STATISTIQUE ET ANALYSE DES DONNÉES

ROBERT SABATIER

Quelques généralisations de l'analyse en composantes principales de variables instrumentales

Statistique et analyse des données, tome 9, nº 3 (1984), p. 75-103.
<http://www.numdam.org/item?id=SAD_1984__9_3_75_0>

Statistiques et Analyse de données
1984 - Vol.9 n°3 - pp. 75-103

QUELQUES GENERALISATIONS DE L'ANALYSE EN COMPOSANTES PRINCIPALES DE VARIABLES INSTRUMENTALES

Robert SABATIER

Laboratoire de Physique Industrielle Pharmaceutique

Faculté de Pharmacie de MONTPELLIER

Résumé : *Le but de cet article est de présenter cinq généralisations possibles de l'Analyse en Composantes Principales de Variables Instumentales (A.C.P.V.I.) introduite par RAO. Ces généralisations sont fondées d'une part sur le formalisme de CAILLIEZ et PAGES et d'autre part sur le coefficient RV d'ESCOUFIER. Ces généralisations considérées comme des problèmes d'optimisation, se ramènent en général à des analyses en composantes principales particulières. Un dernier paragraphe applique ces résultats aux cas des variables qualitatives.*

Abstract : *This paper aims to make five propositions for possible generalizations of principal component analysis of instrumental variables (ACPVI) such as this method has been introduced by RAO. These generalizations are based upon PAGES and CAILLIEZ'formalism and ESCOUFIER's RV coefficient. When considered as optimization problems, they are shown to be equivalent to particular principal component analysis. In the last section results are applied to qualitative variables.*

Mots clés : *Analyse en Composantes Principales, Opérateur, Métrique, Coefficient RV, A.C.P.V.I.*

Indices de classification STMA : *06-070, 06-010, 06-020.*

0 - INTRODUCTION

L'Analyse en Composantes Principales de Variables Instrumentales (A.C.P.V.I.) initialement introduite par RAO, semble actuellement retrouver la faveur des statisticiens : ESCOUFIER a montré comment elle peut s'inscrire comme méthode à part entière de l'Analyse des Données par l'utilisation des opérateurs. Depuis nombre de publications ont montré l'intérêt théorique de cette technique (OBADIA, SCHEKTMAN, CROQUETTE, D'AMBRA - LAURO, SABATIER) ; BONIFAS et al. ont montré également l'intérêt pratique.

Le but de cet article est de montrer, après un bref rappel de la définition et des propriétés fondamentales de l'A.C.P.V.I., comment il est possible de la généraliser à plusieurs groupes de variables de référence. Ces généralisations au nombre de cinq, montrent que l'utilisation d'opérateurs et de triplet statistique, est judicieuse, ramenant ainsi ces A.C.P.V.I. généralisées à des ACP particulières. Une dernière partie permet de retrouver, comme cas particuliers, des techniques que le praticien de l'Analyse des Données connaît bien.

1 - NOTATIONS RAPPELS ET DEFINITIONS

Le formalisme utilisé ici est celui de CAILLIEZ et PAGES. Nous supposerons en outre le lecteur familier avec la notion d'opérateur, de triplet statistique et de coefficient RV (voir [6]).

On suppose réalisée une étude (X_1, Q_1, D) portant sur n individus. X_1 (pxn) est supposé centré pour les poids définissant la matrice diagonale D. On notera par $W_1D = X_1' Q_1 X_1 D$ l'opérateur de $F = \mathbb{R}^n$ des proximités entre individus.

On rappelle que l'espace vectoriel $\mathcal{D}(F)$ des opérateurs auto-adjoints de F peut être muni d'une structure euclidienne par le produit scalaire suivant :

$$(A,B) \in \mathcal{D}(F) \times \mathcal{D}(F) \quad ; \quad \langle A,B \rangle = Tr(A\ B).$$

On notera par $|| \ ||$ la norme déduite de ce produit scalaire. Le coefficient RV entre deux opérateurs A et B de $\mathcal{D}(F)$ est défini par :

$$RV\ (A,\ B) = \frac{\langle A,\ B \rangle}{||A||\ .\ ||B||}$$

1.1 - Definition de l'A.C.P.V.I.

Soit Y un tableau (qxn) de mesures réalisées sur les mêmes n individus que ceux du triplet $(X_1,\ Q_1,\ D)$. On suppose Y centré pour D.

On appelle A.C.P.V.I. d'ordre r de Y par rapport à $(X_1,\ Q_1,\ D)$ (noté $Y/(X_1,\ Q_1,\ D)$), l'ACP de (Y, Q, D) où Q (qxq) est une pseudo-métrique de rang r ($\leqslant$ q) rendant minimale la quantité :

$$||W_1 D - Y'\ Q\ Y\ D|| \qquad (1)$$

1.2 - Conséquences

Thérorème 1.2

> La pseudo-métrique $Q = M_r\ M_r'$ où M_r(qxr) est de rang r minimisant (I) est solution du système aux valeurs propres suivant :
>
> $$(I) \quad \begin{cases} Y\ D\ W_1\ D\ Y'\ M_r = \ Y\ D\ Y'\ M_r\ \Lambda_r \\ M_r'\ Y\ D\ Y'\ M_r \ = \Lambda_r \end{cases}$$

où $\Lambda_r = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_r)$ est la matrice diagonale des r premières valeurs propres, supposées distinctes, rangées en ordre décroissant.

La démonstration de ce théorème peut se trouver dans [1], [6], [10] ou [11].

Remarques :

. Il faut noter le rôle dissymétrique des tableaux X_1 et Y. On remarque aussi que X_1 n'est présent que par l'intermédiaire de son opérateur W_1D.

. L'ACPVI ainsi définie peut être considérée comme une ACP sous contraintes linéaires ([13], [4], [11], [8]).

. RAO définit l'ACPVI comme la recherche de variables non corrélées (combinaisons linéaires de celles définissant Y) prédisant au mieux un ensemble de références. On peut montrer [1] que notre défintion est plus générale.

