

REVUE DE STATISTIQUE APPLIQUÉE

A. ZÁRRAGA

B. GOITISOLO

Étude de la structure inter-tableaux à travers l'analyse simultanée

Revue de statistique appliquée, tome 51, n° 3 (2003), p. 39-60

http://www.numdam.org/item?id=RSA_2003__51_3_39_0

© Société française de statistique, 2003, tous droits réservés.

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

NUMDAM

Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques

<http://www.numdam.org/>

ÉTUDE DE LA STRUCTURE INTER-TABLEAUX À TRAVERS L'ANALYSE SIMULTANÉE

A. ZÁRRAGA et B. GOITISOLO

*Departamento de Economía Aplicada III,
Universidad del País Vasco-Euskal Herriko Unibertsitatea, Bilbao, España*

RÉSUMÉ

L'analyse factorielle des relations internes d'un ensemble de tableaux de contingence, chacun d'eux ayant le même ensemble de lignes et un nombre identique ou différent de colonnes, est résolue dans Zárraga & Goitisoló (2002) à travers l'Analyse Simultanée, méthode qui permet de maintenir les caractéristiques et relations d'association propres à chacun des tableaux. Dans le présent travail, on complète l'analyse antérieure par une comparaison globale des ressemblances et des différences entre les tableaux à travers une étude du nuage de points-tableaux. Pour ce faire, on définit, en dehors du point que représente chaque tableau, des mesures de relation entre les points-groupes et entre les facteurs des différentes analyses, mesures qui sont des adaptations de celles utilisées en Analyse Factorielle Multiple (Escofier & Pagès (1988 (1998 3.éd.))).

Mots-clés : *Analyse factorielle, Juxtaposition de Tableaux de Contingence, Analyse des Correspondances, Analyse Factorielle Multiple, Interstructure, Covariance Généralisée.*

ABSTRACT

Factorial analysis of the internal relationships within a set of contingency tables with the same number of rows and the same or a different number of columns is resolved in Zárraga & Goitisoló (2002) by means of Simultaneous Analysis. This method enables the characteristics of each table and its association relationships with the others to be maintained. This work supplements that analysis with an overall comparison of the similarities and differences between the tables, using a graphical representation for the tables. To that end, not only the point representing each table but also measurements of the relationship between the clouds of points and the factors of the various analyses are described. These measurements are suitable transformations of those used by Multiple Factor Analysis (Escofier & Pagès (1988 (1998 3.éd.))).

Keywords : *Factor Analysis, Contingency Tables Juxtaposition, Correspondence Analysis, Multiple Factor Analysis, Interstructure, Generalized Covariance.*

1. Introduction

L'Analyse Simultanée (Zárraga & Goitisoló (2002)) vient du besoin de développer une méthode factorielle pour l'analyse conjointe des tendances générales de plusieurs tableaux de contingence qui permette, de manière similaire à l'analyse des correspondances, l'étude des ressemblances entre l'ensemble des lignes, des colonnes, et les relations entre les deux ensembles, sans permettre que l'un des tableaux ait une prépondérance sur les autres. Les techniques factorielles classiques appliquées à plusieurs tableaux de contingence peuvent se révéler inadaptées à l'étude des relations internes des tableaux, notamment pour les cas dans lesquels l'effectif des tableaux est différent, quand les poids appliqués à une même ligne diffèrent d'un tableau à un autre ou quand les tableaux possèdent des inerties très dissemblables.

L'Analyse Simultanée préserve les caractéristiques propres à chaque tableau (poids, métrique, etc.). On obtient ainsi une analyse conjointe de tableaux de contingence ou de juxtaposition de tableaux de contingence où sont maintenues les relations d'association à l'intérieur de chaque tableau. Dans un travail antérieur, on analyse en détail l'intrastructure¹ des tableaux et l'on vérifie, à travers un exemple, les avantages de cette méthode face à des techniques classiques d'analyse factorielle. Dans le présent travail, on fait au § 2, une brève présentation de la méthode et de l'obtention des axes d'intrastructure.

Outre l'étude conjointe pour tous les tableaux, des relations entre lignes et colonnes de chaque tableau, la structuration des données en groupes ou en tableaux justifie le besoin d'obtention d'autres résultats qui permettent de comparer globalement les groupes. Pour procéder à cette comparaison on propose, au § 3, la représentation de chaque tableau à travers un point qui caractérise la structure du tableau. La projection de ces points-groupes sur les axes d'intrastructure, ainsi que les aides à l'interprétation des projections habituelles dans les analyses factorielles, facilite l'interprétation des ressemblances et des différences entre groupes, c'est-à-dire, l'interprétation de l'interstructure des tableaux.

Au § 4 on suggère de compléter l'analyse des relations entre les tableaux par des mesures de la relation entre les facteurs obtenus dans l'Analyse Simultanée et ceux obtenus dans les analyses séparées de chaque tableau.

Au § 5 on constate l'utilité de ces aides complémentaires à travers l'application à l'Enquête sur la Société de l'Information relative aux nouvelles technologies de l'information et la communication.

2. L'Analyse Simultanée : Intrastructure

Soit $\mathcal{G} = \{1, \dots, g, \dots, G\}$ l'ensemble des tableaux de contingence à analyser. Chacun d'entre eux classe les réponses de k^g individus à deux variables qualitatives codifiées en modalités. Tous les tableaux ont une des variables (dont les modalités, $\mathcal{I} = \{1, \dots, i, \dots, I\}$, se retrouvent dans les lignes) en commun. L'autre variable de chaque tableau de contingence peut être la même ou différente, codifiée de la

¹ On emploie les termes intrastructure, compromis et interstructure suivant en cela la méthodologie STATIS (L'Hermier des Plantes (1976), Lavit (1988)).

même manière ou d'une manière différente. Les modalités de la seconde variable de chaque tableau g sont $\mathcal{J}_g = \{1, \dots, j, \dots, J_g\}$. En juxtaposant tous ces tableaux de contingence, on obtient un ensemble $\mathcal{J} = \{1, \dots, j, \dots, J\}$ de colonnes.

L'élément k_{ij}^g , correspond au nombre total d'individus qui choisissent simultanément les modalités $i \in \mathcal{I}$ de la première variable et $j \in \mathcal{J}_g$ de la seconde variable (appartenant au tableau $g \in \mathcal{G}$).

L'ajustement des effectifs des tableaux est le premier pas pour équilibrer l'influence des tableaux. Pour cela, chacun des tableaux de contingence est transformé, en le divisant par $k_{..}^g$, total du g -ième tableau, $g \in \mathcal{G}$, en un tableau de fréquences relatives.

On notera :

$$\begin{aligned} f_{ij}^g &= \frac{k_{ij}^g}{k_{..}^g} & f_{.j}^g &= \sum_{i \in \mathcal{I}} f_{ij}^g & i \in \mathcal{I} \\ & & & & j \in \mathcal{J} \\ f_{i.}^g &= \sum_{j \in \mathcal{J}_g} f_{ij}^g & f_{..}^g &= \sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{J}_g} f_{ij}^g = 1 & g \in \mathcal{G} \end{aligned}$$

Afin d'équilibrer l'influence que chaque tableau exerce dans l'analyse conjointe et de ne pas permettre que celle-ci soit dominée par un tableau particulier, dans l'Analyse Simultanée on introduira une pondération sur chaque tableau. Cette pondération dépendra des inerties qui résultent d'une analyse des correspondances simples de chacun des tableaux traités séparément. Ces inerties pour le tableau g , $g \in \mathcal{G}$, seront notées λ_s^g (l'inertie projetée sur l'axe s , $s \in \mathcal{S}$) et λ_{TOT}^g (l'inertie totale). Dans l'Analyse Simultanée on considère $\alpha_g = 1/\lambda_1^g$, $g \in \mathcal{G}$, qui est la pondération effectuée dans l'Analyse Factorielle Multiple (AFM).

