

# REVUE DE STATISTIQUE APPLIQUÉE

MICHEL TENENHAUS

## **La régression qualitative**

*Revue de statistique appliquée*, tome 27, n° 2 (1979), p. 5-21

[http://www.numdam.org/item?id=RSA\\_1979\\_\\_27\\_2\\_5\\_0](http://www.numdam.org/item?id=RSA_1979__27_2_5_0)

© Société française de statistique, 1979, tous droits réservés.

L'accès aux archives de la revue « *Revue de statistique appliquée* » (<http://www.sfds.asso.fr/publicat/rsa.htm>) implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme  
Numérisation de documents anciens mathématiques  
<http://www.numdam.org/>

# LA RÉGRESSION QUALITATIVE

Michel TENENHAUS

CESA 1, rue de la Libération – 78350 JOUY-en-JOSAS

## RESUME

La régression multiple peut se généraliser au cas où la variable dépendante et les variables indépendantes sont de n'importe quel type (nominal, ordinal ou numérique).

La méthode utilisée consiste à coder les variables qualitatives, en respectant leurs structures nominales ou ordinales, de manière à maximiser la corrélation multiple entre la variable dépendante et les variables indépendantes. Il est bien connu que cette méthode permet de retrouver l'analyse de variance ainsi que l'analyse discriminante. Nous montrons qu'elle permet également de retrouver l'analyse de variance monotone (MONANOVA) proposée par J.B. KRUSKAL en 1965.

Nous étudions également dans cet article les programmes ADDALS et MORALS mis au point en 1976 par F.W. YOUNG, J. De LEEUW et Y. TAKANE.

L'objectif de ces programmes est de résoudre le problème général de la régression qualitative (la variable dépendante et les variables indépendantes sont de n'importe quel type) en utilisant la méthode décrite plus haute.

Enfin, nous avons soigneusement étudié les bases géométriques communes à ces méthodes et montrons que celles-ci sont finalement toutes équivalentes à une analyse canonique avec des contraintes sur le signe de certains coefficients.

## INTRODUCTION

La régression multiple peut se généraliser au cas où la variable dépendante  $Y$  et les variables indépendantes  $X_1, \dots, X_k$  sont de n'importe quel type (nominal, ordinal ou numérique).

La méthode utilisée consiste à coder les variables qualitatives, en respectant leurs structures nominales ou ordinales, de manière à maximiser la corrélation multiple entre la variable dépendante et les variables indépendantes.

Il est bien connu que cette méthode permet de retrouver l'analyse de variance ainsi que l'analyse discriminante. Nous montrons qu'elle permet également de retrouver l'analyse de variance monotone (MONANOVA) proposée par J.B. KRUSKAL [6] en 1965.

Nous étudions également dans cet article les programmes ADDALS et MORALS mis au point en 1976 par F.W. YOUNG, J. De LEEUW et Y. TAKANE [3] et [11].

L'objectif de ces programmes est de résoudre le problème général de la régression qualitative (la variable dépendante et les variables indépendantes sont de n'importe quel type) en utilisant la méthode décrite plus haut. Le programme MORALS est général tandis que le programme ADDALS suppose un plan d'expérience orthogonal.

Par ailleurs nous étudions soigneusement les bases géométriques communes à ces méthodes et montrons que celles-ci sont finalement toutes équivalentes à une analyse canonique avec des contraintes sur le signe de certains coefficients.

Enfin nous présentons un exemple d'utilisation de MORALS.

Donnons le plan de cet article :

- 1) Codage d'une variable qualitative
- 2) Equivalence entre différents critères utilisés pour déterminer des codages optimaux
- 3) Un critère général utilisé en régression qualitative
- 4) Trois programmes de régression qualitative : MONANOVA, ADDALS, MORALS.
- 5) Un exemple d'utilisation de MORALS.

## I. CODAGE D'UNE VARIABLE QUALITATIVE

### I.1. Définition

Considérons une population  $E$  de taille  $n$  sur laquelle est définie une variable qualitative  $X$  prenant ses valeurs dans l'ensemble  $\mathcal{X} = \{1, \dots, p\}$  des modalités.

On appelle codage  $\delta$  des modalités toute application  $\delta : \mathcal{X} \rightarrow \mathbf{R}$  et codage  $X^*$  de la variable  $X$  l'application  $X^* = \delta \circ X : E \rightarrow \mathbf{R}$ .

Si la variable  $X$  est nominale ( $\mathcal{X}$  ne possède pas de structure particulière) on impose en général à  $X^*$  d'être centrée ou centrée-réduite.

Si la variable  $X$  est ordinale ( $\mathcal{X}$  est muni d'une structure d'ordre à priori que nous supposons identique à l'ordre naturel  $1, 2, \dots, p$ ) on impose, en plus des éventuelles contraintes ci-dessus, les contraintes  $\delta(1) \leq \delta(2) \leq \dots \leq \delta(p)$ .

### I.2. Interprétation géométrique dans $\mathbf{R}^n$

#### I.2.1. $X$ nominale

Notons  $X_1, \dots, X_p$  les variables indicatrices associées à  $X$ .

Nous avons  $X^* = \delta \circ X = \sum_{i=1}^p \delta(i) X_i$ .

Ainsi si nous nous plaçons maintenant dans  $\mathbf{R}^n$  ( $X^*$  et les  $X_i$  sont assimilées à des vecteurs de  $\mathbf{R}^n$ ) nous avons :

$X^* \in \mathcal{L}(X)$  sous espace vectoriel de  $\mathbf{R}^n$  engendré par  $X_1, \dots, X_p$ .

La contrainte  $X^*$  centrée s'écrit  $X^* \in \mathcal{L}(X) \cap 1^\perp$  où  $1^\perp$  est le sous espace de  $\mathbf{R}^n$  orthogonal au vecteur  $1$  dont toutes les composantes sont égales à un.

