

# REVUE DE STATISTIQUE APPLIQUÉE

BERNARD GOLDFARB

CATHERINE PARDOUX

## **Étude de données multidimensionnelles évolutives et comparaison de codages par l'analyse factorielle multiple**

*Revue de statistique appliquée*, tome 49, n° 1 (2001), p. 97-117

[http://www.numdam.org/item?id=RSA\\_2001\\_\\_49\\_1\\_97\\_0](http://www.numdam.org/item?id=RSA_2001__49_1_97_0)

© Société française de statistique, 2001, 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 DONNÉES MULTIDIMENSIONNELLES ÉVOLUTIVES ET COMPARAISON DE CODAGES PAR L'ANALYSE FACTORIELLE MULTIPLE

Bernard Goldfarb<sup>(1)</sup>, <sup>(2)</sup>, Catherine Pardoux<sup>(1)</sup>

(1) LISE-CEREMADE, Université Paris IX – Dauphine

(2) Laboratoire de Biostatistique, CHU Necker-Enfants Malades, Paris

### RÉSUMÉ

Dans ce papier, nous développons une méthodologie d'analyse exploratoire pour des données multidimensionnelles évolutives, que nous appliquons à un essai thérapeutique prospectif concernant le traitement des malades atteints d'insuffisance rénale chronique. Nous proposons de combiner l'information fournie par la survenue d'une censure et celle provenant des observations réalisées sur des marqueurs quantitatifs en utilisant pour chaque marqueur un recodage. Nous présentons l'utilisation de trois codages auxquels nous associons à chacun un tableau de données. La comparaison des trois tableaux obtenus au travers d'une Analyse Factorielle Multiple, montre l'intérêt du codage flou qui permet de conserver une partie de l'information distributionnelle de chaque variable. L'étude et l'interprétation des trajectoires de sous-groupes de patients sont alors définies à l'aide d'une Analyse des Correspondances réalisée sur le tableau associé à ce codage. Cette approche a permis de dégager des conclusions que les méthodes classiques d'analyse des données de durée n'avaient pas permis de mettre en évidence.

*Mots-clés* : Analyse Factorielle Multiple, Codage Flou, Censure Généralisée, Données Multidimensionnelles Évolutives.

### ABSTRACT

In this paper, we develop an exploratory analysis methodology adapted to multivariate temporal data, with an application to a randomized controlled clinical trials designed for comparing non-diabetic chronic renal failure treatments. The progression of chronic renal disease is assessed through both numerical descriptors and qualitative information concerning censored data when regular dialysis has occurred; analysis thus requires coding. Using a Multiple Factorial Analysis, we compare different types of coding, focusing on the interest of fuzzy coding with its capacity to preserve a large amount of the distributional information from each variable. When Correspondence Analysis is performed on the array produced by fuzzy coding, we obtain mean paths, according to sub-group characteristics, on factorial mappings. These procedures have resulted in new conclusions unobtainable with traditional duration data methodology.

*Keywords* : Fuzzy Coding, Generalized Censoring, Multiple Factorial Analysis, Multivariate Temporal Data.

## 1. Introduction

Nous nous intéressons à l'étude de données multidimensionnelles évolutives avec un faible nombre de dates de point, et pouvant présenter des observations censurées. La notion de censure est habituellement appliquée à des données de durée, dans lesquelles le temps écoulé jusqu'à (ou depuis) la survenue d'un événement donné peut ne pas être déterminable, le plus souvent pour des raisons purement expérimentales. Plus généralement, nous définirons la censure à droite pour des variables numériques d'évolution lorsque le recueil ou l'interprétation de la mesure ne peut être effectué au delà d'un certain seuil. On a donc une ou des variables dont le domaine des valeurs analysables est borné à droite. L'information de l'existence de la censure dans l'analyse des processus évolutifs différencie cette situation d'une donnée manquante.

Dans ce papier, nous proposons l'utilisation de trois codages que nous appliquons à des données obtenues dans un essai thérapeutique prospectif concernant le traitement de malades atteints d'insuffisance rénale chronique. Nous associons un tableau de données à chacun des codages. Nous comparons les trois tableaux obtenus par une Analyse Factorielle Multiple. Avec le codage choisi, nous étudions et interprétons les trajectoires [Goldfarb & Pardoux, 2000].

## 2. Présentation des données

Les données proviennent d'un essai thérapeutique concernant le traitement de l'insuffisance rénale chronique [Hannedouche *et al.*, 1994] dans lequel ont été inclus  $n$  patients ( $n = 130$ ) répartis en  $k$  groupes ( $k = 4$ ) selon le traitement appliqué (expérimental/conventionnel) et l'état tensionnel (hypertendu/normotendu). L'essai a duré 3 ans, avec suivi tous les 6 mois.

Le rein est un organe vital plurifonctionnel. De ce fait, dans les processus pathologiques de dégradation rénale progressive, la mesure de l'altération fonctionnelle repose sur plusieurs indicateurs. Par ailleurs, la signification métrologique de chacun de ces indicateurs ne correspond pas à une interprétation pronostique directe. Seul le dépassement simultané de seuils de ces marqueurs est directement interprétable comme nécessitant une suppléance extrarénale par dialyse continue (pouvant être une première étape avant greffe rénale). Pour ces patients, le passage en dialyse continue résultant de la perte fonctionnelle totale du rein, ne permettait plus de donner un sens aux mesures biochimiques effectuées jusque-là. En effet, les fonctions rénales ne sont alors plus assurées que par « pulses » correspondant aux séances de dialyse, et les paramètres mesurés évoluent seulement par accumulation entre deux séances de dialyse. L'information n'est cependant pas manquante et correspond bien à un phénomène de censure à droite généralisée. Ceci implique qu'on dispose pour chaque patient d'un maximum de  $T$  points d'observation ( $T = 7$ ).

Le suivi a porté sur les mesures de  $p$  variables actives continues correspondant aux marqueurs classiques des fonctions rénales ( $p = 6$ ). Pour chaque patient, on dispose par ailleurs de renseignements d'ordre qualitatif (sexe, traitement, état tensionnel, ...) définissant des variables illustratives.

### 3. Codage des données

L'information fournie par la survenue d'une censure à droite généralisée et celle (numérique) provenant des observations non censurées, justifient l'utilisation d'un codage, car contrairement aux modèles classiques de survie, l'objet de l'étude n'est pas le temps écoulé jusqu'à un terme donné, mais des grandeurs biologiques dont on recherche le ou les modes évolutifs.

La censure est ici le substitut qualitatif de la mesure lorsque la valeur de celle-ci a perdu sa signification métrologique.

#### 3.1. Codages définis par les quartiles

Pour chaque variable active continue, on crée quatre modalités  $\{1, 2, 3, 4\}$  à l'aide des trois quartiles de sa distribution définie par  $n \times T$  ( $130 \times 7 = 910$ ) observations non censurées.

La censure prise en compte sous forme d'une modalité supplémentaire, on obtient 5 modalités pour chacune des variables actives et on peut traiter les données par une Analyse des Correspondances Multiples. La censure correspondant à l'aggravation au-delà du stade matérialisé par la valeur 4, les 5 modalités sont de facto ordonnées. Les variables peuvent alors être considérées comme ordinales donnant lieu à un traitement de variables numériques par Analyse en Composantes Principales.

Ci-dessous nous présentons un exemple extrait des données, d'une part mises sous forme « ordinale » ( $n^\circ$  de la modalité), d'autre part écrites sous forme « qualitative disjonctive ». Il s'agit des 3 premières variables d'abord pour les 4 premiers individus à la 1<sup>ère</sup> date de point, puis pour le 1<sup>er</sup> individu aux 2<sup>ème</sup> et 3<sup>ème</sup> dates de point.

