. [1993]. A method for teaching statistics using n-dimensional geometry. *Amer. Stat.* 40, 3, 205-214.
- SAVILLE D.J., WOOD G.R. [1986]. *Statistical methods : the geometric approach*. New York, Springer, 561 p.
- SCHEFFÉ H. [1959]. *The analysis of variance*. New York, Wiley, 471 p.
- SCHEY H.M. [1985]. A geometric description of orthogonal contrasts in one way analyses of variance. *Amer. Stat.* 39, 2, 104-106.
- WEISBERG S. [1985]. *Applied linear regression*. New York, Wiley, 324 p.