

STATISTIQUE ET ANALYSE DES DONNÉES

THIERRY FOUCART

Sur les suites de tableaux de contingence indexés par le temps

Statistique et analyse des données, tome 3, n° 2 (1978), p. 67-85

http://www.numdam.org/item?id=SAD_1978__3_2_67_0

© Association pour la statistique et ses utilisations, 1978, tous droits réservés.

L'accès aux archives de la revue « Statistique et 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/>

Statistique et Analyse des Données

2 - 1978

SUR LES SUITES DE TABLEAUX DE CONTINGENCE

INDEXES PAR LE TEMPS

Thierry FOUCART

Assistant à l'Université de Haute Bretagne (I.U.T. de Vannes)

On dispose maintenant de méthodes permettant d'analyser une suite de tableaux de données quantitatives indexés par le temps : on étudie la tendance de la suite, définie par la suite des centres de gravité, et les structures des tableaux définies par les matrices des covariances.

Nous avons procédé de façon analogue pour analyser une suite de tableaux de contingence. Nous définissons la tendance de la suite et proposons différentes notions que peut recouvrir le mot structure. L'application de l'Analyse Factorielle des Correspondances et de la méthode STATIS introduite par Y. Escoufier et H. L'hermier des Plantes permet d'obtenir des référentiels communs, c'est-à-dire des systèmes d'axes sur lesquels on peut représenter certaines caractéristiques des tableaux.

NOTATIONS

Dans la première partie de l'article, on étudie un tableau de probabilités P_{IJ} , I désignant l'ensemble des lignes, et J l'ensemble des colonnes.

On pose :

$$n = \text{card } I$$

$$m = \text{card } J$$

$$P_I = \{p_{i.} ; i \in I\}$$

$$p_{i.} = \sum_{j=1}^m p_{ij}$$

$$P_J = \{p_{.j} ; j \in J\}$$

$$p_{.j} = \sum_{i=1}^n p_{ij}$$

} marges du tableau

$$\left. \begin{aligned} p_I^j &= \{ p_i^j ; i \in I \} \\ p_J^i &= \{ p_j^i ; j \in J \} \end{aligned} \right\} \text{Lois conditionnelles}$$

$$\left. \begin{aligned} p_I^J &= \{ p_I^j ; j \in J \} \\ p_J^I &= \{ p_j^i ; i \in I \} \end{aligned} \right\} \text{Tableau des lois conditionnelles} \\ \text{(rangées en colonnes)}$$

$$D_{P_I} = \begin{pmatrix} p_{1.} & & \\ & \ddots & \\ & & p_{n.} \end{pmatrix}$$

$$D_{1/P_I} = \begin{pmatrix} 1/p_{1.} & & \\ & \ddots & \\ & & 1/p_{n.} \end{pmatrix}$$

$$D_{P_J} = \begin{pmatrix} p_{.1} & & \\ & \ddots & \\ & & p_{.m} \end{pmatrix}$$

$$D_{1/P_J} = \begin{pmatrix} 1/p_{.1} & & \\ & \ddots & \\ & & 1/p_{.m} \end{pmatrix}$$

Dans la deuxième partie, nous étudions une suite de tableaux de probabilités $[P_{IJ}^l]$ $l = 1, k$; les notations sont les mêmes que ci-dessus, à l'indice l près.

1 - NOTIONS DE TENDANCE ET DE STRUCTURE DANS UNE SUITE DE TABLEAUX DE PROBABILITES

Comme nous l'avons dit dans l'introduction, il faut préciser les notions de tendance et de structure dans une suite de tableaux de probabilités pour pouvoir s'interroger sur leur évolution et en tenir compte dans les prévisions.

En fait, un tableau de probabilité sur un produit cartésien $I \times J$ donne deux nuages de points définis par les lois conditionnelles, en lignes d'une part, et en colonnes d'autre part. On notera :

- $N(I)$ le nuage des lois conditionnelles (P_J^i) , muni du système de masses p_I
- $N(J)$ le nuage des lois conditionnelles (P_I^j) , muni du système de masses p_J

1.1 - Les tendances

Les nuages $N(I)$ et $N(J)$ admettent pour centres de gravité les lois P_J et P_I respectivement. Il est donc naturel d'associer à une suite de tableaux de probabilités indexés par le temps, définis sur un même produit cartésien deux tendances définies respectivement par les suites des lois marginales P_J et P_I .

Prévoir la tendance d'une suite de tableaux de probabilités, c'est donc prévoir les marges.

1.2 - Les structures

En analyse factorielle des correspondances, à partir des facteurs principaux et des valeurs propres associées, on reconstruit le tableau des données par la formule ([1]) :

$$p_{ij} = p_{i.} \cdot p_{.j} \left[1 + \sum_{\alpha} \sqrt{\lambda_{\alpha}} u_{\alpha}^i v_{\alpha}^j \right]$$

u_{α} et v_{α} étant les facteurs associés à la valeur propre λ_{α} .

Considérer les coefficients $\frac{p_{ij}}{p_{i.} \cdot p_{.j}}$, que les économètres appellent coefficients de structure (Léontieff) ne revient pas à s'affranchir des tendances car il est bien connu ([2]) que la donnée de ces coefficients permet de retrouver les marges.

On peut remarquer également que les bases de facteurs (u_{α}) et (v_{α}) sont caractéristiques des marges, donc des tendances.

Développons maintenant différentes notions que peut recouvrir le mot structure ; on peut dire que les tableaux P_{IJ} et P'_{IJ} ont même structure si :

1.2.1 - Les valeurs propres et les facteurs sont égaux :

$$\lambda_{\alpha} = \lambda'_{\alpha} \quad u_{\alpha} = u'_{\alpha} \quad v_{\alpha} = v'_{\alpha}$$

On en déduit $P_I = P'_I$, $P_J = P'_J$ et donc $P_{IJ} = P'_{IJ}$ (par la formule de reconstitution des données).

Cela revient à dire que les structures sont les mêmes si les tableaux sont égaux.

1.2.2 - Les valeurs propres sont égales :

$$\lambda_\alpha = \lambda'_\alpha$$

les deux tableaux ont alors même "phi deux" $\left(\phi^2 = \frac{\chi^2}{n} : [3]\right)$; seule compte la décomposition de ce phi-deux suivant les valeurs propres. Les formes des nuages se ressemblent (à une rotation près puisque les axes n'ont pas en général les mêmes directions).

1.2.3 - Les facteurs sont égaux :

$$u_\alpha = u'_\alpha \quad v_\alpha = v'_\alpha$$

Les marges sont alors égales. La structure des tableaux est déterminée par la direction des axes ; c'est l'interprétation des axes qui est ici privilégiée.

