CENTRE DE RECHERCHE POUR L'ÉTUDE ET L'OBSERVATION DES CONDITIONS DE VIE
142, RUE DU CHEVALERET - 75013 PARIS - TEL. : 584-14-20

CREDOC

VALIDITE DES RESULTATS EN ANALYSE DES DONNEES

Cadre conceptuel, mise en oeuvre et analyse critique des methodes de validation des resultats dans le traitement des donnees en sciences humaines

Ludovic LEBART

la recherche scientifique et technique

A.C.C. Informatique et sciences humaines

CENTRE DE RECHERCHE POUR L'ÉTUDE ET L'OBSERVATION DES CONDITIONS DE VIE
142, RUE DU CHEVALERET - 75013 PARIS - TEL. : 584-14-20

VALIDITE DES RESULTATS EN ANALYSE DES DONNEES

Cadre conceptuel, mise en oeuvre et analyse critique des methodes de validation des resultats dans le traitement des donnees en sciences humaines

Ludovic LEBART

Compte-rendu de fin d'étude
d'une recherche financée
par la
Délégation Générale
à la recherche scientifique
et technique

Enquêtes :
situations,
perceptions,
aspirations

L.L/cd - n°4465 - 24 Novembre 1975.
Retirage : Novembre 1982

Action complémentaire coordonnée
Décision d'aide n° 74-7-0349

RESUME SIGNALETIQUE

Ce rapport traite de la validation des résultats issus des analyses de données et, plus particulièrement des méthodes d'analyse factorielle descriptives : analyse en composantes principales (A.C.P.) et analyse des correspondances (A.F.C.).

On montre au chapitre I que les valeurs propres issues de l'A.F.C. des tables de contingence peuvent être approchées par celles d'une matrice suivant une loi de *WISHART* particulière. On justifie numériquement et théoriquement l'approximation ; on démontre l'indépendance des taux d'inertie et de la trace, et l'on propose des abaques et des tables originales.

On montre au second chapitre l'inadaptation des taux d'inertie comme critère de qualité des résultats, et l'on discute les rapports de ces notions avec celles d'information au sens mathématique et au sens large , ainsi que des calcul de stabilité par simulation.

Enfin au chapitre III pour éclairer la nature et la valeur des résultats, on analyse les modalités pratiques et effectives d'application des méthodes et les diverses critiques qu'elles suscitent, en insistant sur la contribution qu'elles peuvent apporter aux sciences humaines.

ABSTRACT

This report deals with the validation of output of data analysis methods, especially descriptive factor analysis methods : principal component analysis (A.C.P.) and correspondence analysis (A.F.C.). It is shown that the eigenvalues obtained from correspondence analysis of contingency tables can be approximated by those of a particular *WISHART* matrix ; this approximation is shown to be theoretically and numerically consistent ; the independence between the percentages of variance and the trace is proved and new charts and tables are given. Chapter II deals with percentages

of variance, their ability to describe the quality of results, their connection with the concept of information, and with the control of stability of results by simulation. Then, to illustrate the nature and the consistency of the outputs, the uses and misuses of the methods are discussed ; the contributions to research in the fields of social sciences are emphasized.

S O M M A I R E

INTRODUCTION

Cette étude se scinde en trois chapitres relativement indépendants. Chacun d'eux possède sa propre introduction et ses références bibliographiques. C'est pourquoi nous ne ferons ici, en évoquant l'itinéraire suivi, que justifier brièvement les choix qui ont été opérés lors de la recherche, et qui ont conduit à ce découpage en trois parties.

Pour étudier la validité des résultats en analyse des données, il faut d'abord tenter de définir ce que sont effectivement ces résultats.

S'agit-il de représentation (sous forme de graphiques ou d'arborescence), de la description de ces représentations, des interprétations que le chercheur peut en tirer ? Plutôt que de donner des définitions générales, nous avons préféré considérer des situations pratiques favorables à l'analyse de ces concepts, et élargir ensuite notre champ d'investigation.

Nous commençons ainsi par étudier ce que l'on peut considérer comme le "noyau dur" de l'analyse des données : *l'analyse des correspondances des vrais tableaux de contingence*, pour laquelle nous pensons avoir contribué ici à la résolution de problèmes inférentiels classiques. Après avoir proposé une approximation de la loi des valeurs propres (numériquement satisfaisante), on montre que les taux d'inertie sont indépendants en probabilité de l'inertie totale, ce qui ne sera pas sans conséquences sur l'évaluation des résultats. On propose ensuite des abaques et des tables nouvelles destinées à guider les utilisateurs. Il ne s'agit cependant que d'un domaine d'application relativement étroit et d'une démarche inférentielle limitée, qui a néanmoins l'avantage de faire le lien avec certains chapitres de la statistique multidimensionnelle classique.

On étudie ensuite la mesure classique de la qualité des représentations par les taux d'inertie (chapitre II). Il s'agit apparemment d'un simple aspect technique de l'utilisation des méthodes : des contre-exemples tout d'abord, puis des développements faisant intervenir la théorie statistique de l'information nous montreront en fait l'inadaptation et l'ambiguïté de ces paramètres, et nous obligeront à préciser la nature

des résultats des méthodes d'analyse factorielle descriptive, que l'on pourra soumettre en pratique à des calculs de stabilité. Les modalités d'exécution de ces calculs (qui diffèrent fondamentalement des épreuves de significaiton des sous-espaces abordés au chapitre I) sont spécifiées dans quelques cas correspondant à des applications fréquentes.

Pour apprécier la valeur et l'utilité des matériaux nouveaux produits par les algorithmes d'analyse, il a paru utile, enfin, de procéder à une étude critique de l'utilisation des méthodes dans les sciences humaines. Dans ce domaine d'application vaste et divers, les techniques statistiques sont parfois utilisées sans beaucoup de discernement, et pourtant les critiques sont souvent les plus fines. On analyse quelques-unes de ces critiques, et l'on tente ensuite de dresser un premier bilan de l'apport de ces méthodes en examinant leur contribution et les perspectives qu'elles offrent aux sciences humaines.

Pour plusieurs étapes de cette recherche, nous avons bénéficié des conseils, avis ou critiques de nos collègues du CREDOC et de l'ISUP ; *J-P. FENELON* et *M-O. LEBEAU* à propos d'exploitations lourdes sur ordinateur (chap. I), *A. MORINEAU* pour certains problèmes de simulation (chap. I), *C. HESSE, A. LECLERC, N. TABARD* à propos de l'étude critique de l'utilisation des méthodes.

Chapitre I

LA SIGNIFICATION DES VALEURS PROPRES EN ANALYSE DES CORRESPONDANCES

LA SIGNIFICATION DES VALEURS PROPRES EN ANALYSE DES CORRESPONDANCES

On trouvera ci-dessous quelques développements théoriques et pratiques et un recueil de données empiriques construit par simulation, concernant la signification des valeurs propres issues de l'analyse des correspondances de tableaux de contingence.

Le paragraphe 1, après quelques généralités, donne un aperçu des résultats théoriques disponibles ; ceci sera l'occasion de proposer certains résultats, puis certaines approximations qui seront justifiées, puis confirmées dans les paragraphes suivants.

Le paragraphe 2 traitera plus particulièrement du problème de la génération des nombres pseudo-aléatoires, dans le contexte de nos préoccupations. Il traite également des diverses modalités de remplissage ou de construction des tableaux.

Enfin le paragraphe 3 donne les tables statistiques établies par simulation ; ces tables permettront d'apprécier le degré de signification des cinq premières valeurs propres issues de l'analyse des correspondances de tableaux de contingence depuis la dimension (6 x 6) jusqu'à la dimension (50 x 100). Il s'agit évidemment de tables approchées. Les recommandations du paragraphe 2, ainsi que le mode d'emploi du paragraphe 3 doivent permettre à l'utilisateur d'apprécier aussi la validité et les limites de l'outil que nous mettons à sa disposition.

I - GENERALITES ET RESULTATS THEORIQUES

L'étude de la signification des valeurs propres et des taux d'inertie n'occupe pas une place centrale en analyse des données ; l'utilisation de ces paramètres n'est pas toujours justifiée dans la pratique ; leur interprétation est, de plus, délicate.

L'hypothèse d'indépendance totale des lignes et des colonnes d'un tableau est beaucoup trop générale et sévère pour être réaliste. Dans les recueils de données issues des sciences humaines, elle est pratiquement impossible à observer, dès que les effectifs mis en jeu deviennent importants. Il est beaucoup plus utile, nous l'avons vu, de tester en fait la stabilité d'une configuration, en simulant des données fictives entachées d'erreur ou de fluctuations suggérées par la nature du problème traité et la qualité des mesures.

La grande variété des tableaux analysables (tableaux de mesures, d'incidences, de classements, de comptages, etc...) rend extrêmement délicate l'interprétation de ces valeurs propres et taux d'inertie, dont on sait qu'ils sont étroitement liés au codage des données.

Une situation favorable est fournie par les tableaux de contingence, pour lesquels l'inertie est riche de signification. Bien que l'hypothèse d'indépendance constitue un cas limite sans grande portée pratique, il nous a paru utile de mettre à la disposition des utilisateurs les "garde-fous" que constituent les seuils de signification des valeurs propres et des taux d'inertie, calculés précisément dans le cas où les données analysées ne reflèteraient qu'un "bruit de fond" imputable à des fluctuations d'échantillonnage.

La loi des valeurs propres issues de l'analyse des correspondances d'un tableau de contingence dans l'hypothèse d'indépendance est pratiquement connue de façon implicite, à condition toutefois de procéder à certaines approximations qui paraissent raisonnables, et que justifient d'ailleurs les résultats obtenus par simulation.

Cette loi, fort complexe, a donné lieu à maintes publications erronées. Ainsi, dans le célèbre traité de statistique de *M.G. KENDALL* et *A. STUART* (*The advanced Theory of Statistic, vol.2, 1961, page 574*), les valeurs propres sont supposées suivre des lois du χ^2, comme l'inertie totale. La tentation était forte, en effet, de penser à une décomposition de cette inertie sur les sous-espaces propres, bien que le fait que ces sous-espaces soient calculés après réalisation de l'épreuve soit fort embarrassant...

H.O. LANCASTER ("*Canonical Correlation and Partition of* χ^2", *quart-j . Maths - 14 - pp. 220-224, 1963*) a réfuté sans difficulté ce résultat en montrant que l'espérance mathématique de la première valeur propre est toujours supérieure aux résultats préconisés par les assertions de *M.G. KENDALL* et *A. STUART*.

En fait, la loi cherchée est étroitement liée à celle des valeurs propres des matrices distribuées selon la loi de *WISHART*, dont la densité de probabilité a été explicitée en 1939 par *FISHER* [réf 5] , *GIRSHICK* [réf 6] , *HSU* [réf 7] et *ROY* [réf 17] , puis par *MOOD* [réf 13]. On en trouve la démonstration dans l'ouvrage de *T.W. ANDERSON* [réf 1].

L'intégration de cette densité assez complexe a donné lieu a plusieurs publications parmi lesquelles nous retiendrons principalement celles de *K. PILLAI (1965)* [réf 15] , *P.R. KRISHNAIAH* et *T.C. CHANG (1971)* [réf 8] , de *P.R. KRISHNAIAH* et *V.B. WAIKAR (1971)* [réf 9] , qui s'inspirent des travaux du physicien *M.L. MEHTA* [réf 11, 12].

Une table des seuils de signification de la plus grande et de la plus petite valeur propre a été publiée dès 1968 [réf 3] pour des tableaux dont la plus petite dimension n'excède pas 10.

Une autre tabulation est parue en 1973 [réf 4] qui donne des seuils de signification concernant la loi simultanée de la plus grande et la plus petite valeur propre de matrices distribuées suivant une loi de *WISHART* ; il s'agit de tableaux dont les dimensions n'excèdent pas 20 x 50. On trouvera dans la bibliographie de la page 73 les références de quelques autres travaux connexes.

I.1 - Une approximation de la loi des valeurs propres

Nous utiliserons les notations suivantes : (cf. [réf 2])

Le tableau théorique P_{IJ} n'est autre que le tableau $P_I \otimes P_J$ (Hypothèse d'indépendance probabiliste).

Le tableau observé, correspondant à l'effectif total k, a pour terme général k_{ij} - ses marges sont notées $(k_{i.})$ et $k_{.j}$.

Le tableau de fréquences relatives correspondant a pour terme général $f_{ij} = k_{ij}/k$ - On définit de même $(f_{i.})$ et $(f_{.j})$: $f_{i.} = \sum_j f_{ij}$, $f_{.j} = \sum_i f_{ij}$ (Nous conservons les points pour rendre muets les indices i et j).

On note :

$$f_I^j = \{f_{ij}/f_{.j} \mid i \in I\} \quad , \quad f_I = \{f_{i.} \mid i \in I\} \quad , \quad f_I = \sum\{f_{.j} f_I^j \mid j \in J\}$$

Dans l'espace $\mathbb{R}_I$, muni de la métrique du χ^2 de centre f_I, l'analyse revient à calculer les moments principaux d'inertie du nuage $\mathcal{N}(J)$ de J points-profils de coordonnées $\{f_I^j, j \in J\}$.

Nous ferons le même type d'approximation que lors des calculs de χ^2 : nous supposerons que k est assez grand pour que l'approximation normale de la loi multinomiale puisse s'appliquer ; nous supposerons également que les valeurs calculées pour les marges peuvent être substituées sans dommage aux valeurs théoriques, *sans négliger toutefois les contraintes qui résultent de ces estimations.*

Posons $X_I^j = \sqrt{k_{.j}}(f_I^j - f_I)$ pour $j \in J$

X_I^j est approximativement normal, centré, sphérique pour la métrique choisie, dans $\mathbb{R}^{I-1}$. (On a d'ailleurs : $\sum\{X_i^j \mid i \in I\} = 0, \forall j \in J$)

Nous cherchons les valeurs propres de la matrice d, de terme général

$$d_{ii'} = \frac{1}{k}\sum\left\{ \frac{X_i^j \, X_{i'}^j}{\sqrt{f_{i.} f_{i'.}}} \mid j \in J \right\}$$

Ce sont aussi les valeurs propres de la matrice d'ordre (I-1) x (I-1).

$$d^* = \frac{1}{k} \sum \{Y^j \otimes Y^j \mid j \in J\}$$

où Y^j est le vecteur des (I-1) coordonnées du point j dans une base de $\mathbb{R}^{I-1}$, choisie de façon à ce que celles-ci soient indépendantes centrées réduites.

Or la relation $\sum \sqrt{k_{.j}} \{X_I^j \mid j \in J\} = 0$ implique $\sum \{y^j \mid j \in J\} = 0$
Les cardJ vecteurs Y^j ne sont pas indépendants. On peut les considérer comme suivant la loi conditionnelle de n vecteurs multinormaux centrès sphériques indépendants, connaissant leur somme. On peut alors trouver $J-1$ vecteurs Z^j, Normaux, centrés, indépendants, tels que :

$$d^* = \frac{1}{k} \sum \{Z^j \otimes Z^j \mid j = 1,2,\ldots,J-1\}$$

- D'où le résultat de l'approximation :

Soit λ_u la $u^{ième}$ valeur propre issue de l'analyse des correspondances du tableau de contingence d'ordre I x J, correspondant à l'effectif total k.

Dans l'hypothèse où les lignes et les colonnes du tableau sont indépendantes, $k\lambda_u$ suit approximativement la loi de la $u^{ième}$ valeur propre d'une matrice distribuée suivant une loi de *WISHART* = $W_{I-1}(E,J-1)$ (Loi de $S = \sum \{Z^j \otimes Z^j \mid j = 1, J-1\}$, chaque vecteur Z^j à I-1 composantes étant multinormal, centré, de matrice d'inertie théorique E (matrice unité). Z^j étant indépendant de $Z^{j'}$ si $j \neq j'$).

I.2 - Vérification expérimentale de la qualité de l'approximation.

Les seules tables existantes concernent la plus grande valeur propre et les tableaux dont la plus petite dimension n'excède pas 10. (Tables citées en [réf 3]) . Les tables citées en [réf 4] ne concernent que la loi jointe des valeurs propres extrêmes et sont donc pratiquement inutilisable en analyse des données.

Les tables que nous publions en fin de chapitre sont établies à partir de 100 simulations par cas (p,n), p étant la largeur du tableau, n sa longueur. (p varie de 6 à 50 : de 6 à 20 (pas : 2), de 20 à 50 (pas : 5), n varie de 6 à 100, de 6 à 20 (pas : 2), de 20 à 50 (pas : 5), de 50 à 100 (pas : 10), ce qui fait 170 cas).

Pour chaque simulation, on donne la moyenne, la médiane, l'écart-type, le fractile correspondant à 95% des observations pour chacune des cinq plus grandes valeurs propres, pour chacun des cinq pourcentages d'inertie correspondants, enfin, pour la trace, à titre de vérification, puisque l'on sait que celle-ci suit une loi du χ^2 à (n-1) (p-1) degré de liberté. Afin de mettre à l'épreuve les programmes de génération de séries pseudo-aléatoires et de construction de tableaux vérifiant l'hypothèse d'indépendance, il a été procédé à divers essais correspondant à 1000 simulations concernant plus spécialement les deux cas : (n=8, p=8) et (n=6, p=6), que nous comparerons donc aux tables existantes. Seuls les seuils pourront faire l'objet de comparaisons puisque les tables existantes ne donnent pas de valeurs centrales, ni de caractéristiques de dispersion.

Pour des tableaux dont l'effectif total k vaut 1000, la loi établie par simulation donne :

Tableau 8 x 8 = seuil 0.05 (plus grande valeur propre) = 0.0303
Tableau 6 x 6 = seuil 0.05 (plus grande valeur propre) = 0.0216

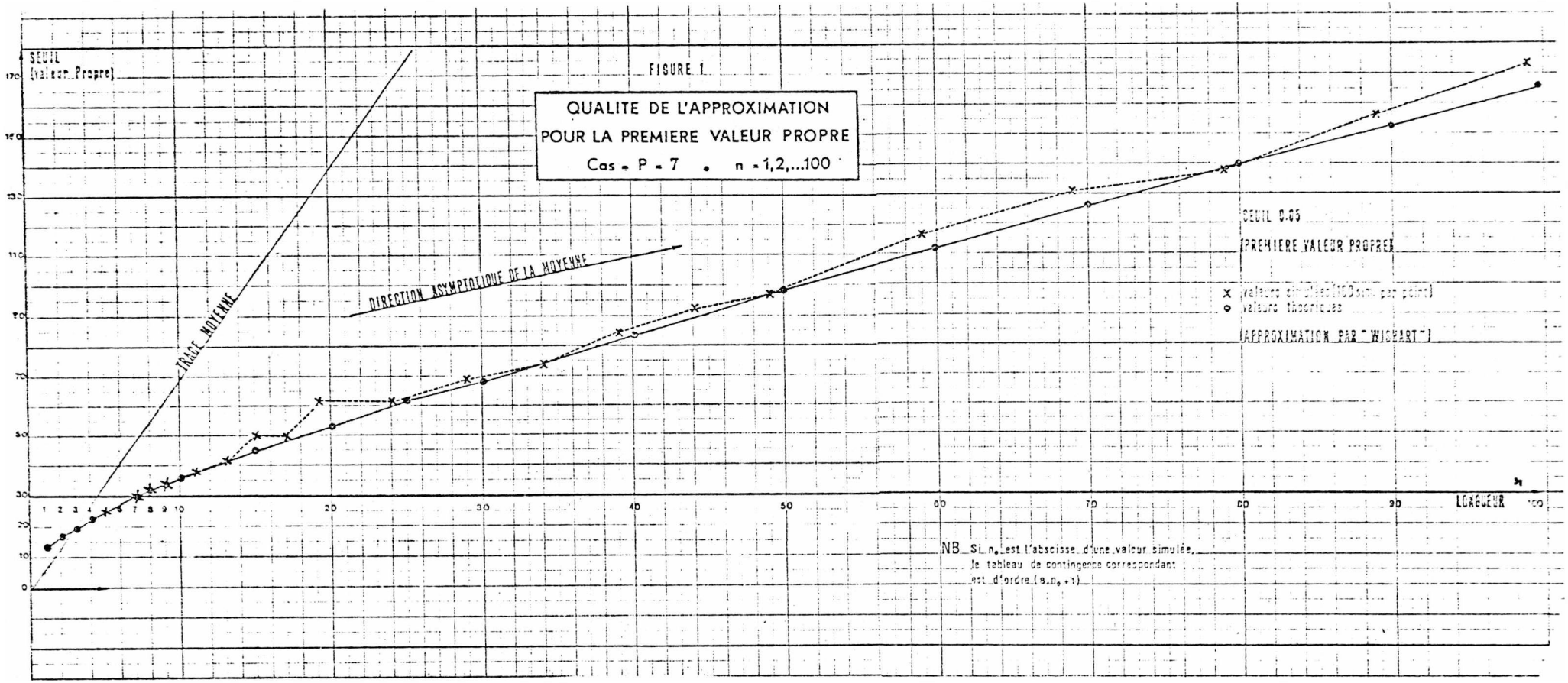
Les tables [réf 3] nous donnent, après division par k(=1000) des seuils :

Tableau 7 x 7 = seuil 0.05 (plus grande valeur propre) = 0.0304
Tableau 5 x 5 = seuil 0.05 (plus grande valeur propre) = 0.0218

On voit que la concordance des deux résultats est satisfaisante.

La figure 1 compare maintenant, pour un tableau d'une largeur donnée (p=7 pour la matrice de *WISHART*, et par conséquent J=8 pour la matrice à diagonaliser en analyse des correspondances) les valeurs empiriques du seuil correspondant à 100 simulations et les valeurs exac-

FIGURE 1


QUALITE DE L'APPROXIMATION
POUR LA PREMIERE VALEUR PROPRE
Cas : P = 7 • n = 1,2,...100
SEUIL
(valeur Propre)
TRACE MOYENNE
DIRECTION ASYMPTOTIQUE DE LA MOYENNE
SEUIL 0.05
(PREMIERE VALEUR PROPRE)
(valeurs simulées (100 sim. par point))
(valeurs théoriques)
(APPROXIMATION PAR "WISHART")
LONGUEUR
NB Si n_0 est l'abscisse d'une valeur simulée,
le tableau de contingence correspondant
est d'ordre $(n, n_0 + 1)$

tes, conformément au modèle retenu.

Les valeurs exactes se situent sur une courbe très tendue que l'on peut assez rapidement assimiler à une droite de pente 1,5 (contre 7, pour la trace, 1 pour la moyenne).

On remarque que pour n=1, on a le seuil correspondant à un χ^2 à 7 degrés de liberté ($s_{0,05} = 14,07$).

Les seuils observés (il s'agit d'estimations plus fragiles que les valeurs centrales) coïncident pratiquement avec les valeurs théoriques pour $n \leqslant 13$. Ils restent très proches de la courbe théorique, en la surpassant toutefois systématiquement, ce qui peut être dû précisément à l'adéquation de l'approximation choisie dans la génération des tableaux. Malgré ce biais, l'ordre de grandeur des écarts reste satisfaisant compte tenu des effectifs mis en jeu.

L'ensemble des paramètres fournis par les tables simulées permettent de détecter les valeurs "exceptionnelles" sans qu'il soit indispensable de procéder à un lissage ou à une régression polynomiale. Ainsi, pour la valeur la plus atypique, correspondant à n=19 (qui correspond donc sur la table au seuil de la plus grande valeur propre issue de l'analyse d'un tableau de contingence 8 x 20), on peut constater un écart exceptionnel de la moyenne et de la médiane, une valeur du seuil de pourcentage d'inertie excessive, puisque ces quantités doivent décroître régulièrement avec n.

I.3 - Aspects techniques de la loi approchée.

Nous esquisserons ci-dessous quelques-uns des problèmes de calcul qui surgissent lors de l'étude théorique de la loi des valeurs propres.

On désignera par $W_p(T,n)$ la loi, dite de *WISHART*, suivie par la matrice aléatoire : $S = \sum\{X^j \otimes X^j \mid j \leqslant n\}$. X^j étant un vecteur multinormal de $\mathbb{R}^p$, de matrice d'inertie théorique T non singulière (X^j indépendant de $X^{j'}$ pour $j \neq j'$).

Classiquement, la densité de probabilité de S désignée par $W_p(S,n,T)$ a la forme suivante, en désignant par $|S|$ et $|T|$ les déterminants des matrices S et T.

$$W_p(S,n,T) = \frac{\left|\frac{1}{2}T\right|^{\frac{n}{2}} |S|^{\frac{n-p-1}{2}} e^{-\frac{1}{2}\mathrm{tr}\, T^{-1}S}}{\pi^{\frac{p(p-1)}{4}} \prod_{i=1}^{p} \Gamma\left(\frac{1}{2}(n-i+1)\right)}$$

Dans le cas qui nous intéresse, où $T = E$ (matrice unité), on exprime facilement cette densité en fonction des seules valeurs propres de S :

$$W_p(S,n,I) = \frac{\prod_{i=1}^{p} \lambda_i^{\frac{(n-p-1)}{2}} \exp\{-\frac{1}{2}\sum \lambda_i\}}{2^{\frac{pn}{2}} \pi^{\frac{p(p-1)}{4}} \prod_{i=1}^{p} \Gamma\left(\frac{n+1-i}{2}\right)}$$

La matrice aléatoire S peut se mettre sous la forme :

$S = P'\Lambda P$, P étant une matrice orthogonale et Λ étant une matrice diagonale dont les éléments diagonaux sont les valeurs propres de S ; en imposant à celles-ci d'être classées par ordre décroissant, et en imposant $p_{i1} \geqslant 0$, il est facile de voir que la correspondance entre S et (Λ, P) est bijective.

Désignons par $J(\Lambda, P)$ le jacobien de cette transformation.

On établit [réf 1] le résultat classique suivant :

$$\int\cdots\int J(\Lambda,P)\, dP = \frac{\pi^{\frac{p(p+1)}{4}} \prod_{i<j} (\lambda_i - \lambda_j)}{\prod \Gamma\left(\frac{1}{2}(p+1-i)\right)}$$

Dans ces conditions comme la densité $W_p(S,n,E)$ ne dépend que

de Λ , la densité des valeurs propres s'écrit : ($\prod$ signifie $\prod_1^p$, sauf spécification contraire).

$$v(\Lambda) = C(n,p) \cdot \prod \lambda_i^{\frac{n-p-1}{2}} \exp\left(-\frac{1}{2}\sum \lambda_i\right) \prod_{i<j} (\lambda_i - \lambda_j)$$

avec $$C(n,p) = \pi^{\frac{p}{2}} / \left(2^{\frac{pn}{2}} \prod \Gamma\left(\frac{1}{2}(n+1-i)\right) \cdot \Gamma\left(\frac{1}{2}(p+1-i)\right)\right)$$

L'essentiel des difficulté de calculs réside dans le calcul des constantes d'intégration [réf 1, 13]. Le calcul de la partie variable du Jacobien (déterminant de VANDERMONDE) ne présente pas de difficulté, mais la simplicité du résultat aurait laissé espérer des démonstrations plus courtes.

I.4 - Modalités pratiques de calcul des seuils.

1) Les approximations établies par *PILLAI* [réf 16] et revues par *R. CHOUDARY HANUMARA* et *W. THOMSON* [réf 3] pour le calcul des seuils correspondant à la seule plus grande valeur propre conduisent à la séquence d'opération suivante : avec $a = \frac{1}{2}(n-p-2)$,

$$\text{on a : } P\{\lambda_{MAX} > u_1\} = \begin{cases} G_p(a,u_1) \text{ si } p \text{ est pair} \\ 1 - K_p(a,u_1) \text{ si } p \text{ est impair} \end{cases}$$

$$\text{avec } G_p(a,t) = \sum_{j=1}^{p} (-1)^{j+1} H_{p-j} e^{-\frac{1}{2}t} \left(\frac{t}{2}\right)^{a+p-j}$$

$$\text{et } K_p(a,t) = \int_0^t \frac{e^{-\frac{1}{2}z} z^a dz}{2^{a+1} + a!} + \sum_{i=1}^{p-1} H_{p-j} e^{-\frac{1}{2}t} \left(\frac{t}{2}\right)^{a+p-j}$$

La fonction H_{p-j} ci-dessus se calculant par la relation de récurrence :

$$H_{p-j} = \frac{\sqrt{\pi}}{\Gamma\left(a + \frac{p}{2} + \frac{1}{2}\right)\Gamma\left(\frac{p}{2}\right)} \binom{p-1}{j-1} \prod_i^{j-1} \left(\frac{2a+p-i+1}{2^{j-1}}\right) - (m+p-j+1) H_{p-j+1}$$

Ces formules sont relativement faciles à programmer, mais elles ne sont elles-mêmes que des approximations. Les essais de calculs directs que nous avons tentés ont été peu fructueux ; la maîtrise de la précision demande un important travail de numéricien.