Afin d'effectuer l'analyse conjointe, on cherche pour chaque ligne un représentant, que nous noterons i^c , $i \in \mathcal{I}$, dit «compromis», qui représente le sous-nuage de profils-lignes formé par tous les points de la même ligne dans les différents tableaux. L'ensemble de tous les représentants sur R^J , muni de la métrique de matrice associée la matrice diagonale des $\alpha^g / f_{.j}^g$, $j \in \mathcal{J}_g \subset \mathcal{J}$, forme le nuage $\mathcal{N}(\mathcal{I}^c)$.

On définit le compromis i^c avec le poids p_{i^c} , $i \in \mathcal{I}$:

$$i^c = \sum_{g \in \mathcal{G}} \frac{\sqrt{f_{i.}^g}}{\sum_{g \in \mathcal{G}} \sqrt{f_{i.}^g}} i^g \quad p_{i^c} = \left(\sum_{g \in \mathcal{G}} \sqrt{f_{i.}^g} \right)^2$$

où i^g dénote le profil correspondant à la ligne i , $i \in \mathcal{I}$, du tableau g , $g \in \mathcal{G}$, représenté dans l'espace R^J dont les coordonnées sont :

$$\begin{cases} \left(\begin{array}{c} \frac{f_{ij}^g}{f_{i.}^g} - f_{.j}^g \\ \end{array} \right) & \text{si } j \in \mathcal{J}_g \\ 0 & \text{sinon} \end{cases} \quad (1)$$

2.1. Obtention des facteurs

Afin de chercher la relation entre les analyses des lignes et des colonnes dans les développements suivants, l'analyse factorielle de l'ensemble des tableaux de contingence s'effectue, comme il est dit dans Zárraga & Goitisoló (2002), en recherchant les valeurs propres (λ_s) et les vecteurs propres (u_s), $s \in \mathcal{S}$, issus de la diagonalisation de la matrice $A = X^T X$ où le terme général de la matrice X est :

$$x_{ij} = \sqrt{\alpha^g} \sqrt{f_{i.}^g} \left(\frac{f_{ij}^g}{f_{i.}^g f_{.j}^g} - 1 \right) \sqrt{f_{.j}^g} \quad \begin{array}{l} i \in \mathcal{I} \\ j \in \mathcal{J} \\ g \in \mathcal{G} \end{array} \quad (2)$$

Cette matrice peut être exprimée comme juxtaposition des matrices X_g , d'ordre $(I \times J_g)$, qui se composent des coordonnées de la matrice X pour l'ensemble des colonnes J_g du tableau g , $g \in \mathcal{G}$:

$$X = [X_1, \dots, X_g, \dots, X_G] \quad (3)$$

De plus, on utilisera les matrices diagonales R d'ordre $I \times I$ et N d'ordre $J \times J$, de termes généraux respectivement :

$$r_{ii} = p_{i^c} \quad n_{jj} = f_{.j}^g \quad \begin{array}{l} i \in \mathcal{I} \\ j \in \mathcal{J}_g \subset \mathcal{J} \\ g \in \mathcal{G} \end{array}$$

On définit aussi la matrice Y , diagonale par blocs, où chaque bloc de la diagonale est la matrice X_g et on définit la matrice Q , diagonale par blocs aussi, où chaque bloc est, de même, une matrice diagonale d'ordre $(I \times I)$ et de terme général $f_{i.}^g$, $i \in \mathcal{I}$, $g \in \mathcal{G}$.

2.2. Relation entre les vecteurs propres des deux analyses

Puisque dans l'analyse des lignes on diagonalise la matrice $X^T X$ et dans celle des colonnes la matrice $X X^T$ la relation entre les valeurs propres et vecteurs directeurs des analyses des lignes (λ_s et $u_s \in R^J$) et colonnes (μ_s et $v_s \in R^I$) est :

$$\lambda_s = \mu_s \quad s \in \mathcal{S} \quad (4)$$

$$u_s = \lambda_s^{-1/2} X^T v_s \quad s \in \mathcal{S}$$

$$v_s = \lambda_s^{-1/2} X u_s \quad s \in \mathcal{S} \quad (5)$$

2.3. Projections des profils

On calcule les projections sur l'axe s , $s \in \mathcal{S}$, des profils-lignes partiels $F_s(i^g)$, des compromis $F_s(i^c)$, $i \in \mathcal{I}$, et des profils-colonnes $G_s(j)$, $j \in \mathcal{J}$. Les projections

de tous les profils-lignes partiels et compromis sont respectivement $F_s(\mathcal{I}^g)$ avec $\mathcal{I}^g = \{\mathcal{I}^g/g \in \mathcal{G}\}$ et $F_s(\mathcal{I}^c)$ et de toutes les colonnes $G_s(\mathcal{J})$. Les projections sur l'axe s , $s \in \mathcal{S}$, des lignes et colonnes sont calculées en sachant qu'il est nécessaire d'éliminer l'effet de l'introduction des poids et les profils-lignes partiels sont projetés comme supplémentaires en éliminant aussi l'effet de l'introduction du poids :

$$F_s(\mathcal{I}^c) = R^{-1/2} X u_s = \lambda_s^{1/2} R^{-1/2} v_s \quad (6)$$

$$G_s(\mathcal{J}) = N^{-1/2} X^T v_s = \lambda_s^{1/2} N^{-1/2} u_s \quad (7)$$

$$F_s(\mathcal{I}^g) = Q^{-1/2} Y u_s \quad (8)$$

3. Comparaison des groupes ou tableaux : Interstructure

Jusqu'à maintenant on a réalisé une étude simultanée des tableaux à partir des projections des profils-lignes partiels et des profils-colonnes, ainsi que des points compromis. Quand le nombre de tableaux est important, l'analyse peut être complétée par une visualisation où chaque point représente l'un des tableaux.

Dans certains cas (Cazes (1982), Torre & Chessel (1995)) les projections de chaque tableau correspondent au centre de gravité des projections des lignes de chaque tableau sur les axes obtenus. Dans d'autres (Qannari, Qannari & Courcoux (1992), Escofier & Pagès (1983), Chessel & Hanafi (1996), Simier, Blanc, Pellegrin & Nandris (1999) et L'Hermier des Plantes (1976) entre autres) on se livre à une analyse globale des ressemblances et différences des tableaux en recherchant des axes sur la base de différents critères.

Ci-après, on recherche un point qui représenterait chacun des tableaux et des axes, en utilisant, comme critère, la maximisation de la somme des projections de ces points sur les axes.

3.1. Représentation de chaque tableau (Analyse dans R^{I^2})

Le nuage de colonnes du tableau g , $g \in \mathcal{G}$, une fois ses profils transformés pour pouvoir être représentés dans l'espace R^I doté de la métrique euclidienne, est caractérisé en utilisant la matrice d'inertie que nous dénoterons par W_g :

$$W_g = X_g X_g^T$$

où X_g est la matrice définie au § 2.1.

