

JAMBU

**Programme de calcul des contributions  
mutuelles entre classes d'une hiérarchie et  
facteurs d'une correspondance**

*Les cahiers de l'analyse des données*, tome 1, n° 1 (1976),  
p. 77-92

[http://www.numdam.org/item?id=CAD\\_1976\\_\\_1\\_1\\_77\\_0](http://www.numdam.org/item?id=CAD_1976__1_1_77_0)

© Les cahiers de l'analyse des données, Dunod, 1976, tous droits réservés.

L'accès aux archives de la revue « Les cahiers de l'analyse des données » 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/>

PROGRAMME DE CALCUL DES CONTRIBUTIONS MUTUELLES  
ENTRE CLASSES D'UNE HIERARCHIE ET FACTEURS D'UNE CORRESPONDANCE

par M. Jambu <sup>(1)</sup>

0. Objet de ces calculs :

L'Analyse de correspondance a le mérite de faire figurer sur un graphique deux ensembles I et J ; ainsi apparaissent avec la structure de I et celle de J une explication, une interprétation de chacune de ces structures par l'autre. Cependant les praticiens de l'analyse des données et notamment les taxinomistes doivent souvent fixer en un système rigide de classes la vision globale qu'offre l'analyse factorielle ; c'est pourquoi ils recourent à la classification automatique. Mais alors, la richesse de l'interprétation mutuelle de I par J et de J par I est perdue ; on a d'une part des classes sur I ; d'autre part des classes sur J. On doit donc revenir à l'Analyse factorielle, figurer les contours des classes sur les plans factoriels afin de commenter celles-ci ; et l'on dira que sur l'axe 2 la classe q s'oppose à la classe q' ; que la classe q'' est très excentrique dans la direction de l'axe 3 ; etc ... De semblables relations entre classes et facteurs, qui ne peuvent être ou'aproximatives, les calculs des contributions proposés ici permettent de chiffrer la portée ; et ils aideront à les découvrir. Ces calculs ont été introduits en 1974 par M. Sadaka (thèse de 3ème cycle) et M. Jambu ; le programme publié ici est celui de M. Jambu. Afin d'aider le lecteur, on a rassemblé en introduction quelques résultats bien connus sur l'analyse des correspondances (cf e.g. [Prat. Corr.] T II A n°2) et sur la classification hiérarchique (cf notamment [C.A.H.] T.I.B. n°4 § 2.5 et [Inf. Tab] T.I.B. n°5 § 2.3). Les calculs de contributions eux-mêmes font dans la deuxième édition du traité l'objet d'une brève note (cf [Inf. Tab], in fine, Note 2).

1. L'analyse des correspondances ; rappel de notations et résultats

Soit I et J deux ensembles finis et  $k_{IJ}$  un tableau de nombres positifs (e.g. un tableau de contingence) sur  $I \times J : k_{IJ} = \{k(i,j) | i \in I, j \in J\}$ . On note classiquement k le total des nombres contenus dans le tableau  $k_{IJ}$  et on associe à  $k_{IJ}$  une loi de fréquence  $f_{IJ}$  dont les lois marginales sont notées  $f_I$  et  $f_J$  :

$$f_{IJ} = \{f_{ij} | i \in I, j \in J\} ; f_{ij} = k(i,j)/k ;$$

$$f_I = \{f_i | i \in I\} ; f_i = \sum \{f_{ij} | j \in J\} ;$$

$$f_J = \{f_j | j \in J\} ; f_j = \sum \{f_{ij} | i \in I\} .$$

---

(1) Attaché de Recherches CNRS - Université Pierre et Marie Curie - Paris

Aux éléments  $i$  de  $I$  et  $j$  de  $J$  sont associés des profils, que dans l'interprétation probabiliste on se représentera comme des probabilités conditionnelles :

$$\begin{aligned} \text{profil de } i &= f_J^i = \left\{ f_j^i \mid j \in J \right\} ; & f_j^i &= f_{ij} / f_i ; \\ \text{profil de } j &= f_I^j = \left\{ f_i^j \mid i \in I \right\} ; & f_i^j &= f_{ij} / f_j ; \end{aligned}$$

Entre éléments d'un même ensemble (soit  $I$ , soit  $J$ ) on prend la distance du  $\chi^2$  entre leurs profils :

$$\begin{aligned} d^2(i, i') &= \|f_J^i - f_J^{i'}\|_{f_J}^2 = \Sigma \left\{ (f_j^i - f_j^{i'})^2 / f_j \mid j \in J \right\} ; \\ d^2(j, j') &= \|f_I^j - f_I^{j'}\|_{f_I}^2 = \Sigma \left\{ (f_i^j - f_i^{j'})^2 / f_i \mid i \in I \right\} \end{aligned}$$

On a ainsi dans  $R_J$  (muni de la distance du  $\chi^2$  de centre  $f_J$ ) le nuage  $N_J(I)$ , des profils  $f_J^i$  chacun affecté de la masse  $f_i$  ; nuage dont le centre de gravité n'est autre que  $f_J$  ; on a de même dans  $R_I$  le nuage  $N_I(J)$ , dont le centre est  $f_I$  :

$$\begin{aligned} N_J(I) &= \left\{ (f_J^i, f_i) \mid i \in I \right\} \subset R_J ; \\ N_I(J) &= \left\{ (f_I^j, f_j) \mid j \in J \right\} \subset R_I ; \end{aligned}$$

dans ces formules on a donné au nuage  $N_J(I)$  associé à l'ensemble  $I$ , un indice inférieur  $J$  afin de rappeler que ce nuage est dans  $R_J$  ; et de même pour  $N_I(J)$ , associé à  $J$  dans  $R_I$ .