Rappelons que ces calculs ne concernent que la plus grande valeur propre, dans le cas où la matrice à diagonaliser S suit une loi de *WISHART* exacte. Ces calculs n'ont été conduits que pour $p \leqslant 10$, ce qui présente évidemment peu d'intérêt en analyse des données.

2) S'inspirant des travaux de *MEHTA* [réf 11, 12] et *WIGNER* [réf 19], ainsi que des travaux déjà cités précédemment, *KRISHNAIAH* et *CHANG* [réf 8] proposent des formules exactes, sinon simples, pour la plus grande valeur propre d'une matrice aléatoire pour laquelle la densité de la loi jointe des valeurs propres est de la forme :

$$v(\Lambda) = C \prod_{i=1}^{p} \Psi(\lambda_i) \prod_{i<j} (\lambda_i - \lambda_j) \qquad a \leqslant \lambda_p \leqslant \lambda_{p-1} \leqslant \lambda_1 \leqslant b$$

(C,a,b étant des constantes connues).

On a alors la relation :

$$P[\lambda_1 \leqslant s] = C\ \rho(\Psi\ ;\ p,\ 0,\ a,\ s)$$

La fonction ρ dépend ici encore de la parité de p.

- Si p est pair :

$$\rho(\Psi\ ;\ p,\ 0,\ a,\ s) = \Delta(\Psi\ ;\ p,\ 0,\ a,\ s)$$

Où Δ est la racine carrée du déterminant de terme général a_{ij} (d'ordre p) $\Delta = |A|^{\frac{1}{2}}$

$$\text{avec} \quad a_{ij} = \oint_{i-1}^{j-1}$$

$$\text{où} \quad \oint_{u}^{t} = F_u^t - F_t^u$$

avec encore : $F_u^t = \int_o^u F_u(\theta)\, \theta^t \quad \psi(\theta)\, d\theta$

Dans cette expression : $F_u(\theta) = \int_a^\infty \psi(x) x^u dx$

ρ est donc la valeur de la racine carrée du déterminant d'une matrice dont chaque élément est la différence de deux intégrales doubles.

- Si p est impair on a :

$$\rho(\psi ; p, 0, a, s) = \sum_{i=0}^{p-1} (-1)^i F_i(s)\, G_{i+1}(\psi ; p, 0, a, s)$$

$F_i(s)$ a déjà été défini.

G_t est la racine carrée du déterminant d'une matrice A_t d'ordre p-1, obtenue en supprimant la $t^{ème}$ ligne et la $t^{ème}$ colonne de A.

Dans le cas qui nous intéresse directement,

$$\psi(\lambda) = \lambda^{\frac{n-p-1}{2}} e^{-\frac{\lambda}{2}}$$

Les équations précédentes permettent de calculer les seuils.

La densité de la plus grande valeur propre s'écrit, quant à elle :

$$v_1(\lambda_1) = C(n,p) \cdot \lambda_1^{qp+(p-1)(1+p/2)} \exp(-\lambda_1/2)\, \rho(\varphi ; p-1, r, 0, 1)$$

avec $q = (n-p-1)/2$ et $\varphi(x) = \exp(-\lambda_1 x/2)\,(1 - x)$

La fonction ρ a été définie plus haut.

Ce sont des formules de ce type qui sont effectivement utilisées par *CLEMM* et alii [réf 4] pour procéder à une tabulation de la loi jointe des valeurs propres extrêmes, ainsi que des seuils (à gauche, c'est-à-dire, par exemple, des valeurs que λ_1 à 95 chances sur 100 de dépasser !). Ces tables ne sont donc d'aucune utilité pour nous,

qui sommes au contraire intéressés par les valeurs que les premières valeurs propres ont peu de chances de dépasser, afin d'apprécier la signification des sous-espaces de visualisation.

Les calculs nécessaires à l'évaluation des seuils concernant des valeurs propres intermédiaires sont sensiblement plus complexes (cf. [réf 9]).

I.5 - Indépendance des taux d'inertie et de la trace - Loi des taux.

Un changement de variable du type : $\begin{cases} \lambda_i = z u_i \quad \text{pour} \quad i < p \\ \lambda_p = (1-u_1-u_2 \ldots u_{p-1})z \end{cases}$

dans la densité de probabilité des valeurs propres nous montre immédiatement la factorisation de $v(\Lambda)$ sous la forme : $v(\Lambda) = v_1(z) \cdot v_2(u_1,\ldots,u_{p-1})$ avec $z = \Sigma\lambda_i$, les u_i étant des taux d'inertie. (Le Jacobien vaut z^{p-1}). ($v(\Lambda)$ est donné par la formule du §-I-3).

On retrouve évidemment pour z la loi du χ^2 à np degrés de liberté :

$$v_1(z) = \frac{1}{2\,\Gamma\left(\frac{np}{2}\right)} \left(\frac{z}{2}\right)^{\frac{np}{2}-1} \exp\{-z/2\}$$

Ceci nous permet d'écrire la loi jointe des $p-1$ premiers taux d'inertie :

$$v_2(u_1,u_2,\ldots,u_{p-1}) = \mathfrak{D}(n,p) \cdot \prod_1^{p-1} u_i \cdot \prod_{i<j<p} (u_i-u_j) \cdot \left(1-\sum_1^{p-1} u_i\right) \cdot \prod_{i<p} (u_1+\ldots$$

$$+2u_i+\ldots+u_{p-1}{}^{-1}) \ .$$

Avec $\mathfrak{D}(n,p) = \mathcal{C}(n,p) \cdot \Gamma\left(\frac{np}{2}\right) \cdot 2^{\frac{np}{2}}$

La constante $\mathcal{C}(n,p)$ étant définie plus haut.

Les domaines d'intégration étant indépendants, on peut donc affirmer que trace et taux d'inertie sont indépendants.

L'orthogonalité de ces deux quantités peut être aisément vérifiée sur les tables.

Ceci nous conduira à ne retenir, pour tester les valeurs propres d'une analyse, que les taux d'inertie, puisque la loi de la trace est connue. Les deux premiers moments des valeurs propres s'obtiendront aisément à partir des moments de la trace (connus exactement) et ceux des taux (donnés par la table).

II - *SUR LA CONSTRUCTION DE CERTAINS TABLEAUX PSEUDO-ALEATOIRES EN VUE DES ETUDES PAR SIMULATION.*

L'approximation de la loi des valeurs propres étudiée au paragraphe précédent implique, dans la mesure où elle est acceptable, que la loi des valeurs propres issues de l'analyse des correspondances d'un tableau de contingence ne dépend que des dimensions du tableau (tout comme la trace) et non des poids "a priori" de chacune des lignes ou colonnes de ce tableau. Cette propriété, qui permet justement de procéder à une tabulation, n'est cependant qu'une propriété asymptotique, dont nous devrons vérifier le réalisme sur quelques exemples.

Une fois admis cette indépendance des poids "a priori", il reste à générer de façon efficace et économique des tables de contingence dont le comportement est analogue - pour notre objet précis d'étude - à de véritables réalisations d'épreuves aléatoires.

Pour cela, divers types de modèles de remplissage sont possibles, chacun d'eux pouvant faire appel à diverses classes de générations.

Or un générateur de nombres pseudo-aléatoires n'a de sens que pour une utilisation particulière, pour laquelle il possède les propriétés statistiques requises : le choix du générateur va s'avérer être indissociable du mode de construction du tableau, et nous devrons par conséquent traiter ces deux questions simultanément. Sur les propriétés statistiques des générateurs, nous avons bénéficié des conseils de A. MORINEAU ; nous renvoyons d'ailleurs le lecteur

à la note [réf 14] pour des informations méthodologiques et bibliographiques plus complètes sur ce sujet.

Pour chaque type de générateur, pour chaque méthode de construction, nous avons procédé à 1000 simulations sur tableau 8 x 8. A l'exception des essais destinés précisément à éprouver l'effet des variations des poids a priori des lignes et des colonnes, tous ces tableaux sont générés en supposant tous les poids théoriques égaux.

II.1 - Indépendance des poids.

Bien que l'approximation du paragraphe 1 soit une forte présomption en faveur du caractère non-paramétrique de la loi des valeurs propres, il nous a paru utile de procéder à l'expérience suivante : pour un tableau 8 x 8, d'effectif total $k = 1000$, généré suivant l'approximation normale (technique spécifiée et justifiée plus bas), on a envisagé les trois cas suivants, correspondant chacun à 1000 analyses.

1) Les poids sont tous égaux.
2) Une colonne i_o est 21 fois plus lourde que les autres ($f_{i_o} = \frac{3}{4}$, $f_i = \frac{1}{28}$, $i \neq i_o$).
3) Les poids des lignes et des colonnes sont proprotionnels à deux séries de nombres pseudo-aléatoires tirés uniformément entre 0 et 1.

Les résultats sont les suivants, pour les trois premières valeurs propres : tableau 8 x 8 (1000 simulations par cas ; k = 1000).

		1) poids égaux	2) disparités de poids (type 2)	3) poids "quelconques" (type 3)
Moyenne	λ_1	0.0212	0.0219	0.0218
	λ_2	0.0127	0.0131	0.0130
	λ_3	0.0077	0.0079	0.0079

	1) poids égaux	2) disparités de poids (type 2)	3) poids "quelconques" (type 3)
Moyenne des pourcentages p_1	43.3	43.5	43.3
p_2	26.0	26.0	26.0
p_3	15.7	15.7	15.8

L'amplitude des variations pourrait être justifiée par les seules fluctuations d'échantillonnage. Le mode de génération des tableaux introduit cependant un léger biais dont nous parlerons plus bas.

Cette expérience, jointe aux considérations théoriques du paragraphe 1, nous confirme la propriété d'indépendance des poids, et permet de procéder à une tabulation générale.

II.2 - Influence de l'effectif total k .

Nous avons vu que, pour un tableau T d'effectif total k dont les dimensions sont spécifiées, la loi de $k\lambda_u$, où λ_u est la $u^{-ème}$ valeur propre issue de l'analyse des correspondances du tableau, ne dépend que des dimensions de ce tableau. Il s'ensuit immédiatement que l'espérance mathématique et l'écart-type de λ_u sont des fonctions hyperboliques de k.

Les pourcentage d'inertie, eux, ne dépendront pas de k. On prendra garde que dans les tables publiées ci-après, l'effectif k prend successivement les valeurs 1000 (tables de dimensions inférieures à 20 x 20) puis 5000 (tables dont une seule des dimensions excède 20), enfin 10 000 (tables dont les deux dimensions excèdent 20) ; ceci évidemment afin d'assurer un "remplissage" correct des tables, conforme aux exigences de l'approximation normale de la loi multinomiale.

Le résultat du paragraphe 1.5 (*Indépendance des taux et de la trace*) nous permet cependant de ne consulter que les taux et les traces. On peut en effet vérifier la relation :

$$E_k(VP_i^k) \simeq E_k\left(\frac{VP_i^k}{T^k}\right) \cdot E_k(T^k)$$

où $\begin{cases} VP_i & \text{désigne la } i^{\text{ème}} \text{ valeur propre} \\ T & \text{" la trace} \\ E_k & \text{" l'opérateur "moyenne empirique".} \end{cases}$

II.3 -Le schéma multinomial exact.

La façon la plus rigoureuse de remplir une table de contingence dont les différents poids (lignes et colonnes) sont égaux a priori, et vérifient l'hypothèse d'indépendance, consiste à tirer successivement l'abscisse et l'ordonnée d'une case - par un tirage uniforme - d'affecter une unité à cette case, et de recommencer l'opération k fois, si k est l'effectif total du tableau.

Ainsi, pour remplir un tableau 20 x 50, par exemple, il faut que k soit de l'ordre de 10 000, pour qu'il y ait un nombre important de cases dont l'effectif dépasse 5 (seuil empirique généralement admis pour les calculs de χ^2).

La simulation d'un seul tableau demande déjà 20 000 tirages, ce qui exclut l'emploi de ce mode de génération pour construire des tables nécessitant la construction de dizaines de milliers de tableaux. Nous avons en fait appliqué ce schéma à 1000 simulations d'un tableau 8 x 8. d'effectif total $k = 1000$.

Une faiblesse des générateurs par congruence.

Les premières simulations sur tableau 8 x 8 d'effectif $k = 1000$ ont produit une trace t moyenne de l'ordre de 0.048 (au lieu de 0.049, comme le veut la théorie, puisque 1000 t est en χ^2 à (8-1) (8-1) degrés de liberté).

Cette différence (significative) nous a plongés dans une certaine perplexité quand A. *MORINEAU* a attiré notre attention sur une faibles-

se déjà étudiée (*Method in Randomness, M. GREENBERGER, M.I.T. Comm. ACM-1965 - vol.8 - p.177-179*) des générateurs par congruence : des zones rectilignes de faible densité apparaissent dans le tableau croisant des couples de tirages successifs.

Cette différence s'est immédiatement estompée dès que nous avons tiré les abscisses et les ordonnées avec des générateurs différents. (*Programmes SEN3A et SEN3B* cités en [réf 14] d'après une méthode proposée par *K.D. SENNE* en 1974). Ces programmes produisent des séries orthogonales, de période 2^{37}, et sont indépendants des caractéristiques des calculatrices.

Des extraits des résultats numériques figurent dans le tableau récapitulatif publié plus bas. Les résultats détaillés et les programmes figurent dans des annexes techniques qui sont à la disposition des chercheurs intéressés.

II.4 -Approximations binomiales et normales.

Il existe des façon beaucoup moins coûteuses de générer des tables de contingences aléatoires, en procédant toutefois à certaines approximations.

2.4.1. Schéma binomial et poissonnien.

Chaque case (i,j) dont la probabilité théorique est $p_i p_j$, se voit affecter une valeur d'une variable binomiale $B(k,p_i p_j)$, k étant l'effectif total.

En fait, dès que $p_i p_j < 0.05$, il est admis que la loi de *POISSON* est une bonne approximation de la loi binomiale. Or, dans les cas qui nous intéressent, ceci est toujours vérifié : pour un tableau (n,p), la probabilité théorique d'une case est $\frac{1}{np}$ (puisque ce n'est pas une contrainte de supposer toutes les marges théoriques égales) et n et p sont supérieurs à 5

Le schéma poissonnien a l'avantage de générer un "comportement

multinomial" exact du tableau, d'après une propriété déjà remarquée par *ABBE (1889)* et *STUDENT (1908)* :

Proposition : Si $y_1, y_2, \ldots, y_n$ sont des réalisations de variables aléatoires de *POISSON* indépendantes de même paramètre a, alors la loi conditionnelle de $y_1, y_2, \ldots, y_n$ connaissant leur somme est une loi multinomiale.

En effet, la probabilité d'un n-uple $(y_1, y_2, \ldots, y_n)$ s'écrit : $p(y_1, y_2, \ldots, y_n) = \prod_i^n \left\{ e^{-a} \, a^{y_i} / y_i! \right\}$. Comme la somme Y est une variable aléatoire de *POISSON* de paramètre na, la loi conditionnelle s'écrit immédiatement :

$$p(y_1, y_2, \ldots, y_n \mid Y) = n^{-y} \, y! / \Pi y_i!$$

ce qui n'est autre qu'une loi multinomiale (cas des modalités équiprobables). Cette propriété n'est cependant vraie que pour la loi de *POISSON*, qu'elle caractérise. Elle est couramment utilisée pour procéder à certains tests dans certains problèmes de dénombrement en hématologie. Le test qui en est issu ne possède cependant aucune robustesse.

En effet, on déduit immédiatement de la propriété précédente que, si $y_1, y_2, \ldots, y_n$ sont des variables aléatoires de *POISSON* indépendantes, si $\bar{y}$ désigne leur moyenne arithmétique, alors $Z = \sum_i (y_i - \bar{y})^2 / \bar{y}$ est une variable aléatoire distribuée suivant une loi du χ^2 à (n-1) degrés de liberté.

Or, pour de petites valeurs de n ($n < 20$), cette propriété n'est vraie que si les variables aléatoires suivent rigoureusement des lois de *POISSON*. En particulier, cette propriété n'est pas valable pour la loi normale, alors que l'on sait que cette loi est une loi limite de la loi de *POISSON*. Des simulations nous ont montré que la variation du χ^2 est très sensible. Avec une loi binomiale très proche d'une loi de *POISSON* ($p = 0.05$, $k = 200$), $E(Z) \simeq n-2$. La valeur (n-1) n'a pu être obtenue empiriquement qu'avec un générateur de loi de *POISSON*.

Le test habituellement utilisé par les biologistes est donc très peu robuste.

2.4.2. Approximation normale.

Comme l'approximation binomiale, l'approximation normale néglige l'autocorrélation négative existant entre les cases du tableau au moment de la construction de celui-ci. L'effectif total est lui-même une variable normale, et la loi conditionnelle du tableau connaissant la somme, qui seule nous intéresse ici, introduit une autocorrélation qui nous rapproche du schéma multinomial. Considérons en effet n variables $X_1, X_2, \ldots, X_n$, aléatoires, normales, indépendantes, de moyenne μ, de variance σ^2. Soit $S' = X_1 + X_2 + \ldots + X_n$.

Nous allons comparer les deux premiers moments d'un vecteur multinomial (modalités équiprobables, somme S), et d'un vecteur multinormal (conditionné par S').

	Schéma normal conditionnel (connaissant $S' = \Sigma X_i$)	Schéma multinomial exact (S fixé, n modal.équip.)
Moyenne X_i	S'/n	S/n
var (X_i)	$\sigma^2(1 - \frac{1}{n})$	$\frac{S}{n}(1 - \frac{1}{n})$
Cov $(X_i X_j)$	$-\sigma^2/n$	$-S/n^2$

Ces deux schémas seront d'autant plus proches que σ^2 sera voisin de (S/n), ce qui nous rapproche du modèle poissonnien.

En pratique on sera assuré d'une bonne concordance en choisissant $\sigma^2 = \frac{S}{n}$ et μ (moyenne théorique) $= \frac{S}{n}$. $(S'/n$ tend vers S/n en probabilité quand $n \to \infty)$.

Le schéma normal, peu coûteux, sera donc adopté pour la construction des tables. Le tableau ci-après compare le schéma multinomial exact, et le schéma normal, avec divers types de générateurs.

2.4.3. Les générateurs. (On se reportera à la [réf 14] pour plus de détails).

Quatre types de générateurs de variables normales à partir de tirages uniformes ont été utilisés successivement.

1) Approximation comme somme de 12 variables uniformes sur [0,1]. Cette méthode, bien que grossière et coûteuse, donne cependant des résultats acceptables.

2) Méthode de *BOX et MULLER (1958)*.
Transformation analytique à partir de deux tirages uniformes.

3) Méthode de *MARSAGLIA - BRAV (1964)*.
Même principe que la précédente.

4) Méthode de *FORSYTH, AMRENS, DIETER et BRENT (1974)*.
Transformation analytique à partir de un ou deux tirages uniformes.

Ces quatre méthodes faisant appel à un ou plusieurs tirages uniformes, il restait à les combiner avec les différents programmes de tirages, qui sont tous fondés sur des méthodes de congruence, mais diffèrent par les germes, la période : seuls sont publiés ci-dessous les quatre versions faisant appel au programme *SENS3A* déjà cité.

Les résultats du tableau ci-dessous sont très convergents. La colonne 1 contient les résultats signalés au paragraphe 2.3 : ils sont entachés par le biais coorélatif à l'utilisation d'un seul générateur. (La trace est significativement faible).

Rappelons que les écarts-types doivent être divisés par $\sqrt{1000}$, pour obtenir les écarts-types des moyennes.

Les autres résultats ne diffèrent apparemment que par des fluctuation d'échantillonnage.

Ces différents générateurs, indiscernables à partir des estimations relatives à 1000 analyses, le seront a fortiori lorsque l'on se restreindra à 100 analyses pour chaque élément de tabulation.

COMPARAISONS DE 6 MODES DE GENERATIONS DE TABLEAUX ALEATOIRES

Incidences sur les 5 premières valeurs propre *

Tableau 8 x 8 - k = 1000
(1000 simulations par cas)

Moyennes et écarts-types des valeurs propres	Loi multinomiale "exacte"		Approximations normales			
	1(biais) (un seul générateur)	2 (deux générateurs)	BOX-MULLER	Loi Limite	MARSAGLIA-BRAY	FORSYTH-BRENT
1 m =	0.0203	0.0209	0.0212	0.0211	0.0213	0.0214
σ =	(0.0053)	(0.0052)	(0.0056)	(0.0054)	(0.0055)	(0.0059)
2 m =	0.0126	0.0129	0.0127	0.0130	0.0129	0.0129
σ =	(0.0032)	(0.0032)	(0.0035)	(0.0034)	(0.0034)	(0.0034)
3 m =	0.0077	0.0079	0.0077	0.0079	0.0079	0.0079
σ =	(0.0023)	(0.0022)	(0.0023)	(0.0023)	(0.0023)	(0.0023)
4 m =	0.0044	0.0045	0.0044	0.0046	0.0045	0.0044
σ =	(0.0016)	(0.0016)	(0.0016)	(0.0017)	(0.0016)	(0.0016)
5 m =	0.0021	0.0022	0.0021	0.0022	0.0022	0.0021
σ =	(0.0010)	(0.0010)	(0.0010)	(0.0010)	(0.0010)	(0.0010)
TRACE						
m =	0.0480	0.0493	0.0489	0.0495	0.0496	0.0496
σ =	(0.0098)	(0.0096)	(0.0104)	(0.0106)	(0.0102)	(0.0106)

* (On pourra trouver les histogrammes des valeurs propres, des pourcentages d'inertie, les valeurs extrêmes dans des annexes techniques non publiées).

III - ABAQUES ET TABLES STATISTIQUES (*)

III.1 - Abaques

Les abaques ci-après ne donnent qu'une toute petite partie de l'information contenue dans les tables du paragraphe III.2, mais sous une forme plus vivante.

On représente tout d'abord les valeurs médianes de la plus grande valeur propre (figure 2) en sélectionnant les largeurs : 6 - 8 - 10 - 20 - 30 - 40 - 50). Puis les estimations des seuils pour les taux d'inertie correspodant au cinq premières valeurs propres (Figures 3, 4, 5, 6,7). On peut constater une certaine dispersion des estimations pour les tableaux de petites dimensions : la variance des taux d'inertie (et par conséquent celle de leur estimation) est en effet une fonction décroissante des dimensions du tableau [cf. tables ci-après].

Les extrêmités des courbes (points : 6 x 6, 8 x 8, 10 x 10) ont été établies à partir de 1000 simulations (au lieu de 100 pour les autres points) afin de préciser leur tracés.

(*) On trouvera des extraits des histogrammes et des matrices des corrélations décrivant certains aspects de la *loi jointe* des valeurs propres et des taux d'inertie en annexe 1.

L'ensemble des résultats figure dans des annexes techniques non publiées.

