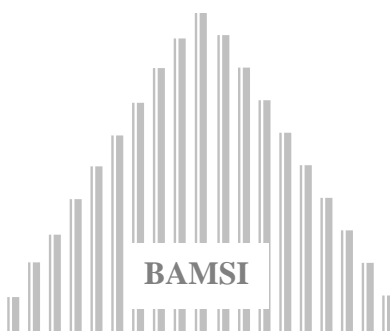


BUREAU D'APPLICATION DES METHODES
STATISTIQUES ET INFORMATIQUES

DT 01/2001

Statis : une méthode d'analyse conjointe
de plusieurs tableaux de données

Samuel AMBAPOUR



BAMSI B.P. 13734 Brazzaville

**Statis : une méthode d'analyse conjointe
de plusieurs tableaux de données**

*Samuel AMBAPOUR**

Résumé : Nous présentons et utilisons une méthode d'analyse conjointe de plusieurs tableaux de données : la méthode STATIS (structuration des tableaux à trois indices de la statistique). Introduite par Y. Escoufier et H. l'Hermier des Plantes, elle s'appuie sur les propriétés des opérateurs associés à différents tableaux de données. Elle est mise en œuvre ici pour étudier l'évolution des tableaux entrées-sorties (T.E.S) du Congo.

Mots clés : Analyse conjointe, Analyse des correspondances, Analyse en composantes principales, interstructure, intrastructure, compromis, opérateur, produit scalaire, T.E.S.

*SNE BP 95, Brazzaville, TEL : 81 05 69, FAX : 81 05 69

BAMSI BP 13734, Brazzaville, FAX : 81 58 64

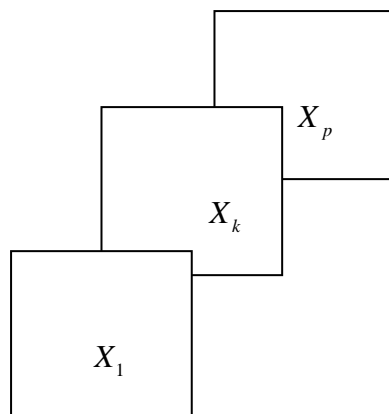
E-mail : ambapour_samuel@yahoo.fr

Je remercie J. Mata et J.C. Okandza pour leurs commentaires.

Ces documents de travail ne reflètent pas la position du BAMSI mais n'engagent que ses auteurs.

Introduction

L'économiste dispose d'un ensemble de tableaux d'échanges correspondant à plusieurs années qu'il souhaite analyser. Il peut pour ce faire, utiliser les techniques classiques d'analyse des données (Ambapour-Kosso, 1992). En effet, les tableaux d'échanges peuvent être considérés comme des tableaux de contingence particuliers associés aux deux caractères qualitatifs : « être envoyé par ... » et « être reçu par ». Dans ce cas, l'analyse factorielle des correspondances permet d'obtenir des images résumées des tableaux considérés isolément. L'étude globale des différents tableaux peut se faire ensuite par la comparaison des images (plans factoriels) correspondant à des années successives. Ce qui permet de saisir ou d'expliquer l'évolution de la structure des échanges. Mais cette façon de procéder peut présenter des inconvénients. L'économiste risque d'être submergé par le nombre important de représentations. Il peut également penser que, chacun des tableaux ne donne qu'une vision partielle de la réalité (Escoufier, 1985). Pour ces deux raisons, il va alors chercher un ensemble de représentation unique. Ce vœu est pleinement comblé par l'existence aujourd'hui de différentes techniques d'analyse conjointe de tableaux de données. On présente dans ce texte, l'une d'entre elles : la méthode STATIS (structuration des tableaux à trois indices de la statistique). C'est une méthode exploratoire d'analyse des données qui s'applique à des données quantitatives : p tableaux de mesures X_k ont été observés en différentes occasions sur n individus.



Cette méthode de répondre aux objectifs suivants :

- déceler les tableaux qui se ressemblent ;
- fournir un tableau résumé de l'ensemble ;
- décrire les écarts entre tableaux en prenant comme référence ce tableau résumé ;
- déterminer si les écarts sont dus aux variables ou aux individus.

Lorsque le temps indice les situations expérimentales, l'objectif peut être de reconnaître les évolutions.

Dans le premier paragraphe de ce texte méthodologique, on définit les notations. On caractérise ensuite les situations rencontrées dans l'application de cette méthode dans le deuxième paragraphe ; les équivalences entre tableaux et les opérateurs qui permettent de rendre opérationnelles ces équivalences sont également introduits. Le troisième paragraphe développe les procédures mathématiques utilisées pour satisfaire les objectifs énumérés précédemment. Enfin dans le dernier paragraphe, la méthode STATIS est appliquée pour étudier l'évolution des T.E.S congolais. Seuls les résultats de l'interstructure sont présentés.

