Une méthode d'investigation : la régression orthogonale

R. Tomassone

▶ **To cite this version:**

R. Tomassone. Une méthode d'investigation : la régression orthogonale. Annales des sciences forestières, 1967, 24 (3), pp.233-258. 10.1051/forest/19670303 . hal-00881975

HAL Id: hal-00881975
https://hal.science/hal-00881975

Submitted on 11 May 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Ann. Sci. forest., 1967, **24** (3), 233-258.

UNE MÉTHODE D'INVESTIGATION : LA RÉGRESSION ORTHOGONALE

R. TOMASSONE

Station de Biométrie,
Centre national de Recherches forestières, 54 - *Nancy*
Institut national de la Recherche agronomique

SOMMAIRE

On étudie l'analyse de régression orthogonale de façon à la fois théorique et pratique : seules les démonstrations qui permettent de mieux comprendre les conditions exactes d'application sont données. On expose les liens avec les méthodes classiques de la régression et de l'analyse des composantes principales, et on dégage l'originalité de la méthode en insistant tout particulièrement sur son intérêt comme outil d'investigation dans la première approche d'un problème nouveau. L'exemple traité étudie la nature des relations entre une variable de production et des variables écologiques.

1. — INTRODUCTION

1.1. — *Méthode d'investigation et méthode statistique*

Lorsqu'un chercheur travaillant sur un matériel vivant désire le confronter avec un modèle mathématique donné a priori, les résultats sont souvent décevants ; rarement, les hypothèses indispensables pour que ce modèle soit applicable sont vérifiées. La confiance que l'on peut avoir dans les résultats de l'analyse statistique est naturellement très limitée. Pourtant, rares sont les publications où, si une interprétation statistique a été effectuée, on ne fasse état de « tests de signification » et « d'intervalle de confiance » ; malheureusement, ces deux caractéristiques sont généralement sensibles aux écarts par rapport aux hypothèses. Sans prendre le contre-pied systématique d'une telle attitude, nous voudrions attirer l'attention sur des techniques d'analyse qui, si elles sont moins fines, permettent en revanche, dans une première phase de recherche, de s'attacher aux causes de variation les plus importantes ; ainsi un projecteur qui n'éclaire qu'un domaine nécessairement restreint, rend quelques services s'il est braqué dans de bonnes directions. Les conclusions partielles d'une analyse de ce type peuvent ultérieurement servir de base à des études plus poussées : des hypothèses formulées à partir de ces conclusions pourront être « testées » : les méthodes statistiques les plus rigoureuses sont alors employées de façon beaucoup plus sûre.

Article disponible sur le site http://www.afs-journal.org ou http://dx.doi.org/10.1051/forest/19670303

1.2. — *Analyse multivariate*

A l'intérieur des techniques statistiques, l'analyse multivariate jouit d'une place à part : souvent, elle peut jouer le rôle du projecteur dont nous venons de parler. Evidemment, les tests statistiques existent, et fort nombreux, mais en abordant un problème où plusieurs variates agissent de façon concomitante tellement de questions se font jour qu'il est naturel de reléguer les tests, pour un temps, au second plan.

Même si les chercheurs qui travaillent sur un matériel vivant reconnaissent qu'il n'est pas raisonnable de traiter séparément les caractères qu'ils enregistrent, rares sont ceux qui utilisent l'analyse multivariate de façon systématique. Cette attitude est due, en grande partie, à ce que par sa nature l'analyse multivariate nécessite des calculs qui, sans être toujours complexes, sont régulièrement fastidieux. A l'heure où n'importe qui, à condition de le vouloir, peut s'assurer les services d'un bureau de calcul doté d'un ordinateur, cette attitude n'est plus admissible.

Dans cet article, nous voudrions faire connaître une méthode peu connue qui fait appel simultanément à l'analyse des composantes principales et à l'analyse de régression. Nous allons voir que l'idée générale qui conduit à cette méthode est relativement simple et pourtant nous en avons trouvé peu d'applications publiées. A titre documentaire, notons que STONE J.R.N. (1945) l'a utilisée dans une étude de consommation de bière en fonction de divers critères économiques, que MASSY W.F. (1965) l'a utilisée pour analyser la demande de certains biens de consommation en fonction des revenus, enfin JEFFERS J.N.R. et SEALE B.E. (1966) dans l'étude d'un problème forestier.

1.3. — *Principe de l'analyse de régression orthogonale*

Nous appellerons tout au long de notre exposé *variables explicatives* un ensemble de p variables $x_1, x_2, \ldots x_p$, et variable expliquée [1] une variable y que nous essaierons d'estimer au moyen des variables x_i $(i = 1, p)$ par une équation linéaire de la forme :

$$y = b_0 + b_1 x_1 + b_2 x_2 + \ldots + b_p x_p \qquad (1)$$

où $b_1, b_2, \ldots b_p$ sont les coefficients de régression que nous nous proposons d'estimer, b_0 le coefficient constant.

Nous supposons que toutes les transformations inspirées, soit pour des raisons statistiques (validité d'hypothèses concernant les variables) soit pour des raisons physiques (connaissance a priori d'une forme d'équation) ont été effectuées, ce sont donc ces nouvelles variables qui figurent dans l'équation (1). L'objectif de l'analyse de la régression orthogonale est double :

[1] Nous préférons ces deux termes à ceux de variables indépendantes et de variable dépendante communément employés, car ils reflètent mieux, à la fois l'objet de l'analyse et les liens entre les deux types de variables ; de plus l'adjectif indépendant s'adapte mal à des variables qui ne le sont en général pas au sens des probabilités. On trouve quelquefois aussi le terme de « régresseurs » pour désigner ces dernières, cf. DUMAS DE RAULY D. (1966).

1° transformer les variables x_i ($i = 1, p$) en un nouvel ensemble de variables explicatives. Nous choisirons les composantes principales z_j ($j = 1, p$) qui possèdent la propriété d'être indépendantes deux à deux ;

2° effectuer une régression de la variable y en fonction des z_j ($j = 1, p$) ; ces variables étant indépendantes, chacune d'elle apporte à l'explication de y une contribution qui lui est propre (1).

1.4. — *Intérêt de cette méthode*

Il est rare dans les applications réelles de ne pas pouvoir interpréter assez aisément les premières composantes principales (2) il est donc particulièrement intéressant de chercher les relations entre y et un nombre de variables explicatives inférieur à p ; cet aspect est particulièrement important dans des analyses d'investigation où le choix des x_i ($i = 1, p$) est très souvent dicté par des raisons empiriques ; la question essentielle est de savoir combien choisir de variables pour expliquer convenablement y, ce choix permettant d'éliminer des variables n'ayant aucun effet sur elle.

Nous verrons qu'en plus, la méthode que nous allons exposer possède des avantages précieux d'un strict point de vue numérique (3).

1.5. — *Présentation de l'exposé*

Nous utiliserons tout au long de notre exposé la notation matricielle ; nous adopterons toutefois une approche heuristique en ne formulant rigoureusement que ce qui est strictement nécessaire à la compréhension et à l'application de la méthode (4).

2. — ASPECT THÉORIQUE ET PROPRIÉTÉS

2.1. — *Position du problème et hypothèses de départ*

2.11. *Restrictions*

Nous savons que les techniques de l'analyse multivariate ne sont pas invariantes lorsqu'on effectue des changements d'échelle sur les variables observées ; les modifications des résultats de l'analyse sont en général mineures, aussi, dans un but

(1) Cette façon d'aborder la régression linéaire est comparable à celle d'aborder la régression curviligne (régression de y en fonction des puissances croissantes d'une seule variable x par un polynôme) au moyen des polynômes orthogonaux, cf. DUMAS DE RAULY D. (1966), pp. 317-321.

(2) A ce propos, cf. JEFFERS J.N.R. (1967), MASSY W.F. (1965).

(3) Ces avantages sont voisins de ceux de la méthode de régression progressive qui consiste à introduire les variables l'une après l'autre en fonction de la part propre qu'elles fournissent à l'explication de y ; cette méthode est exposée par EFROYMSON M.A. (1959) ; nous pouvons fournir des programmes pour ordinateur écrits, soit en Algol, soit en Fortran, cf. TOMASSONE R. (1967).

(4) Tous les ouvrages statistiques de base comportent une partie réservée à la présentation de l'outil matriciel, nous prions le lecteur qui aurait quelques difficultés à suivre le côté technique de notre exposé de s'y reporter ; cf. par exemple RAO C.R. (1965), SCHEFFE H. (1959), SEAL H. (1964).

de simplification, nous supposerons que toutes les variables sont centrées (c'est-à-dire de moyennes nulles) et réduites (de variance égale à 1) (1). Le centrage élimine le coefficient constant b_0 dans l'équation (1) ; tandis que la réduction revient à utiliser non pas les covariances de toutes les variables prises deux à deux mais les coefficients de corrélation.