FIGURE 2

VALEURS MEDIANES DU POURCENTAGE D'INERTIE DE LA PLUS GRANDE VALEUR PROPRE

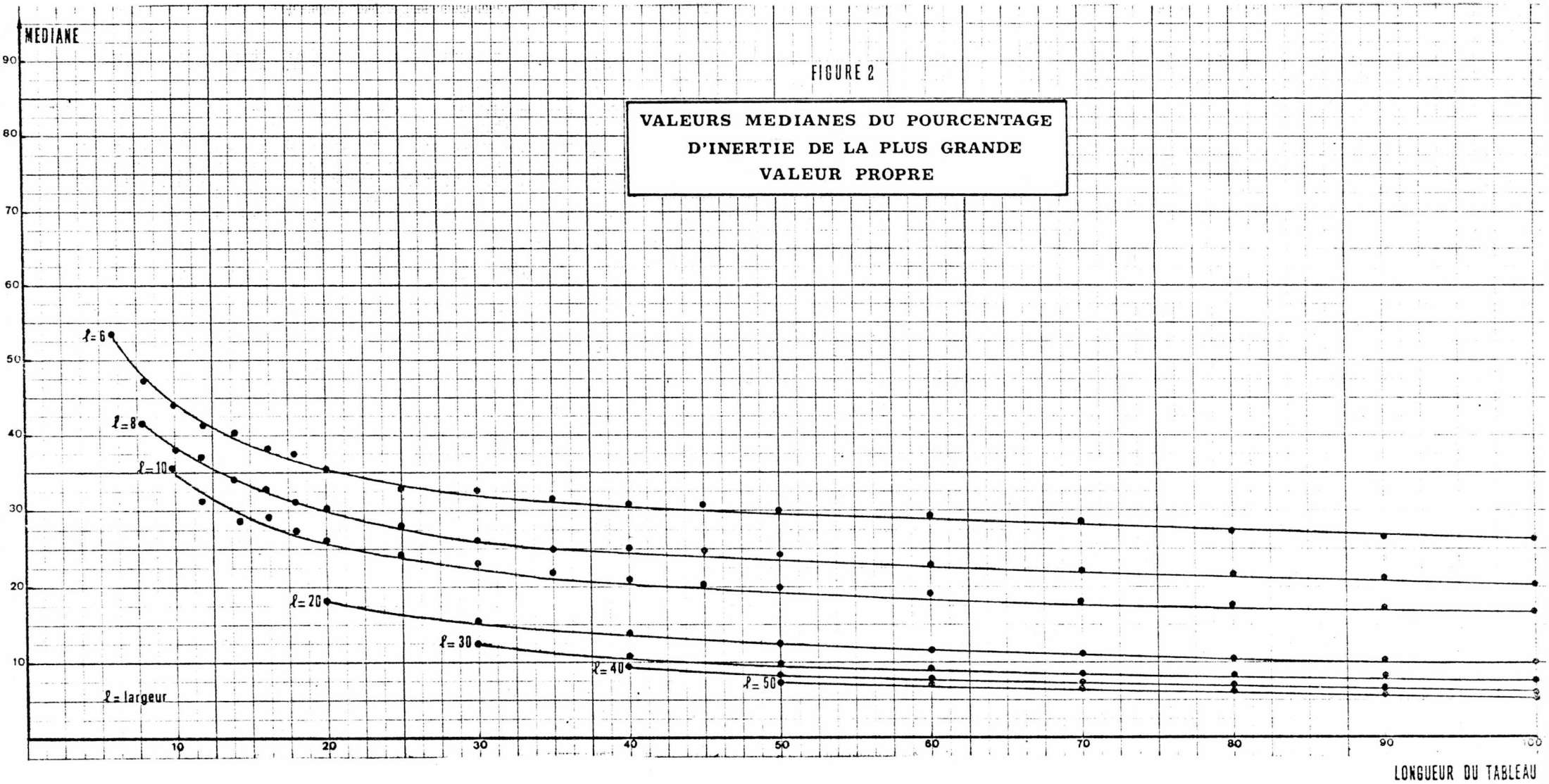


FIGURE 3

SEUIL (0.05 unilatéral) DU POURCENTAGE D'INERTIE DE LA PLUS GRANDE VALEUR PROPRE

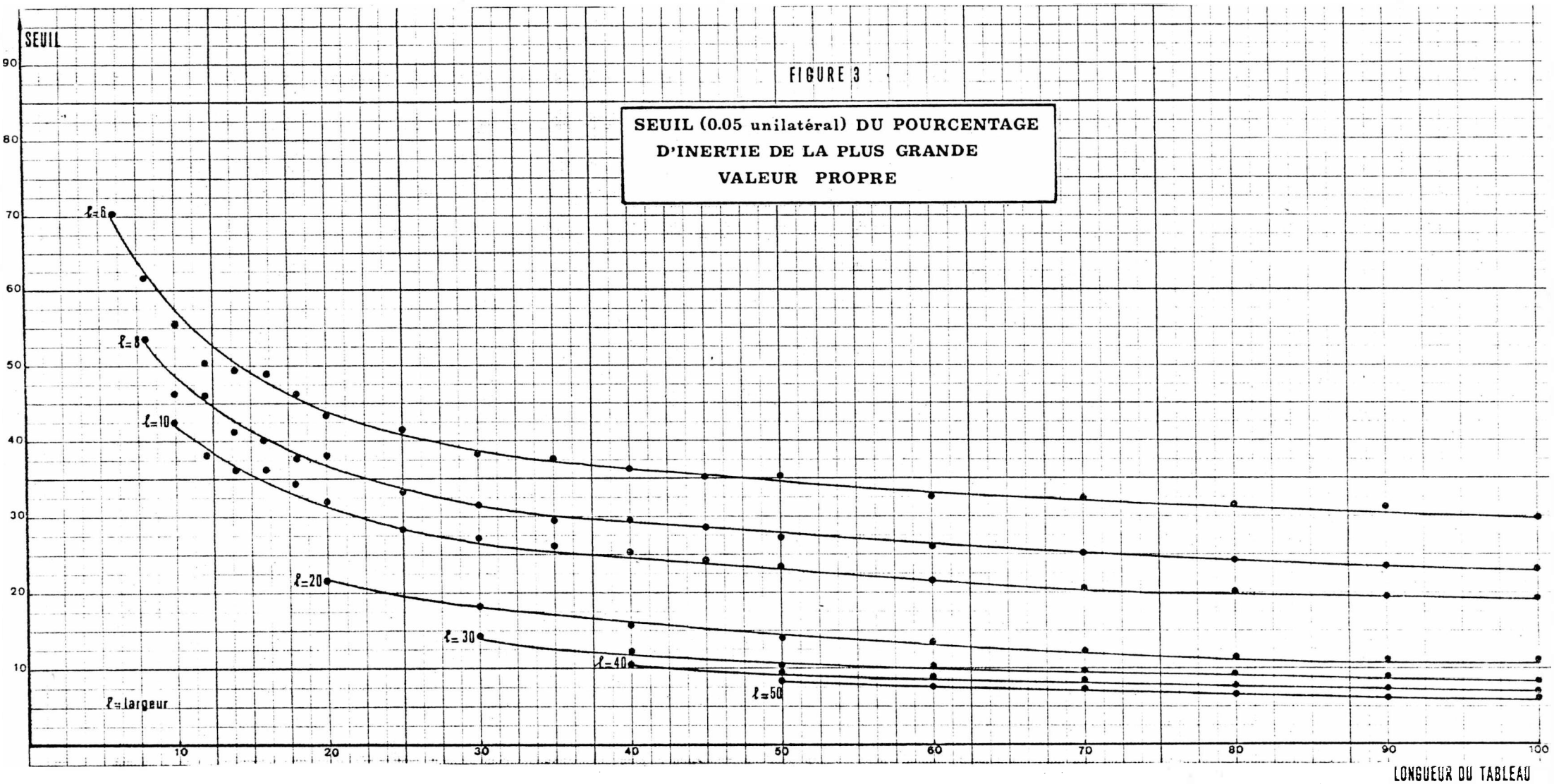


FIGURE 4

SEUIL (0.05 unilatéral) DU POURCENTAGE D'INERTIE DE LA DEUXIEME VALEUR PROPRE

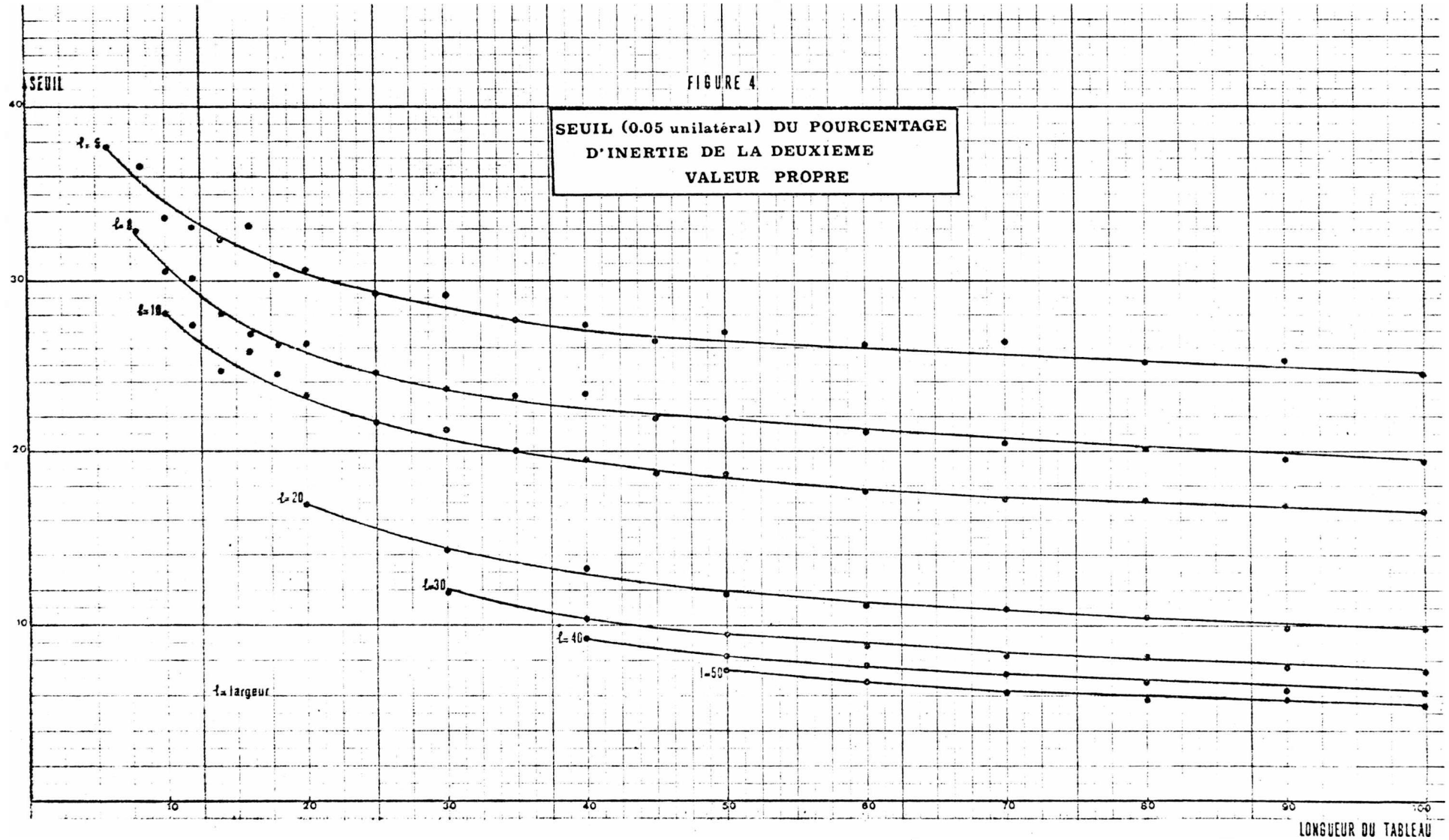


FIGURE 5

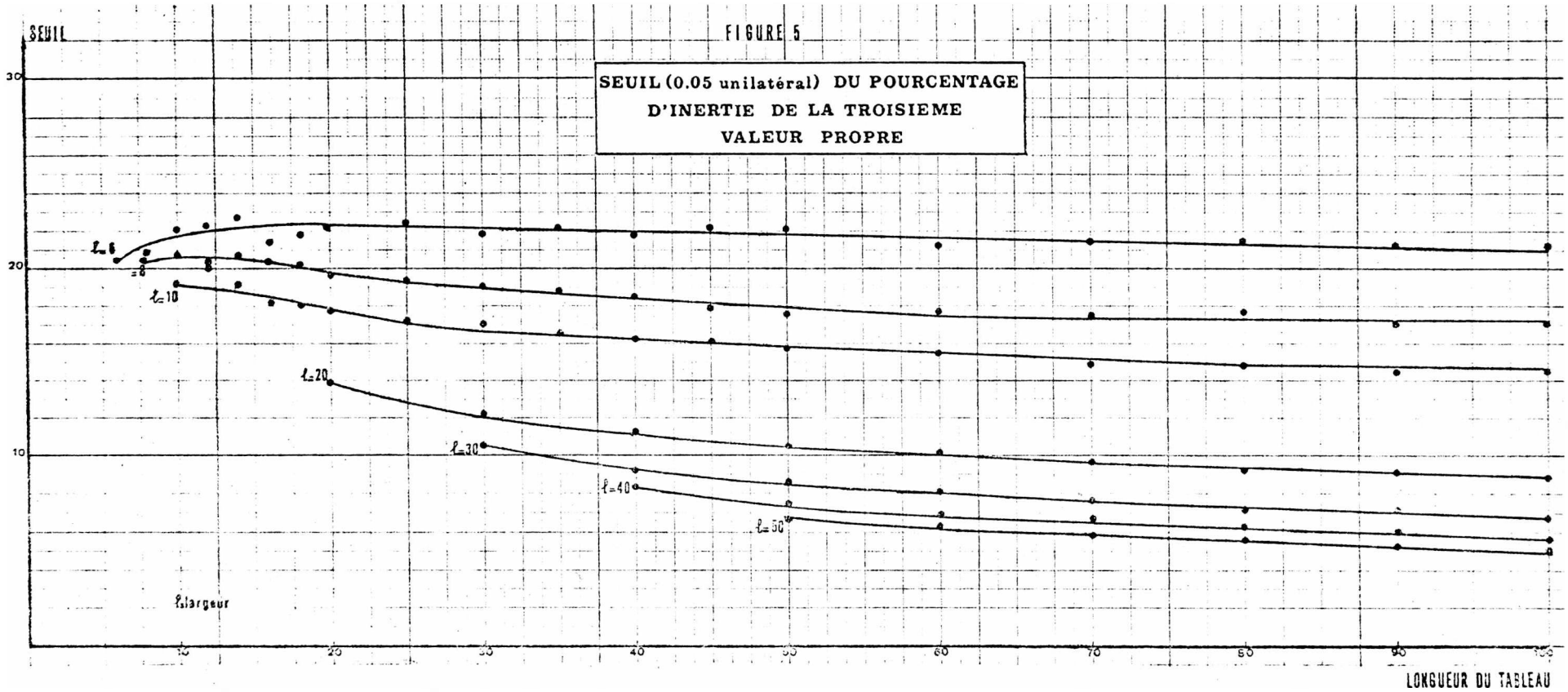

SEUIL (0.05 unilatéral) DU POURCENTAGE
D'INERTIE DE LA TROISIEME
VALEUR PROPRE
SEUIL
30
20
10
ℓ=6
ℓ=8
ℓ=10
ℓ=20
ℓ=30
ℓ=40
ℓ=50
ℓ largeur
10
20
30
40
50
60
70
80
90
100
LONGUEUR DU TABLEAU

FIGURE 6

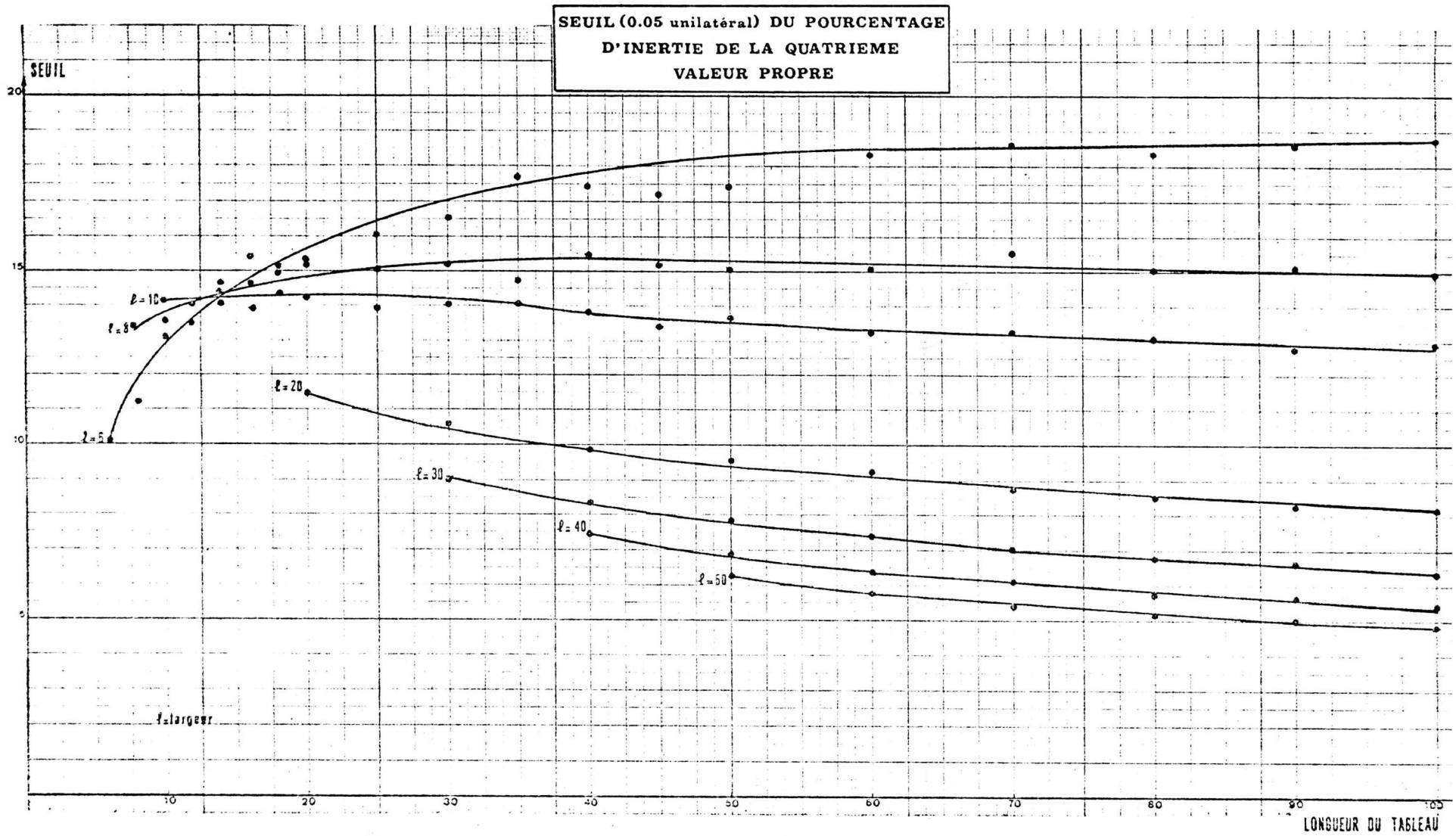
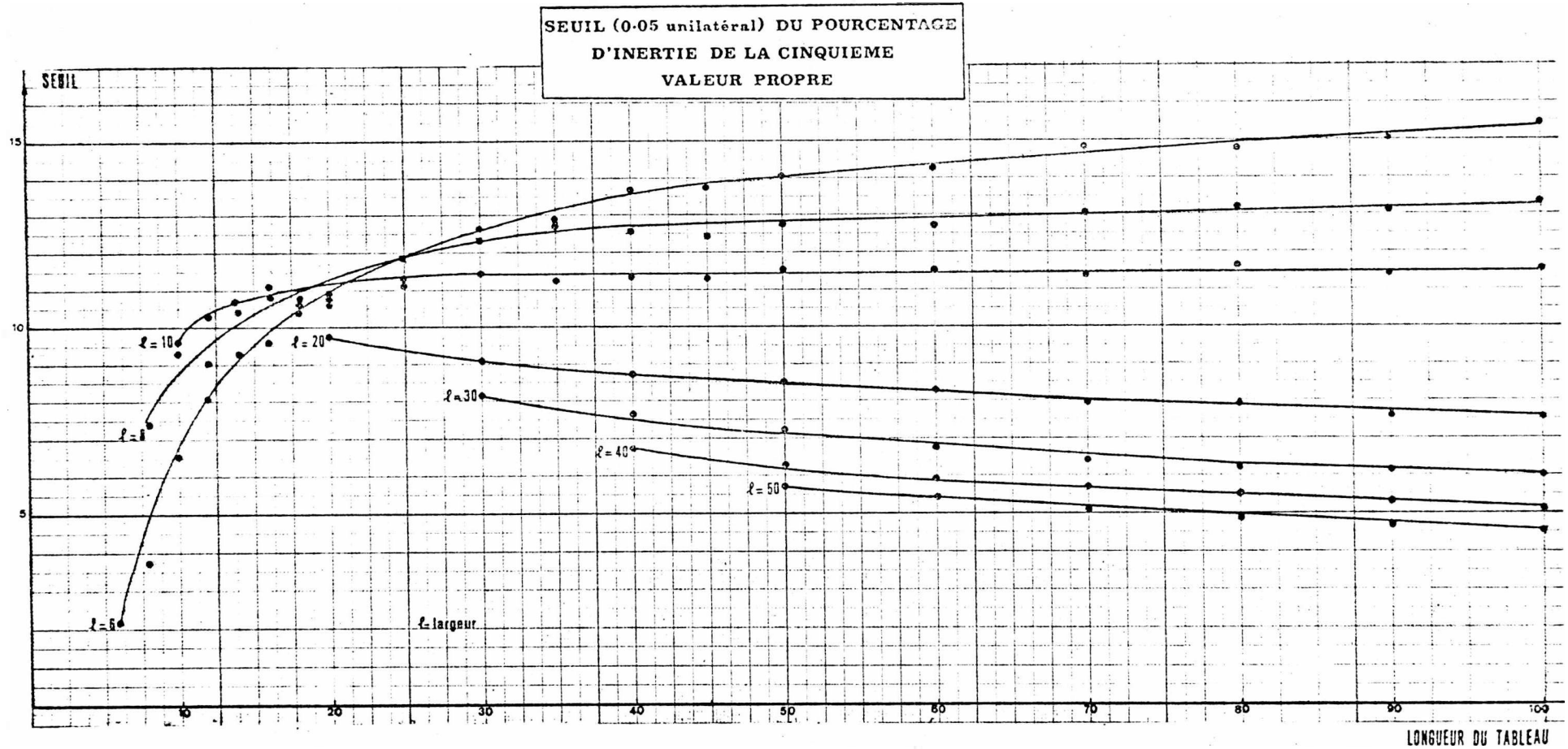

SEUIL (0.05 unilatéral) DU POURCENTAGE
D'INERTIE DE LA QUATRIEME
VALEUR PROPRE
SEUIL
20
15
10
5
ℓ=10
ℓ=8
ℓ=5
ℓ=20
ℓ=30
ℓ=40
ℓ=50
ℓ=largeur
10
20
30
40
50
60
70
80
90
100
LONGUEUR DU TABLEAU

FIGURE 7


SEUIL (0·05 unilatéral) DU POURCENTAGE
D'INERTIE DE LA CINQUIEME
VALEUR PROPRE
SEUIL
15
10
5
ℓ = 10
ℓ = 20
ℓ = 30
ℓ = 40
ℓ = 50
ℓ = 6
ℓ = 5
ℓ = largeur
10
20
30
40
50
60
70
80
90
100
LONGUEUR DU TABLEAU

III.2 - Tables

Analyse des correspondances de tables de contingence

Tables approchées donnant des estimations des moyennes, médianes, écarts-types, seuil (0.05 unilatéral) des cinq premières valeurs propres et des cinq pourcentages d'inertie correspondants.

Tableaux 6 x 6 à 50 x 100

(de 6 à 20 = pas "2")
(de 20 à 50 = pas "5")
(de 50 à 100 = pas "10")

- l'Effectif k est donné en haut de page.
- Les moments et les seuils correspondant à chaque cas sont calculés sur 100 analyses de tableaux pseudo-aléatoires.

Abréviations : VPk = Valeur propre de rang k.
PCk = Pourcentage d'inertie correspondant.
TRA = Trace.

SIGNIFICATION DES VALEURS PROPRES (VP),ET DES POURCENTAGES (PC) EN ANALYSE DES CORRESPONDANCES (EFFECTIFS= 1000.)

LARGEUR DU TABLEAU= 6

LONG.= 6	VP1	VP2	VP3	VP4	VP5	*	PC1	PC2	PC3	PC4	PC5	*	TRA
MOYENNES	0.0139	0.0071	0.0031	0.0012	0.0002	*	54.17	27.97	12.44	4.84	0.59	*	0.0255
MEDIANES	0.0138	0.0065	0.0028	0.0010	0.0001	*	53.21	28.77	11.79	4.45	0.25	*	0.0250
ECARTS-TYPES	0.0047	0.0025	0.0015	0.0007	0.0002	*	8.46	6.18	4.67	2.72	0.71	*	0.0073
SEUIL 0.05	0.0211	0.0116	0.0062	0.0027	0.0006	*	70.10	37.67	20.58	10.12	2.12	*	0.0380

LONG.= 8	VP1	VP2	VP3	VP4	VP5	*	PC1	PC2	PC3	PC4	PC5	*	TRA
MOYENNES	0.0169	0.0096	0.0051	0.0024	0.0006	*	48.79	27.93	14.62	6.83	1.83	*	0.0346
MEDIANES	0.0162	0.0095	0.0049	0.0021	0.0006	*	47.58	27.57	14.52	6.55	1.77	*	0.0344
ECARTS-TYPES	0.0049	0.0030	0.0020	0.0012	0.0005	*	7.34	5.22	4.09	2.56	1.24	*	0.0086
SEUIL 0.05	0.0251	0.0150	0.0083	0.0043	0.0016	*	60.74	36.67	20.55	11.20	3.75	*	0.0486

LONG.= 10	VP1	VP2	VP3	VP4	VP5	*	PC1	PC2	PC3	PC4	PC5	*	TRA
MOYENNES	0.0197	0.0120	0.0073	0.0039	0.0015	*	44.58	27.00	16.37	8.76	3.29	*	0.0444
MEDIANES	0.0190	0.0116	0.0067	0.0037	0.0013	*	44.23	27.09	16.12	8.61	3.01	*	0.0430
ECARTS-TYPES	0.0049	0.0034	0.0024	0.0015	0.0009	*	6.15	4.17	3.49	2.80	1.83	*	0.0097
SEUIL 0.05	0.0273	0.0181	0.0113	0.0061	0.0031	*	55.52	33.68	22.04	13.08	6.59	*	0.0600

LONG.= 12	VP1	VP2	VP3	VP4	VP5	*	PC1	PC2	PC3	PC4	PC5	*	TRA
MOYENNES	0.0228	0.0147	0.0092	0.0053	0.0025	*	42.05	26.98	16.80	9.66	4.51	*	0.0545
MEDIANES	0.0224	0.0141	0.0088	0.0049	0.0021	*	41.62	26.76	16.45	9.23	4.40	*	0.0516
ECARTS-TYPES	0.0053	0.0038	0.0030	0.0019	0.0013	*	5.14	3.85	3.10	2.42	1.90	*	0.0121
SEUIL 0.05	0.0321	0.0208	0.0145	0.0091	0.0049	*	50.35	33.08	22.39	13.57	8.08	*	0.0777

LONG.= 14	VP1	VP2	VP3	VP4	VP5	*	PC1	PC2	PC3	PC4	PC5	*	TRA
MOYENNES	0.0269	0.0172	0.0110	0.0067	0.0035	*	41.12	26.28	16.88	10.38	5.35	*	0.0654
MEDIANES	0.0257	0.0167	0.0111	0.0068	0.0035	*	40.59	26.37	17.06	10.02	5.13	*	0.0641
ECARTS-TYPES	0.0071	0.0046	0.0033	0.0021	0.0015	*	5.33	3.52	2.98	2.37	1.94	*	0.0146
SEUIL 0.05	0.0390	0.0250	0.0169	0.0099	0.0060	*	49.73	32.54	22.67	14.08	9.26	*	0.0910

LONG.= 16	VP1	VP2	VP3	VP4	VP5	*	PC1	PC2	PC3	PC4	PC5	*	TRA
MOYENNES	0.0310	0.0205	0.0136	0.0090	0.0045	*	39.46	26.19	17.28	11.35	5.72	*	0.0786
MEDIANES	0.0297	0.0202	0.0134	0.0088	0.0042	*	38.24	26.09	17.18	11.23	5.28	*	0.0785
ECARTS-TYPES	0.0063	0.0042	0.0030	0.0025	0.0019	*	5.21	3.73	2.67	2.47	2.13	*	0.0120
SEUIL 0.05	0.0412	0.0281	0.0180	0.0134	0.0080	*	48.97	33.16	21.48	15.40	9.60	*	0.0951

LONG.= 18	VP1	VP2	VP3	VP4	VP5	*	PC1	PC2	PC3	PC4	PC5	*	TRA
MOYENNES	0.0330	0.0219	0.0151	0.0102	0.0059	*	38.31	25.46	17.57	11.81	6.85	*	0.0862
MEDIANES	0.0318	0.0213	0.0149	0.0098	0.0057	*	37.61	25.39	17.67	11.84	6.88	*	0.0824
ECARTS-TYPES	0.0071	0.0044	0.0031	0.0025	0.0021	*	4.59	3.09	2.52	2.15	2.07	*	0.0143
SEUIL 0.05	0.0457	0.0302	0.0202	0.0143	0.0097	*	46.02	30.32	21.82	14.92	10.40	*	0.1140

LONG.= 20	VP1	VP2	VP3	VP4	VP5	*	PC1	PC2	PC3	PC4	PC5	*	TRA
MOYENNES	0.0349	0.0246	0.0174	0.0117	0.0070	*	36.51	25.69	18.24	12.24	7.33	*	0.0955
MEDIANES	0.0340	0.0236	0.0172	0.0110	0.0068	*	35.48	25.61	18.14	12.23	7.09	*	0.0937
ECARTS-TYPES	0.0072	0.0054	0.0034	0.0029	0.0022	*	4.32	3.10	2.25	2.17	2.14	*	0.0158
SEUIL 0.05	0.0456	0.0345	0.0231	0.0169	0.0112	*	43.59	30.74	22.15	15.33	10.92	*	0.1220

SIGNIFICATION DES VALEURS PROPRES (VP),ET DES POURCENTAGES (PC) EN ANALYSE DES CORRESPONDANCES (EFFECTIFS= 5000.)

LARGEUR DU TABLEAU= 6

LONG.= 25	VP1	VP2	VP3	VP4	VP5	*	PC1	PC2	PC3	PC4	PC5	*	TRA	*
MOYENNES	0.0082	0.0060	C.0046	0.0032	0.0021	*	34.16	24.86	18.99	13.34	8.65	*	0.0240	*
MEDIANES	0.0081	0.0059	0.0045	0.0033	0.0020	*	33.30	25.10	19.09	13.44	8.82	*	0.0239	*
ECARTS-TYPES	0.0015	0.0010	0.0009	0.0006	0.0005	*	4.07	2.63	2.51	1.87	1.96	*	0.0032	*
SEUIL 0.05	0.0114	0.0080	C.0061	0.0041	0.0029	*	41.61	29.35	22.52	16.09	11.34	*	0.0291	*

LONG.= 30	VP1	VP2	VP3	VP4	VP5	*	PC1	PC2	PC3	PC4	PC5	*	TRA	*
MOYENNES	0.0096	0.0072	0.0053	0.0041	0.0027	*	33.15	25.03	18.46	14.02	9.35	*	0.0289	*
MEDIANES	0.0094	0.0071	0.0053	0.0040	0.0026	*	32.94	24.97	18.48	14.00	9.43	*	0.0290	*
ECARTS-TYPES	0.0015	0.0010	C.0008	0.0007	0.0006	*	3.45	2.24	1.97	1.76	1.84	*	0.0033	*
SEUIL 0.05	0.0120	0.0088	C.0068	0.0052	0.0039	*	38.31	29.16	21.90	16.52	12.63	*	0.0336	*

LONG.= 35	VP1	VP2	VP3	VP4	VP5	*	PC1	PC2	PC3	PC4	PC5	*	TRA	*
MOYENNES	0.0110	0.0085	C.0065	0.0049	0.0034	*	32.05	24.83	18.89	14.32	9.92	*	0.0344	*
MEDIANES	0.0109	0.0085	C.0065	0.0048	0.0033	*	31.66	24.96	18.87	14.12	9.73	*	0.0344	*
ECARTS-TYPES	0.0017	0.0012	0.0009	0.0008	0.0007	*	3.20	2.19	2.04	1.85	1.66	*	0.0036	*
SEUIL 0.05	0.0141	0.0103	C.0080	0.0063	0.0046	*	37.88	27.72	22.07	17.79	12.88	*	0.0403	*

LONG.= 40	VP1	VP2	VP3	VP4	VP5	*	PC1	PC2	PC3	PC4	PC5	*	TRA	*
MOYENNES	0.0122	0.0095	0.0075	0.0057	0.0041	*	31.26	24.26	19.21	14.72	10.55	*	0.0390	*
MEDIANES	0.0120	0.0092	0.0073	0.0057	0.0041	*	30.55	24.30	19.18	14.92	10.29	*	0.0382	*
ECARTS-TYPES	0.0019	0.0014	0.0010	0.0007	0.0008	*	3.03	2.15	1.66	1.68	1.79	*	0.0042	*
SEUIL 0.05	0.0154	0.0116	0.0096	0.0068	0.0055	*	36.24	27.42	21.93	17.40	13.63	*	0.0464	*

LONG.= 45	VP1	VP2	VP3	VP4	VP5	*	PC1	PC2	PC3	PC4	PC5	*	TRA	*
MOYENNES	0.0138	0.0105	C.0085	0.0067	0.0049	*	31.08	23.79	19.13	15.03	10.97	*	0.0445	*
MEDIANES	0.0139	0.0105	0.0084	0.0067	0.0049	*	30.74	24.14	18.99	15.10	11.14	*	0.0451	*
ECARTS-TYPES	0.0016	0.0013	C.0012	0.0009	0.0009	*	2.46	1.73	1.84	1.43	1.62	*	0.0043	*
SEUIL 0.05	0.0161	0.0126	0.0106	0.0082	0.0063	*	35.28	26.56	22.24	17.27	13.78	*	0.0502	*

LONG.= 50	VP1	VP2	VP3	VP4	VP5	*	PC1	PC2	PC3	PC4	PC5	*	TRA	*
MOYENNES	0.0149	0.0118	0.0096	0.0075	0.0054	*	30.27	23.96	19.51	15.30	10.95	*	0.0494	*
MEDIANES	0.0148	0.0118	0.0097	0.0075	0.0053	*	29.74	23.73	19.35	15.61	10.87	*	0.0493	*
ECARTS-TYPES	0.0018	0.0015	C.0012	0.0010	0.0010	*	2.78	1.81	1.53	1.45	1.67	*	0.0046	*
SEUIL 0.05	0.0182	0.0146	0.0117	0.0090	0.0071	*	35.31	27.11	21.72	17.49	13.52	*	0.0561	*

SIGNIFICATION DES VALEURS PROPRES (VP),ET DES POURCENTAGES (PC) EN ANALYSE DES CORRESPONDANCES (EFFECTIFS= 5000.)

