

# CLASSIFICATION FACTORIELLE HIÉRARCHIQUE OPTIMISÉE D'UN TABLEAU DE MESURES

Jean-Jacques DENIMAL<sup>1</sup>

## RÉSUMÉ

L'analyse d'un tableau de mesures est généralement basée sur l'utilisation de l'analyse en composantes principales et de techniques de classification appliquées aux lignes et aux colonnes du tableau. Le dépouillement des résultats générés par ces analyses et surtout leur synthèse représentent souvent pour l'utilisateur un travail long et pénible principalement lorsque les dimensions du tableau sont élevées. La méthodologie proposée dans cet article permet de construire conjointement les axes factoriels et les noeuds des hiérarchies, facilitant ainsi leurs interprétations. Des étapes d'optimisation et d'élagage permettent d'obtenir les meilleures hiérarchies possibles restreintes à leurs noeuds significatifs. Un exemple d'application est également proposé en fin d'article.

*Mots-clés* : Classification hiérarchique, Optimisation, Elagage, Validation croisée, Test de permutation, Plans factoriels.

## ABSTRACT

The analysis of an ordinary table crossing units and variables is generally obtained from the use of different statistical techniques : Principal Component Analysis, Classifications applied to rows and columns of the table. When the dimensions of the table become large, the study of all the results given by these procedures and more particularly their synthesis may represent a long and tedious work. In the methodology proposed in this paper, factorial axes and hierarchy nodes are jointly built , making easier their interpretations. Optimization and pruning techniques allow us to obtain the best possible hierarchies restricted to their significative nodes. An application of the methodology to Russett data is proposed at the end of the paper.

*Keywords* : Hierarchical classification, Optimization, Pruning techniques, Factorial representations, Cross-validation, Permutation test.

---

1. Université des Sciences et Technologies de Lille1, e-mail : jean-jacques.denimal@univ-lille1.fr

## 1. Introduction

L'objet de cet article n'est pas uniquement de fournir une nouvelle méthode de classification mais d'unifier dans une même approche les différentes analyses réalisées classiquement sur un tableau de mesures afin d'extraire d'une manière rapide et aisée l'information significative contenue dans un tableau de mesures. La démarche proposée permettra de réaliser rapidement la synthèse des résultats générés par les représentations factorielles et par les classifications des individus et des variables.

Nous rappelons ci-dessous la suite classique des analyses réalisées sur un tableau de mesures :

- a) Analyse en Composantes Principales du tableau : (Rao, 1964).
- b) Classification des observations obtenue soit par une technique hiérarchique (Benzecri, 1976; Jambu *et al.*, 1978) soit une technique de partitionnement : (Diday, 1979).
- c) Classification des variables : (Varclus, SAS; Lerman, 1991; Qannari *et al.*, 1999).
- d) Aides à l'interprétation établissant les relations existant entre les axes factoriels extraits, les classes-lignes et colonnes (Jambu *et al.*, 1978); Lerman, 1979; Benzecri *et al.*, 1980).

Concernant les étapes b) et c), le couple des partitions sur les lignes et les colonnes du tableau peut être déterminé également par des méthodes de classification croisée (Govaert, 1983), dérivées des nuées dynamiques de Diday. Des généralisations introduisant un modèle probabiliste (modèle de mélange) peuvent être aussi utilisées (Celeux, 1988; Bencheikh, 1992 et 2004).

La méthode proposée dans cet article poursuit trois objectifs principaux :

Le premier objectif est d'harmoniser les résultats de l'analyse factorielle et ceux des classifications.

En effet, plus les dimensions du tableau deviennent élevées, plus la synthèse des différents résultats générés par les analyses des quatre étapes précédentes peut s'avérer laborieuse. On conçoit alors qu'une méthodologie qui ferait en sorte que les axes factoriels et les noeuds des hiérarchies soient construits conjointement faciliterait grandement la tâche de l'utilisateur. Cette idée n'est pas nouvelle. Sous le nom de « correspondances hiérarchiques », Benzecri (1976) et Cazes (1984) proposent un ensemble de tableaux théoriques pour lesquels les axes factoriels extraits sont en complète concordance avec les noeuds d'une hiérarchie donnée à priori.

Le second objectif est de définir un critère dont l'optimisation fournira la meilleure hiérarchie possible.

En effet, l'utilisation des méthodes de classification hiérarchique oblige souvent l'utilisateur à améliorer la partition obtenue après coupure de l'arbre par un algorithme d'échange ou par des itérations à centres mobiles (procédure de consolidation dans le logiciel SPAD). La conséquence est que l'on obtient une partition et que le caractère hiérarchique est perdu. Dans la méthodologie que

l'on propose, une phase d'optimisation permettra par un processus itératif du type k-means d'améliorer la hiérarchie elle même.

Le dernier objectif est d'assurer la signification statistique des noeuds et par conséquent des axes factoriels retenus. La détermination du nombre d'axes factoriels à retenir ainsi que la coupure de chacun des arbres obtenus sont réalisées par des techniques de validation croisée (hiérarchie des variables) ou par des tests de permutations (hiérarchie des individus).

La méthodologie proposée se décompose en 8 étapes :

- Classification hiérarchique initiale des variables (Étape 1).
- Optimisation de cette hiérarchie par une technique du type k-means (Étape 2).
- Élagage de l'arbre obtenu par une technique de validation croisée (Étape 3).
- Représentations factorielles des individus et des variables dans différents plans factoriels en totale association avec les noeuds de la hiérarchie des variables (Étape 4).
- Classification hiérarchique initiale des individus en totale association avec les noeuds de la hiérarchie des variables (Étape 5).
- Optimisation de cette hiérarchie par une technique du type k-means (Étape 6).
- Élagage de l'arbre obtenu par l'utilisation de tests de permutations (Étape 7).
- Interprétation des classes d'individus obtenues par l'utilisation de tests de permutations (Étape 8).

Chacune de ces 8 étapes est détaillée dans la suite de l'article. Un exemple d'application de la méthode aux données de Russett (1964) est ensuite proposée.

## 2. Hiérarchie initiale des variables (Étape 1, Denimal, 2001)

On considère un tableau  $Y = \left( y_i^j \right)_{\substack{j \in J \\ i \in I}}$  croisant un ensemble  $I$  de  $n$  individus et un ensemble  $J$  de  $p$  variables supposées centrées et réduites. On se propose de construire une classification ascendante hiérarchique sur l'ensemble  $J$  des variables.

Chaque individu  $i$  sera muni d'un poids  $p_i$  vérifiant les conditions usuelles :  $p_i > 0$  et  $\sum_{i \in I} p_i = 1$ .

La classification proposée repose sur l'algorithme général de construction d'une classification ascendante hiérarchique, le critère d'agrégation mesurant l'écart entre deux classes étant égal à la seconde valeur propre issue de l'analyse en composantes principales (ACP) du tableau défini par les deux variables représentatives de ces deux classes de variables.

Une série de vecteurs de coefficients de  $R^p$ ,  $a^0, a^1, a^2, \dots, a^{p-1}$ , sera également définie le long des  $(p - 1)$  étapes de l'algorithme.

Ce dernier vecteur  $a^{p-1}$ , noté simplement  $a$  dans la suite, permettra de recalculer les autres vecteurs  $a^k$  avec  $0 \leq k \leq p - 2$ .

Ce vecteur  $a$  jouera un rôle important car toute classe  $q$  de la hiérarchie sera représentée par le vecteur  $z^q = \sum_{j \in q} \frac{a_j}{\sqrt{\sum_{j \in q} (a_j)^2}} . y^j$ . Une phase d'optimisation

de cette hiérarchie initiale (Étape 2) sera, par la suite, engagée recalculant de manière itérative ce vecteur  $a$  de façon à ce que globalement l'ensemble des classes de la hiérarchie ait la meilleure variable représentative possible.

D'autre part, comme chaque classe  $q$  admet une variable représentative définie par la formule donnée ci-dessus, on en déduit que, pour chaque noeud  $(q1, q2)$ , les variables représentatives des trois classes  $q1$ ,  $q2$ ,  $q1 \cup q2$ , sont toujours coplanaires. Il sera alors possible d'obtenir des représentations planes visualisant individus ou variables (Étape 4). Une interprétation graphique du noeud pourra donc être menée facilement.

De manière plus précise, la procédure de construction de la hiérarchie initiale est la suivante :

## 2.1 Algorithme de construction

**Étape 0** : On pose  $a^0 = (a_j^0)_{j \in J}$  telle que  $a_j^0 = 1 \forall j \in J$ .

**Étape 1** : Pour chaque couple de variables  $(y^{j1}, y^{j2})$ , on effectue l'ACP du tableau à 2 colonnes formé par ces deux variables. Les deux variables étant standardisées, les ACP de cette étape sont donc normées. Cette analyse génère deux valeurs propres  $\lambda_1(j1, j2) \geq \lambda_2(j1, j2)$  telles que :

$$\lambda_1(j1, j2) = \sum_{i \in I} p_i \left[ \alpha_1 y_i^{j1} + \beta_1 y_i^{j2} \right]^2 = \text{var}(\alpha_1 y^{j1} + \beta_1 y^{j2}) \text{ avec } \alpha_1^2 + \beta_1^2 = 1.$$

$$\lambda_2(j1, j2) = \text{var}(y^{j1}) + \text{var}(y^{j2}) - \lambda_1(j1, j2) = \text{var}(\beta_1 y^{j1} - \alpha_1 y^{j2})$$

On détermine ensuite le couple  $(j1, j2)$  pour lequel  $\lambda_2(j1, j2)$  est minimum.

On en déduit alors la série de coefficients  $a^1$ .

$$\begin{cases} a^1(j1) = \alpha_1 \\ a^1(j2) = \beta_1 \\ a^1(j) = a^0(j), j \neq j1, j \neq j2 \end{cases}$$

L'indice du premier noeud  $n_1 = (y^{j1}, y^{j2})$  vaut :  $\nu(n_1) = \lambda_2(j1, j2)$

La variable représentative de ce noeud  $n_1$  est :  $z^{\{j1, j2\}} = a^1(j1).y^{j1} + a^1(j2).y^{j2}$ .

**Étape k** : ( $k \in [2, p - 1]$ ). Chaque classe  $q$  obtenue après  $(k - 1)$  étapes est représentée par la variable :  $z^q = \sum_{j \in q} a^{k-1}(j).y^j$ . Pour chaque couple de

variables  $(z^{q1}, z^{q2})$ , on réalise l'ACP non normée du tableau associé. Les deux valeurs propres extraites sont notées :  $\lambda_1(q1, q2) \geq \lambda_2(q1, q2)$ .

$$\lambda_1(q_1, q_2) = \sum_{i \in I} p_i \cdot [\alpha_k \cdot z_i^{q_1} + \beta_k \cdot z_i^{q_2}]^2 = \text{var}(\alpha_k \cdot z^{q_1} + \beta_k \cdot z^{q_2}) \text{ avec } \alpha_k^2 + \beta_k^2 = 1$$

$$\lambda_2(q_1, q_2) = \text{var}(z^{q_1}) + \text{var}(z^{q_2}) - \lambda_1(q_1, q_2) = \text{var}(\beta_k \cdot z^{q_1} - \alpha_k \cdot z^{q_2}).$$

On détermine alors le couple  $(z^{q_1}, z^{q_2})$  pour lequel  $\lambda_2(q_1, q_2)$  est minimum.

On déduit alors :

– une nouvelle série de coefficients  $a^k = (a_j^k)_{j \in J}$  telle que

$$\begin{cases} a_j^k = \alpha_k \cdot a_j^{k-1}, & j \in q_1 \\ a_j^k = \beta_k \cdot a_j^{k-1}, & j \in q_2 \\ a_j^k = a_j^{k-1}, & \text{sinon} \end{cases}$$

– l'indice du nouveau noeud formé :  $\nu(n_k) = \lambda_2(q_1, q_2)$ .

– La variable représentative la classe  $q_1 \cup q_2$  :  $z^{q_1 \cup q_2} = \sum_{j \in q_1 \cup q_2} a_j^k y^j = \alpha_k \cdot z^{q_1} + \beta_k \cdot z^{q_2}$ .

Par définition, la série  $a^{p-1}$  obtenue lors de la dernière étape de l'algorithme sera simplement notée :  $a = (a_j)_{j \in J}$ . Autrement dit,  $\forall j \in J, a_j = a_j^{p-1}$ .

La variable représentative de la classe complète  $J$ ,  $z^J = \sum_{j \in J} a_j \cdot y^j$ , a donc la même interprétation que la première composante principale issue de l'ACP du tableau initial  $Y$ .

**Conséquences :** Chaque noeud de la hiérarchie regroupant deux classes  $q_1$  et  $q_2$  de variables est défini à partir de l'ACP non normée du tableau formé par les deux variables représentatives  $z^{q_1}$  et  $z^{q_2}$ . Le premier axe factoriel est dirigé par la variable représentative  $z^{q_1 \cup q_2}$  de la réunion  $q_1 \cup q_2$ . Nous verrons plus loin que les trois variables  $z^{q_1}, z^{q_2}, z^{q_1 \cup q_2}$  peuvent toujours être choisies de façon à ce qu'elles soient positivement corrélées.

Ce premier axe s'interprète donc comme un axe de compromis entre les deux classes  $q_1$  et  $q_2$ . Le deuxième axe sera dirigé par la variable  $\bar{z}^{(q_1, q_2)} = \beta_k \cdot z^{q_1} - \alpha_k \cdot z^{q_2}$  et s'interprète comme un axe visualisant les différences entre les deux classes  $q_1$  et  $q_2$ .

L'ensemble des  $n$  individus peut être également représenté dans ce plan et se répartit suivant ces deux axes.