2.12. *Modèle et hypothèses*

Nous supposons que les estimations des coefficients de régression s'effectuent à partir d'un échantillon de taille n, ce qui signifie que pour chaque élément du n-échantillon l'équation (1), (où le terme constant a été supprimé pour la raison exposée plus haut), est « pratiquement » satisfaite ; en terme statistique, nous introduisons un terme aléatoire dans cette équation et nous formulons un modèle :

$$y_j = b_1 x_{j1} + b_2 x_{j2} + \ldots + b_p x_{jp} + e_j \qquad (j = 1, n). \tag{2}$$

Dans les conditions habituelles, nous supposons que les termes e_j, qui traduisent l'écart au modèle, sont de moyenne nulle, de variance constante et non corrélés entre eux. A l'aide de l'information fournie par le n-échantillon, nous désirons obtenir une estimation des b_i ($i = 1, p$) et de la variance des fluctuations aléatoires. Pour l'obtenir, nous essayons de rendre minimale l'influence du terme aléatoire du modèle. Si on emploie la méthode d'estimation au sens des moindres carrés, on obtient cette minimisation en rendant minimale la quantité U de l'équation suivante :

$$U = \sum_{j=1}^{j=n} e_j^2$$

En terme matriciel, l'équation du modèle et les différentes hypothèses sont condensées de la façon suivante :

$$\underset{(n\times 1)}{\boldsymbol{Y}} = \underset{(n\times p)}{\boldsymbol{X}} \quad \underset{(p\times 1)}{\boldsymbol{\beta}} + \underset{(n\times 1)}{\boldsymbol{e}} \tag{4}$$

$$U = \boldsymbol{e}' \, \boldsymbol{e}_{\min} \tag{5}$$

$$E(\boldsymbol{e}) = 0\mathbf{1}_n \tag{6}$$

$$\underset{(n\times n)}{\Sigma_e} = \underset{(n\times n)}{\sigma^2 \boldsymbol{I}_n} \tag{7}$$

où — le signe $'$ après une matrice signifie que l'on a effectué une transposition (échange des lignes et des colonnes).

— Les nombres inscrits sous une matrice indiquent sa dimension (nombre de lignes et de colonnes) (2).

(1) Les modifications dues aux changements d'échelle ont rarement été étudiées de façon systématique d'un point de vue à la fois théorique et pratique, cf. à ce sujet ESCOUFFIER Y. (1966) pour quelques comparaisons.

(2) Afin de ne pas alourdir le texte, ces dimensions ne seront indiquées que lors de la première apparition de cette matrice.

— $\boldsymbol{I}_n$ est la matrice unité, carrée, à n lignes et à n colonnes, formée de 1 sur la diagonale principale et de 0 partout ailleurs.

— $\mathbf{1}_n$ un vecteur dont les n composantes sont toutes égales à 1.

— $E(\boldsymbol{e})$ représente l'estimation du vecteur $\boldsymbol{e}$.

— Σ_e la matrice des variances et covariances relative aux variables aléatoires représentées par les composantes du vecteur $\boldsymbol{e}$.

Ainsi l'équation (6) traduit le fait que les moyennes des e_j sont nulles, l'équation (7) celui de l'homoscédascité (égalité des variances) et de la non-corrélation des termes aléatoires puisque tous les termes extérieurs à la diagonale qui représentent les covariances sont nuls et ceux de la diagonale tous égaux à σ^2.

2.13. *Estimation des paramètres*

On démontre que les paramètres définis par les composantes du vecteur β sont définis par l'équation :

$$(\boldsymbol{X}'\boldsymbol{X})\boldsymbol{\beta} = \boldsymbol{X}'\boldsymbol{Y} \tag{8}$$

$\boldsymbol{X}'\boldsymbol{X}$ représente la matrice des variances et covariances des variables x_i ($i = 1, p$), et $\boldsymbol{X}'\boldsymbol{Y}$ les covariances entre y et les x_i ($i = 1, p$).

Si la matrice $(\boldsymbol{X}'\boldsymbol{X})$ possède une inverse il est alors possible :

1° d'estimer les coefficients de régression par :

$$\boldsymbol{\beta} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y} \tag{9}$$

2° d'estimer la matrice de variances et des covariances de β par :

$$\Sigma_\beta = \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1} \tag{10}$$

(la variance résiduelle σ^2 étant estimée comme nous le verrons plus loin).

Nous remarquons immédiatement les deux points essentiels qui constitueront les éléments primordiaux de notre étude :

— l'estimation des coefficients de régression et de leur matrice de variances et covariances n'est possible que si la matrice des coefficients de corrélation des variables explicatives possède une inverse ; nous noterons désormais :

$$\underset{(p\times p)}{\boldsymbol{R}} = \boldsymbol{X}'\boldsymbol{X} \tag{11}$$

et

$$\underset{(p\times 1)}{\boldsymbol{r}_y} = \boldsymbol{X}'\boldsymbol{Y} \tag{12}$$

— les estimations des coefficients sont en général corrélées puisque les éléments extérieurs à la diagonale de leurs matrices de variances et covariances sont différents de zéro.

2.14. *Test relatif aux paramètres*

On démontre que la variation totale $\boldsymbol{Y}'\boldsymbol{Y}$ de la variable expliquée peut se décomposer en deux parties :

— l'une que l'ajustement au modèle permet de contrôler, est égale à $\boldsymbol{\beta}'\boldsymbol{r}_y$,

— l'autre « inexplicable » par le modèle traduit les variations aléatoires non contrôlées, elle est égale à $(\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{\beta})'(\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{\beta})$. Cette dernière divisée par $n-p-1$ fournit une estimation sans biais de la variance résiduelle σ^2.

Jusqu'à présent, nous n'avons fait appel à aucune hypothèse : cela est assez remarquable puisque toutes les estimations sont absolument indépendantes de la forme des distributions des variables aléatoires e_j. Si nous supposons que ces distributions sont normales :

1° tout ce qui a été dit au sujet de la non-corrélation des variables peut être traduit en terme d'indépendance et les estimateurs au sens des moindres carrés sont confondus avec ceux fournis par la méthode du maximum de vraisemblance, (cf. DUMAS de RAULY D., *op. cit.*) ;

2° dans l'équation qui traduit la variation des y_j en deux parties :

$$\boldsymbol{Y}'\boldsymbol{X} = \boldsymbol{\beta}'\boldsymbol{r}_y + (\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{\beta})'(\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{\beta}) \qquad (13)$$

les termes du second membre sont distribués indépendamment : le premier comme un χ^2 à p degrés de liberté et le second comme un χ^2 à $n-p-1$ degrés de liberté. Il est alors possible d'effectuer un test analogue à celui de l'analyse de variance à un facteur contrôlé, test que nous résumons dans le tableau suivant (1) :

Source de variation	Somme des carrés	d. l.	Carré moyen
Ajustement au modèle (due à la régression)	$\boldsymbol{\beta}'\boldsymbol{r}_y$	p	σ^2_R
Résiduelle	(*)	$n-p-1$	σ^2_E
Totale	$\mathbf{Y}'\mathbf{Y}$	$n-1$	

(*) Le terme résiduel est calculé, au moins lorsque les calculs sont faits sur des machines électromécaniques, par différence.

Le test de la valeur de l'ajustement se fait au moyen du rapport des deux carrés moyens qui est un F de Snedecor à p et $n-p-1$ degrés de liberté. On rejette l'hypothèse : « l'ajustement au modèle n'est pas significatif » si cette valeur calculée est supérieure à la valeur théorique $F(p, n-p-1 ; \alpha)$ au seuil de probabilité α choisi. Malheureusement ce test est global, il ne permet pas de juger de l'amélioration apportée par chaque variable indépendamment des autres ; cela est absolument normal puisque les variables observées sont corrélées et il n'est pas possible, à cause de cette corrélation, de juger de leur effet propre. Tout au plus, est-il possible de juger de l'effet partiel de chaque variable x_i en testant son coefficient de régression β_i

(1) Il y a d'ailleurs plus identité qu'analogie : la seule différence avec l'analyse de variance tient à la nature des variables x ; continues dans la régression, discrètes (0 ou 1) dans l'analyse de variance. Seule une étude « terre à terre » des deux méthodes masque leurs liens profonds.

ou son coefficient de corrélation partielle, les expressions analytiques des deux tests statistiques étant absolument identiques. Mais ce test permet uniquement d'analyser l'effet d'une variable quelconque parmi les p variables, toutes les autres étant maintenues constantes.

2.2. — *Simplification des variables explicatives*

2.21. *Rappel sur l'analyse des composantes principales*

Dans l'esprit de l'exposé de notre méthode, nous voyons que la régression multiple sous sa forme « globale » est d'emploi délicat dans la mesure où l'action de chaque variable n'est pas indépendante de celle des autres, sauf cas d'espèce. Il est donc naturel dans une phase d'investigation d'essayer de synthétiser au maximum l'action des variables explicatives. Nous allons donc dans un premier temps porter notre attention sur les variables explicatives seules et effectuer sur elles un changement de variables, les variables transformées seront les composantes principales relatives aux variables initiales. Nous ne reviendrons pas sur cette analyse dans ses détails (1), nous n'allons en donner que les éléments nécessaires pour la suite. Faire une analyse des composantes principales revient à faire sur les variables x_i ($i = 1, p$) une transformation qui les transforment en de nouvelles variables z_j. En général, il y aura autant de variables z_j qu'il y a de variables x_i, mais :

1° les variables z_j ($j = 1, p$) sont non corrélées entre elles,

2° chaque variable z_j traduit une part positive de la variation globale dont la somme est p ; il en résulte que certaines parmi les nouvelles variables ont une participation négligeable dans la variabilité totale des variables explicatives. La transformation sur les variables x_i est définie par une matrice $\underset{(p\times p)}{\boldsymbol{M}}$ telle que pour chaque élément du n-échantillon on puisse faire correspondre aux variables x_i des variables z_j définies par :

$$\underset{(n\times p)}{\boldsymbol{Z}} = \underset{(n\times p)}{\boldsymbol{X}}\ \underset{(p\times p)}{\boldsymbol{M}} \tag{15}$$

