REVUE DE STATISTIQUE APPLIQUÉE

A. FARAJ

Analyse de contiguïté : une analyse discriminante généralisée à plusieurs variables qualitatives

Revue de statistique appliquée, tome 41, n° 3 (1993), p. 73-84

<http://www.numdam.org/item?id=RSA_1993__41_3_73_0>

Rev. Statistique Appliquée, 1993, **XLI** (3), 73-84

ANALYSE DE CONTIGUITÉ : UNE ANALYSE DISCRIMINANTE GÉNÉRALISÉE À PLUSIEURS VARIABLES QUALITATIVES

A. Faraj

Institut Français du Pétrole
1-4, Av. de Bois-Préau
92506 Rueil Malmaison Cedex

RÉSUMÉ

Nous traitons ici un couple de tableaux de données définies sur un même ensemble d'individus qui en constituent les lignes. Le premier croise ces individus avec des variables quantitatives et le deuxième les croise avec des variables qualitatives.

Les variables qualitatives induisent des partitions sur l'ensemble des individus. Nous nous intéressons à étudier les effets des variables qualitatives sur les valeurs prises par les variables quantitatives. Nous proposons un indice d'évaluation de la pertinence des variables qualitatives quant à cet effet. Nous rechercherons, dans un espace de dimension réduite, une représentation des individus aussi proche que possible des partitions définies par les variables qualitatives les plus significatives.

Notre démarche est basée sur les méthodes proposées par L. Lebart [7] pour la prise en compte de structures de contiguïté et par Y. Le Foll [8] concernant la pondération des couples de points.

Nous pondérons chaque individu par l'effectif de ses contigus. Ensuite nous centrons et réduisons les variables quantitatives relativement à cette pondération. C'est sur ces mesures centrées et réduites que nous effectuons l'analyse. La matrice des variances-covariances coïncide alors avec la matrice des corrélations pondérées. Ceci nous permettra de diminuer l'importance des individus isolés, mais surtout de retrouver les équations de l'analyse discriminante dans le cas d'une seule variable qualitative. Notre méthode apparaît alors comme une généralisation de l'analyse discriminante à plusieurs variables qualitatives.

L'Analyse Canonique des Correspondances [13] telle qu'elle est décrite par D. Chessel et al. [5] pourrait s'appliquer à notre problème. Elle donne, d'un point de vue formel tout au moins, des équations comparables aux notres. Cependant elle n'exploite pas les mêmes pondérations que celles utilisées par notre méthode.

Mots clés : *Analyse de contiguïté, ACP, Analyse discriminante, Pondération, Graphe de contiguïté, Covariance locale.*

SUMMARY

Methods are investigated for handling a couple of tables, the lines of each being cases issued from a unique set. In the first table, cases are crossed with quantitative variables whereas qualitative in the second.

Qualitative variables induce partitions of the cases set. Influence of qualitative variables on quantitative ones are studied and a criterion is proposed measuring its relevance. A representation of the individuals in a reduced dimensionnal space is presented, expected to be as close as possible to the partitions defined by the most relevant qualitative variables.

Our process is based, on the one hand, on methods proposed by L. Lebart (see [7]) for handling contiguity structures and, on the other hand, on Y. Le Foll's works (see [8]) on the weightings of couple of points.

In order to calculate variances/covariances, each individual is weighted by the number of its neighbours, what leads to diminish the isolated individuals importance and moreover, to recover the discriminant analysis formalism (in the one-qualitative-variable case). Our method then appears as a generalization of discriminant analysis to the multiple variables case.

Canonical correspondance analysis (see Ter Braak [13]), as described by D. Chessel et al. (see [5]) could be applied to our problem, as far as both formalisms are close.

Key Words : *Contiguity Analysis, PCA, Discriminant Analysis, Weightings, Contiguity graph, Local covariance.*

1. Covariance locale – Costabilité

On se donne un couple X et V de tableaux de données ayant le même nombre n de lignes que nous noterons i ($1 \leqslant i \leqslant n$) et qui constituent l'ensemble des individus. Le premier tableau croise ces individus avec J variables quantitatives x^j ($1 \leqslant j \leqslant J$) et le deuxième les croise avec K variables qualitatives v^k ($1 \leqslant k \leqslant K$).