La contraintes  $X^*$  centrée réduite s'écrit  $X^* \in \mathcal{L}(X) \cap 1^\perp \cap S$  où  $S$  est la sphère centrée sur  $0$  et de rayon  $\sqrt{n}$ .

#### I.2.2. $X$ ordinale

Remarquons que tout codage  $\delta$  vérifiant les contraintes

$$\delta(1) \leq \delta(2) \leq \dots \leq \delta(p)$$

peut s'écrire

$$\begin{aligned} \delta(1) &= -\lambda_1 + \lambda_2 \\ \delta(2) &= -\lambda_1 + \lambda_2 + \lambda_3 \\ &\vdots \\ \delta(p) &= -\lambda_1 + \lambda_2 + \lambda_3 + \dots + \lambda_{p+1} \end{aligned}$$

où les  $\lambda_i, i = 1, \dots, p + 1$ , sont positifs ou nuls.

La variable codée  $X^*$  associée à la variable ordinale  $X$  peut alors s'écrire

$$\begin{aligned} X^* &= \sum_{i=1}^p \delta(i) X_i = \sum_{i=1}^p \left( -\lambda_1 + \sum_{j=2}^{i+1} \lambda_j \right) X_i \\ &= \sum_{k=1}^{p+1} \lambda_k \tilde{X}_k \end{aligned}$$

où les  $\lambda_k$  sont positifs ou nuls et où les  $\tilde{X}_k$  sont définies par  $\tilde{X}_1 = -1$  et  $\tilde{X}_k = \sum_{i=k-1}^p X_i$  pour  $k = 2, \dots, p + 1$ .

Autrement dit  $\tilde{X}_1$  représente la variable codée  $X^*$  telle que toutes les modalités sont codées  $-1$  et  $\tilde{X}_k$  représente la variable codée  $X^*$  telle que les modalités  $1, \dots, k - 2$  sont codées  $0$  et les modalités  $k - 1, \dots, p$  sont codées  $1$ .

Ainsi si nous nous plaçons maintenant dans  $\mathbf{R}^n$  (les  $\tilde{X}_k$  sont assimilées à des vecteurs de  $\mathbf{R}^n$ ) nous avons :

$X^* \in \mathcal{C}(X)$  cône polyédrique convexe de  $\mathbf{R}^n$  engendré par  $\tilde{X}_1, \dots, \tilde{X}_{p+1}$ .

La contrainte  $X^*$  centrée s'écrit  $X^* \in \mathcal{C}(X) \cap 1^\perp$  et la contrainte  $X^*$  centrée réduite s'écrit  $X^* \in \mathcal{C}(X) \cap 1^\perp \cap S$ .

## II. EQUIVALENCES

Les propositions suivantes sont utiles pour montrer que les méthodes proposées par J.B. KRUSKAL [6] puis par F.W. YOUNG, J. De LEEUW et Y. TAKANE [3] et [11] possèdent une base géométrique commune et utilisent un critère commun.

Rappelons tout d'abord quelques propriétés des cônes polyédriques convexes (J. STOER et C. WITZGALL [10], chapitre 2).

Soit  $\mathcal{C}$  un cône polyédrique convexe de  $\mathbf{R}^n$ . Tout vecteur  $M$  de  $\mathbf{R}^n$  possède une projection orthogonale sur  $\mathcal{C}$ , notée  $P_{\mathcal{C}}(M)$ , réalisant le Minimum  $\sum_{X \in \mathcal{C}} \|M - X\|^2$ . Cette projection est unique.

On définit le cône polaire  $\mathcal{C}^p$  de  $\mathcal{C}$  par

$$\mathcal{C}^p = \{Y \in \mathbf{R}^n \mid Y'X \leq 0 \text{ pour tout } X \in \mathcal{C}\}.$$

On a alors la propriété suivante pour tout  $M \in \mathbb{R}^n$  :

$$M = P_e(M) + P_{e^p}(M)$$

avec  $P_e(M)$  et  $P_{e^p}(M)$  orthogonaux.

Et cette décomposition orthogonale de  $M$  en un vecteur de  $\mathcal{C}$  plus un vecteur de  $\mathcal{C}^p$  est unique.

Cette dernière propriété implique donc :

$$1) (M - P_e(M))' P_e(M) = 0$$

$$2) (M - P_e(M))' X \leq 0 \text{ pour tout } X \in \mathcal{C}.$$

### Proposition 1

Soit  $\mathcal{C}$  un cône polyédrique convexe de  $\mathbb{R}^n$ .

Soit  $M$  un vecteur de  $\mathbb{R}^n$  n'appartenant pas à  $\mathcal{C}^p$ .

Soit :

$$A_1 : \text{Minimum}_{A \in \mathcal{C}} \|M - A\| = \|M - A_1\|$$

$$A_2 : \text{Minimum}_{A \in \mathcal{C} \cap S} \|M - A\| = \|M - A_2\|$$

$$A_3 : \text{Maximum}_{A \in \mathcal{C}} \cos(A, M) = \cos(A_3, M)$$

$$A_4 : \text{Maximum}_{A \in \mathcal{C} \cap S} \cos(A, M) = \cos(A_4, M)$$

alors

$$\frac{\sqrt{n}}{\|A_1\|} A_1 = A_2 = A_3 = A_4$$

### Démonstration

$$1) \text{ L'identité } \|M - A\|^2 = \|M\|^2 + \|A\|^2 - 2 \|M\| \|A\| \cos(M, A)$$

implique  $A_2 = A_4$ .

$$2) \text{ Montrons que } \frac{\sqrt{n}}{\|A_1\|} A_1 = A_2.$$

Soit  $A \in \mathcal{C} \cap S$ .