L'analyse factorielle construit simultanément pour chacun des nuages le système des axes principaux d'inertie, ou axes factoriels, indicé par  $\alpha \in A$  ; la distance d'un profil au profil moyen (ou centre de gravité) s'écrit :

$$\begin{aligned} \rho(i)^2 &= \|f_J^i - f_J\|^2 = \Sigma \{ F_\alpha(i)^2 \mid \alpha \in A \} ; \\ \rho(j)^2 &= \|f_I^j - f_I\|^2 = \Sigma \{ G_\alpha(j)^2 \mid \alpha \in A \} . \end{aligned}$$

L'inertie totale ou trace (qui est la même pour l'un est l'autre nuage) s'écrit alors comme une somme double :

$$\begin{aligned} \text{trace} &= \Sigma \{ f_i \|f_J^i - f_J\|^2 \mid i \in I \} = \Sigma \{ f_i F_\alpha(i)^2 \mid i \in I, \alpha \in A \} ; \\ &= \Sigma \{ f_j \|f_I^j - f_I\|^2 \mid j \in J \} = \Sigma \{ f_j G_\alpha(j)^2 \mid j \in J, \alpha \in A \} ; \end{aligned}$$

ou, en considérant les moments d'inertie axe par axe :

$$\begin{aligned} \text{trace} &= \Sigma \{ \lambda_\alpha \mid \alpha \in A \} ; \\ \lambda_\alpha &= \Sigma \{ f_i F_\alpha(i)^2 \mid i \in I \} = \Sigma \{ f_j G_\alpha(j)^2 \mid j \in J \} . \end{aligned}$$

Dans une somme de nombres positifs,  $S = \sum\{s_t | t \in T\}$ , on peut dire que chaque terme  $t$  apporte sa contribution absolue, qui n'est autre que  $s_t$  ; et sa contribution relative  $s_t/S$ . D'où l'usage établi du terme de contribution en analyse factorielle ; et aussi - c'est l'objet de la présente note - en classification hiérarchique. Rappelons que l'on note :

$\tau_\alpha = \lambda_\alpha / \text{trace}$  : contribution relative ou taux d'inertie afférent au facteur  $\alpha$ .

$F_\alpha(i)^2 / \rho(i)^2$  : contribution relative du facteur  $\alpha$  à l'élément  $i$  (i.e. à l'écart du profil  $f_J^i$  de celui-ci par rapport au centre  $f_J$ ) ; cette contribution relative est le carré du cosinus formé par l'axe  $\alpha$  avec le rayon vecteur  $\overline{f_J f_J^i}$  reliant le centre  $f_J$  au profil  $f_J^i$ .

$G_\alpha(j)^2 / \rho(j)^2$  : contribution relative du facteur  $\alpha$  à l'élément  $j$

$f_i F_\alpha(i)^2$  : contribution absolue de l'élément  $i$  au facteur  $\alpha$  (i.e. à la valeur propre  $\lambda_\alpha$  qui est la somme des  $f_i F_\alpha(i)^2$ ) ; on peut encore dire que  $f_i F_\alpha(i)^2$  est la contribution absolue de la paire  $(i, \alpha)$  à la trace, ou inertie totale.

$f_i F_\alpha(i)^2 / \lambda_\alpha$  : contribution relative de l'élément  $i$  au facteur  $\alpha$  ; et de même pour l'ensemble  $J$  : contribution absolue ou relative de  $j$  au facteur  $\alpha$ .

Pour un élément supplémentaire  $s$  qui est placé sur les axes  $\alpha$  après son profil  $f_J^s$  (ou  $f_I^s$  si il s'agit d'un élément adjoint à  $J$ ), mais est traité comme ayant masse nulle, la contribution à  $\lambda_\alpha$  ou à la trace est évidemment nulle ; en revanche on peut encore considérer le rayon vecteur  $\overline{f_J f_J^s}$  et l'angle qu'il forme avec chacun des axes :

$\rho(s)^2 = \sum\{F_\alpha(s)^2 | \alpha \in A\}$  : carré du rayon vecteur de  $s$  ;

$F_\alpha(s)^2 / \rho(s)^2$  : contribution relative du facteur  $\alpha$  à  $s$ .

2. Classification hiérarchique et décomposition de l'inertie d'un nuage :

Rappelons sur une figure la terminologie usuelle de la classification hiérarchique binaire :

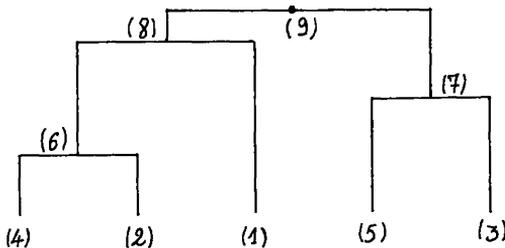


Figure 1 : Un exemple de classification hiérarchique binaire.

$I = \{1,2,3,4,5\}$  : ensemble des éléments sur lesquels est édifée la classification.

$N = \{6,7,8,9\}$  : ensemble des noeuds, ou classes que l'on a construites ; ce sont :

$$6 = \{4,2\} ; 7 = \{5,3\} ; 8 = \{4,2,1\} ; 9 = I = \{1,2,3,4,5\}$$

Le numérotage des noeuds s'effectue dans l'ordre où le programme les a construits : c'est pourquoi sur notre figure le numéro croît avec la hauteur du noeud.

Aîné et Benjamin : chaque noeud  $n$  est formé par réunion de deux classes préexistantes (classe, ou élément isolé) : conventionnellement le programme étiquette l'une de ces classes aîné,  $A(n)$  et l'autre benjamin  $B(n)$  :  $n=A(n) \cup B(n)$ ; le choix de celui qu'on appelle aîné étant indifférent ; sur la figure on pourra prendre l'aîné à gauche :

$$A(9) = 8 ; A(8) = 6 ; A(7) = 5 ; A(6) = 4 ;$$

$$B(9) = 7 ; B(8) = 1 ; B(7) = 3 ; B(6) = 2 ;$$

Supposons maintenant que l'ensemble  $I$  est dans un espace euclidien  $E$  ; notons  $\|i-i'\|^2$  le carré de la distance euclidienne entre deux éléments  $i$  et  $i'$  de  $I$ . Supposons de plus donné un système  $f_I = \{f_i | i \in I\}$  de masses positives. Le centre de gravité  $g$  du système s'écrit (en identifiant  $i$  au point, ou vecteur correspondant de  $E$ ) :