Chacune des variables qualitatives v^k induit sur l'ensemble des individus un graphe de contiguïté $G^k = [g^k_{ii'}]$ de la façon suivante :

$$g^k_{ii'} = \begin{cases} 1 & \text{si } v^k_i = v^k_{i'} \\ 0 & \text{sinon} \end{cases}$$

De sorte que l'ensemble $\{v^k\}$ des variables qualitatives définit le graphe symétrique pondéré $G = [g_{ii'}] = \sum_{k=1}^{K} G^k$ où $g_{ii'} = \sum_{k=1}^{K} g^k_{ii'}$ (compris entre 0 et K) représente le nombre de variables qualitatives qui réunissent i et i' dans une même classe de modalité.

Nous désignerons par $m_k = \sum_{i=1}^{n} \sum_{i'=1}^{n} g^k_{ii'}$ le nombre de connexions du graphe G^k et par $m = \sum_{k=1}^{K} m_k$ le nombre total des connexions des K variables qualitatives.

La matrice des covariances locales du tableau X définie par Lebart [7] relativement au graphe G sera notée Γ_{XX}. Son terme général $\Gamma_{jj'}$ qui représente la covariance locale entre les deux variables x^j et $x^{j'}$ est défini par :

$$\Gamma_{jj'} = \frac{1}{2m} \sum_{i=1}^{n} \sum_{i'=1}^{n} g_{ii'}(x_i^j - x_{i'}^j)(x_i^{j'} - x_{i'}^{j'}).$$

Si $\Gamma_{jj'}^k$ désigne le terme général de la matrice Γ_{XX}^k des covariances locales du tableau X relative au graphe G^k, nous avons la relation suivante :

$$\Gamma_{jj'} = \sum_{k=1}^{K} \frac{m_k}{m} \Gamma_{jj'}^k.$$

Soit matriciellement :

$$\Gamma_{XX} = \sum_{k=1}^{K} \frac{m_k}{m} \Gamma_{XX}^k.$$

La valeur Γ_{jj} du j-ème terme de la diagonale de Γ_{XX} correspond à la variance locale de la variable x^j par rapport au graphe de contiguïté induit par l'ensemble des variables qualitatives.

Notons $n_i = \sum_{i'=1}^{n} g_{ii'}$ l'effectif local au «voisinage» de i, $p_i = \frac{n_i}{m}$ le taux de ses contigus, $N = [n_i]$ et $D_n = [p_i]$ les matrices carrées diagonales d'ordre n des n_i et des p_i respectivement.

Nous supposerons que chaque individu i a pour masse p_i, ce qui revient à pondérer i par le taux de ses contigus (au lieu de la pondération classique égale à $\frac{1}{n}$) et à adopter comme métrique des poids la métrique de matrice D_n.

Dans la suite, nous considérons que les variables x_j sont D_n-centrées et D_n-réduites. Nous prendrons la même notation X pour le tableau des données D_n-normées. Nous noterons C_{XX} la matrice des corrélations relativement à D_n.

Nous définissons *le coefficient de stabilité* d'une variable x^j par :

$$T_{jj} = \frac{1}{m} \sum_{i=1}^{n} \sum_{i'=1}^{n} g_{ii} x_i^j x_{i'}^j.$$

De même *le coefficient de costabilité* de deux variables x^j et $x^{j'}$ par :

$$T_{jj'} = \frac{1}{m} \sum_{i=1}^{n} \sum_{i'=1}^{n} g_{ii'} x_i^j x_{i'}^{j'}.$$

Notons que $T_{jj'}+\Gamma_{jj'} = \text{cor}(x^j, x^{j'})$[1] et plus particulièrement $T_{jj}+\Gamma_{jj} = 1$.

Si $T^k_{jj'}$ désigne le terme général de la matrice T^k_{XX} des costabilités du tableau X relative au graphe G^k et T^k_{jj} le coefficient de stabilité de la variable x^j par rapport à G^k, nous avons la relation suivante :

$$T_{jj'} = \sum_{k=1}^{K} \frac{m_k}{m} T^k_{jj'}.$$

En désignant par T_{XX} la matrice dont le terme général est $T_{jj'}$, nous avons les écritures matricielles suivantes :

$$T_{XX} = \sum_{k=1}^{K} \frac{m_k}{m} T^k_{XX}$$

et

$$T_{XX} + \Gamma_{XX} = C_{XX}.$$

Cette deuxième relation représente une écriture de l'égalité de Huyghens étendue aux classes empiétantes dans laquelle C_{XX} est calculée relativement à la métrique D_n. Lebart en propose une autre [7] généralisée par Mom dans sa thèse [9].