LARGEUR DU TABLEAU= 6

LONG.= 50	VP1	VP2	VP3	VP4	VP5	PC1	PC2	PC3	PC4	PC5	TRA
MOYENNES	0.0149	0.0118	0.0095	0.0076	0.0058	30.15	23.76	19.17	15.30	11.62	0.0496
MEDIANES	0.0147	0.0116	0.0095	0.0076	0.0057	29.97	23.65	19.06	15.39	11.66	0.0491
ECARTS-TYPES	0.0019	0.0015	0.0012	0.0012	0.0010	2.51	1.96	1.49	1.54	1.62	0.0050
SEUIL 0.05	0.0180	0.0142	0.0113	0.0096	0.0075	34.47	26.93	22.23	17.70	14.38	0.0576

LONG.= 60	VP1	VP2	VP3	VP4	VP5	PC1	PC2	PC3	PC4	PC5	TRA
MOYENNES	0.0176	0.0141	0.0115	0.0095	0.0074	29.29	23.40	19.15	15.89	12.28	0.0601
MEDIANES	0.0174	0.0139	0.0114	0.0095	0.0074	29.04	23.25	19.25	15.94	12.41	0.0598
ECARTS-TYPES	0.0019	0.0015	0.0013	0.0011	0.0011	2.23	1.62	1.36	1.40	1.51	0.0048
SEUIL 0.05	0.0205	0.0167	0.0137	0.0113	0.0092	32.76	26.21	21.35	18.31	14.26	0.0681

LONG.= 70	VP1	VP2	VP3	VP4	VP5	PC1	PC2	PC3	PC4	PC5	TRA
MOYENNES	0.0199	0.0163	0.0136	0.0112	0.0087	28.56	23.35	19.50	16.09	12.50	0.0698
MEDIANES	0.0197	0.0163	0.0135	0.0110	0.0087	28.57	23.23	19.71	16.11	12.49	0.0692
ECARTS-TYPES	0.0023	0.0016	0.0016	0.0014	0.0012	2.25	1.50	1.42	1.46	1.39	0.0058
SEUIL 0.05	0.0239	0.0188	0.0160	0.0133	0.0110	32.64	26.40	21.57	18.60	14.88	0.0790

LONG.= 80	VP1	VP2	VP3	VP4	VP5	PC1	PC2	PC3	PC4	PC5	TRA
MOYENNES	0.0223	0.0186	0.0156	0.0131	0.0103	27.94	23.26	19.55	16.39	12.87	0.0798
MEDIANES	0.0224	0.0184	0.0154	0.0128	0.0102	27.86	23.27	19.55	16.33	12.87	0.0797
ECARTS-TYPES	0.0024	0.0018	0.0017	0.0015	0.0014	2.00	1.21	1.38	1.33	1.25	0.0065
SEUIL 0.05	0.0261	0.0214	0.0182	0.0161	0.0124	31.51	25.19	21.44	18.32	14.80	0.0922

LONG.= 90	VP1	VP2	VP3	VP4	VP5	PC1	PC2	PC3	PC4	PC5	TRA
MOYENNES	0.0251	0.0210	0.0178	0.0152	0.0122	27.49	22.97	19.50	16.66	13.38	0.0913
MEDIANES	0.0249	0.0208	0.0176	0.0152	0.0121	27.29	22.98	19.64	16.65	13.58	0.0900
ECARTS-TYPES	0.0026	0.0021	0.0018	0.0017	0.0014	1.96	1.35	1.17	1.30	1.21	0.0069
SEUIL 0.05	0.0303	0.0247	0.0207	0.0177	0.0146	31.40	25.19	21.25	18.68	15.07	0.1026

LONG.=100	VP1	VP2	VP3	VP4	VP5	PC1	PC2	PC3	PC4	PC5	TRA
MOYENNES	0.0272	0.0229	0.0199	0.0169	0.0138	27.05	22.80	19.72	16.74	13.69	0.1007
MEDIANES	0.0269	0.0231	0.0199	0.0169	0.0136	26.88	22.74	19.72	16.80	13.69	0.1006
ECARTS-TYPES	0.0024	0.0018	0.0017	0.0016	0.0015	1.76	1.10	1.02	1.26	1.22	0.0062
SEUIL 0.05	0.0314	0.0255	0.0221	0.0192	0.0165	29.69	24.49	21.22	18.73	15.48	0.1113

SIGNIFICATION DES VALEURS PROPRES (VP),ET DES POURCENTAGES (PC) EN ANALYSE DES CORRESPONDANCES (EFFECTIFS= 1000.)

LARGEUR DU TABLEAU= 8

LONG.= 8	VP1	VP2	VP3	VP4	VP5	PC1	PC2	PC3	PC4	PC5	TRA
MOYENNES	0.0205	0.0129	0.0079	0.0043	0.0021	42.43	26.50	16.35	8.81	4.30	0.0484
MEDIANES	0.0202	0.0120	0.0077	0.0041	0.0019	41.42	26.39	16.46	8.51	4.09	0.0468
ECARTS-TYPES	0.0053	0.0036	0.0021	0.0017	0.0011	5.85	3.75	3.01	2.54	1.85	0.0107
SEUIL 0.05	0.0303	0.0200	0.0115	0.0075	0.0035	53.64	32.99	20.55	13.44	7.41	0.0701

LONG.= 10	VP1	VP2	VP3	VP4	VP5	PC1	PC2	PC3	PC4	PC5	TRA
MOYENNES	0.0246	0.0156	0.0104	0.0064	0.0039	38.81	24.75	16.51	10.23	6.19	0.0633
MEDIANES	0.0240	0.0159	0.0101	0.0065	0.0037	37.77	24.57	16.29	10.18	5.98	0.0611
ECARTS-TYPES	0.0052	0.0034	0.0024	0.0017	0.0014	4.48	3.47	2.41	2.22	1.94	0.0108
SEUIL 0.05	0.0345	0.0226	0.0144	0.0095	0.0067	46.48	30.54	20.69	13.64	9.30	0.0822

LONG.= 12	VP1	VP2	VP3	VP4	VP5	PC1	PC2	PC3	PC4	PC5	TRA
MOYENNES	0.0291	0.0192	0.0128	0.0085	0.0051	37.11	24.52	16.36	10.82	6.43	0.0784
MEDIANES	0.0284	0.0188	0.0124	0.0084	0.0049	36.90	24.81	16.64	10.80	6.53	0.0786
ECARTS-TYPES	0.0061	0.0041	0.0030	0.0021	0.0016	4.80	3.41	2.51	1.88	1.61	0.0133
SEUIL 0.05	0.0395	0.0252	0.0179	0.0119	0.0077	45.79	30.06	20.22	14.05	9.03	0.0979

LONG.= 14	VP1	VP2	VP3	VP4	VP5	PC1	PC2	PC3	PC4	PC5	TRA
MOYENNES	0.0320	0.0220	0.0159	0.0109	0.0071	34.08	23.48	16.82	11.63	7.60	0.0940
MEDIANES	0.0319	0.0217	0.0154	0.0108	0.0070	33.75	23.55	16.38	11.59	7.28	0.0930
ECARTS-TYPES	0.0062	0.0040	0.0035	0.0024	0.0019	3.92	2.59	2.30	1.71	1.67	0.0146
SEUIL 0.05	0.0429	0.0286	0.0220	0.0152	0.0104	41.64	28.07	20.68	14.31	10.48	0.1201

LONG.= 16	VP1	VP2	VP3	VP4	VP5	PC1	PC2	PC3	PC4	PC5	TRA
MOYENNES	0.0367	0.0249	0.0183	0.0129	0.0087	33.41	22.79	16.66	11.85	7.95	0.1095
MEDIANES	0.0358	0.0244	0.0174	0.0127	0.0088	32.85	22.96	16.47	11.83	7.84	0.1079
ECARTS-TYPES	0.0076	0.0046	0.0039	0.0025	0.0022	3.82	2.52	2.16	1.77	1.69	0.0167
SEUIL 0.05	0.0516	0.0326	0.0252	0.0166	0.0131	40.36	26.84	20.30	14.60	10.82	0.1409

LONG.= 18	VP1	VP2	VP3	VP4	VP5	PC1	PC2	PC3	PC4	PC5	TRA
MOYENNES	0.0391	0.0276	0.0202	0.0148	0.0101	31.91	22.58	16.55	12.11	8.29	0.1225
MEDIANES	0.0382	0.0273	0.0193	0.0146	0.0100	31.32	22.50	16.64	12.04	8.10	0.1168
ECARTS-TYPES	0.0074	0.0048	0.0037	0.0032	0.0022	3.72	2.33	2.02	1.78	1.50	0.0181
SEUIL 0.05	0.0509	0.0360	0.0274	0.0203	0.0135	37.92	26.16	20.13	15.13	10.75	0.1537

LONG.= 20	VP1	VP2	VP3	VP4	VP5	PC1	PC2	PC3	PC4	PC5	TRA
MOYENNES	0.0426	0.0303	0.0226	0.0165	0.0113	30.99	22.23	16.62	12.20	8.63	0.1365
MEDIANES	0.0399	0.0296	0.0223	0.0165	0.0117	30.20	22.13	16.72	12.06	8.60	0.1343
ECARTS-TYPES	0.0096	0.0054	0.0038	0.0027	0.0027	3.89	2.28	1.92	1.85	1.45	0.0201
SEUIL 0.05	0.0620	0.0403	0.0289	0.0210	0.0156	38.02	26.35	19.68	15.28	10.68	0.1700

SIGNIFICATION DES VALEURS PROPRES (VP),ET DES POURCENTAGES (PC) EN ANALYSE DES CORRESPONDANCES (EFFECTIFS= 5000.)

LARGEUR DU TABLEAU= 8

LONG.= 25	VP1	VP2	VP3	VP4	VP5	*	PC1	PC2	PC3	PC4	PC5	*	TRA
MOYENNES	0.0095	0.0072	0.0056	0.0044	0.0032	*	28.29	21.42	16.61	12.96	9.45	*	0.0337
MEDIANES	0.0093	0.0072	0.0054	0.0042	0.0031	*	27.98	21.58	16.44	13.15	9.32	*	0.0336
ECARTS-TYPES	0.0016	0.0011	0.0010	0.0008	0.0005	*	3.12	1.88	1.61	1.49	1.31	*	0.0040
SEUIL 0.05	0.0123	0.0088	0.0076	0.0059	0.0040	*	33.34	24.51	19.36	15.08	11.86	*	0.0402

LONG.= 30	VP1	VP2	VP3	VP4	VP5	*	PC1	PC2	PC3	PC4	PC5	*	TRA
MOYENNES	0.0110	0.0085	0.0067	0.0054	0.0042	*	26.80	20.77	16.36	13.23	10.33	*	0.0408
MEDIANES	0.0107	0.0084	0.0067	0.0054	0.0042	*	26.14	20.63	16.22	13.33	10.28	*	0.0407
ECARTS-TYPES	0.0017	0.0011	0.0008	0.0008	0.0007	*	2.72	1.78	1.56	1.35	1.28	*	0.0041
SEUIL 0.05	0.0140	0.0105	0.0079	0.0065	0.0053	*	31.48	23.71	19.03	15.26	12.37	*	0.0479

LONG.= 35	VP1	VP2	VP3	VP4	VP5	*	PC1	PC2	PC3	PC4	PC5	*	TRA
MOYENNES	0.0123	0.0098	0.0080	0.0063	0.0050	*	25.68	20.44	16.62	13.11	10.50	*	0.0481
MEDIANES	0.0121	0.0096	0.0079	0.0062	0.0050	*	25.38	20.50	16.73	13.37	10.51	*	0.0474
ECARTS-TYPES	0.0015	0.0011	0.0010	0.0009	0.0008	*	2.23	1.54	1.37	1.21	1.25	*	0.0046
SEUIL 0.05	0.0151	0.0116	0.0097	0.0080	0.0064	*	29.99	23.22	18.88	14.85	12.67	*	0.0572

LONG.= 40	VP1	VP2	VP3	VP4	VP5	*	PC1	PC2	PC3	PC4	PC5	*	TRA
MOYENNES	0.0140	0.0111	0.0090	0.0074	0.0060	*	25.27	19.98	16.22	13.38	10.85	*	0.0554
MEDIANES	0.0138	0.0108	0.0090	0.0073	0.0060	*	25.04	19.57	16.16	13.33	10.96	*	0.0556
ECARTS-TYPES	0.0017	0.0015	0.0012	0.0009	0.0007	*	2.28	1.86	1.47	1.25	1.05	*	0.0045
SEUIL 0.05	0.0172	0.0142	0.0107	0.0088	0.0070	*	29.47	23.36	18.55	15.47	12.55	*	0.0623

LONG.= 45	VP1	VP2	VP3	VP4	VP5	*	PC1	PC2	PC3	PC4	PC5	*	TRA
MOYENNES	0.0153	0.0122	0.0101	0.0084	0.0068	*	24.54	19.50	16.19	13.47	10.98	*	0.0624
MEDIANES	0.0150	0.0120	0.0100	0.0082	0.0068	*	24.27	19.34	16.28	13.37	11.00	*	0.0614
ECARTS-TYPES	0.0020	0.0014	0.0010	0.0009	0.0008	*	2.08	1.52	1.15	0.98	0.94	*	0.0053
SEUIL 0.05	0.0187	0.0142	0.0116	0.0103	0.0082	*	28.56	21.85	17.93	15.23	12.49	*	0.0712

LONG.= 50	VP1	VP2	VP3	VP4	VP5	*	PC1	PC2	PC3	PC4	PC5	*	TRA
MOYENNES	0.0165	0.0133	0.0111	0.0093	0.0076	*	23.99	19.37	16.14	13.55	11.05	*	0.0688
MEDIANES	0.0162	0.0131	0.0111	0.0094	0.0076	*	23.80	19.42	15.96	13.67	11.03	*	0.0691
ECARTS-TYPES	0.0018	0.0013	0.0011	0.0009	0.0007	*	1.88	1.39	1.02	0.96	0.91	*	0.0048
SEUIL 0.05	0.0196	0.0155	0.0128	0.0107	0.0088	*	27.05	21.82	17.64	14.97	12.45	*	0.0773

SIGNIFICATION DES VALEURS PROPRES (VP),ET DES POURCENTAGES (PC) EN ANALYSE DES CORRESPONDANCES (EFFECTIFS= 5000.)

LARGEUR DU TABLEAU= 8

LONG.= 50	VP1	VP2	VP3	VP4	VP5	*	PC1	PC2	PC3	PC4	PC5	*	TRA	*
MOYENNES	0.0165	0.0136	0.0113	0.0094	0.0078	*	23.62	19.41	16.14	13.49	11.22	*	0.0699	*
MEDIANES	0.0163	0.0137	0.0112	0.0094	0.0078	*	23.12	19.36	16.14	13.50	11.18	*	0.0695	*
ECARTS-TYPES	0.0021	0.0015	0.0013	0.0011	0.0010	*	2.09	1.31	1.14	1.01	1.07	*	0.0058	*
SEUIL 0.05	0.0199	0.0160	0.0134	0.0112	0.0099	*	27.33	21.57	18.21	15.21	13.03	*	0.0789	*

LONG.= 60	VP1	VP2	VP3	VP4	VP5	*	PC1	PC2	PC3	PC4	PC5	*	TRA	*
MOYENNES	0.0192	0.0159	0.0135	0.0113	0.0096	*	23.00	19.05	16.17	13.59	11.47	*	0.0833	*
MEDIANES	0.0188	0.0158	0.0133	0.0113	0.0096	*	22.75	19.12	16.20	13.60	11.47	*	0.0841	*
ECARTS-TYPES	0.0022	0.0016	0.0012	0.0011	0.0011	*	1.76	1.16	0.95	0.87	0.83	*	0.0068	*
SEUIL 0.05	0.0234	0.0188	0.0155	0.0129	0.0110	*	26.10	21.03	17.62	15.10	12.70	*	0.0936	*

LONG.= 70	VP1	VP2	VP3	VP4	VP5	*	PC1	PC2	PC3	PC4	PC5	*	TRA	*
MOYENNES	0.0222	0.0184	0.0158	0.0137	0.0115	*	22.37	18.57	15.96	13.78	11.63	*	0.0992	*
MEDIANES	0.0218	0.0185	0.0158	0.0135	0.0115	*	22.20	18.52	15.91	13.81	11.68	*	0.0989	*
ECARTS-TYPES	0.0025	0.0016	0.0014	0.0012	0.0010	*	1.70	1.02	0.99	0.91	0.81	*	0.0065	*
SEUIL 0.05	0.0262	0.0207	0.0182	0.0155	0.0132	*	25.24	20.53	17.52	15.52	13.09	*	0.1089	*

LONG.= 80	VP1	VP2	VP3	VP4	VP5	*	PC1	PC2	PC3	PC4	PC5	*	TRA	*
MOYENNES	0.0242	0.0205	0.0179	0.0156	0.0134	*	21.62	18.24	15.95	13.89	11.95	*	0.1121	*
MEDIANES	0.0241	0.0202	0.0179	0.0154	0.0131	*	21.52	18.21	15.93	13.94	11.92	*	0.1121	*
ECARTS-TYPES	0.0025	0.0021	0.0017	0.0013	0.0011	*	1.52	1.13	0.89	0.78	0.79	*	0.0073	*
SEUIL 0.05	0.0277	0.0239	0.0204	0.0174	0.0152	*	24.20	20.05	17.64	15.08	13.22	*	0.1248	*

LONG.= 90	VP1	VP2	VP3	VP4	VP5	*	PC1	PC2	PC3	PC4	PC5	*	TRA	*
MOYENNES	0.0270	0.0228	0.0201	0.0174	0.0151	*	21.40	18.12	15.95	13.82	12.01	*	0.1260	*
MEDIANES	0.0269	0.0230	0.0202	0.0174	0.0151	*	21.28	18.04	16.02	13.84	11.99	*	0.1247	*
ECARTS-TYPES	0.0027	0.0017	0.0017	0.0018	0.0014	*	1.48	0.88	0.79	0.78	0.73	*	0.0084	*
SEUIL 0.05	0.0313	0.0256	0.0226	0.0202	0.0180	*	23.68	19.68	17.11	15.12	13.14	*	0.1397	*

LONG.=100	VP1	VP2	VP3	VP4	VP5	*	PC1	PC2	PC3	PC4	PC5	*	TRA	*
MOYENNES	0.0300	0.0253	0.0223	0.0197	0.0173	*	21.09	17.83	15.69	13.84	12.21	*	0.1421	*
MEDIANES	0.0296	0.0251	0.0222	0.0194	0.0173	*	20.87	17.71	15.68	13.84	12.15	*	0.1420	*
ECARTS-TYPES	0.0026	0.0021	0.0017	0.0015	0.0013	*	1.38	0.96	0.83	0.67	0.67	*	0.0078	*
SEUIL 0.05	0.0347	0.0291	0.0253	0.0222	0.0192	*	23.40	19.48	17.11	14.96	13.30	*	0.1558	*

SIGNIFICATION DES VALEURS PROPRES (VP),ET DES POURCENTAGES (PC) EN ANALYSE DES CORRESPONDANCES (EFFECTIFS= 1000.)

LARGEUR DU TABLEAU= 10

LONG.= 10	VP1	VP2	VP3	VP4	VP5	PC1	PC2	PC3	PC4	PC5	TRA
MOYENNES	0.0284	0.0186	C.0128	0.CC89	0.0055	35.63	23.25	15.99	11.17	6.89	0.0799
MEDIANES	0.0285	0.0184	C.0127	0.CC89	0.0053	35.44	23.17	16.11	11.12	6.81	0.0798
ECARTS-TYPES	0.0053	0.0037	C.0028	0.0022	0.0016	4.39	2.83	2.32	2.01	1.62	0.0124
SEUIL 0.05	0.0374	0.0251	C.0171	0.0127	0.0085	42.55	28.00	19.34	14.19	9.64	0.0986

LONG.= 12	VP1	VP2	VP3	VP4	VP5	PC1	PC2	PC3	PC4	PC5	TRA
MOYENNES	0.0329	0.C235	C.0170	0.0118	0.0080	32.02	23.01	16.64	11.48	7.77	0.1024
MEDIANES	0.0316	0.0233	0.0166	0.0113	0.0078	31.76	23.01	16.66	11.45	7.71	0.1019
ECARTS-TYPES	0.0072	0.0043	C.0032	0.0028	0.0021	3.81	2.58	2.05	1.97	1.52	0.0168
SEUIL 0.05	0.0451	0.0307	C.0233	0.0167	0.0114	38.95	27.58	20.15	15.22	10.31	0.1297

LONG.= 14	VP1	VP2	VP3	VP4	VP5	PC1	PC2	PC3	PC4	PC5	TRA
MOYENNES	0.0346	0.0254	C.0189	0.0140	0.0099	29.47	21.61	16.15	12.00	8.46	0.1172
MEDIANES	0.0332	0.0247	C.0190	0.0138	0.0098	28.88	21.63	16.37	11.98	8.25	0.1141
ECARTS-TYPES	0.0074	0.0046	0.0033	0.0025	0.0021	4.03	2.07	1.93	1.60	1.40	0.0168
SEUIL 0.05	0.0498	0.0337	C.0243	0.0186	0.0136	36.56	24.76	19.03	14.60	10.77	0.1474

LONG.= 16	VP1	VP2	VP3	VP4	VP5	PC1	PC2	PC3	PC4	PC5	TRA
MOYENNES	0.0414	0.0294	C.0213	0.C161	0.0118	29.78	21.23	15.38	11.60	8.52	0.1387
MEDIANES	0.0396	0.0282	0.0208	0.0160	0.0113	29.07	20.80	15.50	11.48	8.39	0.1353
ECARTS-TYPES	0.0086	0.0053	C.0035	0.C032	0.0029	3.78	2.46	1.64	1.57	1.57	0.0198
SEUIL 0.05	0.0566	0.0381	0.0275	0.0211	0.0166	36.27	25.94	18.24	13.92	11.10	0.1762

LONG.= 18	VP1	VP2	VP3	VP4	VP5	PC1	PC2	PC3	PC4	PC5	TRA
MOYENNES	0.0446	0.0325	C.0246	0.0190	0.0139	28.26	20.58	15.62	12.02	8.80	0.1579
MEDIANES	0.0429	0.0319	0.0241	0.C190	0.0140	27.72	20.52	15.43	12.10	8.71	0.1568
ECARTS-TYPES	0.0082	0.0053	C.0042	0.0034	0.0028	3.26	1.96	1.54	1.44	1.20	0.0222
SEUIL 0.05	0.0593	0.0412	0.0335	0.0241	0.0178	34.36	24.54	18.14	14.30	10.68	0.1907

LONG.= 20	VP1	VP2	VP3	VP4	VP5	PC1	PC2	PC3	PC4	PC5	TRA
MOYENNES	0.0472	0.0352	C.0272	0.0212	0.0156	26.87	20.06	15.50	12.08	9.03	0.1756
MEDIANES	0.0460	0.0347	0.0267	0.0204	0.0155	26.23	19.98	15.50	12.21	8.93	0.1731
ECARTS-TYPES	0.0086	0.0056	C.0044	0.C037	0.0027	2.96	1.79	1.62	1.34	1.07	0.0233
SEUIL 0.05	0.0651	0.0436	0.0338	0.0283	0.0200	32.45	23.00	17.80	14.22	10.74	0.2148

SIGNIFICATION DES VALEURS PROPRES (VP),ET DES POURCENTAGES (PC) EN ANALYSE DES CORRESPONDANCES (EFFECTIFS= 5000.)

LARGEUR DU TABLEAU= 10

LONG.= 25	VP1	VP2	VP3	VP4	VP5	*	PC1	PC2	PC3	PC4	PC5	*	TRA	*
MOYENNES	0.0106	0.0081	0.0065	0.0052	0.0042	*	24.70	18.92	15.12	12.06	9.75	*	0.0430	*
MEDIANES	0.0107	0.0082	0.0064	0.0051	0.0042	*	24.73	18.84	15.29	12.10	9.77	*	0.0434	*
ECARTS-TYPES	0.0014	0.0010	0.0009	0.0007	0.0005	*	2.24	1.67	1.45	1.14	0.95	*	0.0042	*
SEUIL 0.05	0.0129	0.0097	0.0081	0.0065	0.0051	*	28.53	21.73	17.34	13.93	11.11	*	0.0496	*

LONG.= 30	VP1	VP2	VP3	VP4	VP5	*	PC1	PC2	PC3	PC4	PC5	*	TRA	*
MOYENNES	0.0124	0.0098	0.0079	0.0065	0.0052	*	23.39	18.45	14.90	12.23	9.81	*	0.0528	*
MEDIANES	0.0121	0.0097	0.0078	0.0064	0.0052	*	23.38	18.31	14.75	12.28	9.83	*	0.0525	*
ECARTS-TYPES	0.0018	0.0013	0.0010	0.0008	0.0007	*	2.37	1.55	1.24	1.09	1.00	*	0.0051	*
SEUIL 0.05	0.0157	0.0117	0.0097	0.0079	0.0063	*	27.35	21.29	17.03	14.01	11.41	*	0.0612	*

LONG.= 35	VP1	VP2	VP3	VP4	VP5	*	PC1	PC2	PC3	PC4	PC5	*	TRA	*
MOYENNES	0.0139	0.0110	0.0091	0.0076	0.0062	*	22.47	17.73	14.65	12.29	9.93	*	0.0620	*
MEDIANES	0.0138	0.0109	0.0091	0.0076	0.0061	*	22.29	17.62	14.66	12.28	9.91	*	0.0617	*
ECARTS-TYPES	0.0018	0.0013	0.0010	0.0009	0.0007	*	2.10	1.37	1.08	1.07	0.85	*	0.0051	*
SEUIL 0.05	0.0172	0.0130	0.0106	0.0090	0.0074	*	26.04	20.01	16.41	14.10	11.29	*	0.0717	*

LONG.= 40	VP1	VP2	VP3	VP4	VP5	*	PC1	PC2	PC3	PC4	PC5	*	TRA	*
MOYENNES	0.0151	0.0123	0.0103	0.0086	0.0071	*	21.40	17.35	14.56	12.14	10.00	*	0.0707	*
MEDIANES	0.0149	0.0122	0.0104	0.0086	0.0070	*	21.01	17.32	14.66	12.12	9.89	*	0.0702	*
ECARTS-TYPES	0.0019	0.0014	0.0012	0.0010	0.0008	*	1.99	1.22	1.04	1.01	0.77	*	0.0060	*
SEUIL 0.05	0.0191	0.0149	0.0121	0.0101	0.0085	*	25.12	19.48	16.27	13.83	11.38	*	0.0802	*

LONG.= 45	VP1	VP2	VP3	VP4	VP5	*	PC1	PC2	PC3	PC4	PC5	*	TRA	*
MOYENNES	0.0166	0.0136	0.0116	0.0098	0.0082	*	20.70	16.93	14.43	12.24	10.27	*	0.0801	*
MEDIANES	0.0165	0.0135	0.0116	0.0097	0.0083	*	20.58	17.04	14.32	12.16	10.25	*	0.0801	*
ECARTS-TYPES	0.0018	0.0014	0.0012	0.0009	0.0009	*	1.81	1.13	0.95	0.75	0.79	*	0.0058	*
SEUIL 0.05	0.0196	0.0162	0.0132	0.0113	0.0096	*	24.04	18.78	16.17	13.40	11.35	*	0.0902	*

LONG.= 50	VP1	VP2	VP3	VP4	VP5	*	PC1	PC2	PC3	PC4	PC5	*	TRA	*
MOYENNES	0.0181	0.0149	0.0125	0.0108	0.0091	*	20.39	16.69	14.06	12.17	10.26	*	0.0890	*
MEDIANES	0.0181	0.0148	0.0126	0.0107	0.0091	*	20.33	16.43	14.04	12.08	10.26	*	0.0890	*
ECARTS-TYPES	0.0018	0.0015	0.0011	0.0010	0.0009	*	1.77	1.17	0.93	0.84	0.74	*	0.0059	*
SEUIL 0.05	0.0215	0.0171	0.0140	0.0124	0.0105	*	23.09	18.62	15.50	13.57	11.36	*	0.0973	*

SIGNIFICATION DES VALEURS PROPRES (VP),ET DES POURCENTAGES (PC) EN ANALYSE DES CORRESPONDANCES (EFFECTIFS= 5000.)