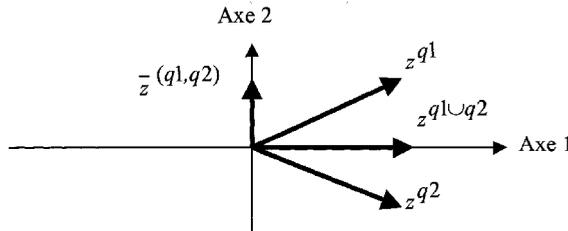


FIG 1. — Plan factoriel associé au noeud  $(q_1, q_2)$ .

## 2.2. Propriétés liées au vecteur $\mathbf{a}$

PROPRIÉTÉ 1. —  $\forall k \in [1, p-1]$ ,  $q$  étant une classe formée à l'étape  $k$ ,

$$a) \sum_{j \in q} (a_j^k)^2 = 1$$

$$\text{En particulier, } \sum_{j \in J} [a_j]^2 = 1$$

$$b) \forall j \in q, a_j^k = \varepsilon_q \cdot \frac{a_j}{\sqrt{\sum_{j \in q} [a_j]^2}}$$

où  $\varepsilon_q$  vaut  $+1$  ou  $-1$  et est indépendant du choix de la variable  $j$  dans la classe  $q$ .

*Démonstration*

a) Le résultat est obtenu par récurrence :

$$- \text{ Pour le premier noeud } n_1 = \{j_1, j_2\}, \text{ on a : } \sum_{j \in n_1} [a_j^1]^2 = (\alpha_1)^2 + (\beta_1)^2 = 1$$

Les autres classes  $q = \{j\}$  sont des singletons. Donc,  $\sum_{j \in q} (a_j^1)^2 = 1$ , (puisque

$$a_j^1 = a_j^0 = 1)$$

- Si la propriété est vraie jusqu'à l'étape  $k-1$ , et si  $q_1$  et  $q_2$  sont les deux classes qui fusionnent au niveau  $k$ , on a :

$$\sum_{j \in q_1 \cup q_2} [a_j^k]^2 = (\alpha_k)^2 \cdot \sum_{j \in q_1} [a_j^{k-1}]^2 + (\beta_k)^2 \cdot \sum_{j \in q_2} [a_j^{k-1}]^2 = (\alpha_k)^2 + (\beta_k)^2 = 1$$

$$\text{Pour toute classe } q \text{ différente de } q_1 \cup q_2, \sum_{j \in q} (a_j^k)^2 = \sum_{j \in q} (a_j^{k-1})^2 = 1.$$

Le cas particulier  $\sum_{j \in J} (a_j)^2 = 1$  est obtenu quand  $k = p-1$ .

b)  $\forall k \in [1, p-1]$ , et  $q$  une classe obtenue au niveau  $k$ .

De la construction de la hiérarchie, on déduit l'existence d'un réel  $\Gamma_q$  dépendant uniquement de  $q$  tel que :

$$\forall j \in q, a_j = \Gamma_q \cdot a_j^k \quad (\text{où } a_j = a_j^{p-1}).$$

En utilisant la propriété précédente a), on obtient :

$$\sum_{j \in q} (a_j)^2 = (\Gamma_q)^2 \cdot \sum_{j \in q} (a_j^k)^2 = (\Gamma_q)^2.$$

$$\text{D'où, } \Gamma_q = \varepsilon_q \cdot \sqrt{\sum_{j \in q} (a_j)^2} \text{ avec } \varepsilon_q = \mp 1.$$

$$\text{et finalement, } \forall j \in q, a_j^k = \frac{a_j}{\Gamma_q} = \varepsilon_q \cdot \frac{a_j}{\sqrt{\sum_{j \in q} (a_j)^2}}.$$

PROPRIÉTÉ 2. —  $\forall k \in [1, p - 1]$ , chaque classe  $q$  formée à l'étape  $k$  peut être représentée par la variable  $z^q = \sum_{j \in q} \frac{a_j}{\sqrt{\sum_{j \in q} (a_j)^2}} \cdot y^j$ .

Démonstration. — Par construction, la variable représentative de la classe  $q$  est :  $\sum_{j \in q} a_j^k \cdot y^j = \varepsilon_q \cdot \sum_{j \in q} \frac{a_j}{\sqrt{\sum_{j \in q} (a_j)^2}} \cdot y^j$ .

(La dernière égalité s'obtient par l'application de la propriété 1b) ci-dessus).

La variable représentative de la classe  $q = (q1, q2)$  est par définition la première composante principale issue d'une ACP particulière. Comme le signe qui lui est attribué est arbitraire, la variable  $z^q = \sum_{j \in q} \frac{a_j}{\sqrt{\sum_{j \in q} (a_j)^2}} \cdot y^j$  peut être

choisie comme variable représentative de la classe  $q$ .

**Conséquences et définition :** Ainsi, à l'issue de l'algorithme de construction de la hiérarchie, le vecteur  $a$  est obtenu et chaque classe  $q$  peut être représentée par la formule  $z^q = \sum_{j \in q} \frac{a_j}{\sqrt{\sum_{j \in q} (a_j)^2}} \cdot y^j$  (Propriété 2). Par définition, dans la

suite de l'article, chaque classe  $q$  sera représentée par cette variable  $z^q$ . Dans ces conditions, si  $q$  est un singleton  $\{j\}$ , on a :  $z^q = \frac{a_j}{|a_j|} \cdot y^j$ . Autrement dit,  $y^j$  sera muni du signe du coefficient  $a_j$ .

Ce système de variables représentatives vérifient alors la propriété suivante

PROPRIÉTÉ 3. —  $\forall k \in [1, p - 1]$ ,  $q1$  et  $q2$  étant les deux classes fusionnant à l'étape  $k$ ,

$$a) \text{ On a : } \begin{cases} \alpha_k^2 = \frac{a_{q1}}{a_{q1} + a_{q2}} \\ \beta_k^2 = \frac{a_{q2}}{a_{q1} + a_{q2}} \end{cases} \text{ où } a_q = \sum_{j \in q} [a_j]^2.$$

b) Les variables représentatives

$$z^{q1} = \sum_{j \in q1} \frac{a_j}{\sqrt{\sum_{j \in q1} (a_j)^2}} \cdot y^j \text{ et } z^{q2} = \sum_{j \in q2} \frac{a_j}{\sqrt{\sum_{j \in q2} (a_j)^2}} \cdot y^j \text{ sont toujours positivement corrélées}$$

Démonstration

a) Les variables représentatives des classes  $q1$ ,  $q2$ ,  $q1 \cup q2$  sont choisies en utilisant la propriété 2.

En conséquence,

$$z^{q1 \cup q2} = \sum_{j \in q1 \cup q2} \frac{a_j}{\sqrt{\sum_{j \in q1 \cup q2} (a_j)^2}} \cdot y^j = \sqrt{\frac{\sum_{j \in q1} (a_j)^2}{\sum_{j \in q1 \cup q2} (a_j)^2}} \cdot z^{q1} + \sqrt{\frac{\sum_{j \in q2} (a_j)^2}{\sum_{j \in q1 \cup q2} (a_j)^2}} \cdot z^{q2}.$$

Donc,  $z^{q1 \cup q2} = \alpha_k \cdot z^{q1} + \beta_k \cdot z^{q2}$  avec  $\alpha_k^2 = \frac{a_{q1}}{a_{q1} + a_{q2}}$  et  $\beta_k^2 = \frac{a_{q2}}{a_{q1} + a_{q2}}$ .

b) En utilisant ce dernier résultat, on peut écrire :  $z^{q1 \cup q2} = \alpha_k \cdot z^{q1} + \beta_k \cdot z^{q2}$  avec  $\alpha_k \geq 0$  et  $\beta_k \geq 0$ . On considère cette fois l'ACP réalisée au niveau  $k$  de la hiérarchie comme celle du tableau composé des deux variables  $z^{q1}$  et  $z^{q2}$ . Dans cette ACP, les deux vecteurs axiaux de  $R^2$  associés aux valeurs propres  $\lambda_1 \geq \lambda_2$  sont :  $u_1(\alpha_k, \beta_k)$  et  $u_2(\beta_k, -\alpha_k)$ . En conséquence, les coordonnées de  $z^{q1}$  et  $z^{q2}$  sur les deux axes factoriels sont respectivement :  $(\sqrt{\lambda_1} \alpha_k, \sqrt{\lambda_2} \beta_k)$  et  $(\sqrt{\lambda_1} \beta_k, -\sqrt{\lambda_2} \alpha_k)$ . La covariance  $cov(z^{q1}, z^{q2})$  étant le produit scalaire de ces deux vecteurs pour la métrique des poids, on déduit :

$$cov(z^{q1}, z^{q2}) = [\sqrt{\lambda_1} \alpha_k \cdot \sqrt{\lambda_1} \beta_k] + [-\sqrt{\lambda_2} \beta_k \cdot \sqrt{\lambda_2} \alpha_k] = (\lambda_1 - \lambda_2) \cdot \alpha_k \cdot \beta_k.$$

La covariance est donc positive ou nulle.

### 2.3. Interprétation du critère d'agrégation utilisé

PROPRIÉTÉ 4. — Si  $n_k$  est le noeud obtenu au niveau  $k$  regroupant les classes  $q1$  et  $q2$ , l'indice de ce noeud peut encore se mettre sous la forme suivante :

$$\nu(n_k) = \frac{a_{q1} \cdot a_{q2}}{a_{q1} + a_{q2}} \cdot \sum_{i \in I} p_i \left[ \sum_{j \in q1} \frac{a_j}{a_{q1}} \cdot y^j - \sum_{j \in q2} \frac{a_j}{a_{q2}} \cdot y^j \right]^2$$

avec  $a_q = \sum_{j \in q} [a_j]^2$ ,  $q \in \{q1, q2\}$ .

*Démonstration.* — Ce résultat se déduit facilement de l'égalité  $\nu(n_k) = var(\beta_k \cdot z^{q1} - \alpha_k \cdot z^{q2})$ , où  $\beta_k$  et  $\alpha_k$  sont remplacés par leurs valeurs données à la propriété 3.

En effet, les deux réels  $\beta_k$  et  $\alpha_k$  sont toujours positifs, compte tenu du choix de la variable représentative de chaque classe  $q$  (Propriété 2). On peut donc écrire (Propriété 3) :

$$\beta_k \cdot z^{q1} = \sqrt{\frac{a_{q2}}{a_{q1} + a_{q2}}} \cdot \left[ \sum_{j \in q1} \frac{a_j}{\sqrt{a_{q1}}} \cdot y^j \right] \text{ et } \alpha_k \cdot z^{q2} = \sqrt{\frac{a_{q1}}{a_{q1} + a_{q2}}} \cdot \left[ \sum_{j \in q2} \frac{a_j}{\sqrt{a_{q2}}} \cdot y^j \right]$$

$$\beta_k \cdot z^{q1} - \alpha_k \cdot z^{q2} = \sqrt{\frac{a_{q1} \cdot a_{q2}}{a_{q1} + a_{q2}}} \cdot \left[ \sum_{j \in q1} \frac{a_j}{a_{q1}} \cdot y^j - \sum_{j \in q2} \frac{a_j}{a_{q2}} \cdot y^j \right]$$

**Conséquence :**

$R^{|I|}$  étant muni de la métrique des poids  $p_i$  des individus  $i$  ( $i \in I$ ), on considère le nuage de  $R^{|I|}$  composé des points  $t^j = \frac{y^j}{a_j}$ ,  $j \in J$ , munis des poids  $m_j = a_j^2$ .

Dans ce cadre, l'indice d'agrégation du noeud  $n_k$  s'écrit alors :

$$\nu(n_k) = \frac{m_{q1} \cdot m_{q2}}{m_{q1} + m_{q2}} \cdot \|t^{q1} - t^{q2}\|^2 \text{ où } m_q = \sum_{j \in q} m_j \text{ et } t^q = \sum_{j \in q} \frac{m_j}{m_q} \cdot t^j.$$

Par conséquent, si les coefficients  $a_j$  étaient connus au début de l'algorithme de construction de la hiérarchie, celle-ci serait obtenue par l'application de la **méthode classique de Ward** appliquée aux points  $t^j$  munis des poids  $m_j$ . L'ensemble des propriétés vérifiées par le critère de Ward sera également vérifié par la présente classification.

PROPRIÉTÉ 5. —

a) La série des indices d'agrégation  $\{\nu(n_k) / 1 \leq k \leq p-1\}$  est croissante. Autrement dit, la hiérarchie ne présentera pas d'inversions.

b)  $\nu(n_{p-1}) \leq \text{var}(\sum_{j \in J} a_j y^j)$  où le second membre représente la variance de la variable représentative de la classe complète.

$$c) p = \sum_{j \in J} \text{var}(y^j) = \text{var}(\sum_{j \in J} a_j \cdot y^j) + \sum_{k=1}^{p-1} \nu(n_k)$$

Démonstration

a) La propriété a) est une propriété classique du critère de Ward.

b) On sait que si  $n$  est un noeud regroupant les deux classes  $q$  et  $q'$ , on a :

$$\nu(n) = \lambda_2(q1, q2) = \text{var}(z^{q1}) + \text{var}(z^{q2}) - \lambda_1(q1, q2)$$

En conséquence, si  $q1$  et  $q2$  sont les deux classes fusionnant au dernier niveau  $k = p-1$ , on a :

$$\nu(n_{p-1}) = \text{var}(z^{q1}) + \text{var}(z^{q2}) - \text{var}(\sum_{j \in J} a_j \cdot y^j)$$

Comme par construction on a :  $\text{var}(z^{q1}) \leq \text{var}(\sum_{j \in J} a_j \cdot y^j)$  et  $\text{var}(z^{q2}) \leq \text{var}(\sum_{j \in J} a_j \cdot y^j)$ ,

on en déduit :  $\nu(n_{p-1}) \leq \text{var}(\sum_{j \in J} a_j \cdot y^j)$ .

c) On considère le nuage de  $R^{|I|}$  composé des points  $t^j = \frac{y^j}{a_j}$ ,  $j \in J$ , munis des poids  $m_j = a_j^2$  de la conséquence exposée ci-dessus. Dans le cadre du critère de Ward, on sait que la somme des indices des noeuds redonne l'inertie totale, à savoir dans ce cadre  $\sum_{j \in J} m_j \|t^j - \bar{t}\|^2$  avec  $\bar{t} = \sum_{j \in J} m_j \cdot \frac{y^j}{a_j} = \sum_{j \in J} a_j \cdot y^j$ .