L'estimation de la matrice des variances et covariances des z_j est obtenue en calculant $\mathrm{E}(\boldsymbol{Z}'\boldsymbol{Z})$. Soit $\boldsymbol{D}$ cette nouvelle matrice ; on voit que $\boldsymbol{M}$ étant une matrice dont les éléments sont fixés :

$$\underset{(p\times p)}{\boldsymbol{D}} = \boldsymbol{M}'\boldsymbol{R}\boldsymbol{M} \tag{16}$$

Pour que les z_j aient les propriétés énoncées plus haut, nous choisissons pour $\boldsymbol{D}$ une matrice diagonale. On sait, d'après les théorèmes connus d'algèbre matricielle, que la transformation $\boldsymbol{M}$ qui permet d'obtenir une matrice diagonale est telle que les colonnes de $\boldsymbol{M}$ sont constituées par les vecteurs propres de la matrice $\boldsymbol{R}$, et que

(1) Se reporter soit aux traités généraux déjà cités plus haut, soit aux exposés plus simples comme ceux de DEBAZAC E.F. et TOMASSONE R. (1965) ou ESCOUFFIER Y. (1966).

les valeurs propres correspondantes sont les éléments diagonaux de $\boldsymbol{D}$. Nous supposons que les valeurs propres sont placées sur la diagonale principale de la plus grande vers la plus petite ($d_{11} \geqslant d_{22} \geqslant \ldots d_{pp}$) (1). D'autre part, puisque la matrice $\boldsymbol{M}$ est une matrice orthogonale réelle son inverse est égale à sa transposée, nous pouvons donc passer très aisément des variables z_j aux variables x_i ; en effet, puisque $\boldsymbol{M}' = \boldsymbol{M}^{-1}$

$$\boldsymbol{X} = \boldsymbol{Z}\boldsymbol{M}' \tag{17}$$

2.22. *Les nouvelles variables explicatives*

L'analyse des composantes principales classique s'arrête en général à ce stade. Il est fréquent, surtout en analyse factorielle, de ne pas utiliser directement les variables z_j ($j = 1, p$) mais les variables divisées par leur écart-type, de telle sorte que les nouvelles variables que nous appellerons composantes réduites f_j ($j = 1, p$) ont toutes une variance égale à 1 (2). En terme matriciel, cette transformation s'écrit :

$$\left\{ \begin{array}{l} \underset{(n \times p)}{\boldsymbol{F}} = \boldsymbol{Z}\boldsymbol{D}^{-1/2} \\ \qquad\quad = \boldsymbol{X}\boldsymbol{M}\boldsymbol{D}^{-1/2} \end{array} \right. \tag{18}$$

selon que nous désirons exprimer les composantes réduites soit en fonction des composantes principales, soit en fonction des variables initiales. Il est souvent intéressant d'exprimer ces dernières en fonction des composantes réduites :

$$\boldsymbol{X} = \boldsymbol{F}(\boldsymbol{M}\boldsymbol{D}^{-1/2})^{-1} \tag{19}$$

Il est facile de démontrer les égalités suivantes qui nous seront utiles dans la suite de l'exposé :

$$\left\{ \begin{array}{l} \boldsymbol{A}\boldsymbol{A}' = \boldsymbol{R} \\ \boldsymbol{A} = (\boldsymbol{M}\boldsymbol{D}^{-1/2})^{-1} \\ \boldsymbol{A} = \boldsymbol{M}'\boldsymbol{D}^{1/2} \\ \boldsymbol{A}^{-1} = \boldsymbol{D}^{-1}\boldsymbol{A}' \end{array} \right. \tag{20}$$

La matrice $\boldsymbol{A}$ est couramment utilisée en analyse factorielle sous le nom de matrice des saturations ; on voit qu'elle permet de passer facilement des composantes réduites aux variables initiales par :

$$\boldsymbol{X} = \boldsymbol{F}\boldsymbol{A} \tag{21}$$

De plus, ses coefficients ont une signification statistique puisque, si nous multiplions à gauche les deux membres de l'équation (21) par $\boldsymbol{F}'$ pour estimer la matrice

(1) Selon la méthode de diagonalisation utilisée les valeurs propres sont obtenues directement par ordre décroissant (méthode de déflation, méthode de Jacobi adaptée par Von Neumann) ou non (méthode de tridiagonalisation) ; dans ce second cas, il est nécessaire de les replacer dans l'ordre dans l'étude que nous faisons.

(2) Cette transformation peut paraître étrange puisque l'utilisation des composantes est essentiellement facilitée par les valeurs décroissantes des variances ; mais c'est à cause de la facilité d'interprétation qui en découlera qu'elle est utilisée, cf. Massy W.F. (1965).

des variances et covariances des composantes et des variables, et si nous prenons les estimations de ces deux membres, nous obtenons :

$$E(\boldsymbol{F}' \boldsymbol{X}) = E(\boldsymbol{F}' \boldsymbol{F})\, \boldsymbol{A} \tag{22}$$

comme la matrice des variances et covariances des f_j est égale à la matrice $\boldsymbol{I}_p$ par définition des variables elles-mêmes [1] nous voyons que les éléments de $\boldsymbol{A}$ sont égaux aux coefficients de corrélation des variables initiales avec les composantes réduites :

$$\left\{\begin{array}{l} E(f_i x_j) = \text{coefficient de corrélation de} \\ \qquad\qquad f_i \text{ et de } x_j \\ \qquad\qquad = a_{ij} \end{array}\right. \tag{23}$$

2.3. — *Régression sur les nouvelles variables*

Etant donné les propriétés des composantes principales que nous avons rappelées, nous voyons que l'étude de la régression de y avec les f_j sera intéressante dans la mesure où elle permettra « d'épurer » l'analyse :

— soit parce que les composantes n'étant pas corrélées il sera plus facile d'analyser leur action indépendamment les unes des autres,

— soit parce que le nombre des variables explicatives étant élevé, il peut être bon de diminuer d'abord la dimension de cette variation : plus petite sera cette dimension, plus simple sera l'interprétation. Formellement, le nouveau modèle dont nous nous proposons d'estimer les paramètres est analogue à celui de l'équation (4) mais les variables f_j remplacent les variables x_i :

$$\underset{n\times 1}{\boldsymbol{Y}} = \underset{(n\times p)}{\boldsymbol{F}} \quad \underset{(p\times 1)}{\boldsymbol{\gamma}} + \underset{(n\times 1)}{\boldsymbol{e}} \tag{24}$$

où $\boldsymbol{\gamma}$ est un vecteur à p dimensions dont les composantes sont les coefficients de régression de y en fonction des composantes principales réduites. L'estimation de $\boldsymbol{\gamma}$ au sens des moindres carrés s'obtient très facilement à l'aide de l'équation (8) où la matrice $\boldsymbol{X}$ est remplacée par la matrice $\boldsymbol{F}$, comme nous avons construit $\boldsymbol{F}$ de telle sorte que $\boldsymbol{F}'\,\boldsymbol{F} = \boldsymbol{I}_p$

$$\left\{\begin{array}{l} \boldsymbol{\gamma} = \boldsymbol{F}'\,\boldsymbol{Y} \\ \quad = (\boldsymbol{X}\boldsymbol{A}^{-1})'\,\boldsymbol{Y} = (\boldsymbol{X}\boldsymbol{D}^{-1}\boldsymbol{A}')'\,\boldsymbol{Y} \\ \boldsymbol{\gamma} = \boldsymbol{A}\boldsymbol{D}^{-1}(\boldsymbol{X}'\,\boldsymbol{Y}) = \boldsymbol{A}\boldsymbol{D}^{-1}\boldsymbol{r}_y \end{array}\right. \tag{25}$$

La première des équations (25) nous permet de dire que les coefficients de régression γ_j représentent les coefficients de corrélation de y avec la composante f_j.

Parallèlement, la matrice des variances et covariances de γ s'obtient grâce à l'homologue de l'équation (10), et toujours en vertu des propriétés des variables f_j cette matrice prend une forme particulièrement simple :

$$\Sigma_\gamma = \sigma^2 \boldsymbol{I}_p \tag{26}$$

[1] Ceci ne serait pas exact en analyse factorielle où les 1 de la diagonalisation principale sont remplacés par les communautés.

cette dernière équation signifie que non seulement les coefficients γ_j sont non corrélés, mais aussi qu'ils ont tous la même variance σ^2. En fonction des coefficients de la matrice $\boldsymbol{A}$ et des valeurs propres de $\boldsymbol{R}$ on retrouve aisément l'expression

$$\gamma_j = \frac{1}{d_{jj}} \sum_{i=1}^{i=p} a_{ji} r_{iy} \tag{27}$$

où r_{iy} est le coefficient de corrélation de y avec la variable explicative x_i. Il ne nous reste donc plus qu'à estimer la variance résiduelle en calculant la part de la variation totale de y (ici supposée égale à 1) que la régression permet de contrôler ; cette part peut se décomposer en p parties indépendantes car :

$$\begin{aligned} \boldsymbol{\gamma}' \boldsymbol{F}' \boldsymbol{Y} &= \boldsymbol{\gamma}' \boldsymbol{\gamma} \\ &= \gamma_1^2 + \gamma_2^2 + \ldots \gamma_p^2 \end{aligned} \tag{28}$$

Chacune de ces parts donne une mesure de l'explication fournie par chaque composante. Un tableau d'analyse de variance calqué sur le tableau (14) permet de synthétiser l'ensemble de ces résultats.