$$g(I) = g = \Sigma \{f_i i | i \in I\}$$

et l'inertie totale, ou moment centré d'ordre 2 de  $I$  est donné par :

$$M_2(I) = \Sigma \{f_i \|i-g\|^2 | i \in I\}$$

Plus généralement, si  $q$  est une partie de  $I$  on pourra noter  $f_q$  sa masse totale,  $g(q)$  ou simplement  $q$  son centre de gravité, et  $M_2(q)$  son moment centré d'ordre 2 :

$$f_q = \Sigma \{f_i | i \in q\} ; q = \Sigma \{(f_i / f_q) i | i \in q\} ;$$

$$M_2(q) = \Sigma \{f_i \|i-q\|^2 | i \in q\} .$$

Ceci posé à chaque noeud d'une classification hiérarchique, on peut associer la différence :

$$\begin{aligned} v(n) &= M_2(n) - (M_2(A(n)) + M_2(B(n))) \\ &= \left( \frac{f_{A(n)} \times f_{B(n)}}{f_n} \right) \|A(n) - B(n)\|^2 ; \end{aligned}$$

en bref, l'inertie totale de la classe  $n$  peut suivant les principes de Huyghens être décomposée en trois termes qui sont  $M_2(A(n))$ ,  $M_2(B(n))$  (inertie des deux classes  $A(n)$  et  $B(n)$  dont  $n$  est la réunion) et  $v(n)$  (terme proportionnel au carré de la distance entre aîné et benjamin du noeud  $n$ ).

Dans la méthode d'agrégation suivant la variance (dite encore de maximisation du moment centré d'ordre 2 d'une partition) c'est justement  $v(n)$  (noté encore  $d(n)$ ) qui est le niveau, ou hauteur à laquelle se place le noeud  $n$  dans la hiérarchie. Mais il est clair que le calcul de  $v(n)$  peut se faire quelle que soit la méthode utilisée pour construire la hiérarchie de classes.

Maintenant les  $v(n)$  fournissent une décomposition de l'inertie totale du Nuage  $M_2(I) = \Sigma\{v(n) | n \in N\}$  (en effet en bref ;  $M_2(I) = v(I) + M_2(A(I)) + M_2(B(I))$  ;  $M_2(A(I))$  peut lui-même être décomposé en descendant l'arbre en  $v(A(I)) + M_2(A(A(I))) + M_2(B(A(I)))$  ; etc...). Donc reprenant la terminologie introduite par l'analyse des correspondances on note :

$v(n)$  : contribution absolue du noeud à l'inertie totale  $M_2(I)$  ;

$\tau(n) = v(n)/M_2(I)$  : contribution relative du noeud  $n$  à l'inertie totale

Tout ce qui est requis pour ces calculs est que  $I$  soit un nuage, un ensemble de points muni de masses dans un espace euclidien. C'est en particulier le cas quand la classification sur  $I$  est effectuée d'après la donnée d'un tableau de correspondance  $k_{IJ}$  : chaque élément  $i$  est identifié à son profil  $f_J^i$ , muni de la masse  $f_i$  ; une classe  $q$  a de même une masse  $f_q$  et un profil  $f_J^q$  qui n'est autre que le centre de gravité des profils de ses points :

$$f_q = \Sigma\{f_i | i \in q\} ; f_{qj} = \Sigma\{f_{ij} | i \in q\} ; f_j^q = f_{qj}/f_q ;$$

$$f_J^q = \Sigma\{f_J^i | j \in J\} = \Sigma\{f_i f_J^i | i \in I\}/f_q ;$$

Evidemment, la formule de distance euclidienne utilisée est la distance du  $\chi^2$  de centre  $f_J$  .

### 3. Contributions mutuelles des facteurs et des noeuds

Soit  $I$  un ensemble d'éléments décrits d'après un tableau de correspondance  $k_{IJ}$  ;  $I$  est identifié au nuage  $N_J(I)$  des profils  $f_J^i \in R_J$  . Une classe  $q$  ou partie de  $I$  a de même une masse  $f_q$  et un profil  $f_J^q$  . Par l'analyse factorielle, l'espace  $R_J$  est muni d'un système d'axes adapté au Nuage  $N_J(I)$  ; sur ces axes, on peut placer la classe  $q$  ; on peut dire que  $f_J^q$  est un élément supplémentaire adjoint à l'ensemble  $I$  ; et plus précisément, comme  $q$  est un barycentre des points  $i$ , on a :

$$F_q(q) = \Sigma\{f_i F_q(i)/f_q | i \in q\}$$

(Il faut prendre garde que  $q$  a une masse  $f_q$  ; mais en le plaçant comme élément supplémentaire, on ne fait pas cas de cette masse ; en ce sens que les axes ont été déterminés sans le point  $q$ ).

Un premier calcul de contribution très simple, et déjà utile est celui des contributions relatives des facteurs aux classes, qu'on préférera appeler  $\text{Cos}^2$  (cosinus carré), pour éviter toute confusion (cf. infra :  $v(n;\alpha)/v(n)$ ) ;

$(F_\alpha(q)/\rho(q))^2 = \text{Cos}^2$  de l'angle formé par le rayon vecteur du noeud  $n$  avec l'axe  $\alpha$ .

On voit ainsi non seulement où est une classe sur les axes (valeurs de  $F_\alpha(q)$ ), mais encore la part relative de chaque axe ou facteur dans l'explication de l'écart entre le profil  $f_J^q$  de cette classe et celui  $f_J$  du centre de gravité du nuage  $N_J(I)$ .