Corollaire 1.2.1

Rechercher la pseudo-métrique $Q = M_r M_r'$, où M_r (qxr) est de rang r, minimisant (1) revient à résoudre le problème suivant :

Rechercher M_r (qxr) de rang r maximisant

$$RV(W_1D, Y' M_r M_r' Y D)$$

et l'on a avec les notations du théorème 1.2

$$\text{Max } RV (W_1D , Y' M_r M_r' Y D) = \sqrt{\frac{tr(\Lambda_r^2)}{||W_1D||^2}}$$

(voir [6])

Corollaire 1.2.2

a) Si S_{YY} est inversible et si l'on pose $Q^* = S_{YY}^{-1} S_{YX}Q_1 S_{XY}S_{YY}^{-1}$ avec $S_{XY} = X_1 D Y' = S_{YX}'$ et $S_{YY} = Y D Y'$, on a :

$$||W_1D - Y'Q^* Y D||^2 = ||W_1D||^2 - ||Y'Q^* Y D||^2 = ||W_1D||^2 - ||P_Y W_1 D P_Y||^2$$

$$= ||W_1D||^2(1 - RV^2(W_1D, Y' Q^* Y D)) \qquad (2)$$

avec $P_Y = Y'(Y D Y')^{-1} Y D$ projecteur D-orthogonal sur $F_Y = Im(Y') \subset F$

b) Pour toute pseudo-métrique $\bar{Q}(pxp)$ on a :

$$||W_1D - Y'\bar{Q} Y D||^2 = ||W_1D - P_Y W_1 D P_Y||^2 + ||P_Y W_1 D P_Y - Y'\bar{Q} Y D||^2$$

c) Si L_r est la matrice des r premiers axes principaux du triplet (Y, Q^*, D) alors $M_r = Q^* L_r$.

La démonstration de a) se trouve dans [11], la partie b) a été démontrée dans [1], c) se prouve par simple substitution dans l'équation aux valeurs propres donnant les axes principaux (voir aussi [1]).

Remarques :

. A la suite du point c) la métrique Q^* est dite métrique d'ACPVI.

. Dans [11] on peut trouver le lien entre l'ACPVI et des méthodes traditionnelles de l'Analyse des Données : l'analyse des covariances partielles de LEBART, l'analyse des redondances de WOLLENBERG, l'analyse canonique, l'analyse des correspondances, l'analyse des correspondances non symétrique de LAURO, l'analyse discriminante.

2 - LES DIFFERENTES GENERALISATIONS POSSIBLES

Ce paragraphe va nous permettre de donner cinq généralisations possibles à l'ACPVI (notées ACPVI 1 à 5) déduites naturellement de la définition de l'ACPVI. Nous verrons également que sous certaines conditions ces généralisations sont équivalentes. Le paragraphe suivant sera une application des résultats énoncés dans cette partie.

2.1 - Introduction

Soients $\mathcal{K} = \{(X_k, Q_k, D) \quad ; \quad X_k(p_k x n) \ , \ Q_k(p_k x p_k) \ ,$

X_k est centré pour D ; $\forall k = 1,2,.., K\}$;

une suite de K triplets statistiques centrés mesurés sur n mêmes individus. Y un tableau (pxn) mesuré sur les mêmes individus que ceux de la suite $\mathcal{K}$. On recherche la pseudo-métrique (pxp) de rang $r \leqslant p$, associée au triplet (Y, Q, D) telle que ce dernier "résume" au mieux la suite.

Dans le paragraphe 1, nous avons vu que pour WD connu la pseudo-métrique de rang r associée à X, solution du problème de l'ACPVI, était celle :

. qui minimisait $||WD - Y' Q Y D||^2$

. qui maximisait $RV(WD, Y' Q Y D)$

. qui maximisait $RV^2(WD, Y' Q Y D)$

On se propose donc de généraliser l'ACPVI sous l'une des cinq formes suivantes :

trouver $\widetilde{Q}$ telle que :

. $\sum_{k=1}^{K} q_k ||W_k D - Y' Q Y D||^2$ soit minimum (ACPVIG-1)

. $\sum_{k=1}^{K} q_k RV(W_k D, Y' Q Y D)$ soit maximum (ACPVIG-2)

$. || \sum_{k=1}^{K} q_k W_k D - Y' Q Y D||^2$ soit minimum (ACPVIG-3)

$. \sum_{k=1}^{K} q_k [tr(W_k D Y' Q Y D)]^2$ soit maximum (ACPVIG-4)

$. \sum_{k=1}^{K} q_k [RV(W_k D, Y' Q Y D)]^2$ soit maximum (ACPVIG-5)

avec $W_k D$ l'opérateur associé au triplet (X_k, Q_k, D)

où $\{q_k ; k=1, 2, \ldots, K\}$ est une suite de coefficients positifs pondérant chaque triplet ; on pose

$$\sum_{k=1}^{K} q_k = 1$$

Il est certain que ces cinq généralisations ne sont pas les seules envisageables, mais les calculs engendrés par celles-ci semblent plus simples que pour d'autres.

2.2 - La première généralisation : l'ACPVIG-1

On recherche donc $\tilde{Q}$ minimisant

$$S(Q) = \sum_{k=1}^{K} q_k ||W_k D - Y' Q Y D||^2$$

Posons $\tilde{Q} = \sum_{k=1}^{K} q_k S_{YY}^{-1} S_{Yk} Q_k S_{kY} S_{YY}^{-1}$ avec $S_{YY} = Y D Y'$

et $S_{Yk} = Y D X_k' = S_{kY}'$

Comme nous allons le vérifier cette première généralisation implique des calculs assez pénibles que l'on peut sauter dans une première lecture ainsi la proposition 2.2 est assez technique. Toutefois le résultat important est donné par (iii) du Corollaire 2.2.

On remarque :

$$\widetilde{Q} = \sum_{k=1}^{K} q_k \, Q_k^* \quad \text{où} \quad Q_k^* \text{ est la métrique d'ACPVI de } Y/(X_k, Q_k, D)$$

Proposition 2.2

i) $$S(Q) = \sum_{k=1}^{K} q_k \, ||W_kD - P_YW_kD\,P_Y||^2 + \sum_{k=1}^{K} q_k \, ||P_YW_kDP_Y - Y'QYD||^2$$

avec $P_Y = Y' S_{YY}^{-1} Y\,D$: projecteur D-orthogonal sur

$F_Y = \mathrm{Im}(Y') \subset F$

ii) $$\sum_{k=1}^{K} q_k \, ||P_YW_k\,DP_Y - Y'QYD||^2 = \left(\sum_{k=1}^{K} q_k||P_YW_kDP_Y - Y'\,\widetilde{Q}\,Y\,D||^2\right) + $$

$$+ \,||Y'\,\widetilde{Q}\,Y\,D - Y'\,Q\,Y\,D||^2$$

iii) $$\sum_{k=1}^{K} q_k \, ||P_YW_kDP_Y - Y'\,\widetilde{Q}\,Y\,D||^2 = \left(\sum_{k=1}^{K} q_k \, ||P_YW_k\,DP_Y||^2\right) -$$

$$- \,||Y'\,\widetilde{Q}\,Y\,D||^2$$

Démonstration

i) évident car $||W_kD - Y'Q\,Y\,D||^2 = ||W_kD - P_YW_k\,DP_Y||^2 +$

$$+ \,||P_YW_kDP_Y - Y'\,Q\,Y\,D||^2$$

d'après le corollaire 1.2.2. b.

