REVUE DE STATISTIQUE APPLIQUÉE

M. CHAVANCE
F. AMIEL-LEBIGRE

Sur le codage des données en analyse factorielle des correspondances. Application à un questionnaire

Revue de statistique appliquée, tome 23, nº 1 (1975), p. 65-81
<http://www.numdam.org/item?id=RSA_1975__23_1_65_0>

© Société française de statistique, 1975, 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/

SUR LE CODAGE DES DONNÉES EN ANALYSE FACTORIELLE DES CORRESPONDANCES APPLICATION A UN QUESTIONNAIRE [1]

M. CHAVANCE [2] et F. AMIEL-LEBIGRE [3]

Le codage des données pour leur traitement statistique est un problème délicat. Il masque souvent, comme nous le verrons plus loin, l'introduction d'hypothèses implicites sur l'importance et la nature des variables étudiées, même lorsque les techniques utilisées ne reposent sur aucune hypothèse stochastique.

Nous nous proposons d'étudier l'influence du codage en analyse factorielle des correspondances. Pour cela, nous énoncerons quelques principes généraux à respecter dans tout codage, puis nous étudierons les propriétés des variables booléennes disjonctives et nous comparerons trois méthodes de codage applicables en particulier aux questionnaires, nous donnerons enfin un exemple pratique.

I – TROIS PRINCIPES QU'IL EST BON DE RESPECTER

On sait que l'analyse factorielle des correspondances, dont nous ne rappellerons pas le principe ici[4], permet de traduire géométriquement les rapports qui existent entre les lignes et les colonnes d'un tableau à double entrée (k_{ij}) $i \in I$, $j \in J$.

Dans le cas d'un questionnaire, les lignes correspondent aux sujets interrogés, les colonnes aux paramètres traduisant les réponses données à chaque question. Mais il importe de respecter certaines règles.

I.1 – Conservation de l'information

On peut remplacer une variable à q modalités par q variables booléennes sans perdre aucun renseignement : le passage inverse est toujours possible.

---

(1) Article remis le 12/12/73, révisé le 21/6/74

(2) Attaché de recherche INSERM U 88. 91 Bd de l'Hopital. Paris

(3) Attachée de recherche CNRS. Laboratoire de psychologie médicale, clinique des maladies mentales et de l'encéphale. Hôpital Ste Anne. Paris

(4) On pourra se reporter à Benzecri, 1973 (réf. 2) ou à Lebart et Fenelon (1971), (réf. 7) pour un exposé détaillé ou à Nakache, 1973 (réf. 8) pour un résumé de la méthodologie.

Par contre, si l'on transforme un paramètre quantitatif en paramètre qualitatif à q modalités, on conçoit bien qu'il y a appauvrissement de l'information. C'est ce que confirme la théorie de l'information (cf. réf. 5 et 8) qui précise qu'en ce cas la perte d'information est minimale lorsque les différentes modalités sont équiprobables. Bien entendu cette perte est d'autant plus forte que le nombre de modalités est plus réduit.

I.2 – Equipondération des sujets.

Chaque sujet k^i est affecté d'un poids $f_{i.}$ proportionnel à la somme des valeurs des paramètres qui le caractérisent :

$$f_{i.} = \frac{\sum_{j \in J} k_{ij}}{\sum_{i \in I} \sum_{J \in J} k_{ij}} = \frac{k_{i.}}{k}$$

(k_{ij}) $i \in I$, $J \in J$, tableau étudié

I ensemble des sujets

J ensemble des paramètres.

Si l'on convient de noter 1 la présence d'un caractère (e.g. réponse positive à une question donnée) et 0 son absence, on élimine pratiquement certains sujets à qui l'on aurait fait jouer un rôle prépondérant en adoptant le point de vue inverse (ou en rédigeant différemment les questions).

Pour éviter cet inconvénient, il suffit d'ajouter des variables redondantes. Par exemple, en plus de la variable "réponse positive à la question k" codée 1 si le sujet a répondu OUI, 0 sinon, on considère la variable "réponse négative à la question k" codée 1 si le sujet a répondu NON, 0 sinon. Ainsi tous les sujets sont affectés d'un même poids $f_{0.}$ et la somme des éléments de chaque ligne est constante et égale à $k_{0.}$:

$$\forall\, i \in I \quad f_{i.} = f_{0.} \quad , \quad k_{i.} = k_{0.}$$

1.3. – Pondération équivalente des variables

Chaque variable k^j est affectés d'un poids $f_{.j}$ proportionnel à la somme des valeurs qu'elle prend pour l'ensemble des sujets :

$$f_{.j} = \frac{\sum_{i \in I} k_{ij}}{\sum_{i \in I} \sum_{j \in J} k_{ij}} = \frac{k_{.j}}{k}$$

Les deux conditions :

$$f_{i.} = f_{i'.} \qquad \forall\, i \in I \qquad \forall\, i' \in I$$

$$f_{.j} = f_{.j'} \qquad \forall\, j \in J \qquad \forall\, j' \in J$$

sont évidemment contradictoires et il est légitime de n'imposer que la première, mais il ne faut pas introduire de distorsions systématiques en juxtaposant maladroitement des variables quantitatives et qualitatives. Une solution consiste à déterminer des classes de valeurs pour les variables quantitatives (de préférence équiprobables pour satisfaire I.1), ce qui revient à les transformer en paramètres qualitatifs.