LARGEUR DU TABLEAU= 10

LONG.= 50	VP1	VP2	VP3	VP4	VP5	PC1	PC2	PC3	PC4	PC5	TRA
MOYENNES	0.0182	0.0150	0.0128	0.0109	0.0093	20.30	16.79	14.22	12.19	10.36	0.0897
MEDIANES	0.0180	0.0148	0.0126	0.0107	0.0092	19.87	16.61	14.21	12.19	10.31	0.0893
ECARTS-TYPES	0.0020	0.0015	0.0012	0.0011	0.0010	1.68	1.30	0.98	0.88	0.89	0.0063
SEUIL 0.05	0.0214	0.0180	0.0145	0.0127	0.0108	23.58	18.96	16.02	13.71	11.94	0.1007

LONG.= 60	VP1	VP2	VP3	VP4	VP5	PC1	PC2	PC3	PC4	PC5	TRA
MOYENNES	0.0210	0.0175	0.0153	0.0133	0.0114	19.28	16.05	14.04	12.24	10.47	0.1090
MEDIANES	0.0205	0.0174	0.0153	0.0132	0.0113	19.09	15.96	13.99	12.24	10.45	0.1092
ECARTS-TYPES	0.0024	0.0016	0.0014	0.0012	0.0011	1.67	0.97	0.81	0.66	0.67	0.0076
SEUIL 0.05	0.0252	0.0200	0.0178	0.0153	0.0135	21.81	17.70	15.55	13.29	11.49	0.1201

LONG.= 70	VP1	VP2	VP3	VP4	VP5	PC1	PC2	PC3	PC4	PC5	TRA
MOYENNES	0.0236	0.0201	0.0176	0.0153	0.0133	18.64	15.88	13.89	12.11	10.54	0.1265
MEDIANES	0.0234	0.0200	0.0174	0.0152	0.0133	18.56	15.82	13.84	12.12	10.56	0.1254
ECARTS-TYPES	0.0021	0.0017	0.0012	0.0012	0.0009	1.13	0.92	0.65	0.66	0.52	0.0071
SEUIL 0.05	0.0277	0.0229	0.0194	0.0175	0.0147	20.50	17.32	14.91	13.28	11.40	0.1392

LONG.= 80	VP1	VP2	VP3	VP4	VP5	PC1	PC2	PC3	PC4	PC5	TRA
MOYENNES	0.0258	0.0224	0.0195	0.0172	0.0153	18.05	15.69	13.59	12.00	10.66	0.1431
MEDIANES	0.0259	0.0224	0.0193	0.0170	0.0152	17.95	15.62	13.59	11.93	10.67	0.1420
ECARTS-TYPES	0.0020	0.0017	0.0016	0.0014	0.0014	1.11	0.89	0.72	0.62	0.60	0.0082
SEUIL 0.05	0.0292	0.0252	0.0220	0.0198	0.0173	20.01	17.01	14.85	13.06	11.61	0.1577

LONG.= 90	VP1	VP2	VP3	VP4	VP5	PC1	PC2	PC3	PC4	PC5	TRA
MOYENNES	0.0289	0.0250	0.0221	0.0196	0.0175	17.65	15.23	13.49	11.96	10.65	0.1640
MEDIANES	0.0292	0.0249	0.0220	0.0196	0.0174	17.46	15.17	13.43	11.96	10.65	0.1616
ECARTS-TYPES	0.0026	0.0022	0.0016	0.0015	0.0014	1.21	0.85	0.66	0.51	0.48	0.0104
SEUIL 0.05	0.0337	0.0288	0.0254	0.0221	0.0193	19.84	16.87	14.50	12.80	11.48	0.1828

LONG.=100	VP1	VP2	VP3	VP4	VP5	PC1	PC2	PC3	PC4	PC5	TRA
MOYENNES	0.0319	0.0277	0.0245	0.0218	0.0194	17.55	15.23	13.45	11.99	10.70	0.1817
MEDIANES	0.0316	0.0275	0.0243	0.0218	0.0195	17.40	15.13	13.40	11.99	10.72	0.1825
ECARTS-TYPES	0.0029	0.0024	0.0020	0.0016	0.0015	1.14	0.85	0.62	0.58	0.54	0.0105
SEUIL 0.05	0.0373	0.0314	0.0281	0.0249	0.0221	19.42	16.47	14.60	12.93	11.55	0.1976

SIGNIFICATION DES VALEURS PROPRES (VP),ET DES POURCENTAGES (PC) EN ANALYSE DES CORRESPONDANCES (EFFECTIFS= 1000.)

LARGEUR DU TABLEAU= 12

LONG.= 12	VP1	VP2	VP3	VP4	VP5	*	PC1	PC2	PC3	PC4	PC5	*	TRA	*
MOYENNES	0.0377	0.0262	0.0193	0.0143	0.0103	*	30.55	21.20	15.52	11.53	8.29	*	0.1237	*
MEDIANES	0.0375	0.0256	0.0191	0.0143	0.0099	*	29.79	21.27	15.41	11.57	8.27	*	0.1224	*
ECARTS-TYPES	0.0066	0.0044	0.0038	0.0029	0.0022	*	4.18	2.14	1.86	1.52	1.32	*	0.0168	*
SEUIL 0.05	0.0508	0.0338	0.0260	0.0192	0.0145	*	39.53	24.63	18.31	13.79	10.44	*	0.1532	*

LONG.= 14	VP1	VP2	VP3	VP4	VP5	*	PC1	PC2	PC3	PC4	PC5	*	TRA	*
MOYENNES	0.0414	0.0294	0.0223	0.0168	0.0123	*	28.62	20.32	15.37	11.60	8.49	*	0.1448	*
MEDIANES	0.0408	0.0289	0.0220	0.0169	0.0122	*	28.78	20.30	15.72	11.52	8.30	*	0.1486	*
ECARTS-TYPES	0.0077	0.0050	0.0041	0.0032	0.0025	*	3.55	2.12	1.75	1.56	1.40	*	0.0192	*
SEUIL 0.05	0.0548	0.0386	0.0299	0.0223	0.0171	*	33.61	23.70	17.85	14.50	10.57	*	0.1810	*

LONG.= 16	VP1	VP2	VP3	VP4	VP5	*	PC1	PC2	PC3	PC4	PC5	*	TRA	*
MOYENNES	0.0461	0.0335	0.0253	0.0192	0.0146	*	27.25	19.83	14.98	11.38	8.64	*	0.1689	*
MEDIANES	0.0452	0.0331	0.0252	0.0190	0.0144	*	26.80	19.61	15.16	11.31	8.43	*	0.1677	*
ECARTS-TYPES	0.0079	0.0052	0.0039	0.0033	0.0025	*	3.19	1.94	1.54	1.48	1.24	*	0.0195	*
SEUIL 0.05	0.0597	0.0428	0.0318	0.0254	0.0192	*	33.89	23.17	17.42	13.91	10.83	*	0.2013	*

LONG.= 18	VP1	VP2	VP3	VP4	VP5	*	PC1	PC2	PC3	PC4	PC5	*	TRA	*
MOYENNES	0.0478	0.0352	0.0281	0.0219	0.0169	*	25.45	18.76	14.95	11.68	9.00	*	0.1879	*
MEDIANES	0.0466	0.0347	0.0278	0.0215	0.0163	*	25.00	18.74	15.09	11.53	9.02	*	0.1856	*
ECARTS-TYPES	0.0079	0.0052	0.0043	0.0033	0.0027	*	3.29	1.84	1.41	1.24	1.04	*	0.0205	*
SEUIL 0.05	0.0625	0.0440	0.0355	0.0282	0.0213	*	31.25	22.20	16.97	13.83	10.32	*	0.2201	*

LONG.= 20	VP1	VP2	VP3	VP4	VP5	*	PC1	PC2	PC3	PC4	PC5	*	TRA	*
MOYENNES	0.0519	0.0400	0.0315	0.0251	0.0196	*	24.01	18.54	14.62	11.64	9.09	*	0.2159	*
MEDIANES	0.0512	0.0407	0.0313	0.0247	0.0198	*	23.70	18.48	14.59	11.60	9.14	*	0.2153	*
ECARTS-TYPES	0.0088	0.0060	0.0044	0.0037	0.0032	*	2.62	1.72	1.22	1.15	1.14	*	0.0260	*
SEUIL 0.05	0.0658	0.0488	0.0392	0.0311	0.0248	*	28.26	21.80	16.89	13.51	11.07	*	0.2592	*

SIGNIFICATION DES VALEURS PROPRES (VP),ET DES POURCENTAGES (PC) EN ANALYSE DES CORRESPONDANCES (EFFECTIFS= 5000.)

LARGEUR DU TABLEAU= 12

LONG.= 25	VP1	VP2	VP3	VP4	VP5	PC1	PC2	PC3	PC4	PC5	TRA
MOYENNES	0.0116	0.0091	0.0073	0.0060	0.0049	22.24	17.50	14.04	11.45	9.47	0.0520
MEDIANES	0.0115	0.0091	0.0073	0.0060	0.0049	22.14	17.43	13.92	11.45	9.32	0.0515
ECARTS-TYPES	0.0016	0.0011	0.0009	0.0007	0.0006	2.20	1.55	1.29	0.93	0.88	0.0045
SEUIL 0.05	0.0141	0.0110	0.0087	0.0071	0.0060	26.14	20.13	16.18	12.87	10.83	0.0592

LONG.= 30	VP1	VP2	VP3	VP4	VP5	PC1	PC2	PC3	PC4	PC5	TRA
MOYENNES	0.0131	0.0106	0.0087	0.0072	0.0060	20.75	16.68	13.80	11.36	9.53	0.0633
MEDIANES	0.0130	0.0106	0.0086	0.0071	0.0060	20.38	16.63	13.82	11.30	9.52	0.0630
ECARTS-TYPES	0.0015	0.0012	0.0010	0.0009	0.0007	2.02	1.29	1.00	0.86	0.83	0.0055
SEUIL 0.05	0.0157	0.0124	0.0104	0.0085	0.0075	24.76	18.52	15.41	12.69	10.66	0.0718

LONG.= 35	VP1	VP2	VP3	VP4	VP5	PC1	PC2	PC3	PC4	PC5	TRA
MOYENNES	0.0148	0.0122	0.0102	0.0086	0.0072	19.70	16.19	13.60	11.40	9.59	0.0752
MEDIANES	0.0147	0.0120	0.0102	0.0085	0.0072	19.32	16.20	13.55	11.47	9.52	0.0755
ECARTS-TYPES	0.0017	0.0012	0.0011	0.0008	0.0007	1.81	1.17	1.08	0.75	0.70	0.0054
SEUIL 0.05	0.0182	0.0141	0.0122	0.0099	0.0083	23.15	17.85	15.26	12.54	10.66	0.0825

LONG.= 40	VP1	VP2	VP3	VP4	VP5	PC1	PC2	PC3	PC4	PC5	TRA
MOYENNES	0.0164	0.0134	0.0115	0.0098	0.0084	18.88	15.39	13.19	11.30	9.65	0.0869
MEDIANES	0.0164	0.0134	0.0115	0.0098	0.0084	18.80	15.31	13.21	11.18	9.76	0.0866
ECARTS-TYPES	0.0019	0.0013	0.0012	0.0011	0.0009	1.64	1.14	0.96	0.80	0.74	0.0066
SEUIL 0.05	0.0198	0.0154	0.0136	0.0117	0.0099	22.29	17.41	14.88	12.69	10.89	0.0975

LONG.= 45	VP1	VP2	VP3	VP4	VP5	PC1	PC2	PC3	PC4	PC5	TRA
MOYENNES	0.0183	0.0149	0.0127	0.0109	0.0095	18.56	15.11	12.92	11.04	9.61	0.0985
MEDIANES	0.0181	0.0147	0.0127	0.0107	0.0093	18.40	14.89	12.91	11.05	9.56	0.0986
ECARTS-TYPES	0.0021	0.0015	0.0012	0.0011	0.0009	1.73	1.05	0.76	0.80	0.64	0.0067
SEUIL 0.05	0.0219	0.0174	0.0150	0.0128	0.0111	22.03	17.50	14.13	12.44	10.85	0.1101

LONG.= 50	VP1	VP2	VP3	VP4	VP5	PC1	PC2	PC3	PC4	PC5	TRA
MOYENNES	0.0192	0.0161	0.0139	0.0120	0.0104	17.75	14.86	12.79	11.10	9.59	0.1084
MEDIANES	0.0186	0.0160	0.0137	0.0120	0.0102	17.49	14.72	12.68	11.06	9.58	0.1081
ECARTS-TYPES	0.0022	0.0015	0.0013	0.0011	0.0010	1.46	0.99	0.73	0.70	0.60	0.0079
SEUIL 0.05	0.0232	0.0186	0.0161	0.0139	0.0119	20.00	16.29	14.05	12.08	10.49	0.1200

SIGNIFICATION DES VALEURS PROPRES (VP),ET DES POURCENTAGES (PC) EN ANALYSE DES CORRESPONDANCES (EFFECTIFS= 5000.)

LARGEUR DU TABLEAU= 12

LONG.= 50	VP1	VP2	VP3	VP4	VP5	PC1	PC2	PC3	PC4	PC5	TRA
MOYENNES	0.0193	0.0161	0.0139	0.0121	0.0105	17.79	14.86	12.86	11.14	9.66	0.1083
MEDIANES	0.0192	0.0161	0.0137	0.0119	0.0102	17.65	14.79	12.80	11.06	9.67	0.1077
ECARTS-TYPES	0.0019	0.0015	0.0013	0.0011	0.0011	1.32	0.94	0.76	0.66	0.64	0.0075
SEUIL 0.05	0.0228	0.0185	0.0161	0.0139	0.0127	19.72	16.32	14.01	12.36	10.77	0.1209

LONG.= 60	VP1	VP2	VP3	VP4	VP5	PC1	PC2	PC3	PC4	PC5	TRA
MOYENNES	0.0224	0.0189	0.0164	0.0145	0.0127	17.00	14.41	12.48	11.01	9.67	0.1315
MEDIANES	0.0219	0.0188	0.0163	0.0144	0.0127	16.94	14.36	12.40	11.01	9.76	0.1314
ECARTS-TYPES	0.0024	0.0015	0.0013	0.0010	0.0011	1.31	0.87	0.68	0.56	0.60	0.0082
SEUIL 0.05	0.0269	0.0215	0.0185	0.0160	0.0143	19.38	15.85	13.70	11.91	10.74	0.1468

LONG.= 70	VP1	VP2	VP3	VP4	VP5	PC1	PC2	PC3	PC4	PC5	TRA
MOYENNES	0.0252	0.0214	0.0188	0.0169	0.0150	16.26	13.83	12.14	10.87	9.68	0.1551
MEDIANES	0.0251	0.0212	0.0188	0.0167	0.0150	16.17	13.71	12.12	10.92	9.66	0.1546
ECARTS-TYPES	0.0025	0.0017	0.0015	0.0013	0.0013	1.14	0.77	0.59	0.53	0.49	0.0097
SEUIL 0.05	0.0296	0.0243	0.0215	0.0188	0.0174	18.25	15.11	13.22	11.68	10.38	0.1731

LONG.= 80	VP1	VP2	VP3	VP4	VP5	PC1	PC2	PC3	PC4	PC5	TRA
MOYENNES	0.0273	0.0241	0.0214	0.0190	0.0172	15.47	13.66	12.11	10.78	9.74	0.1764
MEDIANES	0.0272	0.0240	0.0214	0.0192	0.0171	15.46	13.63	12.10	10.70	9.76	0.1771
ECARTS-TYPES	0.0022	0.0017	0.0014	0.0014	0.0013	0.97	0.70	0.56	0.51	0.50	0.0092
SEUIL 0.05	0.0311	0.0268	0.0235	0.0212	0.0191	17.13	14.88	13.11	11.65	10.56	0.1907

LONG.= 90	VP1	VP2	VP3	VP4	VP5	PC1	PC2	PC3	PC4	PC5	TRA
MOYENNES	0.0311	0.0271	0.0241	0.0215	0.0193	15.45	13.50	11.99	10.69	9.58	0.2011
MEDIANES	0.0306	0.0273	0.0242	0.0213	0.0192	15.48	13.40	12.00	10.71	9.59	0.2001
ECARTS-TYPES	0.0026	0.0021	0.0018	0.0016	0.0013	0.84	0.76	0.58	0.50	0.43	0.0107
SEUIL 0.05	0.0359	0.0300	0.0269	0.0239	0.0217	16.90	14.88	12.94	11.40	10.30	0.2190

LONG.=100	VP1	VP2	VP3	VP4	VP5	PC1	PC2	PC3	PC4	PC5	TRA
MOYENNES	0.0334	0.0292	0.0263	0.0237	0.0213	15.02	13.13	11.82	10.66	9.60	0.2222
MEDIANES	0.0331	0.0293	0.0261	0.0236	0.0214	14.84	13.14	11.74	10.71	9.60	0.2223
ECARTS-TYPES	0.0026	0.0018	0.0016	0.0015	0.0014	0.95	0.57	0.51	0.46	0.46	0.0098
SEUIL 0.05	0.0374	0.0318	0.0290	0.0261	0.0236	16.64	14.02	12.60	11.55	10.28	0.2352

SIGNIFICATION DES VALEURS PROPRES (VP),ET DES POURCENTAGES (PC) EN ANALYSE DES CORRESPONDANCES (EFFECTIFS= 1000.)

LARGEUR DU TABLEAU= 14

LONG.= 14	VP1	VP2	VP3	VP4	VP5	*	PC1	PC2	PC3	PC4	PC5	*	TRA	*
MOYENNES	0.0448	0.0340	0.0260	0.0200	0.0153	*	25.92	19.72	15.06	11.58	8.87	*	0.1727	*
MEDIANES	0.0437	0.0342	0.0259	0.0200	0.0152	*	25.79	19.78	14.80	11.62	8.74	*	0.1711	*
ECARTS-TYPES	0.0080	0.0046	0.0042	0.0031	0.0025	*	3.12	1.75	1.61	1.32	1.07	*	0.0199	*
SEUIL 0.05	0.0569	0.0415	0.0327	0.0254	0.0193	*	31.77	22.35	17.86	13.50	10.77	*	0.2077	*

LONG.= 16	VP1	VP2	VP3	VP4	VP5	*	PC1	PC2	PC3	PC4	PC5	*	TRA	*
MOYENNES	0.0492	0.0374	0.0290	0.0231	0.0182	*	24.46	18.64	14.42	11.52	9.04	*	0.2010	*
MEDIANES	0.0487	0.0370	0.0283	0.0229	0.0183	*	24.20	18.42	14.38	11.51	8.91	*	0.2004	*
ECARTS-TYPES	0.0092	0.0059	0.0049	0.0038	0.0034	*	2.77	1.65	1.30	1.12	1.07	*	0.0275	*
SEUIL 0.05	0.0658	0.0479	0.0381	0.0298	0.0239	*	29.35	21.54	16.64	13.33	10.76	*	0.2538	*

LONG.= 18	VP1	VP2	VP3	VP4	VP5	*	PC1	PC2	PC3	PC4	PC5	*	TRA	*
MOYENNES	0.0520	0.0406	0.0317	0.0255	0.0204	*	23.09	18.02	14.08	11.31	9.03	*	0.2254	*
MEDIANES	0.0517	0.0404	0.0311	0.0252	0.0201	*	22.62	18.03	14.00	11.37	9.05	*	0.2238	*
ECARTS-TYPES	0.0079	0.0055	0.0043	0.0034	0.0030	*	2.57	1.72	1.24	1.06	0.88	*	0.0237	*
SEUIL 0.05	0.0649	0.0497	0.0395	0.0306	0.0253	*	27.25	21.16	15.89	12.98	10.39	*	0.2663	*

LONG.= 20	VP1	VP2	VP3	VP4	VP5	*	PC1	PC2	PC3	PC4	PC5	*	TRA	*
MOYENNES	0.0562	0.0438	0.0350	0.0275	0.0225	*	22.61	17.55	14.06	11.06	9.01	*	0.2491	*
MEDIANES	0.0562	0.0431	0.0348	0.0271	0.0222	*	22.22	17.36	14.19	10.92	9.00	*	0.2485	*
ECARTS-TYPES	0.0078	0.0067	0.0046	0.0040	0.0036	*	2.43	1.47	1.24	1.03	0.86	*	0.0278	*
SEUIL 0.05	0.0703	0.0557	0.0422	0.0341	0.0286	*	26.33	19.95	15.85	12.58	10.64	*	0.2948	*

SIGNIFICATION DES VALEURS PROPRES (VP),ET DES POURCENTAGES (PC) EN ANALYSE DES CORRESPONDANCES (EFFECTIFS= 5000.)

LARGEUR DU TABLEAU= 14

LONG.= 25	VP1	VP2	VP3	VP4	VP5	*	PC1	PC2	PC3	PC4	PC5	*	TRA	*
MOYENNES	0.0130	0.0101	C.CC83	0.CC70	0.0059	*	20.55	15.98	13.19	11.16	9.29	*	0.0630	*
MEDIANES	0.0125	0.0100	C.CC83	0.CC70	0.0058	*	20.30	15.69	13.24	10.98	9.22	*	0.0626	*
ECARTS-TYPES	0.0017	0.0012	C.0009	0.0008	0.0007	*	2.08	1.39	1.10	0.85	C.82	*	0.0049	*
SEUIL 0.05	0.0156	0.0125	C.CC98	0.CC84	0.0069	*	23.90	18.42	14.95	12.70	10.66	*	0.0721	*

LONG.= 30	VP1	VP2	VP3	VP4	VP5	*	PC1	PC2	PC3	PC4	PC5	*	TRA	*
MOYENNES	0.0145	0.0117	0.0096	0.CC81	0.0068	*	19.12	15.40	12.69	10.70	9.03	*	0.0757	*
MEDIANES	0.0143	0.0115	C.CC96	0.CC80	0.0068	*	18.93	15.29	12.71	10.73	9.02	*	0.0762	*
ECARTS-TYPES	0.0019	0.0014	C.0011	0.0010	0.0008	*	1.81	1.16	0.86	0.94	0.70	*	0.0068	*
SEUIL 0.05	0.0178	0.0144	C.0115	0.CC96	0.0082	*	22.12	17.42	14.08	12.34	10.17	*	0.0862	*

LONG.= 35	VP1	VP2	VP3	VP4	VP5	*	PC1	PC2	PC3	PC4	PC5	*	TRA	*
MOYENNES	0.0159	0.0131	C.0112	0.CC95	0.0081	*	17.89	14.68	12.56	10.65	9.06	*	0.0890	*
MEDIANES	0.0157	0.0129	0.0112	0.CC94	0.0080	*	17.87	14.55	12.52	10.70	9.04	*	0.0882	*
ECARTS-TYPES	0.0017	0.0013	0.0010	0.0008	0.0008	*	1.43	0.93	0.74	0.68	0.63	*	0.0062	*
SEUIL 0.05	0.0189	0.0153	C.0128	0.C109	0.0093	*	20.09	16.48	13.87	11.62	10.04	*	0.0958	*

LONG.= 40	VP1	VP2	VP3	VP4	VP5	*	PC1	PC2	PC3	PC4	PC5	*	TRA	*
MOYENNES	0.0171	0.0144	0.0123	0.0105	0.0093	*	16.93	14.22	12.19	10.46	9.21	*	0.1012	*
MEDIANES	0.0169	0.0143	C.0123	0.0107	0.0094	*	16.84	14.17	12.13	10.43	9.18	*	0.1012	*
ECARTS-TYPES	0.0017	0.0012	0.0010	0.CC09	0.0009	*	1.41	0.93	0.74	0.67	0.58	*	0.0059	*
SEUIL 0.05	0.0206	0.0164	C.0140	0.0117	0.0107	*	19.43	15.72	13.29	11.62	10.10	*	0.1111	*

LONG.= 45	VP1	VP2	VP3	VP4	VP5	*	PC1	PC2	PC3	PC4	PC5	*	TRA	*
MOYENNES	0.0188	0.0160	C.0140	0.0120	C.0106	*	16.25	13.84	12.09	10.37	9.16	*	0.1158	*
MEDIANES	0.0185	0.0160	0.0137	0.0118	0.0106	*	16.27	13.78	11.96	10.36	9.27	*	0.1154	*
ECARTS-TYPES	0.0019	0.0015	0.0014	0.0011	0.0009	*	1.34	0.82	0.77	0.59	0.53	*	0.0081	*
SEUIL 0.05	0.0217	0.0189	0.0166	0.0137	0.0122	*	18.59	15.35	13.38	11.26	9.88	*	0.1304	*

LONG.= 50	VP1	VP2	VP3	VP4	VP5	*	PC1	PC2	PC3	PC4	PC5	*	TRA	*
MOYENNES	0.0205	0.0173	C.C153	0.0132	0.0117	*	15.88	13.42	11.82	10.25	9.11	*	0.1291	*
MEDIANES	0.0205	0.0172	0.0152	0.0131	0.0116	*	15.85	13.35	11.84	10.15	9.09	*	0.1292	*
ECARTS-TYPES	0.0022	0.0016	C.0014	0.CC11	0.0011	*	1.19	0.78	0.66	0.59	0.56	*	0.0088	*
SEUIL 0.05	0.0246	0.0198	0.0176	0.0149	0.0133	*	17.58	14.87	13.00	11.24	9.99	*	0.1433	*

SIGNIFICATION DES VALEURS PROPRES (VP),ET DES POURCENTAGES (PC) EN ANALYSE DES CORRESPONDANCES (EFFECTIFS= 5000.)