ii) $||P_YW_kDP_Y - Y'Q\,Y\,D||^2 = ||P_YW_k\,D\,P_Y - Y'\,\widetilde{Q}\,Y\,D||^2 +$

$$||Y'\,\widetilde{Q}\,Y\,D - Y'\,Q\,Y\,D||^2 + 2\,a_k$$

avec $a_k = Tr((P_Y W_k D P_Y - Y' \widetilde{Q} Y D) \quad (Y'\widetilde{Q} Y D - Y' Q Y D)) =$

$= tr(P_Y W_k D P_Y Y' \widetilde{Q} Y D) - tr(P_Y W_k D P_Y Y' Q Y D) -$

$- tr((Y' \widetilde{Q} Y D)^2) + tr(Y'\widetilde{Q} Y D Y' Q Y D)$

En substituant $\widetilde{Q}$ par sa valeur on obtient, compte tenu de ce que

$Y' Q_k Y D = P_Y W_k D P_Y$ et donc $Y' \widetilde{Q} Y D = \sum_{k=1}^{K} q_k P_Y W_k D P_Y$

$$. \ tr(P_Y W_k D P_Y Y' \widetilde{Q} Y D) = \sum_{\ell=1}^{K} q_\ell \, tr(W_k D P_Y W_\ell D P_Y) = tr(W_k D Y' \widetilde{Q} Y D) \quad (3)$$

$$. \ tr((Y' \widetilde{Q} Y D)^2) = ||Y' \widetilde{Q} Y D||^2 = \sum_{k=1}^{K} \sum_{\ell=1}^{K} q_k q_\ell \, tr(W_k D P_Y W_\ell D P_Y) \quad (4)$$

On note aussi que, comme $P_Y Y' = Y'$, $Y D P_Y = Y D$, comme on le vérifie aisément que

$tr(P_Y W_k D P_Y Y'Q Y D) = tr(W_k D Y' Q Y D)$, d'où

$$. \ tr(Y' \widetilde{Q} Y D Y' Q Y D) = \sum_{\ell=1}^{K} q_\ell \, tr(W_\ell D Y' Q Y D) \quad (5)$$

Toutes ces égalités substituées dans a_k permettent de montrer que

$\sum_{k=1}^{K} q_k a_k = 0$ ce qui établit (ii)

iii) $||P_Y W_k D P_Y - Y' \widetilde{Q} Y D||^2 = ||P_Y W_k D P_Y||^2 +$

$+ ||Y' \widetilde{Q} Y D||^2 - 2 \, tr(P_Y W_k D P_Y Y' \widetilde{Q} Y D)$

Considérons $\sum_{k=1}^{K} q_k(||Y'Q\ Y\ D||^2 - 2\ tr(P_Y\ W_k\ D\ P_Y\ Y'\ \widetilde{Q}\ Y\ D))$ (6)

d'après (3) on a : $\sum_{k=1}^{K} q_k\ tr(P_Y\ W_k\ D\ P_Y\ Y'\ \widetilde{Q}\ Y\ D) =$

$$= \sum_{k=1}^{K} \sum_{\ell=1}^{K} q_k\ q_\ell\ tr(W_k\ D\ P_Y\ W_\ell\ D\ P_Y)$$

d'après (4) on a $\sum_{k=1}^{K} q_k\ ||Y'\ \widetilde{Q}\ Y\ D||^2 = \sum_{k=1}^{K} \sum_{\ell=1}^{K} q_k\ q_\ell\ tr(W_kD\ P_Y\ W_\ell D\ P_Y)$

d'où en substituant : (6) $= - \sum_{k=1}^{K} q_k\ ||Y'\ \widetilde{Q}\ Y\ D||^2 = -||Y'\ \widetilde{Q}\ Y\ D||^2$

ce qui établit (iii).

Corollaire 2.2

(i) $S(Q) = \left(\sum_{k=1}^{K} q_k\ ||W_kD - Y'\ \widetilde{Q}\ Y\ D||^2\right) + ||Y'\ \widetilde{Q}\ Y\ D - Y'\ Q\ Y\ D||^2$

(ii) $\sum_{k=1}^{K} q_k\ ||W_kD - Y'\ \widetilde{Q}\ Y\ D||^2 = \left(\sum_{k=1}^{K} q_k||W_kD||^2\right) - ||Y'\ \widetilde{Q}\ Y\ D||^2$

(iii) $\underset{Q}{\text{Min}}\ S(Q) = S(\widetilde{Q}) = \sum_{k=1}^{K} q_k\ ||W_kD - Y'\ \widetilde{Q}\ Y\ D||^2$

et de plus $S(\widetilde{Q}) = \sum_{k=1}^{K} q_k\ ||W_kD||^2 \left(1 - \left[\sum_{\ell=1}^{K} q_\ell\ \frac{||W_\ell D||}{||W_kD||}\ RV(W_\ell D, Y'\ \widetilde{Q}\ Y\ D)\right]^2\right)$

Démonstration

(i) $||W_k D - Y' Q Y D||^2 = ||W_k D - Y' \tilde{Q} Y D||^2 + ||Y' \tilde{Q} Y D - Y' Q Y D||^2 +$

$$+ 2 \, tr((W_k D - Y' \tilde{Q} Y D) Y' R Y D)$$

avec $R = \tilde{Q} - Q$

or $tr(Y' \tilde{Q} Y D Y' R Y D) = \sum_{k=1}^{K} q_k \, tr(W_k D Y' R Y D)$ (7)

d'où $\sum_{k=1}^{K} q_k \, tr(Y' \tilde{Q} Y D Y' R Y D) = \sum_{k=1}^{K} q_k \, tr(W_k D Y' R Y D)$

ce qui démontre le point (i)

(ii) $||W_k D - Y' \tilde{Q} Y D||^2 = ||W_k D||^2 + ||Y' \tilde{Q} Y D||^2 - 2 \, tr(W_k D Y' \tilde{Q} Y D)$

en utilisant (3) et (4) on vérifie que

$\sum_{k=1}^{K} q_k \, tr(W_k D Y' \tilde{Q} Y D)$ est égal à $\sum_{k=1}^{K} q_k \, ||Y' \tilde{Q} Y D||^2$

ce qui démontre (ii).