La diagonalisation de la matrice W_g permet d'obtenir une reconstruction parfaite de la structure du nuage de profils-colonnes du tableau g d'où l'on déduit que la matrice W_g est une bonne représentation du tableau correspondant.

La matrice W_g , d'ordre $(I \times I)$, peut être représentée comme un point dans un espace de dimension I^2 doté de la métrique euclidienne. Les coordonnées de ce point, représenté par le vecteur d'ordre $(1 \times I^2)$ que nous dénoterons \vec{W}_g (Figure 1), s'obtiennent en juxtaposant les lignes de la matrice W_g .

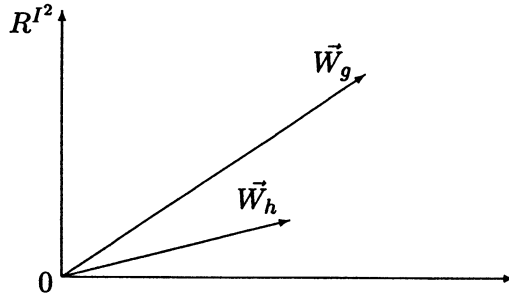


FIGURE 1
Représentation des tableaux en R^{I^2}

3.2. Mesure de la dimensionnalité d'un tableau

On peut se livrer à une première comparaison des tableaux en utilisant comme mesure de la dimensionnalité des tableaux la norme de \vec{W}_g ou distance au carré du point \vec{W}_g à l'origine. Nous dénoterons cette mesure par ² par $\mathcal{L}_g(g)$:

$$\mathcal{L}_g(g) = d^2(\vec{W}_g, 0) = \text{tr}(W_g W_g) \quad (9)$$

où $\text{tr}(A)$ est la trace de la matrice A .

Quand la pondération appliquée aux tableaux est l'inverse de la première valeur propre des analyses de correspondances séparées ($\alpha_g = 1/\lambda_1^g$), cette distance au carré peut être exprimée en fonction des valeurs propres, λ_s^g , de l'analyse de correspondances séparée du tableau g , $g \in \mathcal{G}$:

$$\mathcal{L}_g(g) = \sum_s \left(\frac{\lambda_s^g}{\lambda_1^g} \right)^2 \quad (10)$$

Cette mesure dépend, par conséquent, de la structure du groupe ; elle sera d'autant plus grande que le groupe sera plus multidimensionnel, étant entendu que multidimensionnalité est non le nombre total de dimensions mais le fait qu'à chaque dimension il est assigné une importance en fonction de son inertie. Cette valeur sera proche de 1 s'il existe un premier axe prépondérant dans l'analyse et elle approchera le nombre de dimensions total si chacun des axes possède une inertie semblable.

Pour vérifier l'égalité (10) on dénote par V_g la matrice qui contient en colonnes les vecteurs propres de l'analyse de correspondances des profils-colonnes du tableau g et par Λ_g la matrice diagonale d'inerties projetées associées à cette analyse. Une fois qu'on a introduit la pondération $\alpha_g = 1/\lambda_1^g$, ces inerties projetées sont $(\lambda_s^g/\lambda_1^g)$.

² Par sa similitude avec la mesure utilisée dans l'Analyse Factorielle Multiple.

Les distances au carré de chaque point-tableau à l'origine en R^{I^2} peuvent être obtenues en partant des relations dans l'analyse du tableau g :

$$\begin{aligned} X_g X_g^T V_g &= V_g \Lambda_g \\ W_g V_g &= V_g \Lambda_g \end{aligned}$$

En post-multipliant les matrices de chaque membre de l'équation par leurs transposées :

$$\begin{aligned} W_g V_g V_g^T W_g^T &= V_g \Lambda_g \Lambda_g^T V_g^T \\ W_g W_g &= V_g \Lambda_g^2 V_g^T \end{aligned}$$

étant donné que V_g est une matrice orthogonale comme matrice de vecteurs propres.

En post-multipliant par V_g et en simplifiant le résultat :

$$\begin{aligned} W_g W_g V_g &= V_g \Lambda_g^2 V_g^T V_g = \\ &= V_g \Lambda_g^2 \end{aligned}$$

En conséquence, les vecteurs propres, contenus en V_g , de l'analyse de profils-colonnes du tableau g sont également vecteurs propres de $W_g W_g$ mais associés aux valeurs propres $(\lambda_s^g / \lambda_1^g)^2$.

3.3. Projection des groupes

L'étude de l'affinité entre les groupes ou tableaux pourrait être opérée à travers une analyse factorielle du nuage de points – tous de poids unitaire –, chacun d'entre eux représentant un tableau, dans l'espace de dimension I^2 . La représentation dans un espace de petite dimension permettrait de comparer globalement les tableaux, en décidant lesquels sont similaires et lesquels sont les tableaux qui se différencient le plus du reste. Cependant, il est difficile par cette manière de procéder d'interpréter la cause de ces différences entre les tableaux, étant donné que les axes sont fonction des données initiales, de façon complexe.

Pour obtenir des axes facilement interprétables en fonction des variables initiales, serait-ce aux dépens d'une représentation moins précise des relations entre les tableaux, on impose aux axes (définis à travers le vecteur \vec{z}_s d'ordre $(1 \times I^2)$) d'être des éléments symétriques de rang 1 ($z_s z_s^T$). La démarche adoptée pour représenter les groupes est, en réalité, celle de l'Analyse Factorielle Multiple.

Nous dénotons par $F_s(g)$ les projections des points-groupes (\vec{W}_g) sur le vecteur \vec{z}_s . Dans la mesure où ces projections sont toujours positives, le critère utilisé pour l'obtention des axes consiste à maximiser, sous la contrainte de l'orthonormalité des axes, la somme des projections. Critère plus facile à obtenir, et dans ce cas interprétable, que la maximisation de la somme des carrés des projections.

$$\text{Max} \sum_{g \in \mathcal{G}} F_s(g) = \text{Max} \sum_{g \in \mathcal{G}} \text{Projection de } \vec{W}_g \text{ sur } \vec{z}_s$$

Pour obtenir les axes \vec{z}_s qui maximisent la fonction antérieure nous commencerons par développer la projection de \vec{W}_g sur \vec{z}_s :

$$\begin{aligned} F_s(g) &= \text{Projection de } \vec{W}_g \text{ sur } \vec{z}_s = \\ &= \vec{W}_g \vec{z}_s^T = \\ &= \text{tr}(W_g z_s z_s^T) = \\ &= z_s^T W_g z_s = \\ &= z_s^T X_g N_g^{-1/2} N_g N_g^{-1/2} X_g^T z_s \end{aligned}$$

où N_g est la matrice diagonale d'ordre $(J_g \times J_g)$ qui contient les poids des profils-colonnes du tableau g . En usant de l'équivalence avec la relation (7), ceci peut s'exprimer comme :

$$\begin{aligned} F_s(g) &= \sum_{j \in \mathcal{J}_g} f_{.j}^g (\text{Projection de la modalité } j \text{ sur } z_s)^2 = \\ &= \sum_{j \in \mathcal{J}_g} (\text{Inertie projetée de la modalité } j \text{ sur } z_s) \end{aligned} \quad (11)$$

En faisant la somme dans l'ensemble des groupes :

$$\text{Max} \sum_{g \in \mathcal{G}} F_s(g) = \text{Max} \sum_{g \in \mathcal{G}} \sum_{j \in \mathcal{J}_g} (\text{Inertie projetée de la modalité } j \text{ sur } z_s)$$

Mais cette dernière expression est, en réalité, le critère utilisé dans l'obtention des axes de l'analyse des profils-colonnes, et par conséquent les vecteurs \vec{z}_s obtenus dans l'analyse du nuage de points-groupes peuvent être calculés à partir des axes v_s obtenus dans l'Analyse Simultanée des profils-colonnes de tous les tableaux. Autrement dit, le vecteur ligne de dimension I^2 (\vec{z}_s) est celui résultant de la vectorisation de la matrice $z_s z_s^T = v_s v_s^T$.