LARGEUR DU TABLEAU= 14

LONG.= 50	VP1	VP2	VP3	VP4	VP5	*	PC1	PC2	PC3	PC4	PC5	*	TRA	*
MOYENNES	0.0209	0.0175	0.0152	0.0132	0.0117	*	16.12	13.56	11.79	10.25	9.04	*	0.1293	*
MEDIANES	0.0205	0.0175	0.0150	0.0132	0.0116	*	15.79	13.46	11.69	10.30	9.08	*	0.1295	*
ECARTS-TYPES	0.0022	0.0015	0.0013	0.0011	0.0010	*	1.15	0.88	0.66	0.63	0.51	*	0.0080	*
SEUIL 0.05	0.0250	0.0197	0.0176	0.0151	0.0133	*	18.21	14.96	12.92	11.19	9.91	*	0.1417	*

LONG.= 60	VP1	VP2	VP3	VP4	VP5	*	PC1	PC2	PC3	PC4	PC5	*	TRA	*
MOYENNES	0.0239	0.0203	0.0177	0.0158	0.0140	*	15.30	13.01	11.32	10.08	8.97	*	0.1564	*
MEDIANES	0.0236	0.0201	0.0176	0.0157	0.0139	*	15.32	13.01	11.38	10.03	8.95	*	0.1568	*
ECARTS-TYPES	0.0025	0.0017	0.0014	0.0012	0.0011	*	1.20	0.84	0.63	0.50	0.41	*	0.0093	*
SEUIL 0.05	0.0278	0.0235	0.0202	0.0177	0.0157	*	17.15	14.70	12.44	10.98	9.61	*	0.1674	*

LONG.= 70	VP1	VP2	VP3	VP4	VP5	*	PC1	PC2	PC3	PC4	PC5	*	TRA	*
MOYENNES	0.0268	0.0229	0.0201	0.0181	0.0163	*	14.77	12.62	11.06	9.97	8.98	*	0.1814	*
MEDIANES	0.0266	0.0227	0.0200	0.0180	0.0164	*	14.54	12.46	11.00	9.98	9.00	*	0.1810	*
ECARTS-TYPES	0.0023	0.0018	0.0012	0.0012	0.0011	*	1.10	0.79	0.53	0.48	0.44	*	0.0087	*
SEUIL 0.05	0.0307	0.0262	0.0219	0.0200	0.0179	*	16.45	14.07	11.95	10.76	9.66	*	0.1952	*

LONG.= 80	VP1	VP2	VP3	VP4	VP5	*	PC1	PC2	PC3	PC4	PC5	*	TRA	*
MOYENNES	0.0296	0.0256	0.0231	0.0207	0.0186	*	14.10	12.21	10.98	9.85	8.87	*	0.2101	*
MEDIANES	0.0295	0.0253	0.0227	0.0205	0.0186	*	13.99	12.12	10.94	9.84	8.82	*	0.2088	*
ECARTS-TYPES	0.0027	0.0020	0.0015	0.0014	0.0013	*	0.88	0.62	0.48	0.47	0.43	*	0.0115	*
SEUIL 0.05	0.0340	0.0293	0.0259	0.0232	0.0206	*	15.77	13.28	11.77	10.73	9.59	*	0.2285	*

LONG.= 90	VP1	VP2	VP3	VP4	VP5	*	PC1	PC2	PC3	PC4	PC5	*	TRA	*
MOYENNES	0.0323	0.0284	0.0255	0.0232	0.0210	*	13.61	11.97	10.74	9.77	8.87	*	0.2371	*
MEDIANES	0.0320	0.0282	0.0254	0.0233	0.0212	*	13.52	11.91	10.74	9.77	8.80	*	0.2359	*
ECARTS-TYPES	0.0025	0.0021	0.0017	0.0019	0.0016	*	0.77	0.63	0.42	0.46	0.42	*	0.0131	*
SEUIL 0.05	0.0367	0.0322	0.0282	0.0260	0.0233	*	15.03	13.10	11.45	10.50	9.64	*	0.2557	*

LONG.=100	VP1	VP2	VP3	VP4	VP5	*	PC1	PC2	PC3	PC4	PC5	*	TRA	*
MOYENNES	0.0358	0.0315	0.0281	0.0255	0.0234	*	13.46	11.86	10.56	9.61	8.79	*	0.2659	*
MEDIANES	0.0354	0.0320	0.0279	0.0253	0.0233	*	13.33	11.88	10.56	9.63	8.83	*	0.2659	*
ECARTS-TYPES	0.0030	0.0023	0.0020	0.0017	0.0017	*	0.83	0.60	0.46	0.40	0.40	*	0.0133	*
SEUIL 0.05	0.0404	0.0347	0.0313	0.0285	0.0263	*	14.92	12.78	11.32	10.20	9.36	*	0.2865	*

SIGNIFICATION DES VALEURS PROPRES (VP),ET DES POURCENTAGES (PC) EN ANALYSE DES CORRESPONDANCES (EFFECTIFS= 1000.)

LARGEUR DU TABLEAU= 16

LONG.= 16	VP1	VP2	VP3	VP4	VP5	PC1	PC2	PC3	PC4	PC5	TRA
MOYENNES	0.0532	0.0403	0.0325	0.0262	0.0207	23.09	17.47	14.06	11.35	8.98	0.2309
MEDIANES	0.0527	0.0396	0.0322	0.0259	0.0203	22.71	17.27	13.95	11.32	8.97	0.2306
ECARTS-TYPES	0.0080	0.0052	0.0046	0.0033	0.0030	2.78	1.45	1.33	1.15	0.97	0.0233
SEUIL 0.05	0.0679	0.0496	0.0401	0.0319	0.0256	28.13	19.95	16.15	13.43	10.63	0.2706

LONG.= 18	VP1	VP2	VP3	VP4	VP5	PC1	PC2	PC3	PC4	PC5	TRA
MOYENNES	0.0574	0.0451	0.0363	0.0293	0.0238	21.80	17.07	13.74	11.10	9.03	0.2638
MEDIANES	0.0564	0.0452	0.0358	0.0292	0.0236	21.79	16.81	13.57	11.18	9.14	0.2660
ECARTS-TYPES	0.0078	0.0066	0.0048	0.0040	0.0032	2.35	1.74	1.16	0.99	0.86	0.0246
SEUIL 0.05	0.0721	0.0559	0.0453	0.0362	0.0289	26.71	20.31	15.82	12.73	10.27	0.2984

LONG.= 20	VP1	VP2	VP3	VP4	VP5	PC1	PC2	PC3	PC4	PC5	TRA
MOYENNES	0.0632	0.0483	0.0391	0.0325	0.0268	21.27	16.22	13.15	10.93	9.03	0.2972
MEDIANES	0.0623	0.0477	0.0387	0.0322	0.0266	21.07	16.26	13.20	10.97	9.03	0.2957
ECARTS-TYPES	0.0091	0.0065	0.0052	0.0045	0.0034	2.39	1.38	1.14	0.97	0.81	0.0277
SEUIL 0.05	0.0780	0.0595	0.0482	0.0403	0.0331	25.78	18.42	14.83	12.55	10.36	0.3473

LONG.= 25	VP1	VP2	VP3	VP4	VP5	PC1	PC2	PC3	PC4	PC5	TRA
MOYENNES	0.0136	0.0110	0.0090	0.0076	0.0064	18.94	15.29	12.57	10.50	8.94	0.0720
MEDIANES	0.0134	0.0110	0.0088	0.0075	0.0064	18.80	15.29	12.48	10.53	8.92	0.0713
ECARTS-TYPES	0.0016	0.0013	0.0011	0.0009	0.0007	1.60	1.33	0.93	0.70	0.63	0.0064
SEUIL 0.05	0.0162	0.0130	0.0108	0.0092	0.0078	21.66	17.66	14.15	11.46	9.89	0.0830

LONG.= 30	VP1	VP2	VP3	VP4	VP5	PC1	PC2	PC3	PC4	PC5	TRA
MOYENNES	0.0153	0.0124	0.0105	0.0091	0.0078	17.59	14.26	12.04	10.42	8.95	0.0870
MEDIANES	0.0151	0.0124	0.0104	0.0090	0.0077	17.51	14.19	11.90	10.36	8.91	0.0866
ECARTS-TYPES	0.0019	0.0013	0.0011	0.0008	0.0007	1.77	1.06	0.88	0.68	0.61	0.0058
SEUIL 0.05	0.0186	0.0148	0.0125	0.0104	0.0091	20.71	16.03	13.44	11.47	10.06	0.0986

LONG.= 35	VP1	VP2	VP3	VP4	VP5	PC1	PC2	PC3	PC4	PC5	TRA
MOYENNES	0.0173	0.0144	0.0122	0.0105	0.0092	16.63	13.78	11.68	10.09	8.82	0.1042
MEDIANES	0.0171	0.0142	0.0121	0.0104	0.0092	16.54	13.66	11.61	10.03	8.81	0.1044
ECARTS-TYPES	0.0021	0.0014	0.0011	0.0009	0.0009	1.46	1.03	0.74	0.56	0.61	0.0069
SEUIL 0.05	0.0207	0.0170	0.0138	0.0119	0.0106	19.33	15.62	12.86	11.10	9.65	0.1151

LONG.= 40	VP1	VP2	VP3	VP4	VP5	PC1	PC2	PC3	PC4	PC5	TRA
MOYENNES	0.0187	0.0157	0.0136	0.0119	0.0105	15.63	13.19	11.39	9.93	8.79	0.1194
MEDIANES	0.0185	0.0155	0.0136	0.0117	0.0105	15.47	13.10	11.36	9.88	8.83	0.1191
ECARTS-TYPES	0.0019	0.0014	0.0011	0.0011	0.0009	1.17	0.81	0.66	0.64	0.52	0.0076
SEUIL 0.05	0.0221	0.0182	0.0156	0.0138	0.0120	17.92	14.43	12.39	11.04	9.57	0.1314

LONG.= 45	VP1	VP2	VP3	VP4	VP5	PC1	PC2	PC3	PC4	PC5	TRA
MOYENNES	0.0203	0.0172	0.0149	0.0131	0.0116	15.13	12.87	11.14	9.83	8.67	0.1337
MEDIANES	0.0199	0.0168	0.0147	0.0129	0.0116	14.99	12.72	11.13	9.81	8.66	0.1330
ECARTS-TYPES	0.0022	0.0015	0.0010	0.0009	0.0008	1.14	0.81	0.64	0.57	0.50	0.0082
SEUIL 0.05	0.0242	0.0203	0.0167	0.0148	0.0129	17.05	14.12	12.19	10.78	9.43	0.1488

LONG.= 50	VP1	VP2	VP3	VP4	VP5	PC1	PC2	PC3	PC4	PC5	TRA
MOYENNES	0.0216	0.0184	0.0161	0.0143	0.0127	14.60	12.43	10.86	9.68	8.58	0.1482
MEDIANES	0.0213	0.0183	0.0161	0.0144	0.0127	14.48	12.46	10.89	9.74	8.55	0.1473
ECARTS-TYPES	0.0021	0.0014	0.0013	0.0011	0.0010	1.14	0.70	0.56	0.54	0.48	0.0085
SEUIL 0.05	0.0256	0.0209	0.0182	0.0161	0.0145	16.49	13.52	11.87	10.53	9.32	0.1631

SIGNIFICATION DES VALEURS PROPRES (VP),ET DES POURCENTAGES (PC) EN ANALYSE DES CORRESPONDANCES (EFFECTIFS= 5000.)

LARGEUR DU TABLEAU= 16

LONG.= 50	VP1	VP2	VP3	VP4	VP5	PC1	PC2	PC3	PC4	PC5	TRA
MOYENNES	0.0217	0.0186	0.0163	0.0145	0.0127	14.57	12.46	10.96	9.72	8.54	0.1491
MEDIANES	0.0216	0.0185	0.0161	0.0147	0.0128	14.45	12.34	10.90	9.71	8.51	0.1489
ECARTS-TYPES	0.0020	0.0015	0.0014	0.0011	0.0009	1.01	0.76	0.68	0.53	0.43	0.0076
SEUIL 0.05	0.0249	0.0208	0.0183	0.0161	0.0142	16.33	13.65	12.20	10.62	9.24	0.1614

LONG.= 60	VP1	VP2	VP3	VP4	VP5	PC1	PC2	PC3	PC4	PC5	TRA
MOYENNES	0.0249	0.0213	0.0190	0.0169	0.0151	13.91	11.89	10.58	9.44	8.40	0.1793
MEDIANES	0.0249	0.0213	0.0191	0.0168	0.0150	13.97	11.87	10.57	9.48	8.35	0.1790
ECARTS-TYPES	0.0020	0.0016	0.0013	0.0013	0.0012	0.88	0.63	0.53	0.52	0.47	0.0096
SEUIL 0.05	0.0281	0.0238	0.0209	0.0189	0.0170	15.39	12.91	11.54	10.17	9.14	0.1951

LONG.= 70	VP1	VP2	VP3	VP4	VP5	PC1	PC2	PC3	PC4	PC5	TRA
MOYENNES	0.0278	0.0243	0.0216	0.0195	0.0176	13.23	11.55	10.29	9.29	8.38	0.2102
MEDIANES	0.0275	0.0242	0.0217	0.0195	0.0176	13.04	11.50	10.32	9.31	8.43	0.2098
ECARTS-TYPES	0.0022	0.0017	0.0016	0.0015	0.0012	0.83	0.60	0.55	0.45	0.40	0.0106
SEUIL 0.05	0.0317	0.0272	0.0242	0.0218	0.0195	14.85	12.57	11.27	9.98	9.02	0.2263

LONG.= 80	VP1	VP2	VP3	VP4	VP5	PC1	PC2	PC3	PC4	PC5	TRA
MOYENNES	0.0310	0.0271	0.0243	0.0221	0.0199	12.81	11.19	10.05	9.14	8.22	0.2420
MEDIANES	0.0307	0.0269	0.0245	0.0222	0.0197	12.65	11.16	9.96	9.21	8.23	0.2419
ECARTS-TYPES	0.0027	0.0019	0.0017	0.0017	0.0013	0.87	0.54	0.44	0.43	0.35	0.0124
SEUIL 0.05	0.0357	0.0304	0.0270	0.0250	0.0223	14.28	12.07	10.97	9.81	8.76	0.2610

LONG.= 90	VP1	VP2	VP3	VP4	VP5	PC1	PC2	PC3	PC4	PC5	TRA
MOYENNES	0.0333	0.0299	0.0271	0.0245	0.0224	12.20	10.95	9.93	9.00	8.23	0.2726
MEDIANES	0.0331	0.0297	0.0267	0.0246	0.0224	12.14	10.96	9.91	8.98	8.23	0.2713
ECARTS-TYPES	0.0024	0.0019	0.0015	0.0014	0.0013	0.67	0.49	0.42	0.39	0.27	0.0122
SEUIL 0.05	0.0372	0.0327	0.0294	0.0269	0.0246	13.46	11.72	10.63	9.58	8.64	0.2948

LONG.=100	VP1	VP2	VP3	VP4	VP5	PC1	PC2	PC3	PC4	PC5	TRA
MOYENNES	0.0362	0.0324	0.0296	0.0269	0.0248	11.91	10.66	9.74	8.86	8.17	0.3038
MEDIANES	0.0360	0.0322	0.0295	0.0269	0.0249	11.86	10.66	9.70	8.84	8.19	0.3043
ECARTS-TYPES	0.0023	0.0018	0.0016	0.0017	0.0014	0.53	0.42	0.37	0.39	0.30	0.0142
SEUIL 0.05	0.0398	0.0358	0.0326	0.0298	0.0272	12.87	11.29	10.34	9.53	8.63	0.3312

SIGNIFICATION DES VALEURS PROPRES (VP),ET DES POURCENTAGES (PC) EN ANALYSE DES CORRESPONDANCES (EFFECTIFS= 5000.)

LARGEUR DU TABLEAU= 18

LONG.= 18	VP1	VP2	VP3	VP4	VP5	PC1	PC2	PC3	PC4	PC5	TRA
MOYENNES	0.0617	0.0477	0.0394	0.0328	0.0270	20.69	16.01	13.21	11.00	9.05	0.2981
MEDIANES	0.0609	0.0475	0.0391	0.0329	0.0268	20.44	15.89	13.20	11.03	9.00	0.3022
ECARTS-TYPES	0.0093	0.0058	0.0051	0.0043	0.0038	2.08	1.25	1.04	0.93	0.78	0.0312
SEUIL 0.05	0.0790	0.0565	0.0491	0.0397	0.0332	24.12	18.09	14.75	12.76	10.35	0.3465

LONG.= 20	VP1	VP2	VP3	VP4	VP5	PC1	PC2	PC3	PC4	PC5	TRA
MOYENNES	0.0677	0.0530	0.0437	0.0364	0.0301	19.99	15.64	12.90	10.74	8.87	0.3388
MEDIANES	0.0664	0.0527	0.0443	0.0362	0.0299	19.61	15.47	12.94	10.58	8.76	0.3403
ECARTS-TYPES	0.0097	0.0064	0.0048	0.0045	0.0037	2.20	1.14	0.98	0.96	0.70	0.0318
SEUIL 0.05	0.0851	0.0631	0.0504	0.0436	0.0369	24.24	17.45	14.51	12.64	10.10	0.3874

LONG.= 25	VP1	VP2	VP3	VP4	VP5	PC1	PC2	PC3	PC4	PC5	TRA
MOYENNES	0.0147	0.0119	0.0099	0.0085	0.0072	17.90	14.41	12.05	10.34	8.76	0.0823
MEDIANES	0.0145	0.0118	0.0099	0.0084	0.0072	17.88	14.37	11.97	10.30	8.70	0.0813
ECARTS-TYPES	0.0018	0.0013	0.0010	0.0009	0.0008	1.66	1.01	0.83	0.64	0.62	0.0064
SEUIL 0.05	0.0178	0.0139	0.0113	0.0100	0.0086	20.75	16.07	13.67	11.25	9.65	0.0926

LONG.= 30	VP1	VP2	VP3	VP4	VP5	PC1	PC2	PC3	PC4	PC5	TRA
MOYENNES	0.0164	0.0134	0.0113	0.0100	0.0086	16.40	13.43	11.35	9.97	8.60	0.0999
MEDIANES	0.0163	0.0133	0.0113	0.0099	0.0085	16.27	13.37	11.32	10.03	8.51	0.1007
ECARTS-TYPES	0.0016	0.0013	0.0010	0.0009	0.0008	1.36	0.97	0.80	0.69	0.58	0.0055
SEUIL 0.05	0.0192	0.0154	0.0129	0.0115	0.0100	18.89	15.05	12.74	11.07	9.61	0.1080

LONG.= 35	VP1	VP2	VP3	VP4	VP5	PC1	PC2	PC3	PC4	PC5	TRA
MOYENNES	0.0181	0.0151	0.0130	0.0113	0.0100	15.43	12.89	11.05	9.64	8.54	0.1172
MEDIANES	0.0179	0.0150	0.0128	0.0112	0.0101	15.31	12.86	10.95	9.59	8.56	0.1178
ECARTS-TYPES	0.0019	0.0014	0.0014	0.0010	0.0009	1.23	0.82	0.75	0.59	0.51	0.0083
SEUIL 0.05	0.0213	0.0176	0.0153	0.0128	0.0115	17.70	14.17	12.29	10.53	9.28	0.1303

LONG.= 40	VP1	VP2	VP3	VP4	VP5	PC1	PC2	PC3	PC4	PC5	TRA
MOYENNES	0.0195	0.0165	0.0144	0.0127	0.0113	14.49	12.27	10.73	9.47	8.38	0.1346
MEDIANES	0.0192	0.0165	0.0143	0.0127	0.0111	14.17	12.25	10.69	9.47	8.34	0.1347
ECARTS-TYPES	0.0019	0.0012	0.0011	0.0010	0.0009	1.17	0.70	0.65	0.55	0.51	0.0072
SEUIL 0.05	0.0228	0.0184	0.0163	0.0144	0.0131	16.71	13.37	11.71	10.49	9.12	0.1475

LONG.= 45	VP1	VP2	VP3	VP4	VP5	PC1	PC2	PC3	PC4	PC5	TRA
MOYENNES	0.0213	0.0181	0.0158	0.0140	0.0125	14.06	11.93	10.43	9.26	8.24	0.1515
MEDIANES	0.0210	0.0179	0.0157	0.0140	0.0124	13.96	11.72	10.42	9.21	8.21	0.1513
ECARTS-TYPES	0.0023	0.0017	0.0014	0.0011	0.0011	1.11	0.86	0.60	0.50	0.43	0.0101
SEUIL 0.05	0.0256	0.0209	0.0179	0.0160	0.0139	15.64	13.51	11.33	10.08	9.03	0.1665

LONG.= 50	VP1	VP2	VP3	VP4	VP5	PC1	PC2	PC3	PC4	PC5	TRA
MOYENNES	0.0228	0.0196	0.0174	0.0155	0.0139	13.49	11.60	10.26	9.15	8.19	0.1692
MEDIANES	0.0230	0.0195	0.0173	0.0154	0.0140	13.43	11.53	10.25	9.17	8.18	0.1706
ECARTS-TYPES	0.0017	0.0015	0.0013	0.0011	0.0011	0.76	0.67	0.56	0.43	0.41	0.0097
SEUIL 0.05	0.0257	0.0222	0.0192	0.0174	0.0154	14.74	12.82	11.30	9.95	8.85	0.1831

SIGNIFICATION DES VALEURS PROPRES (VP),ET DES POURCENTAGES (PC) EN ANALYSE DES CORRESPONDANCES (EFFECTIFS= 5000.)

LARGEUR DU TABLEAU= 18

LONG.= 50	VP1	VP2	VP3	VP4	VP5	PC1	PC2	PC3	PC4	PC5	TRA
MOYENNES	0.0228	0.0197	0.0175	0.0155	0.0136	13.51	11.70	10.37	9.21	8.09	0.1684
MEDIANES	0.0223	0.0196	0.0173	0.0154	0.0135	13.36	11.64	10.34	9.23	8.10	0.1679
ECARTS-TYPES	0.0021	0.0014	0.0012	0.0011	0.0009	1.03	0.62	0.53	0.48	0.38	0.0082
SEUIL 0.05	0.0261	0.0221	0.0196	0.0174	0.0152	15.14	12.74	11.35	9.97	8.69	0.1833

LONG.= 60	VP1	VP2	VP3	VP4	VP5	PC1	PC2	PC3	PC4	PC5	TRA
MOYENNES	0.0265	0.0228	0.0201	0.0182	0.0164	12.91	11.12	9.82	8.87	8.00	0.2049
MEDIANES	0.0262	0.0229	0.0201	0.0183	0.0163	12.85	11.04	9.83	8.88	7.99	0.2060
ECARTS-TYPES	0.0024	0.0016	0.0015	0.0012	0.0013	0.91	0.61	0.50	0.36	0.40	0.0109
SEUIL 0.05	0.0309	0.0251	0.0226	0.0202	0.0185	14.53	12.22	10.59	9.38	8.66	0.2221

LONG.= 70	VP1	VP2	VP3	VP4	VP5	PC1	PC2	PC3	PC4	PC5	TRA
MOYENNES	0.0291	0.0253	0.0227	0.0207	0.0189	12.23	10.64	9.55	8.69	7.91	0.2383
MEDIANES	0.0291	0.0251	0.0226	0.0208	0.0189	12.15	10.63	9.50	8.68	7.90	0.2382
ECARTS-TYPES	0.0023	0.0018	0.0016	0.0014	0.0014	0.85	0.53	0.47	0.40	0.35	0.0126
SEUIL 0.05	0.0326	0.0287	0.0256	0.0230	0.0213	13.64	11.69	10.26	9.44	8.46	0.2597

LONG.= 80	VP1	VP2	VP3	VP4	VP5	PC1	PC2	PC3	PC4	PC5	TRA
MOYENNES	0.0324	0.0285	0.0259	0.0234	0.0214	11.78	10.35	9.40	8.52	7.78	0.2751
MEDIANES	0.0319	0.0284	0.0258	0.0233	0.0214	11.75	10.26	9.41	8.52	7.80	0.2722
ECARTS-TYPES	0.0029	0.0021	0.0018	0.0016	0.0014	0.77	0.53	0.46	0.37	0.30	0.0144
SEUIL 0.05	0.0373	0.0318	0.0289	0.0260	0.0237	13.14	11.21	10.14	9.12	8.22	0.2999

LONG.= 90	VP1	VP2	VP3	VP4	VP5	PC1	PC2	PC3	PC4	PC5	TRA
MOYENNES	0.0351	0.0312	0.0283	0.0259	0.0239	11.32	10.05	9.11	8.34	7.70	0.3102
MEDIANES	0.0351	0.0310	0.0283	0.0258	0.0237	11.28	10.03	9.06	8.33	7.70	0.3091
ECARTS-TYPES	0.0026	0.0021	0.0018	0.0015	0.0014	0.64	0.47	0.34	0.27	0.28	0.0142
SEUIL 0.05	0.0393	0.0344	0.0316	0.0286	0.0264	12.35	10.89	9.75	8.79	8.15	0.3341

LONG.=100	VP1	VP2	VP3	VP4	VP5	PC1	PC2	PC3	PC4	PC5	TRA
MOYENNES	0.0380	0.0342	0.0312	0.0284	0.0262	11.05	9.95	9.06	8.26	7.62	0.3438
MEDIANES	0.0377	0.0341	0.0312	0.0284	0.0261	11.02	9.95	9.10	8.23	7.60	0.3411
ECARTS-TYPES	0.0028	0.0020	0.0020	0.0018	0.0015	0.55	0.41	0.39	0.36	0.28	0.0166
SEUIL 0.05	0.0425	0.0376	0.0344	0.0313	0.0286	12.18	10.48	9.59	8.81	8.08	0.3749

SIGNIFICATION DES VALEURS PROPRES (VP),ET DES POURCENTAGES (PC) EN ANALYSE DES CORRESPONDANCES (EFFECTIFS=10000.)

LARGEUR DU TABLEAU= 20

LONG.= 20	VP1	VP2	VP3	VP4	VP5	PC1	PC2	PC3	PC4	PC5	TRA
MOYENNES	0.0068	0.0054	0.0045	0.0038	0.0032	18.64	14.86	12.31	10.42	8.75	0.0364
MEDIANES	0.0068	0.0054	0.0044	0.0038	0.0031	18.35	14.70	12.25	10.37	8.75	0.0363
ECARTS-TYPES	0.0008	0.0006	0.0005	0.0004	0.0003	1.77	1.23	0.83	0.76	0.62	0.0027
SEUIL 0.05	0.0082	0.0064	0.0052	0.0043	0.0037	21.73	16.89	13.92	11.47	9.73	0.0410

LONG.= 30	VP1	VP2	VP3	VP4	VP5	PC1	PC2	PC3	PC4	PC5	TRA
MOYENNES	0.0086	0.0072	0.0061	0.0053	0.0046	15.57	12.98	11.06	9.53	8.33	0.0553
MEDIANES	0.0085	0.0071	0.0062	0.0052	0.0046	15.29	12.90	11.02	9.47	8.34	0.0554
ECARTS-TYPES	0.0010	0.0006	0.0006	0.0005	0.0004	1.37	0.91	0.71	0.64	0.51	0.0036
SEUIL 0.05	0.0103	0.0082	0.0070	0.0061	0.0053	18.34	14.31	12.23	10.64	9.07	0.0612

LONG.= 40	VP1	VP2	VP3	VP4	VP5	PC1	PC2	PC3	PC4	PC5	TRA
MOYENNES	0.0105	0.0089	0.0077	0.0069	0.0061	13.89	11.76	10.28	9.11	8.08	0.0754
MEDIANES	0.0105	0.0088	0.0077	0.0067	0.0061	13.93	11.66	10.21	9.17	8.07	0.0749
ECARTS-TYPES	0.0010	0.0008	0.0006	0.0006	0.0005	1.07	0.76	0.51	0.50	0.47	0.0045
SEUIL 0.05	0.0121	0.0102	0.0087	0.0079	0.0070	15.61	13.15	11.20	9.89	8.79	0.0830

LONG.= 50	VP1	VP2	VP3	VP4	VP5	PC1	PC2	PC3	PC4	PC5	TRA
MOYENNES	0.0118	0.0101	0.0090	0.0081	0.0073	12.70	10.89	9.65	8.70	7.81	0.0932
MEDIANES	0.0117	0.0101	0.0090	0.0081	0.0072	12.58	10.88	9.61	8.69	7.82	0.0929
ECARTS-TYPES	0.0011	0.0007	0.0007	0.0006	0.0005	0.97	0.55	0.50	0.49	0.38	0.0047
SEUIL 0.05	0.0139	0.0114	0.0102	0.0093	0.0082	14.43	11.90	10.47	9.52	8.47	0.1001

LONG.= 60	VP1	VP2	VP3	VP4	VP5	PC1	PC2	PC3	PC4	PC5	TRA
MOYENNES	0.0135	0.0117	0.0105	0.0095	0.0086	11.97	10.39	9.34	8.43	7.64	0.1127
MEDIANES	0.0134	0.0117	0.0105	0.0095	0.0086	11.90	10.33	9.30	8.45	7.63	0.1130
ECARTS-TYPES	0.0011	0.0008	0.0007	0.0007	0.0006	0.78	0.56	0.45	0.42	0.38	0.0052
SEUIL 0.05	0.0158	0.0130	0.0118	0.0105	0.0096	13.67	11.26	10.05	9.25	8.31	0.1207

LONG.= 70	VP1	VP2	VP3	VP4	VP5	PC1	PC2	PC3	PC4	PC5	TRA
MOYENNES	0.0148	0.0132	0.0120	0.0109	0.0098	11.21	9.99	9.04	8.20	7.42	0.1324
MEDIANES	0.0147	0.0131	0.0121	0.0108	0.0098	11.10	9.98	9.04	8.20	7.42	0.1320
ECARTS-TYPES	0.0011	0.0009	0.0008	0.0007	0.0006	0.66	0.50	0.40	0.33	0.35	0.0057
SEUIL 0.05	0.0166	0.0149	0.0133	0.0119	0.0109	12.27	10.91	9.60	8.79	7.98	0.1414

LONG.= 80	VP1	VP2	VP3	VP4	VP5	PC1	PC2	PC3	PC4	PC5	TRA
MOYENNES	0.0164	0.0145	0.0132	0.0121	0.0112	10.84	9.56	8.71	7.99	7.39	0.1514
MEDIANES	0.0164	0.0143	0.0132	0.0120	0.0112	10.78	9.53	8.71	8.00	7.38	0.1513
ECARTS-TYPES	0.0012	0.0010	0.0008	0.0007	0.0007	0.65	0.49	0.36	0.35	0.32	0.0061
SEUIL 0.05	0.0188	0.0163	0.0144	0.0131	0.0123	11.95	10.44	9.36	8.56	7.90	0.1599

LONG.= 90	VP1	VP2	VP3	VP4	VP5	PC1	PC2	PC3	PC4	PC5	TRA
MOYENNES	0.0179	0.0161	0.0146	0.0134	0.0124	10.50	9.40	8.57	7.85	7.25	0.1708
MEDIANES	0.0178	0.0161	0.0146	0.0134	0.0124	10.51	9.39	8.61	7.83	7.29	0.1703
ECARTS-TYPES	0.0013	0.0009	0.0008	0.0007	0.0007	0.57	0.37	0.35	0.29	0.26	0.0071
SEUIL 0.05	0.0201	0.0175	0.0158	0.0143	0.0133	11.41	9.95	9.08	8.28	7.67	0.1817

LONG.=100	VP1	VP2	VP3	VP4	VP5	PC1	PC2	PC3	PC4	PC5	TRA
MOYENNES	0.0198	0.0176	0.0161	0.0147	0.0137	10.30	9.13	8.39	7.67	7.14	0.1922
MEDIANES	0.0195	0.0175	0.0162	0.0147	0.0137	10.26	9.10	8.38	7.67	7.16	0.1923
ECARTS-TYPES	0.0015	0.0011	0.0009	0.0008	0.0007	0.60	0.40	0.32	0.28	0.23	0.0074
SEUIL 0.05	0.0225	0.0196	0.0175	0.0162	0.0148	11.45	9.87	8.85	8.16	7.53	0.2056

SIGNIFICATION DES VALEURS PROPRES (VP),ET DES POURCENTAGES (PC) EN ANALYSE DES CORRESPONDANCES (EFFECTIFS=10000.)