1.2.4 - Les valeurs propres et un système de facteurs sont égaux :

$$\lambda_\alpha = \lambda'_\alpha \quad u_\alpha = u'_\alpha \quad (\text{ou } v_\alpha = v'_\alpha)$$

On a alors $P_I = P'_I$; ici, on privilégie, au niveau de l'interprétation, un système d'axes et, puisque $\lambda_\alpha = \lambda'_\alpha$, le nuage $N(J)$ associé à P_{IJ} est identique au nuage $N'(J)$ associé à P'_{IJ} .

1.2.5 - Les nuages $N(J)$ et $N'(J)$ sont identiques :

$$P_I^J = P'^J_I$$

Les facteurs et les valeurs propres diffèrent d'un tableau à l'autre. Si l'on est en présence de tableaux d'échanges entre régions, par exemple, dire que l'on a identité des structures revient à dire que la part des exportations de la région i dans les importations de la région j est la même dans chaque tableau pour toutes les régions.

Cette notion de structure a été utilisée par Caussinus ([4]) dans son introduction au modèle Markovien.

1.2.6. - Les valeurs propres sont égales et un des tableaux des lois conditionnelles est constant :

$$\lambda_\alpha = \lambda'_\alpha \quad P_I^J = P'^J_I$$

Comme en 1.2.2, les tableaux ont même phi-deux ; l'interprétation de l'égalité $p_i^j = p'^j_i$ est la même qu'en 1.2.5. Mais la fonction permettant de passer du système d'axes (u_α) au système (v_α) est la même pour chaque tableau :

$$P_I^J u_\alpha = \sqrt{\lambda_\alpha} v_\alpha$$

(formule de transition : [1]).

2 - NOTIONS DE REFERENTIEL COMMUN ASSOCIE A UNE SUITE DE TABLEAUX DE PROBABILITES

Un référentiel commun, associé à une suite dans le temps de tableaux de probabilités définis sur un même ensemble $I \times J$, est un système d'axes permettant de représenter simultanément certaines caractéristiques de ces tableaux. Trois types de référentiels sont ici introduits ; ils permettent :

- d'analyser l'évolution des tendances (1^{er} référentiel)
- de représenter les nuages $N(I)$ et $N(J)$ (2^{ème} référentiel)
- de représenter les couples (i, j) (3^{ème} référentiel)

De plus, on discute de la méthode STATIS, introduite par Y. Escoufier et l'Hermier des Plantes ([5]), qui permet d'obtenir une famille de référentiels.

Chaque méthode a été appliquée aux tableaux d'échanges de blé entre 16 régions françaises, de 1967 à 1974 (sauf 1968) ([2] et [6]) ; les résultats sont donnés à la suite de l'exposé de la méthode correspondante.

La définition géographique des 16 régions figure en annexe.

Dans certains cas, il a fallu réduire les tableaux à 8 régions (la définition de ces 8 régions est également donnée en annexe).

2.1 - Premier référentiel commun

Nous avons défini en 1.1 la tendance d'une suite de tableaux de probabilités à l'aide des marges, c'est-à-dire ici des suites :

$$(P_I^1)_{l=1,k} \quad \text{et} \quad (P_J^1)_{l=1,k} \quad (\text{centres de gravité}).$$

La suite $(P_I^1)_{l=1,k}$ est une suite de lois de probabilités définies sur le même ensemble I ; pour mesurer les proximités entre ces lois, il est naturel d'utiliser la distance du chi-deux de centre :

$$P_I = \frac{1}{k} \sum_{l=1}^k P_I^l$$

L'analyse factorielle sur le tableau des distances obtenu est équivalente à l'analyse factorielle des correspondances appliquée sur le tableau ci-après :

P_{IJ}

P_{IT}

$$\begin{bmatrix} p_{1.}^1 & p_{2.}^1 & p_{n.}^1 \\ p_{1.}^2 & p_{2.}^2 & p_{n.}^2 \\ \vdots & \vdots & \vdots \\ p_{1.}^k & p_{2.}^k & p_{n.}^k \end{bmatrix}$$

La ligne 1 du tableau correspondant à la marge $p_{1.}^1$.

Nous obtenons donc, à l'aide d'une analyse factorielle des correspondances, notre premier référentiel qui permet :

- une représentation de l'évolution de la tendance considérée (les lignes) ;
- une représentation de chaque élément de la tendance (les colonnes).

L'interprétation des proximités entre lignes et colonnes permet de déterminer quels sont les éléments qui contribuent à l'évolution de la tendance.

Application aux tableaux d'échanges de blé (16 régions) :

Nous avons analysé par la méthode ci-dessus la tendance associée aux exportations. Les marges correspondant à chaque année sont notées P_{67} , P_{69} , P_{70} ... Le long du premier axe, que l'on peut appeler "axe des temps", la tendance évolue régulièrement ; les proximités de Nancy et Orléans avec les années 72, 73, 74 montrent que les parts de ces régions dans les exportations totales augmentent, contrairement par exemple à celles de Lille et Amiens (voir Figure n° 1 page suivante).

2.2 - Deuxième référentiel commun

On sait que le tableau de probabilité le plus proche des tableaux P_{IJ}^l , au sens de toute distance euclidienne, est le tableau moyen $\pi_{IJ} = \frac{1}{k} \sum_l P_{IJ}^l$.

L'analyse factorielle des correspondances effectuée sur le tableau π_{IJ} donne une représentation des lois conditionnelles en lignes et en colonnes :

$$\left(\pi_J^i \right)_{i=1,n} \quad \text{et} \quad \left(\pi_I^j \right)_{j=1,m}$$

Les axes principaux définissent un "référentiel commun" ; on projetera dans le système des axes principaux obtenus les lois conditionnelles :

$$\left(p_J^{il} \right)_{i=1,n}, \quad \left(p_I^{jl} \right)_{j=1,m}$$

pour chaque valeur de l .