De la sorte,

$$F_s(g) = \sum_{j \in \mathcal{J}_g} f_{.j}^g G_s^2(j) = \text{Inertie}_s(g)$$

où $\text{Inertie}_s(g)$ représente l'inertie projetée de l'ensemble de colonnes du groupe g sur l'axe s .

Rappelant que, dans l'analyse séparée du groupe g , l'inertie maximum projetée sur un axe correspond à sa première valeur propre (λ_1^g) et que, après la pondération appliquée aux groupes, l'inertie maximum projetée sur n'importe lequel des axes est l'unité. La coordonnée d'un groupe peut donc également être interprétée comme la proportion de l'inertie maximale possible en projection sur l'axe s de l'ensemble des colonnes du groupe g . Une valeur proche de 1 indique que l'axe de l'analyse conjointe est pratiquement le même que dans l'analyse séparée alors qu'une valeur proche de

0 indique orthogonalité entre les premiers axes des analyses séparées par rapport à l'Analyse Simultanée.

En faisant la somme des projections de tous les groupes on obtient l'inertie projetée sur l'axe :

$$\sum_{g \in \mathcal{G}} F_s(g) = \sum_{g \in \mathcal{G}} \text{Inertie}_s(g) = \lambda_s$$

La projection de chaque groupe sur le premier axe étant inférieure ou égale à l'unité, l'inertie projetée sur lui est comprise entre zéro et G , le nombre de groupes.

Aides à l'interprétation des groupes en R^{I^2}

La contribution de chaque groupe à la formation d'un axe s , $s \in \mathcal{S}$, et sa qualité de représentation sur celui-ci, peut se calculer, respectivement, comme :

$$\begin{aligned} \text{cta}_s(g) &= \frac{F_s(g)}{\lambda_s} \\ \text{ctr}_s(g) &= \frac{F_s^2(g)}{d^2(g, 0)} \end{aligned}$$

$d^2(g, 0)$ étant la distance du point-groupe à l'origine obtenue en (9), où elle est dénotée par $\mathcal{L}_g(g)$.

3.4. Mesure de la relation entre les groupes

On peut utiliser le produit scalaire entre deux points-groupes comme mesure de la relation entre les groupes. Cette mesure sera dénotée par $\mathcal{L}(g, h)$:

$$\mathcal{L}(g, h) = \langle \vec{W}_g, \vec{W}_h \rangle \quad g, h \in \mathcal{G}$$

Pour justifier pourquoi ce produit scalaire est une bonne mesure de la relation entre deux points-groupes nous définissons la relation entre un groupe g et un vecteur z_s , et nous la dénotons par $\mathcal{L}(g, z_s)$, comme la projection du point-groupe sur le vecteur \vec{z}_s dans l'espace de dimension I^2 , $F_s(g)$. En (11) nous avons vu que cette projection peut s'exprimer en fonction des inerties projetées par les colonnes :

$$\mathcal{L}(g, z_s) = F_s(g) = \sum_{j \in \mathcal{J}_g} (\text{Inertie projetée de la modalité } j \text{ sur } z_s)$$

De plus, la matrice d'inertie du nuage de profils-colonnes, pondérés par $\sqrt{\alpha_h}$, du tableau h peut s'exprimer comme :

$$W_h = X_h X_h^T = V_h \Lambda_h V_h^T = \sum_s \lambda_s^h v_s^h v_s^{hT}$$

λ_s^h et v_s^h étant les valeurs et vecteurs propres (normalisés pour la métrique euclidienne usuelle), d'ordre s , obtenus dans l'analyse de correspondances séparée des profils-colonnes, pondérés, du groupe h et où Λ_h et V_h sont les matrices qui les contiennent.

En d'autres termes :

$$\begin{aligned} \mathcal{L}(g, h) &= \langle \vec{W}_g, \vec{W}_h \rangle = \\ &= \sum_s \lambda_s^h v_s^{hT} W_g v_s^h = \\ &= \sum_s \left(\lambda_s^h \sum_{j \in \mathcal{J}_g} \text{Inertie projetée de la modalité } j \text{ sur } v_s^h \right) = \\ &= \sum_s \lambda_s^h \mathcal{L}(g, v_s^h) \qquad g, h \in \mathcal{G} \end{aligned}$$

Cette mesure, toujours positive, sera d'autant plus grande que chacune des modalités d'un groupe g est plus reliée aux vecteurs propres de l'autre groupe h , cette relation étant pondérée par la valeur propre de l'axe correspondant. Les vecteurs propres étant les variables qui synthétisent les relations entre les variables de chaque groupe, le produit scalaire entre deux points-groupes mesure la relation de toutes les modalités de l'un des groupes avec l'ensemble de modalités de l'autre groupe.

4. Autres aides à l'interprétation : Relation entre facteurs des différentes analyses

Pour mesurer le degré de similitude de chacun des facteurs dans l'Analyse Simultanée avec les facteurs des analyses séparées de chaque tableau et de ces derniers entre eux, on peut envisager de calculer leur coefficient de corrélation. Cette manière de procéder est habituelle quand les facteurs proviennent de l'analyse de tableaux où les lignes représentent des individus, par exemple, Torre & Chessel (1995) et Escofier & Pagès (1988 (1998 3.éd.)) ou des groupes d'individus auxquels on associe le même poids dans les différents tableaux (Bastin, Benzécri, Bourgarit & Cazes (1980), Cazes (1982)). Cependant, il nous faut rappeler que dans ce travail on considère G tableaux et à chaque ligne du tableau général est associé un groupe de G profils-lignes différents, un par tableau. Cazes (1982) propose de calculer le coefficient de corrélation entre facteurs en prenant pour poids des lignes, les poids issus de l'analyse des correspondances d'un de ces tableaux, ces poids dépendant donc du choix du tableau de référence retenu pour assigner les poids. Nous proposons de résoudre ce problème des poids en élargissant le concept de covariance généralisée (Méot & Leclerc (1997)) à celui de corrélation généralisée.

Bien que le nombre d'axes dans chacune des analyses séparées et dans l'Analyse Simultanée puisse être différent, nous dénoterons par s un axe général de n'importe laquelle d'entre elles. Les mesures proposées ne pourront être calculées que pour des axes dont l'inertie projetée sera différente de zéro.

• **Relation entre facteurs des analyses séparées**

La relation entre les facteurs s et s' des groupes g et h , respectivement, se calculerait comme :

$$r(F_s^g(\mathcal{I}), F_{s'}^h(\mathcal{I})) = \sum_{i \in \mathcal{I}} \sqrt{f_i^g} \frac{F_s^g(i)}{\sqrt{\lambda_s^g}} \sqrt{f_i^h} \frac{F_{s'}^h(i)}{\sqrt{\lambda_{s'}^h}} \quad g, h \in \mathcal{G} \quad (12)$$

où $F_s^g(i)$ et $F_{s'}^h(i)$ sont les projections sur les axes s et s' des analyses séparées dans les groupes g et h respectivement et où λ_s^g et $\lambda_{s'}^h$ sont les inerties associées à ces axes.