II – PROPRIETES DES VARIABLES BOOLEENNES DISJONCTIVES EN ANALYSE FACTORIELLE DES CORRESPONDANCE

Définition 1 : Deux variables booléennes k^j et k^s sont dites disjonctives si elles ne peuvent prendre simultanément la valeur 1

$$\forall i \in I \quad k_{ij} = 1 \quad \Rightarrow k_{ij} = 0$$

$$\forall i \in I \quad k_{is} = 1 \quad \Rightarrow k_{ij} = 0$$

Les seuls couples de valeurs possibles sont donc (1, 0), (0, 1) et (0, 0).

Définition 2 : Un ensemble de q variables booléennes deux à deux disjonctives $k^{j1}, k^{j2}, \ldots, k^{jq}$ est dit complet si pour tout sujet il existe une variable de l'ensemble (nécessairement unique) prenant la valeur 1

$$\forall i \in I \quad \exists l \in [1, q]\, k_{ij_l} = 1$$

II.1. – Distance du chi-deux entre deux variables booléennes disjonctives

Soient k^j et k^s deux variables booléennes disjonctives,

alors la formule usuelle :

$$d^2(k^j, k^s) = \sum_{i \in I} \frac{k}{k_{i.}} \left(\frac{k_{ij}}{k_{.j}} - \frac{k_{is}}{k_{.s}}\right)^2$$

devient

$$d^2(k^j, k^s) = \frac{k}{k_{0.}} \left(\frac{1}{k_{.j}} + \frac{1}{k_{.s}}\right)$$

car en posant

$$I_j = \{i \in I \mid k_{ij} = 1\}$$

$$I_s = \{i \in I \mid k_{is} = 1\}$$

il vient

$$d^2(k^j, k^s) \sum_{i \in I_j} \frac{k}{k_{i.}} \left(\frac{1}{k_{.j}} - 0\right)^2 + \sum_{i \in I_s} \frac{k}{k_{i.}} \left(0 - \frac{1}{k_{.s}}\right)^2$$

mais étant donnée la nature du tableau, nous avons d'une part

$$\forall\, i \in I \qquad \frac{k}{k_{i.}} = \frac{k}{k_{0.}}$$

d'autre part

$$[I_j] = k_{.j} \qquad \text{et} \qquad [I_s] = k_{.s}$$

nous obtenons donc

$$d^2(k^j, k^s) = \frac{k}{k_{0.}} \left[k_{.j}\left(\frac{1}{k_{.j}}\right)^2 + k_{.s}\left(\frac{1}{k_{.s}}\right)^2 \right]$$

$$= \frac{k}{k_{0.}} \left(\frac{1}{k_{.j}} + \frac{1}{k_{.s}}\right)$$

Cette expression peut également se mettre sous la forme

$$d^2(k^j, k^s) = \frac{k}{k_{0.}} \frac{1}{p(1-p)(k_{.j} + k_{.s})}$$

en posant : $$p = \frac{k_{.j}}{k_{.j} + k_{.s}}$$

Cette formule montre que

a) pour k et p fixés, cette distance est inversement proportionnelle à $(k_{.j} + k_{.s})$. Donc, *en moyenne, les points représentatifs d'un ensemble complet de variables disjonctives sont d'autant plus éloignés les uns des autres que le cardinal de cet ensemble est plus élevé.*

b) pour $(k_{.j} + k_{.s})$ fixé, la distance est minimum quand $k_{.j} = k_{.s}$

II.2 – Principe du centre de gravité.

Proposition : Sur chaque axe factoriel, le centre de gravité d'un ensemble complet de variables booléennes disjonctives se trouve a l'origine.

La coordonnée d'une variable k^j sur l'axe factoriel φ^I de composantes $\{\varphi^i, i \in I\}$ est :

$$\psi^j = \sum_{i \in I} \varphi^i \frac{k_{ij}}{k_{.j}}$$

Le centre de gravité d'un ensemble E de variables booléennes disjonctives est donc :

$$\sum_{j \in E} \frac{k_{.j}}{k'} \psi^j = \sum_{j \in E} \sum_{i \in I} \frac{k_{.j}}{k'} \frac{\varphi^i k_{ij}}{k_{.j}} = \sum_{i \in I} \frac{\varphi^i}{k'}$$

avec $$k' = \sum_{j \in E} k_{.j}$$

Mais $\sum_{i \in I} \varphi^i = 0$ (car on sait que φ^I est orthogonal au vecteur de composantes $\{1, 1, \ldots, 1\}$ vecteur propre trivial de l'analyse associé à la valeur propre 1).

Une conséquence de cette propriété est que *les variables de l'ensemble sont d'autant plus éloignées de l'origine qu'elles sont affectées d'un poids $k_{.j}$ plus faible.*

II.3 – Contribution d'un ensemble complet de variables booléennes disjonctives à l'inertie du nuage des variables.

Proposition : La contribution d'un ensemble complet de q variables booléennes disjonctives à l'inertie du nuage est proportionnelle à $q - 1$.