Remarque

Si $K = 1$, T_{XX} et G_{XX} correspondent respectivement aux matrices de covariances inter et intra-classes associées à la partition définie par la seule variable qualitative de l'analyse, chaque individu étant pondéré par la masse de la classe à laquelle il appartient.

Dans ce cas, si $\text{cov}_c(x^j, x^{j'})$ désigne la covariance usuelle entre les variables x^j et $x^{j'}$ dans la classe c, μ^j_c le centre de gravité de cette classe pour la variable x^j, et n_c son effectif, on a : $m = \sum_c n_c^2$, $\Gamma_{jj'} = \frac{1}{m}\sum_c n_c^2 \text{cov}_c(x^j, x^{j'})$ et $T_{jj'} = \frac{1}{m}\sum_c n_c^2 \mu^j_c \mu^{j'}_c$. On voit ainsi apparaître, du fait des poids adoptés, les pondérations $\frac{n_c^2}{m}$ au lieu des pondérations $\frac{n_c}{n}$ usuelles.

---

(1) En effet $T_{jj'} + \Gamma_{jj'} = \frac{1}{2m}\sum_{i=1}^{n}\sum_{i'=1}^{n} g_{ii'}[2x^j_i x^{j'}_{i'} + (x^j_i - x^j_{i'})(x^{j'}_i - x^{j'}_{i'})]$. Soit, compte tenu du fait que $g_{ii'} = g_{i'i}$:

$$T_{jj'}+\Gamma_{jj'} = \frac{1}{m}\sum_{i=1}^{n}\sum_{i'=1}^{n} g_{ii'} x^j_i x^{j'}_i = \frac{1}{m}\sum_{i=1}^{n} x^j_i x^{j'}_i \sum_{i'=1}^{n} g_{ii'} = \sum_{i=1}^{n} \frac{n_i}{m} x^j_i x^{j'}_i = \sum_{i=1}^{n} p_i x^j_i x^{j'}_i.$$

Et, puisque les variables sont centrées et réduites par rapport à $D_n = [p_i]$, $\sum_{i=1}^{n} p_i x^j_i x^{j'}_i$ est le terme général de la matrice C_{XX} des corrélations calculée relativement à D_n.

2. La méthode

Une variable x^j pour laquelle le coefficient $T_{jj} \approx 1$ présente en moyenne de faibles dispersions (ou variabilités) à l'intérieur des classes induites par les variables qualitatives. Par contre, si $T_{jj} \approx 0$, x^j présentera de fortes variabilités à l'intérieur de ces classes.

Nous appliquons l'analyse de la contiguïté (cf [7]). Elle consiste à rechercher une variable f (composante principale), combinaison linéaire des variables initiales x^j ($f = Xu$), pour laquelle la variance locale est minimale. Ce qui est équivalent à rechercher celle pour laquelle le rapport de stabilité $\frac{T_f}{\sigma_f^2} = \frac{u^t T_{XX} u}{u^t C_{XX} u}$ est maximal.

Ceci revient à déterminer les vecteurs propres $(u^l)_{l=1,\ldots,J}$ de la matrice $C_{XX}^{-1} T_{XX}$ associés aux valeurs propres $\lambda_1 \geqslant \lambda_2 \geqslant \ldots \geqslant \lambda_J \cdot \lambda_l$ – coefficient de stabilité de la l-ème composante – s'écrit :

$$\lambda_l = \frac{T_l}{\sigma_l^2} = \frac{(u^l)^t T_{XX} u^l}{(u^l)^t C_{XX} u^l}$$

ou encore $\lambda_l = \sum_{l=1}^{K} \frac{m_k}{m} \frac{(u^l)^t T_{XX}^k u^l}{(u^l)^t C_{XX} u^l}$.