- Codage ordinal

		X1	X2	X3
101	1	2	1	1
102	1	2	3	2
103	1	1	3	2
104	1	3	4	1
101	2	4	1	4
101	3	5	5	5

- Codage disjonctif complet

		X1					X2					X3					
101	1	0	1	0	0	0	1	0	0	0	0	1	0	0	0	0	0
102	1	0	1	0	0	0	0	0	1	0	0	0	1	0	0	0	0
103	1	1	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0
104	1	0	0	1	0	0	0	0	0	1	0	1	0	0	0	0	0
101	2	0	0	0	1	0	1	0	0	0	0	0	0	0	1	0	0
101	3	0	0	0	0	1	0	0	0	0	1	0	0	0	0	1	0

### 3.2. Codage flou

Une deuxième approche conduisant à un recodage en 4 classes utilise le principe du codage flou suivant 3 modalités [Cazes, 1990a]. Nous ajoutons une 4<sup>ème</sup> modalité correspondant à la censure. On s'affranchit alors de l'arbitraire des valeurs de coupure (les quartiles) et on conserve une partie de l'information distributionnelle pour chaque variable.

Nous avons retenu l'un des codages flous les plus utilisés. Il consiste à éclater pour chaque individu  $\omega_i$  la variable  $j$  en trois modalités  $j_+, j_-, j_ =$  auxquelles on adjoint la censure  $j_C$  en définissant pour cette variable  $j$ , 4 valeurs numériques :  $k(i, j_+)$  et  $k(i, j_-)$ ,  $k(i, j_ =)$  et  $k(i, j_C)$ . Ces 4 valeurs représentent en fait la distribution de fréquences de la variable sur ces 4 modalités.

La réalisation procède ainsi :

- pour chaque variable  $j$ , on détermine à l'aide des observations non censurées son minimum  $m_j$ , son maximum  $M_j$  et sa moyenne  $\bar{x}_j$ ,

- pour une observation non censurée  $x_{ij} < \bar{x}_j$  : 
$$\begin{cases} k(i, j_+) = k(i, j_C) = 0 \\ k(i, j_-) = \frac{\bar{x}_j - x_{ij}}{\bar{x}_j - m_j} \\ k(i, j_ =) = 1 - k(i, j_-) \end{cases}$$

- pour une observation non censurée  $x_{ij} \geq \bar{x}_j$  : 
$$\begin{cases} k(i, j_ =) = \frac{x_{ij} - \bar{x}_j}{M_j - \bar{x}_j} \\ k(i, j_+) = k(i, j_C) = 0 \\ k(i, j_-) = 1 - k(i, j_ =) \end{cases}$$

- pour une observation censurée  $x_{ij}$  : 
$$\begin{cases} k(i, j_+) = k(i, j_-) = k(i, j_ =) = 0 \\ k(i, j_C) = 1 \end{cases}$$

Pour tout individu  $\omega_i$ , la somme de ces quatre valeurs de la variables  $j$  est égale à 1, comme pour un codage disjonctif complet :

$$k(i, j_+) + k(i, j_-) + k(i, j_ =) + k(i, j_C) = 1$$

De plus, comme pour un tableau disjonctif complet, la somme en ligne est constante et égale au nombre de variables.

Le tableau suivant donne le résultat de ce codage flou pour les mêmes individus que dans les tableaux précédents. Pour la  $j^{\text{ème}}$  variable, les 4 modalités  $j_+, j_-, j_ =$  et  $j_C$  sont notées  $Xj_-P, Xj_-M, Xj_-E$  et  $Xj_-C$ .

		X1				X2				X3			
		X1_P	X1_M	X1_E	X1_C	X2_P	X2_M	X2_E	X2_C	X3_P	X3_M	X3_E	X3_C
101	1	0.000	0.458	0.542	0.000	0.000	0.204	0.796	0.000	0.000	0.471	0.529	0.000
102	1	0.000	0.355	0.645	0.000	0.083	0.000	0.917	0.000	0.013	0.000	0.987	0.000
103	1	0.000	0.510	0.490	0.000	0.083	0.000	0.917	0.000	0.013	0.000	0.987	0.000
104	1	0.000	0.154	0.846	0.000	0.312	0.000	0.688	0.000	0.000	0.317	0.683	0.000
101	2	0.776	0.000	0.224	0.000	0.000	0.363	0.637	0.000	0.638	0.000	0.362	0.000
101	3	0.000	0.000	0.000	1.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000	1.000

#### 4. Traitement de données évolutives – Tableaux ternaires

Les individus et les variables étant les mêmes à chaque date d'observation, on dispose de  $T$  tableaux à  $n$  lignes et  $p$  colonnes (variables actives). Le théorème d'Eckart et Young [Lebart *et al.*, 1995] sur lequel reposent les méthodes factorielles n'admet pas de généralisation en ce sens qu'il n'existe pas de décomposition optimale unique d'un tableau à trois entrées en tableaux de rangs 1.

Parmi les méthodologies proposées, nous avons choisi d'étudier l'évolution des  $n$  individus en utilisant une juxtaposition verticale des  $T$  tableaux, réalisant pour notre application un tableau à 910 lignes [Cazes, 1990b].

Le traitement de ce tableau repose, selon la nature des variables actives, sur une analyse en composantes principales (codage ordinal) ou une analyse des correspondances (codage disjonctif complet, codage flou).

Ce traitement des tableaux ternaires par juxtaposition verticale permet de suivre les individus ou des groupes d'individus homogènes (repérés par leur centre de gravité) vis-à-vis d'une variable supplémentaire et autorise des changements sur les sujets observés, aux différentes dates.

#### 5. Comparaison des codages par l'Analyse Factorielle Multiple

Avec le codage disjonctif complet définissant des variables qualitatives, le tableau à traiter par l'analyse des correspondances multiples est un tableau disjonctif complet à 910 lignes et 30 ( $6 \times 5$ ) colonnes actives.

Avec le codage quartile définissant des variables ordinales, le tableau à traiter par l'analyse en composantes principales est un tableau à 910 lignes et 6 colonnes actives. Avec le codage « flou », le tableau à traiter par l'analyse des correspondances multiples est un tableau disjonctif flou à 910 lignes et 24 ( $6 \times 4$ ) colonnes actives.

La méthode privilégiée pour comparer des codages différents pour un même ensemble de variables en donnant une représentation optimale dans un référentiel commun des facteurs des analyses associées à chaque codage, est l'Analyse Factorielle Multiple (AFM).

##### 5.1. L'Analyse Factorielle Multiple

L'Analyse Factorielle Multiple [Escofier & Pagès, 1998] traite des tableaux où un même ensemble d'individus est défini par plusieurs groupes de variables. Dans notre application, nous avons trois tableaux différents définis par les mêmes variables avec un codage différent pour chaque tableau.

Le principe de l'AFM repose sur une analyse factorielle globale de l'ensemble des groupes, les variables de chaque groupe étant pondérées de façon à équilibrer leur rôle : le coefficient de pondération de chaque variable d'un même groupe étant égal à l'inverse de la racine carrée de la plus grande valeur propre de l'analyse factorielle séparée du groupe, la plus grande valeur propre (ou inertie axiale maximum) de chaque groupe est égale à la valeur 1.