$\frac{1}{|T|} P_{IJ}$

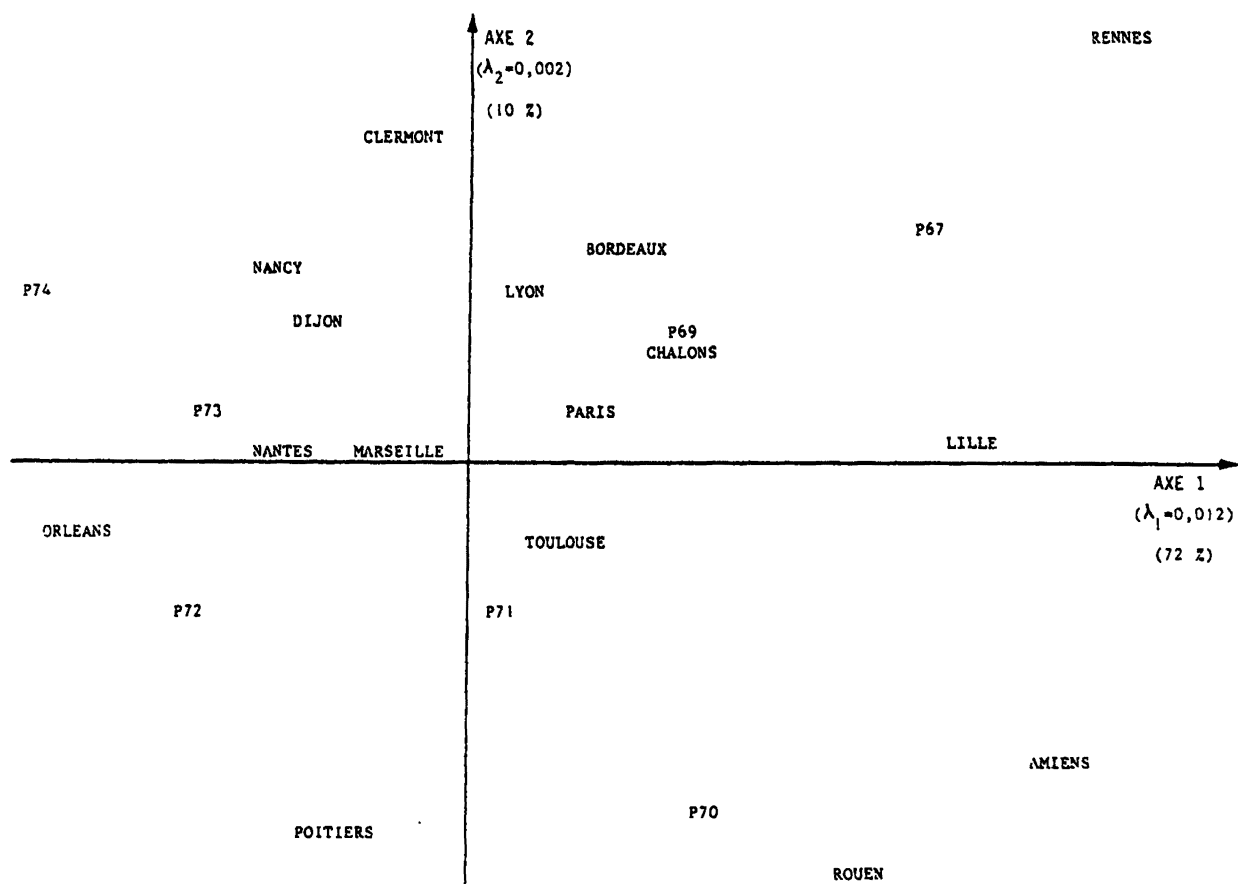
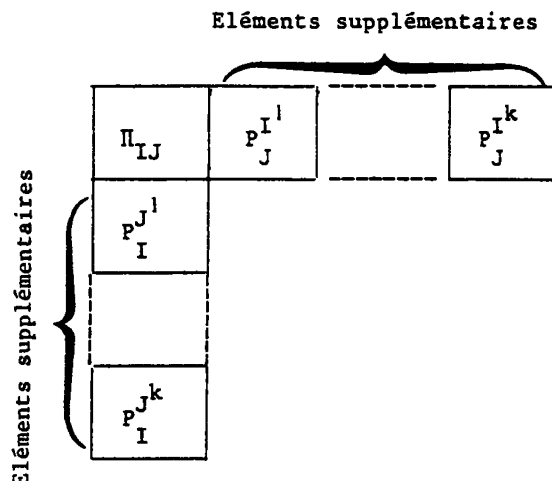


Figure n° 1 - Analyse factorielle des correspondances - Premier référentiel : évolution des exportations (16 lignes).

Cette procédure est à comparer à celle consistant à effectuer une analyse factorielle des correspondances sur les tableaux de probabilités juxtaposés en ligne ou en colonne ; on a l'avantage ici de traiter de façon simultanée et symétrique les deux nuages de lois conditionnelles.

$$\begin{matrix} \sqrt{P_{11}} & P_{12} \\ P_{21} & P_{22} \end{matrix}$$

8/5 P



Application aux tableaux d'échanges de blé (8 régions)

Le plan principal obtenu par l'analyse des correspondances est donné ci-après (Figure n° 2). Chaque région figure sur ce plan, à la fois comme exportatrice (nuage $N(I)$) et comme importatrice (nuage $N(J)$), les deux points étant pratiquement confondus pour toutes les régions.

On constate que les régions Orléans, Paris, Clermont, Lyon, sont très proches ; cela signifie que leurs profils, tant à l'exportation qu'à l'importation, sont très voisins, et que les échanges entre elles sont particulièrement importants.

Nous avons projeté en points supplémentaires les profils d'exportations et d'importations de chaque région pour chaque année. Ils se trouvent dans la zone délimitée autour de la région correspondante. Nous constatons donc sur ce schéma une grande stabilité des échanges, qui n'est pas du tout évidente lorsqu'on compare entre elles les analyses faites sur chaque tableau.

Pour expliquer cette instabilité qui apparaît entre les différentes analyses, nous avons effectué une analyse en composantes principales sur le tableau formé par les 5 premiers facteurs obtenus par chacune d'elles.