(iii) la première partie découle de (i). Pour la deuxième :

$$S(\tilde{Q}) = \sum_{k=1}^{K} q_k \, ||W_k D||^2 - ||Y' \tilde{Q} Y D||^2 = \sum_{k=1}^{K} q_k \, (||W_k D||^2 - ||Y' \tilde{Q} Y D||^2) =$$

$$= \sum_{k=1}^{K} q_k \, ||W_k D||^2 \left(1 - \frac{\left(\sum_{\ell=1}^{K} q_\ell \, tr(W_\ell D Y' \tilde{Q} Y D)\right)^2}{||W_k D||^2 ||Y' \tilde{Q} Y D||^2}\right) =$$

$$= \sum_{k=1}^{K} q_k \, ||W_k D||^2 \left(1 - \left[\sum_{k=1}^{K} q_\ell \frac{||W_\ell D||}{||W_k D||} \times \frac{tr(W_\ell D Y' \tilde{Q} Y D)}{||W_\ell D|| \; ||Y' \tilde{Q} Y D||}\right]^2\right)$$

Remarques :

. Les points (i) et (ii) du corollaire 2.2 montrent que l'opérateur $Y' \tilde{Q} Y D$ vérifie des propriétés analogues à celles que l'on obtient pour le moment d'ordre deux d'une suite de nombres : minimal quand il est calculé au point moyen et égal à la variance.

. Le point trois du corollaire montre que minimiser S(Q) n'est en général pas équivalent à maximiser

$$\sum_{k=1}^{K} q_k \, RV(W_k D, Y' Q Y D)$$

Toutefois on remarque que si $||W_k D|| = C, \quad \forall k$, il y a équivalence.

. D'après (iii) de la proposition 2.2 on a

$$\sum_{k=1}^{K} q_k \, ||P_Y W_k D P_Y||^2 \geq ||Y' \tilde{Q} Y D||^2$$

Ce qui veut dire que la moyenne des carrés des normes des opérateurs projetés est supérieure au carré de la norme de l'opérateur "compromis" $Y' \tilde{Q} Y D$.

. Si $K = 1$ on retrouve l'ACPVI.

. Si $p_k = 1 \quad \forall k, \quad k=1,2,\dots,K$. Posons alors $X_k = X^k$, $Q_k = 1$

$$\tilde{Q} = \sum_{k=1}^{K} q_k \, S_{YY}^{-1} S_{Yk} Q_k S_{kY} S_{YY}^{-1} = S_{YY}^{-1} YD \left(\sum_{k=1}^{K} q_k \, X'^k X^k \right) D Y' S_{YY}^{-1}$$

Si on pose

$$X_{[K]} = \begin{pmatrix} \sqrt{q_1} X^1 \\ \sqrt{q_2} X^2 \\ \vdots \\ \sqrt{q_K} X^k \end{pmatrix} \quad \text{alors } X'_{[K]} X_{[K]} = \sum_{k=1}^{K} q_k \, X'^k X^k$$

Donc $\tilde{Q}$ est égal au Q^* trouvé par l'ACPVI de $Y/(X_{[K]}, Id_K, D)$

. Dans ce même cas particulier, on peut montrer que

$$RV(Y' \tilde{Q} Y D, W_k D) = \alpha \cdot R^2_{k.Y}$$

où $R^2_{k.Y}$ = carré du coefficient de corrélation multiple entre X^k et Y.

α = une constante dépendant des X^k ; k=1,..., K et de Y. ($\alpha = 1$ si $p = 1$ et $Q_k = \dfrac{1}{Var(X^k)}$)

Il est donc clair que cette première généralisation de l'ACPVI est équivalente à effectuer l'ACP du triplet suivant : $(Y, \tilde{Q}, D)$.

2.3 - Deuxième généralisation : ACPVIG-2

Ici la généralisation nous est dictée par 2 impératifs : pouvoir de manière analogue à l'ACPVI rechercher une suite de pseudo-métrique Q_r de rang croissant et montrer par une autre démarche que celle du corollaire 2.2 (iii) que l'équation (2) du corollaire 1.2.2 ne se généralise pas aussi simplement.

On recherche donc Q_r maximisant :

$$R_r(Q) = \sum_{k=1}^{K} q_k \, RV(W_k D, Y' Q_r Y D) \quad \text{où} \quad Q_r : (pxp) \text{ de rang } r \leqslant p$$

Proposition 2.3 : ACPVIG-2 d'ordre r

i) La pseudo-métrique $\bar{Q}_r = \bar{M}_r \bar{M}_r'$ avec $\bar{M}_r(pxr)$ et $rg(\bar{Q}_r) = rg(\bar{M}_r) = r$, maximisant $R_r(Q)$ est solution du système aux valeurs propres (II) suivant

$$(II) \quad \begin{cases} \left(\sum_{k=1}^{K} \alpha_k \, S_{Yk} \, Q_k \, S_{kY} \right) \bar{M}_r = S_{YY} \, \bar{M}_r \, \Lambda_r \\ \bar{M}_r' \, S_{YY} \, \bar{M}_r = \Lambda_r \end{cases}$$

avec $\alpha_k = \dfrac{q_k}{||W_k D||}$; $S_{Yk} = Y\ D\ X_k'$; $\Lambda_r = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_r)$

où Λ_r est la matrice diagonale des r premières valeurs propres, supposées distinctes, rangées en ordre décroissant.

ii) de plus $\mathrm{Max}(R_r(Q)) = \sqrt{\sum_{k=1}^{r} \lambda_k^2}$

La démonstration de cette proposition, analogue à celle du théorème 1.2 sera laissée au soin du lecteur.

Corollaire 2.3.1

Le point c) du corollaire 1.2.2 montre que l'ACPVIG-2 est équivalente à effectuer l'ACP du triplet (Y, $\bar{Q}$, D) avec

$$\bar{Q} = \sum_{k=1}^{K} \alpha_k\ S_{YY}^{-1}\ S_{Yk}\ Q_k\ S_{kY}\ S_{YY}^{-1}$$

De plus $\bar{M}_r = \bar{Q}\ L_r$ où L_r est la matrice des r premiers axes principaux du triplet (Y, $\bar{Q}$, D).

Remarque :

. On sait que $\forall (A,B) \in \mathcal{D}(F) \times \mathcal{D}(F)$; $0 \leqslant RV(A,B) \leqslant 1$

donc $\forall Q$, $\forall r \leqslant p$; $0 \leqslant R_r(Q) \leqslant 1$

donc d'après ii) on en déduit : $\forall k$; $k \leqslant r$; $\lambda_k \leqslant \dfrac{1}{\sqrt{k}}$

Corollaire 2.3.2

Pour $\bar{M}_r$ vérifiant le système (II) on a

$$\text{(i)}\quad ||Y'\bar{M}_r\ \bar{M}_r'\ Y\ D - \sum_{k=1}^{K} \alpha_k\ W_k D||^2 = ||\sum_{k=1}^{K} \alpha_k\ W_k D||^2 -$$

$$- \left\{ \sum_{k=1}^{K} q_k\ RV(W_k D,\ Y'\ \bar{M}_r\ \bar{M}_r'\ Y\ D) \right\}^2$$