Soit $k^{j_1}, k^{j_2}, \ldots, k^{j_q}$ un ensemble complet de variables disjonctives, la distance du chi-deux entre la variable k^{j_l} et le centre de gravité G, de coordonnées $\left(\frac{k_{i.}}{k}\right)_{i \in I}$ du nuage des variables est :

$$d^2(G, k^{j_l}) = \sum_{i \in I} \frac{k}{k_{i.}} \left(\frac{k_{i.}}{k} - \frac{k_{ij_l}}{k_{.j_l}}\right)^2$$

$$= k \frac{k_{.j_l}}{k_{0.}} \left(\frac{k_{0.}}{k} - \frac{1}{k_{.j_l}}\right)^2$$

$$+ k \frac{k_{.j_1} + \ldots k_{.j_{l-1}} + k_{.j_{l+1}} + \ldots k_{.j_q}}{k_{0.}} \left(\frac{k_{0.}}{k}\right)^2$$

La contribution de l'ensemble à l'inertie du nuage est donc :

$$C = \frac{k}{k_{0.}} \sum_{l=1}^{q} \left[\frac{k_{.j_l}^2}{k} \left(\frac{k_{0.}}{k} - \frac{1}{k_{.j_l}}\right)^2 + \frac{k_{.j_l}}{k} (k_{.j_1} + k_{.j_2} + \ldots + k_{.j_{l-1}} + k_{.j_{l+1}} + \ldots + k_{.j_q}) \left(\frac{k_{0.}}{k}\right)^2 \right]$$

$$= \frac{1}{k_{0.}} \left[\left(\sum_{l=1}^{q} k_{.j_l}^2 + \sum_{l=2}^{q} \sum_{m=1}^{l-1} 2\, k_{.j_l} k_{.j_m} \right) \left(\frac{k_{0.}}{k}\right)^2 - 2 \sum_{l=1}^{q} k_{.j_l} \frac{k_{0.}}{k} + \sum_{l=1}^{q} \left(\frac{k_{.j_l}}{k_{.j_l}}\right)^2 \right]$$

$$= \frac{1}{k_{0.}} \left[n^2 \frac{k_{0.}^2}{k^2} - 2n\, k_{0.} + q \right] = \frac{1}{k_{0.}} \left[\left(\frac{n k_{0.}}{k} - 1\right)^2 + q - 1 \right] = \frac{q-1}{k_{0.}}$$

En effet, puisque l'on a affaire à un ensemble complet $\sum_{l=1}^{q} k_{.j_l} = n$, nombre de lignes du tableau, et la somme des éléments de chaque ligne étant constante :

$$n\, k_{0.} = k$$

III – COMPARAISON DE TROIS CODAGES

Nous allons maintenant procéder à la comparaison de trois procédés de codage communément utilisés pour les données provenant d'un questionnaire fermé.

III.1 – Codage polytomique

Si à chaque question on fait correspondre un ensemble complet de variables disjonctives en associant une variable à chaque modalité de réponse, on ne perd aucune information et on a l'avantage de pouvoir faire apparaître la spécificité de telle ou telle modalité de réponse. Bien entendu les non-réponses, quand il en existe, doivent être considérées comme des modalités particulières, au même titre que les réponses effectives.

Nous avons vu en II.3 que la contribution d'un ensemble complet de q variables booléennes disjonctives à l'inertie du nuage est proportionnelle à (q – 1). Certaines questions jouent alors un rôle deux, trois ou quatre fois plus important que d'autres.

Nous avons en outre montré en II.1 que les points représentatifs d'un ensemble complet de variables disjonctives sont, en moyenne, d'autant plus éloignés de l'origine que le cardinal de l'ensemble est plus élevé. Certains points sont alors rejetés aux extrémités des axes ou (et) mal représentés. Cela peut rendre l'interprétation délicate.

III.2 – Codage dichotomique

Si, pour les questions à réponses multiples, les différentes modalités peuvent être ordonnées, il est possible d'utiliser deux procédés de codage dichotomique.

III.2.1 – Codage dichotomique disjonctif

Si l'on veut que toutes les questions aient un poids égal dans l'analyse, on fait correspondre à chacune un couple de variables booléennes disjonctives en associant les premières modalités à l'une, les dernières à l'autre. En raison de I.1, cela se traduit par une perte d'information que l'on minimise en veillant à ce que, pour chaque variable, les deux valeurs 0 et 1 soient aussi équiprobables que possible.

Exemple :

Si à une question les proportions de réponse ont été :

jamais	20 %
quelquefois	45 %
souvent	35 %

il faut identifier les réponses "jamais" et "quelquefois" et les opposer à "souvent".

III.2.2. – Codage dichotomique pondéré

On peut éviter toute perte d'information en associant à chaque modalité de réponse une note comprise entre 0 et 1. Pour respecter l'équipondération

des sujets, il faut créer une variable redondante qui prend pour valeur le complément à 1 de cette note.

Dans l'exemple précédent, si l'on estime que la réponse "quelquefois" est équidistante de "jamais" et de "souvent", on crée une variable k'^j et une variable k'^s codées :

	k'^j	k'^s
jamais	1	0
quelquefois	1/2	1/2
souvent	0	1

Soulignons l'arbitraire d'une telle attitude, on aurait pu aussi bien faire correspondre à la réponse "quelquefois" les valeurs 1/3 et 2/3 ou 3/4 et 1/4.

Nous n'avons plus affaire à des variables booléennes disjonctives, les résultats démontrés en II ne sont donc plus valables. Nous allons reprendre les calculs en nous plaçant dans le cas particulier le plus simple, celui d'une question à trois modalités de réponse en posant comme règle de passage du codage polytomique au dichotomique :

$$k'^j = k^j + \frac{1}{2} k^q \qquad k'^s = k^s + \frac{1}{2} k^q$$

k^q variable associée à la réponse intermédiaire, k^j et k^s variables associées aux réponses extrêmes dans le codage polytomique,

k'^j et k'^s variables du codage dichotomique pondéré.