L'application du théorème de Huyghens nous permet d'écrire :  $\sum_{j \in J} m_j \|t^j\|^2 =$

$\sum_{j \in J} m_j \|t^j - \bar{t}\|^2 + \|\bar{t}\|^2$ . En conséquence, la somme des indices des noeuds

vaut :  $\sum_{j \in J} m_j \|t^j\|^2 - \|\bar{t}\|^2 = \sum_{j \in J} \text{var}(y^j) - \text{var}(\sum_{j \in J} a_j \cdot y^j)$ .

*Remarque 1.* — La différence entre la somme des variances  $\sum_{j \in J} \text{var}(y^j)$  et la

variance  $\text{var}\left(\sum_{j \in J} a_j \cdot y^j\right)$  de la variable représentative de la classe complète représente la perte de variance lorsque les  $p$  variables initiales sont remplacées par cette variable représentative. La propriété précédente montre que cette perte de variance se décompose sur les noeuds de la hiérarchie.

*Remarque 2.* — Dans le cadre du critère de Ward, des techniques particulières permettant l'accélération de la construction de la hiérarchie ont été proposées : La technique des graphes réductibles (Bruynhooge 1978) ou la technique des voisins réciproques (Juan, 1982). En conséquence, ces deux approches peuvent être également appliquées dans ce cadre. Ce qui permet de classer rapidement un grand nombre de variables. La méthode des graphes réductibles a été appliquée dans le logiciel mettant en oeuvre la technique exposée dans cet article.

#### 2.4. Base orthonormée de $R^p$ associée à la hiérarchie

DÉFINITION et PROPRIÉTÉ 6. — Soit  $a = (a_j)_{j \in J}$  un vecteur de  $R^p$  et  $H$  une hiérarchie sur  $J$ .

On considère le système suivant de  $p$  vecteurs  $\{e_J^1, e_J^2, \dots, e_J^p\}$  de  $R^p$  :

–  $\forall k \in [1, p-1]$ , si  $n_k = (q1, q2)$  représente le noeud obtenu à l'étape  $k$  et fusionnant les deux classes  $q1$  and  $q2$ , on pose :

$$e_J^k(j) = \begin{cases} -\beta_k \cdot a_j^{k-1} = -\sqrt{\frac{a_{q2}}{a_{q1} \cdot (a_{q1} + a_{q2})}} \cdot a_j, & j \in q1 \\ \alpha_k \cdot a_j^{k-1} = \sqrt{\frac{a_{q1}}{a_{q2} \cdot (a_{q1} + a_{q2})}} \cdot a_j, & j \in q2 \\ 0, & \text{ailleurs} \end{cases}$$

–  $\forall j \in J$ ,  $e_J^p(j) = a_j$ .

Dans ces formules, la notation  $a_q$  pour une classe  $q$  signifie  $a_q = \sum_{j \in q} a_j^2$ , ( $q = q1$  ou  $q2$ ).

On montre facilement que  $\mathcal{B}_J = \{e_J^1, e_J^2, \dots, e_J^p\}$  est une base orthonormée de  $R^p$ , celui-ci étant muni de la métrique euclidienne classique.

On considère, à présent, la hiérarchie initiale construite précédemment sur  $J$  et  $a$  le vecteur de coefficients associé. On déduit alors facilement la propriété 7 suivante ainsi que sa conséquence.

PROPRIÉTÉ 7. — Soit  $F_k(i)$  la coordonnée du vecteur  $y_i$  représentant l'individu  $i$  du tableau initial  $Y$  sur le vecteur de base  $e_k$ .

a)  $\forall k \in [1, p-1]$ ,  $F_k(i) = -\beta_k \cdot z_i^{q1} + \alpha_k \cdot z_i^{q2} = \bar{z}_i^{n_k}$

où  $n_k = (q1, q2)$  représente le noeud obtenu au niveau  $k$  et où  $\bar{z}^{(q1, q2)}$  représente la deuxième composante principale issu de l'ACP définissant le noeud  $q = (q1, q2)$ .

$$b) F_p(i) = \sum_{j \in J} a_j \cdot y_i^j$$

### Conséquences

En utilisant la base orthonormée  $\mathcal{B}_J = \{e_J^1, e_J^2, \dots, e_J^p\}$ , on peut écrire :

$$\forall i \in I, y_i = \sum_{k=1}^p F_k(i) \cdot e_J^k$$

$$\text{Ce qui s'écrit encore : } \forall j \in J, y^j = \sum_{k=1}^p e_J^k \cdot F_k = a_j \cdot F_p + \sum_{k=1}^{p-1} e_J^k \cdot \bar{z}^{n_k}$$

Ainsi, chaque variable  $y^j$  se décompose sous la forme d'une combinaison linéaire de la variable représentative de la classe complète  $J$  et des secondes composantes principales des ACP définissant chaque noeud.

De plus,  $\forall k \in [1, p-1]$ ,  $\text{var}(\bar{z}^{n_k}) = \lambda_2(n_k) = \nu(n_k)$

où  $n_k$  est le noeud formé au niveau  $k$  et où  $\lambda_2(n_k)$  et  $\nu(n_k)$  sont respectivement la seconde valeur propre de l'ACP réalisée à ce niveau et l'indice du noeud correspondant.

D'après la propriété 5, on a :  $\text{var}(F_p) \geq \text{var}(\bar{z}^{n_{p-1}}) \geq \text{var}(\bar{z}^{n_{p-2}}) \geq \dots \geq \text{var}(\bar{z}^{n_1})$

Ces inégalités peuvent s'interpréter comme traduisant l'importance décroissante des variables intervenant dans la combinaison linéaire définissant  $y^j$ . Cette suite décroissante de variances s'interprète aussi comme la suite des inerties projetées du nuage des individus sur les vecteurs  $e_J^p, e_J^{p-1}, \dots, e_J^1$ .

## 3. Optimisation et élagage de la hiérarchie des variables

La phase d'optimisation de la hiérarchie initiale a pour but de recalculer le contenu des classes et le vecteur de coefficients  $a = (a_j)_{j \in J}$  de façon à ce que

$$\text{la variable représentative } z^q = \sum_{j \in q} \frac{a_j}{\sqrt{\sum_{j \in q} (a_j)^2}} \cdot y^j \text{ de chaque classe } q \text{ soit de}$$

variance maximum.

### 3.1. Optimisation de la hiérarchie des variables (Étape 2)

Une hiérarchie peut être vue sous la forme d'une succession de partitions  $\mathcal{P}_0, \mathcal{P}_1, \dots, \mathcal{P}_{p-1}$ , où  $\mathcal{P}_0$  et  $\mathcal{P}_{p-1}$  sont respectivement la partition des singletons et la partition composée de la seule classe  $J$ . Chaque partition  $\mathcal{P}_k$  correspond à un niveau de la hiérarchie mesuré par l'indice  $\nu(k)$  du noeud  $n_k$  formé à ce niveau et obtenu lors de l'étape de la hiérarchie initiale.

Le critère que l'on se propose de maximiser est :

$$Q = \sum_{k=1}^{p-1} \nu(k) \cdot \sum_{q \in \mathcal{P}_k} \text{var}(z^q) \text{ avec } a_q = \sum_{j \in q} (a_j)^2$$

$$\text{et } z^q = \sum_{j \in q} \frac{a_j}{\sqrt{\sum_{j \in q} (a_j)^2}} \cdot y^j = \sum_{j \in q} \frac{a_j}{\sqrt{a_q}} \cdot y^j$$

$\nu(k)$  s'interprète comme l'importance (ou « poids ») de la séparation réalisée au niveau  $k$  considéré.

*Remarques :*

1) Les poids  $\nu(k)$  des différents niveaux, égaux aux indices des noeuds de la hiérarchie initiale, resteront constants lors de la phase d'optimisation. Leur rôle est de favoriser les niveaux les plus hauts qui marquent les principales séparations.

2) Pour la classe complète  $q = J$  obtenue à l'étape  $p - 1$ , déterminer  $a = (a_j)_{j \in J}$  tel que  $\text{var}(z^q)$  soit maximum sous la contrainte  $\sum_{j \in J} a_j^2 = 1$  revient à rechercher le vecteur propre de la matrice  ${}^t Y \cdot Y$  associé à la plus grande valeur propre (ACP de  $Y$ ).

3) Pour chaque classe  $q$  de variables, on note  $Y_q$  la matrice de dimensions  $(n, p)$  égale à  $Y$  à l'exception des colonnes associées aux variables n'appartenant pas à  $q$  qui sont remplacées par 0.

On peut alors écrire  $Q$  sous la forme :

$$Q = \sum_{k=1}^{p-1} \nu(k) \cdot \sum_{q \in \mathcal{P}_k} \left( \frac{{}^t a \cdot {}^t Y_q \cdot Y_q \cdot a}{{}^t a \cdot I_q \cdot a} \right)$$

où  $I_q$  est une matrice diagonale telle que  $I_q(j, j) = \begin{cases} 1, & j \in q \\ 0, & \text{ailleurs} \end{cases}$

PROPRIÉTÉ 8. —  $\forall k \in [1, p - 1]$ , on a :

$$\sum_{q \in \mathcal{P}_k} \text{var}(z^q) = p - \sum_{q \in \mathcal{P}_k} \sum_{j \in q} a_j^2 \left\| \frac{y^j}{a_j} - \sum_{j \in q} \frac{a_j}{a_q} \cdot y^j \right\|^2$$

*Démonstration.* —  $\forall q \in \mathcal{P}_k$ , on considère le sous nuage  $\mathcal{N}_q$  composé des points  $\frac{y^j}{a_j}$ ,  $j \in q$ , munis des poids  $a_j^2$  :

$$\mathcal{N}_q = \left\{ \left( \frac{y^j}{a_j}, a_j^2 \right) / j \in q \right\}$$

En appliquant le théorème de Huyghens, on peut écrire, avec  $a_q = \sum_{j \in q} (a_j)^2$  :

$$\sum_{j \in q} a_j^2 \left\| \frac{y^j}{a_j} \right\|^2 = \sum_{j \in q} a_j^2 \left\| \frac{y^j}{a_j} - \sum_{j \in q} \frac{a_j^2}{a_q} \cdot \frac{y^j}{a_j} \right\|^2 + a_q \cdot \left\| \sum_{j \in q} \frac{a_j^2}{a_q} \cdot \frac{y^j}{a_j} \right\|^2$$

Comme  $\|y^j\|^2 = \text{var}(y^j) = 1$ , on déduit que :

$$\text{Card}(q) = \sum_{j \in q} a_j^2 \left\| \frac{y^j}{a_j} - \sum_{j \in q} \frac{a_j}{a_q} \cdot y^j \right\|^2 + \left\| \sum_{j \in q} \frac{a_j}{\sqrt{a_q}} \cdot y^j \right\|^2$$

En sommant sur  $q \in \mathcal{P}_k$ , on a finalement :

$$p = \sum_{q \in \mathcal{P}_k} \sum_{j \in q} a_j^2 \left\| \frac{y^j}{a_j} - \sum_{j \in q} \frac{a_j}{a_q} \cdot y^j \right\|^2 + \sum_{q \in \mathcal{P}_k} \text{var}(z^q)$$

### Conséquences :

1) La quantité  $Q$  à maximiser devient donc :

$$Q = \left[ \sum_{k=1}^{p-1} \nu(k) \right] \cdot p - \sum_{k=1}^{p-1} \nu(k) \sum_{q \in \mathcal{P}_k} \sum_{j \in q} a_j^2 \left\| \frac{y^j}{a_j} - \sum_{j \in q} \frac{a_j}{a_q} \cdot y^j \right\|^2$$

$$Q = \left[ \sum_{k=1}^{p-1} \nu(k) \right] \cdot p - \sum_{k=1}^{p-1} \nu(k) \sum_{q \in \mathcal{P}_k} \sum_{j \in q} a_j^2 \left\| \frac{y^j}{a_j} - \frac{1}{\sqrt{a_q}} \cdot z^q \right\|^2$$

Maximiser  $Q$  équivaut à minimiser  $Q' = \sum_{k=1}^{p-1} \nu(k) \sum_{q \in \mathcal{P}_k} \sum_{j \in q} a_j^2 \left\| \frac{y^j}{a_j} - \frac{1}{\sqrt{a_q}} \cdot z^q \right\|^2$

2) La quantité  $Q$  est donc majorée par  $\left[ \sum_{k=1}^{p-1} \nu(k) \right] \cdot p$ , et donc par  $p^2$ .

3)  $\forall j \in J$ , on pose  $q_j(k)$  la classe de la partition  $\mathcal{P}_k$  contenant  $j$ , la quantité  $Q'$  s'écrit encore :

$$Q' = \sum_{j \in J} a_j^2 \cdot \left[ \sum_{k=1}^{p-1} \nu(k) \cdot \left\| \frac{y^j}{a_j} - \frac{1}{\sqrt{a_{q_j(k)}}} \cdot z^{q_j(k)} \right\|^2 \right]$$

### Processus d'optimisation

Ce processus recherche une hiérarchie sur  $J$  et un vecteur  $a$  associé maximisant la quantité  $Q$ .