Comme en Analyse Canonique Généralisée [Carroll, 1968], on cherche en AFM, une suite de *variables générales*, orthogonales entre elles, les plus liées à l'ensemble des groupes de variables et, pour chaque variable générale et dans chaque groupe, des *variables canoniques* les plus liées aux variables générales. La mesure de liaison entre une variable  $z$  et un groupe de variables n'est pas en AFM le coefficient de détermination comme en ACG, mais la mesure de liaison  $Lg$  égale à la valeur de l'inertie projetée des variables du groupe sur  $z$ . L'AFM s'applique à des tableaux formés de groupes de variables quantitatives, qualitatives ou de type fréquence, à condition qu'à l'intérieur d'un groupe, les variables soient de même type.

L'AFM prend en compte la structure en groupes de variables, permet la comparaison globale des groupes, la mise en évidence de facteurs communs aux différents groupes et une visualisation des ressemblances entre groupes.

## 5.2. Mise en œuvre

L'Analyse Factorielle Multiple a été réalisée à l'aide de la procédure AFMULT de LADDAD. Le programme commence par réaliser une ACP séparée de chaque groupe. Dans les groupes qualitatifs, les indicatrices des modalités sont centrées et pondérées de telle sorte que leur ACP est équivalente à une ACM. Les axes factoriels extraits de ces analyses séparées, dits « axes partiels », sont considérés comme des variables supplémentaires (cf. annexe § 3).

### 5.2.1. Analyses séparées des groupes

- Le tableau associé au codage disjonctif complet a été traité par une Analyse des Correspondances. La 1<sup>ère</sup> valeur propre, voisine de la valeur 1, indique que le nuage des points-individus est séparé en deux sous-nuages et que les modalités des variables se décomposent en deux groupes [Lebart *et al.*, 1995]. On peut constater, à l'aide de l'interprétation des facteurs, que l'un des sous-groupes de modalités est constitué des modalités relatives à la censure.

ANALYSE DU GROUPE 1 : codage quartile qualitatif (ACM)  
 INERTIE TOTALE (= SOMME DES VALEURS PROPRES) : 4.091  
 DIAGRAMME COUPE A 5 VALEURS

NUM	INERTIE	POURCENT	CUMUL	VARIATION
1	.9977	24.387	24.387	24.387
2	.5783	14.135	38.522	10.251
3	.3618	8.843	47.365	5.292
4	.2890	7.064	54.430	1.779
5	.2382	5.823	60.252	1.241

- Le tableau associé au codage ordinal a été traité par une Analyse en Composantes Principales non normée. Le 1<sup>er</sup> facteur résume à lui seul 73 % de l'information. Les variables étant toutes corrélées fortement et positivement au 1<sup>er</sup> facteur, cette situation n'est autre que l'« effet taille » [Escofier & Pagès, 1998]. Le codage ordinal traite l'état « censure » comme étape ultime de la maladie, et ne traite pas cet état en tant que « donnée spéciale » comme le font les deux autres codages.

Ce 1<sup>er</sup> facteur apparaît comme un facteur de confusion lié à la fois à l'état de gravité de la maladie et à la censure.

ANALYSE DU GROUPE 2 : codage quartile ordinal (ACP non normée)

INERTIE TOTALE (= SOMME DES VALEURS PROPRES) : 10.279

DIAGRAMME COUPE A 5 VALEURS

NUM	INERTIE	POURCENT	CUMUL	VARIATION
1	7.4982	72.949	72.949	72.949
2	1.2775	12.428	85.378	60.521
3	.8220	7.997	93.375	4.431
4	.3921	3.815	97.189	4.182
5	.2168	2.109	99.298	1.706

On aurait pu songer à s'affranchir de cet « effet taille » en utilisant une Analyse des Correspondances (AC) au lieu d'une Analyse en Composantes Principales (ACP).

La pondération apportée par l'AC à chaque individu, égale à la fréquence marginale, aurait alors été reportée pour les autres tableaux <sup>1</sup>. Mis à part la disparition de l'« effet taille », les résultats obtenus pour cette analyse par une AC appliquée à ce tableau sont parfaitement homogènes avec ceux de l'ACP.

- Le tableau associé au codage « flou » a été traité par une Analyse des Correspondances du tableau défini au § 3.2. On retrouve une 1<sup>ère</sup> valeur propre voisine de 1 associée à l'ensemble des modalités « censure ».

Pour cette analyse, la représentation dans le premier plan factoriel donne un excellent résumé de l'information (73 % de l'inertie). Toutes les modalités « censure » ayant sur le 1<sup>er</sup> axe factoriel une qualité de représentation égale à 1 (tableau 1), le 1<sup>er</sup> facteur est défini par la censure. Pour chaque variable, les modalités « + » et « - » ayant des coordonnées opposées sur le 2<sup>nd</sup> axe factoriel, le 2<sup>nd</sup> facteur est lié à l'état de gravité de la maladie, hors ce qui est de la censure.

ANALYSE DU GROUPE 3 : codage flou (ACM)

INERTIE TOTALE (= SOMME DES VALEURS PROPRES) : 1.693

DIAGRAMME COUPE A 5 VALEURS

NUM	INERTIE	POURCENT	CUMUL	VARIATION
1	.9977	58.929	58.929	58.929
2	.2404	14.202	73.130	44.727
3	.0805	4.753	77.883	9.449
4	.0780	4.608	82.491	.144
5	.0719	4.249	86.741	.359

Avec le codage flou, on conserve une partie de l'information distributionnelle pour chaque variable, et l'état « censure » se distingue encore plus nettement des autres états qu'avec le codage disjonctif complet : c'est, en effet, le seul état pour

<sup>1</sup> En Analyse Factorielle Multiple, tous les individus ont la même pondération quel que soit le groupe



lequel on a obligatoirement une séquence du type « 0 0 0 1 » (cf. § 3.1 et § 3.3). Ce sont les raisons pour lesquelles on a besoin de beaucoup plus de facteurs avec le codage quartile qualitatif pour une restitution de l'information équivalente à celle du codage flou.

### 5.2.2. Analyse globale

L'analyse globale du tableau complet à 910 lignes et 60 (30 + 6 + 24) colonnes est une ACP pondérée de l'ensemble des variables, le coefficient de pondération de chaque variable d'un même groupe étant égal à l'inverse de la racine carrée de la plus grande valeur propre de l'analyse factorielle séparée de ce groupe.

En AFM, la liaison  $Lg$  entre une variable quantitative  $z$  et un groupe de variables est égale à la valeur de l'inertie projetée des variables du groupe sur  $z$ . La 1<sup>ère</sup> variable générale  $z_1$  maximise la somme de ses liaisons avec chacun des groupes. La 2<sup>nde</sup> variable générale  $z_2$ , orthogonale à  $z_1$ , maximise la somme de ses liaisons avec chacun des groupes...

On montre que les variables générales sont les composantes principales normées de l'ACP pondérée du tableau complet. En notant  $\lambda_\alpha$  la valeur propre de rang  $\alpha$  et  $\Psi_\alpha$  le facteur associé, on a :  $\Psi_\alpha = \sqrt{\lambda_\alpha} z_\alpha$ .

A chaque groupe  $g$  de variables, on associe un nuage d'individus dit « partiel » qui contient les individus vus par ce groupe de variables, et on projette les nuages partiels sur les axes principaux de l'analyse factorielle globale. La projection du nuage associé au groupe  $g$  sur l'axe factoriel de rang  $\alpha$  de l'analyse globale définit la variable canonique  $\Psi_\alpha^g$  du groupe  $g$  qui n'est autre que la combinaison linéaire des variables du groupe la plus liée à la variable générale  $z_\alpha$ .