(ii) $$RV\left(\sum_{k=1}^{K} \alpha_k W_k D,\ Y'\bar{M}_r\bar{M}_r' YD\right) = \sqrt{\frac{\sum_{k=1}^{K} q_k\ RV(W_k D,\ Y'\bar{M}_r\ \bar{M}_r'\ Y\ D)}{\sum_{k=1}^{K}\sum_{\ell=1}^{K} q_k q_\ell\ RV(W_k D,\ W_\ell D)}}$$

Démonstration :

(i) $$||Y'\bar{M}_r\bar{M}_r'\ Y\ D - \sum_{k=1}^{K} \alpha_k W_k D||^2 = ||Y'\bar{M}_r\bar{M}_r'\ Y\ D||^2 + ||\sum_{k=1}^{K} \alpha_k W_k D||^2 -$$

$$- 2\ tr\left(\sum_{k=1}^{K} \alpha_k\ Y'\bar{M}_r\ \bar{M}_r'\ Y\ D\ W_k D\right)$$

or $$tr\left(\sum_{k=1}^{K} \alpha_k\ Y'\bar{M}_r\ \bar{M}_r'\ Y\ D\ W_k D\right) = \sum_{k=1}^{K} \alpha_k\ tr(\bar{M}_r'\ S_{Yk}\ Q_k\ S_{kY}\ \bar{M}_r) =$$

$$= tr(\bar{M}_r'\ S_{YY}\ \bar{M}_r\ \Lambda_r) = tr(\Lambda_r^2) = ||Y'\ \bar{M}_r\ \bar{M}_r'\ Y\ D||^2 \qquad (8)$$

donc $$||Y'\bar{M}_r\ \bar{M}_r'\ Y\ D - \sum_{k=1}^{K} \alpha_k W_k D||^2 = ||\sum_{k=1}^{K} \alpha_k W_k D||^2 - ||Y'\bar{M}_r\ \bar{M}_r'\ YD||^2 =$$

$$= ||\sum_{k=1}^{K} \alpha_k\ W_k D||^2 - \left\{\sum_{k=1}^{K} q_k\ RV(W_k D,\ Y'\ \bar{M}_r\ \bar{M}_r'\ Y\ D)\right\}^2$$

(ii) $$RV\left(\sum_{k=1}^{K} \alpha_k W_k D,\ Y'\bar{M}_r\ \bar{M}_r'\ YD\right) = \frac{tr\left(\sum_{k=1}^{K} \alpha_k W_k D\ Y'\ \bar{M}_r\ \bar{M}_r'\ Y\ D\right)}{\sqrt{tr\left(\left(\sum_{k=1}^{K} \alpha_k W_k D\right)^2\right)\ tr((Y'\bar{M}_r\bar{M}_r' YD)^2)}}$$

or d'après (8) on peut écrire

$$RV\left(\sum_{k=1}^{K} \alpha_k W_k D,\ Y'\ \bar{M}_r\bar{M}_r'\ YD\right) = \sqrt{\frac{tr\left(\sum_{k=1}^{K} \alpha_k W_k D\ Y'\bar{M}_r\bar{M}_r'\ Y\ D\right)}{tr\left(\left(\sum_{k=1}^{K} \alpha_k\ W_k\ D\right)^2\right)}}$$

En effectuant les calculs pour le numérateur et le dénominateur, et en remplaçant α_k par sa valeur, on trouve le résultat annoncé.

C.Q.F.D

Remarques :

. La relation (i) de la proposition 2.3 montre que la recherche de Q maximisant $R_r(Q)$ est équivalente à la recherche de Q minimisant une combinaison linéaire particulière des opérateurs W_kD (chaque W_kD est pondéré par α_k). Cette remarque justifie la troisième généralisation de l'ACPVI (où chaque W_kD est pondéré par q_k).

. Le point (ii) du corollaire 2.3.2 est le seul cas connu où l'on peut décomposer un RV, entre un opérateur et une combinaison linéaire d'opérateurs, comme fonction des RV entre opérateurs initiaux.

. Si K = 1, on retrouve l'ACPVI.

. Si $p_k = 1$, $\forall k$, $k=1,2,\dots,K$. Posons alors $X_k = X^k$, $Q_k = 1$

$$X_{[K]} = \begin{pmatrix} \sqrt{q_1}X^1 \\ \sqrt{q_2}X^2 \\ \vdots \\ \sqrt{q_K}X^K \end{pmatrix} \qquad \text{et} \quad Q_{[K]} = \operatorname{diag}\left(\frac{1}{s_1^2}, \frac{1}{s_2^2}, \dots, \frac{1}{s_K^2}\right)$$

$$\text{où} \quad s_k^2 = \operatorname{var}(X^k)$$

alors l'ACPVIG-2 d'ordre r est équivalente à l'ACPVI d'ordre r de $Y/ (X_{[K]}, Q_{[K]}, D)$.

. La comparaison de (i) du corollaire 2.3.2 et (iii) du corollaire 2.2 montre que si $||W_kD|| = C$ $\forall k$, $k=1,\dots,K$, la minimisation de S(Q) est équivalente à la maximisation de $R_r(Q)$ pour $rg(\widetilde{Q}) = rg(\widetilde{Q}_r) = r$

Cas où Q_k : Id_k ; $k = 1, 2, \ldots, K$

Dans ce paragraphe on notera

$X_{i_k} = i^{\text{ème}}$ variable du $k^{\text{ième}}$ tableau

$C_j = (\bar{M}_r' Y)_j = j^{\text{ième}}$ composante principale

$= \bar{M}_j' Y$

$\text{cov}(X_{i_k}, C_j) = X_{i_k} D Y' \bar{M}_j$

donc $\sum_{i_k=1}^{p_k} \left[\text{cov}(X_{i_k}, C_j)\right]^2 = \bar{M}_j' Y D X_k' X_k D Y' \bar{M}_j = \bar{M}_j' S_{Yk} S_{kY} \bar{M}_j$

d'où $\sum_{k=1}^{K} \sum_{i_k=1}^{p_k} \alpha_k \left[\text{cov}(X_{i_k}, C_j)\right]^2 = \lambda_j^2$ (d'après (i) de la proposition 2.3)

Ce qui nous permet d'énoncer la proposition suivante

Proposition 2.4

La recherche de Z_j, combinaison linéaire des variables définissant Y vérifiant les conditions suivantes

. $\sum_{k=1}^{K} \sum_{i_k=1}^{p_k} \alpha_k \left[\text{cov}(X_{i_k}, Z_j)\right]^2$ maximum (9)

. $Z_j D Z_i' = \delta_i^j$ (symbole de KRONECKER)

avec $\alpha_k = \dfrac{q_k}{||S_{kk}||}$ et $S_{kk} = X_k D X_k'$

est équivalente à la résolution du système (II). De plus, au maximum, (9) est égal à λ_j^2 et $Z_j = \frac{1}{\sqrt{\lambda_j}} M_j' Y$ (si $\lambda_j \neq 0$).