Ce processus se traduit par la succession de trois étapes appelées ci-dessous agrégation ou affectation :

#### – Étape d'agrégation 1

La hiérarchie sur  $J$  étant fixée, on recherche un vecteur  $a$  maximisant  $Q$  sachant que  $\sum_{j \in J} a_j^2 = 1$ .

Il s'agit donc de rechercher un vecteur  $a$  tel que (voir remarques 3) ci-dessus) :

$$Q = \sum_{k=1}^{p-1} \nu(k) \cdot \sum_{q \in \mathcal{P}_k} \left( \frac{{}^t a \cdot {}^t Y_q \cdot Y_q \cdot a}{{}^t a \cdot I_q \cdot a} \right) \text{ soit maximum avec la contrainte } \sum_{j \in J} a_j^2 = 1.$$

Ce problème d'optimisation est traité en annexe.

#### – Étape d'affectation

Le vecteur  $a$  étant fixé, on recherche la meilleure hiérarchie associée. On utilisera le fait que maximiser  $Q$  équivaut à minimiser  $Q'$ . Cette étape dite d'affectation peut s'interpréter comme celle d'une procédure de type k-means. Afin de simplifier les notations employées, nous supposons que les classes terminales de la hiérarchie initiale considérée sont constituées des singletons de la partition  $\mathcal{P}_0$ . A chaque classe terminale  $q = \{j\}$  de la hiérarchie initiale, on

peut associer une suite de noyaux constituées des  $(p-1)$  classes des partitions  $\mathcal{P}_1, \dots, \mathcal{P}_{p-1}$  contenant  $q$ . Nous noterons alors :

$\mathcal{N}_j = \{q_j(k), 1 \leq k \leq p-1\}$  l'ensemble des noyaux associés à la variable  $j$ .

$\mathcal{N}_J = \{\mathcal{N}_j / j \in J\}$

La quantité  $Q'$  à minimiser s'écrit (voir Conséquences 3) ci-dessus) :

$$Q' = \sum_{j \in J} a_j^2 \cdot \left[ \sum_{k=1}^{p-1} \nu(k) \cdot \left\| \frac{y^j}{a_j} - \frac{1}{\sqrt{a_{q_j(k)}}} \cdot z^{q_j(k)} \right\|^2 \right]$$

Pour chaque variable  $j \in J$ , on recherche, parmi  $\mathcal{N}_J$ , la suite de noyaux

$$\mathcal{N} = \{q(k), 1 \leq k \leq p-1\} \text{ minimisant } \left[ \sum_{k=1}^{p-1} \nu(k) \cdot \left\| \frac{y^j}{a_j} - \frac{1}{\sqrt{a_{q(k)}}} \cdot z^{q(k)} \right\|^2 \right].$$

La variable  $j$  est alors affectée à cette suite de noyaux.

Une nouvelle hiérarchie sur  $J$  est ainsi construite. La succession des partitions associées sera notée :

$\mathcal{P}'_0, \mathcal{P}'_1, \dots, \mathcal{P}'_{p-1}$ , où  $\mathcal{P}'_0$  et  $\mathcal{P}'_{p-1}$  sont respectivement la partition des singletons et la partition composée de la seule classe  $J$ .

### – Étape d'agrégation 2 :

À l'issue de l'étape d'affectation, la quantité  $Q'$  à minimiser s'écrit :

$$Q' = \sum_{k=1}^{p-1} \nu(k) \sum_{q' \in \mathcal{P}'_k} \sum_{j \in q'} a_j^2 \left\| \frac{y^j}{a_j} - \frac{1}{\sqrt{a_{q'}}} \cdot z^{q'} \right\|^2 \text{ où } q \text{ est l'ancien noyau associé à}$$

la nouvelle classe  $q'$ .

Par l'application du théorème de Huyghens, on minimise à nouveau  $Q'$  en

remplaçant  $\frac{1}{\sqrt{a_{q'}}} \cdot z^{q'}$  par le centre de gravité  $\frac{1}{\sqrt{a_{q'}}} \cdot z^{q'} = \sum_{j \in q'} \frac{a_j^2}{a_{q'}} \cdot \frac{y^j}{a_j}$ . On obtient

alors :

$$Q' = \sum_{k=1}^{p-1} \nu(k) \sum_{q' \in \mathcal{P}'_k} \sum_{j \in q'} a_j^2 \left\| \frac{y^j}{a_j} - \frac{1}{\sqrt{a_{q'}}} \cdot z^{q'} \right\|^2$$

Ces partitions  $\mathcal{P}'_0, \mathcal{P}'_1, \dots, \mathcal{P}'_{p-1}$  définissent un nouvel ensemble  $\mathcal{N}_J$  permettant de relancer l'étape d'agrégation 1.

### – Critère d'arrêt :

Le processus s'arrête dès que le rapport  $\frac{Q_{t+1} - Q_t}{Q_t}$  devient inférieur à un certain seuil,  $Q_t$  (respectivement  $Q_{t+1}$ ) étant la valeur de  $Q$  à l'itération  $t$  (respectivement  $t+1$ ). Dans le logiciel, ce critère est appliqué à l'issue de l'étape d'agrégation 1.

*Remarque.* — L'indice d'un noeud  $n = (q1, q2)$  était défini dans la hiérarchie initiale par la quantité :  $var(z^{q1}) + var(z^{q2}) - var(z^{q1 \cup q2})$ . Il est possible qu'après la phase d'optimisation, les valeurs de cet indice, rangées du sommet vers le bas de la hiérarchie, ne forment plus tout à fait une suite décroissante et

qu'il présente quelques inversions. L'arbre pourrait néanmoins être représenté en prenant par exemple pour indice à chaque niveau  $k$ , la somme  $\sum_{q \in \mathcal{P}_k} var(z^q)$ .

### 3.2. Élagage de la hiérarchie des variables (Étape 3)

À l'issue de la phase d'optimisation, un certain nombre de noeuds va déjà être supprimé, aucune variable ne leur ayant été affectée. Il s'agit maintenant de tester la signification des noeuds restants.

Pour chaque noeud retenu, fusionnant deux classes  $(q1, q2)$  de variables, trois variables représentatives des classes  $q1, q2, q1 \cup q2$  ont été obtenues suite à l'algorithme d'optimisation. Chacune de ces variables vaut :  $z^q = \sum_{j \in q} \frac{a_j}{\sqrt{\sum_{j \in q} (a_j)^2}} \cdot y^j, q \in \{q1, q2, q1 \cup q2\}$ .

On appellera  $\bar{z}^{(q1, q2)}$  la variable associée au noeud  $(q1, q2)$  comme étant le vecteur normé du sous espace engendré par  $z^{q1}$  et  $z^{q2}$  et orthogonal à  $z^{q1 \cup q2}$ . L'élagage de l'arbre se définit comme suit

- 1) Pour chaque variable  $y^j$ ,
  - on considère une régression linéaire multiple entre chaque variable  $y^j$  et l'ensemble des variables explicatives constituées d'abord de la variable représentative de l'ensemble complet  $J$ , à savoir  $z^J$ , puis de l'ensemble des variables  $\bar{z}$  associées aux noeuds de  $\mathcal{N}_j$ .
  - On teste l'apport d'un noeud supplémentaire représenté par une variable  $\bar{z}$  par une technique identique à celle utilisée en régression PLS pour tester l'apport d'un nouveau facteur PLS (Tenenhaus, 1998; Wold H., 1985). Cette technique basée sur la validation croisée est explicitée plus bas.
- 2) Un noeud sera déclaré significatif s'il l'est pour au moins l'une des variables  $y^j$  le concernant.
- 3) On peut procéder de manière ascendante ou descendante. Dans le cas ascendant, on part du bas de la hiérarchie, et on teste la signification des noeuds les plus bas et on s'arrête dès que l'on rencontre un noeud significatif. Dans le cas descendant, on part du haut de la hiérarchie et on s'arrête dès que l'on rencontre un noeud non significatif (option choisie dans le logiciel).
- 4) Rappel de l'indice utilisé (Tenenhaus, 1998; SIMCA-P Software)

On considère, par exemple, la variable  $y^j$  et une approche descendante. En retenant  $h+1$  noeuds les plus hauts, on calcule la prédiction  $\hat{y}_{h+1, (-i)}^j$  à l'aide du modèle incluant  $h+1$  noeuds mais sans utiliser l'observation  $i$ , puis la quantité  $PRESS_{h+1} = \sum_{i \in I} \left( y_i^j - \hat{y}_{h+1, (-i)}^j \right)^2$ .

Sans le dernier noeud dont on veut tester la signification (le noeud le plus bas parmi les  $h+1$  noeuds considérés), on calcule la prédiction  $\hat{y}_{h, i}^j$  à l'aide

du modèle incluant  $h$  noeuds et l'ensemble des observations, puis on calcule  $RSS_h = \sum_{i \in I} (y_i^j - \hat{y}_{h,i}^j)^2$ .

Le noeud  $h + 1$  est retenu si :  $\sqrt{PRESS_{h+1}} \leq 0.95 \cdot \sqrt{RSS_h}$ .

### 3.3. Représentations factorielles associées à chaque noeud retenu (Étape 4)

Pour chaque noeud retenu, fusionnant deux classes  $(q1, q2)$  de variables, trois variables  $z^{q1}, z^{q2}, z^{q1 \cup q2}$  représentatives des classes  $q1, q2, q1 \cup q2$  ont été obtenues suite à l'algorithme d'optimisation ainsi que la variable normée  $\bar{z}^{(q1, q2)}$  (voir § 3.2).

$var(z^{q1}) + var(z^{q2}) - var(z^{q1 \cup q2})$  représente la perte de variance associée à la formation du noeud  $(q1, q2)$ .

Cette quantité est une mesure de l'importance de la séparation de  $q1 \cup q2$  en ses deux successeurs  $q1$  et  $q2$ .

On multiplie la variable normée  $\bar{z}^{(q1, q2)}$  par  $\sqrt{\lambda_2} = \sqrt{var(z^{q1}) + var(z^{q2}) - var(z^{q1 \cup q2})}$ .

Les deux variables  $z^{q1 \cup q2}$  et  $\sqrt{\lambda_2} \cdot \bar{z}^{(q1, q2)}$  forment un plan dans lequel peuvent être projetées les variables  $y^j, j \in q1 \cup q2$ . De même, les individus peuvent y être représentés en considérant les valeurs des deux variables  $z^{q1 \cup q2}$  et  $\sqrt{\lambda_2} \cdot \bar{z}^{(q1, q2)}$  pour chaque individu  $i$ .

## 4. Classification hiérarchique des individus

### 4.1. Classification hiérarchique initiale des individus (Étape 5)

Cette classification initiale est obtenue à partir de la hiérarchie des variables. Supposons qu'après les phases d'optimisation et d'élagage de la hiérarchie des variables, on conserve un ensemble  $N_0$  de noeuds.

La hiérarchie initiale des individus est obtenue à partir de la variable représentative de la classe complète  $z^J$  et des variables notées  $\bar{z}^n$  associées aux noeuds  $n = (q1, q2)$  de  $N_0$  (voir §3.2).

À chacune de ces variables, correspond un niveau de la hiérarchie des individus. Chacun de ces niveaux pourra contenir plusieurs noeuds ou séparations. De manière plus précise :

- Niveau 0 : La variable représentative  $z^J$  de la classe complète partage l'ensemble des individus en deux classes

$$\{z^J \geq 0\} \text{ et } \{z^J \leq 0\}.$$

- Niveau 1 : On considère le noeud le plus haut  $(q1, q2)$  représenté par la variable  $\bar{z}^{(q1, q2)}$ . Cette variable repartage chacune de ces deux classes en deux nouvelles sous-classes suivant le signe de  $\bar{z}^{(q1, q2)}$ .

- Niveau 2 et suivantes : Chacune de ces quatre classes d'individus est ensuite rescindée suivant le signe de la variable  $\bar{z}$  associée au noeud suivant. Il est alors

possible que pour l'une des quatre classes obtenues à l'étape 2, le signe de la nouvelle variable  $\bar{z}$  reste constant auquel cas, cette classe n'est pas repartagée. Ainsi, à l'étape 3,  $2^3$  sous classes au maximum peuvent être obtenues. Les étapes 4 et suivantes se définissent de manière analogue.

## 4.2. Optimisation de la classification des individus (Étape 6)

À chacun des niveaux précédents, une partition des individus est ainsi obtenue. L'optimisation consiste à faire en sorte que leurs inerties intra classes soient globalement minimisées.

Chaque niveau sera muni d'un poids et engendrera une partition des individus. Les poids des différents niveaux seront fixés et égaux aux indices des noeuds correspondants de la hiérarchie des variables. De manière plus précise :

- Le premier niveau (niveau 0) est basée sur la variable  $z^J$ , variable représentative de la classe complète. Le poids associé à ce niveau est  $\nu(0) = \text{var}(z^J)$ . La partition des individus notée  $\mathcal{P}_0$  est composée de 2 classes.
- Le second niveau (niveau 1) est basé sur la variable  $\bar{z}^{(q1,q2)}$  associée au noeud le plus haut  $n = (q1, q2)$ . Le poids associé à ce niveau est  $\nu(1) = \text{var}(z^{q1}) + \text{var}(z^{q2}) - \text{var}(z^{q1 \cup q2})$ . La partition des individus associée, notée  $\mathcal{P}_1$ , est composée de 4 classes.
- Les niveaux  $k$  suivants sont associés aux noeuds suivants  $n_k$  représentés par les variables  $\bar{z}^{n_k}$ . En posant de même  $n_k = (q1, q2)$ , le poids associé au niveau  $k$  vaut de même  $\nu(k) = \text{var}(z^{q1}) + \text{var}(z^{q2}) - \text{var}(z^{q1 \cup q2})$  et la partition  $\mathcal{P}_k$  obtenue à ce niveau admet au maximum  $2^k$  classes.