1. Notations – Schéma de dualité

Les notations utilisées sont celles de Caillez et Pages (1976). On note par : $X = \{x_j^i \mid i \in I; j \in J\}$, le tableau de données du type individus caractères ; où :

I est appelé ensemble des individus ; avec $CardI = n$;

J Ensemble des caractères ; avec $CardJ = p$.

Ce tableau de format (p, n) admet deux représentations :

- une représentation du nuage M des n "points individus" dans l'espace vectoriel $E = R^p$; E est "l'espace des individus" ;
- une représentation du nuage N des p "points caractères" dans l'espace vectoriel $F = R^n$; F est "l'espace des caractères".

E et F sont munis de leur base canonique :

$$\left\{ \underline{e}_j \mid j = 1, \dots, p \right\}$$

$$\left\{ \underline{f}_i \mid i = 1, \dots, n \right\}$$

E^* et F^* espaces duaux de E et F auront pour base :

$$\left\{ \begin{array}{l} \underline{e}_j^* \\ \underline{f}_i^* \end{array} \mid \begin{array}{l} j = 1, \dots, p \\ i = 1, \dots, n \end{array} \right\}$$

On peut alors considérer le tableau X comme la matrice associée à l'application linéaire de F^* dans E définie par :

$$X(\underline{f}_i^*) = \underline{x}_i$$

De même la transposée X' est la matrice de l'application linéaire de E^* dans F est définie par :

$$X'(\underline{e}_j^*) = \underline{x}^j$$

Pour mesurer les proximités entre individus (ou entre caractères), on munit E (respectivement F) d'une métrique euclidienne M (respectivement N) :

$M : E \rightarrow E^*$; M définit la distance entre individus ;

$N : F \rightarrow F^*$; N définit la distance entre les caractères. Généralement, on choisit pour N , la métrique des poids D_p définie par la matrice diagonale :

$$D_p = \begin{pmatrix} p_1 & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & p_n \end{pmatrix}$$

X est centré pour D_p si $X D \underline{1} = 0$

Où :

$$\underline{1} = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$$

Et finalement, au tableau des données X correspond le schéma de dualité ci-après résumant les applications précédentes :

$$\begin{array}{ccc} & \xleftarrow{X} & \\ E = R^p & & F^* \\ \downarrow M & \uparrow V & \downarrow W \\ & & \uparrow D_p \\ E^* & \xrightarrow{X'} & F = R^n \end{array}$$

avec :

$$W = X' M X \quad \text{et} \quad V = X D_p X'$$

tels que :

$$V(\underline{e}_j^*, \underline{e}_k^*) = D_p(\underline{x}^j, \underline{x}^k) = D_p(X' \underline{e}_j^*, X' \underline{e}_k^*)$$

$$W(\underline{f}_i^*, \underline{f}_k^*) = M(\underline{x}_i, \underline{x}_k) = M(X \underline{f}_i^*, X \underline{f}_k^*)$$

W est la matrice de produits scalaires entre individus de E .

V est la matrice de produits scalaires entre variables de F . Si X est centré, V est la matrice des variances et covariances entre les p variables du tableau X .

2. Les données — Caractérisation des situations rencontrées

On dispose de k tableaux de données quantitatives obtenus à différentes époques, et que l'on désire comparer. Lorsque l'on applique la méthode STATIS, on peut se trouver dans l'une des trois situations décrites ci-dessous. On indique par ailleurs, comment à chaque cas, on doit associer à chaque tableau, un élément caractéristique.

2.1. Les situations

2.1.1. Situation 1

k tableaux de mesures concernant les mêmes individus affectés des mêmes poids. On a donc k triplets (X_k, M_k, D) , où X_k la matrice $p_k \times n$ centrée pour D et M_k une métrique sur R^{p_k} . On pose :

$$W_k = X' M_k X$$

2.1.2. Situation 2

k tableaux de mesures résultant de l'observation des mêmes variables. On a donc k triplets (X_k, M, D_k) , où X_k est une matrice $p \times n_k$ centrée pour D , et M une métrique sur R^p . On pose :

$$V_k = X_k D_k X_k'$$

2.1.3. Situation 3

k tableaux de mesures résultant de l'observation des mêmes variables sur les mêmes individus. C'est le cas par exemple des données chronologiques. La situation 3 comme

on peut le constater, englobe donc 1 et 2. Ainsi, suivant que l'on s'intéresse aux individus ou aux variables, on se considère dans l'une des deux premières situations décrites. A chacune de ces trois situations, correspond une stratégie appropriée. La première privilégie les positions relatives des individus ; la deuxième étudie les relations entre les variables. La troisième situation qui permet d'étudier des tableaux croisant les mêmes individus avec les variables, applique les deux premières stratégies.