Cette mesure permet de vérifier si les facteurs des analyses séparées sont semblables et les possibles rotations qui se produisent.

Comparer les projections entre les groupes g et h équivaut à étudier le cosinus entre le vecteur directeur de l'analyse séparée du groupe g (v_s^g) et le vecteur associé au groupe h ($v_{s'}^h$). Sachant que ces vecteurs sont représentés dans des espaces dotés de métriques différentes, il sera nécessaire (de même qu'il est justifié en Zárrega & Goitiso (2002)) de rechercher un espace commun. Pour ce faire, on les représente dans un espace doté de la métrique euclidienne dans lequel il est possible de calculer leur cosinus. Ce cosinus coïncide avec l'expression (12).

• **Relation entre facteurs de l'Analyse Simultanée et facteurs des analyses séparées**

De manière similaire, on peut calculer pour chaque facteur s de l'Analyse Simultanée, la relation avec chacun des facteurs s' des analyses séparées des différents groupes :

$$r(F_{s'}^g(\mathcal{I}), F_s(\mathcal{I}^c)) = \sum_{i \in \mathcal{I}} \sqrt{f_i^g} \frac{F_{s'}^g(i)}{\sqrt{\lambda_{s'}^g}} \left(\sum_{g \in \mathcal{G}} \sqrt{f_i^g} \right) \frac{F_s(i^c)}{\sqrt{\lambda_s}} \quad g \in \mathcal{G} \quad (13)$$

Si chacun des tableaux de fréquences analysés possède la même marginale colonne cette mesure se réduit, puisqu'on a alors $p_{i^c} = G^2 f_i^g$ et compte tenu de (6), à :

$$\begin{aligned} r(F_{s'}^g(\mathcal{I}), F_s(\mathcal{I}^c)) &= \sum_{i \in \mathcal{I}} \frac{\sqrt{f_i^g} F_{s'}^g(i)}{\sqrt{\sum_{i \in \mathcal{I}} f_i^g (F_{s'}^g(i))^2}} \frac{G \sqrt{f_i^g} F_s(i^c)}{\sqrt{\sum_{i \in \mathcal{I}} G^2 f_i^g (F_s(i^c))^2}} = \\ &= \sum_{i \in \mathcal{I}} \frac{f_i^g F_{s'}^g(i) F_s(i^c)}{\sqrt{\sum_{i \in \mathcal{I}} f_i^g (F_{s'}^g(i))^2} \sqrt{\sum_{i \in \mathcal{I}} f_i^g (F_s(i^c))^2}} \end{aligned}$$

c'est-à-dire, le coefficient de corrélation entre les facteurs des analyses séparées et les facteurs de l'Analyse Simultanée.

• **Relation entre le nuage compromis et les nuages de profils-lignes partiels**

Pour vérifier si un groupe est bien représenté sur chacun des facteurs de l'Analyse Simultanée on peut calculer la corrélation généralisée entre les projections de chaque sous-nuage de profils-lignes partiels avec les projections des compromis.

$$r(F_s(\mathcal{I}^g), F_s(\mathcal{I}^c)) = \sum_{i \in \mathcal{I}} \frac{\sqrt{f_{i.}^g} F_s(i^g)}{\sqrt{\sum_{i \in \mathcal{I}} f_{i.}^g (F_s(i^g))^2}} \left(\sum_{g \in \mathcal{G}} \sqrt{f_{i.}^g} \right) \frac{F_s(i^c)}{\sqrt{\lambda_s}} \quad g \in \mathcal{G}$$

On notera que dans les analyses factorielles les facteurs sont déterminés au signe près. Donc, dans toutes les mesures de liaisons précédentes on ne peut interpréter que la valeur absolue et pas le signe.

5. Application de l'Analyse Simultanée Proposée

Ci-après, on se livre à une Analyse Simultanée³ sur des données obtenues de l'Enquête sur la Société de l'Information -ESI- Familles correspondant au quatrième trimestre 2000. Réalisée à intervalle semestriel par l'Institut Basque de la Statistique -EUSTAT-,⁴ cette enquête se propose de recueillir de l'information sur les nouvelles technologies de l'information et de la communication dans la population des 15 ans et plus résidant dans la Communauté Autonome du Pays Basque. Outre l'information relative à l'enquête proprement dite on dispose, pour chacun des individus, de variables sociodémographiques extraites de l'Enquête démographique en relation avec l'activité également fournie par EUSTAT sur la même période.

Pour l'application qui se présente, les individus ont été classés, conformément à leur âge, en 5 groupes : 15-17 ans, 18-24, 25-34, 35-44 et plus de 44 ans. De plus, de l'Enquête sur la Société de l'Information on a sélectionné les variables relatives à la fréquence d'utilisation de quatre services techniques d'Internet (tableau 1), ainsi que les thèmes sondés, pour lesquels les individus montrent un intérêt ou non lorsqu'ils vont sur Internet (tableau 2). Compte tenu du faible effectif des individus qui déclarent ne jamais se connecter sur les pages web (modalité 5 de la variable UWE), cette dernière a été jointe à la modalité 4 pour cette variable.

Dans les graphiques, chacune des modalités sera identifiée par le libellé de la variable, suivie du numéro de la modalité. Dans le cas des profils-lignes partiels et des profils-colonnes, cette notation sera précédée du numéro indicateur du groupe auquel appartient ce même profil. Par exemple, 3UCE2 fait référence à l'utilisation du courrier électronique plusieurs fois par mois chez les individus du groupe 3 (entre 25 et 34 ans). Pour faciliter la localisation des modalités mentionnées dans l'interprétation on a, compte tenu du grand nombre de projections, augmenté la taille des libellés des modalités qui contribuent le plus à la création des axes et représenté sur des

³ Cette application a été extraite de Goitiso (2002), où l'on peut consulter plus de détails sur l'application et la méthode.

⁴ L'Eustat est un organisme public du Pays Basque auprès duquel il est possible d'obtenir plus d'information à l'adresse «www.eustat.es».

TABLEAU 1
Variables d'Utilisation des Services Techniques d'Internet

Fréquence d'Utilisation des Services Techniques d'Internet		
Variables	UWE	Utilisation du Web, consultations
	UCE	Utilisation de courrier électronique
	UTF	Utilisation de transfert de fichiers (FTP)
	UCC	Utilisation de « chats », conversations en direct entre connectés
Modalités	1	Plusieurs fois par semaine
	2	Plusieurs fois par mois
	3	Avec moins de fréquence
	4	Quasiment jamais
	5	Jamais

TABLEAU 2
Variables Thèmes jugés d'intérêt dans l'Accès à Internet

Thèmes jugés d'intérêt dans l'Accès à Internet		
Variables	TMU	Musées virtuels
	TPO	Contacts et débats politiques
	TED	Télé-enseignement
	TME	Télé-médecine, télé-diagnostic
	TSA	Services administratifs et citoyens
	TVI	Préparation de voyages
	TNO	Journaux électroniques, informations
	TBE	Recherche d'emploi
	TDC	Droits des consommateurs
	TFI	Gestion bancaire, de la bourse et financière
	TBI	Bibliothèque, documentation
	TJU	Jeux, hobbies
Modalités	1	Intéressé
	2	Pas intéressé

graphiques différents les projections des compromis et des profils-colonnes. On a aussi représenté ce derniers sur des graphiques différents, en fonction de la variable à laquelle ils sont associés.