$\frac{1}{\lambda_l} \frac{m_k}{m} \frac{(u^l)^t T_{XX}^k u^l}{(u^l)^t C_{XX} u^l}$ est la contribution de la variable v^k à la construction de la l-ème composante. $R_l^k = \frac{(u^l)^t T_{XX}^k u^l}{(u^l)^t C_{XX} u^l}$ coïncide – pour $K = 1$ – avec le rapport de la variance inter-classes de la partition induite par la variable v^k sur la variance totale de la l-ème composante et possède un «comportement» analogue à ce dernier pour $K \geq 2$. R_l^k est d'autant plus fort que les classes de cette partition sont homogènes et bien isolées pour les valeurs de la lème composante. Ceci nous conduit à une méthode d'élimination pas à pas des variables qualitatives les moins discriminantes pour ne garder que celles dont la ségrégation des mesures des variables quantitatives est significative. Nous conserverons comme variables qualitatives significatives celles pour lesquelles ce rapport – pour les premières composantes factorielles retenues – est proche de 1.

3. Décomposition du coefficient de stabilité

On montre que les composantes factorielles sont D_n-normées et leurs coefficients respectifs de stabilité égaux aux valeurs propres de $C_{XX}^{-1} T_{XX}$.

Soit $r_l^j = \text{cor}(x^j, f^l)$ le coefficient de corrélation de la j-ème variable initiale avec la l-ème composante, on montre que le coefficient de stabilité de x^j se met sous

la forme :

$$T_{jJ} = \sum_{l=1}^{J} (r_l^j)^2 \lambda_l.$$

Donc, du fait que les variables initiales sont D_n-normées (i.e. $\sum_{l=1}^{J} (r_l^j)^2 = \sigma_j^2 = 1$), le coefficient de stabilité d'une variable s'écrit sous la forme d'une somme pondérée des coefficients de stabilité des composantes factorielles. $(r_l^j)^2$ – carré du coefficient de corrélation – est la contribution de la l-ème composante factorielle à la stabilité de la variable x^j.

4. Un exemple d'application

Les données que nous traitons nous ont été communiquées par E. Stemmelen de la SOFRES. L'application de l'analyse de la contiguïté à ce type de données nous permet d'illustrer notre méthode. La collecte de ces données a été faite dans le cadre de la «Sémiométrie»[6]. Une liste de 66 mots est présentée à un échantillon de 970 élèves de grandes écoles auxquels il est demandé d'attribuer à chacun des mots une note d'agrément (allant de 1 à 7) selon qu'ils n'aiment pas ou qu'ils aiment le mot en question. Les réponses sont rangées dans un tableau où les 970 élèves sont en lignes et les 66 mots en colonnes.

Il existe, par ailleurs, des informations signalétiques a priori sur les élèves répertoriées en tant que variables qualitatives : type d'école (2 modalités de réponses : écoles d'ingénieurs ou de commerce), région (2 modalités : Paris/Région Parisienne ou Province), sexe (2 modalités), lieu où l'élève désire commencer à travailler (3 modalités : en France, à l'étranger ou «ne sait pas»), année de scolarité de l'élève (2 modalités : dernière année ou non).

Le but est de définir, dans un premier temps, un espace de représentation qui fournisse une organisation conceptuelle des 66 mots. On recherche, dans un deuxième temps, des inter-dépendances entre ces mots-concepts et les modalités des variables qualitatives signalétiques des individus. Quelles notes accorderait – par exemple – un élève de sexe masculin étudiant à Paris en dernière année d'école de commerce aux mots *Mariage* ou *Honneur*? Existe-t-il une relation entre ces modalités combinées et les notes attribuées aux mots indiqués?

L'analyse discriminante généralisée (ADG) apporte des éléments de réponse à l'analyse de telles données. Non seulement elle établit le modèle conceptuel des 66 mots en relation avec les variables qualitatives, mais elle nous fournit aussi des critères d'évaluation de l'effet de ces variables dans l'élaboration du modèle.

D'autres méthodes pourraient cependant être appliquées pour l'analyse de tels types de données. La plus simple – voire la plus classique – serait l'analyse en composantes principales. Elle déterminerait dans un premier temps l'espace de représentation des variables quantitatives. L'importance de l'effet des variables qualitatives serait illustrée, dans un deuxième temps, par la projection des centres de

classes de leurs modalités dans l'espace des individus. Les positions de ces centres permettent alors de les rattacher aux significations des axes factoriels (cf [6]). Notons que l'ACP ne donne aucune indication chiffrée ni sur l'importance relative des variables qualitatives quant à leur influence sur les valeurs prises par les variables quantitatives, ni sur la construction des axes factoriels.