LARGEUR DU TABLEAU= 25

LONG.= 30	VP1	VP2	VP3	VP4	VP5	*	PC1	PC2	PC3	PC4	PC5	*	TRA	*
MOYENNES	0.0097	0.0081	0.0070	0.0062	0.0055	*	13.95	11.62	10.11	8.86	7.83	*	0.0697	*
MEDIANES	0.0096	0.0080	0.0070	0.0062	0.0055	*	13.84	11.62	10.03	8.87	7.77	*	0.0696	*
ECARTS-TYPES	0.0010	0.0006	0.0006	0.0005	0.0005	*	1.17	0.64	0.55	0.46	0.43	*	0.0042	*
SEUIL 0.05	0.0116	0.0091	0.0082	0.0069	0.0062	*	15.62	12.78	11.03	9.60	8.57	*	0.0766	*

LONG.= 40	VP1	VP2	VP3	VP4	VP5	*	PC1	PC2	PC3	PC4	PC5	*	TRA	*
MOYENNES	0.0113	0.0097	0.0086	0.0077	0.0069	*	12.11	10.39	9.23	8.25	7.39	*	0.0936	*
MEDIANES	0.0112	0.0097	0.0087	0.0077	0.0069	*	12.00	10.35	9.21	8.23	7.38	*	0.0930	*
ECARTS-TYPES	0.0010	0.0007	0.0006	0.0006	0.0005	*	0.88	0.54	0.40	0.40	0.39	*	0.0049	*
SEUIL 0.05	0.0129	0.0108	0.0095	0.0087	0.0079	*	13.62	11.35	9.85	8.89	8.08	*	0.1022	*

LONG.= 50	VP1	VP2	VP3	VP4	VP5	*	PC1	PC2	PC3	PC4	PC5	*	TRA	*
MOYENNES	0.0131	0.0114	0.0102	0.0092	0.0083	*	11.17	9.68	8.67	7.80	7.10	*	0.1174	*
MEDIANES	0.0130	0.0113	0.0101	0.0092	0.0083	*	11.00	9.67	8.65	7.76	7.10	*	0.1175	*
ECARTS-TYPES	0.0012	0.0008	0.0007	0.0007	0.0006	*	0.86	0.54	0.39	0.41	0.32	*	0.0060	*
SEUIL 0.05	0.0150	0.0128	0.0115	0.0102	0.0093	*	12.51	10.50	9.27	8.45	7.63	*	0.1273	*

LONG.= 60	VP1	VP2	VP3	VP4	VP5	*	PC1	PC2	PC3	PC4	PC5	*	TRA	*
MOYENNES	0.0148	0.0132	0.0118	0.0108	0.0099	*	10.28	9.15	8.21	7.49	6.89	*	0.1440	*
MEDIANES	0.0148	0.0131	0.0117	0.0109	0.0100	*	10.24	9.18	8.16	7.47	6.90	*	0.1442	*
ECARTS-TYPES	0.0011	0.0009	0.0007	0.0006	0.0006	*	0.63	0.44	0.36	0.29	0.30	*	0.0065	*
SEUIL 0.05	0.0166	0.0147	0.0131	0.0118	0.0109	*	11.22	9.80	8.89	8.01	7.37	*	0.1554	*

LONG.= 70	VP1	VP2	VP3	VP4	VP5	*	PC1	PC2	PC3	PC4	PC5	*	TRA	*
MOYENNES	0.0165	0.0146	0.0133	0.0121	0.0111	*	9.87	8.73	7.95	7.25	6.67	*	0.1668	*
MEDIANES	0.0165	0.0144	0.0133	0.0121	0.0111	*	9.85	8.72	7.90	7.20	6.67	*	0.1670	*
ECARTS-TYPES	0.0011	0.0010	0.0008	0.0007	0.0006	*	0.52	0.42	0.33	0.27	0.26	*	0.0073	*
SEUIL 0.05	0.0181	0.0166	0.0144	0.0133	0.0120	*	10.70	9.49	8.58	7.72	7.10	*	0.1795	*

LONG.= 80	VP1	VP2	VP3	VP4	VP5	*	PC1	PC2	PC3	PC4	PC5	*	TRA	*
MOYENNES	0.0181	0.0162	0.0147	0.0135	0.0125	*	9.42	8.42	7.64	7.05	6.53	*	0.1921	*
MEDIANES	0.0181	0.0161	0.0146	0.0135	0.0125	*	9.39	8.37	7.63	7.01	6.52	*	0.1908	*
ECARTS-TYPES	0.0011	0.0011	0.0009	0.0008	0.0006	*	0.51	0.43	0.32	0.29	0.26	*	0.0077	*
SEUIL 0.05	0.0200	0.0183	0.0161	0.0148	0.0137	*	10.45	9.14	8.19	7.57	6.97	*	0.2054	*

LONG.= 90	VP1	VP2	VP3	VP4	VP5	*	PC1	PC2	PC3	PC4	PC5	*	TRA	*
MOYENNES	0.0195	0.0175	0.0161	0.0150	0.0139	*	9.03	8.12	7.49	6.94	6.43	*	0.2156	*
MEDIANES	0.0194	0.0174	0.0160	0.0149	0.0138	*	8.99	8.10	7.49	6.91	6.40	*	0.2156	*
ECARTS-TYPES	0.0012	0.0009	0.0009	0.0008	0.0007	*	0.51	0.32	0.29	0.25	0.22	*	0.0077	*
SEUIL 0.05	0.0216	0.0192	0.0174	0.0162	0.0150	*	9.92	8.62	7.96	7.35	6.78	*	0.2312	*

LONG.=100	VP1	VP2	VP3	VP4	VP5	*	PC1	PC2	PC3	PC4	PC5	*	TRA	*
MOYENNES	0.0212	0.0191	0.0175	0.0163	0.0150	*	8.79	7.94	7.28	6.75	6.25	*	0.2408	*
MEDIANES	0.0211	0.0191	0.0175	0.0163	0.0150	*	8.82	7.87	7.26	6.73	6.25	*	0.2396	*
ECARTS-TYPES	0.0014	0.0011	0.0009	0.0008	0.0007	*	0.45	0.37	0.30	0.23	0.17	*	0.0092	*
SEUIL 0.05	0.0234	0.0208	0.0189	0.0176	0.0161	*	9.57	8.62	7.81	7.13	6.51	*	0.2566	*

SIGNIFICATION DES VALEURS PROPRES (VP),ET DES POURCENTAGES (PC) EN ANALYSE DES CORRESPONDANCES (EFFECTIFS=10000.)

LARGEUR DU TABLEAU= 30

LONG.= 30	VP1	VP2	VP3	VP4	VP5	PC1	PC2	PC3	PC4	PC5	TRA
MOYENNES	0.0107	0.0092	0.0081	0.0071	0.0064	12.59	10.85	9.54	8.43	7.55	0.0848
MEDIANES	0.0107	0.0091	0.0081	0.0071	0.0064	12.46	10.79	9.51	8.45	7.58	0.0841
ECARTS-TYPES	0.0010	0.0007	0.0006	0.0005	0.0005	0.98	0.68	0.60	0.40	0.37	0.0043
SEUIL 0.05	0.0123	0.0102	0.0092	0.0079	0.0072	14.09	11.99	10.55	9.01	8.13	0.0913

LONG.= 40	VP1	VP2	VP3	VP4	VP5	PC1	PC2	PC3	PC4	PC5	TRA
MOYENNES	0.0126	0.0111	0.0098	0.0089	0.0081	11.01	9.63	8.56	7.73	7.05	0.1148
MEDIANES	0.0126	0.0111	0.0098	0.0089	0.0081	10.90	9.65	8.56	7.71	7.03	0.1149
ECARTS-TYPES	0.0010	0.0008	0.0007	0.0006	0.0006	0.73	0.51	0.42	0.36	0.33	0.0050
SEUIL 0.05	0.0142	0.0123	0.0109	0.0099	0.0091	12.42	10.38	9.24	8.37	7.64	0.1231

LONG.= 50	VP1	VP2	VP3	VP4	VP5	PC1	PC2	PC3	PC4	PC5	TRA
MOYENNES	0.0143	0.0126	0.0114	0.0105	0.0097	9.99	8.78	7.98	7.31	6.74	0.1433
MEDIANES	0.0143	0.0126	0.0115	0.0104	0.0097	9.93	8.77	7.94	7.30	6.68	0.1434
ECARTS-TYPES	0.0011	0.0009	0.0007	0.0006	0.0006	0.66	0.44	0.33	0.31	0.29	0.0055
SEUIL 0.05	0.0160	0.0141	0.0125	0.0116	0.0105	10.96	9.46	8.56	7.85	7.20	0.1525

LONG.= 60	VP1	VP2	VP3	VP4	VP5	PC1	PC2	PC3	PC4	PC5	TRA
MOYENNES	0.0161	0.0143	0.0130	0.0119	0.0109	9.34	8.29	7.56	6.92	6.32	0.1724
MEDIANES	0.0160	0.0143	0.0130	0.0119	0.0109	9.29	8.23	7.59	6.94	6.31	0.1718
ECARTS-TYPES	0.0010	0.0007	0.0006	0.0006	0.0006	0.50	0.35	0.31	0.28	0.26	0.0058
SEUIL 0.05	0.0178	0.0155	0.0140	0.0129	0.0118	10.26	8.91	8.06	7.40	6.83	0.1815

LONG.= 70	VP1	VP2	VP3	VP4	VP5	PC1	PC2	PC3	PC4	PC5	TRA
MOYENNES	0.0177	0.0159	0.0145	0.0134	0.0124	8.76	7.86	7.18	6.63	6.13	0.2021
MEDIANES	0.0176	0.0160	0.0145	0.0134	0.0124	8.70	7.84	7.16	6.61	6.11	0.2028
ECARTS-TYPES	0.0014	0.0010	0.0008	0.0007	0.0007	0.54	0.35	0.29	0.22	0.22	0.0077
SEUIL 0.05	0.0202	0.0173	0.0158	0.0144	0.0134	9.87	8.45	7.67	7.03	6.48	0.2148

LONG.= 80	VP1	VP2	VP3	VP4	VP5	PC1	PC2	PC3	PC4	PC5	TRA
MOYENNES	0.0195	0.0175	0.0160	0.0149	0.0138	8.43	7.55	6.90	6.43	5.97	0.2312
MEDIANES	0.0193	0.0175	0.0159	0.0149	0.0138	8.35	7.51	6.90	6.41	5.96	0.2312
ECARTS-TYPES	0.0013	0.0010	0.0008	0.0008	0.0007	0.49	0.32	0.25	0.22	0.21	0.0078
SEUIL 0.05	0.0221	0.0191	0.0173	0.0160	0.0150	9.28	8.11	7.27	6.84	6.33	0.2435

LONG.= 90	VP1	VP2	VP3	VP4	VP5	PC1	PC2	PC3	PC4	PC5	TRA
MOYENNES	0.0209	0.0189	0.0174	0.0162	0.0152	8.06	7.28	6.69	6.24	5.85	0.2595
MEDIANES	0.0209	0.0189	0.0174	0.0160	0.0151	8.03	7.28	6.70	6.22	5.86	0.2588
ECARTS-TYPES	0.0013	0.0011	0.0009	0.0008	0.0007	0.47	0.34	0.29	0.23	0.22	0.0083
SEUIL 0.05	0.0236	0.0206	0.0189	0.0174	0.0164	8.88	7.75	7.15	6.63	6.15	0.2727

LONG.=100	VP1	VP2	VP3	VP4	VP5	PC1	PC2	PC3	PC4	PC5	TRA
MOYENNES	0.0228	0.0207	0.0191	0.0177	0.0166	7.83	7.11	6.57	6.09	5.70	0.2912
MEDIANES	0.0227	0.0206	0.0191	0.0177	0.0166	7.82	7.08	6.56	6.11	5.69	0.2916
ECARTS-TYPES	0.0012	0.0010	0.0009	0.0008	0.0007	0.36	0.28	0.22	0.20	0.18	0.0097
SEUIL 0.05	0.0248	0.0221	0.0205	0.0189	0.0177	8.42	7.58	6.94	6.38	6.05	0.3086

SIGNIFICATION DES VALEURS PROPRES (VP),ET DES POURCENTAGES (PC) EN ANALYSE DES CORRESPONDANCES (EFFECTIFS=10000.)

LARGEUR DU TABLEAU= 35

LONG.= 40	VP1	VP2	VP3	VP4	VP5	PC1	PC2	PC3	PC4	PC5	TRA
MOYENNES	0.0135	0.0120	0.0108	0.0098	0.0088	10.12	8.95	8.05	7.30	6.58	0.1339
MEDIANES	0.0134	0.0119	0.0107	0.0098	0.0088	10.01	8.88	8.01	7.28	6.59	0.1337
ECARTS-TYPES	0.0010	0.0008	0.0006	0.0006	0.0006	0.65	0.47	0.35	0.32	0.31	0.0059
SEUIL 0.05	0.0152	0.0133	0.0119	0.0107	0.0097	11.27	9.83	8.63	7.90	7.09	0.1427

LONG.= 50	VP1	VP2	VP3	VP4	VP5	PC1	PC2	PC3	PC4	PC5	TRA
MOYENNES	0.0155	0.0138	0.0124	0.0114	0.0105	9.24	8.22	7.40	6.78	6.24	0.1678
MEDIANES	0.0154	0.0138	0.0125	0.0114	0.0105	9.18	8.20	7.44	6.73	6.21	0.1675
ECARTS-TYPES	0.0011	0.0009	0.0008	0.0007	0.0006	0.57	0.40	0.31	0.30	0.26	0.0074
SEUIL 0.05	0.0176	0.0153	0.0135	0.0126	0.0115	10.28	8.88	7.87	7.27	6.73	0.1783

LONG.= 60	VP1	VP2	VP3	VP4	VP5	PC1	PC2	PC3	PC4	PC5	TRA
MOYENNES	0.0171	0.0153	0.0141	0.0129	0.0121	8.44	7.57	6.95	6.38	5.96	0.2022
MEDIANES	0.0170	0.0153	0.0140	0.0129	0.0120	8.32	7.62	6.94	6.44	5.94	0.2012
ECARTS-TYPES	0.0012	0.0010	0.0008	0.0007	0.0006	0.50	0.35	0.31	0.25	0.20	0.0079
SEUIL 0.05	0.0191	0.0169	0.0154	0.0142	0.0131	9.28	8.11	7.54	6.77	6.29	0.2157

LONG.= 70	VP1	VP2	VP3	VP4	VP5	PC1	PC2	PC3	PC4	PC5	TRA
MOYENNES	0.0190	0.0171	0.0157	0.0145	0.0136	7.99	7.21	6.62	6.12	5.71	0.2374
MEDIANES	0.0188	0.0170	0.0157	0.0146	0.0135	7.96	7.18	6.60	6.11	5.70	0.2369
ECARTS-TYPES	0.0012	0.0009	0.0008	0.0007	0.0007	0.43	0.30	0.25	0.23	0.19	0.0086
SEUIL 0.05	0.0212	0.0187	0.0169	0.0156	0.0147	8.73	7.72	7.03	6.49	6.04	0.2526

LONG.= 80	VP1	VP2	VP3	VP4	VP5	PC1	PC2	PC3	PC4	PC5	TRA
MOYENNES	0.0207	0.0187	0.0172	0.0160	0.0149	7.68	6.92	6.39	5.91	5.52	0.2699
MEDIANES	0.0208	0.0187	0.0171	0.0158	0.0147	7.66	6.93	6.39	5.90	5.50	0.2691
ECARTS-TYPES	0.0012	0.0010	0.0009	0.0008	0.0008	0.36	0.28	0.23	0.19	0.21	0.0089
SEUIL 0.05	0.0226	0.0206	0.0188	0.0173	0.0162	8.26	7.37	6.77	6.22	5.87	0.2839

LONG.= 90	VP1	VP2	VP3	VP4	VP5	PC1	PC2	PC3	PC4	PC5	TRA
MOYENNES	0.0222	0.0202	0.0188	0.0175	0.0164	7.25	6.63	6.15	5.73	5.38	0.3056
MEDIANES	0.0220	0.0202	0.0189	0.0175	0.0164	7.23	6.61	6.16	5.71	5.36	0.3041
ECARTS-TYPES	0.0013	0.0010	0.0010	0.0008	0.0007	0.30	0.25	0.23	0.18	0.16	0.0113
SEUIL 0.05	0.0240	0.0218	0.0205	0.0187	0.0179	7.75	7.03	6.54	6.05	5.65	0.3276

LONG.=100	VP1	VP2	VP3	VP4	VP5	PC1	PC2	PC3	PC4	PC5	TRA
MOYENNES	0.0242	0.0219	0.0204	0.0191	0.0180	7.09	6.43	5.98	5.61	5.28	0.3409
MEDIANES	0.0240	0.0219	0.0205	0.0193	0.0179	7.12	6.42	5.97	5.62	5.26	0.3423
ECARTS-TYPES	0.0014	0.0011	0.0011	0.0009	0.0009	0.35	0.24	0.20	0.17	0.18	0.0115
SEUIL 0.05	0.0266	0.0236	0.0222	0.0203	0.0192	7.64	6.79	6.28	5.87	5.61	0.3591

SIGNIFICATION DES VALEURS PROPRES (VP),ET DES POURCENTAGES (PC) EN ANALYSE DES CORRESPONDANCES (EFFECTIFS=10000.)

LARGEUR DU TABLEAU= 40

LONG.= 40	VP1	VP2	VP3	VP4	VP5	*	PC1	PC2	PC3	PC4	PC5	*	TRA
MOYENNES	0.0147	0.0130	0.0117	0.0107	0.0098	*	9.58	8.47	7.63	6.94	6.37	*	0.1537
MEDIANES	0.0147	0.0129	0.0117	0.0106	0.0097	*	9.57	8.42	7.59	6.92	6.36	*	0.1537
ECARTS-TYPES	0.0011	0.0008	0.0007	0.0006	0.0005	*	0.58	0.43	0.36	0.30	0.27	*	0.0064
SEUIL 0.05	0.0166	0.0148	0.0129	0.0117	0.0107	*	10.64	9.39	8.30	7.44	6.86	*	0.1635

LONG.= 50	VP1	VP2	VP3	VP4	VP5	*	PC1	PC2	PC3	PC4	PC5	*	TRA
MOYENNES	0.0165	0.0147	0.0134	0.0123	0.0114	*	8.60	7.65	6.99	6.41	5.93	*	0.1916
MEDIANES	0.0164	0.0146	0.0134	0.0121	0.0113	*	8.57	7.66	6.99	6.43	5.91	*	0.1912
ECARTS-TYPES	0.0012	0.0009	0.0008	0.0007	0.0007	*	0.49	0.35	0.29	0.26	0.24	*	0.0071
SEUIL 0.05	0.0182	0.0160	0.0147	0.0135	0.0125	*	9.30	8.26	7.43	6.80	6.37	*	0.2020

LONG.= 60	VP1	VP2	VP3	VP4	VP5	*	PC1	PC2	PC3	PC4	PC5	*	TRA
MOYENNES	0.0185	0.0166	0.0152	0.0140	0.0130	*	7.94	7.13	6.53	6.04	5.61	*	0.2323
MEDIANES	0.0184	0.0165	0.0151	0.0140	0.0130	*	7.89	7.10	6.52	6.00	5.59	*	0.2321
ECARTS-TYPES	0.0012	0.0008	0.0008	0.0007	0.0006	*	0.44	0.30	0.25	0.25	0.20	*	0.0074
SEUIL 0.05	0.0206	0.0180	0.0164	0.0151	0.0140	*	8.76	7.60	6.92	6.44	5.96	*	0.2414

LONG.= 70	VP1	VP2	VP3	VP4	VP5	*	PC1	PC2	PC3	PC4	PC5	*	TRA
MOYENNES	0.0203	0.0185	0.0170	0.0158	0.0147	*	7.45	6.77	6.23	5.78	5.40	*	0.2732
MEDIANES	0.0203	0.0185	0.0169	0.0157	0.0147	*	7.42	6.76	6.20	5.75	5.36	*	0.2730
ECARTS-TYPES	0.0012	0.0009	0.0010	0.0007	0.0006	*	0.39	0.26	0.28	0.21	0.17	*	0.0076
SEUIL 0.05	0.0224	0.0200	0.0184	0.0170	0.0158	*	8.07	7.29	6.66	6.10	5.71	*	0.2665

LONG.= 80	VP1	VP2	VP3	VP4	VP5	*	PC1	PC2	PC3	PC4	PC5	*	TRA
MOYENNES	0.0218	0.0199	0.0184	0.0172	0.0162	*	7.01	6.41	5.93	5.54	5.21	*	0.3110
MEDIANES	0.0216	0.0198	0.0183	0.0171	0.0161	*	7.02	6.40	5.93	5.56	5.21	*	0.3104
ECARTS-TYPES	0.0014	0.0011	0.0009	0.0008	0.0007	*	0.36	0.29	0.22	0.18	0.18	*	0.0095
SEUIL 0.05	0.0244	0.0221	0.0200	0.0185	0.0174	*	7.58	6.94	6.24	5.78	5.50	*	0.3263

LONG.= 90	VP1	VP2	VP3	VP4	VP5	*	PC1	PC2	PC3	PC4	PC5	*	TRA
MOYENNES	0.0237	0.0216	0.0201	0.0189	0.0177	*	6.75	6.16	5.74	5.39	5.04	*	0.3507
MEDIANES	0.0235	0.0216	0.0201	0.0188	0.0177	*	6.67	6.15	5.76	5.40	5.03	*	0.3510
ECARTS-TYPES	0.0015	0.0011	0.0010	0.0008	0.0007	*	0.34	0.24	0.19	0.18	0.15	*	0.0106
SEUIL 0.05	0.0257	0.0234	0.0216	0.0204	0.0188	*	7.38	6.51	6.00	5.67	5.31	*	0.3656

LONG.=100	VP1	VP2	VP3	VP4	VP5	*	PC1	PC2	PC3	PC4	PC5	*	TRA
MOYENNES	0.0255	0.0233	0.0216	0.0203	0.0191	*	6.51	5.94	5.51	5.19	4.87	*	0.3914
MEDIANES	0.0254	0.0233	0.0214	0.0202	0.0190	*	6.51	5.93	5.51	5.20	4.86	*	0.3900
ECARTS-TYPES	0.0014	0.0011	0.0010	0.0009	0.0008	*	0.28	0.20	0.18	0.16	0.14	*	0.0127
SEUIL 0.05	0.0279	0.0251	0.0233	0.0219	0.0203	*	7.03	6.26	5.79	5.41	5.16	*	0.4118

SIGNIFICATION DES VALEURS PROPRES (VP),ET DES POURCENTAGES (PC) EN ANALYSE DES CORRESPONDANCES (EFFECTIFS=10000.)

LARGEUR DU TABLEAU= 45

LONG.= 50	VP1	VP2	VP3	VP4	VP5	PC1	PC2	PC3	PC4	PC5	TRA
MOYENNES	0.0176	0.0157	0.0144	0.0134	0.0124	8.08	7.24	6.63	6.14	5.70	0.2175
MEDIANES	0.0176	0.0157	0.0144	0.0134	0.0123	8.02	7.15	6.61	6.11	5.69	0.2171
ECARTS-TYPES	0.0012	0.0009	0.0007	0.0005	0.0006	0.44	0.34	0.25	0.20	0.18	0.0074
SEUIL 0.05	0.0196	0.0172	0.0155	0.0142	0.0134	8.84	7.77	7.07	6.47	6.00	0.2296

LONG.= 60	VP1	VP2	VP3	VP4	VP5	PC1	PC2	PC3	PC4	PC5	TRA
MOYENNES	0.0196	0.0176	0.0164	0.0152	0.0141	7.45	6.70	6.22	5.77	5.36	0.2628
MEDIANES	0.0194	0.0175	0.0164	0.0151	0.0140	7.46	6.72	6.23	5.77	5.34	0.2631
ECARTS-TYPES	0.0012	0.0010	0.0008	0.0008	0.0007	0.41	0.30	0.23	0.20	0.20	0.0096
SEUIL 0.05	0.0214	0.0192	0.0178	0.0164	0.0154	8.10	7.19	6.59	6.11	5.67	0.2777

LONG.= 70	VP1	VP2	VP3	VP4	VP5	PC1	PC2	PC3	PC4	PC5	TRA
MOYENNES	0.0215	0.0194	0.0180	0.0168	0.0158	6.97	6.30	5.84	5.46	5.11	0.3080
MEDIANES	0.0215	0.0193	0.0179	0.0167	0.0158	6.91	6.26	5.83	5.43	5.10	0.3073
ECARTS-TYPES	0.0015	0.0008	0.0007	0.0006	0.0006	0.40	0.20	0.19	0.18	0.15	0.0095
SEUIL 0.05	0.0241	0.0208	0.0192	0.0178	0.0168	7.72	6.66	6.13	5.77	5.36	0.3233

LONG.= 80	VP1	VP2	VP3	VP4	VP5	PC1	PC2	PC3	PC4	PC5	TRA
MOYENNES	0.0233	0.0212	0.0197	0.0184	0.0173	6.62	6.04	5.62	5.24	4.92	0.3510
MEDIANES	0.0233	0.0211	0.0196	0.0184	0.0172	6.59	6.00	5.62	5.24	4.91	0.3508
ECARTS-TYPES	0.0014	0.0011	0.0009	0.0008	0.0008	0.31	0.22	0.18	0.16	0.15	0.0113
SEUIL 0.05	0.0252	0.0232	0.0214	0.0197	0.0185	7.15	6.44	5.92	5.49	5.15	0.3678

LONG.= 90	VP1	VP2	VP3	VP4	VP5	PC1	PC2	PC3	PC4	PC5	TRA
MOYENNES	0.0253	0.0230	0.0215	0.0201	0.0190	6.35	5.79	5.40	5.06	4.77	0.3975
MEDIANES	0.0254	0.0229	0.0214	0.0201	0.0189	6.34	5.78	5.39	5.05	4.79	0.3985
ECARTS-TYPES	0.0014	0.0011	0.0010	0.0008	0.0008	0.28	0.22	0.19	0.15	0.14	0.0120
SEUIL 0.05	0.0276	0.0252	0.0233	0.0215	0.0202	6.83	6.18	5.73	5.31	4.98	0.4181

LONG.=100	VP1	VP2	VP3	VP4	VP5	PC1	PC2	PC3	PC4	PC5	TRA
MOYENNES	0.0271	0.0246	0.0229	0.0217	0.0205	6.11	5.55	5.17	4.90	4.63	0.4433
MEDIANES	0.0270	0.0246	0.0229	0.0217	0.0205	6.06	5.57	5.16	4.89	4.65	0.4419
ECARTS-TYPES	0.0015	0.0012	0.0010	0.0009	0.0008	0.29	0.21	0.16	0.16	0.12	0.0124
SEUIL 0.05	0.0296	0.0270	0.0246	0.0231	0.0216	6.70	5.91	5.48	5.17	4.81	0.4645

SIGNIFICATION DES VALEURS PROPRES (VP),ET DES POURCENTAGES (PC) EN ANALYSE DES CORRESPONDANCES (EFFECTIFS=10000.)