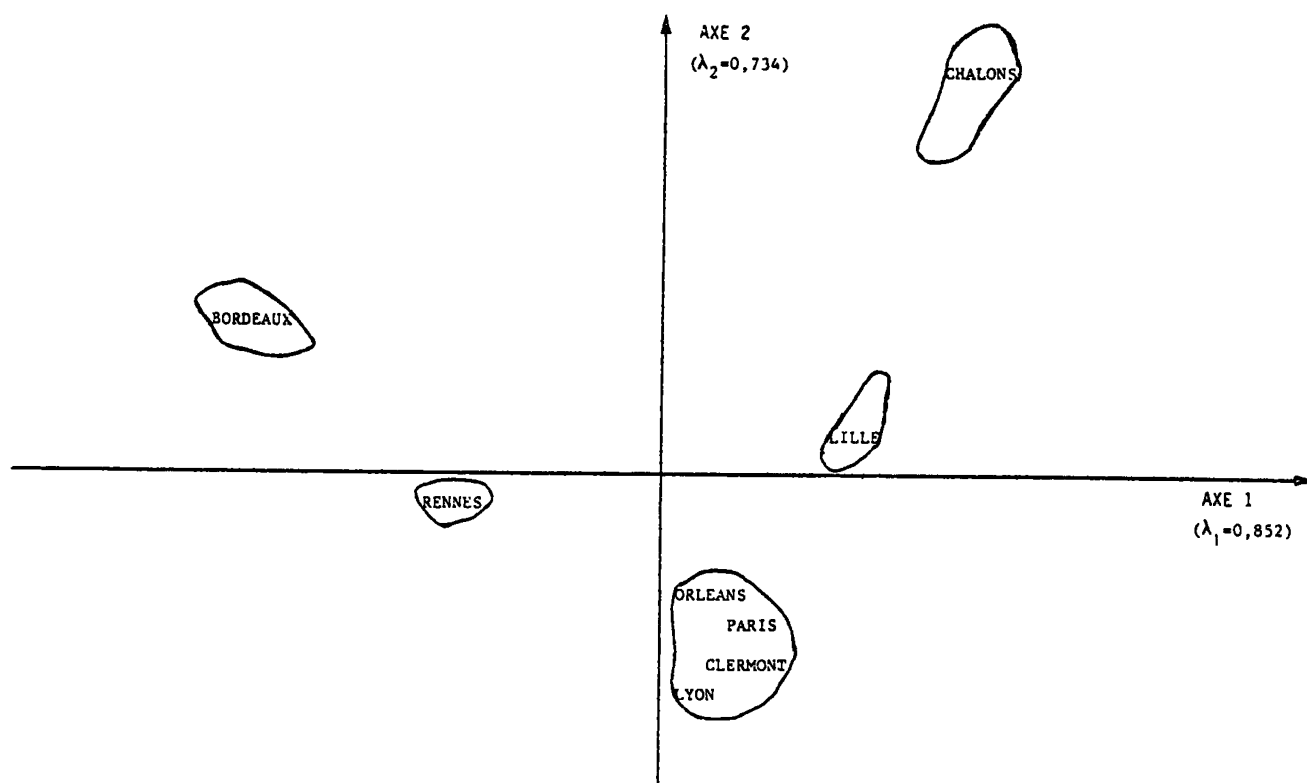


Figure n° 2 - Analyse des correspondances - Deuxième référentiel : analyse du tableau moyen

On constate sur le plan 1-2 (Figure n° 3) une grande stabilité des facteurs, sauf en 1972 et en 1973. L'examen des plans 1-3 et 2-3 ne permet pas d'agréger les 2e et 3e facteurs du tableau 1972 aux groupes déjà formés. Mais, on constate que le 2e facteur du tableau 1973 est bien dans le groupe des 3e facteurs des autres tableaux, et que, inversement, le 3e facteur du tableau 1973 est proche des 2e facteurs des autres années. La signification des facteurs ne change pas, sauf en 1972, les 2e et 3e valeurs propres permutant en 1973. On voit donc que la structure (au sens de la notion développée en 1.2.3) n'évolue pas ; seule celle du tableau 1972 présente des anomalies.

L'instabilité constatée dans les études successives résulte de l'anomalie de structure affectant l'année 1972 et de l'intervention pour l'année 1973 des 2e et 3e valeurs propres.

(voir Figure n° 3 page suivante)

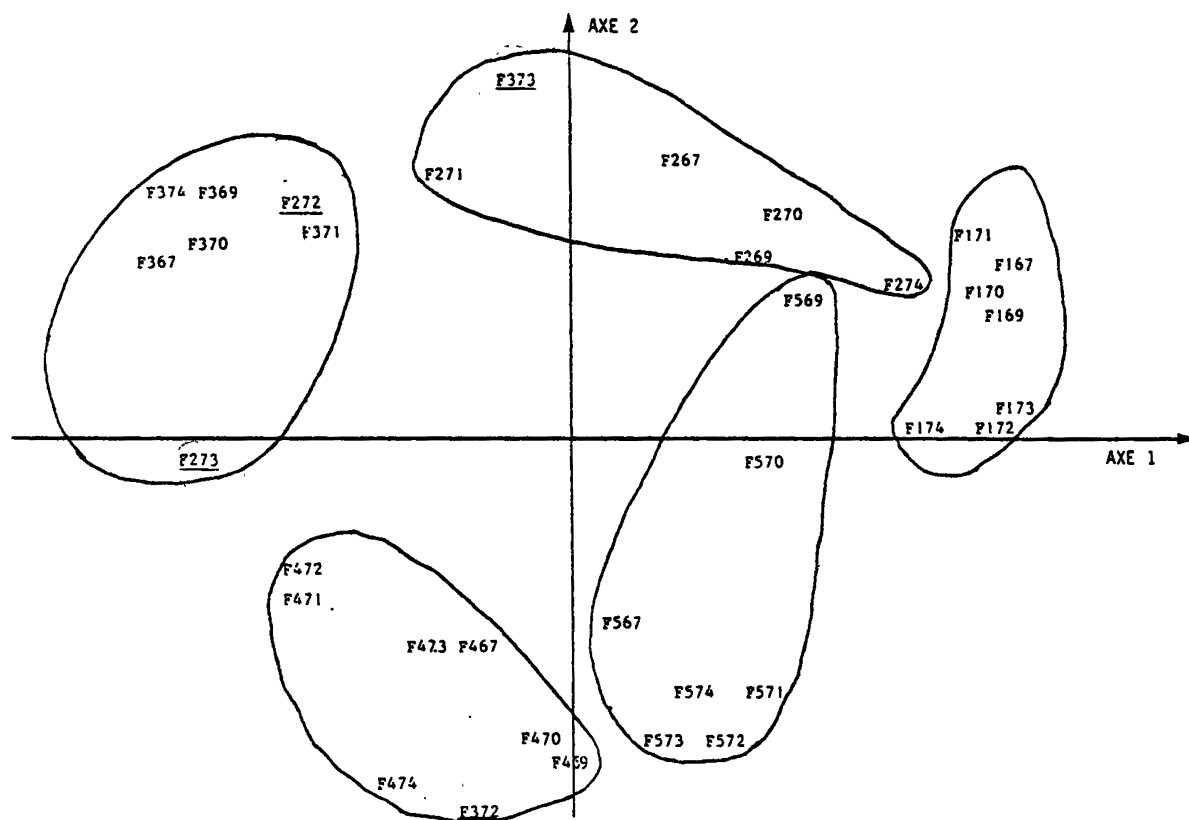


Figure n° 3 - Analyse en composantes principales : stabilité des facteurs.