Mais il y a plus ; chaque terme  $v(n)$  associé à un noeud dans l'inertie totale ou trace, peut lui-même être fractionné suivant les axes factoriels ; on a en effet :

$$\begin{aligned} v(n) &= (f_{A_n} \cdot f_{B_n} / f_n) \left\| f_J^{A_n} - f_J^{B_n} \right\|^2 \\ &= (f_{A_n} \cdot f_{B_n} / f_n) \sum \{ (F_\alpha(A_n) - F_\alpha(B_n))^2 \mid \alpha \in A \} \end{aligned}$$

où on a noté en bref  $A_n$  et  $B_n$  pour  $A(n)$  et  $B(n)$  ; on posera donc :  $v(n;\alpha) = (f_{A_n} \cdot f_{B_n} / f_n) (F_\alpha(A_n) - F_\alpha(B_n))^2$ . Ce terme  $v(n;\alpha)$  peut être appelé contribution absolue mutuelle du noeud  $n$  et du facteur  $\alpha$  car il intervient dans la décomposition de  $v(n)$ , de  $\lambda_\alpha$  et de l'inertie totale ou trace ; on a

$$\begin{aligned} v(n) &= \sum \{ v(n;\alpha) \mid \alpha \in A \} ; \\ v_\alpha &= \sum \{ v(n;\alpha) \mid n \in N \} ; \\ M_2(I) = \text{trace} &= \sum \{ v(n;\alpha) \mid n \in N, \alpha \in A \} ; \end{aligned}$$

(la décomposition pour  $\lambda_\alpha$  résulte de ce que  $\lambda_\alpha$  n'est autre que l'analogie du moment d'ordre 2,  $M_2(I)$ , mais pour le nuage projeté sur l'axe  $\alpha$ ).

On reprendra donc les termes de contributions absolues et relatives :

$v(n;\alpha)$  : Contribution absolue mutuelle de  $n$  et  $\alpha$   
 Contribution absolue de la paire  $(n,\alpha)$  à la trace  
 Contribution absolue du noeud  $n$  à l'axe  $\alpha$   
 Contribution absolue de l'axe  $\alpha$  au noeud  $n$

(toutes ces expressions peuvent être justifiées selon le total dont  $v(n;\alpha)$  est considéré comme un membre ; le plus simple est de parler toujours de contribution absolue mutuelle).

$v(n;\alpha)/v_\alpha =$  contribution relative du noeud  $n$  à l'axe  $\alpha$

$v(n;\alpha)/v(n) =$  contribution relative de l'axe  $\alpha$  au noeud  $n$  qu'on se gardera de confondre avec  $(F_\alpha(n)/\rho(n))^2 = \text{Cos}^2$  de l'angle du rayon vecteur de la classe  $n$  avec l'axe  $\alpha$  (cf supra).

Si  $v(n;\alpha)/v(n)$  est proche de 1 c'est que la séparation entre classes  $A_n$  et  $B_n$  (aîné et benjamin dont  $n$  est la réunion) se fait presque exclusivement dans la direction de l'axe  $\alpha$ .

Si  $v(n;\alpha)/\lambda_\alpha \neq 1$ ; c'est que sur l'axe  $\alpha$  la dispersion du nuage est due quasi exclusivement aux points de la classe  $n$ ; donc :

1°) Si  $n$  n'est pas I tout entier,  $F_\alpha(I-n)$  est concentré à l'origine.

2°)  $F_\alpha(A_n)$  et  $F_\alpha(B_n)$  sont concentrés en des pôles opposés de l'axe.

De telles relations apparaissent en effet communément sur les graphiques entre les premiers facteurs et les noeuds les plus hauts; il est curieux de constater que la consultation des contributions relatives permettent parfois d'en découvrir assez bas dans la hiérarchie. En général, on gagnera à faire figurer au moins comme éléments supplémentaires sur les plans factoriels les noeuds supérieurs de la hiérarchie, en reliant par une flèche chaque noeud à ses successeurs immédiats  $A_n$  et  $B_n$ : à l'origine est le sommet I, relié par deux flèches opposées à  $A_I$  et  $B_I$ ; et de même pour chaque noeud  $n$  les flèches  $\overrightarrow{nA_n}$  et  $\overrightarrow{nB_n}$  s'opposent car  $n$  est barycentre de  $A_n$  et  $B_n$ .

Il serait intéressant d'étudier la structure du tableau  $v(n;\alpha)$ ; on peut montrer que  $v(I)$  (niveau du noeud le plus haut, qui n'est autre que I tout entier) est inférieur (ou au plus égal) à la plus forte valeur propre  $\lambda_1$  (car autrement le nuage projeté sur l'axe joignant les centres des classes  $A(I)$  et  $B(I)$  aurait encore une inertie  $\geq v(I)$  donc  $> \lambda_1$ , ce qui est absurde!; le premier axe étant celui sur lequel l'inertie du nuage projeté est la plus forte). N'existe-t-il pas des relations analogues, mais plus complexes entre noeuds et facteurs suivants?

Remarque 1: La double décomposition de l'inertie considérée ici, vaut en général quand le nuage I, muni d'une hiérarchie binaire, est dans un espace euclidien E, lui-même muni d'un système d'axes orthonormés (qu'il s'agisse ou non des axes factoriels issus d'une analyse des correspondances).

Remarque 2: Dans le cas d'un tableau de correspondance il peut être également utile de décomposer le vecteur  $(f_J^{A_n} - f_J^{B_n})$  joignant l'aîné et le benjamin du noeud  $n$ , suivant les axes mêmes de  $R_J$ ; puisque

$$\|f_J^{A_n} - f_J^{B_n}\|^2 = \sum \left\{ \left( f_j^{A_n} - f_j^{B_n} \right)^2 / f_j \mid j \in J \right\}$$

il est naturel, par exemple d'appeler le quotient

$$\left( f_j^{A_n} - f_j^{B_n} \right)^2 / \left( f_j \| f_J^{A_n} - f_J^{B_n} \|^2 \right)$$

contribution relative de l'élément  $j$  au noeud  $n$  (i.e. à l'écart entre ses successeurs  $A(n)$  et  $B(n)$ ).

Il est possible de calculer ces contributions à l'aide du programme donné ci-dessous ; il suffit d'introduire en guise de tableaux de coordonnées des points, non le tableau des facteurs mais le tableau des  $f_j^i \cdot (f_j)^{-\frac{1}{2}}$  qui sont aussi des coordonnées orthonormées pour le nuage I.