Source de variation	Somme des carrés	d. l.	Carré moyen
Due à la régression ..	$\gamma'\gamma = \gamma^2{}_1 + \gamma^2{}_2 + \ldots \gamma^2{}_p$	p	$\sigma^2{}_R$
composante 1	$\gamma^2{}_1$	1	$\sigma^2{}_{R1}$
composante 2	$\gamma^2{}_2$	1	$\sigma^2{}_{R2}$
.	.	.	.
.	.	.	.
composante p	$\gamma^2{}_p$	1	$\sigma^2{}_{Rp}$
Résiduelle	(par différence)	$n-p-1$	$\sigma^2{}_E$
Totale	1	$n-1$	

(29)

On voit alors qu'il est facile d'estimer :

— le coefficient de corrélation multiple R^2 qui mesure le pourcentage de la variation totale imputable à la régression :

$$R^2 = \gamma_1^2 + \gamma_2^2 + \ldots \gamma_p^2 \tag{30}$$

— l'écart type résiduel

$$\sigma_E^2 = \frac{1 - R^2}{n - p - 1} \tag{31}$$

Si nous voulons ensuite étudier les écarts entre les valeurs de la variable explicative et les valeurs estimées au moyen de l'équation de régression, c'est la valeur précédente multipliée par $(n-1)$ qu'il faut utiliser car la variation totale de y est égale à $\Sigma y^2{}_i = (n-1)$, puisque les y ont été réduits.

2.4. — *Retour aux variables initiales*

Avant d'aborder les propriétés découlant de notre étude, nous voulons donner ici les expressions des paramètres β_i coefficients de régression sur les variables initiales. Quel que soit l'élément du n-échantillon $E(Y)$ est fournie par une des deux équations : (4) si l'on désire exprimer y en fonction des x_i, (24) si c'est en fonction des f_j ; ainsi

$$\begin{aligned} \boldsymbol{\beta} &= \boldsymbol{D}^{-1}\boldsymbol{A}'\boldsymbol{\gamma} \\ &= \boldsymbol{A}^{-1}\boldsymbol{\gamma} = \boldsymbol{A}^{-1}(\boldsymbol{A}^{-1})'\boldsymbol{r}_y \\ &= \boldsymbol{R}^{-1}\boldsymbol{r}_y \end{aligned} \tag{32}$$

nous retrouvons bien l'estimation classique des coefficients de régression sur les variables réduites. On peut s'assurer que la part de la variation imputable à la régression est bien la même selon que l'on utilise l'un ou l'autre des groupes de variables explicatives. Enfin, la matrice des variances et covariances des β_i s'obtient

$$\begin{aligned} \Sigma_\beta &= \boldsymbol{E}(\boldsymbol{\beta}\boldsymbol{\beta}') = \boldsymbol{E}(\boldsymbol{D}^{-1}\boldsymbol{A}'\boldsymbol{\gamma}(\boldsymbol{D}^{-1}\boldsymbol{A}'\boldsymbol{\gamma})') \\ &= \boldsymbol{D}^{-1}\boldsymbol{A}'\ \boldsymbol{E}(\boldsymbol{\gamma}\boldsymbol{\gamma}')\ \boldsymbol{A}\boldsymbol{D}^{-1} \\ &= \sigma_E^2\boldsymbol{A}'\boldsymbol{D}^{-2}\boldsymbol{A} \end{aligned} \tag{33}$$

soit

$$\operatorname{cov}(\beta_i\beta_j) = \sigma_E^2 \sum_k \frac{a_{ki}a_{kj}}{d_{kk}^2} \ . \tag{34}$$

Toutes ces quantités, comme nous le verrons plus loin, peuvent se calculer sans aucune difficulté sur un ordinateur.

2.5. — *Propriétés de l'analyse de régression orthogonale*

Au paragraphe 2.13, nous avons dit qu'il était nécessaire, pour calculer les coefficients de régression β_i, d'inverser la matrice des coefficients de corrélation. Si cette inversion est impossible, les coefficients β_i sont indéterminés. Cette situation peut se réaliser lorsqu'il existe des relations linéaires entre les variables x_i, en d'autres termes lorsque le rang de la matrice des coefficients de corrélation est inférieur à p. Dans la pratique, ce cas extrême est relativement rare ; par contre, il peut exister des liaisons suffisamment fortes entre les variables pour que numériquement il ne soit pas possible d'obtenir une précision suffisante dans le calcul d'inversion. Nous allons examiner successivement ces deux cas et voir comment la régression orthogonale permet d'étudier ces problèmes.

2.51. *Prise en compte des liaisons linéaires entre variables*

Lorsque le rang de la matrice des coefficients de corrélation est q ($q < p$), parmi les p valeurs propres, $p-q$ sont nulles et la dimension de l'espace de variation des x est q. En d'autres termes, l'espace est réduit ; par exemple, si nous avons deux

variables x_1 et x_2 et s'il existe une liaison linéaire entre les deux, la variation des x au lieu de se situer dans un plan sera concentrée sur une droite, et c'est en fonction de cette nouvelle variable que nous pourrons essayer d'expliquer la variable y (1).

Dans ce cas, la matrice A n'est plus carrée mais rectangulaire à q lignes et p colonnes. On peut alors démontrer que toutes les estimations faites sur les γ_j sont valables, mais nous calculons q seulement de ces quantités ; il suffit d'effectuer une rotation dans l'espace initial pour se placer dans l'espace des vecteurs propres dont les valeurs propres sont différentes de zéro et il est alors possible d'estimer les β_i dans ce nouvel espace (2).

2.52. *Choix des variables explicatives*

Nous avons dit plus haut que l'interprétation serait d'autant plus facile que la dimension de l'espace initial des variables explicatives serait plus réduite. C'est-à-dire que nous porterions notre attention sur un nombre limité de composantes. Deux attitudes sont alors possibles pour éliminer des composantes :

1. *Suppression en fonction des valeurs propres*

Nous nous plaçons dans ce cas dans l'espace des variables explicatives et nous ne conservons des p composantes que celles ayant une part de variation significative. Cette attitude, dictée par le bon sens, se justifie dans la mesure où l'on désire expliquer y en fonction de variables interprétables (3) ; en effet, l'expérience pratique que nous avons de l'analyse des composantes principales nous permet de penser que, dans la majorité des cas, ce sont les premières composantes que l'on peut interpréter de la façon la plus satisfaisante. De plus :

— la précision numérique des faibles valeurs propres est souvent mal connue, il n'est pas sage d'essayer d'interpréter des composantes alors que des variations, mêmes faibles sur les variables de base, pourraient les modifier profondément,

— les erreurs d'échantillonnage de ces nouvelles variables n'ont pas été étudiées.

Les règles de suppression sont empiriques ; théoriquement, lorsque les distributions sont normales, il est possible, une fois q valeurs propres calculées, de tester l'égalité des $p-q$ suivantes ; si cette hypothèse est vérifiée, il n'existe pas de direction privilégiée dans l'espace des $p-q$ vecteurs propres (4) et l'on dit alors que les variations y sont purement aléatoires.

---

(1) Pour une discussion plus détaillée et une représentation graphique, cf. NAMKOONG G., (1967).

(2) Pour les études d'inverse généralisée d'une matrice cf. RAO C.R. (1965) et pour l'application d'inverse à droite ou à gauche de matrice rectangulaire cf. MASSY W.F. (1965). Ce problème peut être étudié d'une façon différente en ramenant sa résolution à celui de l'estimation des paramètres dans un modèle linéaire lorsque l'on impose des contraintes à ces paramètres, cf. JUDGE C.C. and TAKAYAMA T. (1966), NAGAR A.L. and KAKWANI N.C. (1965), THEIL H. (1963).

(3) Dans l'analyse d'investigation où nous nous plaçons il peut de toute façon être intéressant de voir que y n'est pas lié à des « facteurs » facilement interprétables, ou éventuellement que la forme linéaire pour l'étude de la liaison n'est pas adéquate.

(4) Pour plus de détails, cf. SEAL H. (1964) citant le test proposé par Barlett et amélioré par Lawley.

Mais, ce test n'est valable que dans l'hypothèse où la distribution est p-normale ; dans les cas pratiques, il est prudent de l'utiliser avec précaution, c'est-à-dire de s'en servir comme d'un indicateur sans attacher trop d'importance au seuil de probabilité choisi. Pratiquement, nous préférons choisir une règle plus simple et ne conserver que les vecteurs propres dont les valeurs propres sont supérieures à l'unité [1] ; dans ce cas, nous pouvons dire que nous conservons les composantes dont la contribution à la variabilité est supérieure à celle des variables prises individuellement.

2. *Suppression en fonction des coefficients gamma*

Rien ne nous permet de supposer, a priori, que y est liée à des composantes ayant une part importante dans la variabilité ; il est logique de ne conserver dans la régression sur les composantes réduites que les composantes pour lesquelles la valeur du coefficient de régression gamma est significative.

2.53. *Différences entre les deux stratégies*

Analytiquement, les deux problèmes sont identiques : nous nous plaçons dans des sous-espaces Eq de Ep ($q < p$) ; mais ces sous-espaces ne sont pas les mêmes, et rien ne nous permet de supposer que les résultats soient voisins :

— dans le premier cas, nous axons notre recherche sur les relations éventuelles entre y et des composantes ayant une signification physique ;

— dans le second, nous essayons de chercher uniquement des liaisons sans nous préoccuper de leur interprétation.

Dans les deux cas, nous ne conservons dans la matrice $\boldsymbol{A}$ qu'un certain nombre de lignes sélectionnées suivant un des deux critères énoncés ci-dessus. Une des équations (32) permet de calculer les p coefficients $\beta_i (i = 1, p)$. D'autre part, l'équation (21) fournit la part de variation de chacune des variables x_i ($i = 1, p$) qui entre en compte dans la variabilité totale : c'est la somme des carrés des coefficients des q composantes réduites qui sont conservées relatifs à la variable x_i [2].