Remarques :

. Si Q_k ; k=1, 2, ..., K sont quelconques (9) devient (9)'

$$\sum_{m=1}^{k} \sum_{i_k=1}^{p_m} \sum_{i_\ell=1}^{p_m} \alpha_m Q_m (i_\ell, i_k) \operatorname{cov}(X_{i_k}, Z_j) \operatorname{cov}(X_{i_\ell}, Z_j)$$

. Si k=1 alors (9) nous fournit bien la définition de l'ACPVI donnée par RAO.

2.4. Troisième généralisation : ACPVIG-3

Recherche de Q minimisant

$$D_r(Q) = ||Y' Q Y D - \sum_{k=1}^{K} q_k W_k D||^2$$

Proposition 2.5 : ACPVIG-3 d'ordre r

La pseudo-métrique $\hat{Q}_r = \hat{M}_r \hat{M}_r'$ avec $\hat{M}_r(qxr)$ et $rg(\hat{Q}_r) = rg(\hat{M}_r) = r$, minimisant $D_r(Q)$ est solution du système (III) suivant :

$$\text{(III)} \quad \begin{cases} \left(\sum_{k=1}^{K} q_k S_{Yk} Q_k S_{kY} \right) \hat{M}_r = S_{YY} \hat{M}_r \Lambda_r \\ \hat{M}_r' S_{YY} \hat{M}_r = \Lambda_r \end{cases}$$

avec $S_{Yk} = Y D X_k' = S_{kY}'$; $\Lambda_r = \operatorname{diag}(\lambda_1, \lambda_2, \ldots, \lambda_r)$

où Λ_r est la matrice diagonale des r premières valeurs propres, supposées distinctes, rangées en ordre décroissant.

La démonstration devient évidente en substituant dans le système (I) du Théorème 1.2 W_1D par

$$\sum_{k=1}^{K} q_k W_k D$$

Corollaire 2.5.1

Il est donc clair que l'ACPVIG-3 d'ordre r est équivalente à effectuer l'ACP de $(Y, \tilde{Q}, D)$ et $\hat{M}_r = \tilde{Q}\, L_r$ où L_r est la matrice des r premiers axes principaux. Donc l'ACPVIG-1 est équivalente à l'ACPVIG-3 d'ordre r pour $r = rg(\tilde{Q})$.

Remarques :

. Il faut noter que l'unique différence entre les systèmes (II) et (III) provient des coefficients α_k et q_k. Donc si $\forall k$, $k=1,2,\dots, K$ $||W_kD|| = 1$, la minimisation de $D_r(Q)$ est équivalente à la maximisation de $R_r(Q)$.

. Au minimum de $D_r(Q)$ on a :

$$D_r(\hat{Q}_r) = ||\sum_{k=1}^{K} q_k W_k D||^2 \left(1 - RV^2\left(\sum_{k=1}^{K} q_k W_k D,\ Y'\hat{M}_r\ \hat{M}_r'\ Y\ D\right)\right)$$

par simple application de (2). Toutefois le RV ne se décompose selon une relation analogue à la relation (ii) du corollaire 2.3, que si $||W_kD|| = 1$; $\forall k = 1,2,\dots, K$

La propriété suivante montre les liens entre les généralisations 2 et 3.

Corollaire 2.5.2

$$D_r(Q) = ||Y'QYD - \sum_{k=1}^{K} \alpha_k W_k D||^2 - 2\, R_r(Q) \times \left(\sum_{k=1}^{K} q_k(||W_kD|| - 1) \times\right.$$

$$\left. \times\, RV(Y'QYD,\ W_kD)\right) + \sum_{j=1}^{K} \sum_{k=1}^{K} q_j q_k(||W_jD||-1)(||W_kD||-3)\ RV(W_jD,\ W_kD)$$

Démonstration

$$D_r(Q) = ||Y^{'}QYD - \sum_{k=1}^{K} \alpha_k W_k D||^2 + ||\sum_{k=1}^{K} (\alpha_k - q_k)\, W_k D||^2 - 2a \qquad (10)$$

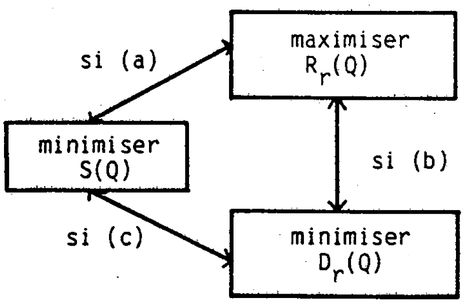
avec
$$a = \sum_{k=1}^{K} (\alpha_k - q_k)\, tr(Y^{'}QYD\, W_k D) + \sum_{j=1}^{K} \sum_{k=1}^{K} \alpha_k q_j\, tr(W_k D\, W_j D) -$$

$$- \sum_{j=1}^{K} \sum_{k=1}^{K} \alpha_k\, \alpha_j\, tr(W_k D\, W_j D)$$

En développant dans (10) le deuxième terme et en substituant la valeur de a, on obtient le résultat annoncé.

Remarques :

. Les équivalences entre les trois premières généralisations peuvent se résumer par le dessin suivant :



(a) : $||W_k D|| = C$; $k=1,2,\ldots, K$
et $rg(\tilde{Q}) = r = rg(\tilde{Q}_r)$

(b) : $||W_k D|| = 1$; $k=1,2,\ldots, K$

(c) : $||W_k D|| = C$; $k=1,2,\ldots, K$
et $rg(\tilde{Q}) = r = rg(\hat{Q}_r)$

. On peut également énoncer une proposition analogue à la proposition 2.4 de la troisième généralisation. Dans son énoncé il suffit de remplacer α_k par q_k.

2.5 - Quatrième généralisation : ACPVIG-4

On recherche $\underline{Q} = \sum_{k=1}^{K} \ell_k \, Q_k^*$ maximisant :

$$C(Q) = \sum_{k=1}^{K} q_k \left[tr(W_k D \, Y' Q \, Y \, D \right]^2$$

Remarque :

. Pour ne pas être gêné dans les calculs par les coefficients q_k dans la suite des calculs, on conviendra de les intégrer dans chaque métrique Q_k, qui deviendra $\sqrt{q_k}\, Q_k$.