La distance du chi deux entre les deux nouvelles variables est alors :

$$d^2(k'^j, k'^s) = \frac{k}{k_0} \left[\frac{k_{.j}}{\left(k_{.j} + \frac{k_{.q}}{2}\right)^2} + \frac{k_{.s}}{\left(k_{.s} + \frac{k_{.q}}{2}\right)^2} \right.$$

$$\left. \frac{k_{.q}}{4} \frac{(k_{.j} - k_{.s})^2}{\left(k_{.j} + \frac{k_{.q}}{2}\right)^2 \left(k_{.s} + \frac{k_{.q}}{2}\right)^2} \right]$$

Un calcul simple montre que cette distance est inférieure à celle qui séparait les variables correspondantes dans le codage polytomique :

$$d^2(k^j, k^s) - d^2(k'^j, k'^s) > 0$$

$$d^2(k^s, k^q) - d^2(k'^j, k'^s) > 0$$

$$d^2(k^q, k^j) - d^2(k'^j, k'^s) > 0$$

Dans le développement de ces expression, tous les termes affectés d'un signe négatif figurent également munis du signe opposé. Il ne reste finalement que des signes positifs

La contribution des deux nouvelles variables à l'inertie du nuage est :

$$C' = \frac{k_{.j} + \frac{1}{2} k_{.q}}{k} d^2(G, k'^j) + \frac{k_{.s} + \frac{1}{2} k_{.q}}{k} d^2(G, k'^s)$$

$$= \frac{1}{k_{0.}} \left[\frac{k_{.j}}{k_{.j} + \frac{k_{.q}}{2}} + \frac{k_{.s}}{k_{.s} + \frac{k_{.q}}{2}} + \frac{k_{.q}}{4\left(k_{.j} + \frac{1}{2} k_{.q}\right)} + \frac{k_{.q}}{4\left(k_{.s} + \frac{1}{2} k_{.q}\right)} - 1 \right]$$

$$= \frac{1}{k_{0.}} \left[\frac{k_{.j} + \frac{k_{.q}}{4}}{k_{.j} + \frac{k_{.q}}{2}} + \frac{k_{.s} + \frac{k_{.q}}{4}}{k_{.s} + \frac{k_{.q}}{2}} - 1 \right]$$

On voit que C' est une fonction décroissante et bornée de $k_{.q}$ puisque tous les termes sont positifs et que, lorsque $k_{.q}$ augmente, le dénominateur croit plus vite que le numérateur. En particulier, on a

$$0 < C' < \frac{1}{k_{0.}}$$

Soulignons que C' est inférieure à $\frac{1}{k_{0.}}$, contribution d'un ensemble complet de deux variables disjonctives. Ce sont donc les questions pour lesquelles deux modalités de réponse seulement sont offertes qui vont jouer un rôle plus important.

On peut se demander quelle est la contribution des deux nouvelles variables k'^j et k'^s quand on utilise une pondération différente, du type ρ, $(1 - \rho)$. Nous devons reconnaître que nous n'avons pas effectué les calculs.

III.2.3. – On voit que le codage polytomique est celui qui rend le mieux compte de la richesse du questionnaire, au prix cependant d'une certaine complication. La fécondité de ce codage est à mettre en relation avec une propriété particulière du tableau qui est alors étudié : J.P. Benzecri (réf. 2, pages 241 à 243) a montré que son analyse est équivalente à celle du tableau de contingence donnant pour chaque couple de réponses possibles le nombre de sujets ayant donné simultanément ces deux réponses. Toutefois, L. Lebart et J.P. Benzecri (réf. 3) ont montré que ce codage ne suffit pas à rendre compte de ce qu'on peut appeler les interrelations d'ordre supérieur ou égal à trois entre les ensembles de réponses possibles aux différentes questions d'un questionnaire.

Le codage dichotomique pondéré simplifie, donc appauvrit, l'interprétation ; en particulier on ne voit plus apparaître la spécificité de telle ou telle modalité de réponse.

Le codage dichotomique disjonctif, en raison de la perte d'information qu'il entraîne, est le moins intéressant. L'essentiel des résultats peut d'ailleurs être retrouvé à partir de l'analyse polytomique en utilisant le principe de la représentation simultanée en analyse des correspondances : *les points représentatifs des variables se trouvent au barycentre des points représentatifs des sujets pour lesquels cette variable est présente* (prend la valeur 1).

Dans la mesure où le changement de codage ne modifie pas trop la répartition des sujets sur les axes, on peut prévoir qu'en identifiant plusieurs variables que l'on remplace alors par une seule, le point représentatif de celle-ci coïncidera avec les points représentatifs de celles-là (affectées de leurs poids respectifs).

Or on sait que tout codage respectant les principes énoncés en I permet de restituer l'essentiel de la structure des données étudiées. C'est ce que l'on va vérifier à travers l'exemple suivant.

IV – ANALYSE DE LA STRUCTURE D'UN QUESTIONNAIRE : LE T.S.T.

Le T.S.T. (test de santé totale) est un questionnaire de 22 items qui vise à apprécier la santé mentale tant sous l'angle des fonctions psychophysiologiques que des fonctions psychiques. Il a été établi par le psychologue américain Langner (1962) et adapté en français par R. Amiel et F. Lebigre (1970). (cf. réf. 1 et 6).

La liste des questions est donnée en annexe. Pour chaque sujet, un score est obtenu en faisant la somme des réponses marquées d'un astérisque.

Le T.S.T. est utilisé dans un but de dépistage, dans des enquêtes épidémiologiques ou pour comparer la psychopathologie de plusieurs sous-populations.

Dans le cadre d'une étude sur l'utilisation du T.S.T. (réf. 1 bis) nous avons été amenés à étudier la structure du test à partir d'un échantillon de 488 sujets de la population générale. Nous avons eu recours, en particulier à l'analyse factorielle des correspondances en utilisant les trois codages étudiés plus haut.