A un coefficient près, la représentation d'un individu  $\omega_i$  par sa coordonnée est au barycentre de ses représentations par les  $\Psi_\alpha^g$  : on dit que l'image globale d'un individu est au centre de gravité de ses images partielles. On retrouve cette propriété pour les modalités des variables qualitatives puisqu'on représente une modalité par le centre de gravité des individus qui la possèdent.

ANALYSE DE L'ENSEMBLE DES GROUPES (AFM)  
INERTIE TOTALE (= SOMME DES VALEURS PROPRES) : 7.168  
DIAGRAMME COUPE A 10 VALEURS

NUM	INERTIE	POURCENT	CUMUL	VARIATION
1	2.7481	38.336	38.336	38.336
2	1.1104	15.490	53.826	22.846
3	.4267	5.953	59.779	9.537
4	.3653	5.097	64.876	.856
5	.3154	4.400	69.275	.697
6	.2615	3.648	72.923	.752
7	.2443	3.408	76.331	.240
8	.1985	2.768	79.099	.639
9	.1794	2.503	81.602	.265
10	.1773	2.474	84.076	.029

La 1<sup>ère</sup> valeur propre, égale à 2,75, est très élevée puisque très proche de sa valeur maximale possible, à savoir 3. Ceci signifie que les 1<sup>ers</sup> axes de chaque groupe sont très proches du 1<sup>er</sup> axe de l'analyse globale. Cette analyse fournit une représentation des individus et des variables qui s'interprète selon les règles usuelles de l'analyse factorielle. En examinant les contributions et les qualités de représentation des colonnes actives et supplémentaires ainsi que leur représentation dans le premier plan factoriel, on retrouve la même interprétation pour les deux premières variables générales que pour les deux premiers facteurs du codage flou, à savoir une 1<sup>ère</sup> variable générale liée à la censure et une 2<sup>nde</sup> variable générale liée à l'état de gravité de la maladie, pour ce qui n'est pas de la censure (cf. annexe § 1, 2 et 4). Au-delà de cette analyse globale, la prise en considération de la notion de groupe de variables enrichit l'interprétation des résultats.

### 5.2.3. Liens entre l'analyse globale et les analyses par groupe

Les valeurs des mesures de liaison  $Lg$  et  $RV$  entre les groupes de variables pris deux à deux [Escofier & Pagès, 1998, Robert & Escoufier, 1976], nous montrent que :

- le 1<sup>er</sup> groupe présente la plus grande dimensionnalité avec  $Lg(1, 1) = 1, 883$ ,
- les groupes 1 et 3 ont les structures les plus proches ( $RV = 0, 842$ ),
- la structure commune aux groupes 1 et 3 est la plus riche avec  $Lg(1, 3) = 1, 186$ .

	$Lg$			$RV$		
	1	2	3	1	2	3
1 : codage disjonctif	1.883			1.000		
2 : codage ordinal	.899	1.045		.650	1.000	
3 : codage flou	1.186	.713	1.083	.842	.670	1.000

Les variables canoniques du groupe  $g$  étant les projections du nuage associé à ce groupe sur les axes de rang correspondant de l'analyse globale, les valeurs des coefficients de corrélation entre les variables canoniques d'un groupe et les variables générales montrent dans quelle mesure les variables générales de l'analyse globale peuvent être considérées comme des structures du groupe  $g$ . Ici, la 1<sup>ère</sup> variable générale est commune aux trois groupes, et la seconde aux 1<sup>er</sup> et 3<sup>ème</sup> groupes.

### CORRELATIONS ENTRE LES VARIABLES CANONIQUES ET LES VARIABLES GÉNÉRALES

	1=F	2=F	3=F	4=F	5=F
GR 1	992	980	985	950	948
GR 2	927	516	251	731	588
GR 3	973	805	824	793	500

Ces résultats confirment la ressemblance plus forte entre les groupes codage disjonctif complet et codage flou qu'entre l'un de ces groupes et le groupe codage ordinal.

Le rapport de corrélation, défini comme le quotient de l'inertie inter par l'inertie totale, est calculé pour chaque axe de la façon suivante : à chaque individu actif, est associé une classe regroupant l'ensemble des points représentant cet individu, ici chaque classe contient trois points puisqu'on a trois groupes actifs. Ce rapport montre dans quelle mesure chaque axe met en évidence une structure commune aux groupes.

Le tableau suivant montre à nouveau que le premier facteur, lié à la censure, est commun aux trois groupes. Ce rapport de corrélation est plus faible pour les axes suivants qui ne sont communs qu'aux 1<sup>er</sup> et 3<sup>ème</sup> groupes.

RAPPORT : INERTIE INTER/INERTIE TOTALE

1=F	2=F	3=F	4=F	5=F
930	606	453	618	459

#### 5.2.4. Représentation des groupes

A chaque groupe, on associe sa matrice des produits scalaires entre individus qui est un élément de l'espace  $R^{N^2}$ ,  $N$  étant égal à  $nT$  (ici 910). L'Analyse Factorielle Multiple fait aussi intervenir les matrices  $z_\alpha z'_\alpha$  associées à chaque variable générale  $z_\alpha$ . Ces éléments forment une base orthonormée d'un sous-espace de  $R^{N^2}$ . C'est sur cette base que sont projetés les points-groupes pour visualiser leurs proximités. L'intérêt de cet espace de projection est que les axes sur lesquels on projette possèdent la même interprétation que les axes de l'analyse globale.

On montre que la coordonnée du groupe  $g$  sur l'axe de rang  $\alpha$  est égale à  $Lg(z_\alpha, g)$ , cette mesure de liaison étant toujours comprise entre 0 et 1 du fait de la pondération des variables à l'intérieur des groupes. La somme des coordonnées des groupes sur un axe donné est ainsi égale à l'inertie de l'axe correspondant dans l'analyse globale.

COORDONNEES ET AIDES A L'INTERPRETATION DES GROUPES ACTIFS

GR	QLT	NCOL	INR	1=F	COR	CTR	2=F	COR	CTR	3=F	COR	CTR	4=F	COR	CTR
GR 1	829	30	463	962	504	350	613	205	552	346	65	811	213	25	584
GR 2	751	6	264	855	700	311	201	39	181	17	0	40	92	8	253
GR 3	889	24	273	931	801	339	297	81	267	63	4	149	60	3	163
	825		1000	2748	637	1000	1110	127	1000	427	31	1000	365	15	1000

Les valeurs de la qualité de représentation des groupes dans 1<sup>er</sup> plan principal montrent que les deux premiers facteurs de l'analyse globale rendent mieux compte de la structure du 3<sup>ème</sup> groupe que de celle des 1<sup>er</sup> et 2<sup>ème</sup> groupes (figure 1).