5.1. Analyse de l'Infrastructure

L'Analyse Simultanée met en évidence l'excessive importance, dans la détermination des axes, des modalités de transfert de fichiers -UTF- pour les plus jeunes (15-17 ans) et celles correspondant à l'usage de «chats» -UCC- dans les groupes les plus âgés (35-44 et plus de 44 ans). L'analyse des poids de ces modalités révèle la raison d'être de cette influence, uniquement 31 % des plus jeunes affirment utiliser à certaines occasions le transfert de fichiers (modalités UTF1 à UTF4) et il en va à peu près de même parmi les groupes plus âgés pour l'emploi des «chats», 79 % des utilisateurs d'Internet de 35 à 44 ans répondent ne jamais utiliser Internet pour converser (ce pourcentage s'élève à 80 % pour les personnes de plus de 44 ans). On décide, pour cette raison, d'éliminer ces variables et les modalités associées pour les groupes d'âge correspondants.

Ces premières précisions étant fournies, nous allons analyser en détail le tableau reproduit dans le tableau 3 dans lequel k_{ij}^g représente le nombre d'individus appartenant au tableau d'âge g qui répondent à la modalité i sur l'intérêt pour les thèmes et j sur l'emploi de services techniques d'Internet.

TABLEAU 3
Tableau analysé

	15-17			18-24				25-34				35-44			45 ou +		
	UWE1	UCE3	UCC5	UWE1	UCE3	UTF3	UCC5	UWE1	UCE3	UTF3	UCC5	UWE1	UCE3	UTF5	UWE1	UCE3	UTF5
TMU1	k_{11}^1							k_{ij}^g									
TMU2																	
⋮																	
TJU2																	

L'inertie projetée (tableau 4) sur le premier axe de l'Analyse Simultanée du tableau précédent (4.323, 45.096 % de l'inertie totale) indique que, même s'il n'atteint pas sa valeur maximum (le nombre de groupes qui est cinq), il s'agit d'un axe d'inertie important dans les cinq groupes analysés.

Ci-après, nous limiterons l'interprétation au premier plan factoriel dans lequel est représenté 56.046 % de l'inertie totale. A grands traits, on peut dire que le premier facteur manifeste l'intérêt ou le désintérêt pour tous les thèmes figurant dans l'enquête. L'intérêt pour les journaux électroniques et les informations -TNO-, pour la gestion bancaire, de la bourse et des services financiers -TFI-, pour les bibliothèques et documentation -TBI- et pour les services administratifs et citoyens -TSA- étant de

TABLEAU 4
Inerties des 5 premiers axes

	Valeur Propre	Pourcent.	Pourcent. Cumule
1	4.3234	45.0958	45.0958
2	1.0498	10.9499	56.0457
3	0.8470	8.8349	64.8806
4	0.6311	6.5826	71.4631
5	0.5763	6.0115	77.4746

la plus haute importance pour la détermination de ce premier facteur. Les modalités indicatrices du désintérêt pour les thèmes cités se projettent dans la partie positive du premier axe et celles qui représentent de l'intérêt, dans la partie négative de l'axe (figure 2).

Le recours aux consultations sur le Web -UWE-, à 39.4 %, ainsi que l'utilisation du transfert de fichiers -UTF-, à 28.96 %, est ce qui apparaît de plus déterminant dans ce facteur. Dans la figure 3, et plus clairement dans la figure 4, où le plan 1-2 est éclaté par services techniques, on observe comment les positions extrêmes (plusieurs fois par semaine et jamais), dans l'utilisation de tous les services techniques d'Internet consultés dans l'enquête (web et consultations -UWE-, courrier électronique -UCE-, transfert de fichiers (FTP) -UTF- et «chats», conversations -UCC-) sont toujours associées. Du côté négatif de l'axe se positionnent les modalités qui indiquent l'utilisation plusieurs fois par semaine des services -modalités 1- et dans la partie positive celles qui montrent leur absence totale (ou pratiquement dans le cas du web, consultations) d'utilisation.

Par conséquent, on note qu'il existe un lien entre l'emploi fréquent de tous les services techniques d'Internet et l'intérêt généralisé pour les thèmes ; sachant que les enquêtés les plus attirés par la gestion bancaire, de la bourse et des services financiers -TFI1- ainsi que par les services administratifs et citoyens -TSA1- sont aussi ceux qui utilisent Internet le plus intensément. De même, les enquêtés qui ne manifestent pas d'intérêt pour les journaux électroniques et les informations -TNO2-, ou pour les bibliothèques et la documentation -TBI2-, sont ceux qui emploient le moins Internet.

Le second facteur distingue, d'une part, l'intérêt ou le désintérêt pour des thèmes déterminés comme les droits des consommateurs -TDC- avec une contribution de 29.10 %, la recherche d'emploi -TBE- avec une contribution de 23.57 % et les journaux électroniques et informations -TNO- et les bibliothèques et documentation -TBI-, avec des contributions de 15.71 % et 9.57 % respectivement et, d'autre part, la fréquence de l'utilisation du Web et des consultations -UWE- et du courrier électronique -UCE- (avec 44.09 % et 33.51 % respectivement de l'inertie du nuage de modalités projetée sur le second axe).