Le tableau I donne les coefficients de corrélation linéaire entre les cinq premiers facteurs (coordonnées des individus sur les axes) de ces trois analyses. On remarque que le premier axe n'est pas perturbé par les changements de codage (coefficients de corrélation supérieurs ou égaux à 0,98), que le second n'est guère modifié quand on passe du codage dichotomique disjonctif au codage dichotomique pondéré mais qu'il n'est est pas de même quand on utilise le codage polytomique, une combinaison linéaire des facteurs deux et trois permet cependant de retrouver le facteur deux des analyses dichotomiques.

(Plus précisément, il existe dans le plan factoriel 2 × 3 de l'analyse polytomique, un axe dont la corrélation avec le deuxième facteur de l'analyse dichotomique vaut :

$$\sqrt{0{,}62^2 + 0{,}71^2} = 0.94).$$

Tous les autres facteurs sont également des combinaisons linéaires de plusieurs facteurs correspondant à un autre codage. Ils ne nous ont pas paru intéressants.

Ces chiffres montrent que les interprétations des trois analyses doivent être analogues. On constate en effet que :

Tableau I - Coefficients de corrélation linéaire entre les cinq premiers facteurs des trois analyses

	DP1	DP2	DP3	DP4	DP5	P1	P2	P3	P4	P5	D1	D2	D3	D4	D5
DP1	1,00														
DP2	−0,00	1,00*													
DP3	−0,00	−0,00	1,00*												
DP4	0,00	−0,00	0,00	1,00*											
DP5	0,00	0,00	−0,00	−0,00	1,00*										
P1	0,98	−0,14	−0,04	0,03	0,06	1,00*									
P2	0,11	0,53	0,07	0,28	0,02	−0,00	1,00								
P3	−0,15	−0,68	−0,13	0,34	0,22	0,00	0,00	1,00*							
P4	−0,00	0,09	0,09	0,45	0,46	0,00	0,00	0,00	1,00*						
P5	0,03	−0,27	0,62	−0,03	−0,19	−0,00	0,00	0,00	−0,00	1,00*					
D1	0,98	−0,10	−0,02	0,02	0,05	0,99	−0,02	−0,08	0,05	0,03	1,00				
D2	0,14	0,91	0,04	−0,09	−0,19	−0,03	−0,62	−0,71	−0,11	−0,18	−0,00	1,00*			
D3	−0,03	0,20	−0,46	0,50	0,48	0,00	0,13	0,16	0,68	−0,52	0,00	−0,00	1,00*		
D4	0,00	0,00	0,83	0,41	0,24	−0,01	0,19	0,03	0,45	0,45	0,00	−0,00	0,00	1,00*	
D5	0,03	−0,15	−0,10	0,56	0,53	0,01	0,28	−0,26	0,09	0,07	−0,00	0,00	−0,00	0,00	1,00*

DP1 – Codage dichotomique pondéré, facteur 1
DP2 – " " " " 2
DP3 – " " " " 3
DP4 – " " " " 4
DP5 – " " " " 5

P1 – " polytomique, facteur 1
P2 – " " " 2
P3 – " " " 3
P4 – " " " 4
P5 – " " " 5

D1 – " dichotomique disjonctif, facteur 1
D2 – " " " " 2
D3 – " " " " 3
D4 – " " " " 4
D5 – " " " " 5

a) Le premier axe oppose toujours la présence et l'absence des troubles, ce qui confirme la significativité du score habituellement calculé ;

b) Une composante dépressive se détache nettement. On la trouve dans le quadrant supérieur droit des figures 1 et 2, et dans le quadrant inférieur droit de la figure 4 ; elle se détache moins bien sur la figure 3. Elle correspond aux variables :

H –	question 6	: humeur triste (ou très triste)
I +	14	: des journées sans s'occuper
P +	19	: soucis qui rendent physiquement malade
J +	20	: sensation d'isolement
Z +	21	: les choses tournent toujours mal
D +	22	: plus rien ne vaut la peine.

Cela justifie que soit calculée une note partielle de dépression.

c) un groupe de variables que l'on peut qualifier d'anxiété somatisée s'oppose à la composante dépressive dans les plans 1 x 2 des analyses dichotomiques et 2 x 3 de l'analyse polytomique. Mais la frontière entre ce groupe et le reste du nuage est imprécise. Le calcul d'une seconde note partielle ne s'impose pas.

Certains résultats ne pouvaient apparaître que grâce au codage polytomique :

a) Toutes les réponses intermédiaires (appétit passable, troubles apparaîssant quelquefois . . .) sortent sur le premier axe du côté des variables "pathologiques" sauf la réponse quelquefois à la question 7 (sentiment de nervosité NQ). Il est donc légitime, à cette expression près, de les faire intervenir dans le calcul du score, en les comptant par exemple pour 1/2 point.

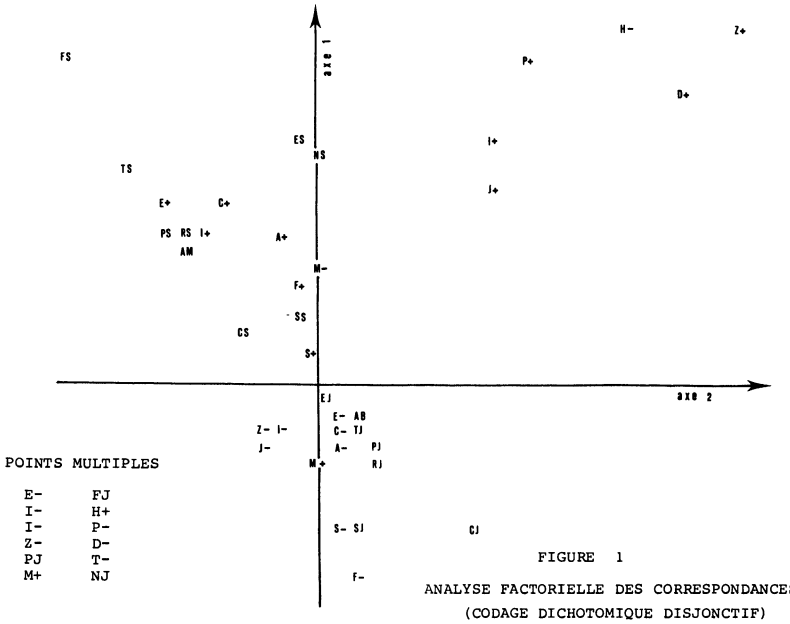
b) les réponses "souvent" aux questions :