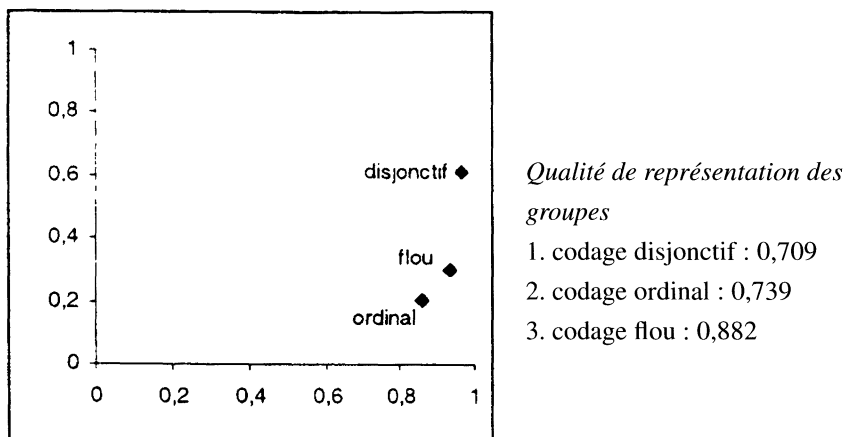


FIGURE 1  
*Représentation des groupes dans le premier plan factoriel*

### 5.2.5. Bilan

L'analyse factorielle multiple est l'outil privilégié pour comparer des codages différents pour un même ensemble de variables en donnant dans un référentiel commun une représentation optimale des facteurs des analyses séparées des groupes associés à chaque codage (cf. annexe § 3).

Les codages disjonctif et flou ont déjà été comparés, mais dans le cadre de données exclusivement quantitatives [Gallego, 1982]. L'auteur a pu essentiellement conclure à une plus grande clarté des résultats, ainsi qu'à une facilité de lecture avec le codage flou.

Nos principaux résultats concernant le codage flou :

- sauvegarde d'une partie de l'information distributionnelle pour chaque variable,
- qualité de la représentation du groupe «codage flou » dans le 1er plan factoriel,
- valeur du pourcentage de l'information restituée dans le plan factoriel de l'analyse séparée, et qualité de représentation des deux premiers axes partiels actifs dans le premier plan factoriel de l'analyse globale,

conduisent à le considérer comme le codage le plus pertinent pour notre application.

C'est ce codage que nous choisissons par conséquent pour construire les trajectoires.

## 6. Représentation et interprétation des trajectoires de groupes d'individus

Une analyse factorielle des correspondances a été réalisée à l'aide du logiciel SPAD sur le tableau associé au codage flou à 910 lignes, 24 colonnes actives et 5 variables nominales supplémentaires.

Les valeurs propres et leurs pourcentages d'inertie associés ont été donnés au § 5.2.1.

Les 5 variables nominales illustratives ont été définies au § 2 :

- la date (7 modalités  $T_1$  à  $T_7$ ),
- le sexe (Homme, Femme),
- l'état tensionnel (normotendu NT, hypertendu HT),
- le traitement (hypertenseur H, placebo P, expérimental E),
- et une variable croisant l'état tensionnel et le traitement (NT\*P, NT\*E, HT\*H, HT\*E).

Cette dernière variable ne comporte que 4 modalités puisque pour les normotendus, ne sont possibles que les traitements placebo et expérimental, et pour les hypertendus, les traitements hypertenseur et expérimental.

On dispose pour chaque individu  $\omega_i$  de  $T$  coordonnées (7 coordonnées pour notre application) sur chaque axe factoriel. Soient  $\Psi_{1i}(t)$  et  $\Psi_{2i}(t)$ , les coordonnées de  $\omega_i$  dans le premier plan factoriel à la date  $t$ .

	Axe 1			...	Axe $q$		
1	$\Psi_{11}(1)$	...	$\Psi_{11}(T)$	...	$\Psi_{q1}(1)$	...	$\Psi_{q1}(T)$
$i$	$\Psi_{1i}(1)$	...	$\Psi_{1i}(T)$	...	$\Psi_{qi}(1)$	...	$\Psi_{qi}(T)$
$n$	$\Psi_{1n}(1)$	...	$\Psi_{1n}(T)$	...	$\Psi_{qn}(1)$	...	$\Psi_{qn}(T)$

Les  $T$  points reliés par la chronologie définissent la trajectoire de  $\omega_i$  dans le premier plan factoriel. Les trajectoires ainsi obtenues constituent un résumé projectif de l'évolution des individus sur l'ensemble des variables actives. On peut aussi représenter pour chaque date les centres de gravité de modalités d'une variable supplémentaire, les points moyens selon le sexe, par exemple (figure 2).

Les contributions et cosinus carrés des modalités actives (tableau 1) complètent les résultats donnés au § 5.2.1, à savoir, un 1<sup>er</sup> facteur défini par la censure et un 2<sup>nd</sup> facteur lié à l'état de gravité de la maladie, pour ce qui n'est pas de la censure.

Pour compléter l'interprétation, une analyse de variance à deux critères « Temps » et « Sexe » avec interaction a été réalisée pour chacun des deux premiers facteurs de l'analyse des correspondances (tableaux 2 et 3) : l'effet « Temps » est significatif pour les deux facteurs, l'effet « Sexe » n'est significatif que pour le 2<sup>nd</sup> facteur, et leur interaction n'est significative pour aucun des deux facteurs.

Ces conclusions sont confirmées par l'examen des valeurs-test [Lebart *et al.*, 1995] des modalités illustratives (tableau 4).

Ces résultats illustrés par la représentation des trajectoires selon le sexe dans le 1<sup>er</sup> plan principal (figure 2) font évoquer un biais de recrutement : les femmes de l'échantillon avaient, à leur entrée dans l'étude, un potentiel d'évolution de la maladie plus accentué que les hommes. Les premières analyses publiées [Hannedouche *et al.*, 1994] ont en réalité buté sur ces particularités des patients inclus dans l'étude, et qui n'étaient pas prévisibles dans l'état actuel des connaissances, malgré un dispositif d'inclusion rigoureux. Les modèles utilisés ne correspondaient pas à une bonne représentation des données issues de cet essai thérapeutique. Notre approche exploratoire se révèle être une étape nécessaire avant modélisation.

TABLEAU 1

*Coordonnées, contributions et qualités de représentation des fréquences actives (pour les libellés des modalités, se reporter § 3.2)*