Remarquons, pourtant, que l'absence de l'a priori des variables qualitatives dans l'utilisation de l'ACP pourrait, dans certains cas, constituer un avantage de celle-ci sur l'ADG. Elle laisse champ libre à l'interprétation des résultats vis-à-vis de nouvelles variables qualitatives. De ce point de vue, l'ADG offre moins de souplesse. Mais nous dirons simplement que la «meilleure» méthode est celle qui répond aux besoins de l'utilisateur. Pour lui permettre le choix entre l'une ou l'autre, nous les appliquerons succesivement à notre jeu de données et comparerons les résultats auxquels elles aboutissent. Nous ne préconiserons pas l'efficacité de l'une des deux analyses par rapport à l'autre. Bien qu'elles conduisent à des résultats fondamentalement différents, elles seraient plutôt complémentaires que concurrentes du fait qu'elles offrent, chacune à sa manière, un regard privilégié sur les données. C'est ce que nous nous efforcerons de montrer dans la suite de cet article.

Les résultats de l'ADG

	Axe 1	**Axe 2**
Type d'école (Ingénieurs/Commerce)	48 %	12 %
Région (Paris/Province)	4 %	2 %
Sexe	37 %	17 %
Lieu où l'élève désire commencer à travailler (en France/à l'étranger/NSP)	3 %	61 %
Année de scolarité de l'élève (Dernière ou pas dernière année)	8 %	8 %
Coefficients de stabilité des facteurs (ie : valeurs propres)	0,083	0,032

Les valeurs proches de 0 (0,083 et 0,032) des coefficients de stabilité des deux premiers facteurs montrent qu'il n'existe pas une typologie nette des élèves – selon les notes qu'ils attribuent aux mots – qui soit reliée aux variables qualitatives intervenant dans l'analyse. Cependant, nous pouvons considérer qu'il existe des effets de certaines de ces variables qualitatives sur les notes attribuées. Il s'agit en particulier du type d'école, du sexe et du lieu où l'élève désire commencer à travailler qui présentent des contributions assez élevées. Ceci montre les tendances qu'ont les individus de ces catégories à noter très fortement ou très faiblement certains mots.

La projection des centres de classes des modalités de toutes les variables (figure 1) montre en effet l'importance des trois variables en question et du moindre effet des deux autres (Région et Année de scolarité de l'élève). La figure 2 permet de rattacher

chacune de ces modalités à une famille de concept. Ainsi les filles auraient tendance à noter fortement les mots à caractère féminin («Mode» et «Bijou») et ceux en rapport avec le confort intérieur («Nid», «Mariage», «Moelleux», ...) ou ce qui s'y rapporte (tel que le mot «Propriété»). Les garçons tout en ayant moins de penchant vers ces mots auraient davantage tendance à surnoter des mots de morbidité ou d'instabilité tels que «Fusil», «Mort», «Angoisse», «Labyrinthe» et «Vide». Ce premier axe qui oppose *le plaisir de faire* (à droite) à l'*instabilité-Morbidité* (à gauche) sépare aussi les élèves des écoles de commerce à ceux des écoles d'ingénieurs.

Le deuxième axe est surtout caractérisé par la variable «Lieu où l'élève désire commencer à travailler» dont la modalité «à l'étranger» se trouve du côté des mots révélant une certaine curiosité de l'esprit tels que «Différent», «Étranger», «Interroger», «Aventurier», «Original», ... Alors que la modalité «en France» se trouve du côté des mots que nous avons défini autour du concept «Confort intérieur». Avec cette modalité, le mot «intérieur» prend son sens le plus large.