Proposition 2.6

i) Les coefficients (ℓ_k ; k=1,2,..., K) de la pseudo-métrique $\underline{Q}$ maximisant C(Q) sont donnés par les K composantes du vecteur propre normé à l'unité, associé à la plus grande valeur propre, obtenue par diagonalisation de la matrice symétrique $\mathcal{C}$.

avec $\mathcal{C} = (\mathcal{C}_{ij})$ i,j = 1,2,..., K

et $\mathcal{C}_{ij} = tr(W_i D \, P_X \, W_j D \, P_X)$.

ii) Au maximum on a :

$\text{Max} \;\; C(Q) = \lambda_1^2$ avec λ_1 plus grande valeur propre de $\mathcal{C}$

$\text{Max} \;\; ||Y' QYD||^2 = \lambda_1$

Démonstration

i) si $Q = \sum_{k=1}^{K} \ell_k Q_k^*$

$$tr(W_k D Y' Q Y D) = \sum_{k=1}^{K} \ell_j \; tr(W_k D Y' Q_j^* Y D)$$

d'où $$\left[tr(W_k D Y' D Y D)\right]^2 = \sum_{i=1}^{K} \sum_{j=1}^{K} \ell_i \ell_j \; tr(W_i D P_Y W_k D P_Y) \times$$

$$\times \; tr(W_j D P_Y W_k D P_Y)$$

donc $$\sum_{k=1}^{K} \left[tr(W_k D Y' Q Y D)\right]^2 = \sum_{i=1}^{K} \sum_{j=1}^{K} \ell_i \ell_j (\mathcal{C}^2)_{ij} = \ell' \mathcal{C}^2 \ell$$

avec $\ell' = (\ell_1, \ell_2, \ldots, \ell_K)$

Il est clair que maximiser $\ell' \mathcal{C}^2 \ell$ n'a de sens que par exemple si $\ell'\ell = 1$ donc d'après le théorème de COURANT-FISHER ([12], [4], [11]) le maximum de C(Q) sera obtenu par $\ell' = (\ell_1, \ell_2, \ldots, \ell_k)$ vecteur propre normé de $\mathcal{C}^2$ associé à la plus grande valeur propre de $\mathcal{C}^2$, donc de $\mathcal{C}$.

ii) Première égalité évidente. La deuxième correspond au calcul suivant :

$$||Y' QYD||^2 = \sum_{i=1}^{K} \sum_{j=1}^{K} \ell_i \ell_j \; tr(W_i DP_Y W_j DP_Y) = \ell' \mathcal{C} \ell$$

et on conclut par le même argument que i).

Il faut remarquer que la symétrie de $\mathcal{C}$ est évidente. De plus puisque $\mathcal{C}_{ij} = tr(W_i D P_Y W_j D P_Y) \geq 0$, d'après le théorème de PERRON-FROBENIUS il en découle que les composantes du premier vecteur propre de $\mathcal{C}$ sont toutes de même signe, on peut alors les choisir positives, et $\underline{Q}$ est une pseudo-métrique.

C.Q.F.D.

Remarques :

. Il est évident que effectuer l'ACPVIG-4 est équivalent à effectuer l'ACP du triplet $(Y, \underline{Q}^*, D)$ avec

$$\underline{Q}^* = \sum_{k=1}^{K} \sqrt{q_k}\, \ell_k\, Q_k^*$$

. Le point (ii) de la proposition précédente montre que la pseudo-métrique $\underline{Q}$ maximisant $C(Q)$ est aussi parmi les combinaisons linéaires des métriques Q_k^* celle de norme maximale.

. Cette méthode n'est rien d'autre que la méthode STATIS (voir [6]) appliquée à la suite $\mathcal{K}' = \{(X_k\, Q_k^*, D)\ ;\ k=1,2,\ldots,K\}$ déduite de $\mathcal{K}$ par ACPVI.

. Dans le cas où $p_k = 1$; $k=1,2,\ldots, K$. Si l'on pose $Q_{[K]} = \text{diag}(\ell_1, \ell_2, \ldots, \ell_k)$ alors l'ACPVIG-4 est analogue à l'ACPVI de $Y/(X_{[K]}, Q_{[K]}, D)$. On note que $Q_{[K]}$ correspond à la métrique diagonale associée à l'opérateur de norme maximale.

. Comme l'on a intégré les q_k dans chaque métrique, en fait $\mathcal{C}_{ij} = \sqrt{q_i q_j}\ \text{tr}(W_i D\, P_Y\, W_j\, D\, P_Y)$.

2.6 - Cinquième proposition : ACPVIG-5

Recherche de $\underset{\sim}{Q} = \sum_{k=1}^{K} \ell_k\, Q_k^*$ maximisant

$$R^2(\underset{\sim}{Q}) = \sum_{k=1}^{K} q_k\, RV^2(W_k D,\ Y'\underset{\sim}{Q}\, Y\, D)$$

Proposition 2.7

i) Les coefficients $\{\ell_k \; ; \; k=1,2,\dots,K\}$ de $\underset{\sim}{Q}$ maximisant $R^2(Q)$ sont donnés par les K composantes du vecteur propre normé à l'unité, associé à la plus grande valeur propre, obtenue par diagonalisation de la matrice $B\,\mathcal{C}$

avec $B = \text{diag}\left(\dfrac{q_k}{||W_k D||^2}\right) \; ; \quad k=1,2,\dots, K$

ii) au maximum on a : $\text{Max } R^2(\underset{\sim}{Q}) = \lambda_1$

avec λ_1 plus grande valeur propre de $B\,\mathcal{C}$.

La démonstration de cette proposition, analogue à celle de la proposition 2.6, sera laissée au soin du lecteur.

Remarques :

. Si $||W_k D|| = \sqrt{q_k}$ alors $B = I_K$ donc maximiser $R^2(Q)$ est équivalent à maximiser $C(Q)$

. On obtient le même résultat si $\forall k$, $q_k = \frac{1}{K}$ et $||W_k D|| = C$

. Il est clair que $0 \leqslant \lambda_1 \leqslant 1$

3 - APPLICATIONS DE CES GENERALISATIONS AUX VARIABLES QUALITATIVES

3.1 - Soient $u_1, u_2, \dots, u_K$, K variables qualitatives respectivement à I_1, $I_2, \dots, I_K$ modalités, dont les matrices des variables indicatrices mesurées sur n individus sont $U_1(I_1 x n)$, $U_2(I_2 x n)$, ..., $U_K(I_K x n)$. Soit $D = \frac{1}{n} I_n$ la métrique des poids.

Soit u_0 une (K+1)ième variable qualitative à I_0 modalités mesurée sur les mêmes individus dont la matrice associée est $U_0(I_0 x n)$. Soit Y une matrice de p variables quantitatives centrées mesurées sur les mêmes individus.