FREQUENCES				COORDONNEES			CONTRIBUTIONS			COSINUS CARRES		
IDEN	LIBELLE	P.REL	DISTO	1	2	3	1	2	3	1	2	3
AAAG	- X1_P	1.11	4.87	-0.41	1.39	-0.81	0.2	<b>9.2</b>	9.4	0.03	<b>0.40</b>	0.14
AAAH	- X1_M	3.52	1.40	-0.41	-0.80	-0.21	0.6	<b>9.6</b>	2.0	0.12	<b>0.45</b>	0.03
AAAI	- X1_E	9.57	0.30	-0.42	0.13	0.17	1.7	0.7	3.6	0.57	0.06	0.10
AAAJ	- X1_C	2.47	5.68	2.38	-0.00	0.00	<b>14.4</b>	0.0	0.0	<b>1.00</b>	0.00	0.00
AAAK	- X2_P	1.28	2.86	-0.41	-0.05	0.65	0.2	0.0	7.0	0.06	0.00	0.15
AAAL	- X2_M	1.78	2.75	-0.41	0.07	-0.65	0.3	0.0	9.5	0.06	0.00	0.15
AAAM	- X2_E	11.12	0.23	-0.42	-0.01	0.03	2.0	0.0	0.1	0.74	0.00	0.00
AAAN	- X2_C	2.49	5.66	2.37	-0.00	0.00	<b>14.3</b>	0.0	0.0	<b>0.99</b>	0.00	0.00
AAAO	- X3_P	2.82	1.90	-0.42	1.11	-0.33	0.5	<b>14.9</b>	4.0	0.09	<b>0.65</b>	0.06
AAAP	- X3_M	1.47	3.43	-0.42	-1.09	-0.79	0.3	<b>7.5</b>	11.8	0.05	<b>0.35</b>	0.18
AAAQ	- X3_E	9.87	0.29	-0.41	-0.16	0.21	1.7	1.0	5.8	0.60	0.08	0.16
AAAR	- X3_C	2.51	5.49	2.34	-0.00	-0.01	<b>14.1</b>	0.0	0.0	<b>1.00</b>	0.00	0.00
AAAS	- X4_P	3.15	1.80	-0.42	1.06	-0.27	0.6	<b>15.1</b>	3.0	0.10	<b>0.62</b>	0.04
AAAT	- X4_M	1.23	3.74	-0.42	-1.13	-0.78	0.2	<b>6.7</b>	9.6	0.05	<b>0.34</b>	0.16
AAAU	- X4_E	9.72	0.31	-0.42	-0.20	0.19	1.7	1.7	4.3	0.56	0.13	0.11
AAAV	- X4_C	2.57	5.22	2.28	-0.00	0.00	<b>13.7</b>	0.0	0.0	<b>0.99</b>	0.00	0.00
AAAW	- X5_P	1.71	2.67	-0.41	1.15	-0.45	0.3	<b>9.7</b>	4.4	0.06	<b>0.50</b>	0.08
AAAX	- X5_M	1.92	2.72	-0.41	-1.06	-0.59	0.3	<b>9.2</b>	8.6	0.06	<b>0.41</b>	0.13
AAAY	- X5_E	10.56	0.25	-0.42	0.01	0.18	1.9	0.0	4.3	0.68	0.00	0.13
AAAZ	- X5_C	2.49	5.61	2.37	-0.00	0.00	<b>14.3</b>	0.0	0.0	<b>1.00</b>	0.00	0.00
AABA	- X6_P	1.13	2.75	-0.41	1.15	-0.46	0.2	<b>6.4</b>	3.0	0.06	<b>0.48</b>	0.08
AABB	- X6_M	1.71	2.71	-0.42	-1.07	-0.57	0.3	<b>8.4</b>	7.2	0.06	<b>0.42</b>	0.12
AABC	- X6_E	11.34	0.22	-0.42	0.05	0.13	2.0	0.1	2.5	0.78	0.01	0.08
AABD	- X6_C	2.49	5.61	2.37	-0.00	0.00	<b>14.3</b>	0.0	0.0	<b>1.00</b>	0.00	0.00

TABLEAU 2  
Analyse de variance pour le 1<sup>er</sup> facteur

Effet	Somme des carrés	ddl	Carré moyen	F	Signification
Temps	88,449	6	14,741	16,714	0,000
Sexe	0,915	1	0,915	1,038	0,309
Temps * Sexe	0,611	6	0,102	0,1162	0,995
Résiduel	732,929	831	0,882		

TABLEAU 3  
Analyse de variance pour le 2<sup>nd</sup> facteur

Effet	Somme des carrés	ddl	Carré moyen	F	Signification
Temps	9,300	6	1,550	8,111	0,000
Sexe	29,066	1	29,066	152,087	0,000
Temps * Sexe	0,350	6	0,058	0,305	0,934
Résiduel	158,819	831	0,191		

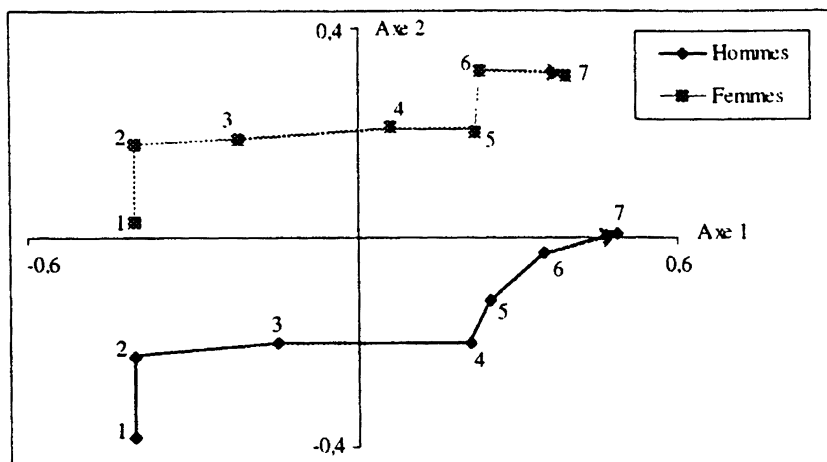


FIGURE 2  
Trajectoires selon le sexe dans le premier plan factoriel

TABLEAU 4  
Valeurs-test et coordonnées des modalités supplémentaires

MODALITES			VALEURS-TEST			COORDONNEES			DISTO.
IDEN	LIBELLE	EFF. P. ABS	1	2	3	1	2	3	
1 . Date									
AM01	- T1	130 780.00	<b>-5.2</b>	<b>-4.9</b>	1.6	-0.42	-0.39	0.13	5.50
AM02	- T2	130 780.00	<b>-5.2</b>	-1.1	0.2	-0.42	-0.09	0.02	5.50
AM03	- T3	122 732.00	<b>-2.2</b>	-0.4	0.6	-0.18	-0.03	0.05	5.93
AM04	- T4	116 696.00	1.7	-0.4	0.3	0.15	-0.03	0.03	6.28
AM05	- T5	117 702.00	<b>2.8</b>	0.6	0.3	0.24	0.05	0.03	6.22
AM06	- T6	116 696.00	<b>3.5</b>	<b>3.1</b>	-1.2	0.30	0.26	-0.11	6.28
AM07	- T7	114 684.00	<b>5.2</b>	<b>3.5</b>	-2.0	0.45	0.30	-0.18	6.41
2 .Sexe									
AA01	- Homme	457 2742.00	1.0	<b>-11.2</b>	-2.5	0.03	-0.35	-0.08	0.85
AA02	- Femme	388 2328.00	-1.0	<b>11.2</b>	2.5	-0.04	0.42	0.09	1.18
5 . Etat tensionnel									
AE01	- NT	236 1416.00	<b>-4.6</b>	-0.4	-2.1	-0.26	-0.02	-0.12	2.58
AE02	- HT	609 3654.00	<b>4.6</b>	0.4	2.1	0.10	0.01	0.05	0.39
6 . Traitement									
AF01	- H	290 1740.00	<b>4.8</b>	0.8	-0.1	0.23	0.04	0.00	1.91
AF02	- P	131 786.00	<b>-3.7</b>	-0.8	0.2	-0.30	-0.06	0.01	5.45
AF03	- E	424 2544.00	-1.9	-0.2	-0.1	-0.06	-0.01	0.00	0.99
7 . Etat*Traitement									
DM01	- NT*P	131 786.00	<b>-3.7</b>	-0.8	0.2	-0.30	-0.06	0.01	5.45
DM02	- NT*E	105 630.00	<b>-2.2</b>	0.2	-3.1	-0.20	0.02	-0.28	7.05
DM03	- HT*H	290 1740.00	4.8	0.8	-0.1	0.23	0.04	0.00	1.91
DM04	- HT*E	319 1914.00	-0.4	-0.3	2.1	-0.02	-0.01	0.09	1.65

## 7. Conclusion

L'application de cette méthodologie factorielle aux données de l'essai thérapeutique a permis de montrer l'importance du choix du codage et de mettre en évidence l'avantage du codage flou qui sauvegarde une partie de l'information distributionnelle des variables. Par ailleurs, elle a permis de dégager des conclusions (dont un effet sexe) qui n'avaient pas pu être caractérisées dans l'analyse initiale réalisée dans le cadre classique des modèles de données de durée : modèle de Cox, analyse en données répétées [Hannedouche *et al.*, 1994]. L'approche classique modélisante se heurte à des difficultés aussi bien d'identification que d'estimation du fait du très grand nombre de paramètres en regard du nombre limité de points d'observation. Même dans un cas de modèle identifiable, les estimateurs peuvent être fortement imprécis.