2.2. Equivalence entre tableaux de données : élément caractéristique d'une étude

2.2.1. Définition d'une étude

Une étude est un triplet (X, M, D) , où :

X est le tableau de données croisant individus et variables ;

M est la métrique permettant le calcul des distances entre individus ;

D est la métrique des poids, permettant le calcul des distances entre variables.

Une étude peut être caractérisée ou représentée par deux "objets" :

- l'objet $W = X' M X$;

- où l'objet $V = X D X'$

2.2.2. Opérateurs

On est dans la situation 1, c'est-à-dire que l'on veut comparer les individus d'un tableau à l'autre. On compare ainsi les matrices W_k . L'élément caractéristique d'une étude sera ici W_k . Au triplet (X_k, M_k, D) , on associe l'opérateur $U_k = W_k D$. Cet opérateur permet d'introduire très simplement des équivalences entre tableaux de données dans une optique de description.

Equivalence 1

Les triplets (X_1, M_1, D) et (X_2, M_2, D) sont équivalents si les opérateurs associés $U_1 = W_1 D$ et $U_2 = W_2 D$ sont identiques. Si l'on fait l'analyse en composantes principales des deux triplets, on aura la même représentation des individus.

Equivalence 2

Les triplets (X_1, M_1, D) et (X_2, M_2, D) sont équivalents si les opérateurs associés sont homothétiques :

$$\begin{aligned} U_1 &= \alpha U_2 \\ W_1 D &= \alpha W_2 D \quad \alpha > 0 \end{aligned}$$

Si l'on fait l'analyse en composantes principales des deux triplets, on aura la représentation des individus homothétiques.

Produit scalaire et distance entre opérateurs

Tout opérateur $U = WD_p$ est une application D_p -symétrique de F dans F ($F = R^n$).

L'ensemble de ces applications définit un sous espace G de $\mathcal{L}(F, F)$ (ensemble des applications linéaires de F dans F). La forme bilinéaire symétrique P définie sur $\mathcal{L}(F, F)$ par (Pages, Escoufier, Cazes, 1976) :

$A \in \mathcal{L}(F, F)$, $B \in \mathcal{L}(F, F) \Rightarrow P(A, B) = \text{tr}(AB)$ est positive sur G si on la restreint à G car :

$$U \in G \Rightarrow P(U, U) = \text{tr}(U^2) = \|U\|^2 = \sum \lambda_i^2$$

où λ_i désigne la i -ème valeur propre de l'opérateur U qui étant D_p symétrique est diagonalisable. La forme bilinéaire P permet donc de définir la distance et l'angle (par le cosinus) entre deux opérateurs :

- Distance

$$U_1 \in G ; U_2 \in G ; \Rightarrow d(U_1, U_2) = \|U_1 - U_2\|$$

- Angle

$$U_1 \in G ; U_2 \in G ; \Rightarrow R_{u_1 u_2} = \frac{P(U_1, U_2)}{\|U_1, U_2\|} = \frac{\text{tr}(U_1 U_2)}{\sqrt{\text{tr}(U_1)^2 \text{tr}(U_2)^2}}$$

On peut faire remarquer que le produit scalaire permet de retrouver les indices classiques utilisés par les statisticiens pour mesurer les liaisons entre caractères. Ainsi, on note :

$\text{Cov}(A, B)$ pour $\text{tr}(AB)$; l'équivalent d'une covariance ;

$\text{Var}(A)$ pour $\text{tr}(A)^2$; l'équivalent d'une variance ;

$R(A, B)$ pour $\text{tr}(AB) / \left(\text{tr}(A)^2 \text{tr}(B)^2 \right)^{1/2}$; l'équivalent d'une corrélation.

2.2.3. Matrice de variances covariances

Dans le cas d'études résultant de l'observation des mêmes variables, c'est à dire la situation 2, on prend comme élément caractéristique la matrice V_k de variances covariances :

$$V_k = X_k D_k X_k'$$

Lorsque les données sont centrées et réduites par étude, V_k est la matrice de corrélation.

Equivalence 1

Les deux triplets (X_1, M_1, D) et (X_2, M_2, D) sont équivalents si $V_1 = V_2$. Si l'on fait l'analyse en composantes principales des deux triplets, on aura la même représentation des caractères.

Equivalence 2

Les deux triplets (X_1, M, D_1) et (X_2, M, D_2) sont équivalents s'il existe $\alpha > 0$ tel que $V_1 = \alpha V_2$. Si l'on fait l'analyse en composantes principales des deux triplets, on obtiendra des représentations homothétiques des variables.