On a, puisque  $(M - A_1)' A_1 = 0$ ,  $P_S(A_1) = \frac{\sqrt{n}}{\|A_1\|} A_1$  (où  $P_S(A_1)$  est la projection orthogonale de  $A_1$  sur  $S$ ) et

$(M - A_1)' (A - A_1) \leq 0$ , les relations suivantes :

$$\begin{aligned} \|M - \frac{\sqrt{n}}{\|A_1\|} A_1\|^2 &= \|M - A_1\|^2 + \|A_1 - \frac{\sqrt{n}}{\|A_1\|} A_1\|^2 \leq \|M - A_1\|^2 + \|A_1 - A\|^2 \\ &\leq \|M - A_1\|^2 + \|A_1 - A\|^2 + 2 (M - A_1)' (A_1 - A) = \|M - A\|^2 \end{aligned}$$

Et par conséquent

$$A_2 = \frac{\sqrt{n}}{\|A_1\|} A_1.$$

3) Le cosinus étant indépendant de la norme des vecteurs on peut choisir  $A_3 = A_4$ .

**Proposition 2**

Soient  $\mathcal{C}_1$  et  $\mathcal{C}_2$  deux cônes polyédriques convexes de  $\mathbb{R}^n$ .

Soit  $\mathcal{M} = \mathcal{C}_2 \cap S$ .

Supposons que  $\mathcal{M}$  ne soit pas entièrement contenu dans  $\mathcal{C}_1^p$ .

Soit  $A_1, M_1$  : Minimum  $\|M - A\| = \|M_1 - A_1\|$ .

$$\begin{matrix} A \in \mathcal{C}_1 \\ M \in \mathcal{M} \end{matrix}$$

Alors Maximum  $\cos(A, M) = \cos(A_1, M_1)$ ,

$$\begin{matrix} A \in \mathcal{C}_1 \\ M \in \mathcal{M} \end{matrix}$$

**Démonstration**

Soit  $A_2, M_2$  : Maximum  $\cos(A, M) = \cos(A_2, M_2)$ .

$$\begin{matrix} A \in \mathcal{C}_1 \\ M \in \mathcal{M} \end{matrix}$$

Nous avons  $A_1 = P_{\mathcal{C}_1}(M_1)$ . L'hypothèse  $\mathcal{M} \not\subset \mathcal{C}_1^p$  implique  $A_1 \neq 0$ . La proposition 1 nous permet de choisir  $A_2 = P_{\mathcal{C}_1}(M_2)$ .

Par conséquent nous avons

$$\|M_1 - A_1\|^2 = n(1 - \cos^2(M_1, A_1))$$

et

$$\|M_2 - A_2\|^2 = n(1 - \cos^2(M_2, A_2))$$

D'où la proposition puisque  $\|M_1 - A_1\|^2 \leq \|M_2 - A_2\|^2$  implique  $\cos^2(M_1, A_1) \geq \cos^2(M_2, A_2)$ .

Dans la proposition suivante et son corollaire nous reprenons, en l'approfondissant, la note théorique écrite par J.B. KRUSKAL dans le mode d'emploi du programme MONANOVA [7].

**Proposition 3**

Soit  $M$  un vecteur de  $\mathbb{R}^n$ .

Soit  $\mathcal{A}$  un sous-espace vectoriel de  $\mathbb{R}^n$

Soit  $S(A, M) = \frac{\|M - A\|}{\|A\|}$  (Stress de Kruskal lorsque  $\bar{A} = 0$ ).

Soit  $A_1$  : Minimum  $S(A, M) = S(A_1, M)$ .

$$A \in \mathcal{A}$$

Alors

$$1) \text{ Maximum } \cos(A, M) = \cos(A_1, M)$$

$$A \in \mathcal{A}$$

$$2) S(A_1, M) = \sqrt{1 - \cos^2(A_1, M)}$$

$$3) A_1 = \frac{1}{\cos^2(P_{\mathcal{A}}(M), M)} P_{\mathcal{A}}(M)$$

### Démonstration

Nous avons :

$$\begin{aligned} \cos^2(A, M) &= 1 - \frac{\text{Min}_{\alpha} \|\alpha M - A\|^2}{\|A\|^2} = 1 - \text{Min}_{\alpha} \frac{\|M - \alpha A\|^2}{\|\alpha A\|^2} \\ &= 1 - \text{Min}_{\alpha} S(\alpha A, M)^2 \end{aligned}$$

Soit  $A_2$  tel que  $\text{Maximum}_{A \in \alpha} \cos(A, M) = \cos(A_2, M)$

Montrons que  $\cos(A_1, M) = \cos(A_2, M)$  en vérifiant que  $\cos(A_1, M) \geq \cos(A_2, M)$  :

$$\begin{aligned} \cos^2(A_1, M) &= 1 - \text{Min}_{\alpha} S(\alpha A_1, M)^2 = 1 - S(A_1, M)^2 \\ &\geq 1 - \text{Min}_{\alpha} S(\alpha A_2, M)^2 = \cos^2(A_2, M). \end{aligned}$$

D'où les relations (1) et (2).

La relation (1) implique  $A_1 = \lambda_1 P_{\alpha}(M)$ .

Montrons que  $\lambda_1 = \frac{1}{\cos^2(P_{\alpha}(M), M)}$  :

$$\begin{aligned} S(A_1, M) &= \text{Minimum}_{\lambda} S(\lambda P_{\alpha}(M), M) \\ &= \text{Minimum}_{\lambda} \frac{\|M - \lambda P_{\alpha}(M)\|}{\|\lambda P_{\alpha}(M)\|} \\ &= \text{Minimum}_{\lambda} \frac{\|\frac{1}{\lambda} M - P_{\alpha}(M)\|}{\|P_{\alpha}(M)\|} \end{aligned}$$

Il faut donc rechercher le Minimum  $\|\frac{1}{\lambda} M - P_{\alpha}(M)\|$

$$\text{D'où } \frac{1}{\lambda_1} = \frac{M' P_{\alpha}(M)}{\|M\|^2} = \frac{\cos(M, P_{\alpha}(M)) \|P_{\alpha}(M)\|}{\|M\|} = \cos^2(P_{\alpha}(M), M)$$

et la relation (3) est démontrée.