### Références

- CARROLL J.D. (1968) *A generalization of canonical correlation analysis to three or more sets of variables*, Proc. 76th Conv. Amer. Psych. Assoc., 227-228.
- CAZES P. (1990a) *Codage d'une variable continue en vue de l'analyse des correspondances*, RSA, 38, 3, 35-51.
- CAZES P. (1990b) *Quelques considérations sur l'utilisation des éléments supplémentaires en Analyse Factorielle*, Revue MODULAD n° 5, 29-44.
- ESCOFIER B., PAGÈS J. (1998) *Analyses factorielles simples et multiples*, 3<sup>ème</sup> éd., Dunod.
- GALLEGO F.J. (1982) *Codage flou en Analyse des correspondances*, Les Cahiers de l'Analyse des Données, Vol. VII, n° 4, 413-430.
- GOLDFARB B., PARDOUX C. (2000) *Données multidimensionnelles évolutives avec comparaison de codages par l'Analyse Factorielle Multiple*, XXXII<sup>èmes</sup> Journées de Statistique, SFdS, Fès, Maroc.
- HANNEDOUCHE T., LANDAIS P., GOLDFARB B. *et al.* (1994) *Blood pressure control in non diabetic chronic renal failure : a prospective controlled study*, British Medical Journal, 309 : 833-837.
- LEBART L., MORINEAU A., PIRON M. (1999) *Statistique exploratoire multidimensionnelle*, 2<sup>ème</sup> édition, Dunod.
- ROBERT P., ESCOUFIER Y. (1976) *A unifying tool for linear multivariate methods : the RV coefficient*, Applied Statistics, 25, 3, 257-265

### Logiciels

- LADDAD (1997) *AFMULT, Programme d'Analyse Factorielle Multiple*, version 3-1997, ADDAD, 22 rue Charcot, 75013-Paris.
- SPAD (1999) *Système portable pour l'analyse des données*, version 4.0, CISIA.

## Annexe

## Résultats de l'analyse globale : procédure AFMULT de LADDAD

## 1. Coordonnées et aides à l'interprétation des colonnes actives

- Modalités du groupe 1 « codage disjonctif complet » : Q11 (1<sup>ère</sup> modalité de la 1<sup>ère</sup> variable), ... , Q65 (5<sup>ème</sup> modalité de la 6<sup>ème</sup> variable).
- Variables du groupe 2 « codage ordinal » : X1, X2, X3, X4, X5, X6.
- Modalités du groupe 3 « codage flou » : notations définies au § 3.2.

GR	COL	QLT	POID	INR	1=F	COR	CTR	2=F	COR	CTR	3=F	COR	CTR		
1	Q11	495	33	19	-748	136	7	898	<b>196</b>	24	-675	111	35		
1	Q12	153	33	19	-550	73	4	356	31	4	408	41	13		
1	Q13	179	33	19	-294	21	1	-468	54	6	566	79	25		
1	Q14	437	32	19	-59	1	0	-1279	<b>395</b>	48	-231	13	4		
1	Q15	986	23	20	2387	<b>899</b>	<b>47</b>	729	84	11	-75	1	0		
1	Q21	527	36	18	-461	59	.3	-252	18	2	-430	51	16		
1	Q22	107	37	18	-422	50	2	-183	9	1	-181	9	3		
1	Q23	122	31	19	-404	38	2	-33	0	0	112	3	1		
1	Q24	533	27	20	-345	23	1	23	0	0	749	106	35		
1	Q25	986	23	20	2387	<b>899</b>	<b>47</b>	729	84	11	-75	1	0		
1	Q31	656	34	19	-779	156	8	992	<b>253</b>	30	-812	170	53		
1	Q32	268	31	19	-541	67	3	332	25	3	801	146	47		
1	Q33	218	32	19	-304	22	1	-414	41	5	794	150	47		
1	Q34	656	31	19	-3	0	0	-1518	<b>525</b>	64	-639	93	30		
1	Q35	986	23	20	2387	<b>899</b>	<b>47</b>	729	84	11	-75	1	0		
1	Q41	609	32	19	-777	146	7	1012	<b>247</b>	30	-785	149	47		
1	Q42	243	32	19	-544	70	3	316	24	3	709	119	38		
1	Q43	220	31	19	-303	21	1	-431	43	5	820	155	49		
1	Q44	644	29	19	0	0	0	-1540	<b>502</b>	62	-683	99	32		
1	Q45	986	23	20	2387	<b>899</b>	<b>47</b>	729	84	11	-75	1	0		
1	Q51	763	35	18	-764	153	7	1011	<b>268</b>	32	-645	109	34		
1	Q52	536	34	19	-525	69	3	213	11	1	841	178	56		
1	Q53	590	30	19	-313	21	1	-415	38	5	622	84	27		
1	Q54	640	32	19	-6	0	0	-1467	<b>501</b>	61	-715	119	38		
1	Q55	986	23	20	2387	<b>899</b>	<b>47</b>	729	84	11	-75	1	0		
1	Q61	775	33	19	-787	155	8	1077	<b>290</b>	35	-732	134	42		
1	Q62	590	32	19	-515	64	3	172	7	1	882	187	59		
1	Q63	605	32	19	-318	24	1	-373	33	4	698	114	36		
1	Q64	641	32	19	-23	0	0	-1428	<b>489</b>	59	-767	141	45		
1	Q65	986	23	20	2387	<b>899</b>	<b>47</b>	729	84	11	-75	1	0		
GRUPE	1	577		572			234	350			149	552		84	811

GR	COL	QLT	POID	INR	1=F	COR	CTR	2=F	COR	CTR	3=F	COR	CTR
2	X1	816	133	32	1040	635	52	-470	130	27	119	8	4
2	X2	931	133	32	890	455	38	353	72	15	325	61	33
2	X3	881	133	32	1070	668	56	-550	176	36	45	1	1
2	X4	868	133	31	1059	672	54	-523	164	33	31	1	0
2	X5	930	133	32	1073	662	56	-541	168	35	-57	2	1
2	X6	932	133	32	1061	657	55	-541	171	35	-47	1	1
GROUPE	2	893		191	624	311		147	181		12	40	
3	X1.P	438	10	8	-27	0	0	-1417	<b>368</b>	<b>18</b>	-511	48	6
3	X1.M	671	32	7	-668	273	5	672	<b>277</b>	13	-287	50	6
3	X1.E	595	88	5	-364	314	4	-265	166	6	192	87	8
3	X1.C	986	23	20	2387	<b>899</b>	<b>47</b>	729	84	11	-75	1	0
3	X2.P	633	12	5	-357	40	1	-3	0	0	604	113	10
3	X2.M	550	16	7	-445	64	1	-275	24	1	-454	67	8
3	X2.E	564	103	5	-415	504	6	-114	38	1	14	1	0
3	X2.C	986	23	20	2387	<b>899</b>	<b>47</b>	729	84	11	-75	1	0
3	X3.P	764	26	8	-79	3	0	-1243	<b>685</b>	<b>36</b>	-292	38	5
3	X3.M	573	13	7	-759	144	3	926	<b>214</b>	10	-711	127	16
3	X3.E	604	89	5	-460	488	7	37	3	0	221	112	10
3	X3.C	986	23	20	2387	<b>899</b>	<b>47</b>	729	84	11	-75	1	0
3	X4.P	712	28	9	-93	4	0	-1205	<b>639</b>	<b>36</b>	-231	23	3
3	X4.M	499	11	7	-762	127	2	975	<b>207</b>	9	-712	111	13
3	X4.E	553	86	6	-478	459	7	88	15	1	197	78	8
3	X4.C	986	23	20	2387	<b>899</b>	<b>47</b>	729	84	11	-75	1	0
3	X5.P	683	16	7	-61	1	0	-1299	<b>551</b>	<b>24</b>	-529	91	10
3	X5.M	734	18	8	-740	176	3	928	<b>276</b>	14	-533	91	12
3	X5.E	626	97	5	-412	449	6	-129	44	1	202	108	9
3	X5.C	986	23	20	2387	<b>899</b>	<b>47</b>	729	84	11	-75	1	0
3	X6.P	670	10	5	-59	1	0	-1301	<b>535</b>	16	-547	95	7
3	X6.M	740	16	7	-747	179	3	945	<b>287</b>	13	-515	85	10
3	X6.E	641	104	5	-400	469	6	-171	86	3	146	62	5
3	X6.C	986	23	20	2387	<b>899</b>	<b>47</b>	729	84	11	-75	1	0
GROUPE	3	811		237	549	339		175	267		37	149	
				1000			1000			1000			1000