2.2.4. Situation 3

C'est le cas où les mêmes variables sont observées sur les mêmes individus. On peut à la fois comparer les W_k , les V_k . Selon l'optique de l'étude, on peut être amené à préférer l'une ou l'autre des équivalences correspondantes, selon que l'on désire étudier l'évolution des liaisons entre les individus, ou entre les variables.

Soit (X_1, M, D) et (X_2, M, D) deux études portant sur les mêmes individus et les mêmes variables. S'il existe $\alpha > 0$ tel que $X_1' M_1 X_1 = \alpha X_2' M_2 X_2$, et $\beta > 0$ tel que $X_1 D X_1' = \beta X_2 D X_2'$ alors $\alpha = \beta$, $X_1 = \sqrt{\alpha_2} X_2$ (au signe près).

3. Interstructure – Compromis – Infrastructure

Les objectifs définis en introduction, peuvent être résumés par trois mots : interstructure, compromis, infrastructure. Ce sont là aussi, les principales étapes de la méthode STATIS. L'exposé qui suit, s'inspire des formulations présentées par Escoufier (1978), Glacon (1981) et Lavit (1988).

3.1. Interstructure

L'étude de l'interstructure correspond à l'analyse globale : l'objectif est de comparer k tableaux de données. Pour ce faire, on construit la matrice \mathfrak{C} , $k \times k$ de produits scalaires entre objets caractéristiques. La diagonalisation de \mathfrak{C} fournit alors une représentation globale, analogue à une représentation d'analyse en composantes principales.

3.1.2. La matrice \mathfrak{C}

a) *L'objet caractéristique de l'étude est la matrice W_k*

\mathfrak{C} est la matrice $k \times k$ dont l'élément \mathfrak{C}_{kl} est :

$$tr(W_k D W_l D) = tr(X_k' M_k X_k D X_l' M_l D)$$

Par ailleurs :

$$-tr(W_k D_l D) \geq 0 ;$$

$-tr(W_k D W_l D) = 0 \Leftrightarrow (X_k D X_l = 0)$; i.e les variables des deux tableaux sont non corrélées.

Si :

- $M_k = I_{p_k}$ et $M_l = I_{p_l}$ (métrique euclidienne classique), alors :

$$tr(W_k D W_l D) = \sum \sum (\text{cov}(x_k, x_l))^2 ;$$

- $M_k = D_{1/\sigma_k^2}$ et $M_l = D_{1/\sigma_l^2}$ (matrices diagonales des inverses des variances), alors :

$$tr(W_k D W_l D) = \sum \sum (\text{cor}(x_k, x_l))^2 .$$

b) *L'objet caractéristique de l'étude est la matrice V_k*

\mathfrak{C} est la matrice $k \times k$ dont l'élément \mathfrak{C}_{kl} est :

$$tr(V_k M V_l M) = tr(X_k D_k X_k' M X_l D_l X_l' M)$$

On a de façon analogue :

$$-tr(V_k M V_l M) \geq 0$$

$-tr(V_k M V_l M) = 0 \Leftrightarrow X_k' M X_l = 0$; cela veut dire que les individus engendrent des sous espaces vectoriels de E , M – orthogonaux.

3.1.3. Description des études variables

Soit O_k l'objet caractéristique de la k –ème étude. Le produit scalaire φ entre objets est :

$$\mathfrak{C}_{kl} = \langle O_k O_l \rangle_{\varphi}$$

Chaque O_k est une matrice $p_k p_l$ à laquelle on peut associer le vecteur \bar{O}_k de $R^{p_k p_l}$ obtenu en juxtaposant les colonnes de O_k . On note aussi par φ la métrique sur $M_k \otimes M_l$, où M_k est la métrique sur $E_k = R^{p_k}$ et M_l la métrique sur $E_l = R^{p_l}$.

$$\text{Soit } \mathfrak{X} \text{ le tableau } k \times (p_k p_l) ; \mathfrak{X} = \begin{pmatrix} \bar{O}_1 \\ \vdots \\ \bar{O}_k \end{pmatrix}$$

On a le schéma de dualité suivant :

$$\begin{array}{ccc} & \xleftarrow{\mathfrak{X}} & \\ \begin{array}{c} \uparrow R^k \\ \downarrow R^{k*} \end{array} & & \begin{array}{c} \uparrow R^{p_k p_l*} \\ \downarrow R^{p_k p_l} \end{array} \\ \begin{array}{c} I \\ \mathfrak{C} \end{array} & & \varphi \\ & \xrightarrow{\mathfrak{X}'} & \end{array}$$