2.3 - Troisième référentiel commun

Les tableaux (P_{IJ}^l) définissent des lois de probabilités sur $I \times J$. On peut donc, comme au § 2.1, calculer la distance du χ^2 de centre $\Pi_{IJ} = \frac{1}{k} \sum_l P_{IJ}^l$ entre ces tableaux ; l'analyse factorielle sur ce tableau des distances est équivalente à l'analyse des correspondances sur le tableau ci-dessous :

$$\begin{bmatrix} p_{1,1}^1 & \dots & p_{i,j}^1 & \dots & p_{n,m}^1 \\ p_{1,1}^2 & \dots & p_{i,j}^2 & \dots & p_{n,m}^2 \\ \vdots & & \vdots & & \vdots \\ p_{1,1}^k & \dots & p_{i,j}^k & \dots & p_{n,m}^k \end{bmatrix}$$

Ce tableau est constitué en "empilant" les tableaux P_{IJ}^l écrits en lignes.

L'analyse des correspondances donne une représentation simultanée des lignes, correspondant à chaque tableau, et des colonnes correspondant à l'évolution des termes $p_{i,j}^l$; le système d'axes ainsi obtenu définit le troisième référentiel commun.

Pour pouvoir juger de l'évolution des marges, on ajoutera en éléments supplémentaires les vecteurs :

$$p_{i.} = (p_{i.}^1)_{l=1,k} \quad \text{et} \quad p_{.j} = (p_{.j}^1)_{l=1,k}$$

L'interprétation des proximités se fait de la façon suivante :

- deux colonnes (i,j) et (i',j') sont proches si les p_{ij}^1 et $p_{i'j'}^1$, ont évolué de façon semblable au cours du temps.
- deux lignes l, l' sont proches lorsque les tableaux P_{IJ}^l et $P_{IJ}^{l'}$ sont peu différents au sens de la distance du χ^2 de centre Π_{IJ} .
- une colonne (i,j) sera d'autant plus proche d'une ligne l (aux homothéties près) que le terme $p_{i,j}^l$ est grand.

Application aux tableaux d'échanges de blé (8 régions)

L'interprétation des résultats obtenus par cette méthode en termes de tableaux d'échanges (cf. figure n° 4) est facile ; deux tableaux sont proches lorsque les échanges ont peu varié ; c'est le cas pour les tableaux des années 1969, 1970, 1971 ; deux colonnes sont proches lorsque les échanges ont évolué de la même façon : c'est le cas des exportations de Paris à Lyon, de Paris à Nancy. D'autre part, les exportations de Lille à Bordeaux et de Rennes à Lille ont augmenté de 67 à 74.

LYON — LILLE

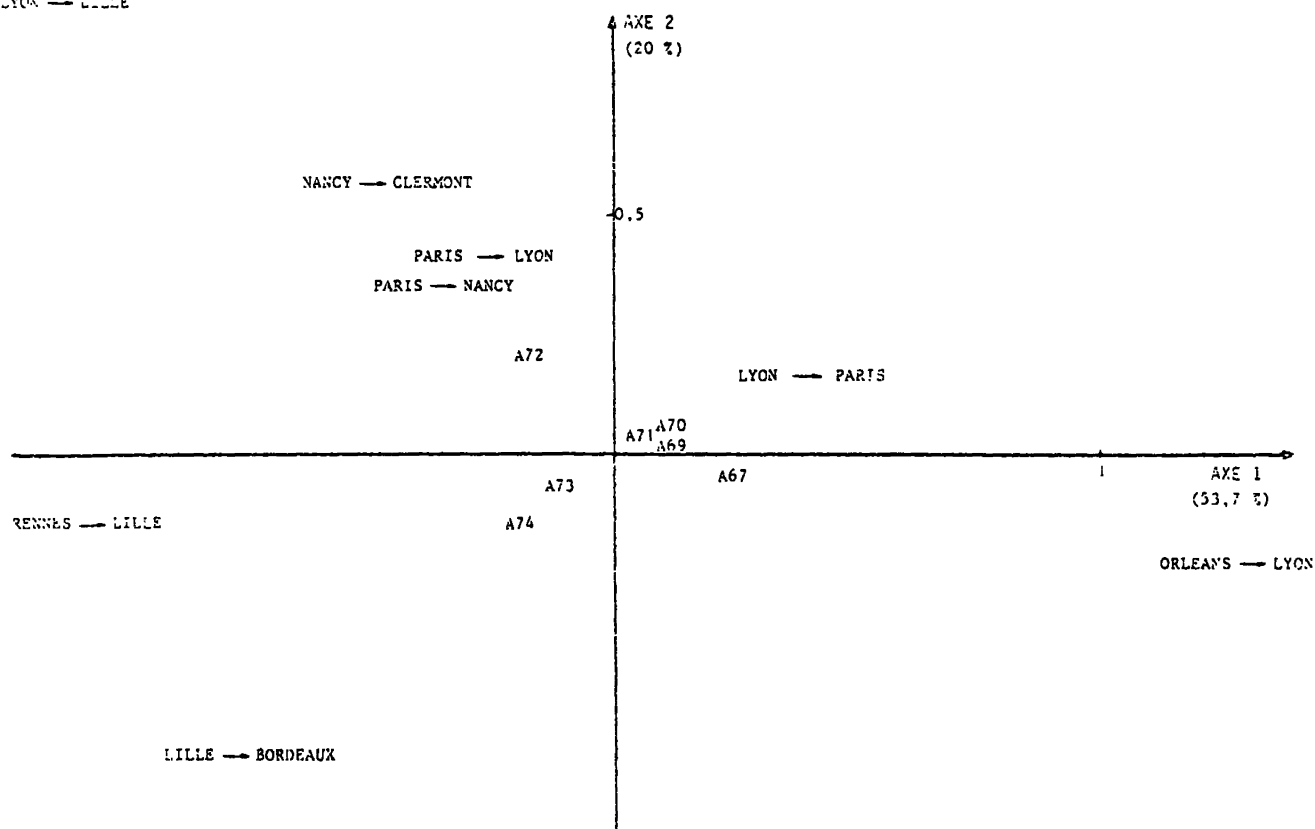


Figure n° 4 : Analyse factorielle des correspondances - Troisième référentiel : tableaux mis en lignes et empilés.

L'axe 1, que l'on peut appeler axe des temps, donne une représentation de l'évolution des tableaux ; la particularité de l'année 1972 s'explique par une augmentation soudaine des exportations de Lyon vers Lille et de Nancy vers Clermont.