### Corollaire

Soit  $\mathcal{A}$  un sous-espace vectoriel de  $\mathbf{R}^n$  et  $\mathcal{C}$  un cône polyédrique convexe.

Soit  $A_1, M_1$  :  $\text{Minimum}_{\substack{A \in \mathcal{A} \\ M \in \mathcal{C}}} S(A, M) = S(A_1, M_1)$

Nous avons  $M_1 = P_{\mathcal{C}}(A_1)$  et  $A_1 = \frac{1}{\cos^2(P_{\mathcal{A}}(M_1), M_1)} P_{\mathcal{A}}(M_1)$

**Démonstration**

$$1) S(A_1, M_1) = \underset{M \in \mathcal{C}}{\text{Minimum}} \frac{\|M - A_1\|}{\|A_1\|}. \text{ D'où } M_1 = P_{\mathcal{C}}(A_1).$$

$$2) S(A_1, M_1) = \underset{A \in \mathcal{A}}{\text{Minimum}} S(A, M_1).$$

$$\text{D'où } A_1 = \frac{1}{\cos^2(P_{\mathcal{C}}(M_1), M_1)} P_{\mathcal{A}}(M_1).$$

**Proposition 4**

Soit  $\mathcal{C}(Y)$  le cône polyédrique convexe associé à une variable ordinaire  $Y$ .

Si  $M$  est un vecteur centré on a  $P_{\mathcal{C}(Y)}(M) = P_{\mathcal{C}(Y) \cap 1^\perp}(M)$ .

**Démonstration**

Il suffit de montrer que si  $M$  est centré, alors sa projection sur  $\mathcal{C}(Y)$  l'est aussi.

Or les vecteurs  $1$  et  $-1$  appartiennent à  $\mathcal{C}(Y)$ .

Par conséquent

$$\begin{aligned} (M - P_{\mathcal{C}(Y)}(M))' 1 &= -P_{\mathcal{C}(Y)}(M)' 1 \leq 0 \\ \text{et } (M - P_{\mathcal{C}(Y)}(M))' (-1) &= P_{\mathcal{C}(Y)}(M)' 1 \leq 0 \end{aligned}$$

D'où nous déduisons que  $P_{\mathcal{C}(Y)}(M)$  est centré.

**III – UN CRITERE GENERAL UTILISE EN REGRESSION QUALITATIVE**

Les variables  $Y, X_1, \dots, X_k$  sont d'un type quelconque (nominal, ordinal ou numérique).

Notons  $\mathcal{C}, \mathcal{C}_1, \dots, \mathcal{C}_k$  les sous-espaces vectoriels ou les cônes polyédriques convexes de  $\mathbb{R}^n$  associés à ces variables.

Un critère général consiste à rechercher les codages  $Y^* \in \mathcal{C}, X_1^* \in \mathcal{C}_1, \dots, X_k^* \in \mathcal{C}_k$  maximisant la corrélation multiple  $R(Y^*; X_1^*, \dots, X_k^*)$  entre  $Y^*$  et les  $X_i^*$ .

Pour éviter des solutions dégénérées on élimine les codages constants.

Comme  $Y^* \in \mathcal{C}$  implique  $\frac{Y^* - \bar{Y}^*}{S_Y} \in \mathcal{C}$ , et de même pour les  $X_i^*$ , le critère général peut aussi s'écrire :

$$\begin{aligned} &\text{Maximiser } R(Y^*; X_1^*, \dots, X_k^*) \\ &Y^* \in \mathcal{C} \cap 1^\perp \cap \mathcal{S} \\ &X_i^* \in \mathcal{C}_i \cap 1^\perp \cap \mathcal{S} \\ &i = 1, \dots, k \end{aligned}$$

On recherche donc les codages centrés-réduits des variables  $Y, X_1, \dots, X_k$  maximisant la corrélation multiple.

Le critère général se ramène à une analyse canonique sous contrainte.



En effet, en utilisant les résultats de la partie I de cet article, nous constatons que tout codage d'une variable qualitative s'écrit comme combinaison linéaire de variables bien définies :

$$X^* = \sum_{i=1}^m \alpha_i \check{X}_i$$

puisque, si X est nominale,  $X^* = \sum_{i=1}^p \delta(i) X_i$

et, si X est ordinale,  $X^* = \sum_{i=1}^{p+1} \lambda_i \tilde{X}_i$  avec  $\lambda_1, \dots, \lambda_{p+1} \geq 0$ .

Les coefficients  $\alpha_i$  sont donc quelconques si X est nominale et positifs si X est ordinale.

Le critère général est donc équivalent à :

$$\text{Maximiser}_{\alpha_j, \tau_i, \beta_{i1}} \text{cor} \left( \sum_{j=1}^{m_0} \alpha_j \check{Y}_j, \sum_{i=1}^k \tau_i \sum_{l=1}^{m_i} \beta_{il} \check{X}_{il} \right)$$

avec les contraintes de positivité adéquates.

Si on pose  $\gamma_{i1} = \tau_i \beta_{i1}$  le critère devient

$$\text{Maximiser}_{\alpha_j, \gamma_{i1}} \text{cor} \left( \sum_{j=1}^{m_0} \alpha_j \check{Y}_j, \sum_{i=1}^k \sum_{l=1}^{m_i} \gamma_{il} \check{X}_{il} \right)$$

avec les contraintes

- les  $\alpha_j$  associés à une variable Y ordinale sont positifs
- les  $\gamma_{i1}$  associés à une variable  $X_i$  ordinale sont de même signe.

*Remarque :*

En analyse des préférences on étudie souvent le problème d'expliquer une variable ordinale Y à l'aide de variable ordinales  $X_i$  et on impose aux  $\tau_i$  d'être positifs.

Ce problème d'analyse multicritère se ramène donc au problème ci-dessus avec les contraintes plus simples car linéaires : les  $\alpha_j$  et les  $\gamma_{i1}$  doivent être positifs.