Ici $\mathfrak{C} = \mathfrak{X} \varphi \mathfrak{X}'$; \mathfrak{C}^* peut être considéré comme une matrice de variances covariances entre les k variables O_k . On peut encore écrire \mathfrak{C} sous la forme : $\mathfrak{C} = Y Y'$, où Y est la matrice $k \times r$ ($r = \text{rang de } \mathfrak{C}$) dont les colonnes sont des vecteurs propres de \mathfrak{C} , de norme égale à la racine carré de la valeur propre correspondante. Les lignes de \mathfrak{C} fournissent une représentation de k études analogues à celles des variables en analyse en composantes principales ; chaque étude est représentée par un vecteur qui, s'il est bien représenté, a pour norme $\|O_k\|_{\varphi}$; deux vecteurs bien représentés auront un angle proche de celui de O_k et O_l selon φ .

Si l'on veut représenter de façon identique des objets proportionnels, on diagonalise dans ce cas la matrice R des cosinus entre objets :

$$R_{kl} = \frac{\langle O_k O_l \rangle}{\|O_k\|_{\varphi} \|O_l\|_{\varphi}}$$

Ce qui revient à normer toutes les variables O_k . L'interprétation des résultats est alors analogue à celle de l'analyse en composantes principales normées. Les R_{kl} sont suivant les cas de la forme :

$$R_{kl} = \frac{tr(W_k D W_l D)}{\sqrt{tr(W_k D)^2 tr(W_l D)^2}} = R_v(W_k D, W_l D) ; \text{ si } O_k \text{ est de la forme } W_k ;$$

$$R_{kl} = \frac{tr(V_k M V_l M)}{\sqrt{tr(V_k M)^2 tr(V_l M)^2}} ; \text{ si } O_k \text{ est de la forme } V_k$$

3.1.4. Description des études individus

On centre la matrice \mathfrak{C} par rapport à la matrice diagonale des poids D_k . On substitue donc à \mathfrak{C} la matrice centrée :

$$\mathfrak{C}^* = (I - \underline{1} \underline{1}' D_k) \mathfrak{C} (I - D_k \underline{1} \underline{1}')$$

où $\underline{1}$ est le vecteur $\begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$ de R^k

De toute évidence, on ne modifie pas les distances entre études en substituant à \mathfrak{C} la matrice \mathfrak{C}^* . On a le schéma de dualité ci-dessous.

$$\begin{array}{ccc} & \mathcal{Y} & \\ & \longleftarrow & \\ R^{p_k p_l} & & R^{k*} \\ \downarrow \varphi & & \uparrow D_k \\ & \mathcal{Y}' & \\ R^{p_k p_l^*} & \longrightarrow & R^k \end{array}$$

$$\mathfrak{C}^* = \mathcal{Y}' \varphi \mathcal{Y}$$

Où

$$\mathcal{Y} = \mathfrak{X}' (I - D_k \underline{1} \underline{1}') = (\bar{O}_1 \dots \bar{O}_k) (I - D_k \underline{1} \underline{1}')$$

La diagonalisation de $\mathfrak{C}^* D_k$ fournit une représentation analogue à la représentation des individus en analyse en composantes principales.

3.2. Compromis

On cherche un compromis, de même nature que les objets pour représenter les études, qui soit un “bon” résumé de l’ensemble des objets. En d’autres termes on veut résumer les k objets O_1, \dots, O_k par un seul, considéré comme “représentatif” de l’ensemble. On peut choisir comme compromis :

$$O = \sum l_k O_k$$

où :

$$\underline{l} = \begin{pmatrix} l_1 \\ \vdots \\ l_k \end{pmatrix} \text{ est un vecteur propre de } \mathfrak{C}, \text{ normé, associé à la grande valeur propre } \lambda_1.$$

Les deux propositions suivantes sont démontrées (Escoufier, 1978 ; Glacon, 1981) : soit α' un vecteur de composantes $(\alpha_1, \dots, \alpha_k)$ positives telles que $\alpha' \alpha = 1$, on a :

$$1- \sum \langle \sum \alpha_k O_k, O_k \rangle_\varphi^2 \leq \sum \langle O, O_k \rangle_\varphi^2 = \lambda_1^2$$

$$2- \left\| \sum \alpha_k O_k \right\|_\varphi^2 \leq \|O\|_\varphi^2 = \lambda_1$$