On remarquera que le point représentatif des exportations d'Orléans à Lyon se trouve très éloigné le long de l'axe 1 ; ceci ne traduit pas une évolution significative de l'échange : l'échange, faible en 1967, est devenu nul les années suivantes (faible variation absolue, mais forte variation relative).

Remarque : Seuls certains échanges sont représentés sur la figure n° 4.

2.4 - A propos de la méthode STATIS

Cette méthode, introduite par Y. Escoufier et H. L'Hermier des Plantes ([5],[7]), s'appuie sur les propriétés des opérateurs associés à différents tableaux de données de type individus x caractères. Elle peut être utilisée de deux façons différentes, suivant les tableaux étudiés.

2.4.1 - Les deux aspects de la méthode

Quand on a à comparer des tableaux individus x caractères, deux cas se présentent :

- Premier cas : les individus considérés sont les mêmes pour tous les tableaux ; ces individus sont décrits, quand on passe d'un tableau à l'autre, par des ensembles de caractères différents.
- Deuxième cas : chaque tableau correspond à un lot différent d'individus décrits par un même ensemble de caractères.

Considérons le schéma de dualité associé au triplet (X, M, D_p)

$$\begin{array}{ccc}
 E = \mathbb{R}^P & \xleftarrow{X} & F^* \\
 \begin{array}{c} \downarrow M \\ \downarrow V \end{array} & & \begin{array}{c} \uparrow D_p \\ \uparrow W \end{array} \\
 E^* & \xrightarrow{X'} & F = \mathbb{R}^P
 \end{array}$$

Les opérateurs $W_o D_p$ et $V_o M$ qui sont respectivement D_p^- et M^- symétriques sont caractéristiques du triplet (X, M, D_p) .

Quelles que soient les métriques M considérées, comparer des tableaux individus x caractères revient, si on se trouve dans le premier cas, à comparer des opérateurs $W_o D_p$ (opérateurs U_q). Ces opérateurs ne sont autres que les projecteurs sur les sous espaces X' (E^*) si les métriques considérées sont les métriques de Mahalanobis $M = V^{-1}$.

Pour comparer ces opérateurs, Y. Escoufier a proposé la procédure suivante :

- a) Diagonaliser la matrice P des produits scalaires entre opérateurs. Rappelons que le produit scalaire entre deux opérateurs $U_1 = W_1 \circ D_p$ et $U_2 = W_2 \circ D_p$ est défini à partir de la trace (tr) de leur produit de composition :

$$P(U_1, U_2) = \text{tr}(U_1 \circ U_2)$$

Diagonaliser P revient à faire une analyse en composantes principales sur l'ensemble des opérateurs U_ℓ , considérés comme caractères ; en procédant ainsi, on obtient une base d'opérateurs (en composantes principales) dans laquelle les opérateurs U_ℓ peuvent être représentés (représentation de l'inter-structure).

b) Considérer la première "composante principale" $U = \sum \alpha_\ell U_\ell$.

Les coefficients α_ℓ , coordonnées du premier facteur principal, sont tous positifs, les éléments de la matrice P étant positifs. L'opérateur U_ℓ , opérateur "le plus représentatif" des opérateurs U_ℓ car il rend maximum la quantité

$$\sum_\ell P \left(\sum_\ell \beta_\ell U_\ell, U_\ell \right) \quad \left(\text{avec } \sum \beta_\ell^2 = \sum \alpha_\ell^2 \right)$$

est appelé "opérateur compromis".

c) Diagonaliser l'opérateur compromis U .

Dans le système des deux premiers vecteurs propres de U (composantes principales), on peut représenter l'ensemble de toutes les variables considérées.

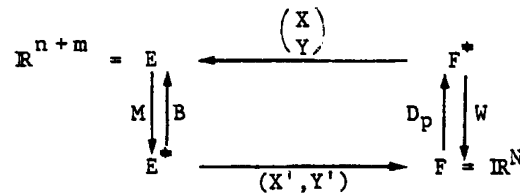
De façon duale, les coordonnées de ces deux composantes principales définissent les coordonnées des individus dans le système des axes principaux "compromis" ; dans ce système, on peut représenter autant de fois les individus que l'on considère de tableaux (représentation des intra-structures).

L'utilisation de l'opérateur compromis U , dans le cas où les opérateurs U_ℓ sont des projecteurs (métriques de Mahalanobis), paraît tout à fait justifiée ; on montre que, dans le cas de variables qualitatives, diagonaliser U revient à faire une analyse factorielle des correspondances optimale ([9]).

Pour procéder de la même façon dans le deuxième cas, il est nécessaire que la métrique définie sur E soit la même pour chaque tableau ; soit la métrique sera définie sur la population totale (métrique D_{1/σ^2} par exemple, où les écarts-type sont calculés dans la population totale), soit on choisira pour métrique la métrique identité. La procédure décrite permet alors de comparer des matrices de variances covariances, des matrices de corrélations, ... ou des tableaux de probabilités.

2.4.2 - Application à l'étude d'une suite de tableaux de contingence ou de probabilités

On sait que l'analyse factorielle des correspondances effectuée sur un tableau de contingence croisant deux variables qualitatives X et Y possédant respectivement n et m modalités, est équivalente à une analyse en composantes principales effectuée sur le tableau disjonctif complet $\begin{pmatrix} X \\ Y \end{pmatrix}$; les éléments de cette dernière analyse sont donnés par le schéma de dualité ci-dessous :



où :

- N est la somme du tableau de contingence (c'est-à-dire le nombre d'individus considérés).
- B est le tableau de Burt défini par

$$B = \begin{bmatrix} D_{P_I} & P_{IJ} \\ P'_{IJ} & D_{P_J} \end{bmatrix}$$

- M est la métrique définie matriciellement par :

$$M = \begin{bmatrix} D_{1/P_I} & 0 \\ 0 & D_{1/P_J} \end{bmatrix}$$

Pour effectuer l'analyse en composantes principales sur le tableau disjonctif complet, on choisit donc sur E une métrique qui dépend des individus considérés. Compte tenu de la remarque faite au paragraphe précédent, on ne peut appliquer directement la méthode STATIS.

Par contre, le tableau de Burt B est la matrice des moments non centrés d'ordre 2 de la variable $Z = (X_1, \dots, X_i, \dots, X_n, Y_1, \dots, Y_j, \dots, Y_m)$, où X_i (resp. Y_j) est la variable de Bernoulli associée à la i^e modalité de X (resp. à la j^e modalité de Y). Il est symétrique et contient toute l'information nécessaire pour effectuer l'analyse en composantes principales.