10 – gêne respiratoire (RS)
15 – palpitations (PS)
16 – évanouissements (ES)
17 – sueurs froides (FS)
18 – tremblement des moins (TS)

se détachent nettement sur les trois premiers axes. Cela est vraisemblablement dû davantage à une attitude de réponse qu'à la plus grande fréquence de ces troubles chez certains sujets.

Soulignons que l'ensemble des résultats est en accord avec les calculs effectués plus haut. Ainsi les déplacements relatifs des variables d'une analyse dichotomique à l'autre correspondent aux variations de leur contribution à l'inertie ; pour le groupe "anxiété somatisée", par exemple, on voit que sur le deuxième axe (figures 1 et 2) les variables F S, T S, P S, R S, se rapprochent de l'origine quand on passe du codage dichotomique disjonctif au codage pondéré, c'est-à-dire quand on diminue leur contribution à l'inertie du nuage (cf. III.3).

En outre, comme on s'y attendait, les résultats essentiels de l'analyse dichotomique disjonctive peuvent se déduire de l'analyse polytomique. Sur

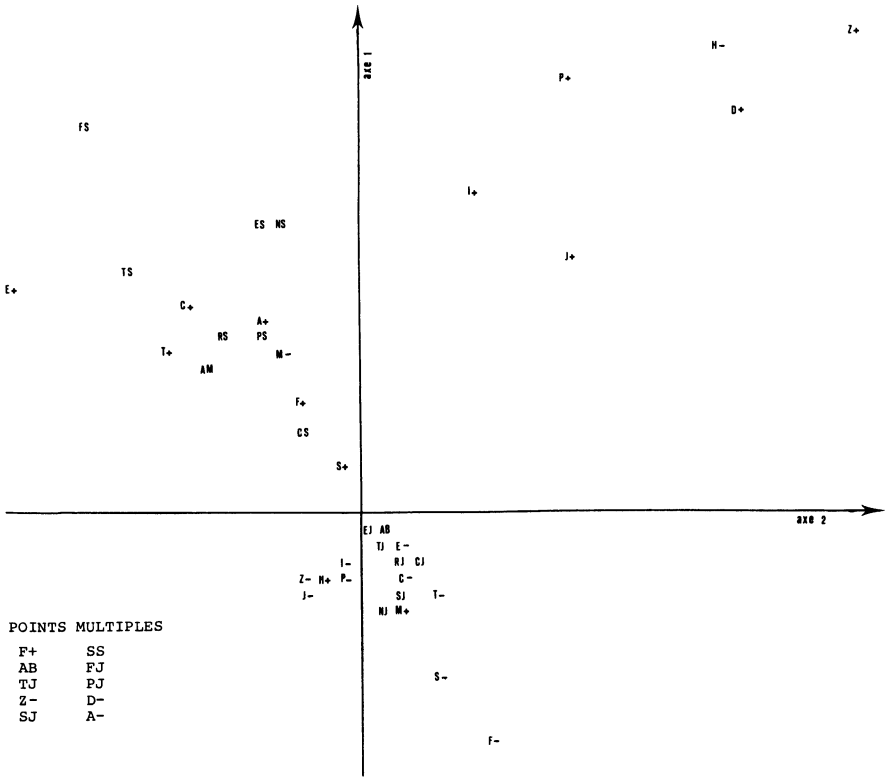


POINTS MULTIPLES

E-	FJ
I-	H+
I-	P-
Z-	D-
PJ	T-
M+	NJ

FIGURE 1

ANALYSE FACTORIELLE DES CORRESPONDANCES
(CODAGE DICHOTOMIQUE DISJONCTIF)