Pour chaque niveau  $k$  ( $k \geq 0$ ), on calcule l'inertie intra-classes de la partition  $\mathcal{P}_k$ . Si on note simplement  $z^k$  la variable associée à ce niveau, l'inertie intra-classes vaut :

$$\text{intra}(\mathcal{P}_k) = \sum_{c \in \mathcal{P}_k} \sum_{i \in c} p_i \cdot \|z_i^k - \bar{z}_c^k\|^2$$

où  $p_i$  est le poids attribué à priori à l'individu  $i$  et où  $\bar{z}_c^k = \sum_{i \in c} \frac{p_i}{p_c} z_i^k$  avec

$$p_c = \sum_{i \in c} p_i$$

Le critère à minimiser s'écrit :  $Q = \sum_{k \geq 0} \nu(k) \cdot \text{intra}(\mathcal{P}_k)$ .

Pour tout individu  $i$ , on note  $c_k(i)$  la classe de la partition  $\mathcal{P}_k$  qui contient  $i$ . On obtient donc un ensemble de noyaux  $\mathcal{N}_i$  associés à l'individu  $i$  :  $\mathcal{N}_i = \{c_k(i) / k \geq 0\}$ . On notera :  $\mathcal{N}_I = \{\mathcal{N}_i / i \in I\}$ . Le critère à minimiser s'écrit alors :

$$Q = \sum_{i \in I} p_i \sum_{k \geq 0} \nu(k) \cdot \|z_i^k - \bar{z}_{c_k(i)}^k\|^2$$

L'étape d'optimisation se décompose en la succession des deux phases suivantes :

- **phase d'affectation** : Pour chaque individu  $i$  et pour chaque suite de noyaux

$$\mathcal{N} = \{c_k / k \geq 0\} \text{ de } \mathcal{N}_I, \text{ on calcule } Q_i(\mathcal{N}) = \sum_{k \geq 0} \nu(k) \cdot \|z_i^k - \bar{z}_{c_k}^k\|^2$$

On recherche la suite de noyaux  $\mathcal{N}_{\min}$  de  $\mathcal{N}_I$  pour laquelle  $Q_i(\mathcal{N}_{\min})$  est minimum.

On affecte l'individu  $i$  aux classes de  $\mathcal{N}_{\min}$ .

Dans ces conditions, en notant  $\mathcal{P}'_0, \mathcal{P}'_1, \dots, \mathcal{P}'_k, \dots$  les nouvelles partitions obtenues, le critère à minimiser devient :

$$Q = \sum_{k \geq 0} \nu(k) \cdot \sum_{c' \in \mathcal{P}'_k} \sum_{i \in c'} p_i \cdot \|z_i^k - \bar{z}_{c'}^k\|^2 \text{ où } \bar{z}_{c'}^k \text{ est la variable associée à l'ancien}$$

noyau  $c$  et où  $c'$  est la nouvelle classe obtenue et associée à  $c$ .

– **phase d'agrégation** : Pour chaque classe  $c'$  d'individus et chaque niveau  $k$ , on recalcule classiquement le centre de gravité  $\bar{z}_{c'}^k$ . Le théorème de Huyghens nous assure une nouvelle diminution de  $Q$  en écrivant :

$$Q = \sum_{k \geq 0} \nu(k) \cdot \sum_{c' \in \mathcal{P}'_k} \sum_{i \in c'} p_i \cdot \|z_i^k - \bar{z}_{c'}^k\|^2$$

– **Critère d'arrêt** : Le processus s'arrête dès que le rapport  $\frac{Q_t - Q_{t+1}}{Q_t}$  devient inférieur à un certain seuil,  $Q_t$  (respectivement  $Q_{t+1}$ ) étant la valeur de  $Q$  à l'itération  $t$  (respectivement  $t + 1$ ).

## 5. Élagage et interprétation des classes d'individus (Étapes 7 et 8)

Il est également possible que certains noeuds de la hiérarchie initiale des individus soient supprimés lors de la phase précédente d'optimisation, puisque certaines suites de noyaux  $N \in \mathcal{N}_I$  peuvent ne recevoir aucun individu lors d'une phase d'affectation. Il s'agit maintenant de tester la signification des noeuds restants.

### 5.1. Élagage de la hiérarchie des individus (Étape 7)

Chaque noeud de la hiérarchie des individus définit deux sous classes de tailles  $n_1$  et  $n_2$  et se localise à un certain niveau associé à un noeud  $(q_1, q_2)$  de la hiérarchie des variables. Il s'agit de vérifier s'il existe une variable  $y^j$  avec  $j \in q_1 \cup q_2$  pour laquelle ces deux classes d'individus de tailles  $n_1$  et  $n_2$  prennent des valeurs en moyenne différentes.

Un test de permutation est réalisé sur les  $N = n_1 + n_2$  valeurs prises par  $y^j$  notées :  $y_1, y_2, \dots, y_N$

L'hypothèse nulle  $H_0$  de ce test suppose que toutes les permutations des  $N$  éléments sont équiprobables. Sous cette hypothèse, la probabilité d'obtenir  $y_k$  au  $l^{ieme}$  tirage vaut :  $P[T_l = y_k] = \frac{1}{N}$ .

On peut alors calculer la moyenne notée  $\bar{Y}_{n_1}$  des  $n_1$  valeurs tirées, sous  $H_0$ , au hasard et sans remise dans l'ensemble des  $N$  observations. On calcule alors facilement l'espérance et la variance sous  $H_0$  de  $\bar{Y}_{n_1}$  (voir, par exemple, théorie des sondages, Tillé Y., 2001) :

$$E_{H_0} [\bar{Y}_{n1}] = \frac{1}{N} \sum_{i=1}^N y_i = \mu$$

$$var_{H_0}(\bar{Y}_{n1}) = \frac{N - n1}{n1(N - 1)} \cdot \sigma^2 \text{ avec } \sigma^2 = \frac{1}{N} \sum_{i=1}^N (y_i - \mu)^2$$

La statistique du test est la variable centrée réduite  $T = \frac{\bar{Y}_{n1} - \mu}{\sqrt{var_{H_0}(\bar{Y}_{n1})}}$

On montre par ailleurs que pour  $n1$  et  $n2$  assez grands, la loi de  $T$  peut être approximée, sous des conditions générales, par une loi normale centrée réduite (voir par exemple Van der Vaart, 2000).

Ainsi, pour chaque variable  $y^j$  de  $q1 \cup q2$ , est obtenue la valeur observée  $Tobs_j$ . On dira alors qu'un noeud  $(n1, n2)$  d'individus est jugé significatif s'il existe une variable  $y^j$  pour laquelle  $|Tobs_j| \geq 2$ . Dans le cas contraire, le noeud  $(n1, n2)$  n'est pas significatif et par suite est supprimé.

L'élagage de l'arbre s'effectue en balayant l'ensemble des branches. Si  $n^*$  est le noeud significatif le plus bas d'une branche donnée, alors tout noeud inférieur à  $n^*$  est supprimé.

## 5.2. Interprétation des classes et noeuds de la hiérarchie des individus (Étape 8)

Deux indices sont proposés, le premier interprétant les noeuds de la hiérarchie et le second chacune de ses classes.

### 5.2.1. Interprétation des noeuds

Pour chaque noeud  $(n1, n2)$  de la hiérarchie des individus, on associe le noeud  $(q1, q2)$  de la hiérarchie des variables. Un premier tableau rassemble les valeurs  $|Tobs_j|$  obtenues ci-dessus pour chaque noeud  $(n1, n2)$  et chaque variable  $y^j$  de  $q1 \cup q2$ . L'examen de ce tableau permet d'identifier les variables intervenant dans l'explication du noeud  $(n1, n2)$ .

### 5.2.2. Interprétation des classes

Considérons une classe d'individus obtenue à un niveau donné ainsi que le noeud  $(q1, q2)$  de variables associé. Pour chaque variable  $y^j$  de  $q1 \cup q2$ , un même test de permutation est appliqué aux individus de la classe concernée, considérés cette fois comme choisis au hasard au sein de la classe complète  $I$ . Une valeur  $Tobs_j$  est calculée de même et permet d'identifier les variables caractéristiques de la classe d'individus donnée.

## 6. Application à un exemple

Les données de cet exemple proviennent d'un article de Russett (1964) et sont analysées à la fois dans Gifi (1990) et dans Tenenhaus (1998).

Ce tableau croise 46 pays et 8 variables mesurant l'instabilité économique ou politique.

On trouvera ci-dessous le détail des 8 variables :

GINI : Indice de concentration de GINI

FARM : % de fermiers possédant la moitié des terres en commençant par les petites surfaces

RENT : % de fermiers locataires de leurs terres

GNPR : Produit national brut

LABO : % de personnes actives travaillant dans l'agriculture

INST : Indice de stabilité politique variant entre 0 (très stable) à 17 (très instable)

ECKS : Nombre de conflits violents

DEAT : Nombre de personnes tuées lors de manifestations

Le tableau est donné en fin d'article (Tableau 13).

### 6.1. Classification initiale des variables (Étape 1)

Le tableau 1 restitue les valeurs des coefficients  $a_j$ . Dans le tableau 2, l'analyse en composantes principales associée à chaque noeud est décrite par les 2 valeurs propres extraites à savoir lamda1 et lamda2 ainsi que par leurs pourcentages associées %lamda1 et %lamda2. D'après la propriété 5, la somme des secondes valeurs propres augmentée de la première valeur propre issue de l'ACP associée au noeud le plus haut est égale à la somme des variances des 8 variables initiales du tableau, à savoir 8, puisque ces 8 variables ont été centrées et réduites. Ainsi, les pourcentages de chaque seconde valeur propre par rapport à ce total de 8, sont donnés dans la colonne « %inertie ». Ce qui permet d'exprimer les indices des noeuds sous la forme d'un taux d'inertie.

Le tableau 3 traduit également cette même propriété 5.

La figure 2 donne la représentation de la hiérarchie initiale.

TABLEAU 1. — Coefficients  $(a_j)_{j \in J}$ .

Variabes	Gini	Farm	Rent	Gnpr	Labo	Inst	Ecks	Deat
Numérotation j	1	2	3	4	5	6	7	8
$a_j$	0.43	0.43	0.10	-0.43	0.43	0.23	0.31	0.31

CLASSIFICATION FACTORIELLE HIÉRARCHIQUE OPTIMISÉE

TABLEAU 2. — Décomposition de la hiérarchie initiale des variables.

noeud	ainé	benjamin	%lamda1	%lamda2	lamda1	lamda2	%inertie	%cumulée
9	1	2	96.88	3.12	1.94	0.06	0.78	0.78
10	4	5	90.11	9.89	1.80	0.20	2.47	3.25
11	7	8	66.81	33.19	1.34	0.66	8.30	11.55
12	10	11	73.54	26.46	2.31	0.83	10.38	21.93
13	9	6	71.52	28.48	2.10	0.84	10.46	32.39
14	3	13	68.56	31.44	2.13	0.97	12.19	44.57
15	14	12	65.98	34.02	2.93	1.51	18.85	63.43
total						5.07	63.43	

TABLEAU 3. — Résultats de la hiérarchie initiale des variables.

Total lamda2	5.07
Variance de la variable représentative de la classe complète	2.93
Somme des variances des 8 variables initiales	8

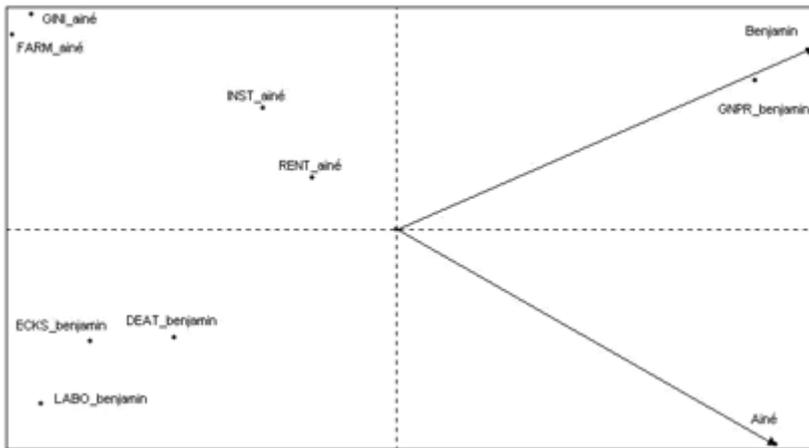


FIG 2. — Hiérarchie initiale des variables.

## 6.2. Optimisation de la hiérarchie des variables (Étape 2)

Le tableau 4 précise les différentes valeurs prises successivement par le critère à maximiser.

À l'issue de l'étape d'optimisation, un nouveau coefficient  $a$  est obtenu. L'ensemble de ses valeurs est donné dans le tableau 5. Chaque classe  $q$  est alors représentée par la nouvelle variable  $z^q = \sum_{j \in q} \frac{a_j}{\sqrt{\sum_{j \in q} (a_j)^2}} \cdot y^j$ .

Pour un noeud  $(q1, q2)$ , les dénominations lamda1 et lamda2 représentent respectivement la variance de la variable  $z^{q1 \cup q2}$  et la quantité  $var(z^{q1}) + var(z^{q2}) - var(z^{q1 \cup q2})$ . On peut remarquer que  $var(z^{q1 \cup q2})$  est toujours inférieure ou égale à la première valeur propre issue de l'ACP constituée par les deux variables  $z^{q1}$  et  $z^{q2}$ . Le tableau 6 rassemble les valeurs lamda1 et lamda2 ainsi que %lamda1 et %lamda2 pour les deux noeuds 15 et 14 retenus.

TABLEAU 4. — Les étapes de la phase d'optimisation.