Aussi, pourquoi ne pas appliquer la méthode STATIS à la suite des tableaux de Burt B_ℓ ($\ell=1, \dots, k$), considérée comme une suite de matrices de variances-covariances ?

On obtiendra ainsi une représentation de la suite des tableaux de Burt par factorisation de la matrice P des produits scalaires définis comme précédemment à partir de la trace (l'interstructure). Le compromis obtenu s'écrit alors :

$$B_c = \sum_{\ell=1}^k \alpha_\ell B_\ell$$

où les coefficients α_ℓ sont, comme précédemment, les coordonnées du vecteur propre associé à la plus grande valeur propre de la matrice P .

Si l'on applique rigoureusement la méthode STATIS, on est conduit à factoriser le compromis B_c pour trouver le référentiel commun. Cela n'étant pas équivalent à effectuer une analyse factorielle des correspondances, on procèdera différemment :

On choisira le vecteur propre $(\alpha_1, \alpha_2, \dots, \alpha_k)$ de façon que le compromis B_c possède la structure d'un tableau de Burt $\left(\sum_{\ell} \alpha_\ell = 1 \right)$. Le compromis B_c s'écrit alors :

$\left\{ \begin{array}{l} \text{STATIS} \\ \rightarrow MB_c \text{ diag} \\ \text{réf.} \end{array} \right.$
 $AFC \quad M \text{ diag } 1/B_c$

stable $\rightarrow$ de B_c et $B_c M$

AFS
de $B_c \rightarrow \text{diag}(11B_c) B_c \text{ et } B_c \text{diag}(11B_c) \quad 81$

$$B_c = \begin{bmatrix} D_{Q_I} & Q_{IJ} \\ Q'_{IJ} & D_{Q_J} \end{bmatrix}$$

Q_{IJ} étant un tableau de probabilités, dont les marges sont les lois D_{Q_I} et D_{Q_J} . Le référentiel commun est alors obtenu par diagonalisation de l'opérateur $O_c = B_c \cdot M_c$, où M_c est défini par :

$$M_c = \begin{bmatrix} D_{I/Q_I} & 0 \\ 0 & D_{I/Q_J} \end{bmatrix}$$

Cela revient à effectuer l'analyse factorielle des correspondances du tableau Q_{IJ} , les intra-structures étant obtenues en projetant en éléments supplémentaires les lignes et les colonnes de chaque tableau de contingence.

Application aux tableaux d'échanges de blé (16 régions)

Nous avons appliqué la méthode précédente à la suite des tableaux d'échanges de blé entre les 16 régions françaises.

On constate, en examinant la représentation sur le plan 1-2 de l'inter-structure (figure 5) une évolution régulière des opérateurs définis par les tableaux de Burt associés à chaque année, la particularité de l'année 1972 (cf. 2.2) étant mise en évidence.

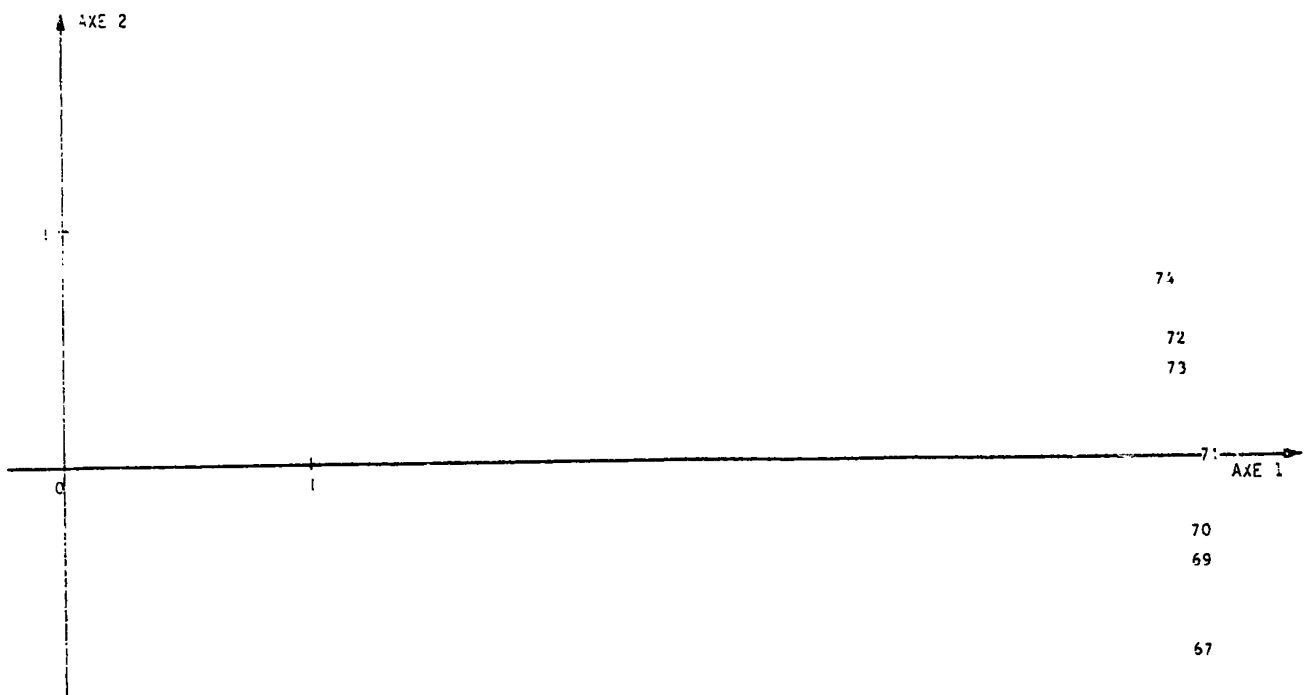


Figure n° 5 : Interstructure (Analyse des opérateurs)

La figure 6 ci-dessous donne le plan 1-2 obtenu en appliquant l'analyse factorielle des correspondances au tableau de probabilités Q_{IJ} calculé comme ci-dessus. Ce tableau est d'ailleurs très proche du tableau de probabilités moyen.