Par exemple si les O_k sont de la forme W_k , en posant : $WD = \sum l_k W_k D$ et $W_\alpha D = \sum \alpha_k W_k D$, les deux propositions précédentes s’écrivent respectivement :

$$\sum [tr(W_\alpha D W_k D)]^2 \leq \sum [tr(W D W_k D)]^2 = \lambda_1^2$$

$$tr(W_\alpha D)^2 \leq tr(W D)^2 = \lambda_1$$

Les résultats obtenus justifient, d’une part le nom de compromis donné à WD et, montrent d’autre part, que ce compromis est une combinaison des $W_h D (h=1, k)$ de norme maximum. Par ailleurs, étant une combinaison linéaire positive d’opérateurs semi définis positifs, WD est lui même semi définitif. Sa diagonalisation permet de trouver une matrice Y , tel que $WD = Y' Y D$ qui permet une représentation des n objets résumant au mieux les visions fournies par chacune des k études. Signalons enfin que, dans le cas des W_k , comme tous les éléments de \mathfrak{C} sont positifs, les composantes de \underline{l} sont donc toutes de même signe et on peut les choisir toutes positives. O (ainsi

que \bar{O}) est une combinaison linéaire à coefficients positifs de matrices semi définies positives. Ce compromis est donc semi défini positif et centré pour D .

3.3. Intrastructure

La diagonalisation du compromis nous donne :

- dans le cas des W_k une représentation des individus "intermédiaire" entre les représentations que pourrait fournir chaque étude ;
- dans le cas des V_k , une représentation "moyenne" des variables.

3.3.1. O_k est de la forme W_k

C'est le cas où l'on a les mêmes individus.

Représentation des variables.

Notons par :

Λ : la matrice diagonale des valeurs propres non nulles de WD ; $\Lambda = YDY'$;

Y : la matrice dont les lignes sont les composantes principales, vecteurs propres de WD de norme égale à la racine carrée de la valeur propre correspondante.

Si l'on effectue l'analyse en composantes principales du triplet (X, M, D) , une représentation des variables est donnée par :

$$\hat{Z} = XDY'\Lambda^{-1/2}$$

Les variables et les composantes principales de chaque étude sont représentées respectivement par :

$$\hat{Z}_k = X_kDY'\Lambda^{-1/2} ;$$

$$\hat{Y}_k = Y_kDY'\Lambda^{-1/2}$$

Cette façon de procéder conduit à réaliser la représentation des X_k et des Y_k sur le principe qui permet de représenter les variables supplémentaires en analyse en composantes principales. Quand Y est entier, on a une représentation exacte de toutes les variables et les composantes principales. Dans la pratique, on a la meilleure représentation plane simultanée de toutes les variables en ne conservant que les deux premières lignes de Y .

Représentation des individus

On veut trouver une représentation graphique permettant de comparer les positions respectives des individus tels qu'ils sont décrits par les différents tableaux ou de comparer leurs évolutions. On peut représenter les individus par les colonnes de :

$$\tilde{Y}_k = Y_k D Y' (Y D Y')^{-1} Y = Y_k D Y' \Lambda^{-1} Y$$

\tilde{Y}_k est l'approximation au sens des moindres carrés de Y_k fournie par Y . Si l'on ne garde pour Y que les deux premières lignes, on peut représenter les n objets tels qu'ils sont vus par le compromis. Par ailleurs, en superposant les différentes représentations que fournissent les deux premières lignes des \tilde{Y}_k , on peut comparer visuellement les différentes visions des objets.

3.3.2. O_k est de la forme V_k

On est dans la situation où l'on a les mêmes variables. C'est le cas dual du précédent.

On peut représenter :

- les variables, par la méthode duale de la précédente ;

- les individus, par :

$$\mathcal{Y}_k = \Lambda^{-1/2} Z' M Z X_k$$

où, Z est tel que :

$$V = \sum V_k = Z Z'$$

4. application de statis à l'étude de l'évolution des T.E.S congolais

Les données sont celles utilisées par Mlle L.J. Ondong (1996) pour son mémoire de maîtrise. Il s'agit des tableaux entrées-sorties congolais (T.E.S) correspondant aux années 1989 et 1993. L'objectif de l'étude était de comparer les différents tableaux, afin de déceler si possible, l'existence d'une tendance de la structure industrielle congolaise ; ce qui permettrait de faire des prévisions. Nous allons commenter les principaux résultats d'un programme d'analyse conjointe de tableaux quantitatifs (logiciel ACT ; Lavit, 1989) pour l'analyse des cinq tableaux entrées-sorties congolais. On peut faire remarquer que tous les tableaux ont la même structure : 16 lignes et 16 colonnes. On est dans le cas où l'on a les mêmes individus et les mêmes variables. On peut alors adopter soit la situation 1, soit la situation 2. Nous considérerons la

première. Afin d'équilibrer l'influence de différents tableaux, on utilise les objets normés pour caractériser chaque étude. Les résultats de l'interstructure sont les suivants :

4.1. Matrice $R_v(k,l) = \frac{tr(W_k D, W_l D)}{\sqrt{tr(W_k D)^2 tr(W_l D)^2}}$

C'est la matrice de corrélation vectorielle entre études. Elle est donnée dans le tableau 1 ci-dessous. On constate que les coefficients associés aux cinq tableaux sont très élevés ; cela signifie que la structure des produits dans les différents tableaux n'a pratiquement pas évolué entre 1989 et 1993. on peut néanmoins signaler que la structure du T.E.S de 1992 apparaît un peu différente des autres.