## 2. Coordonnées et aides à l'interprétation des colonnes supplémentaires

Les modalités des variables illustratives ont été définies au § 6.

GR	COL	QLT	POID	INR	1=F	COR	CTR	2=F	COR	CTR	3=F	COR	CTR
4	T1	66	51	42	-538	48	5	262	11	3	95	1	1
4	T2	37	51	42	-445	33	4	-33	0	0	-39	0	0
4	T3	7	51	42	-175	5	1	-36	0	0	36	0	0
4	T4	4	51	42	124	3	0	54	0	0	38	0	0
4	T5	11	51	42	232	9	1	16	0	0	24	0	0
4	T6	25	51	42	335	19	2	-143	3	1	-70	1	1
4	T7	42	51	42	468	36	4	-120	2	1	-84	1	1
4	Homme	167	191	23	-75	7	0	358	150	22	7	0	0
4	Femme	167	164	27	87	7	0	-418	150	26	-8	0	0
4	NT	47	101	35	-223	20	2	-97	4	1	-204	17	10
4	HT	47	254	14	89	20	1	38	4	0	81	17	4
4	H	57	120	33	196	20	2	128	8	2	59	2	1
4	P	30	55	42	-254	12	1	-121	3	1	-89	1	1
4	E	32	180	24	-53	3	0	-48	2	0	-12	0	0
4	NT*P	30	55	42	-254	12	1	-121	3	1	-89	1	1
4	NT*E	25	46	43	-186	5	1	-67	1	0	-340	17	13
4	HT*H	57	120	33	196	20	2	128	8	2	59	2	1
4	HT*E	27	134	31	-7	0	0	-42	1	0	101	6	3

### 3. Coordonnées et aides à l'interprétation des axes partiels

Les axes partiels sont issus des analyses séparées des groupes. Ils sont considérés comme des variables supplémentaires. Pour chacun des groupes, on dispose des coordonnées, qualités de représentation et contributions des vecteurs unitaires des 5 premiers axes factoriels du groupe sur les 5 premiers axes factoriels de l'analyse globale.

GR	COL	QLT	POID	INR	1=F	COR	CTR	2=F	COR	CTR	3=F	COR	CTR
1	101	1000	1000	140	955	911	332	291	85	76	-31	1	2
1	102	997	580	81	292	85	18	-954	909	475	-35	1	2
1	103	990	363	51	28	1	0	-37	1	0	950	902	766
1	104	625	290	40	14	0	0	-16	0	0	-14	0	0
1	105	405	239	33	24	1	0	10	0	0	186	35	19
GROUPE 1		540		345		234	350	149	551		82	790	
2	201	999	1000	140	-922	850	309	376	142	128	-47	2	5
2	202	929	170	24	-171	29	2	-582	339	52	-287	82	33
2	203	607	110	15	-32	1	0	-119	14	1	78	6	2
2	204	8	52	7	5	0	0	1	0	0	-68	5	1
2	205	6	29	4	25	1	0	47	2	0	2	0	0
GROUPE 2		893		190		624	311	147	181		12	40	
3	301	1000	1000	140	955	911	332	291	85	76	-31	1	2
3	302	963	241	34	286	82	7	-937	878	191	10	0	0
3	303	705	81	11	-25	1	0	-24	1	0	-831	691	131
3	304	68	78	11	-7	0	0	-11	0	0	-255	65	12
3	305	749	72	10	14	0	0	40	2	0	11	0	0
GROUPE 3		794		205		549	339	175	267		36	145	
				740			1000			1000			975

La qualité de représentation du premier axe factoriel dans le premier plan factoriel de l'analyse globale est égale à : 0,996 pour le groupe 1 (codage disjonctif complet), 0,992 pour le groupe 2 (codage ordinal) et 0,996 pour le groupe 3 (codage flou). La qualité de représentation du deuxième axe factoriel dans le premier plan factoriel de l'analyse globale est égale à : 0,994 pour le groupe 1, 0,368 pour le groupe 2 et 0,960 pour le groupe 3.

L'indice COR du groupe  $g$  pour l'axe  $\alpha$  de l'analyse globale est égal à l'inertie des 5 premiers axes du groupe  $g$  sur l'axe  $\alpha$  divisée par l'inertie totale du groupe  $g$  projeté sur l'axe  $\alpha$ .

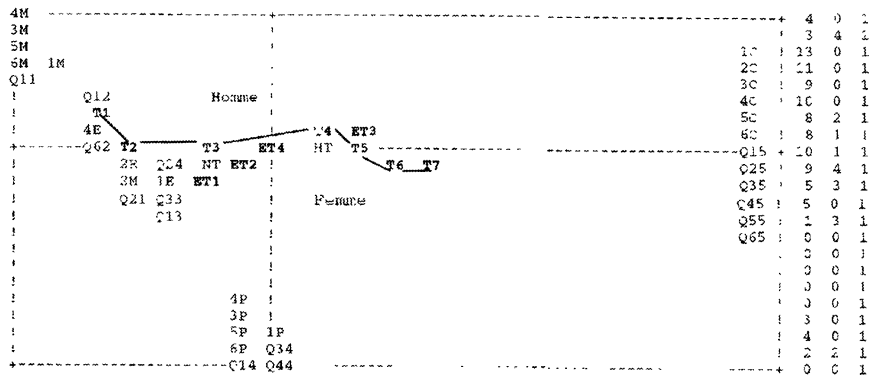
Les deux premiers axes factoriels de chaque analyse séparée rendent compte de 38,5 % de l'inertie du 1<sup>er</sup> groupe, 85,4 % de l'inertie du 2<sup>nd</sup> groupe et 73,1 % de l'inertie du 3<sup>ème</sup> groupe (cf. § 5.2.1).

**4. Comparaison des trois codages :  
codage disjonctif complet, codage ordinal et codage flow**

Pour alléger le graphique :

- les modalités X1\_P, X2\_M, ... sont notées 1P, 2M, ...
- les modalités illustratives NT\*P, NT\*E, HT\*H, HT\*E sont notées respectivement ET1, ET2, ET3 et ET4.

Certaines variables n'ont pas pu être représentées en raison de points multiples.



*Représentation des colonnes actives et supplémentaires dans le premier plan principal de l'analyse globale.*