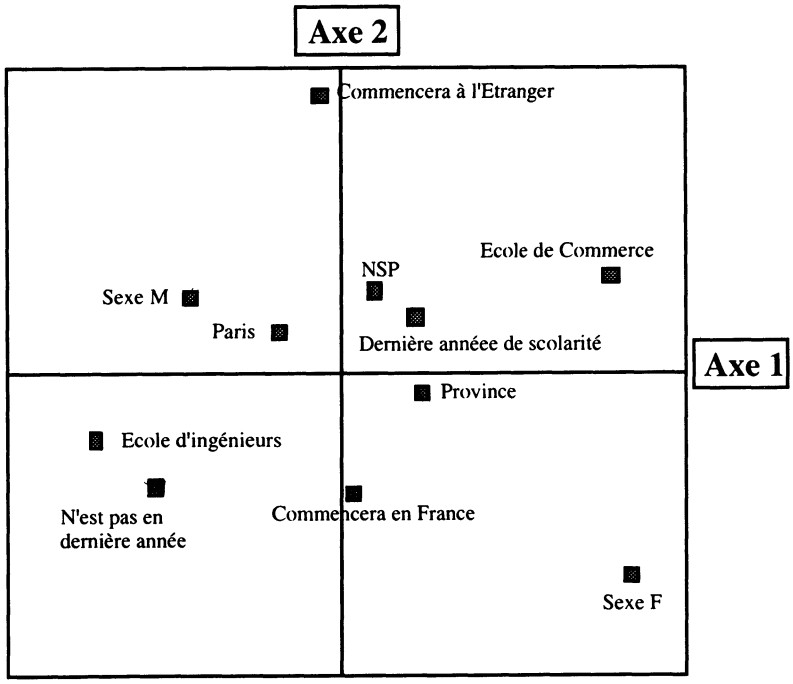

FIGURE 1
Individus – A.D.G. – Projection des centres de classes

La proximité entre deux centres de classes est à interpréter en tenant compte de l'espace de représentation des mots (figure 2). Ainsi l'éloignement entre les classes «Sexe F» et «Commencera à l'étranger» s'expliquerait par le fait que les élèves de sexe féminin auraient plus tendance à surnoter les mots «Nid», «Mariage»,

«Moelleux», ... (qui sont regroupés sur la figure 2 autour des concepts «Chez-soi» et «Confort intérieur») et sous-noter les mots «Différent» et «Étranger» contrairement aux individus de la classe «Commencera à l'étranger».

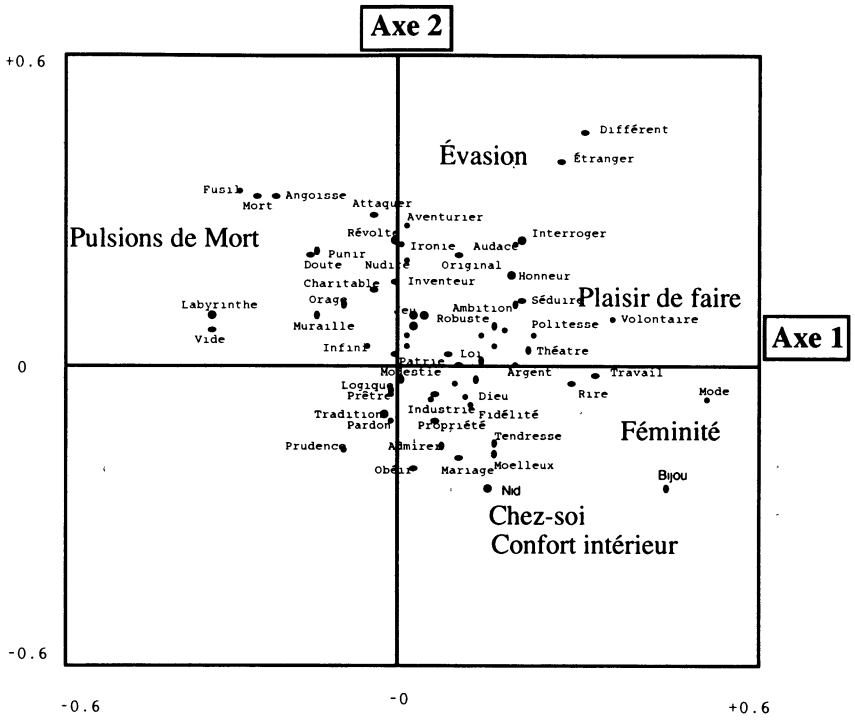

FIGURE 2
Variables – A.D.G. – Corrélations avec les deux premières composantes

La position du centre de la classe «Sexe M» à côté de «Action Morbide» ne signifie pas que les hommes soient plus morbides que les femmes. Les faibles valeurs des coefficients de stabilité des premiers facteurs montrent qu'il n'y a pas une typologie très nette des individus allant dans ce sens. Cela traduit que les individus qui donnent des notes élevées à «Fusil», «Mort», «Angoisse», ... sont plutôt de sexe masculin que féminin. A l'inverse, ceux qui surnotent les mots «Bijou» et «Mode» proviennent en majorité de la classe «Sexe F».

Les résultats de l'ACP

Ici, contrairement à l'ADG, les variables qualitatives ne nous seront d'aucune aide dans la description préliminaire des résultats. Elles ont un rôle illustratif *a posteriori*.