Étapes	valeurs du critère	Étapes	valeurs du critère
initial	13836.348	affectation	13949.604
agrégation	13949.604	agrégation	13949.604

TABLEAU 5. — Coefficients  $(a_j)_{j \in J}$ .

Variabes	Gini	Farm	Rent	Gnpr	Labo	Inst	Ecks	Deat
Numérotation j	1	2	3	4	5	6	7	8
$a_j$	0.44	0.45	0.10	-0.43	0.43	0.16	0.37	0.26

TABLEAU 6. — Hiérarchie des variables après optimisation : ACP associées.

noeud	ainé	benjamin	%lamda1	%lambda2	lamda1	lamda2
12	11	10	74.07	25.93	2.31	0.81
13	9	6	70.90	29.10	2.08	0.85
14	13	3	68.55	31.45	2.11	0.97
15	14	12	66.72	33.28	2.95	1.47

À l'issue de l'étape d'optimisation, seuls les noeuds 15 à 12 sont conservés. Dans le cadre du tableau 7, L'ainé et le benjamin d'un noeud sont représentés par les entiers 1 et 2, munis éventuellement d'un signe. En effet chaque classe est composée de variables corrélées positivement ou négativement à la variable représentative de la classe. L'attribution d'un signe permet de distinguer ces deux catégories de variables. Lorsque le signe n'apparaît pas, les corrélations des variables de la classe avec leur variable représentative sont de signe constant.

On trouvera dans le tableau 7 les contenus des classes de la hiérarchie des variables obtenue après optimisation et à la figure 3, la représentation de la hiérarchie optimisée.

TABLEAU 7. — Hiérarchie des variables après optimisation.

variables	Classes	Variable	noeud	noeud	noeud	noeud
	terminales	représentative de J	15	14	13	12
Gnpr	10	-1	2	0	0	2
Labo	10	1	-2	0	0	-2
Deat	11	1	-2	0	0	1
Ecks	11	1	-2	0	0	1
Rent	3	1	1	2	0	0
Farm	9	1	1	1	1	0
Inst	6	1	1	1	2	0
Gini	9	1	1	1	1	0

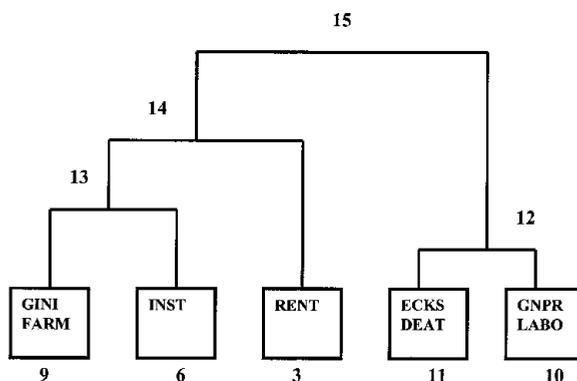


FIG 3. — Hiérarchie des variables après optimisation.

### 6.3. Élagage de la hiérarchie des variables (Étape 3)

Afin de simplifier la présentation des résultats, nous avons pris, comme variables à expliquer, les variables représentatives des classes terminales (à savoir 11, 10, 9, 6, 3) et non l'ensemble des variables  $y^j$  comme cela est présenté dans l'article.

La quantité  $Q^2$  calculée testant l'introduction d'un noeud supplémentaire  $h$  vaut :

$$Q_h^2 = 1 - \frac{PRESS_h}{RSSS_{h-1}}$$

Le noeud est significatif si  $Q_h^2 \geq 0.0975$  (voir, Tenenhaus ou SIMCA-P)

On retrouvera dans le tableau ci-dessous le modèle linéaire considéré avec les notations explicitées dans l'article. Le processus de détermination des noeuds significatifs est réalisé de manière descendante et s'arrête pour une variable à

expliquer donnée dès l'apparition d'un noeud non significatif, ou si la totalité des variables explicatives a été balayée.

TABLEAU 8. — Les modèles linéaires de la phase d'élagage.

Variable à expliquer	Variables explicatives	noeud testé	$Q^2$	conclusion
$z^{11}$	$z^J, \bar{z}^{15}$	<b>15</b>	0.15	<b>significatif</b>
$z^{11}$	$z^J, \bar{z}^{15}, \bar{z}^{12}$	<b>12</b>	1.00	<b>significatif</b>
$z^{10}$	$z^J, \bar{z}^{15}$	<b>15</b>	0.53	<b>significatif</b>
$z^{10}$	$z^J, \bar{z}^{15}, \bar{z}^{12}$	<b>12</b>	1.00	<b>significatif</b>
$z^9$	$z^J, \bar{z}^{15}$	<b>15</b>	0.91	<b>significatif</b>
$z^9$	$z^J, \bar{z}^{15}, \bar{z}^{14}$	14	0.03	non significatif
$z^6$	$z^J, \bar{z}^{15}$	15	0.04	non significatif
$z^3$	$z^J, \bar{z}^{15}$	15	-0.02	non significatif

Seuls, les noeuds 15 et 12 seront conservés.

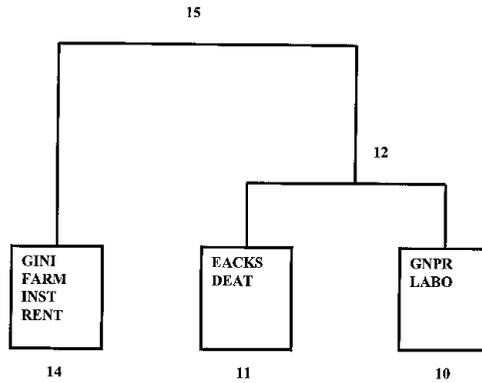


FIG 4. — Hiérarchie des variables après optimisation et élagage.

#### 6.4. Représentations factorielles (Étape 4)

Elles sont données et commentées au paragraphe 6.6 (figures 7, 8, 9, 10). Dans les représentations des individus, autrement dit des pays (figures 7 et 9), l'identificateur de chaque pays comprend le nom du pays ainsi que le numéro de la classe dans lequel il se trouve au niveau considéré. Ainsi, à partir du noeud 15 de la hiérarchie des variables (voir figure 4), une partition en quatre classes de l'ensemble des pays sera obtenue (§ 6.5) et visualisée dans la représentation factorielle correspondante (figure 7). De même, le noeud suivant de la hiérarchie des variables, à savoir le noeud 12, générera de nouvelles classes d'individus représentées par leurs numéros dans la figure 9.

**6.5. Hiérarchie initiale des individus, optimisation et élagage (étapes 5, 6, 7)**

La hiérarchie des individus est déduite de celle des variables. Comme la hiérarchie des variables ne comporte que deux noeuds 15 et 12, trois variables seront utilisées dans la construction de la hiérarchie des individus à savoir  $z^J$ ,  $\bar{z}^{15}$ ,  $\bar{z}^{12}$ . La hiérarchie initiale comportera donc 8 classes terminales.

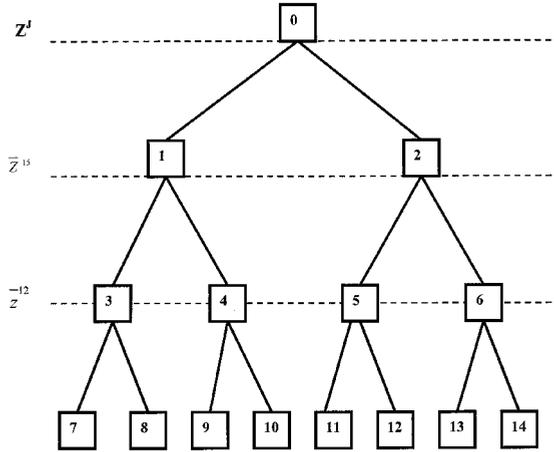


FIG 5. — Hiérarchie initiale des individus.

Les trois hiérarchies des individus obtenues respectivement initialement, après optimisation, puis après élagage sont données en fin d'article (tableau 12). On vérifiera qu'à l'issue des 3 niveaux définis respectivement par la variable représentative de la classe complète ( $z^J$ ) et par les deux variables associées aux noeuds 15 et 12, on obtient (tableau 12) une partition en 8 classes (hiérarchie initiale), également en 8 classes (après optimisation), et enfin en 7 classes (après élagage et optimisation).

La phase d'optimisation s'effectue en 10 étapes. Nous donnons ci-dessous les valeurs du critère à minimiser (tableau 9). La hiérarchie des individus obtenue après optimisation reste identique dans sa présentation (Figure 5). Seul, le contenu des classes a été modifié (voir tableau 12).

La phase d'élagage s'effectue à partir d'un test de permutation. Pour chacun des trois niveaux (variable représentative de la classe complète et les deux noeuds-variables 15 et 12), on observe un certain nombre de noeuds-individus dont il s'agit de tester la signification. Chaque noeud-individus s'interprète ici comme un dipole constitué des deux sous classes générées par ce noeud. Les numéros attribués à ces deux sous-classes sont donnés dans le tableau décrivant la hiérarchie des individus après optimisation (tableaux 10 ou 12) ou peuvent être lus sur la figure 5.

Pour chacun d'eux et chacune des variables  $y^j$  concernée, est calculée la valeur absolue  $|T_{obs}|$  de la valeur observée de la statistique T du test.

TABLEAU 9. — Les étapes de la phase d'optimisation.

Étapes	Phases	Valeurs du critère
1	Affectation	1.616
2	Agrégation	1.606
3	Affectation	1.593
4	Agrégation	1.581
5	Affectation	1.572
6	Agrégation	1.559
7	Affectation	1.553
8	Agrégation	1.545
9	Affectation	1.545
10	Agrégation	1.545

Les notations  $z^J$ ,  $n15$ ,  $n12$  correspondent aux trois niveaux de la hiérarchie. Seul le quatrième noeud du niveau  $n12$  n'est pas significatif car  $\max(|T_{obs}|) \leq 2$ . Les classes 13 et 14 seront donc regroupées. Leur réunion conservera le numéro 6. On obtient alors une partition en 7 classes après élagage. On trouvera ci-dessous la représentation de cette hiérarchie (Figure 6). Le contenu des classes peut être consulté dans le tableau 12.

 TABLEAU 10. — Phase d'élagage de la hiérarchie : valeurs  $|T_{obs}|$ .

niv	dipoles	Gini	Farm	Rent	Gnpr	Labo	Inst	Ecks	Deat	$\max( T_{obs} )$
$z^J$	$0 \rightarrow (1, 2)$	4.29	4.63	0.87	4.65	4.65	1.72	3.84	2.26	4.65
$n15$	$1 \rightarrow (3, 4)$	3.12	2.96	2.00	1.64	2.76	0.62	1.65	1.25	3.12
$n15$	$2 \rightarrow (5, 6)$	3.35	3.05	0.19	1.83	2.13	0.97	1.10	1.93	3.35
$n12$	$3 \rightarrow (7, 8)$				1.87	1.63		2.75	0.47	2.75
$n12$	$4 \rightarrow (9, 10)$				1.96	1.51		2.28	0.20	2.28
$n12$	$5 \rightarrow (11, 12)$				2.62	2.45		2.02	1.13	2.62
$n12$	$6 \rightarrow (13, 14)$				0.24	0.59		1.90	1.39	1.90

## 6.6. Interprétation des classes-individus (Étape 8)

Cette interprétation s'effectue à partir des résultats du tableau 11 et permettra d'étiqueter la hiérarchie finale des individus (Figure 6). Pour chaque noeud de chaque niveau, on trouve tout d'abord, dans ce tableau, la ligne correspondant à l'indice «dipole». Cette ligne contient l'ensemble des  $|T_{obs}|$  déjà données dans le tableau 10 et testant l'opposition entre l'ainé et le benjamin du noeud considéré. On trouve ensuite deux lignes associées à l'indice «classe».

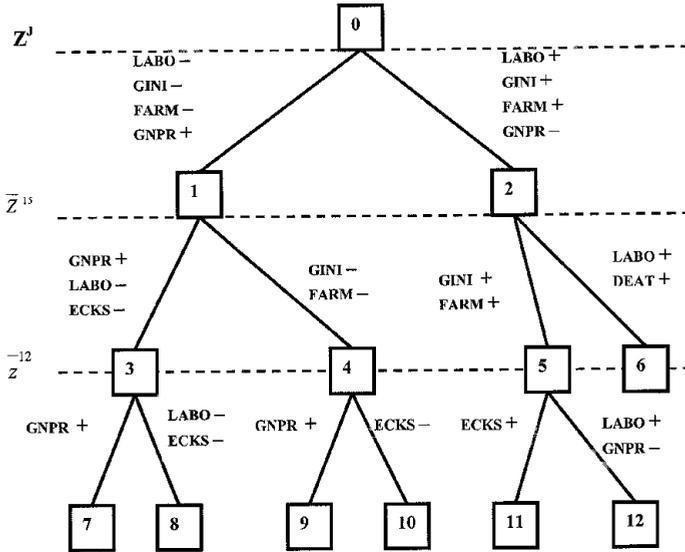


FIG 6. — Hiérarchie des individus après optimisation, élagage et étiquetage.

Chacune de ces deux lignes rassemble les valeurs  $T_{obs}$  testant cette fois les valeurs prises par la classe considérée par rapport à la totalité des valeurs prises par l'ensemble des individus. Compte tenu de la normalité asymptotique, chaque classe d'individus sera caractérisée par les variables dont la valeur  $|T_{obs}|$  est supérieure à 2. L'examen de ces valeurs et des représentations factorielles trouvées (Figures 7, 8, 9, 10) permet d'obtenir facilement un descriptif synthétique et hiérarchisé de l'information significative extraite du tableau de Russett.