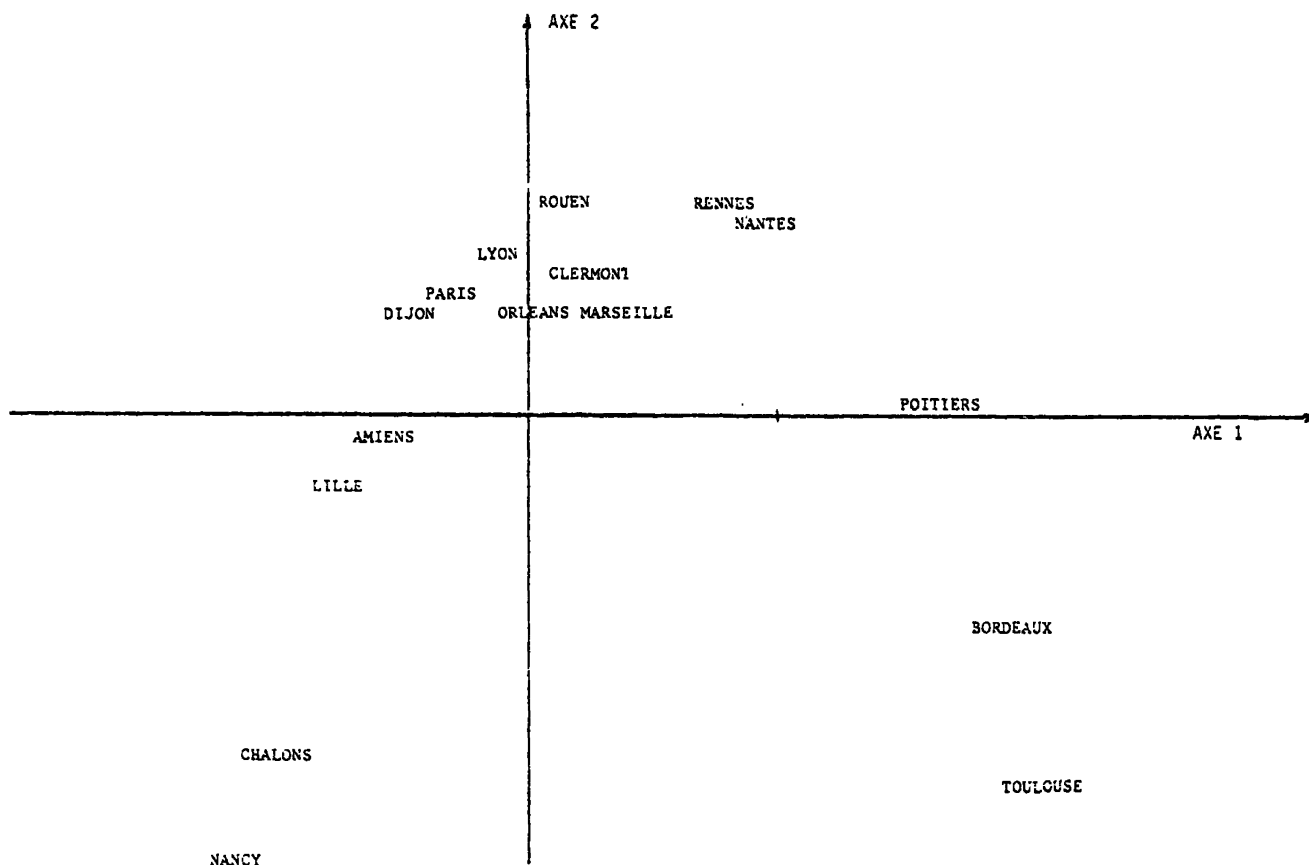


Figure n° 6 : Compromis (Analyse sur les tableaux de Burt)

EN GUISE DE CONCLUSION

Dans cet article, nous avons proposé des définitions de tendance et de structure susceptibles d'être utilisées dans l'étude des suites de tableaux de probabilités indexées par le temps. Si la technique simple d'analyse des évolutions des tendances repose effectivement sur la définition que nous en avons donnée, il n'en est malheureusement pas de même en ce qui concerne les techniques d'étude des évolutions de la structure : nous ne sommes pas partis des équivalences entre structures pour mettre au point les méthodes qui ont été décrites. Si ce manque de cohérence nuit à la qualité de notre exposé, il ne diminue en rien l'intérêt de ces équivalences et l'efficacité de ces méthodes.

BIBLIOGRAPHIE

- 1 - J.P. BENZECRI
 "L'Analyse des données - Tome 2 : l'Analyse Factorielle des Correspondances"
 Dunod 1973.
- 2 - C. MACCIA, E. STEMMELEN et Coll.
 "Contamination par la chaîne alimentaire. Prévision des échanges agricoles et calcul des doses délivrées"
 Rapport C.E.A. R 4885 - 1977.
- 3 - F. CAILLEZ, J.P. PAGES
 "Introduction à l'analyse des données"
 SMASH 1976
- 4 - H. CAUSSINUS
 "Quelques points de vue sur l'analyse des tableaux d'échanges"
 Annales de l'I.N.S.E.E. n° 22-23 ; 1976
- 5 - Y. ESCOUFIER
 "Opérateur associé à un tableau de données"
 Annales de l'I.N.S.E.E. n° 22-23 ; 1976
- 6 - E. STEMMELEN
 "Tableaux d'échanges : description et prévision"
 Cahiers du BURO, série Recherche n° 28. 1977.
- 7 - H. L'HERMIER des PLANTES
 "Structuration des tableaux à trois indices de la statistique : théorie et application d'une méthode d'analyse conjointe"
 Thèse de 3e Cycle présentée à l'Université des Sciences et Techniques du Languedoc
 Montpellier - 1976.
- 8 - P. CAZES, A. BAUMERDER et Coll.
 "Codage et analyse des tableaux logiques. Introduction à la pratique des variables qualitatives"
 Cahiers du BURO, série Recherche n° 27 - 1977
- 9 - P. CAZES, S. BONNEFOUS et Coll
 "Description cohérente des variables qualitatives prises globalement et de leurs modalités"
 Statistiques et Analyse des données, 2e Semestre 1976.

ANNEXE

Nous donnons ci-dessous la liste des régions, les regroupements qui ont été faits, et la carte des régions.

- 1 - Bordeaux
- 2 - Clermont Ferrand
- 3 - Dijon
- 4 - Lille
- 5 - Amiens
- 6 - Lyon
- 7 - Marseille
- 8 - Chalons sur Marne
- 9 - Nancy
- 10 - Rennes
- 11 - Nantes
- 12 - Orléans
- 13 - Paris
- 14 - Poitiers
- 15 - Rouen
- 16 - Toulouse

Nouvelles régions	Régions réunies
1 - Orléans	Orléans
2 - Paris	Paris-Rouen
3 - Rennes	Rennes-Nantes
4 - Bordeaux	Bordeaux-Poitiers-Toulouse
5 - Nancy	Chalons-Nancy
6 - Lille	Lille-Amiens
7 - Clermont	Clermont-Dijon
8 - Lyon	Lyon-Marseille

DIVISION REGIONALE (d'après l'O.N.I.C.)