3. — EXEMPLE

Nous avons choisi [3] un exemple d'étude de la liaison entre treize variables de milieu (variables explicatives) et des variables de production (variables expliquées). Nous avons essayé d'expliquer les cinq variables de production séparément en fonction du même ensemble de variables de milieu [4].

---

(1) ou voisines, s'il existe des valeurs propres de l'ordre de 0.8 ou 0.9.

(2) En analyse factorielle, cette quantité est égale à la communauté.

(3) Nous remercions à la fois M. GODRON du Centre d'Etudes phytosociologiques et écologiques, et M. DECOURT de la Station de Sylviculture et Production du C.N.R.F. de nous avoir communiqué ces éléments.

(4) Il aurait été possible d'analyser les deux ensembles globalement par une analyse canonique cf. SEAL H. (1964) mais nous avons volontairement limité notre étude à la régression orthogonale.

3.1. — *Présentation des données*

Variables explicatives

1 Position topographique
2 Pente
3 Recouvrement de la strate muscinale
4 — — intermédiaire
5 — — dominante
6 Couvert angulaire
7 Epaisseur de l'horizon A_0
8 Epaisseur de l'horizon A_1
9 Profondeur de l'horizon engorgé en hiver
10 Profondeur de l'horizon argileux
11 Activité biologique
12 Humidité
13 Age.

Variables expliquées

— Indice de fertilité (hauteur à cinquante ans)
— Nombre d'arbres à l'hectare
— Surface terrière moyenne
— Volume
— Hauteur moyenne

3.2. — *Résultats*

Nous ne désirons pas entrer dans les détails de l'interprétation biologique qui seront l'objet d'une publication commune ; nous voulons simplement montrer :

— la valeur synthétique de la méthode,

— que des variables expliquées différentes sont liées de façon diverse à un même ensemble de variables explicatives.

Il est assez frappant, dans le tableau 1, que seules les variables 11 (activité biologique) ou 13 (âge) aient un coefficient de régression partielle significatif, et même si nous ne portons notre attention que sur les valeurs très élevées du t de Student, seul l'âge paraît avoir une influence lorsque les autres facteurs sont maintenus constants.

Par contre, les résultats de la régression orthogonale, bien qu'ils soient délicats à interpréter, offrent une diversité plus grande. Si nous portons notre attention sur

les résultats concernant la hauteur moyenne (bas du tableau 2) les composantes 2, 3, 6 et 7 paraissent exercer une influence :

— la *seconde*, à laquelle la hauteur est liée positivement, fait intervenir en plus de l'âge (coefficient positif) le recouvrement de la strate intermédiaire (positif), le recouvrement de la strate dominante (négatif), le couvert angulaire (négatif), les épaisseurs des horizons A_0 et A_1 (positif) ;

— la *troisième* (liaison positive) est interprétable en fonction de la position topographique et de la pente ;

— la *sixième* (liaison négative) fait intervenir l'âge (négatif) mais aussi les variables X_1, X_3, X_4 et X_8 qui s'opposent à l'effet de l'âge ;

— la *septième* (liaison positive) associe positivement le couvert angulaire de façon importante puis l'âge et l'activité biologique.

TABLEAU 1

Résultats de la régression multiple normale :
coefficients de régression (β_i) et test de Student ($t(\beta_i)$)

variables expliquées / variables explicatives	Indice		Nombre		Surface terrière		Volume		Hauteur	
	β_i	$t(\beta_i)$	β_i	$t(\beta_i)$	β_i	$t(\beta_i)$	β_i	$t(\beta_i)$	β_i	$t(\beta_i)$
X 1	−0.132	0.55	−0.058	0.30	−0.003	0.01	−0.023	0.14	−0.020	0.16
X 2	0.100	0.41	0.044	0.23	−0.103	0.45	−0.030	0.18	0.017	0.14
X 3	−0.085	0.37	0.020	0.11	−0.008	0.03	0.032	0.20	−0.011	0.10
X 4	−0.059	0.25	−0.003	0.01	−0.156	0.71	−0.132	0.85	−0.062	0.54
X 5	0.091	0.32	0.040	0.17	0.068	0.25	−0.020	0.10	0.041	0.29
X 6	0.000	0.00	0.154	0.83	0.098	0.44	0.080	0.51	−0.005	−0.04
X 7	0.271	0.87	−0.026	0.10	0.277	0.93	0.248	1.18	0.152	0.98
X 8	0.101	0.43	−0.049	0.26	0.037	0.16	0.022	0.14	0.037	0.32
X 9	0.017	0.04	−0.009	0.03	0.111	0.32	0.133	0.55	0.054	0.30
X10	−0.054	0.18	0.178	0.77	0.042	0.15	−0.014	0.07	−0.089	0.62
X11	0.790	2.38*	−0.085	0.32	0.365	1.16	0.371	1.66	0.350	2.12*
X12	−0.281	0.86	0.039	0.15	−0.217	0.70	−0.229	1.04	−0.157	0.97
X13	−0.243	0.97	−0.694	3.46**	0.604	2.55*	0.792	4.69**	0.888	7.13**

Nous voyons que la régression orthogonale permet de séparer des contributions propres à chaque composante. Ainsi la variable X13, qui paraît jouer un rôle important dans cette étude, apparaît différemment dans les deux méthodes :

— dans la régression multiple habituelle son effet est fourni globalement,

— dans la régression orthogonale, les différents aspects de sa contribution sont mis en évidence dans les diverses composantes qui ont un effet significatif.

TABLEAU 2

Résultats de la régression orthogonale (coefficients γ_i)

Composante i	1	2	3	4	5	6	7	8	9	10	11	12	13		
valeur propre	2.96157	1.74137	1.54278	1.41930	1.24487	0.96063	0.74179	0.63290	0.56259	0.41151	0.38772	0.21400	0.17873		
% cumulé	22.78	36.18	48.04	58.96	68.54	75.93	81.63	86.50	90.83	93.99	96.98	98.62	100.00		
X 1	−0.09	0.05	0.58	0.17	0.13	0.39	−0.29	0.34	0.27	0.12	−0.37	−0.12	−0.13		
X 2	−0.08	−0.02	0.64	−0.20	0.02	0.15	0.07	−0.24	−0.56	0.11	0.35	0.05	0.11		
X 3	0.04	−0.20	−0.21	0.33	−0.53	0.39	0.13	0.29	−0.48	0.04	−0.14	0.01	−0.14		
X 4	0.13	0.41	0.04	0.08	−0.42	0.43	0.15	−0.38	0.44	−0.09	0.27	0.03	−0.07		
X 5	−0.36	−0.34	0.18	0.27	−0.07	−0.04	−0.19	−0.18	0.04	−0.68	0.06	0.34	0.03		
X 6	−0.10	−0.40	0.20	0.38	0.07	0.17	0.62	−0.25	0.22	0.29	−0.09	−0.13	−0.09		
X 7	0.11	0.33	0.04	0.66	0.02	−0.22	−0.17	0.01	−0.10	0.11	0.07	−0.04	0.57		
X 8	−0.01	0.29	−0.20	0.16	0.59	0.37	0.23	−0.28	−0.27	−0.19	−0.31	0.13	−0.10		
X 9	−0.48	0.25	0.01	0.04	−0.09	−0.14	−0.06	−0.14	−0.20	−0.16	0.04	−0.73	−0.23		
X10	−0.42	0.12	−0.02	−0.17	−0.30	−0.07	−0.17	−0.40	−0.08	0.39	−0.48	0.31	0.14		
X11	−0.44	0.05	−0.02	−0.25	0.01	0.25	0.43	0.33	0.10	−0.11	0.05	−0.05	0.60		
X12	−0.44	0.25	−0.05	0.18	0.15	−0.05	0.03	0.28	−0.01	0.30	0.46	0.41	−0.37		
X13	0.15	0.42	0.30	−0.07	−0.22	−0.43	0.39	0.24	−0.07	−0.30	−0.31	0.20	−0.17	R^2	σ^2
indice γ_i	−0.404	−0.086	−0.082	−0.091	0.165	0.195	0.198	0.024	−0.021	−0.115	0.052	−0.085	0.337	0.428	0.04095
$t(\gamma_i)$	1.99	0.42	0.40	0.44	0.81	0.96	0.97	0.11	0.10	0.56	0.25	0.41	1.66		
nombre γ_i	−0.314	−0.496	−0.287	0.111	0.073	0.198	−0.206	−0.231	0.060	0.201	0.091	−0.031	0.025	0.633	0.02629
$t(\gamma_i)$	1.93	3.05**	1.77	0.68	0.44	1.21	1.26	1.42	0.37	1.24	0.55	0.18	0.15		
surface γ_i	−0.132	0.307	0.182	0.093	−0.098	−0.321	0.301	0.169	0.040	−0.181	−0.221	−0.027	0.137	0.488	0.03665
$t(\gamma_i)$	0.68	0.60	0.94	0.48	0.51	1.67	1.57	0.88	0.21	0.94	1.15	0.13	0.71		
volume γ_i	0.016	0.438	0.266	0.021	−0.170	−0.363	0.396	0.221	−0.093	−0.198	−0.222	−0.036	0.116	0.741	0.01858
$t(\gamma_i)$	0.11	3.21**	1.95	0.15	1.24	2.66*	2.90*	1.62	0.68	1.44	1.62	0.26	0.84		
hauteur γ_i	0.097	0.505	0.338	−0.066	−1.058	−0.337	0.406	0.248	−0.072	−0.263	−0.175	0.022	0.078	0.859	0.01009
$t(\gamma_i)$	0.96	5.02**	3.36**	0.65	1.57	3.35**	4.04**	2.46*	0.71	2.62*	1.74	0.21	0.77		

Reçu pour publication en octobre 1967

SUMMARY

AN INVESTIGATORY METHOD : THE ORTHOGONAL REGRESSION.