#### IV. TROIS PROGRAMMES DE REGRESSION QUALITATIVE

Nous allons étudier dans ce paragraphe les programmes MONANOVA, ADDALS et MORALS.

Nous montrerons que le critère général s'applique exactement à MONANOVA et à MORALS et, avec une contraintes supplémentaire, à ADDALS.

#### IV.1. MONANOVA\*

Le programme MONANOVA mis au point par J.B. KRUSKAL [6] et J.B. KRUSKAL et F. CARMONE [7] s'applique au cas où Y est ordinale et les  $X_i$  nominales.

##### IV.1.1. Le critère utilisé dans MONANOVA est équivalent au critère général

Notons :

$$\pi = \mathcal{C}(Y) \cap 1^\perp \cap S$$

l'ensemble des codages centrés réduits de Y,

$$\alpha_i = \mathcal{L}(X_i) \cap 1^\perp$$

l'ensemble des codages centrés de  $X_i$ ,

$$\alpha = \bigoplus_{i=1}^k \alpha_i$$

somme directe des  $\alpha_i$  (c'est l'ensemble des combinaisons linéaires de  $X_i^*$  appartenant à  $\alpha_i$ ).

$$\text{J.B. KRUSKAL définit le Stress } S(A, M) = \frac{\|M - A\|}{\|A\|}$$

L'objectif de MONANOVA est de rechercher le

$$\text{Minimum}_{\substack{A \in \alpha \\ M \in \pi}} S(A, M)$$

La proposition suivante montre que ce critère est équivalent au critère général.

##### Proposition 5

$$\text{Soit } M_1, A_1 \text{ tels que } \text{Minimum}_{\substack{M \in \pi \\ A \in \alpha}} S(A, M) = S(A_1, M_1).$$

$$\text{Alors } \text{Maximum}_{\substack{M \in \pi \\ A \in \alpha}} \cos(A, M) = \cos(A_1, M_1)$$

$$\text{et de plus, si on pose } A_1 = \sum_{i=1}^k A_{1i} \text{ où } A_{1i} \in \alpha_i,$$

on a :

$$\text{Maximum}_{\substack{Y^* \in \pi \\ X^* \in \alpha_i \\ i=1, \dots, k}} R(Y_1^*; X_1^* \dots X_k^*) = R(M_1; A_{11}, \dots, A_{1k})$$

##### Démonstration

1) Montrons que

$$\text{Maximum}_{\substack{M \in \pi \\ A \in \alpha}} \cos(A, M) = \cos(A_1, M_1)$$

---

(\*) MONANOVA = Monotone Analysis of Variance

D'après la proposition 3 nous avons :

$$\underset{A \in \alpha}{\text{Minimum}} S(A, M) = \sqrt{1 - \underset{A \in \alpha}{\text{Maximum}} \cos^2(A, M)}$$

$$\text{D'où } \cos(A_1, M_1) = \underset{\substack{M \in \pi \\ A \in \alpha}}{\text{Maximum}} \cos(A, M)$$

et 
$$S(A_1, M_1) = \sqrt{1 - \cos^2(A_1, M_1)}$$

2) La corrélation multiple  $R(Y^*; X_1^*, \dots, X_k^*)$  étant égale au

$$\cos(Y^*; P_{\mathcal{E}(X_1^*, \dots, X_k^*)}(Y^*))$$

où  $Y^* \in \mathcal{N}\zeta$  et  $P_{\mathcal{E}(X_1^*, \dots, X_k^*)}(Y^*) \in \mathcal{A}$

on a bien le résultat annoncé dans la proposition.

#### IV.1.2. L'algorithme

L'algorithme est itératif et permet de construire une suite de vecteurs  $A^t, M^t$  tels que la suite  $S(A^t, M^t)$  soit décroissante.

La suite peut converger vers un optimum local et l'optimum atteint dépend sans doute de la solution initiale choisie.

Décrivons l'algorithme.

##### Initialisation

On choisit arbitrairement  $M^0 \in \mathcal{N}\zeta$ , codage centré réduit de  $Y$ .

##### Etape $t$

A l'étape  $t - 1$  on a obtenu les vecteurs

$$M^{t-1} \in \mathcal{N}\zeta, A_1^{t-1} \in \mathcal{A}_1, \dots, A_k^{t-1} \in \mathcal{A}_k \quad \text{codages de } Y, X_1, \dots, X_k.$$

Et on construit

$$A^{t-1} = \sum_{i=1}^k A_i^{t-1}$$

##### Détermination des $A_i^t$

Il s'agit de rechercher le  $\underset{A \in \alpha}{\text{Minimum}} S(A, M^{t-1})$ .

J.B. KRUSKAL utilise une procédure du gradient pour obtenir  $A^t$ .

Il nous semble préférable d'utiliser la proposition 3 donnant :

$$\begin{aligned} A^t &= \frac{1}{\cos^2(M^{t-1}, P_{\mathcal{A}}(M^{t-1}))} P_{\mathcal{A}}(M^{t-1}) \\ &= \sum_{i=1}^k A_i^t. \end{aligned}$$

##### Détermination de $M^t$

Il s'agit de rechercher le  $\underset{M \in \pi}{\text{Minimum}} S(A^t, M)$ .

Comme  $S(A^t, M) = \frac{\|M - A^t\|}{\|A^t\|}$  il suffit de rechercher le Minimum  $\|M - A^t\|$ .  
 $M \in \pi$

D'après la proposition 1 ce minimum s'obtient en deux étapes :

1) Rechercher le Minimum  $\|M - A^t\|$ .

$$M \in \mathcal{C}(Y) \cap 1^\perp$$

2) Normaliser le résultat obtenu.

Par ailleurs nous avons vu (Proposition 4) que si un vecteur A est centré sa projection sur  $\mathcal{C}(Y)$  l'est aussi.