#### 4. Les accès au programme de calcul des contributions

4.1. Entrées : Elles sont relatives à un ensemble J dont les éléments sont appelés "variables" et le cardinal est noté CARDJ. L'ensemble des variables est muni d'un système de masses (dont le total est supposé 1) ; d'une classification hiérarchique binaire ; d'un système de coordonnées dans un espace euclidien. Il est naturel de penser que le critère de classification hiérarchique ait été l'agrégation suivant la variance d'une partition avec pour distance la distance  $\chi^2$  ; et que les coordonnées euclidiennes ne sont autres que les facteurs (ou seulement les premiers facteurs) issus d'une analyse des correspondances ; mais cela n'est pas indispensable pour le bon déroulement du programme. On entrera donc :

{NOM (J) | J = 1, CARDJ} : les noms ou sigles abrégatifs des variables.

{PRJ (J) | J = 1, CARDJ} : les poids ou probabilités notés  $f_j$  dans le texte mathématique

{A(N), B(N) | N = LCARD, KCARD} ; les noeuds de la hiérarchie étant (cf § 2) numérotés à partir du bas, de  $CARDJ+1=LCARD$ , à  $2 \cdot CARDJ-1=KCARD$ , on désigne comme à l'ordinaire par A(N) et B(N), les numéros de l'aîné et du benjamin ; {FACJ(A,J) | A = 1, CARFAC ; J = 1, CARDJ} ; les coordonnées des variables dans l'espace euclidien ; si ces coordonnées sont les CARFAC premiers facteurs issus d'une analyse de correspondance on a  $FACJ(A,J) = G_A(J)$  mais répétons-le, cela n'est pas indispensable ; en revanche les calculs effectués perdent leur sens si le centre de gravité n'est pas à l'origine (i.e. les FACJ(A,.) doivent avoir moyenne nulle).

#### 4.2. Les tableaux créés par le programme. Ce sont :

{LAM(A) | A = 1, CARFAC} ; les variances des coordonnées ; celles-ci sont recalculées sous l'hypothèse que les coordonnées ont moyenne nulle et que la somme des masses PRJ(J) est 1 ; évidemment, pour des facteurs ces variances ne sont autres que les valeurs propres.

{PRJ(N) | N = LCARD, KCARD} : les poids des noeuds, calculés en remontant la hiérarchie, comme des poids de l'aîné et du benjamin.

{FACJ(A,N) | A = 1, CARFAC ; N = LCARD, KCARD} : les valeurs des facteurs pour les noeuds de la hiérarchie ; ces valeurs sont calculées suivant la formule :

$$FACJ(A,N) = (PRJ(A(N)) * FACJ(A,A(N)) +$$

PRJ(B(N)) \* FACJ(A,B(N)))/PRJ(N) ; comme pour le calcul des PRJ(N), on doit remonter la hiérarchie afin que les valeurs relatives à A(N) et B(N) soient déjà connues quand on aborde le noeud N.

$\{FAI(A,N) \mid A = 1, CARFAC ; N = LCARD, KCARD\}$  : les contributions mutuelles calculées suivant la formule du §3 :

$FAI(A,N) = (FACJ(A,A(N)) - FACJ(A,B(N)))^2 * QJ(N)$ , où  $QJ(N)$  désigne le coefficient  $(f_{An} \cdot f_{Bn}) / f_n$ , i.e. :

$QJ(N) = PRJ(A(N)) * PRJ(B(N)) / PRJ(N)$ , (le tableau des  $QJ(N)$  étant lui-même gardé en mémoire).

$\{D(N) \mid N = LCARD, KCARD\}$  : ce sont les niveaux  $v(n)$  des noeuds ; si la hiérarchie a été édifée avec pour critère l'agrégation suivant la variance d'une partition, ces niveaux  $v$  sont les indices calculés en construisant les classes ; mais ici les  $v(n)$  sont recalculés suivant la formule :

$D(N) = \Sigma\{FAI(A,N) \mid A = 1, CARFAC\}$ , et c'est pourquoi le programme peut être utilisé quel que soit le critère d'agrégation pris en C.A.H.

#### 4.3. Les sorties :

Ce sont les coordonnées  $FACJ$  des variables et des noeuds dans l'espace factoriel (espace euclidien de dimension  $CARFAC$ ) ; et quatre tableaux de contributions définis au § 3.

. Cosinus carrés des angles des rayons des classes avec les axes ; (le programme omet ce calcul si le noeud est trop proche du centre de gravité).

. Contributions absolues mutuelles des noeuds et des facteurs.

. Contributions relatives des noeuds aux facteurs.

. Contributions relatives des facteurs aux noeuds de plus en marge aux tableaux de contributions, sont donnés les poids et les niveaux des noeuds ( $PRJ(N)$  et  $D(N)$ ).

Nous donnons ici un listage du programme, puis les sorties correspondant à un tableau 28 x 10 relatif à l'état du commerce extérieur de la France en 1832. Sur cet exemple très simple, dont l'analyse est commentée dans le présent cahier, le lecteur pourra s'assurer qu'il a bien suivi le déroulement de l'algorithme ; et s'initier à l'emploi des calculs de contribution dans l'interprétation conjointe des facteurs et des classes.

Pour un autre exemple de calcul de contributions mutuelles entre noeuds et facteurs, voir dans ce cahier l'Analyse de Données sur l'Art Préhistorique ; notamment la figure 9.

### 5. Bibliographie :

Outre le traité sur l'Analyse des Données, partout cité dans ces cahiers, rappelons quelques travaux relatifs aux méthodes et programmes utilisés ici :

- M. Jambu - Techniques de classification appliquées à des données de sciences humaines. Thèse de 3ème cycle - 1972 - L.S.M.
- Programme de construction de classification ascendante hiérarchique (C.A.H.) - Publications du Laboratoire de Statistique - Université Pierre et Marie Curie (Paris VI) - 1974.
  - Quelques calculs utiles à l'interprétation simultanée d'une classification ascendante hiérarchique et d'une analyse de correspondances - Publications du Laboratoire de Statistique - Université Pierre et Marie Curie (Paris VI) - 1974.
- N. Tabet - [BEN. TAB. 75] Programme d'analyse des correspondances - Publications du Laboratoire de Statistique - Université Pierre et Marie Curie (Paris VI) - 1975. Une seconde version du programme figure dans la 2ème édition du traité sur l'Analyse des Données (T.II.D. nos 1, 2, 3).
- M. Sadaka - Comparaison des classes hiérarchiques aux axes factoriels - Thèse 3ème cycle - Université Pierre et Marie Curie (Paris VI) - 1974 - L.S.M.