The theoretical and the practical aspects of the orthogonal regression are studied : the only demonstrations which are necessary for the understanding of the conditions of application are given. The links with the classical methods of regression and principal component analysis are set off ; the particular interest as an investigatory tool for the first approach of a new problem are stromgly emphasyzed. An example treats of the relations between a production's variable and a set of ecological data.

ZUSAMMENFASSUNG

DIE ORTHOGONALE REGRESSION.

In der vorliegenden Arbeit wird die Analyse der orthogonalen Regression sowohl theoretisch als auch an einem praktischen Beispiel untersucht ; es werden dabei nur die für das Verständnis der exakten Anwendungsbedingungen notwendigen Demonstrationen angeführt.

Die Beziehungen mit den klassischen Methoden der Regressions — und Prinzipalkomponentenanalyse werden dargestellt und die Originalität der Methode hervorgehoben. Die Bedeutung dieser Methode als Hilfsmittel für eine erste, angenäherte Untersuchung eines neuen Problemes wird besonders unterstrichen. Das angeführte Beispiel behandelt die Art der Beziehungen zwischen einer Zuwachsgrösse und ökologischen Variabeln.

RÉFÉRENCES BIBLIOGRAPHIQUES

ANDERSSON T.W., 1958. — *Introduction to Multivariate Statistical Analysis*, John Wiley, New York.

DEBAZAC E.F. et TOMASSONE R., 1965. — Contribution à une étude comparée des pins méditerranéens de la section halepensis. *Ann. Sci. forest.*, **22**, (2), 215-256.

DIXON W.J., 1965. — *Regression on principal components, dans Biomedical Computer Programs*, pp. 159-168. University of California, Los Angeles.

DUMAS DE RAULY D., 1966. — *L'estimation statistique*. Gauthier-Villars, Paris.

EFROYMSON M.A., 1959. — Multiple regression analysis, dans RALSTON A. et WILF H.S. *Mathematical Methods for Computers*, John Wiley, New York.

ESCOUFIER Y., 1966. — *Analyse des composantes principales : utilisation des groupes de variables dans la recherche de la solution.* Thèse 3e cycle, Montpellier.

JEFFERS J.N.R., 1963. — Orthogonalysed regression. *Pegasus Autocode Program n° 22, Stn. Sec. Forest. Com. Alice Holt.*

JEFFERS J.N.R., 1967. — Multivariate analysis of progeny and provenance trials. *I.U.F.R.O., Munich Section*, **22**.

JEFFERS J.N.R. and SEALE B.E., 1966. — A multivariate analysis of relationship between staff and work-load, dans Institutionen för skolig matematisk statistik, Skogshökolan. *Research Note n° 9, Stockholm*, pp. 170-203.

JUDGE C.C. and TAKAYAMA T., 1966. — Inequality restrictions in regression analysis. *Jal. Amer. Stn. Ass.*, **61**, (315), 166-181.

KENDALL M.G., 1957. — *A course in multivariate analysis*. Griffin, Londres.

MASSY W.F., 1965. — Principal components regression in exploratory statistical research. *Jal. Amer. Stn. Ass.*, **60**, 309, pp. 234-256.

NAGAR A.L. and KAKWANI N.C., 1965. — Note on the use of prior information in statistical estimation of economic relations. *Sankhya Série A*, **27**, (1), 105-112.

NAMKOONG G., 1967. — Multivariate methods for multiple regression provenance analysis. *I.U.F.R.O. Munich, Section*, **22**, 308-318.

RAO C.R., 1965. — *Linear statistical inference and its applications*. John Wiley, New York.

SCHEFFE H., 1959. — *The analysis of variance*. John Wiley, New York.

SEAL H., 1964. — *Multivariate statistical analysis for biologist*. Methuen, Londres.

STONE J.R.N., 1945. — The analysis of market demand. *Jal. R. Stn. Soc.* (*A*), 286-382.

THEIL H., 1963. — On the use of incomplete prior information in regression analysis. *Jal. Amer. Stn. Ass.*, **58**, 401-414.

TOMASSONE R., 1967. — Regression multiple progressive, Station de Biométrie, *C.N.R.F. Programme AG.*, 65.007.

WILLIAMS E.J., 1967. — The analysis of association among many variates. *Jal. R. Stn. Soc.* (*B*)29, **2**, 199-242.

ANNEXE

PRÉSENTATION DU PROGRAMME DE RÉGRESSION ORTHOGONALE

La suite des opérations effectuées par le programme est indiquée dans les cartes « commentaire » ; nous invitons donc le lecteur qui voudrait suivre le détail des opérations à s'y reporter. Nous voulons simplement donner ici une aide pour des modifications éventuelles que certains voudraient apporter au programme.

1. Le programme a été testé sur l'ordinateur 1130 IBM de la Station de Biométrie du C.N.R.F. (8 K à disque). Le disque est utilisé si on désire calculer les valeurs estimées par l'équation et les composantes principales de chaque élément de l'échantillon étudié. (Programme écrit en FORTRAN).

2. Il fait appel à quatre subroutines ou fonctions :

a) la fonction permettant le calcul de la transformation [1] Arcsin $\sqrt{x}$, qui n'est pas une fonction standard sur le système 1130 ..... carte 128

b) calcul des valeurs propres et des vecteurs propres d'une matrice réelle symétrique (EIGEN) ..... carte 191

c) recherche du nombre de valeurs propres supérieures à une valeur fixée par l'utilisateur (TRACE) ..... carte 199

d) formation de la matrice A cf. texte (LOAD) ..... carte 223

3. Pour atteindre la taille indiquée (25 variables explicatives), il faut utiliser sur le 1130 la possibilité du traitement LOCAL. Douze constantes seulement peuvent être incluses dans les transformations.

4. Si on désire obtenir les éléments A et non les composantes des vecteurs propres, il suffit de placer les cartes 210-217 après la carte 223.

5. Si les résidus sont désirés, ce sont les composantes réduites que le programme calcule ; si on veut les composantes principales, il faut effectuer auparavant l'inverse de l'opération effectuée par le subroutine LOAD.