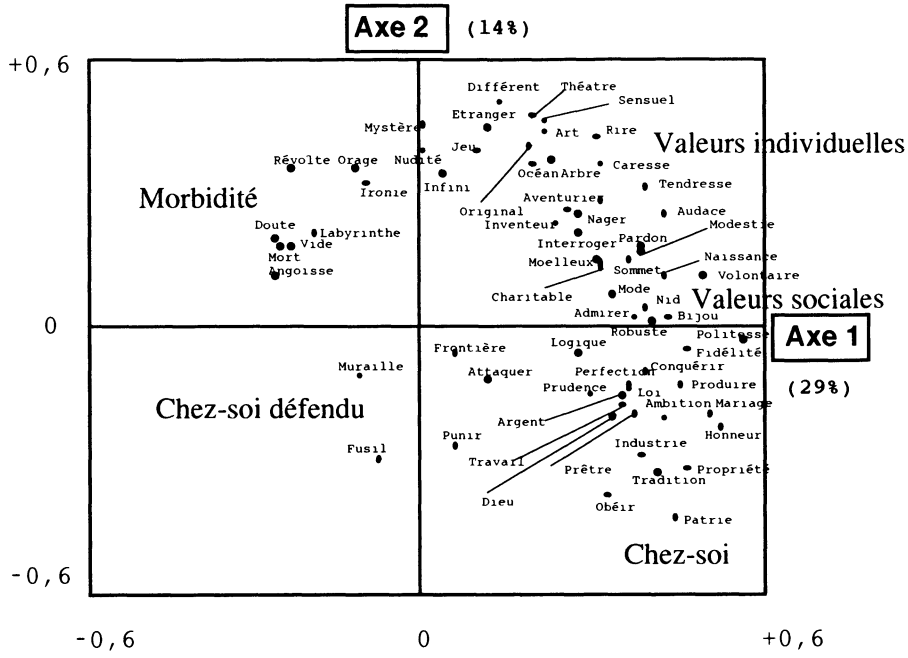

FIGURE 3
Variables – ACP – Corrélations avec les deux premières composantes

En examinant la figure 3, nous voyons que d'une manière isolée, certains mots s'assemblent deux à deux de la même façon qu'en ADG; mais globalement, leurs regroupements avec d'autres offrent un «modèle conceptuel» différent. Les mots suivants sont réunis aussi bien par l'ADG que par l'ACP : «Angoisse» et «Mort»; «Labyrinthe» et «Vide»; «Différent» et «Étranger»; ... Leur similitude ne dépend donc pas des variables qualitatives utilisées en ADG.

Par ailleurs d'autres mots (rapprochés par l'ADG) sont séparés par l'ACP faisant ainsi apparaître une interprétation nouvelle des résultats. C'est le cas de «Travail» et «Rire»; «Argent» et «Théatre»; «Mariage» et «Tendresse». Et alors que les uns («Travail», «Argent» et «Mariage»), en se combinant avec d'autres mots, font apparaître le concept «Valeurs de société», les autres («Rire», «Théâtre» et «Tendresse») apparaissent plutôt comme des valeurs individuelles.

Les centres de classes, à échelle égale, se trouvent dans un rectangle de dimensions plus faible par rapport à la figure 1. La typologie données par l'ACP est moins conforme aux variables qualitatives que celle donnée par l'ADG.