Les lecteurs désireux d'utiliser ce programme, peuvent acquérir un jeu de cartes perforées en s'adressant à l'A.D.D.A.D. (Association pour le Développement et la Diffusion de l'Analyse des Données), Laboratoire de Statistique, 2ème étage - Tour 45-55. Bureau 10 ; Université Pierre et Marie Curie ; 4, Place Jussieu 75005 - Paris.



```
SUBROUTINE CONTRI (CARDJ,CARFAC,KCARD,KJFAC,IJFAC,FACJ,QJ,PRI,PRJ,
1,LAM,A,B,NOM,FAI,D,TRI)
```

```
C
C
C
C
C
C
C
C
```

```
-----
| SOUS PROGRAMME DE CALCUL DES CONTRIBUTIONS MUTUELLES ENTRE |
| CLASSES D'UNE CLASSIFICATION HIERARCHIQUE ET LES AXES DE |
| L'ANALYSE DE CORRESPONDANCE |
|-----
```

```
INTEGER CARDJ,CARFAC
DIMENSION NOM(CARDJ),TRI(CARFAC)
INTEGER A(KCARD),B(KCARD),PRI(CARFAC)
REAL FACJ(KJFAC), PRJ(KCARD),LAM(CARFAC),QJ(KCARD)
REAL FAI(IJFAC),D(KCARD)
KLEX(I,J,M)=(I-1)*M+J
LCARD=CARDJ+1
MPCARD=CARDJ-1
MKCARD=KCARD-1
76 FORMAT(//)
DO 2 K=1,CARFAC
2 LAM(K)=0
PRINT 152
152 FORMAT(1X,'TABLEAU DES VALEURS PROPRES *100'//)
DO 5 J=1,CARDJ
DO 5 K=1,CARFAC
IJ=KLEX(J,K,CARFAC)
5 LAM(K)= LAM(K)+FACJ(IJ)*FACJ(IJ)*PRJ(J)
DO 6 K=1,CARFAC
6 PRI(K)=100 *LAM(K)
PRINT 300
PRINT 76
PRINT 151,(K,K=1,CARFAC)
PRINT 76
151 FORMAT(1X,10(4X,I2)//)
PRINT 150 ,(PRI(K),K=1,CARFAC)
150 FORMAT(1X,10(1X,I5))
DO 3 J=LCARD,KCARD
IA=A(J)
IB=B(J)
PRJ(J)=PRJ(IA)+PRJ(IB)
DO 3 K=1,CARFAC
IJ=KLEX(J,K,CARFAC)
IJA=KLEX(IA,K,CARFAC)
IJB=KLEX(IB,K,CARFAC)
3 FACJ(IJ)=(PRJ(IA)*FACJ(IJA)+PRJ(IB)*FACJ(IJB))/PRJ(J)
PRINT 75
PRINT 207
207 FORMAT(1X,'TABLEAU DES COORDONNEES DES VARIABLES '/3X,'DANS L'ESPA
1CE FACTORIEL*100'//)
PRINT 300
300 FORMAT(16X,'AXES FACTORIELS')
PRINT 13,(I,I=1,CARFAC)
PRINT 76
DO 20 J=1,CARDJ
DO 21 K=1,CARFAC
IJ=KLEX(J,K,CARFAC)
21 PRI(K)=FACJ(IJ)*100
20 PRINT 120,NOM(J),(PRI(K),K=1,CARFAC)
PRINT 75
PRINT 208
208 FORMAT(1X,'TABLEAU DES COORDONNEES DES CLASSES '/3X,'DANS L'ESPA
1CE FACTORIEL*100'//)
PRINT 300
PRINT 76
PRINT 180,(I,I=1,CARFAC)
PRINT 76
180 FORMAT(1X,4X,10(3X,I2,1X)//)
DO 30 J=LCARD,KCARD
DO 31 K=1,CARFAC
IJ=KLEX(J,K,CARFAC)
31 PRI(K)=100 *FACJ(IJ)
30 PRINT 12 ,J,(PRI(K),K=1,CARFAC)
PRINT 75
PRINT 209
209 FORMAT(1X,'TABLEAU DES COS CARRES DES ANGLES'/3X,'DES RAYONS DES
1CLASSES AVEC LES AXES*100'//)
PRINT 300
PRINT 76
PRINT 181,(I,I=1,CARFAC)
PRINT 76
181 FORMAT(1X,4X,' R02',10(3X,I2,1X)//)
```

```

DD 40 J=LCARD,MCARD
RO2=0
DD 41 K=1,CARFAC
IJ=KLEX(J,K,CARFAC)
TRI(K)=FACJ(IJ)*FACJ(IJ)
41 RO2=RO2+TRI(K)
EPS=0.00001
IF(RO2-EPS) 400,400,401
400 DD 402 K=1,CARFAC
402 PRI(K)=0
GO TO 43
401 DD 42 K=1,CARFAC
42 PRI(K)=100*TRI(K)/RO2
43 IRO2=RO2*100
40 PRINT 160,J,IRO2,(PRI(K),K=1,CARFAC)
160 FORMAT(1X,'N',I3,10(1X,I5))
DD 4 J=LCARD,KCARD
IA=A(J)
IB=B(J)
IQ=J-CARDJ
QJ(J)=PRJ(IA)*PRJ(IB)/PRJ(J)
D(J)=0
DD 4 K=1,CARFAC
IJQ=KLEX(IQ,K,CARFAC)
IJA=KLEX(IA,K,CARFAC)
IJB=KLEX(IB,K,CARFAC)
FAI(IJQ)=QJ(J)*(FACJ(IJA)-FACJ(IJB))*#2
4 D(J)=D(J)+FAI(IJQ)
PRINT 75
PRINT 15
15 FORMAT(5X,'TABLEAU DES CONTRIBUTIONS ABSOLUES MUTUELLES * 1000'/)
75 FORMAT(1H1)
PRINT 301
301 FORMAT(26X,'AXES FACTORIELS'/)
PRINT 130,(I,I=1,CARFAC)
PRINT 76
130 FORMAT(1X,4X,3X,'PRJ',2X,'NIVEAU',10(1X,I2,3X))
DD 10 J=1,MMCARD
IQ=CARDJ+J
QJ(IQ)=1000*QJ(IQ)
D(IQ)=1000*D(IQ)
DD 11 K=1,CARFAC
IJ=KLEX(J,K,CARFAC)
11 PRI(K)= FAI(IJ)*1000
IDIQ=D(IQ)
IQQJ=QJ(IQ)
10 PRINT 12 ,IQ,IQQJ ,IDIQ ,(PRI(K),K=1,CARFAC)
120 FORMAT(1X,A4,10(1X,I5))
12 FORMAT(1X,'N',I3,10(1X,I5))
13 FORMAT(/1X,4X,10(3X,I2,1X)/)
PRINT 75
PRINT 16
16 FORMAT(5X,'TABLEAU DES CONTRIBUTIONS RELATIVES'/8X,'DES NOEUDS AUX
1 FACTEURS*1000'//)
PRINT 301
PRINT 130,(I,I=1,CARFAC)
PRINT 76
DD 100J=1,MMCARD
IQ=CARDJ+J
DD 111K=1,CARFAC
IJ=KLEX(J,K,CARFAC)
111 PRI(K)=1000*FAI(IJ)/LAM(K)
IDIQ=D(IQ)
IQQJ=QJ(IQ)
100 PRINT 12 ,IQ,IQQJ ,IDIQ ,(PRI(K),K=1,CARFAC)
PRINT 75
PRINT 17
17 FORMAT(5X,'TABLEAU DES CONTRIBUTIONS RELATIVES'/8X,'DES FACTEURS
1AUX NOEUDS*1000'//)
PRINT 301
PRINT 130,(I,I=1,CARFAC)
PRINT 76
DD 200J=1,MMCARD
IQ=CARDJ+J
IF (D(IQ).LE.EPS) GO TO 200
DD 211K=1,CARFAC
IJ=KLEX(J,K,CARFAC)
211 PRI(K)=(1000*FAI(IJ)/D(IQ))*1000
IQQJ=QJ(IQ)
IDIQ=D(IQ)
PRINT 12 ,IQ,IQQJ ,IDIQ ,(PRI(K),K=1,CARFAC)
200 CONTINUE
RETURN
END