[1] Toutes les transformations possibles grâce à ce programme permettent de recouvrir un grand nombre de cas intéressants dans la pratique statistique.

```
      DIMENSION DATA(26),ITRAN(26),JTRAN(26),KTRAN(26),LTRAN(26),XBAR
     1(26),SIGMA(26),D(25),GAMMA(25),CONST(12),AID(18),B(25),R(325),V
     2(625)
      COMMON MX,MY
      DEFINE FILE 1(1000,60,U,IFA)
    1 FORMAT(2I2)
    2 FORMAT(18A4)
    3 FORMAT(I5,4I2,F6.0,I1)
    4 FORMAT(36I2)
    5 FORMAT(12F6.3)
    6 FORMAT(12F6.0)
    7 FORMAT(1H1,18A4)
    8 FORMAT(34H VAR        MOYENNES    ECART-TYPES,/)
    9 FORMAT(1H ,I3,2F15.3)
   10 FORMAT(41H0 MATRICE DES COEFFICIENTS DE CORRELATION,/)
   11 FORMAT(/16H VALEURS PROPRES/(10F12.5))
   12 FORMAT(/19H POURCENTAGE CUMULE/(10F12.5))
   13 FORMAT(/17H VECTEURS PROPRES)
   14 FORMAT(/18H COMPOSANTE NUMERO,I3/(10F12.5))
   15 FORMAT(//8H RCARRE=,F10.5,20HVARIANCE RESIDUELLE=,F10.5)
   16 FORMAT(27H COMPOSANTE GAMMA    STUDENT)
   17 FORMAT(1H ,I8,F9.5,F10.2)
   18 FORMAT(27H VARIABLE    BETA    STUDENT)
   19 FORMAT(23H  OBS   YOBS YCALC   DIFF,16(4H    Z,I2),/,23X,14(4H    Z,I2
     1))
   20 FORMAT(I5,19F6.2,/,23X,14F6.2)
   21 FORMAT(1H 20F6.2)
C
C        LECTURE DES UNITES DE SORTIE ET D'ENTREE
C
      READ(2,1)MX,MY
  999 IFA=1
C
C        LECTURE DE LA CARTE ENTETE, CODE ALPHANUMERIQUE DES
C        COLONNES 1 A 72
C
      READ(MY,2)(AID(I),I=1,18)
C
C        LECTURE DE LA CARTE PARAMETRE CONTENANT
C             COLONNE 1 A 5 ....NOMBRE D'OBSERVATIONS
C             COLONNES 6 - 7 ....NOMBRE DE VARIABLES OBSERVEES
C             COLONNES 8 - 9 ....NOMBRE DE VARIABLES DANS LA REGRESSION
C             COLONNES 10-11 .NOMBRE DE TRANSFORMATIONS SUR LES
C             VARIABLES
C             COLONNES 12-13 ....NOMBRE DE CONSTANTES DANS LES
C             TRANSFORMATIONS
C             COLONNES 14-19 ....VALEUR MINIMUM DE LA VALEUR PROPRE
C             A CONSERVER
C             COLONNE 20     ....1 SI ON VEUT CALCULER LES RESIDUS
C                             ...0 SINON
C
      READ(MY,3)NOBS,NVIN,NP,NTRAN,NCON,CON,IRES
      NVAR=NP+1
      IF(NTRAN)730,730,700
C
C        LECTURE DES CARTES TRANSFORMATION S'IL Y EN A
C        ITRAN(I) DEFINIT LE TYPE DE TRANSFORMATION
C               (VOIR ADRESSE 761 A 771)
C        JTRAN EST LE NOUVEL INDICE DE LA VARIABLE QUI REMPLACE
C             SOIT KTRAN (VARIABLE)
C             SOIT KTRAN ET LTRAN (VARIABLE)
C             SOIT KTRAN ET LA CONSTANTE LTRAN
C
  700 READ(MY,4)(ITRAN(I),JTRAN(I),KTRAN(I),LTRAN(I),I=1,NTRAN)
      IF(NCON)730,730,720
C
C       LECTURE DES CARTES CONSTANTES S'IL Y EN A
```

```
C                                                                       68
  720 READ(MY,5)(CONST(I),I=1,NCON)                                     69
C                                                                       70
C       INITIALISATION                                                  71
C                                                                       72
  730 OBS=NOBS                                                          73
      DO 90 I=1,NVIN                                                    74
      SIGMA(I)=0.0                                                      75
   90 XBAR(I)=0.0                                                       76
      NP1=(NP*NP+NP)/2                                                  77
      DO 91 I=1,NP1                                                     78
   91 R(I)=0.0                                                          79
      DO 92 I=1,NP                                                      80
   92 B(I)=0.0                                                          81
C                                                                       82
C       LECTURE DES DONNEES,TRANSFORMATION S'IL Y A LIEU ET FORMATION   83
C       DES PREMIERES STATISTIQUES                                      84
C                                                                       85
      DO 110 I=1,NOBS                                                   86
      READ(MY,6)(DATA(J),J=1,NVIN)                                      87
      IF(NTRAN)860,860,750                                              88
C                                                                       89
C       TRANSFORMATION DES DONNEES INITIALES                            90
C                                                                       91
  750 DO 850 M=1,NTRAN                                                  92
      II=ITRAN(M)                                                       93
      JJ=JTRAN(M)                                                       94
      KK=KTRAN(M)                                                       95
      LL=LTRAN(M)                                                       96
      GO TO(761,762,763,764,765,766,767,768,769,770,771),II             97
C                                                                       98
C       LES ONZE TRANSFORMATIONS POSSIBLES SONT FONCTION DU CODE DE LA  99
C       CARTE TRANSFORMATION                                           100
C                                                                      101
  761 DATA(JJ)=DATA(KK)                                                102
      GO TO 850                                                        103
  762 DATA(JJ)=-DATA(KK)                                               104
      GO TO 850                                                        105
  763 DATA(JJ)=ALOG(DATA(KK))                                          106
      GO TO 850                                                        107
  764 DATA(JJ)=1.0/DATA(KK)                                            108
      GO TO 850                                                        109
  765 DATA(JJ)=DATA(KK)+DATA(LL)                                       110
      GO TO 850                                                        111
  766 DATA(JJ)=DATA(KK)*DATA(LL)                                       112
      GO TO 850                                                        113
  767 DATA(JJ)=DATA(KK)/DATA(LL)                                       114
      GO TO 850                                                        115
  768 DATA(JJ)=DATA(KK)+CONST(LL)                                      116
      GO TO 850                                                        117
  769 DATA(JJ)=DATA(KK)+CONST(LL)                                      118
      GO TO 850                                                        119
  770 DATA(JJ)=(DATA(KK))**.5                                          120
      GO TO 850                                                        121
C                                                                      122
C       LA FONCTION ASRAC EST LA FONCTION ARCSINUS RACINE CARREE (X)   123
C       OU X EST EXPRIMEE EN POURCENTAGE, C'EST LA SEUILE FONCTION UTI-124
C       LISEE DANS LES TRANSFORMATIONS QUI N'EST PAS UNE FONCTION      125
C       STANDARD SUR LE 1130 IBM                                       126
C                                                                      127
  771 DATA(JJ)=ASRAC(DATA(KK))                                         128
  850 CONTINUE                                                         129
C                                                                      130
C       SI ON DESIRE CALCULER LES ESTIMEES ET LES COMPOSANTES PRINCIPA-131
C       LES ON ECRIT LES DONNEES SUR LE DISQUE                         132
C                                                                      133
  860 IF(IRES)880,880,870                                              134
  870 WRITE(1'IFA)(DATA(J),J=1,NVAR)                                   135
  880 DO 100 J=1,NVAR                                                  136
      XBAR(J)=XBAR(J)+DATA(J)                                          137
  100 SIGMA(J)=SIGMA(J)+DATA(J)*DATA(J)                                138
      DO 101 J=1,NP                                                    139
      DO 101 K=J,NP                                                    140
C                                                                      141
C       L'ADRESSE DE LA LIGNE J ET DE LA COLONNE K (K SUG J) D'UNE     142
C     MATRICE STOCKEE SUIVANT LE MODE1 DES S.S.P. EST K(K-1)/2+J       143
C                                                                      144
```

```
      L=(K*K-K)/2+J                                                     145
  101 R(L)=R(L)+DATA(J)*DATA(K)                                         146
C                                                                       147
C       ON STOCKE UNIQUEMENT LES ELEMENTS NECESSAIRES AU CALCUL DES     148
C       COEFFICIENTS DE CORRELATION DE LA VARIABLE EXPLIQUEE AVEC LES   149
C       VARIABLES EXPLICATIVES                                          150
C                                                                       151
      DO 102 J=1,NP                                                     152
  102 B(J)=B(J)+DATA(J)*DATA(NVAR)                                      153
  110 CONTINUE                                                          154
C                                                                       155
C       CALCUL DES ECART-TYPE ET DES COEFFICIENTS DE CORRELATION        156
C                                                                       157
      DO 120 I=1,NVAR                                                   158
  120 SIGMA(I)=(SIGMA(I)-XBAR(I)*XBAR(I)/OBS)**.5                       159
      DO 121 I=1,NP                                                     160
      DO 121 J=I,NP                                                     161
      K=(J*J-J)/2+I                                                     162
  121 R(K)=(R(K)-XBAR(I)*XBAR(J)/OBS)/(SIGMA(I)*SIGMA(J))               163
      DO 122 J=1,NP                                                     164
  122 B(J)=(B(J)-XBAR(J)*XBAR(NVAR)/OBS)/(SIGMA(J)*SIGMA(NVAR))         165
C                                                                       166
C       OASSAGE A LA PAGE SUIVANTE (SI MX=3) ET IMPRESSION DES          167
C       PREMIERES STATISTIQUES                                          168
C                                                                       169
      WRITE(MX,7)(AID(I),I=1,18)                                        170
      WRITE(MX,8)                                                       171
      DO 123 I=1,NVAR                                                   172
      XBAR(I)=XBAR(I)/OBS                                               173
      SIGMA(I)=SIGMA(I)/(OBS-1.0)**.5                                   174
  123 WRITE(MX,9)I,XBAR(I),SIGMA(I)                                     175
      WRITE(MX,10)                                                      176
      DO 124 J=1,NP                                                     177
      DO 125 I=1,J                                                      178
      L=I+(J*J-J)/2                                                     179
  125 D(I)=R(L)                                                         180
  124 WRITE(MX,21)(D(I),I=1,J)                                          181
      WRITE(MX,21)(B(I),I=1,NP)                                         182
      MV=0                                                              183
C                                                                       184
C       CALCUL DES VECTEURS PROPRES ET DES VALEURS PROPRES D'UNE MATRICE 185
C       R SYMETRIQUE NPXNP SEULE LA PARTIE SUPERIEURE DROITE EST FOURNIE 186
C       APRES LE CALCUL LES NP VALEURS PROPRES SONT SUR LA DIAGONALE DE 187
C       R; LES VECTEURS PROPRES CORRESPONDANTS SONT LES COLONNES DE V   188
C       MV EST UN PARAMETRE PROPRE A LA SUBROUTINE CF. SSP IBM 1130 NOTI 189
C       CE PAGES 62-63                                                  190
C                                                                       191
      CALL EIGEN(R,V,NP,MV)                                             192
C                                                                       193
C       RECHERCHE DES VALEURS PROPRES SUPERIEURES A LA VALEUR CON DONNEE 194
C       LE NOMBRE DE CES VALEURS PROPRES CALCULEES PAR LA SUBROUTINE EST 195
C       K, LES POURCENTAGES CUMULES SONT LES COMPOSANTES DU VECTEUR D   196
C       (DIMENSION NP)                                                  197
C                                                                       198
      CALL TRACE (NP,R,CON,K,D)                                         199
C                                                                       200
C       IMPRESSION DES VALEURS PROPRES DU POURCENTAGE CUMULE ET DES     201
C       VECTEURS PROPRES                                                202
C                                                                       203
      DO 130 I=1,K                                                      204
      L=I+(I*I-I)/2                                                     205
  130 DATA(I)=R(L)                                                      206
      WRITE(MX,11)(DATA(J),J=1,K)                                       207
      WRITE(MX,12)(D(J),J=1,K)                                          208
      WRITE(MX,13)                                                      209
      L=0                                                               210
      DO 150 J=1,K                                                      211
      DO 140 I=1,NP                                                     212
      L=L+1                                                             213
  140 D(I)=V(L)                                                         214
  150 WRITE(MX,14)J,(D(I),I=1,NP)                                       215
      DO 160 J=1,K                                                      216
  160 GAMMA(J)=0.0                                                      217
C                                                                       218
C       FORMATION DE LA MATRICE DES FACTEURS:CHAQUE COMPOSANTE D'UN     219
C       VECTEUR PROPRE EST MULTIPLIEE PAR LA RACINE CARREE DE SA VALEUR 220
C       PROPRE                                                          221
```

```
C                                                                       222
                                                                        223
      CALL LOAD(NP,K,R,V)                                               224
C                                                                       225
C       A PARTIR DE LA MATRICE DES FACTEURS V,DES VALEURS PROPRES DES    226
C       COMPOSANTES ET DES  COEFFICIENTS DE CORRELATION ENTRE LA VARIA-  227
C       BLE EXPLIQUEE ET LES VARIABLES EXPLICATIVES CALCUL DE TOUS LES   228
C       ELEMENTS DE LA REGRESSION                                        229
C       R.CARRE ET ECART-TYPE RESIDUEL                                   230
C                                                                       231
C             COEFFICIENTS GAMMA ET LEUR TEST (REGRESSION SUR LES COMPO  232
C       SANTES REDUITES                                                  233
C             COEFFICIENTS BETA ET LEUR TEST (REGRESSION SUR LES VARIA    234
C             BLES)                                                      235
C                                                                       236
      R2=0.0                                                            237
      L=0                                                               238
      DO 180 J=1,K                                                      239
      DO 170 I=1,NP                                                     240
      L=L+1                                                             241
  170 GAMMA(J)=GAMMA(J)+B(I)*V(L)                                       242
      L1=J+(J*J-J)/2                                                    243
      GAMMA(J)=GAMMA(J)/R(L1)                                           244
  180 R2=R2+GAMMA(J)*GAMMA(J)                                           245
      ST2=(1.0-R2)/(OBS-1.0-FLOAT(NP))                                  246
      WRITE(MX,15)R2,ST2                                                247
      ST2=SQRT(ST2)                                                     248
      WRITE(MX,16)                                                      249
      DO 190 J=1,K                                                      250
      T=GAMMA(J)/ST2                                                    251
  190 WRITE(MX,17)J,GAMMA(J),T                                          252
      WRITE(MX,18)                                                      253
      DO 200 I=1,NP                                                     254
      L=I-NP                                                            255
      D(I)=0.0                                                          256
      T=0.0                                                             257
      DO 201 J=1,K                                                      258
      L1=J+(J*J-J)/2                                                    259
      L=L+NP                                                            260
      D(I)=D(I)+V(L)*GAMMA(J)/R(L1)                                     261
  201 T=T+V(L)*V(L)/R(L1)/R(L1)                                         262
      T=D(I)/ST2/SQRT(T)                                                263
      WRITE(MX,17)I,D(I),T                                              264
  200 CONTINUE                                                          265
C                                                                       266
C       IRES EST EGAL A 1 ON CALCULE                                    267
C       LA VALEUR CENTREE REDUITE DE LA VARIABLE EXPLIQUEE              268
C       LA VALEUR ESTIMEE                                               269
C       LA DIFFERENCE SUR L'ECART-TYPE RESIDUEL                         270
C       LES VALEURS DES K COMPOSANTES CONSERVEES                        271
C                                                                       272
      IF(IRES)999,999,610                                               273
  610 IFA=1                                                             274
      WRITE(MX,19)(I,I=1,K)                                             275
      ST2=ST2*SQRT(OBS-1.0)                                             276
      DO 630 L1=1,NOBS                                                  277
      READ(1'IFA)(DATA(I),I=1,NVAR)                                     278
      T=0.0                                                             279
      DO 631 J=1,K                                                      280
  631 GAMMA(J)=0.0                                                      281
      DO 632 I=1,NP                                                     282
      DATA(I)=(DATA(I)-XBAR(I))/SIGMA(I)                                283
  632 T=T+D(I)*DATA(I)                                                  284
      L=0                                                               285
      DO 633 J=1,K                                                      286
      DO 633 I=1,NP                                                     287
      L=L+1                                                             288
  633 GAMMA(J)=GAMMA(J)+DATA(I)*V(L)                                    289
      DATA(NVAR)=(DATA(NVAR)-XBAR(NVAR))/SIGMA(NVAR)                    290
      R2=(DATA(NVAR)-T)/ST2                                             291
      WRITE(MX,20)L1,DATA(NVAR),T,R2,(GAMMA(J),J=1,K)                   292
  630 CONTINUE                                                          293
      GO TO 999                                                         294
      END
```