Le minimum cherché peut donc s'obtenir en modifiant l'étape 1.

Soit :

1) Rechercher le Minimum  $\|M - A^t\|$ .

$$M \in \mathcal{C}(Y)$$

2) Normaliser le résultat obtenu.

D'où :

$$M^t = \frac{\sqrt{n}}{\|P_{\mathcal{C}(Y)}(A^t)\|} P_{\mathcal{C}(Y)}(A^t).$$

L'algorithme permettant d'obtenir  $P_{\mathcal{C}(Y)}(A^t)$  est décrit dans [1] et [5].

#### Convergence de l'algorithme

La suite  $S(A^t, M^t)$  est décroissante :

$$S(A^t, M^t) \leq S(A^t, M^{t-1}) \leq S(A^{t-1}, M^{t-1}).$$

L'algorithme s'arrête lorsque le critère  $S(A^t, M^t)$  se stabilise.

## IV.2. ADDALS\*

Le programme ADDALS, mis au point en 1976 par J. De LEEUW, F.W. YOUNG et Y. TAKANE [3] s'applique au cas où Y est de type quelconque et les  $X_i$  nominales ou ordinales. De plus le plan d'expérience associé aux  $X_i$  doit être orthogonal.

### IV.2.1. Le Critère utilisé dans ADDALS est équivalent au critère général avec une contrainte supplémentaire

Notons  $\mathcal{N} = \mathcal{C}(Y) \cap 1^\perp \cap S$  l'ensemble des codages centrés-réduits de la variable Y.

$\mathcal{C}(Y)$  est un sous-espace vectoriel ou un cône polyédrique convexe de  $\mathbb{R}^n$  suivant la nature de Y.

Notons :

$$\mathcal{B}_i = \begin{cases} \mathcal{A}_i = \mathcal{L}(X_i) \cap 1^\perp & \text{si } X_i \text{ est nominale} \\ \mathcal{C}_i = \mathcal{C}(X_i) \cap 1^\perp & \text{si } X_i \text{ est ordinale} \end{cases}$$

On suppose que le plan d'expérience associé aux  $X_i$  est orthogonal :

$$\mathcal{A}_i \perp \mathcal{A}_j \quad \text{pour } i \neq j.$$

(\*) ADDALS = Additivity analysis by alternating least Squares.

On pose  $\mathcal{A} = \bigoplus_{i=1}^k \mathcal{A}_i$  et  $\mathcal{B} = \bigoplus_{i=1}^k \mathcal{B}_i$

Le critère utilisé dans ADDALS est :

$$\lambda(M, B) = \|M - B\|^2$$

et l'objectif est de rechercher le

$$\text{Minimum}_{\substack{M \in \mathcal{M} \\ B \in \mathcal{B}}} \lambda(M, B)$$

D'après la proposition 2 il est équivalent de rechercher le

$$\text{Maximum}_{\substack{M \in \mathcal{M} \\ B \in \mathcal{B}}} \cos(M, B),$$

soit de rechercher le

$$\text{Maximum}_{\substack{Y^* \in \mathcal{Y} \\ X_1^* \in \mathcal{A}_1, \dots, X_k^* \in \mathcal{A}_k}} \text{cor}(Y^*, \sum_{i=1}^k X_i^*).$$

Ce critère est bien équivalent au critère général avec la contrainte supplémentaire : les coefficients de régression des variables ordinales sont positifs.

#### IV.2.2. L'algorithme

L'algorithme est itératif et permet de construire une suite de vecteurs  $M^t$ ,  $B^t$  tels que la suite  $\lambda(M^t, B^t)$  soit décroissante.

Comme pour MONANOVA cette suite peut probablement converger vers un optimum local et l'optimum atteint dépend sans doute de la solution initiale choisie.

Décrivons l'algorithme.

##### Initialisation

On choisit arbitrairement  $M^0 \in \mathcal{M}$ , codage centré réduit de  $Y$ .

##### Etape $t$

A l'étape  $t - 1$  on a obtenu les vecteurs  $M^{t-1} \in \mathcal{M}$ ,  $B_1^{t-1} \in \mathcal{B}_1, \dots, B_k^{t-1} \in \mathcal{B}_k$  codages de  $Y, X_1, \dots, X_k$ . Et on construit  $B^{t-1} = \sum_{i=1}^k B_i^{t-1}$ .

##### Détermination des $B_i^t$

La procédure est basée sur l'identité suivante. Pour tout  $A = \sum_{i=1}^k A_i$ , avec  $A_i \in \mathcal{A}_i$ , nous avons :

$$\begin{aligned} \|M - A\|^2 &= \|M - P_{\mathcal{A}}(M)\|^2 + \|P_{\mathcal{A}}(M) - A\|^2 \\ &= \|M - P_{\mathcal{A}}(M)\|^2 + \sum_{i=1}^k \|P_{\mathcal{A}_i}(M) - A_i\|^2 \end{aligned}$$

puisque les  $\mathcal{A}_i$  sont orthogonaux.

D'où nous déduisons que rechercher le

$$\text{Minimum}_{B \in \mathcal{B}} \|M^{t-1} - B\|^2$$

revient à rechercher, pour  $i = 1, \dots, k$ , le

$$\text{Minimum}_{B_i \in \mathcal{B}_i} \|P_{\alpha_i}(M^{t-1}) - B_i\|^2$$

Par conséquent

$$B_i^t = \begin{cases} P_{\alpha_i}(M^{t-1}) & \text{si } X_i \text{ est nominale} \\ P_{\mathcal{C}_i}(M^{t-1}) & \text{si } X_i \text{ est ordinale} \end{cases}$$

puisque

$$P_{\mathcal{B}_i} \circ P_{\alpha_i}(M^{t-1}) = P_{\mathcal{B}_i}(M^{t-1}).$$

Et

$$B^t = \sum_{i=1}^k B_i^t.$$

*Détermination de  $M^t$*

Il s'agit de rechercher le  $\text{Minimum}_{M \in \mathcal{M}} \|M - B^t\|^2$ .