Le premier niveau ( $z^j$ ) est sans surprise et oppose pays riches (classe 1) et pauvres (classe 2). Les pays riches sont caractérisés par un produit national brut élevé (GNPR+) , un pourcentage de personnes actives dans l'agriculture plus faible (LABO-) et globalement une répartition plus égalitaire des terres (GINI- et FARM-). Les pays pauvres sont caractérisés par les caractéristiques inverses. Cette opposition se retrouve le long du premier axe des représentations factorielles associées au noeud 15 (Figures 7 et 8).

L'ensemble des pays riches n'est cependant pas homogène et se décompose en quatre classes notées 7, 8, 9, 10. Une première séparation oppose les classes 7 et 8 aux classes 9 et 10. Elle s'explique par les variables GINI et FARM prenant des valeurs plus faibles (répartition plus égalitaire encore des terres) pour les classes 10 et 9={France, Canada}. Une seconde séparation oppose d'abord les classes 7 et 8, puis les classes 9 et 10. L'explication de ces oppositions est identique dans les deux cas et se traduit par un produit national brut plus élevé (GNPR+) pour les classes 7={USA, Royaume Uni} et 9={France, Canada} et un nombre plus faible de conflits violents (ECKS-) pour les deux autres classes 8 et 10.

CLASSIFICATION FACTORIELLE HIÉRARCHIQUE OPTIMISÉE

TABLEAU 11. — Valeurs  $|T_{obs}|$  (testdipole), et valeurs  $T_{obs}$  (testclasses).

Niv	test	Gini	Farm	Rent	Gnpr	Labo	Inst	Ecks	Deat
$z^J$	Dipole(1, 2)	4.29	4.63	0.83	4.65	4.65	1.72	3.84	2.26
$z^J$	Classe 1	-4.29	-4.63	-0.87	4.65	-4.65	-1.72	-3.84	-2.26
$z^J$	Classe 2	4.29	4.63	0.87	-4.65	4.65	1.72	3.84	2.26
n15	Dipole(3, 4)	3.12	2.96	2.00	1.64	2.76	0.62	1.65	1.25
n15	Classe 3	-0.53	0.64	1.17	4.17	-4.59	-0.52	-3.09	-1.39
n15	Classe 4	-4.63	-4.94	-2.25	1.31	-0.86	-1.54	-1.44	-1.29
n15	Dipole(5, 6)	3.35	3.05	0.19	1.83	2.13	0.97	1.10	1.93
n15	Classe 5	4.71	4.32	0.74	-2.85	2.25	1.81	1.92	-0.19
n15	Classe 6	-0.37	0.58	0.20	-2.49	3.24	-0.04	2.61	3.21
n12	Dipole(7, 8)				1.87	1.63		2.75	0.47
n12	Classe 7				3.24	-2.29		-0.33	-0.53
n12	Classe 8				2.82	-3.76		-3.15	-1.22
n12	Dipole(9, 10)				1.96	1.51		2.28	0.20
n12	Classe 9				2.32	-1.54		0.88	-0.53
n12	Classe 10				0.17	-0.11		-2.05	-1.13
n12	Dipole(11, 12)				2.62	2.45		2.02	1.13
n12	Classe 11				-0.14	-0.85		2.33	0.53
n12	Classe 12				-2.91	2.82		0.76	-0.48

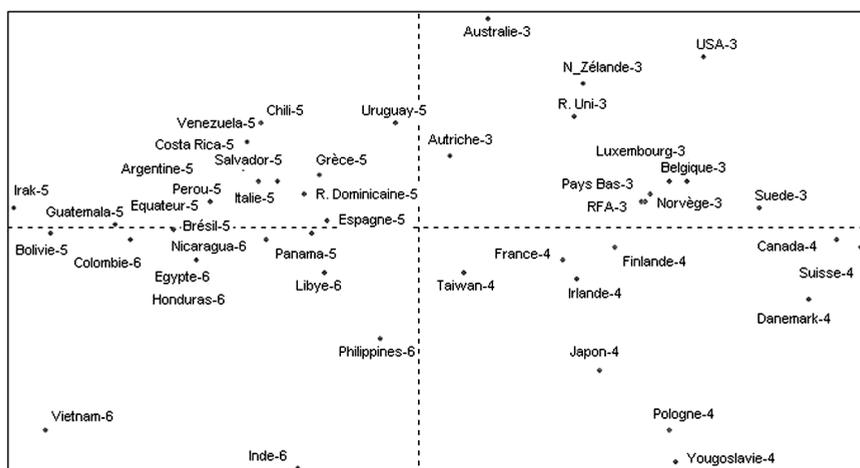


FIG 7. — Représentation des individus dans le plan factoriel associé au noeud 15.

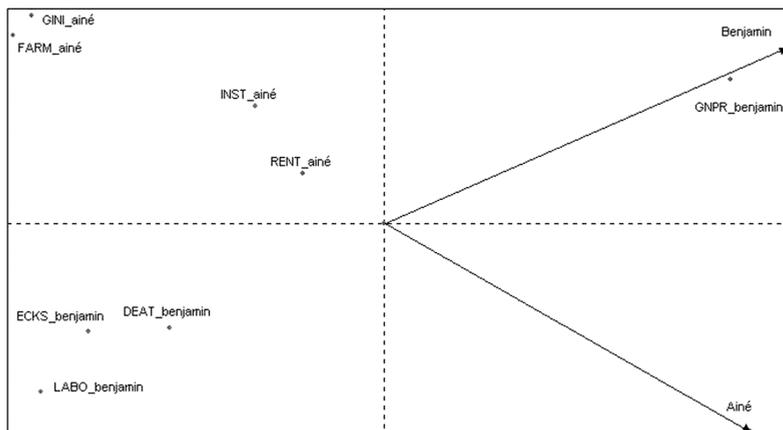


FIG 8. — Représentation des variables du noeud 15 dans le plan factoriel associé.

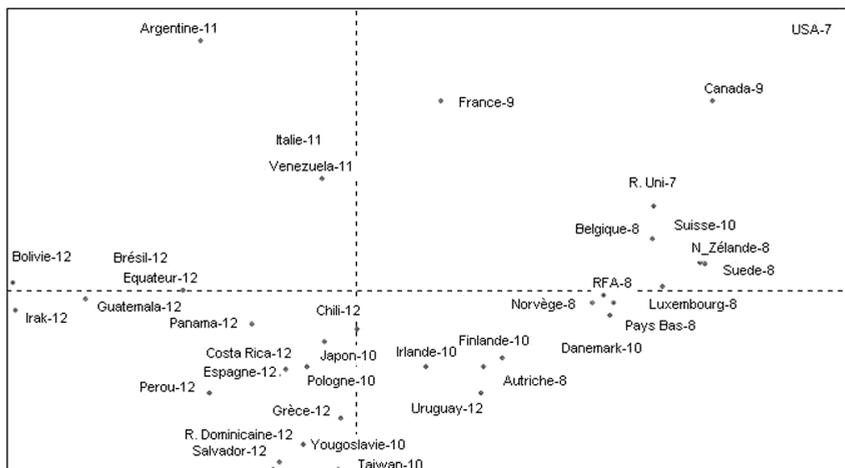


FIG 9. — Représentation des individus des classes 7, 8, 9, 10, 11, 12 dans le plan factoriel associé au noeud 12.

L'ensemble des pays pauvres (Classe 2) isole d'abord la classe 6 constituée des pays présentant un nombre élevé de conflits violents (ECKS+) et de morts lors de manifestations (DEAT +). Cette classe apparaît clairement dans les représentations factorielles (figures 7 et 8). Le Vietnam et l'Inde sont caractéristiques de cette classe. Les deux autres classes de pays pauvres (11 et 12), présentant une répartition inégalitaire des terres (GINI+ et FARM+), s'opposent au dernier niveau de la hiérarchie. Les pays de la classe 11 se distinguent par de nombreux conflits violents (ECKS+). Les trois pays qui composent cette classe 11 (Argentine, Italie, Vénézuéla) apparaissent de manière marquée dans les représentations factorielles (figures 9 et 10). La

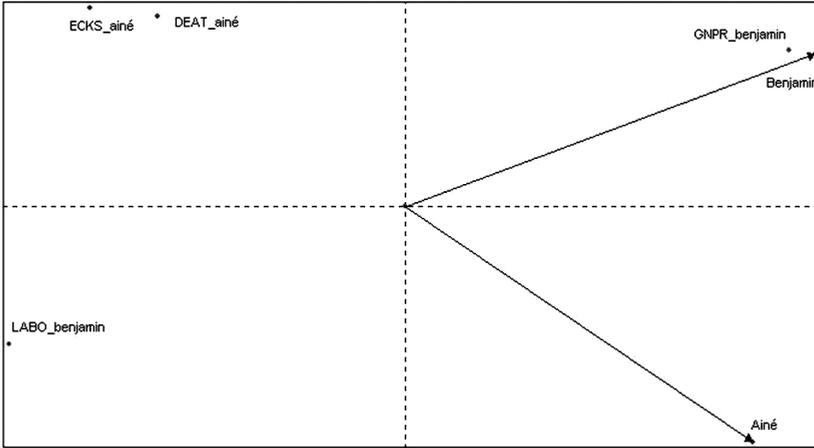


FIG 10. — Représentation des variables du noeud 12 dans le plan factoriel associé.

dernière classe de pays pauvres (12) est caractérisée par un nombre élevé de personnes travaillant dans l’agriculture (LABO+) et par un climat social plus paisible (DEAT et ECKS moins élevés).

## 7. Conclusion

Le fait d’unifier en une même méthode les deux approches d’analyse d’un tableau que sont l’analyse factorielle et les classifications hiérarchiques de ses lignes ou colonnes permet d’obtenir des résultats plus synthétiques et donc plus rapidement interprétables. Les problèmes liés à la détermination du nombre d’axes factoriels à retenir ou à la localisation des coupures des arbres sont à présent réalisés de manière automatique par des procédures d’optimisation et d’élégage ce qui simplifie également la tâche de l’utilisateur. Enfin, le logiciel mettant en oeuvre la méthodologie exposée dans cet article a été écrit en fortran et sera disponible sur la page internet de l’auteur avec des extensions de la méthode aux tableaux de contingence simples et multiples.

## 8. Annexe

Dans le cadre de la phase d’optimisation la hiérarchie des variables, la hiérarchie  $\mathcal{H}$  étant fixée, un vecteur  $a = (a_j)_{j \in J}$  est recherché maximisant sous la contrainte  $\sum_{j \in J} a_j^2 = 1$  la quantité  $Q = \sum_{k=1}^{p-1} \nu(k) \cdot \sum_{q \in \mathcal{P}_k} \left( \frac{{}^t a \cdot {}^t Y_q \cdot Y_q \cdot a}{{}^t a \cdot I_q \cdot a} \right)$ .

On note :

–  $Y_q$  la matrice de dimensions  $(n, p)$  dont les colonnes  $j$  valent  $y^j$  lorsque  $j$  appartient à la classe  $q$  et valent 0 sinon.

–  $Y_q^*$  la matrice de dimensions  $(n, \text{card}(q))$  dont les colonnes  $j$  ( $j \in q$ ) valent  $y^j$

–  $I_q$  est la matrice diagonale de dimensions  $(p, p)$  telle que  $I_q(j, j) = \begin{cases} 1, & j \in q \\ 0, & \text{ailleurs} \end{cases}$

Ce problème d'optimisation est traité dans cet article par une heuristique basée sur les remarques données ci-dessous. Cette heuristique a été appliquée à de nombreux exemples et a montré son efficacité. Deux programmes de la librairie IMSL mettant en oeuvre des techniques d'analyses numériques d'optimisation sous contraintes ont été également testés. Malheureusement, des problèmes liés à la non convergence vers l'extrémum des processus utilisés sont apparus et n'ont pas permis leur emploi.

1)  $\forall k \in [1, p-1], \forall q \in \mathcal{P}_k$ , chaque terme  $\frac{{}^t a \cdot {}^t Y_q \cdot Y_q \cdot a}{{}^t a \cdot I_q \cdot a}$  de  $Q$  admet pour maximum la plus grande valeur propre  $\lambda_1(q, k)$  de la matrice  ${}^t Y_q \cdot Y_q$

2) La quantité  $Q$  à maximiser est donc majorée par  $\sum_{k=1}^{p-1} \nu(k) \cdot \sum_{q \in \mathcal{P}_k} \lambda_1(q, k)$

3)  $\forall k \in [1, p-1]$ , notons  $\lambda_1(k) = \sum_{q \in \mathcal{P}_k} \lambda_1(q, k)$ .

On vérifie facilement que :  $\text{Sup}_{\substack{a \in R^p \\ {}^t a \cdot a = 1}} \left( \sum_{q \in \mathcal{P}_k} \left( \frac{{}^t a \cdot {}^t Y_q \cdot Y_q \cdot a}{{}^t a \cdot I_q \cdot a} \right) \right) = \lambda_1(k)$

En effet, cette borne supérieure est atteinte pour un vecteur normé notée  $a_{\max}(k)$  tel que :

$\forall q \in \mathcal{P}_k$ , le sous-vecteur de  $a_{\max}(k)$  composé uniquement des composantes d'indice  $j$  appartenant à  $q$  est vecteur propre de  ${}^t Y_q \cdot Y_q$  associé à la plus grande valeur propre  $\lambda_1(q, k)$ .