Tableau 1 : Matrice des coefficients R_v

Années	1989	1990	1991	1992	1993
1989	1,000	0,987	0,985	0,938	0,991
1990	0,987	1,000	0,974	0,944	0,989
1991	0,985	0,974	1,000	0,940	0,987
1992	0,938	0,944	0,940	1,000	0,943
1993	0,991	0,989	0,987	0,943	1,000

4.2. Valeurs propres non nulles de R_v

Comme on l'a déjà indiqué, la proximité entre tableaux est représentée par l'inter structure. La diagonalisation de la matrice des coefficients R_v fournit un système d'axes associé aux cinq valeurs propres du tableau 2.

Tableau 2 : Valeurs propres de l'inter structure

Axes	Valeurs propres	% d'inertie	% cumulé
1	4,87	97,43	97,43
2	0,08	1,70	99,13
3	0,03	0,52	99,66
4	0,00	0,20	99,85
5	0,00	0,15	100,00

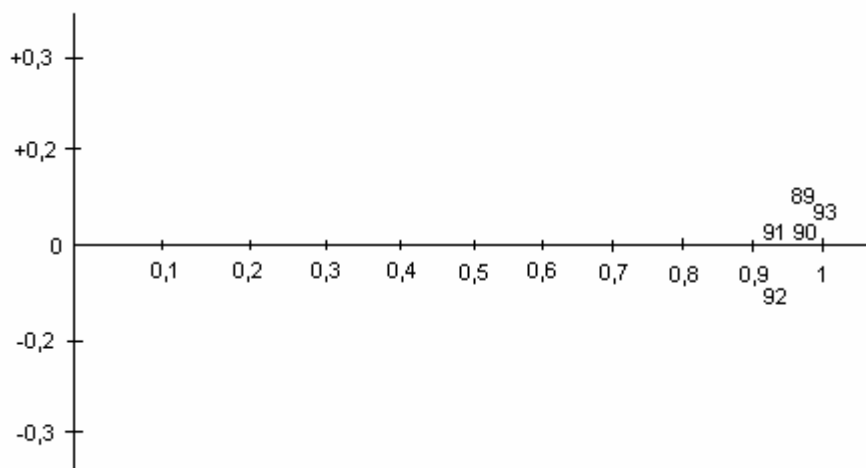
Le premier axe explique 97,43% de l'inertie. En principe, une représentation sur ce premier axe suffit. Cependant, pour marquer la légère différence entre le tableau de 1992 et les autres, nous utilisons le premier plan factoriel qui explique 99,13% de l'inertie.

4.3. Coordonnées des tableaux sur les 5 premiers axes

Tableau 3 : Coordonnées des tableaux

Axes	1989	1990	1991	1992	1993
1	0,993	0,991	0,990	0,965	0,995
2	0,081	0,049	0,059	0,261	0,065
3	0,007	0,109	0,119	0,003	0,005
4	0,083	0,038	0,031	0,007	0,020
5	0,001	0,035	0,029	0,002	0,072

4.4. Représentation graphique



La représentation graphique témoigne, comme on l'a déjà indiqué d'une très forte proximité entre les différents tableaux. Ce qui confirme la stabilité de la structure étudiée.

4.5. Qualité de la représentation

La qualité de la représentation se mesure par la norme des tableaux en projection. Comme on peut le constater (tableau 4), la représentation dans le premier plan factoriel est de très bonne qualité.