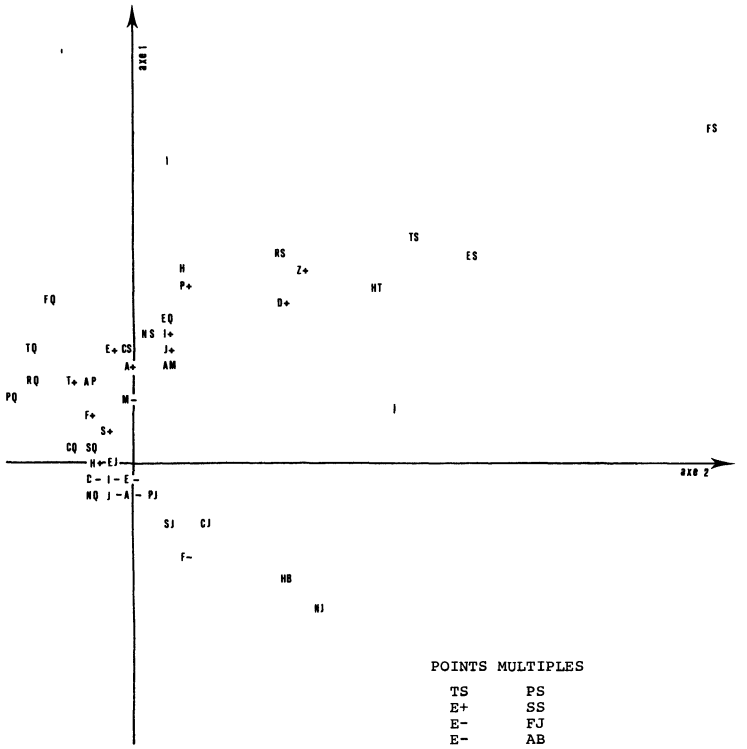


POINTS MULTIPLES

F+	SS
AB	FJ
TJ	PJ
Z-	D-
SJ	A-

FIGURE 2

ANALYSE FACTORIELLE DES CORRESPONDANCES
(CODAGE DICHOTOMIQUE PONDERE)



POINTS MULTIPLES

TS	PS
E+	SS
E-	FJ
E-	AB
I-	Z-
E-	C-
I-	P-
E-	TJ
PJ	T-
A-	M+
PJ	RJ
SJ	S-
E+	C+

FIGURE 3

ANALYSE FACTORIELLE DES CORRESPONDANCES
(CODAGE POLYTOMIQUE)

la figure 4, qui représente le plan factoriel 2 x 3 de l'analyse polytomique, nous avons placé une flèche au centre de gravité de quatre couples de variables identifiées dans l'analyse dichotomique disjonctive. On constate :

a) Que dans ce plan l'axe d'allongement principal du nuage n'est plus l'axe horizontal mais un axe voisin de la seconde bissectrice.

b) Que la succession des variables est identique aux extrémités de cet axe et du second axe factoriel de l'analyse dichotomique disjonctive : F S, T S, P S, R S d'une part, Z +, D +, P + d'autre part.

Ces résultats établissent bien la supériorité du codage polytomique sur les codages dichotomiques étudiés dans cet article.

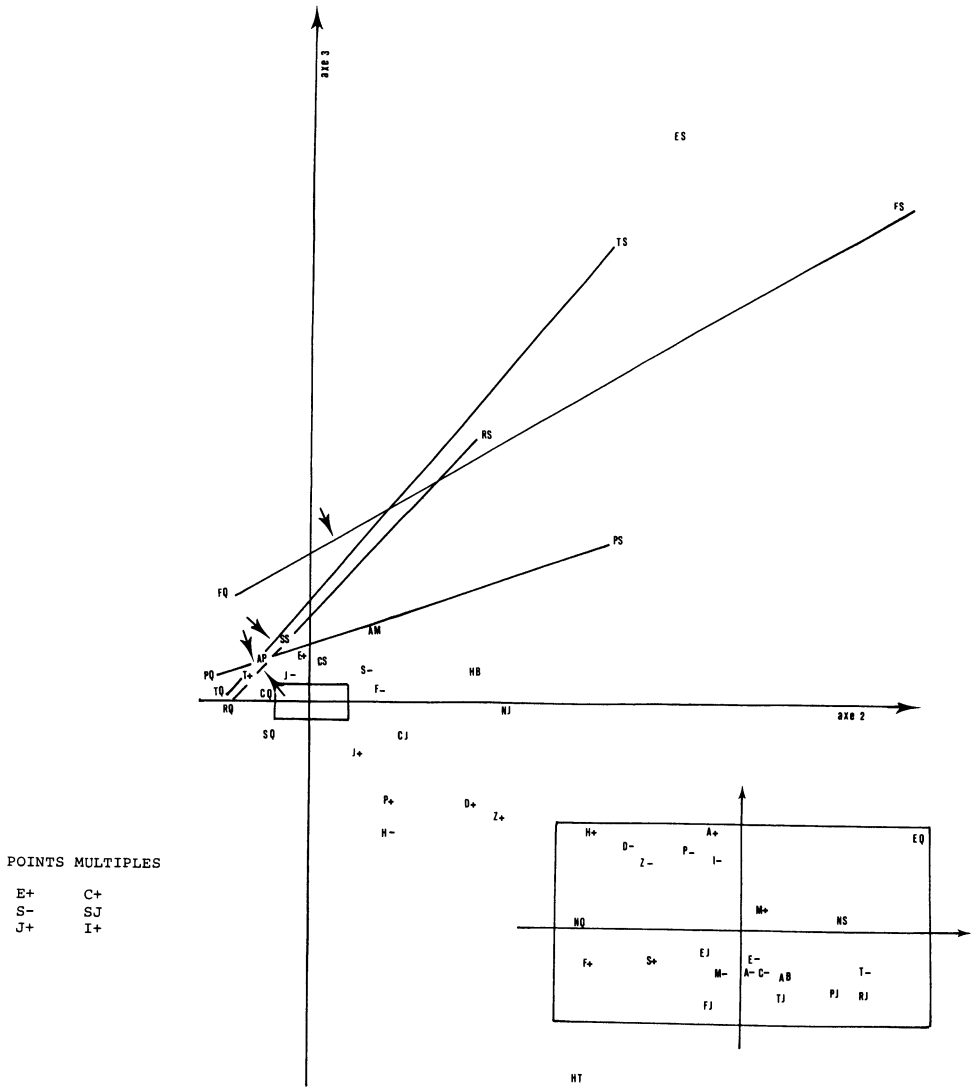


FIGURE 4

ANALYSE FACTORIELLE DES CORRESPONDANCES
(CODAGE POLYTOMIQUE)

BIBLIOGRAPHIE

[1] AMIEL R. et LEBIGRE F. – Un nouveau test rapide d'appréciation de la santé mentale. *Ann. Med. psychol.* – Paris 1970, 1, 4, 565-580.