```

7. Listage des Résultats Obtenus

Par l'analyse de l'Etat du commerce extérieur de la France en 1832.

TABLEAU DES VALEURS PROPRES  $\times 100$ 

AXES FACTORIELS					
1	2	3	4	5	6
53	21	12	2	2	1

TABLEAU DES COORDONNEES DES VARIABLES  
DANS L'ESPACE FACTORIEL  $\times 100$ 

	AXES FACTORIELS					
	1	2	3	4	5	6
DENR	-74	-112	-10	-21	-2	0
EXOT	-72	-72	1	44	2	17
TEIN	-63	-64	-2	73	34	-1
SUCS	-45	-66	-3	12	-3	5
ESP.	-18	-41	-7	39	12	9
MATI	-77	-33	3	48	12	24
PECH	-64	-26	7	56	0	5
META	-54	-17	0	18	11	-1
FRI	-76	24	46	-17	-5	-9
FARI	-61	40	24	-11	-28	-2
OR.A	-71	42	24	3	3	2
BOIS	-55	64	9	-22	0	3
SUB.	-41	36	-13	-9	44	9
PIER	-17	49	-66	-12	38	6
ANIM	-11	33	-78	-35	51	-10
DEPO	-10	33	-91	6	-19	0
FLS	16	28	-51	-11	57	-19
PROD	3	16	-46	-35	30	-17
CHIM	10	-38	18	-12	0	0
FRII	24	0	24	-3	-1	-3
TEIN	60	-12	22	1	21	-2
COUL	57	7	-14	0	11	-42
QUVR	85	-11	13	8	-7	-22
PAPI	89	-1	5	13	-22	-35
TISS	92	0	7	-4	7	7
VITR	94	-3	3	5	-27	-12
COMP	104	-6	19	10	0	-43
ALCO	107	-8	23	3	-28	19

TABLEAU DES COORDONNEES DES CLASSES  
DANS L'ESPACE FACTORIEL  $\times 100$ 

	AXES FACTORIELS					
	1	2	3	4	5	6
N 29	-19	47	-66	-12	39	6
N 30	-67	-27	6	55	1	8
N 31	85	0	2	11	-18	-35
N 32	89	-1	2	8	-22	-25
N 33	-68	-37	5	52	2	10
N 34	12	24	-50	-19	48	-18
N 35	86	-3	11	8	-10	-23
N 36	-48	-20	-1	21	11	0
N 37	19	-15	22	-6	-1	-2
N 38	-14	38	-74	-26	46	-3
N 39	28	-74	22	-4	3	-2
N 40	-8	35	-69	-25	47	-7
N 41	90	-8	12	9	-8	-27
N 42	-55	-26	1	32	8	3
N 43	-59	47	20	-14	-20	0
N 44	-50	-46	-1	22	0	4
N 45	-67	36	33	-15	-13	-4
N 46	-70	40	27	-2	-1	0
N 47	-53	-50	-1	33	6	3
N 48	95	-8	15	7	-14	-13
N 49	93	-3	10	0	0	0
N 50	88	-4	11	0	0	0
N 51	-10	34	-84	-2	0	-1
N 52	-63	-79	-5	7	2	1
N 53	-51	38	-7	-2	-1	0
N 54	-55	2	-6	0	0	0
N 55	0	0	0	0	0	0