D'où, d'après la proposition 1 et les propriétés de la projection,

$$M^t = P_{\mathcal{M}}(B^t) = \frac{\sqrt{n}}{\|P_{\mathcal{C}(Y)}(B^t)\|} P_{\mathcal{C}(Y)}(B^t).$$

*Convergence de l'algorithme*

La suite  $\lambda(M^t, B^t)$  est décroissante :

$$\|M^t - B^t\|^2 \leq \|M^{t-1} - B^t\|^2 \leq \|M^{t-1} - B^{t-1}\|^2.$$

L'algorithme s'arrête lorsque le critère  $\lambda(M^t, B^t)$  se stabilise.

### IV – 3. MORALS\*

Le programme MORALS mis au point en 1976 par F.W. Young, J. De Leeuw et Y. Takane [11] s'applique au cas où les variables  $Y, X_1, \dots, X_k$  sont de type quelconque.

#### IV – 3.1. Le critère utilisé dans MORALS est équivalent au critère général

Notons  $\mathcal{A}_i, \mathcal{A}_1, \dots, \mathcal{A}_k$  les espaces associés aux variables  $Y, X_1, \dots, X_k$  représentant les codages centrés-réduits de ces différentes variables.

Notons  $\mathcal{A} = \sum_{i=1}^k \mathcal{L}(\mathcal{A}_i)$  le sous-espace vectoriel engendré par les  $\mathcal{A}_i$ .

Le critère utilisé dans MORALS est  $\lambda(M, A) = \|M - A\|^2$  et l'objectif de

(\*) MORALS = Multiple optimal regression by alternating least squares.

l'algorithme est de recherche le

$$\begin{aligned} & \text{Minimum } \|M - A\|^2. \\ & M \in \mathcal{M} \\ & A \in \alpha \end{aligned}$$

D'après la proposition 2 il est équivalent de rechercher le

$$\begin{aligned} & \text{Maximum } \cos(M, A) \\ & M \in \mathcal{M} \\ & A \in \alpha \end{aligned}$$

Il s'agit donc de rechercher le

$$\begin{aligned} & \text{Maximum } R(Y^* ; X_1^*, \dots, X_k^*). \\ & Y^* \in \mathcal{M} \\ & X_i^* \in \alpha_i \\ & i = 1, \dots, k \end{aligned}$$

#### IV – 3.2. L'algorithme de MORALS

Comme dans MONANOVA et ADDALS l'algorithme est itératif et permet de construire une suite de vecteurs  $M^t, A^t$  tels que la suite  $\lambda(M^t, A^t)$  soit décroissante.

La convergence peut être locale et dépendre de la solution initiale choisie. Décrivons l'algorithme.

##### 1) Initialisation

On choisit  $M^0$  arbitraire appartenant à  $\mathcal{M}$ , codage centré-réduit de  $Y$ .

##### 2) Etape $t$

A l'étape  $t - 1$  on a défini  $M^{t-1}, A_1^{t-1}, \dots, A_k^{t-1}$  appartenant respectivement à  $\mathcal{M}, \alpha_1, \dots, \alpha_k$ .

On définit également  $A^{t-1} = P_{\mathcal{E}(A_1^{t-1}, \dots, A_k^{t-1})}(M^{t-1}) = \sum_{i=1}^k \alpha_i^{t-1} A_i^{t-1}$ .

##### a) Détermination de $M^t$

Il s'agit de rechercher le Minimum  $\|M - A^{t-1}\|^2$ .

$$M \in \mathcal{M}$$

Donc  $M^t = P_{\mathcal{M}}(A^{t-1})$ .

##### b) Détermination des $A_i^t$

On détermine progressivement les  $A_i^t$ .

##### 1) Détermination de $A_1^t$

On recherche le Minimum  $\|M^t - \alpha_1^{t-1} A_1 - \sum_{i=2}^k \alpha_i^{t-1} A_i^{t-1}\|^2$ .

$$A_1 \in \alpha_1$$

Donc  $A_1^t = P_{\alpha_1} \left( \frac{1}{\alpha_1^{t-1}} \left( M^t - \sum_{i=2}^k \alpha_i^{t-1} A_i^{t-1} \right) \right)$ .

2) *Détermination de  $A_2^t$ .*

On recherche de la même manière le

$$\text{Minimum}_{A_2 \in \mathcal{C}_2} \|M^t - \alpha_1^{t-1} A_1^t - \alpha_2^{t-1} A_2 - \sum_{i=3}^k \alpha_i^{t-1} A_i^{t-1}\|^2$$

D'où 
$$A_2^t = P_{\alpha_2} \left( \frac{1}{\alpha_2^{t-1}} \left( M^t - \alpha_1^{t-1} A_1^t - \sum_{i=3}^k \alpha_i^{t-1} A_i^{t-1} \right) \right)$$

3) On procède de la même manière pour déterminer  $A_3^t, \dots, A_k^t$ .

c) *Détermination de  $A^t$*

On choisit 
$$A^t = P_{\mathcal{E}(A_1^t, \dots, A_k^t)} (M^t) = \sum_{i=1}^k \alpha_i^t A_i^t.$$

3) *Convergence de l'algorithme*

La suite  $\lambda(M^t, A^t) = \|M^t - A^t\|^2$  est décroissante :

$$\begin{aligned} \|M^t - A^t\|^2 &\leq \|M^t - \sum_{i=1}^k \alpha_i^{t-1} A_i^t\|^2 \leq \|M^t - \sum_{i=1}^{k-1} \alpha_i^{t-1} A_i^t - \alpha_k^{t-1} A_k^{t-1}\|^2 \\ &\leq \dots \leq \|M^t - A^{t-1}\|^2 \leq \|M^{t-1} - A^{t-1}\|^2. \end{aligned}$$

*Remarque*

Nous n'avons pas décrit toutes les possibilités des programmes ADDALS et MORALS et nous renvoyons le lecteur intéressé aux articles originaux et à la documentation des programmes.