[1bis] AMIEL, LEBIGRE F. et CHAVANCE M. – Etude d'un questionnaire pour l'appréciation de la santé mentale. – *Rev. Psychol. Appl.* Paris 1974, XX4, 4, 251-265.

[2] BENZECRI J.P. – L'analyse des données – Tome II : L'analyse des correspondances Dunod 73.

[3] BENZECRI J.P. – Sur l'analyse des tableaux binaires associés à une correspondance multiple 1971 – Laboratoire de Statistique – Université Paris VI (polycopié).

[4] CAZES P. – Etude du dédoublement d'un tableau en analyse factorielle des correspondances. 1972 – Laboratoire de Statistique – Université Paris VI (polycopié).

[5] DELABRE M. – Contribution à l'étude de la classification automatique. – Thèse pour le doctorat en médecine – Lille 1971.

[6] LANGNER T. – A twenty two item screening score of psychiatric symptoms indicating impairment – *Health Hum. Behav.* 1962, 3, 269-276.

[7] LEBART L. et FENELON J.P. – Statistique et Informatique appliquées – Dunod 1971.

[8] NAKACHE J.P. – Influence du codage des données en analyse factorielle des correspondances. *Rev. Stat. Appl.* 1973, vol. XXI N° 2, p. 57 à 70.

ANNEXE

QUESTIONS	REPONSES	COD POLYC	COD DIC
1 – (Ces derniers temps) en général, votre appétit est-il bon, passable ou mauvais ?	bon passable mauvais*	AB AP AM	AB AM
2 – Souffrez-vous d'aigreurs d'estomac plusieurs fois par semaine ?	oui* non	E^+ E^-	E^+ E^-
3 – Souffrez-vous de maux de tête : souvent, quelquefois ou jamais ?	souvent* quelquefois jamais	CS CQ CJ	CS CJ
4 – (Ces derniers temps) avez-vous du mal à vous endormir ou à rester endormi : souvent, parfois ou jamais ?	souvent* quelquefois jamais	SS SQ SJ	SS SJ
5 – Dans l'ensemble (ces derniers temps) votre mémoire vous donne-t-elle satisfaction ?	oui non*	M^+ M^-	M^+ M^-
6 – En général (ces derniers temps) êtes-vous de très bonne humeur, d'assez bonne humeur, assez triste ou très triste ?	très bonne h. assez bonne h. assez triste* très triste*	HB H^+ H^- HT	H^+ H^-
7 – (Ces derniers temps) vous arrive-t-il d'être gêné par un état de nervosité, d'irritabilité ou de tension : souvent quelquefois ou jamais ?	souvent* quelquefois jamais	NS NQ NJ	NS NJ
8 – Etes-vous d'un tempérament soucieux ?	oui* non	S^+ S^-	S^+ S^-
9 – (Ces derniers temps) vous arrive-t-il parfois d'être envahi brusquement par une sensation de chaleur ?	oui* non	C^+ C^-	C^+ C^-
10 – (Ces derniers temps) vous est-il arrivé d'être gêné par un essoufflement, une sensation d'étouffement, sans que vous ayez fait d'efforts physiques : souvent, quelquefois ou jamais ?	souvent* quelquefois jamais	RS RQ RJ	RS RJ
11 – (Ces derniers temps) avez-vous assez souvent des sensations de tête lourde ou de nez bouché ?	oui* non	T^+ T^-	T^+ T^-
12 – (Ces derniers temps) avez-vous des moments d'agitation au point de ne pas pouvoir rester en place un certain temps ?	oui* non	A^+ A^-	A^+ A^-

ANNEXE

QUESTIONS	REPONSES	COD POLYC	COD DIC
13 – (Ces derniers temps) vous arrive-t-il assez souvent d'être envahi par une sensation de fatigue générale ?	oui* non	F^+ F^-	F^+ F^-
14 – (Ces derniers temps) vous arrive-t-il de passer des journées, des semaines ou même des mois, sans pouvoir vous occuper de rien parce que vous n'arrivez pas à vous y mettre ?	oui* non	I^+ I^-	I^+ I^-
15 – (Ces derniers temps) avez-vous été gêné par des palpitations : souvent, quelquefois ou jamais ?	souvent* quelquefois jamais	PS PQ PJ	PS PJ
16 – (Ces derniers temps) vous est-il arrivé de vous évanouir : souvent, quelquefois ou jamais ?	souvent* Quelquefois jamais	ES EQ EJ	ES EJ
17 – Vous est-il arrivé (ces derniers temps) d'être gêné par des sueurs froides : souvent, quelquefois ou jamais ?	souvent* quelquefois jamais	FS FQ FJ	FS FJ
18 – (Ces derniers temps) vous arrive t-il d'avoir les mains qui tremblent au point de vous gêner : souvent, quelquefois ou jamais ?	souvent* Quelquefois jamais	TS TQ TJ	TS TJ
19 – (Ces derniers temps) avez-vous des soucis qui vous rendent physiquement malade ?	oui* non	P^+ P^-	P^+ P^-
20 – Vous sentez-vous un peu isolé, un peu seul, même parmi des amis ?	oui* non	J^+ J^-	J^+ J^-
21 – Avez-vous l'impression que les choses tournent toujours mal pour vous ?	oui* non	Z^+ Z^-	Z^+ Z^-
22 – Avez-vous parfois l'impression que plus rien ne vaut plus la peine ?	oui* non	D^+ D^-	D^+ D^-