On posera

. $U = \begin{pmatrix} U_1 \\ \vdots \\ U_K \end{pmatrix}$ de dimensions $\left(\left(\sum_{k=1}^{K} I_k\right) x\ n\right)$

. $D_k = U_k D U_k'$

. $\mathcal{K}_M = \{(U_k, D_k^{-1}, D) \quad ; \quad k=1,2,\ldots,K\}$

. $P_k = U_k' D_k^{-1} U_k D$, opérateur associé au triplet (U_k, D_k^{-1}, D)

Proposition

Les trois conditions suivantes sont équivalentes :

1) maximiser : $\frac{1}{K} \sum_{k=1}^{K} \sqrt{I_k}\ RV(P_k, U_0' M_r M_r' U_0 D)$

2) minimiser : $\left|\left|\frac{1}{K} \sum_{k=1}^{K} P_k - U_0' M_r M_r' U_0 D\right|\right|^2$

3) M_r est égal à la juxtaposition des r premiers axes principaux fournis par l'analyse des correspondances du sous tableau de BURT $\mathcal{B}$, avec $\mathcal{B} = U_0 D U'$.

Démonstration immédiate d'après les systèmes (II) et (III).

Remarques :

. Pour 1) on note que q_k étant égal à $\frac{1}{K}\sqrt{I_k}$, on n'a pas la condition $\sum_{k=1}^{K} q_k = 1$, ce qui n'influe pas sur la suite des calculs.

. La démonstration repose sur le fait que l'analyse des correspondances du sous tableau de BURT est l'ACP du triplet

$(\frac{1}{K}\ D_0^{-1}\ \mathcal{B}\, Q^{-1},\ K\ D_0,\ Q)$ avec

$$Q = \begin{pmatrix} D_1 & & & 0 \\ & D_2 & & \\ & & \ddots & \\ 0 & & & D_K \end{pmatrix}$$

(vérification immédiate en calculant les opérateurs VQ et WD associés au triplet)

. Dans la proposition les tableaux U_k ; k=0,1,..., K n'ont pas été centrés.

. G. SAPORTA, dans un article récent [12] à l'aide de pondération de projecteurs généralise la pratique des opérateurs dans un but descriptif puis dans le cas prévisionnel arrive à une impasse. Il semble évident ici que la raison de son échec vient de l'utilisation systématique de l'opérateur de projection. En effet nous venons de voir que ce problème posé en termes de recherche de métriques admet un certain nombre de solutions.

3.2 - ACPVIG de Y par rapport à $\mathcal{K}_M$

Proposition

Effectuer l'ACPVIG d'ordre r de Y par rapport à $\mathcal{K}_M$ c'est rechercher la pseudo-métrique $Q_r = M_r M_r'$ de rang r vérifiant l'une des trois conditions équivalentes suivantes :

1) maximiser : $\sum_{k=1}^{K} \sqrt{I_k}\ \ RV(P_k,\ Y'\ M_r\ M_r'\ Y\ D)$

2) minimiser : $\left|\left| \sum_{k=1}^{K} P_k - Y' M_r M_r' Y D \right|\right|^2$

3) effectuer l'ACP du triplet $(\mu_1, \mu_2, \ldots, \mu_K), V^{-1}, Q)$

avec $\mu_K = Y D U_k' D_k^{-1}$; $V = Y D Y'$

$$Q = \begin{pmatrix} D_1 & & & 0 \\ & D_2 & & \\ & & \ddots & \\ 0 & & & D_K \end{pmatrix}$$

et en posant $M_r = (\sqrt{\lambda_1}, L_1, \sqrt{\lambda_2}\, L_2, \ldots, \sqrt{\lambda_r} L_r)$

où $L_k = k^{\text{ième}}$ facteur de l'ACP.

De plus la représentation des individus dans les deux ACPVIG est identique au coefficient $\sqrt{\lambda_i}$ près à celle fournie par l'ACP.

Remarque :

Une telle ACPVIG est équivalente à une analyse de variance multivariée à plusieurs facteurs sans interactions.

CONCLUSION

Cet article bien que très formel nous a permis de montrer que des généralisations de l'ACPVI pouvaient être envisagées. De telles généralisations sont possibles en utilisant deux outils fondamentaux : les métriques et les opérateurs. Loin d'apporter que des réponses il est clair toutefois que notre effort à venir doit envisager la notion d'interaction, utile pour l'analyse des correspondances multiples.

BIBLIOGRAPHIE

[1] BONIFAS, L., ESCOUFIER, Y., GONZALEZ, P.L., SABATIER, R., "Choix de variables en analyse en composantes principales", R.S.A. XXXII, 1984, n° 2, pp. 5-15.

[2] CAILLIEZ, F., PAGES, J.P., Introduction à l'analyse des données, Smash, 9, rue Duban, 75016 Paris, 1979.

[3] CAZES, P., BAUMERDER, A., BONNEFOU, S., PAGES, J.P., "Codage et analyse des tableaux logiques. Introduction à la pratique des variables qualitatives",Cahiers du BURO, Université Paris VI, 1977, n° 27.

[4] CROQUETTE, A., Quelques résultats synthétiques en analyse des données multidimensionnelles. Optimalité et métriques à effets relationnels, Thèse de 3e cycle, Université Paul Sabatier, Toulouse, 1980.

[5] D'AMBRA, L., LAURO, N., "L'analyse non symétrique des correspondances", Troisièmes Journées Internationales d'Analyse des Données et Informatique, I.N.R.I.A., 1983.

[6] ESCOUFIER, Y., Operators related to a data matrix. Recent developments in statistics, J.R. Barra et al. editors, North Holland Publishing Company, 1977.

[7] LEBART, L., MORINEAU, A., FENELON, J.P., Traitement des données statistiques : Méthodes et Programmes, Dunod, 1979.

[8] OBADIA, J., "L'analyse en composantes explicatives", R.S.A., XXVI, 1978, n° 4, pp. 5-28.

[9] RAO, C.R., "The use and the interpretation of principal component analysis in applied research", Sankya, 1964, Ser. A, 26, pp. 329-359

[10] ROBERT, P., ESCOUFIER, Y., Choosing variables and metrics optimizing the RV coefficient, Academis Press, INC-JS, Rustagi, 1979.

[11] SABATIER, R., Approximation d'un tableau de données, application à la reconstitution des paléoclimats, Thèse de 3ème cycle, Université des Sciences et Techniques du Languedoc, Montpellier, 1983.

[12] SAPORTA, G., "Pondération optimale de variables qualitatives en Analyse des Données", Statistique et Analyse des Données, 1979, n° 3, pp. 19-31.

[13] SCHEKTMAN, Y., Contribution à la mesure en facteurs dans les sciences expérimentales et à la mise en oeuvre des calculs statistiques, Thèse d'Etat, Université Paul Sabatier, Toulouse, 1978.

[14] WOLLENBERG, A.L.V.D., "Redundancy analysis an alternative for canonical correlation analysis", Psychometrika, 1977, Vol. 42, n° 2, pp. 207-221.