## V – UN EXEMPLE D'UTILISATION DE MORALS

Nous avons utilisé le programme MORALS sur les données de G. de Peslouan [9] concernant les femmes ingénieurs diplômées en France.

Cet exemple a déjà été traité par la segmentation [2]. Il s'agissait d'analyser la liaison entre une note d'ambition et des facteurs socioéconomiques.

MORALS permet de résoudre le même problème que la segmentation, mais de manière plus rapide et plus synthétique.

Précisons les variables utilisées dans cet exemple donné à titre illustratif.

Y : *Ambition (ordinale)*

- 1 beaucoup d'ambition
- .
- .
- .
- 5 très peu d'ambition



$X_1$ : *Ecole (nominale)*  
 1 école polytechnique féminine  
 2 école d'ingénieurs mixte

$X_2$ : *Enfants (numérique)*  
 0 enfant  
 1 enfant  
 2 enfants  
 3 enfants  
 4 enfants  
 5 enfants et plus

$X_3$ : *Emploi (nominale)*  
 1 recherche  
 2 bureau d'études  
 3 laboratoire  
 4 informatique  
 5 autres

Le programme MORALS nous a permis d'obtenir les codages suivants des variables qualitatives :

$Y^*$ : <i>Ambition</i>	$X_1^*$ : <i>Ecole</i>	$X_3^*$ : <i>Emploi</i>
1 → - 0,62	1 → - 1	1 → - 0,48
2 → - 0,62	2 → + 1	2 → 1,86
3 → - 0,62		3 → - 0,13
4 → 1,6		4 → - 1,52
5 → 1,6		5 → 0,69

La régression multiple entre  $Y^*$  et  $X_1^*$ ,  $X_2$ ,  $X_3^*$  donne

$$\hat{Y}^* = - 0,08 + 0,19 X_1^* + 0,08 X_2 + 0,16 X_3^*$$

avec

$$R(Y^* ; X_1^*, X_2, X_3^*) = 0,29.$$

Ainsi une femme sortie de l'E.P.F., ayant 2 enfants et travaillant en informatique a une note d'ambition estimée à

$$\begin{aligned} & - 0,08 + 0,19 \times (- 1) + 0,08 \times 2 + 0,16 \times (- 1,52) \\ & = - 0,35 \end{aligned}$$

Nous la classerions dans la catégorie des femmes ambitieuses.

## CONCLUSION

La régression qualitative nous permet donc de généraliser la régression multiple au cas où les variables  $Y$ ,  $X_1$ , ...,  $X_k$  sont de n'importe quel type.

Lorsque toutes les variables sont quantitatives ou qualitatives nominales le critère général proposé dans la troisième partie de cet article est en effet équivalent

au critère de l'analyse canonique et on retrouve donc toutes les méthodes qui sont des cas particuliers de l'analyse canonique.

Mais lorsque certaines variables sont ordinales nous aboutissons à des méthodes nouvelles.

Des programmes existent : MONANOVA fréquemment utilisé en marketing, ADDALS et MORALS plus récents.

Les algorithmes utilisés reposent tous sur des méthodes aux moindres carrés alternés et sont convergents dans un sens très étroit : le critère utilisé converge. Nous savons peu de choses sur l'importance du choix de la solution initiale et sur la possibilité de converger vers un optimum local.

De nouveaux algorithmes basés sur une formulation du problème en terme de programmation mathématique linéaire (E. Jacquet-Lagrèze et J. Siskos [4]) ou non linéaire (A. Macquin [8]) semblent très prometteurs aussi bien sur le plan théorique que pratique.

Disons pour conclure que toutes ces méthodes nous ont paru fort intéressantes et qu'elles devraient se révéler très utiles dans le domaine de l'analyse des données qualitatives.

## REFERENCES

- [1] BARLOW R.E., BARTHOLOMEW D.J., BREMNER J.M., BRUNK H.D. — Statistical inference under order restrictions. London : Wiley, 1972.
- [2] BERTIER P. et BOUROCHE J.M. — Analyse des données multidimensionnelles. PUF, 1975.
- [3] De LEEUW J., YOUNG F.W., TAKANE Y. — Additive structure in qualitative data : an alternating least squares method with optimal scaling features. *Psychometrika*, vol. 41, n° 4, Déc. 1976.
- [4] JACQUET-LAGREZE E. et SISKOS J. — Une méthode de construction d'une fonction d'utilité additive explicative d'une préférence globale pré-ordinaire. *Cahier du LAMSADE* n° 16, 1978.
- [5] KRUSKAL J.B. — Non metric multidimensional scaling : a numerical method. *Psychometrika*, vol. 29, 1964.
- [6] KRUSKAL J.B. — Analysis of factorial experiments by estimating monotone transformations of the data. *Journal of the Royal Statistical Society*, Series B, 1965, 27, 251-263.
- [7] KRUSKAL J.B. et CARMONE F. — Use and theory of Monanova, a program to analyse factorial experiments by estimating monotone transformation of the data. Unpublished paper, Bell Telephone Laboratories, Murray Hill, New Jersey, 1968.
- [8] MACQUIN A. — Utilisation de la programmation non linéaire en régression qualitative. Thèse de 3<sup>e</sup> cycle, en cours.
- [9] de PESLOUAN G. — Les femmes ingénieurs diplômées en France. Thèse de 3<sup>e</sup> cycle, Paris 1971.
- [10] STOER J. et WITZGALL C. — Convexity and optimization in finite dimensions I. Springer — Verlag, Berlin, 1970.
- [11] YOUNG F.W., De LEEUW J., TAKANE Y. — Regression with qualitative and quantitative variables : an alternating least squares method with optimal scaling features. *Psychometrika*, vol. 41, n° 4, Déc. 1976.