Analyse Simultanée par groupes d'âge

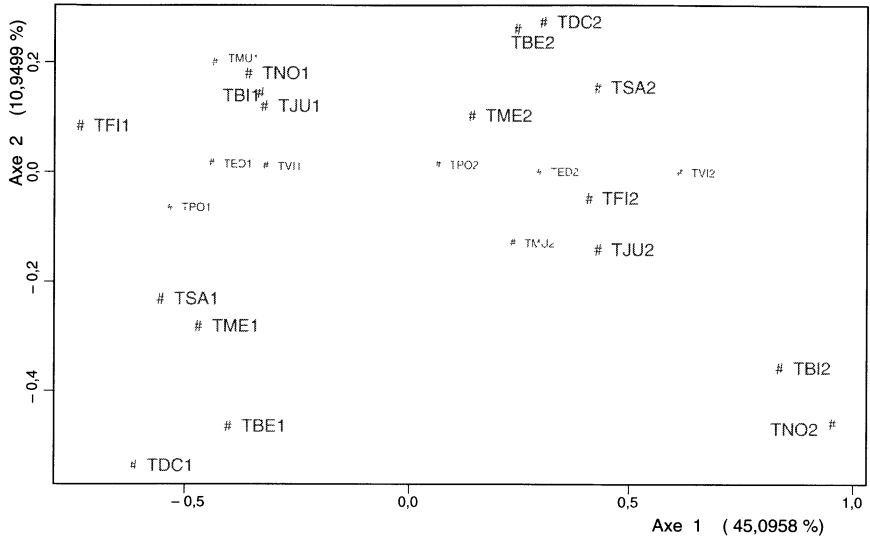


FIGURE 2
Projection des compromis

Analyse Simultanée par groupes d'âge

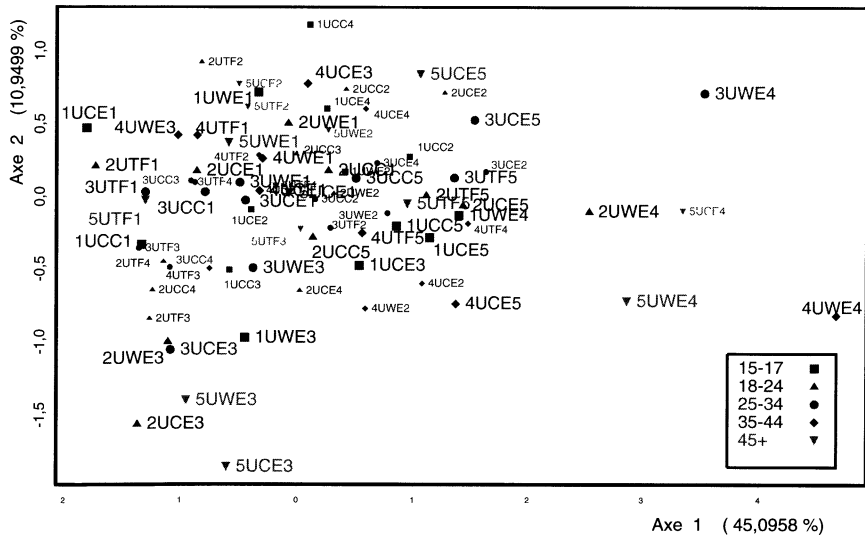


FIGURE 3
Projection des profils-colonnes

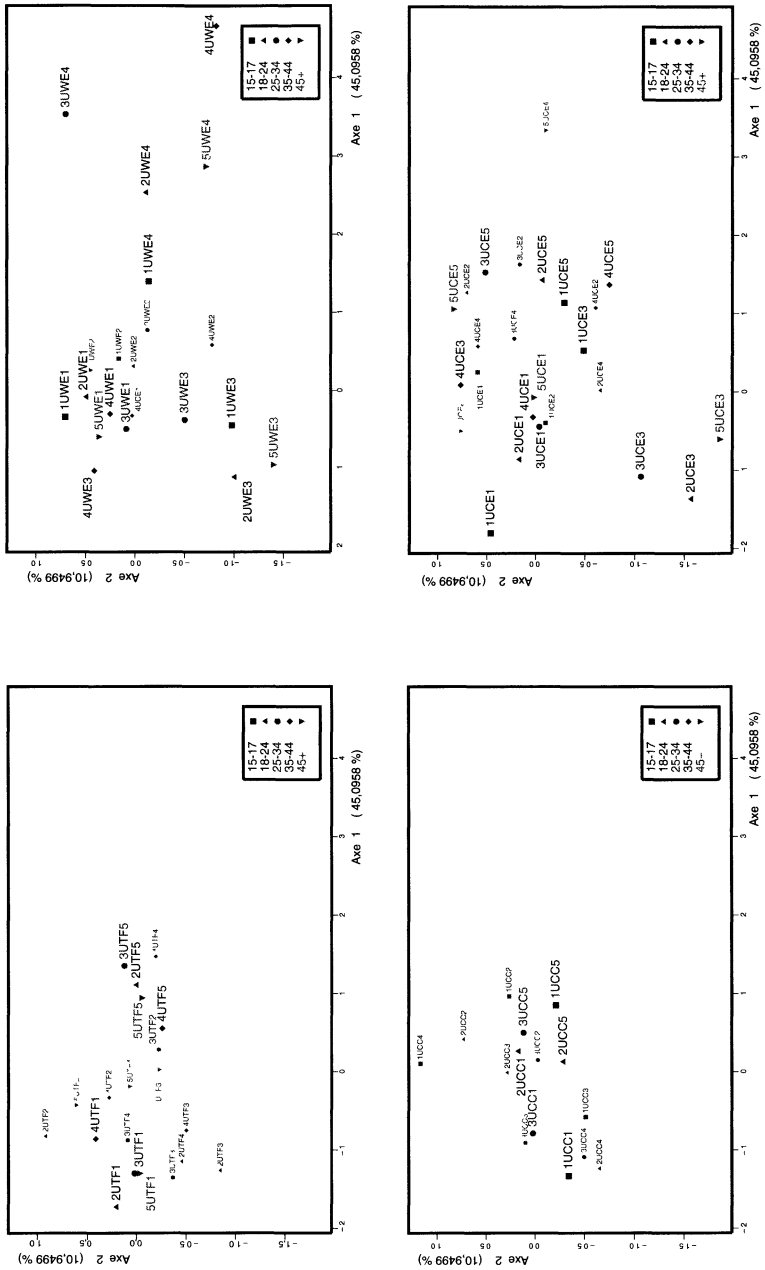


FIGURE 4

Plans factoriels 1-2 pour utilisation de services techniques d'Internet

On observe, par exemple, comment l'intérêt pour les journaux électroniques et les informations -TNO1-, pour les bibliothèques et documentation -TBI1- et pour les jeux et hobbies -TJU1- est lié aux personnes, de tous âges, qui consultent le web plusieurs fois par semaine -UWE1-, tandis que celles qui ne se sentent pas attirées par les journaux électroniques et les nouvelles sur Internet -TNO2-, pas plus que par les bibliothèques et la documentation -TBI2-, n'utilisent non plus jamais ou pratiquement jamais Internet pour des consultations -UWE4-. Jamais le font-elles pour le courrier électronique -UCE5-, pour le transfert de fichiers -UTF5- ou encore pour le «chat» et converser -UCC5-. L'intérêt pour la recherche d'emploi -TBE1- et pour les droits des consommateurs -TDC1- se manifeste chez les personnes qui utilisent de manière sporadique le web -UWE3- et le courrier électronique -UCE3-.

5.2. Analyse de l'Interstructure

Dans l'étude de l'interstructure l'objectif est de comparer globalement les cinq tableaux ou groupes d'âge, ce qui peut faire l'objet de mise en évidence à l'aide d'un graphique sur lequel les groupes sont représentés chacun par un point. Ces projections, par le fait de la pondération retenue pour les groupes, vont toujours se trouver, dans un plan, dans un carré de côté 1 (figure 5).

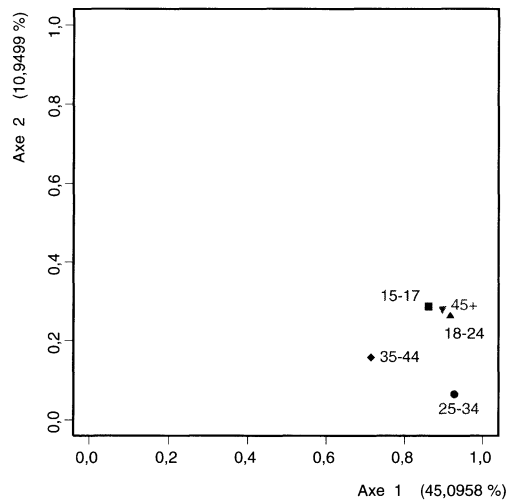


FIGURE 5
Projection des groupes sur les deux premiers axes

Comme inertie projetée du groupe, la coordonnée sur un axe indique l'importance du groupe dans la détermination de l'axe. Ainsi, la proximité des cinq projections à la valeur 1 indique que la première direction de l'Analyse Simultanée est une direction d'inertie importante de chacun des cinq nuages considérés dans l'application.