TABLEAU DES COS CARRES DES ANGLÉS  
DES RAYONS DES CLASSES AVEC LES AXES=100

AXES FACTORIELS							
	RO2	1	2	3	4	5	6
N 29	88	4	25	50	1	17	0
N 30	84	53	9	0	35	0	0
N 31	91	80	0	0	1	3	14
N 32	92	86	0	0	0	5	6
N 33	90	51	15	0	31	0	1
N 34	63	2	9	39	6	36	5
N 35	84	88	0	1	0	1	6
N 36	34	69	12	0	13	3	0
N 37	11	31	20	43	3	0	0
N 38	102	2	14	54	6	21	0
N 39	15	51	13	32	1	0	0
N 40	90	0	14	53	6	24	0
N 41	92	87	0	1	0	0	7
N 42	49	62	14	0	21	1	0
N 43	88	51	32	5	3	6	0
N 44	52	49	40	0	10	0	0
N 45	74	61	17	14	3	2	0
N 46	73	67	22	10	0	0	0
N 47	65	43	38	0	17	0	0
N 48	98	92	0	2	0	2	1
N 49	89	98	0	1	0	0	0
N 50	79	98	0	1	0	0	0
N 51	84	1	13	84	0	0	0
N 52	104	38	60	0	0	0	0
N 53	42	62	35	1	0	0	0
N 54	30	98	0	1	0	0	0

TABLEAU DES CONTRIBUTIONS ABSOLUES MUTUELLES = 1000

AXES FACTORIELS								
	PRJ	NIVEAU	1	2	3	4	5	6
N 29	0	0	0	0	0	0	0	0
N 30	1	0	0	0	0	0	0	0
N 31	0	0	0	0	0	0	0	0
N 32	3	0	0	0	0	0	0	0
N 33	2	0	0	0	0	0	0	0
N 34	1	0	0	0	0	0	0	0
N 35	11	0	0	0	0	0	0	0
N 36	3	0	0	0	0	0	0	0
N 37	5	0	0	0	0	0	0	0
N 38	7	1	0	0	0	0	0	0
N 39	5	1	0	0	0	0	0	0
N 40	6	1	0	0	0	0	0	0
N 41	13	1	0	0	0	0	0	0
N 42	9	1	0	0	0	0	0	0
N 43	10	1	0	0	0	0	0	0
N 44	19	4	0	3	0	0	0	0
N 45	24	4	0	1	1	0	0	0
N 46	65	5	0	0	0	2	1	0
N 47	16	7	0	0	0	4	1	0
N 48	28	7	0	0	0	0	1	5
N 49	80	9	0	0	0	1	4	3
N 50	28	13	12	0	0	0	0	0
N 51	27	16	0	0	1	2	12	0
N 52	46	35	1	18	0	14	0	0
N 53	91	148	32	0	114	0	0	0
N 54	130	185	1	181	0	1	0	0
N 55	236	495	485	1	8	0	0	0

TABLEAU DES CONTRIBUTIONS RELATIVES  
DES NOEUDS AUX FACTEURS#1000

PRJ	NIVEAU	AXES FACTORIELS					
		1	2	3	4	5	6
N 29	0	0	0	0	0	0	0
N 30	1	0	0	0	0	0	5
N 31	0	0	0	0	0	0	0
N 32	3	0	0	0	0	0	17
N 33	2	0	0	2	0	0	1
N 34	1	0	0	0	0	3	5
N 35	11	0	0	0	1	0	11
N 36	3	0	0	0	0	5	0
N 37	5	0	0	3	0	1	0
N 38	7	1	0	0	0	13	4
N 39	5	1	1	0	0	1	10
N 40	6	1	0	0	3	1	0
N 41	13	1	0	0	0	0	5
N 42	9	1	0	1	0	31	3
N 43	10	1	0	2	1	3	35
N 44	19	4	0	14	0	25	24
N 45	24	4	1	5	13	0	22
N 46	65	5	0	1	4	87	74
N 47	16	7	0	2	0	144	77
N 48	26	7	1	0	2	2	41
N 49	80	9	0	2	4	38	161
N 50	28	13	22	1	3	2	1
N 51	27	16	0	0	10	98	488
N 52	46	35	3	86	2	486	16
N 53	91	148	60	1	888	0	1
N 54	130	185	3	863	0	49	6
N 55	236	495	899	5	62	0	0

TABLEAU DES CONTRIBUTIONS RELATIVES  
DES FACTEURS AUX NOEUDS#1000

PRJ	NIVEAU	AXES FACTORIELS						
		1	2	3	4	5	6	
N 29	0	0	677	191	71	13	37	7
N 30	1	0	182	58	18	83	186	480
N 31	0	0	356	30	128	71	385	17
N 32	3	0	103	8	0	53	118	714
N 33	2	0	11	895	10	44	0	37
N 34	1	0	102	87	19	336	450	2
N 35	11	0	33	182	249	0	522	12
N 36	3	0	520	237	20	178	0	44
N 37	5	0	98	824	18	50	0	6
N 38	7	1	41	154	118	378	98	207
N 39	5	1	757	2	0	26	212	0
N 40	6	1	401	107	335	28	2	126
N 41	13	1	349	4	71	4	109	461
N 42	9	1	201	142	22	527	49	56
N 43	10	1	15	316	118	57	472	18
N 44	19	4	41	661	8	155	133	0
N 45	24	4	156	282	390	2	128	38
N 46	65	5	15	51	93	444	332	63
N 47	16	7	39	79	0	582	280	7
N 48	26	7	100	0	33	11	136	718
N 49	80	9	4	57	56	111	414	355
N 50	28	13	930	27	30	5	3	1
N 51	27	16	0	0	80	169	739	8
N 52	46	35	56	519	10	400	11	1
N 53	91	148	222	2	774	0	0	0
N 54	130	185	9	980	0	7	0	0
N 55	236	495	981	2	16	0	0	0