1REGRESSION ORTHOGONALE EXPLICATION DE LA HAUTEUR MOYENNE
0 MATRICE DES COEFFICIENTS DE CORRELATION

1.00												
0.36	1.00											
-0.09	-0.15	1.00										
0.05	0.03	0.14	1.00									
0.22	0.00	0.10	-0.24	1.00								
0.08	0.06	0.10	-0.20	0.39	1.00							
0.12	-0.12	0.10	0.21	-0.05	0.05	1.00						
0.00	-0.10	-0.14	0.05	-0.18	-0.08	0.19	1.00					
0.06	-0.08	-0.09	-0.02	0.40	-0.02	0.04	0.06	1.00				
0.00	-0.06	-0.03	0.02	0.30	-0.04	-0.18	-0.12	0.64	1.00			
0.10	-0.06	-0.03	-0.07	0.25	0.04	-0.40	0.07	0.52	0.43	1.00		
0.14	-0.18	-0.10	-0.09	0.28	0.04	0.15	0.17	0.65	0.41	0.54	1.00	
0.03	0.21	-0.15	0.25	-0.30	-0.12	0.25	-0.10	0.01	-0.06	-0.09	-0.08	1.00
0.06	0.18	-0.14	0.17	-0.19	-0.07	0.21	-0.02	0.11	-0.05	0.12	0.00	0.88

VALEURS PROPRES

2.96157	1.74137	1.54278	1.41930	1.24487	0.96063	0.74179	0.63290	0.56259	0.41151
0.38772	0.21400	0.17873							

POURCENTAGE CUMULE

0.22781	0.36176	0.48044	0.58961	0.68537	0.75927	0.81633	0.86501	0.90829	0.93994
0.96977	0.98623	0.99998							

VECTEURS PROPRES

COMPOSANTE NUMERO 1

-0.09133	0.07948	0.04055	0.12966	-0.35530	-0.09724	0.11347	-0.00610	-0.48244	-0.41898
-0.43645	-0.44482	0.14693							

COMPOSANTE NUMERO 2

0.05020	-0.02367	-0.19932	0.40899	-0.34286	-0.39982	0.33206	0.29274	0.24660	0.11598
0.04759	0.24516	0.42411							

COMPOSANTE NUMERO 3

0.58242	0.63923	-0.21252	0.03923	0.17724	0.19673	0.04183	-0.20222	0.00533	-0.01991

-0.02291 -0.05035 0.29862

COMPOSANTE NUMERO 4
0.16848 -0.20307 0.33371 0.08418 0.26970 0.38054 0.66415 0.15677 0.04448 -0.16715
-0.24710 0.18286 -0.06519

COMPOSANTE NUMERO 5
0.12999 0.02448 -0.52541 -0.42321 -0.07335 0.06660 0.01735 0.59029 -0.09172 -0.30074
0.01127 0.14571 -0.22027

COMPOSANTE NUMERO 6
0.38957 0.14658 0.38889 0.42692 -0.03763 -0.16732 -0.22174 0.37045 -0.14354 -0.07145
0.25152 -0.05494 -0.43058

COMPOSANTE NUMERO 7
-0.28760 0.07326 0.13406 0.14638 -0.18635 0.61987 -0.17311 0.22879 -0.06262 -0.17229
0.43025 0.02974 0.38936

COMPOSANTE NUMERO 8
0.33844 -0.24249 0.28878 -0.38170 -0.17605 -0.25484 0.00760 -0.28095 -0.13593 -0.39530
0.33235 0.27876 0.24430

COMPOSANTE NUMERO 9
0.27303 -0.55627 -0.48433 0.43611 0.04405 0.21615 -0.10325 -0.26734 -0.19879 -0.07503
0.09759 -0.01179 -0.07482

COMPOSANTE NUMERO 10
0.11881 0.11155 0.03614 -0.08845 -0.67865 0.29373 0.10500 -0.19398 -0.16188 0.38767
-0.11100 0.30094 -0.29899

COMPOSANTE NUMERO 11
-0.36905 0.34793 -0.13636 0.27317 0.05863 -0.08812 0.07249 -0.30963 0.03971 -0.47861
0.04914 0.45874 -0.30602

COMPOSANTE NUMERO 12
-0.11913 0.04996 0.00955 0.02522 0.33590 -0.12620 -0.04218 0.12858 -0.72835 0.30847
-0.05276 0.40685 0.20378

COMPOSANTE NUMERO 13
-0.12852 0.10741 -0.13721 -0.07401 0.02692 -0.09397 0.57075 -0.10143 -0.23492 0.14487
0.59968 -0.36946 -0.17233

RCARRE= 0.85905VARIANCE RESIDUELLE= 0.01006

COMPOSANTE	GAMMA	STUDENT
1	0.09706	0.96
2	0.50482	5.03
3	0.33778	3.36
4	-0.06596	-0.65
5	-0.15788	-1.57
6	-0.33680	-3.35
7	0.40597	4.04
8	0.24786	2.47
9	-0.07181	-0.71
10	-0.26336	-2.62
11	-0.17503	-1.74
12	0.02152	0.21
13	0.07815	0.77

VARIABLE	BETA	STUDENT
1	-0.01974	-0.16
2	0.01748	0.14
3	-0.01141	-0.10
4	-0.06168	-0.54
5	0.04120	0.29
6	-0.00490	-0.04
7	0.15249	0.98
8	0.03695	0.32
9	0.05357	0.30
10	-0.08898	-0.63
11	0.35018	2.13
12	-0.15673	-0.97
13	0.88769	7.15