La faible coordonnée sur le second axe pourrait donner à penser qu'il s'agit d'une direction de faible inertie pour chaque nuage. Cependant, dans notre exemple,

la consultation des coefficients de relation entre les facteurs de l'Analyse Simultanée et des A.C.⁵ séparées (tableau 5 et figure 6) révèle que ce second axe est très lié avec les seconds axes des analyses séparées des premier, second et cinquième groupes (15-17, 18-24 et plus de 44 ans respectivement), ces derniers étant ceux qui contribuent le plus à la création du second axe (tableau 6).

TABLEAU 5
Relation entre les premiers axes des A.C. séparées et l'Analyse Simultanée

		An. Simultanée	
		Axe 1	Axe 2
A.C. 15-17 ans	Axe 1	0.9258	-0.0421
	Axe 2	0.0370	0.8250
A.C. 18-24 ans	Axe 1	0.9504	0.2192
	Axe 2	0.2010	-0.8280
A.C. 25-34 ans	Axe 1	0.9620	0.1519
	Axe 2	0.0704	-0.3079
A.C. 35-44 ans	Axe 1	0.8181	-0.2767
	Axe 2	-0.0998	0.2895
A.C. 45 ans et plus	Axe 1	0.9435	-0.0483
	Axe 2	-0.0337	0.7389

TABLEAU 6
Contributions ($cta_s(g)$) de chaque groupe à l'inertie des 2 premiers axes

	Axe 1	Axe 2
15-17 ans	0.1995	0.2736
18-24 ans	0.2125	0.2507
25-34 ans	0.2148	0.0611
35-44 ans	0.1643	0.1503
45 ans et plus	0.2089	0.2643

Le tableau 6 montre également comment tous les groupes contribuent de manière significative et équilibrée à la création du premier axe. Ce qui, joint aux fortes relations entre celui-ci et les premiers axes des analyses séparées (tableau 5), nous permet de conclure que la structure représentée par le premier axe, association entre l'utilisation fréquente des services techniques d'Internet et l'intérêt pour les

⁵ A.C. représente l'Analyse de Correspondances.

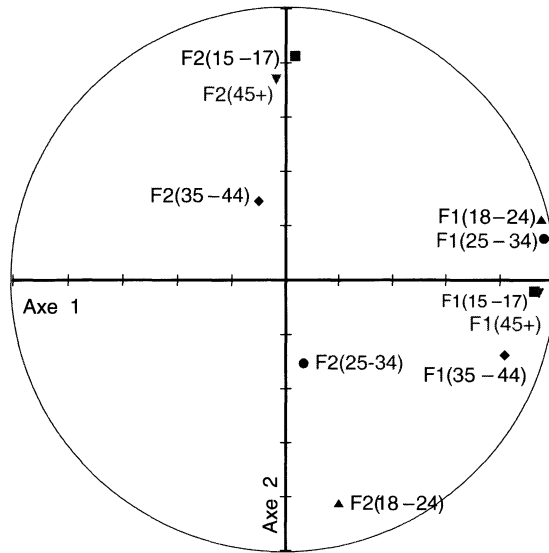


FIGURE 6

Représentation des deux premiers facteurs des A.C. séparées de chaque groupe par leur cosinus avec les deux premiers axes de l'Analyse Simultanée.

Fs(g) : facteur de rang s de l'A.C. séparée du groupe g

différents thèmes d'accès à Internet face à l'association entre le désintérêt pour les thèmes et la faible utilisation des services techniques d'Internet, est une structure commune à chacun des groupes d'âge.

Parallèlement, la structure représentée par le second axe de l'Analyse Simultanée, association entre l'utilisation plusieurs fois par semaine de consultations sur le web avec l'intérêt manifesté pour les journaux électroniques et les informations et pour les bibliothèques et documentation face à la relation entre le désintérêt pour la recherche d'emploi et pour les droits des consommateurs avec une utilisation plus sporadique, est une structure plus présente parmi les individus appartenant aux groupes d'âges de moins de 24 ans et de plus de 44 ans.

A signaler enfin que le quatrième groupe (35-44 ans), celui de plus grande dimensionnalité (tableau 7), est aussi le moins relié, en général, avec les autres groupes d'âge.

TABLEAU 7
 $\mathcal{L}(g, h)$

	15-17	18-24	25-34	35-44	45 ou +
15-17	1.2444				
18-24	0.8295	1.1639			
25-34	0.8966	0.9206	1.0535		
35-44	0.6761	0.7385	0.5979	1.2716	
45 ou +	0.8315	0.9220	0.8314	0.7242	1.2547

Remerciements

Ce travail a été financé par le projet d'investigation UPV 038.321-HA041/99 de l'Université du Pays Basque (UPV/EHU) et UPV 00038.321-13631/2001 et le projet PB98-0149 de la Direction Générale de l'Enseignement supérieur et de la Recherche Scientifique du Ministère Espagnol de l'Éducation et de la Culture.

Références

- BASTIN CH., BENZÉCRI J.P., BOURGARIT CH., CAZES P. (1980), *Pratique de L'Analyse Des Données. 2 : Abrégé Théorique. Études de cas modèle*, Dunod.
- CAZES P. (1982), « Note sur les éléments supplémentaires en analyse des correspondances II. Tableaux multiples », *Les Cahiers de l'Analyse des Données VII*(2), 133-154.
- CHEssel D., HANAFI M. (1996), « Analyses de la co-inertie de K nuages de points », *Revue de Statistique Appliquée XLIV*(2), 35-60.
- ESCOFIER B., PAGÈS J. (1983), « Méthode pour l'analyse de plusieurs groupes de variables. Application à la caractérisation de vins rouges du Val de Loire », *Revue de Statistique Appliquée XXXI*(2) 43-60.
- ESCOFIER B., PAGÈS J. (1988 (1998 3.éd.)), *Analyses Factorielles Simples et Multiples. Objectifs, méthodes et interprétation*, Dunod.
- GOITISOLO B. (2002), *El Análisis Simultáneo. Propuesta y Aplicación de un Nuevo Método de Análisis Factorial de Tablas de Contingencia*, Tesis Doctoral, Universidad del País Vasco.
- LAVIT C. (1988), *Analyse Conjointe de Tableaux Quantitatifs*, Masson. Paris.
- L'HERMIER DES PLANTES H. (1976), *STATIS : Structuration de Tableaux À Trois Indices de la Statistique*, Thèse (3c), USTL, Montpellier.
- MÉOT A., LECLERC B. (1997), « Voisinages A Priori et Analyses Factorielles : Illustration dans le cas de Proximités Géographiques », *Revue de Statistique Appliquée XLV*, 25-44.

- QANNARI A., QANNARI E.M, COURCOUX P. (1992), «Analyse d'un tableau ternaire d'évaluations sensorielles», *Rev. Statistique Appliquée* **XXXX**(4), 45-58.
- SIMIER M., BLANC L., PELLEGRIN F., NANDRIS D. (1999), «Approche simultanée de K Couples de Tableaux : Application à l'étude des relations pathologie végétale-environnement», *Revue de Statistique Appliquée* **XLVII**(1), 31-46.
- TORRE F., CHESSEL D. (1995), «Co-structure de deux tableaux totalement appariés», *Revue de Statistique Appliquée* **XLIII**(1), 109-122.
- ZÁRRAGA A., GOITISOLO B. (2002), «Méthode factorielle pour l'analyse simultanée de tableaux de contingence», *Revue de Statistique Appliquée* **L**(2), 47-70.