Tableau 4 : Qualité de la représentation

Années	Norme exacte	Norme projetée	$\cos^2 x \times 1000$
1989	1	1	992,94
1990	1	0,99	985,38
1991	1	0,99	983,98
1992	1	1	999,94
1993	1	1	994,41

Références

- Ambapour-Kosso, S.**, (1992). Introduction à l'analyse des données. *Cahiers du CASP* n° spécial n°3-4.
- Aris, H., Rio, P.**, (1992). Une méthode tridimensionnelle basée sur le rapprochement des tableaux considérés. *Journal de la Société Statistique de Paris*, Vol 133, n°1/2.
- Braun, J.M.**, (1973). Séries chronologiques multiples : recherche d'indicateurs. *Revue de Statistique Appliquée*, Vol XXI, n°1.
- Bernard, M.C., Lavit, C.**, (1985). L'évolution des sociétés rurales du Languedoc Roussillon (1954-1975) : application de la méthode Statis. 110^{ème} congrès des sociétés savantes. *Géographie*.
- Bonnefous, S., Cloutier, E., Roy, E.**, (1978). Analyse multidimensionnelle des proximités. *Rapport CEA*, R-4910.
- Cailles, F., Pages, J.P.**, (1976). *Introduction à l'analyse des données*, Smash.
- Cazes, P., Bauumeder, A., Bonnefous, S., Pages, J.P.**, (1977). Codage des tableaux logiques. Introduction à la pratique des variables qualitatives. *Cahiers du BURQ*, série recherche n°27. Université P & M Curie.
- Dazy, F., Lebarzic, J.F.**, (1996). *Analyse des données évolutives : méthodes et applications*, Editions Technip.
- Escoufier, Y.**, (1976). Opérateur associé à un tableau de données. *Annales de l'INSEE*, n°22-23.
- Escoufier, Y.**, (1978). Cours d'Analyse des Données. *CRIG*, Montpellier.
- Escoufier, Y.**, (1985). Objectifs et procédures de l'analyse conjointe de plusieurs tableaux de données. *Statistique et Analyse des Données*, Vol 10, n°1.
- Escoufier, Y.**, (1986). A propos du choix des variables en analyse des données. *Métron*, n° XLIV.
- Foucart, T.**, (1978). Sur les suites de tableaux de contingence indexés par le temps. *Statistique et Analyse des Données*, n°2.
- Foucart, T.**, (1984). *Analyse factorielle de tableaux multiples*, Masson.
- Foucart, T.**, (1989). Une nouvelle approche de la méthode statis. *Revue de Statistique Appliquée*, Vol XXXI, n°2.
- Glacon, F.**, (1981). Analyse conjointe de plusieurs matrices de données. *Thèse de 3^{ème} cycle*, Université scientifique et médicale de Grenoble.

- Khodja, H., Lubes-Niel, H., et ali** (1998). Analyse spatio-temporelle de données pluviométriques en Afrique de l'Ouest. Une alternative intéressante : les tests de permutation. *Revue de statistique et Analyse des Données*, Vol XLVI, n°1.
- Lavit, C.**, (1985). Application de la méthode statis. *Statistique et Analyse des Données*, Vol 10, n°1.
- Lavit, C.**, (1988), *Analyse conjointe de tableaux quantitatifs*, Masson.
- Lavit, C.**, (1989). *ACT* : Logiciel d'Analyse Conjointe de Tableaux quantitatifs ; méthode statis, CISIA.
- Le Chevalier, F.**, (1990). L'analyse en composantes conjointes d'une famille de triplets indexés. *Statistique et Analyse des Données*, Vol 15, n°2.
- L'hermier des Plantes, H., Thiebaut B.**, (1977). Etude de la pluviosité au moyen de la méthode statis. *Revue de statistique Appliquée*, Vol XXV, n°2.
- Mailles, J.P.**, (1980). Analyse des tableaux de proximité. Cahier du BURQ, série recherche, n°33, Université P & M Curie.
- Mailles, J.P., Mailles, D., Bonnefous, S.**, (1976). Analyse des données et APL : techniques, bibliothèque, exemples. Rapport CEA. *R-4753*.
- Ondong, L.J.**, (1996). Analyse factorielle des matrices d'échanges : cas des TES congolais de 1989-1993, *Mémoire de maîtrise*, Université M. Ngouabi.
- Pages, J.**, (1996). Eléments de comparaison entre l'analyse factorielle multiple et la méthode statis. *Revue de Statistique Appliquée*, Vol XLIV, n°4.
- Pages, J.P., Escoufier, Y., Cazes, P.**, (1976). Opérateurs et analyse des tableaux à plus de deux dimensions. *Cahiers du BURQ*, série recherche, n°25, Université P & M Curie.
- Saporta, G.**, (1981). Méthodes exploratoires d'analyse de données temporelles. *Cahier du Buro*, série recherche, n°25, Université P & M Curie.
- Stemmelen, E.**, (1979). Tableaux d'échanges : description et prévision. *Cahier du BURQ*, n°28, Université P & M Curie.
- Tennenhaus, M., Prieuret, B.**, (1974). analyse des séries chronologiques multidimensionnelles. *RAIRO*, Vol 2, mai.
- Thorson, H.**, (1981). Une contribution de l'analyse statistique des tableaux d'échanges. *Thèse de 3^{ème} cycle*, Montpellier.