$\forall q \in \mathcal{P}_k$ , ce même vecteur propre est aussi vecteur propre de  $\frac{\lambda_1(k)}{\lambda_1(q, k)} \cdot {}^t Y_q \cdot Y_q$  associé à la plus grande valeur propre  $\lambda_1(k)$ . Autrement dit, puisque  $\mathcal{P}_k$  est une partition de  $J$ ,  $a_{\max}(k)$  sera aussi vecteur propre de  $\sum_{q \in \mathcal{P}_k} \frac{\lambda_1(k)}{\lambda_1(q, k)} \cdot {}^t Y_q \cdot Y_q$  associé à la plus grande valeur propre  $\lambda_1(k)$ . On peut donc écrire :

$$\text{Sup}_{\substack{a \in R^p \\ {}^t a \cdot a = 1}} \left( \sum_{q \in \mathcal{P}_k} \left( \frac{{}^t a \cdot {}^t Y_q \cdot Y_q \cdot a}{{}^t a \cdot I_q \cdot a} \right) \right) = \text{Sup}_{\substack{a \in R^p \\ {}^t a \cdot a = 1}} \left( \sum_{q \in \mathcal{P}_k} \frac{\lambda_1(k)}{\lambda_1(q, k)} \cdot {}^t a \cdot {}^t Y_q \cdot Y_q \cdot a \right)$$

4) Compte tenu des remarques précédentes, on propose de prendre comme solution approchée du problème d'optimisation initial le vecteur  $a$  qui maximise la quantité

$$\tilde{Q} = \sum_{k=1}^{p-1} \nu(k) \cdot \sum_{q \in \mathcal{P}_k} \frac{\lambda_1(k)}{\lambda_1(q, k)} \cdot {}^t a \cdot {}^t Y_q \cdot Y_q \cdot a \text{ sous la contrainte } \sum_{j \in J} a_j^2 = 1.$$

Ce qui revient à chercher le vecteur propre de  $\sum_{k=1}^{p-1} \nu(k) \cdot \sum_{q \in \mathcal{P}_k} \frac{\lambda_1(k)}{\lambda_1(q, k)} \cdot {}^t Y_q \cdot Y_q$  associé à la plus grande valeur propre.

L'algorithme d'optimisation de la hiérarchie des variables est adapté comme suit :

- Si  $a$  est le vecteur obtenu à l'issue de la hiérarchie initiale, on applique la phase d'agrégation (recherche d'un nouveau vecteur  $a$ ), si ce nouveau vecteur est meilleur (valeur  $Q$  plus élevée), on aborde la phase d'affectation avec ce vecteur, sinon on aborde la phase d'affectation avec le vecteur obtenu à l'issue de la construction de la hiérarchie initiale.
- Si  $a$  est un vecteur obtenu au cours de la phase d'optimisation et que l'on cherche à améliorer lors d'une nouvelle phase d'agrégation, si le nouveau vecteur est meilleur (valeur  $Q$  plus élevée), on aborde une nouvelle phase d'affectation avec ce vecteur, sinon on arrête le processus d'optimisation.

## 9. Références

- [1] BENCHEIKH Y.K. (1992). Classification automatique et modèles. Thèse Université de Metz.
- [2] BENCHEIKH Y.K. (2004). Classification croisée et mélanges sur données quantitatives. *Revue de Statistique Appliquée*, LII(2), 71-86.
- [3] BENZECRI J.-P. (1976). *L'Analyse des Données (Volumes I and II)*. Dunod, Paris.
- [4] BENZECRI J.-P., JAMBU M., LEBEAUX M.O. (1980). Aides à l'interprétation en classification automatique. *Les cahiers de l'analyse des données*, 5, 1, pp.101-123.
- [5] BRUYNOOGHE M. (1978). Large data set clustering methods using the concept of space contraction. *Compstat. 3*, Physika Verlag, Vienna, pp 239-245.
- [6] CAZES P. (1984). Correspondances hiérarchiques et ensembles associés. *Les cahiers du Bureau Universitaire de Recherche Opérationnelle*, n° 43-44, pp 43-142.
- [7] CELEUX G. (1988). Classification et modèles. *Revue de Statistique Appliquée*, XXXVI (4), 43-58
- [8] DENIMAL J.J. (2000). Correspondances hiérarchiques : une nouvelle approche. XXXII<sup>ième</sup> Journées de Statistiques, 15-19 mai 2000. Fès, Maroc.
- [9] DENIMAL J.J. (2001). Hierarchical factorial analysis. 10th International Symposium on Applied Stochastic Models and Data Analysis. 12-15 juin 2001. Compiègne.
- [10] DIDAY E. (1979). Optimisation en classification automatique. INRIA, Paris.
- [11] GIFI A. (1990). Non linear multivariate analysis. John Wiley&sons, chichester.
- [12] GOVAERT G. (1983). Classification croisée. Thèse d'Etat, Université Pierre et Marie Curie, Paris 6.
- [13] JAMBU M., LEBEAUX M.O. (1978). Classification automatique pour l'analyse des données. Dunod, Paris.
- [14] JUAN J. (1982). Classification automatique hiérarchique selon les voisins réciproques. *Les cahiers de l'analyse des données*, Vol 7, n°2.
- [15] LEBART L., MORINEAU A., WARWICK K.M. (1984). *Multivariate descriptive statistical analysis*. New York, Wiley.
- [16] LERMAN I.C. (1979). Les représentations factorielles de la classification. *R.A.I.R.O.*, Vol 13, n°2 et n°3.

- [17] LERMAN I.C. (1991). Foundations of the Likelihood Linkage Analysis (LLA) classification method. Applied Stochastic Models and Data Analysis, Vol.7, pp 63-76.
- [18] QANNARI E.M., VIGNEAU F., COURCOUX PH. (1999). Classification des variables autour de composantes principales. Applications. XXXI<sup>ième</sup> Journées de Statistiques, 17-21 mai 1999. Grenoble.
- [19] RAO C.R., (1964). The use and interpretation of Principal Component Analysis in Applied Research. *Sankhya A*, 26, pp 329-358.
- [20] RUSSETT B.M. (1964). Inequality and Instability. World Politics, 21, pp 442, 454.
- [21] SAS/STAT (1996). User's guide, SAS institute Inc., Cary, Nc, USA.
- [22] TENENHAUS M. (1998). La régression PLS. Editions Technip.
- [23] TILLE Y. (2001). Théorie des sondages. Dunod.
- [24] UMETRICS ACADEMY (1996). SIMCA-P for WINDOWS. Graphical Software for multivariate Process Modeling.
- [25] VAN DER VAART A.W. (2000). Asymptotic Statistics. Cambridge University Press.
- [26] WARD J.H., (1963). Hierarchical grouping to optimize an objective function. Journal of the American Statistical Association, 58, pp 236-244.
- [27] WOLD H. (1985). Partial least squares. Encyclopedia of statistical Sciences, vol 6, Kotz S.&Johnson N.L.(Eds), John Wiley&sons, New York, pp 581, 591.

CLASSIFICATION FACTORIELLE HIÉRARCHIQUE OPTIMISÉE

TABLEAU 12. — Hiérarchies des individus initiale, après optimisation, après élagage.

<i>hiérarchie initiale</i>				<i>Après optimisation</i>				<i>Après élagage</i>			
	$z^J$	15	12		$z^J$	15	12		$z^J$	15	12
Australie	1	3	7	USA	1	3	7	USA	1	3	7
Belgique	1	3	7	Roy. Uni	1	3	7	Roy. Uni	1	3	7
USA	1	3	7	Australie	1	3	8	Australie	1	3	8
Luxembourg	1	3	7	Belgique	1	3	8	Belgique	1	3	8
N. Zelande	1	3	7	Luxembourg	1	3	8	Luxembourg	1	3	8
Roy. Uni	1	3	7	Norvège	1	3	8	Norvège	1	3	8
Suède	1	3	7	N. Zelande	1	3	8	N. Zelande	1	3	8
Norvège	1	3	8	Pays Bas	1	3	8	Pays Bas	1	3	8
Pays Bas	1	3	8	Suède	1	3	8	Suède	1	3	8
Autriche	1	3	8	Autriche	1	3	8	Autriche	1	3	8
RFA	1	3	8	RFA	1	3	8	RFA	1	3	8
Canada	1	4	9	Canada	1	4	9	Canada	1	4	9
Suisse	1	4	9	France	1	4	9	France	1	4	9
France	1	4	9	Danemark	1	4	10	Danemark	1	4	10
Danemark	1	4	10	Irlande	1	4	10	Irlande	1	4	10
Irlande	1	4	10	Suisse	1	4	10	Suisse	1	4	10
Finlande	1	4	10	Finlande	1	4	10	Finlande	1	4	10
Japon	1	4	10	Japon	1	4	10	Japon	1	4	10
Pologne	1	4	10	Pologne	1	4	10	Pologne	1	4	10
Taiwan	1	4	10	Taiwan	1	4	10	Taiwan	1	4	10
Yougoslavie	1	4	10	Yougoslavie	1	4	10	Yougoslavie	1	4	10
Argentine	2	5	11	Argentine	2	5	11	Argentine	2	5	11
Italie	2	5	11	Italie	2	5	11	Italie	2	5	11
Equateur	2	5	11	Venezuela	2	5	11	Venezuela	2	5	11
Venezuela	2	5	11	Uruguay	2	5	12	Uruguay	2	5	12
Uruguay	2	5	12	Brésil	2	5	12	Brésil	2	5	12
Chili	2	5	12	Chili	2	5	12	Chili	2	5	12
Costa-Rica	2	5	12	Costa Rica	2	5	12	Costa Rica	2	5	12
Grèce	2	5	12	Grèce	2	5	12	Grèce	2	5	12
Rep. Domi.	2	5	12	Bolivie	2	5	12	Bolivie	2	5	12
Espagne	2	5	12	R. Domi.	2	5	12	R. Domi.	2	5	12
Guatemala	2	5	12	Equateur	2	5	12	Equateur	2	5	12
Irak	2	5	12	Espagne	2	5	12	Espagne	2	5	12
Perou	2	5	12	Guatemala	2	5	12	Guatemala	2	5	12
Salvador	2	5	12	Irak	2	5	12	Irak	2	5	12
Inde	2	6	13	Panama	2	5	12	Panama	2	5	12
Brésil	2	6	13	Perou	2	5	12	Perou	2	5	12
Colombie	2	6	13	Salvador	2	5	12	Salvador	2	5	12
Bolivie	2	6	13	Inde	2	6	13	Inde	2	6	6
Honduras	2	6	13	Colombie	2	6	13	Colombie	2	6	6
Vietnam	2	6	13	Vietnam	2	6	13	Vietnam	2	6	6
Egypte	2	6	14	Egypte	2	6	14	Egypte	2	6	6
Libye	2	6	14	Honduras	2	6	14	Honduras	2	6	6
Nicaragua	2	6	14	Libye	2	6	14	Libye	2	6	6
Panama	2	6	14	Nicaragua	2	6	14	Nicaragua	2	6	6
Philippines	2	6	14	Philippines	2	6	14	Philippines	2	6	6

## CLASSIFICATION FACTORIELLE HIÉRARCHIQUE OPTIMISÉE

TABLEAU 13. — Tableau des données.

	Gini	Farm	Rent	Gnpr	Labo	Inst	Ecks	Deat
Argentine	86.3	98.2	32.9	374	25	13.6	57	217
Australie	92.9	99.6	22.3	1215	14	11.3	0	0
Autriche	74	97.4	10.7	532	32	12.8	4	0
Belgique	58.7	85.8	62.3	1015	10	15.5	8	1
Bolivie	93.8	97.7	20	66	72	15.3	53	663
Brésil	83.7	98.5	9.1	262	61	15.5	49	1
Canada	49.7	82.9	7.2	1667	12	11.3	22	0
Chili	93.8	99.7	13.4	180	30	14.2	21	2
Colombie	84.9	98.1	12.1	330	55	14.6	47	316
Costa Rica	88.1	99.1	5.4	307	55	14.6	19	24
Danemark	45.8	79.3	3.5	913	23	14.6	0	0
Rep. Domi.	79.5	98.5	20.8	205	56	11.3	6	31
Equateur	86.4	99.3	14.6	204	53	15.1	41	18
Egypte	74	98.1	11.6	133	64	15.8	45	2
Espagne	78	99.5	43.7	254	50	0	22	1
Etats-Unis	70.5	95.4	20.4	2343	10	12.8	22	0
Finlande	59.9	86.3	2.4	941	46	15.6	4	0
France	58.3	86.1	26	1046	26	16.3	46	1
Guatemala	86	99.7	17	179	68	14.9	45	57
Grece	74.7	99.4	17.7	239	48	15.8	9	2
Honduras	75.5	97.4	16.7	137	66	13.6	45	111
Inde	52.2	86.9	53	72	71	3	83	14
Irak	88.1	99.3	75	195	81	16.2	24	344
Irlande	59.8	85.9	2.5	509	40	14.2	9	0
Italie	80.3	98	23.8	442	29	15.5	51	1
Japon	47	81.5	2.9	240	40	15.7	22	1
Libye	70	93	8.5	90	75	14.8	8	0
Luxembourg	63.8	87.7	18.8	1194	23	12.8	0	0
Nicaragua	75.7	96.4		254	68	12.8	16	116
Norvege	66.9	87.5	7.5	969	26	12.8	4	0
Nle Zelande	77.3	95.5	22.3	1259	16	12.8	0	0
Panama	73.7	95	12.3	350	54	15.6	29	25
Pays-bas	60.5	86.2	53.3	708	11	13.6	2	0
Pérou	87.5	96.9	20	140	60	14.6	23	26
Philippines	56.4	88.2	37.3	201	59	14	15	292
Pologne	45	77.7	0	468	57	8.5	19	5
RFA	67.4	93	5.7	762	14	3	4	0
Royaume Uni	71	93.4	44.5	998	5	13.6	12	0
Salvador	82.8	98.8	15.1	244	63	15.1	9	2
Sud-Vietnam	67.1	94.6	20	133	65	10	50	1000
Suède	57.7	87.2	18.9	1165	13	8.5	0	0
Suisse	49.8	81.5	18.9	1229	10	8.5	0	0
Taiwan	65.2	94.1	40	132	50	0	3	0
Uruguay	81.7	96.6	34.7	569	37	14.6	1	1
Venezuela	90	99.3	20.6	762	42	14.9	36	111
Yougoslavie	43.7	79.8	0	297	67	0	9	0