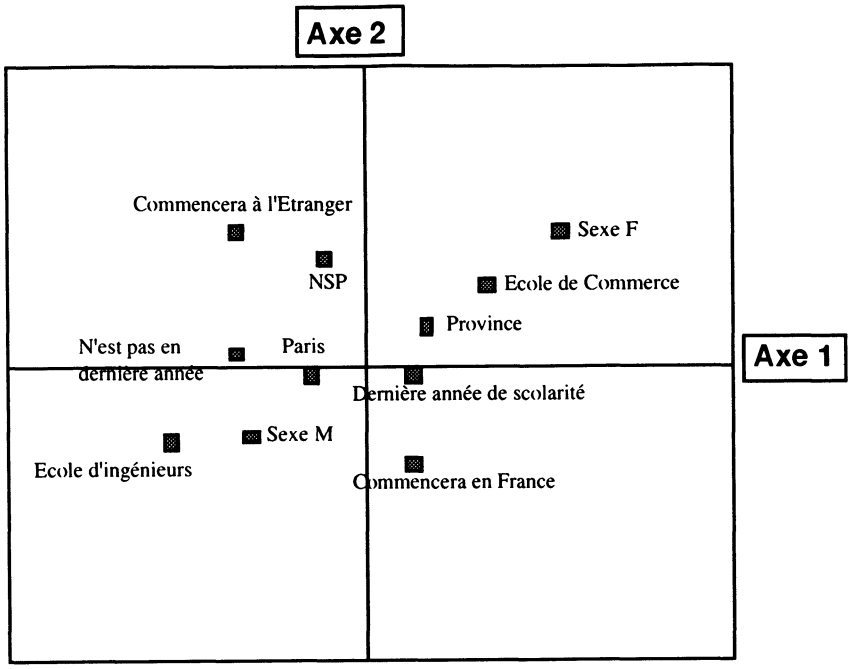

FIGURE 4
Individus – ACP – Projection des centres de classes

5. Conclusion

La pondération des paires d'individus par le nombre de variables qualitatives qui les réunissent offre une méthologie permettant l'analyse simultanée d'un couple de tableaux; l'un quantitatif et l'autre qualitatif. Elle permet d'accepter ou de réfuter l'influence des variables qualitatives sur les valeurs prises par les variables quantitatives.

En pondérant, d'autre part, chacun des individus par le nombre de ses «contigus», nous établissons les équations d'une analyse factorielle discriminante généralisée à plusieurs variables qualitatives.

Signalons cependant que l'analyse canonique des correspondances mise en place par Ter Braak [13] et telle qu'elle est présentée par Chessel *et al.* [5] permet d'analyser un couple de tableaux; le premier relevant de l'ACP et le deuxième de l'analyse des correspondances – en l'occurrence le tableau disjonctif complet des variables qualitatives éclatées. Elle peut donc répondre à notre problème. Elle remplace la notion de contiguïté par celle de moyenne conditionnelle par classe. Elle permet comme l'ADG d'indiquer l'influence des modalités des variables qualitatives sur la construction de l'espace de représentation des variables quantitatives.

Je remercie Monsieur L. Lebart pour ses conseils qui m'ont permis de réaliser ce travail.

Références

[1] Benali H., (1989), Analyse statistique spatiale : Application à la mortalité par cancer chez l'homme, XIVème réunion, Vevey, Suisse, 4-5 mai 1989, pp. 101-114.

[2] Burtchy B., Lebart L. (1991), Contiguity analysis and Projection Pursuit, 5th International Symposium on Applied Stochastic Models and Data Analysis, Granada, April 23-26 1991.

[3] Carlier A. (1985), Application de l'analyse factorielle des évolutions et de l'analyse intra- période, Statistique et Analyse des données, Vol. 10, n° 1, pp. 13-31.

[4] Cazes P., Moreau J. (1991), Analysis of a contingency table in which the rows and columns have a graph structure, Journées Internationales de Versailles, 18-20 septembre 1991, pp. 271-280.

[5] Chessel D., Lebreton J.D., Yoccoz N. (1987), Propriétés de l'analyse canonique des correspondances, Rev. de Statistique Appliquée, Vol. 35 (4), pp. 57-72.

[6] Deutsch E. (1989), Sémiométrie : une nouvelle approche de positionnement et de segmentation, Revue Française du Marketing, n° 25, pp. 5-16.

[7] Lebart L. (1974), Description statistique de certaines relations binaires, Actes du 3ème colloque sur l'analyse des données en Géographie, Besançon, pp. 75-101.

[8] Le Foll Y. (1982), Pondération des distances en analyse factorielle, Statistique et Analyse des données, pp. 13-31.

[9] Mom A. (1988), Méthodologie statistique de la classification des réseaux de transports, Thèse USTL, Montpellier.

[10] Monestiez P. (1978), Présentation de deux méthodes utilisant la contiguïté pour l'analyse des données géographiques, Thèse de docteur ingénieur, Paris 6.

[11] Romeder J.M. (1973), Méthodes et programmes d'analyse discriminante, Dunod, Paris.

[12] Royer J.J. (1988), Analyse multivariable et filtrage des données régionalisées, Thèse de Doctorat, INPL, Nancy.

[13] Ter Braak C.J.F. (1986), Canonical correspondance analysis : a new eigenvector technique for multivariate direct gradient analysis, Ecology, 67 (5), pp. 1167-1179.