LARGEUR DU TABLEAU= 50

LONG.= 50	VP1	VP2	VP3	VP4	VP5	*	PC1	PC2	PC3	PC4	PC5	*	TRA	*
MOYENNES	0.0188	0.0168	C.0154	0.0143	0.0133	*	7.73	6.92	6.37	5.89	5.48	*	0.2425	*
MEDIANES	0.0188	0.0166	C.0154	0.0142	0.0133	*	7.72	6.90	6.33	5.89	5.47	*	0.2414	*
ECARTS-TYPES	0.0013	0.0010	C.0008	0.0007	0.0007	*	0.44	0.28	0.24	0.20	0.19	*	0.0095	*
SEUIL 0.05	0.0205	0.0183	C.0167	0.0155	0.0143	*	8.40	7.44	6.76	6.23	5.76	*	0.2589	*

LONG.= 60	VP1	VP2	VP3	VP4	VP5	*	PC1	PC2	PC3	PC4	PC5	*	TRA	*
MOYENNES	0.0207	0.0187	C.0173	0.0160	0.0150	*	7.09	6.39	5.92	5.49	5.14	*	0.2920	*
MEDIANES	0.0206	0.0186	0.0172	0.0159	0.0150	*	7.05	6.38	5.90	5.47	5.11	*	0.2919	*
ECARTS-TYPES	0.0013	0.0010	C.0009	0.0008	0.0007	*	0.34	0.26	0.23	0.20	0.18	*	0.0091	*
SEUIL 0.05	0.0227	0.0204	C.0190	0.0174	0.0160	*	7.66	6.80	6.30	5.83	5.46	*	0.3057	*

LONG.= 70	VP1	VP2	VP3	VP4	VP5	*	PC1	PC2	PC3	PC4	PC5	*	TRA	*
MOYENNES	0.0225	0.0205	C.0191	0.0179	0.0168	*	6.58	5.99	5.57	5.22	4.91	*	0.3427	*
MEDIANES	0.0224	0.0206	C.0191	0.0173	0.0167	*	6.48	5.95	5.57	5.20	4.91	*	0.3421	*
ECARTS-TYPES	0.0013	0.0011	C.0009	0.0008	0.0008	*	0.35	0.24	0.19	0.17	0.15	*	0.0104	*
SEUIL 0.05	0.0254	0.0221	C.0205	0.0194	0.0182	*	7.17	6.39	5.91	5.49	5.17	*	0.3590	*

LONG.= 80	VP1	VP2	VP3	VP4	VP5	*	PC1	PC2	PC3	PC4	PC5	*	TRA	*
MOYENNES	0.0243	0.0222	C.0208	0.0196	0.0185	*	6.18	5.65	5.29	4.98	4.70	*	0.3937	*
MEDIANES	0.0241	0.0221	0.0207	0.0196	0.0184	*	6.17	5.64	5.26	4.96	4.70	*	0.3922	*
ECARTS-TYPES	0.0014	0.0010	C.0010	0.0008	0.0008	*	0.29	0.20	0.19	0.14	0.13	*	0.0119	*
SEUIL 0.05	0.0267	0.0241	C.0226	0.0208	0.0198	*	6.60	5.99	5.63	5.22	4.92	*	0.4127	*

LONG.= 90	VP1	VP2	VP3	VP4	VP5	*	PC1	PC2	PC3	PC4	PC5	*	TRA	*
MOYENNES	0.0264	0.0241	C.0226	0.0211	0.0200	*	5.98	5.46	5.10	4.78	4.53	*	0.4419	*
MEDIANES	0.0263	0.0241	C.0225	0.0209	0.0200	*	5.99	5.47	5.09	4.77	4.53	*	0.4424	*
ECARTS-TYPES	0.0013	0.0013	C.0011	0.0009	0.0008	*	0.24	0.21	0.18	0.14	0.13	*	0.0129	*
SEUIL 0.05	0.0286	0.0265	C.0245	0.0228	0.0214	*	6.40	5.80	5.38	5.02	4.73	*	0.4648	*

LONG.=100	VP1	VP2	VP3	VP4	VP5	*	PC1	PC2	PC3	PC4	PC5	*	TRA	*
MOYENNES	0.0280	0.0258	C.0242	0.0228	0.0216	*	5.68	5.23	4.89	4.62	4.38	*	0.4935	*
MEDIANES	0.0278	0.0257	0.0242	0.0229	0.0215	*	5.69	5.22	4.90	4.60	4.38	*	0.4918	*
ECARTS-TYPES	0.0015	0.0013	C.0011	0.0009	0.0009	*	0.24	0.21	0.15	0.14	0.12	*	0.0145	*
SEUIL 0.05	0.0308	0.0277	0.0259	0.0242	0.0228	*	6.05	5.57	5.10	4.86	4.55	*	0.5175	*

REFERENCES BIBLIOGRAPHIQUES DU CHAPITRE I

[1] ANDERSON T.W. (1958) : An Introduction to Multivariate Statistical Analysis - Wiley and Son - New-York.

[2] BENZECRI J.P. (1973) : Analyse des données - Tome 2 - Dunod.

[3] CHOUDARY HANUMARA R., THOMPSON W.A. (1968) : Percentage point of the extreme roots of the WISHART Matrix - Biometrika - 55 - p. 505-512.

[4] CLEMM D.S., KRISHNAIAH P.R., WAIKAR V.B. (1973) : Tables of the extreme roots of a WISHART Matrix - J. Statist. Comput. Simul. 2- p. 65-92.

[5] FISHER R.A. (1939) : "The Sampling Distribution of some Statistics obtained from non linear Equations - Ann. Eugen 7 - p. 179-188.

[6] GIRSHICK M.A. (1939) : "On the sampling Theory of roots of determinantal Equations" - Ann. Math. Stat. 10 - p. 203-224.

[7] HSU P.L. (1939) : "On the Distribution of the roots of certain determinantal Equations - Ann. Eugen. 9 - p. 250-258.

[8] KRISHNAIAH P.R., CHANG T.C. (1971) : "On the exact Distribution of the extreme roots of the WISHART and MANOVA Matrices - J. of Mult. An. 1-1. p. 108-116.

[9] KRISHNAIAH P.R., WAIKAR V.B. (1971) : Exact joint Distribution of any few ordered roots of a class of Random Matrices - J. of Multi. An. 1-3. p. 308-315.

[10] LANCASTER H.O. (1963) : "Canonical Correlation and Partition of χ^2" - Quart. J. Maths. 14 - p. 220-224.

[11] MEHTA M.L. (1960) : On the Statistical Properties of the level spacing in nuclear spectra - Nucl. Phys. 18 - p. 395-419.

[12] MEHTA M.L. (1967) : Random Matrices and the Statistical Theory of Energy Levels - Academic Press - New-York.

[13] MOOD A.M. (1951) : "On the Distribution of the Charasteristic roots of normal second moment Matrices" - Ann. Math. Stat. 22 - p. 266-273.

[14] MORINEAU A. (1975) : "Quelques générateurs de nombres pseudo-aléatoires" Méth. et Prog. Stat. - Bull. Technique du CESIA n°1.

[15] PILLAI K.C.S. (1965) : On the Distribution of the largest root of a Matrix in multivariate Analysis - Biometrika - 52 - p.405-414.

[16] PILLAI K.C.S. (1967) : Upper Percentage point of the largest root of a Matrix in multivariate Analysis - Biometrika - 54 - p. 189-194.

[17] ROY S.N. (1939) : "p - Statistics or some Generalisations on Analysis of Variance appropriate to Multivariate Problems". Sankhya 4 - p. 381-396.

[18] SUGIYAMA T. (1966) : On the Distribution of the largest latent root and the corresponding latent vector for principal component Analysis - Ann. Math. Stat. 37 - p. 995-1001.

[19] WIGNER E.P. (1967) : Random Matrices in Physics - S I A M Rev. 9 - p.1-23.

Chapitre II

INFORMATION ET VALIDITE DES RESULTATS EN ANALYSE DES DONNEES

I - Etude de quatre situations caractéristiques.

II - Mesures classiques de l'Information.

III - Valeur pratique de l'Information.

IV - Références bibliographiques.

INFORMATION ET VALIDITE DES RESULTATS EN ANALYSE DES DONNEES

Les techniques d'analyse des données consomment et produisent de l'information. Mais de quel type d'information s'agit-il, et comment évaluer la qualité des résultats ?

Le premier paragraphe nous dissuade d'utiliser les classiques taux d'inertie pour "noter" la qualité d'une représentation ; quatre situations caractéristiques ayant valeur de contre-exemples mettent en évidence l'inadaptation générale de ces quantités qui ne peuvent donner qu'une idée pessimiste de la valeur des résultats.

Le second paragraphe étudie quelle peut être la contribution de l'information de *SHANNON-WIENER* à ces problèmes d'évaluation. Les conclusions n'en sont pas totalement négatives, bien que les résultats disponibles n'aient pour l'instant que peu d'intérêt pour les praticiens.

Le troisième paragraphe, enfin, tente de lever certaines ambiguïtés du concept d'information en se restreignant cependant au domaine de l'analyse des données.

Dans ce domaine, en bref, l'activité principale consiste à découvrir ou reconnaître des formes, issues d'une opération analogue à un filtrage. Les problèmes de mesure passent alors au second plan. Il faut cependant s'assurer de la stabilité relative de ces formes, c'est-à-dire de la réalité de ce que l'on observe.

I - ETUDE DE QUATRE SITUATIONS CARACTERISTIQUES

Les quatre exemples généraux que nous allons évoquer invalident de façon assez nette les éventuelles interprétations systématiques des pourcentages d'inertie en terme de "part d'information".

Leur critique va nous permettre de préciser les relations existant entre le concept d'information et celui de variance.

Le premier exemple nous montre que, pour un même recueil d'observations, pour une même représentation finale, certaines options dans le calcul modifient de façon radicale les paramètres usuellement interprétés en terme de "pouvoir explicatif".

Le second exemple a trait aux problèmes d'étalonnages de techniques : on peut reconstituer intégralement la structure (en un sens à préciser) de tableaux particuliers à partir de sous-espaces dont la part d'inertie peut être rendue dans certaines conditions aussi petite que l'on veut.

Le troisième exemple [cf. réf 1, chap. II-B.9] montre comment la part d'inertie peut refléter les conditions mêmes de l'expérience qui a servi à la collecte des données.

Enfin, le quatrième exemple illustre l'influence du choix des variables sur l'inertie expliquée.

I.1 - *Tableau d'incidence, de contingence classique, de contingence de BURT.*

Désignons par $Z = (Z_1, Z_2)$ un tableau d'incidence à S lignes (individus, sujets, observations) et $J = J_1 + J_2$ colonnes. Chaque bloc Z_i décrit les réponses à une question mise sous forme disjonctive complète. Il est connu que l'analyse des correspondances du tableau Z donne les mêmes facteurs de norme 1 que ceux issus de l'analyse de la table de contingence : $C = Z_1'Z_2$, ainsi que ceux issus de l'analyse du tableau de contingence de BURT : $B = Z'Z$.

Rappelons brièvement cette propriété importante :

Désignons par D la matrice diagonale ayant les mêmes éléments diagonaux que B (D comporte deux blocs diagonaux D_1 et D_2, dont les éléments ne sont autres que les marges de C) :

$D = \begin{bmatrix} D_1 & 0 \\ 0 & D_2 \end{bmatrix}$, avec d'ailleurs $D_1 = Z_1'Z_1$ et $D_2 = Z_2'Z_2$

- μ, λ, β désignent des valeurs propres de même rang issues respectivement des analyses de Z, C, B.

L'équation aux valeurs propres de *l'analyse de Z* s'écrit, en décomposant $Z'Z$ en quatre blocs, et les facteurs en deux blocs :

$$(1) \qquad \frac{1}{2} D^{-1} \begin{bmatrix} Z_1'Z_1 & Z_1'Z_2 \\ Z_2'Z_1 & Z_2'Z_2 \end{bmatrix} \begin{bmatrix} \varphi_1 \\ \varphi_2 \end{bmatrix} = \mu \begin{bmatrix} \varphi_1 \\ \varphi_2 \end{bmatrix}$$

Soit :
$$\begin{cases} \frac{1}{2} D_1^{-1} C\varphi_2 = (\mu - \frac{1}{2}) \varphi_1 \\ \frac{1}{2} D_2^{-1} C'\varphi_1 = (\mu - \frac{1}{2}) \varphi_2 \end{cases}$$

Ces relations ne sont autres que les relations de transition existant entre les facteurs de *l'analyse de C*, où le coefficient classique $\sqrt{\lambda}$ vaut $2\mu - 1$.

La relation matricielle (1) est également le système des relations de transition de l'analyse de B (avec cette fois : $\sqrt{\beta} = \mu$)

Ainsi, pour des facteurs de norme 1 identiques, les analyses de Z, C, B donnent respectivement pour valeurs propres λ, $\frac{1+\sqrt{\lambda}}{2}$, $\left(\frac{1+\sqrt{\lambda}}{2}\right)^2$

Il est clair que les pourcentages d'inertie les plus forts seront ceux relatifs à l'analyse de C, les plus faibles ceux relatifs à celle de Z.

On sait de plus que, lors de l'analyse d'un tableau disjonctif Z formé de Qblocs (questions), le $q^{\text{ième}}$ bloc contenant J_q colonnes (modalités de réponse), la trace issue de l'analyse de Z vaut :

$$T(Z) = \frac{J}{Q} - 1 \quad , \quad \text{avec} \quad J = \Sigma\{J_q \mid q \in Q\}$$

Comme les valeurs propres sont nécessairement inférieures à 1, le premier facteur de l'analyse d'un tel tableau ne peut expliquer plus de $\frac{100\ Q}{J-Q}$ % de l'inertie.

Prenons l'exemple publié ailleurs [réf 7] de l'étude du tableau de contingence croisant 373 communes de la région parisienne et 29 catégories d'activité. Le premier facteur extrait explique 50% de l'inertie totale. Le résultat ci-dessus nous prouve que l'analyse du tableau disjonctif correspondant ne pouvait donner un premier facteur expliquant plus de $\frac{100 \times 2}{373 + 29 - 2} = 0{,}5\%$ de l'inertie. C'est-à-dire un pourcentage plus de cent fois plus faible, pour un même facteur de variance 1, c'est-à-dire pour une représentation sur l'axe simplement homothétique de la précédente.

Bien entendu, des considérations géométriques simples nous montrent que le calcul des taux d'inertie dans le cas disjonctif ne sont pas pertinents. Il n'en reste pas moins vrai que pour de nombreux utilisateurs, cette différence de taux d'inertie, pour un même recueil de données initial, et pour une même représentation finale, apparaît paradoxale.

I-2. Etalonnage de la méthode : cas du "cycle".

Nous parlons ailleurs de procédures d'étalonnage de techniques de visualisation à partir de tableaux ayant une structure connue : il s'agit principalement de matrices associées à des graphes particuliers.

Dans la plupart des cas [cf. réf 1], un calcul analytique exact peut être fait, sans recours à l'ordinateur. Il est alors intéressant d'étudier analytiquement les variations des représentations en fonctions des différents codages de la matrice associée. Nous étudierons ici plus particulièrement le comportement des taux d'inertie. Désignons par n le nombre de sommets du graphe. La relation de transition s'écrit ici : $\frac{1}{2} M\varphi = \varepsilon(\varphi) \cdot \sqrt{\lambda}\, \varphi$.

(M est la matrice associée au cycle : matrice symétrique n'ayant que deux éléments non nuls par ligne (égaux à 1), $\varepsilon(\varphi) = \pm 1$ selon la parité du facteur).

La relation précédente s'écrit encore : $\frac{1}{2}\left(\varphi_{j-1} + \varphi_{j+1}\right) = \varepsilon(\varphi)\sqrt{\lambda}$

$\varphi^p(j) = \cos\left(\frac{2j\ p\pi}{n}\right)$ et $\psi^p(j) = \sin\left(\frac{2j\ p\pi}{n}\right)$ sont donc les $j^{ème}$ composantes des deux facteurs associés à la valeur propre double :

$$\lambda_p = \cos^2 \frac{2p\pi}{n} \quad .$$

On obtient dans le plan des deux premiers facteurs l'équation paramétrique d'un cercle, et donc une reconstitution satisfaisante de la structure dont le tableau M représente un codage particulier.

La trace de la matrice à diagonaliser : $tr\left(\frac{1}{4} M^2\right) = \frac{n}{2}$

Le taux d'inertie du premier sous-espace propre est donc :

$$\tau = \frac{1}{n} \cos^2\left(\frac{2p\pi}{n}\right)$$

Le résultat en apparence paradoxal est le suivant : le taux d'inertie du sous espace qui "restitue" la structure initiale peut être rendu aussi petit que l'on veut, pourvu de choisir un cycle assez long. (Si $n = 10^3$, $\tau \simeq 10^{-3}$).

Particulièrement simple analytiquement dans le cas du cycle, ce résultat s'étend aux chaînes, aux réseaux à mailles carrées, etc... Cette absence de signification des taux d'inertie est due en fait au caractère "local" du codage. Si, au lieu de construire M en plaçant dans la case (i,j) 1 ou 0 selon que les sommets i et j sont contigüs ou non-contigüs, on met dans la case (i,j) le nombre de chemins de longueur k joignant sur le graphe les sommets i et j (k étant un entier supérieur à 1), on obtient un tableau de codage M, avec $M' = M^k$. Les valeurs propres de l'analyse de M' sont alors les puissances $k^{ième}$ de celles issues de l'analyse de M. Pour k suffisamment grand, le taux d'inertie du plan de visualisation peut alors avoisiner 100%.

I-3. Analyse de certaines matrices symétriques.

Le taux d'inertie et la parité des facteurs peuvent être modifiés par des variations minimes des conditions de recueil des données, dans le cas notamment des matrices de confusion [cf. réf 1, chap. II-B.9]

Considérons en effet un tableau de contingence (n,n) symétrique C, de terme général c_{ij}. Désignons par c_i le terme général d'une marge de C. A la correspondance C, on associe la correspondance D, de terme général $d_{ij} = \delta_i^j c_i$ et la correspondance H de terme général : $h_{ij} = c_i c_j$.

Considérons la correspondance artificielle symétrique décrite par le tableau $A = uC + vD + wH$ $\left\{\begin{array}{l}\text{avec } u > 0,\ v \geqslant 0,\ w \geqslant 0,\\ \qquad u + v + w = 1\end{array}\right.$

Soient λ et α deux valeurs propres de même rang issues des analyses de C et A.

A étant symétrique et ayant les mêmes marges que C, on aura, x désignant un facteur de parité $\varepsilon(x)$, tel que $\Sigma c_i x_i = 0$:

$$D^{-1} A x = \varepsilon(x)\sqrt{\alpha}\cdot x \quad [\,\varepsilon(x) = +1 \text{ ou } -1 \text{ selon que x est direct ou inverse}]$$

D'où :

$$uD^{-1}Cx + vx = \varepsilon(x)\sqrt{\alpha}\cdot x \qquad (Hx = 0)$$

Comme $u \neq 0$, $D^{-1}Cx = \left(\dfrac{\sqrt{\alpha}\,\varepsilon(x) - v}{u}\right) = x$

D'où la relation, $\eta(x)$ désignant la parité de x pour l'analyse de C :

$$\eta(x)\sqrt{\lambda} = \frac{\sqrt{\alpha}\cdot\varepsilon(x) - v}{u}$$

Soit $\varepsilon(x)\sqrt{\alpha} = u\eta(x)\sqrt{\lambda} + v$

Ainsi, la modulation des coefficients u,v,w, permet de rendre les nouvelles valeurs propres pratiquement nulles ($w \simeq 1$) ou égales ($v \simeq 1$).

Les nouveaux pourcentages d'inertie τ_i s'écrivent, en fonction des valeurs propres de C, des coefficients u et v, et de l'indicateur de parité η_i du $i^{\text{ème}}$ facteur de C :

$$\tau_i = \frac{u^2\lambda_i + v^2 + 2uv\sqrt{\lambda_i}\cdot\eta_i}{u^2\,\Sigma\lambda_i + nv^2 + 2uv\,\Sigma\sqrt{\lambda_i}\,\eta_i}$$

Si v est voisin de 0, quel que soit w, les pourcentages d'inertie sont très peu modifiés.

Si au contraire u est voisin de 0, toujours quel que soit w, ces pourcentages tendent tous à être égaux à $\frac{1}{n}$.

Il est intéressant de noter que la transformation de C en A peut correspondre à des modalités pratiques de recueil des données : dans le cas des matrices de confusion (qui ne sont pas exactement symétriques), la correspondance D correspond à une reconnaissance exacte des signaux (les réponses se concentrent sur la diagonale), alors que la correspondance H représente le cas de la confusion la plus totale : il y a indépendance entre le signal émis et le signal reconnu. La matrice symétrique A correspond à un dosage particulier de ces situations extrêmes. Un coefficient v élevé indiquera des signaux relativement faciles à distinguer. Dans tous les cas, les facteurs de norme 1 restent les mêmes ; seules les valeurs propres et les taux d'inertie sont influencés par l'organisation de l'expérience.

Comme dans l'exemple 1.1, mais dans un contexte théorique radicalement différent, à une représentation stable correspondent des valeurs propres et des taux variables.

I-4. Influence du choix des variables.

Il s'agit ici du problème classique de la complétion d'un tableau par des lignes (ou des colonnes) de "bruit blanc". Bien entendu, la part d'inertie expliquée par les facteurs décroît , alors que ceux-ci sont inchangés.

Il en est ainsi lorsque le nombre potentiel des variables est très grand (par exemple : présence d'espèces animales ou végétales rares dans des relevés écologiques). En principe, une certaine discipline dans le choix du recueil des données (critères d'homogénéité, d'exhaustivité) doit permettre d'éviter ces inconvénients. Mais d'une part le

statisticien n'a pas toujours la maîtrise de la collecte des données ni - il faut le déplorer - une connaissance suffisante du domaine d'application, et d'autre part, ces critères sont eux-mêmes trop qualitatifs et trop généraux pour définir de façon rigoureuse un tableau optimal, parmi tous les tableaux potentiels. Ici encore, les taux d'inertie seront donc beaucoup plus sensibles que les facteurs aux différentes modalités pratiques qui précèdent ou accompagnent une analyse. Certains praticiens, afin d'obtenir des taux d'inertie "honorables", n'hésitent pas à recommencer une analyse en supprimant les variables ou observations contribuant peu aux premiers facteurs ; cette manipulation opportuniste n'est pas sans fondement : il est exact que les taux d'inertie finaux représentent mieux la "part d'information" expliquée par les facteurs ; cette démarche relève, selon nous, d'une déformation qui est peut-être héritée de la théorie de la régression multiple : on juge la qualité globale d'une régression par le coefficient de corrélation multiple, dont le carré représente la part de variance expliquée par la liaison exprimée par le modèle.

Or un coefficient de corrélation multiple *ne peut qu'augmenter* si l'on ajoute des variables explicatives (exogènes), fussent-elles des nombres au hasard. Précisément dans le cas de complétion du tableau par des lignes de "bruit", le taux d'inertie des *premiers facteurs* d'une analyse en composantes principales, par exemple, ne peut, en général, que diminuer. Ces propriétés qui découlent de considérations géométriques élémentaires font que le coefficient de corrélation multiple est une mesure *optimiste* de la qualité d'une régression, alors que les taux d'inertie mesurent de façon *pessimiste* la qualité d'une représentation. Un coefficient de corrélation multiple voisin de 1 ne signifie pas forcément que la régression est de bonne qualité (une seule des variables exogènes peut être colinéaire à la variable endogène), alors qu'un taux d'inertie voisin de 100% pour un sous-espace signifie nécessairement que ce sous-espace permet une excellente représentation de l'ensemble des variables et individus analysés. A l'inverse, comme le montrent en particulier les quatre situations caractéristiques que nous venons d'examiner, des taux d'inertie faibles peuvent laisser espérer des représentations de bonne qualité. Il est cependant fâcheux que dans les deux cas, l'on parle de "part de variance expliquée" pour désigner des choses si différentes. Cette ter-

minologie n'est acceptable que dans le cas de la régression multiple, encore qu'il serait plus licite et plus prudent de parler de "part de variance résorbée". Nous reparlerons de ce problème plus bas à propos de la comparaison des descriptions et des recherches de relations.

II - MESURES CLASSIQUES DE L'INFORMATION

II-1. Taux d'inertie et divergence de JEFFREYS.

Une mesure de la distance entre deux hypothèses est fournies par la divergence de *JEFFREYS*, étroitement apparentée à l'Information de *SHANNON-WIENER* [cf. réf 4,6].

Rappelons que l'on définit, pour deux hypothèses H_1 et H_2 pouvant donner lieu à la réalisation d'un échantillon x, la divergence $J(H_1,H_2)$ comme la différence :

$$J(H_1,H_2) = \int \log \frac{P(H_1|x)}{P(H_2|x)} \, d\mu_1(x) \; - \; \int \log \frac{P(H_2|x)}{P(H_1|x)} \, d\,\mu_2(x)$$

μ_1 et μ_2 étant les mesures associées aux hypothèses H_1 et H_2 ; $P(H_i|x)$ étant la probabilité conditionnelle que H_i soit vraie connaissant x.

Dans le cas de densités continues $f_1(x)$ et $f_2(x)$, on a :

$$J(H_1,H_2) = \int \Big(f_1(x) - f_2(x)\Big) \log \frac{f_1(x)}{f_2(x)} \, dx$$

Nous appliquerons cette notion à un échantillon multidimensionnel x, censé être issu de l'un des deux schéma multinormaux suivants : (Lois normales dans $\mathbb{R}^p$) :

H_1 = Hypothèse d'indépendance $\begin{cases} \text{(Moyenne théorique = } \mu_1\text{)} \\ \text{(Matrice des covariances théorique = } \sigma^2 \text{ I)} \end{cases}$

H_2 = Cas général $\begin{cases} \text{(Moyenne théorique = } \mu_2\text{)} \\ \text{(Matrice des covariances théorique = S, supposée ici régulière)} \end{cases}$

La densité de probabilité de l'échantillon x s'écrit, pour une matrice des covariances théorique S_i, et un vecteur moyen μ_i :

$$f_i(x) = \frac{1}{|2\pi S_i|^{1/2}} \cdot \exp\left\{-\frac{1}{2}(x-\mu_i)' \, S_i^{-1}(x-\mu_i)\right\}$$

d'où :

$$\log \frac{f_1(x)}{f_2(x)} = \frac{1}{2} \log \frac{|S_2|}{|S_1|} - \frac{1}{2} \operatorname{tr} \left(S_1^{-1} (x-\mu_1)(x-\mu_1)' \right)$$

$$+ \frac{1}{2} \operatorname{tr} \left(S_2^{-1} (x-\mu_2)(x-\mu_2)' \right)$$

Le premier terme de $J(H_1,H_2)$ n'est autre que $I(1;2)$, (Information moyenne apportée par l'échantillon selon H_1, en vue de discriminer en faveur de H_1 contre H_2).

$$I(1;2) = \int f_1(x) \log \frac{f_1(x)}{f_2(x)} dx$$

$$= \frac{1}{2} \log \frac{|S_1|}{|S_2|} + \frac{1}{2} \operatorname{tr} S_1 \left(S_2^{-1} - S_1^{-1} \right) + \frac{1}{2} \operatorname{tr} S_2^{-1} (\mu_1-\mu_2)(\mu_1-\mu_2)'$$

(On a en effet posé : $x-\mu_2 = x-\mu_1 + \mu_1-\mu_2$)

d'où $J(H_1,H_2)$ $\left(= I(1;2) + I(2;1)\right)$

$$J(H_1,H_2) = \frac{1}{2} \operatorname{tr} \left(S_1-S_2\right) \left(S_2^{-1}-S_1^{-1}\right) + \frac{1}{2} \operatorname{tr} \left(S_1^{-1} + S_2^{-1}\right) (\mu_1-\mu_2)(\mu_1-\mu_2)'$$

Dans le cas où $S_1 = S_2$, $J(H_1,H_2)$ est proportionnel au D^2 de *MAHALANOBIS*, ou distance généralisée entre les populations théoriques 1 et 2. Mais le cas qui nous intéresse est celui pour lequel $\mu_1 = \mu_2$, $S_1 = \sigma^2 I$, $S_2 = S$. On notera, en abrégé :

$$J(\sigma^2 I,S) = \frac{1}{2} \operatorname{tr} (\sigma^2 I - S) \left(S^{-1} - \frac{1}{\sigma^2} I \right) = \frac{\sigma^2}{2} \operatorname{tr} \left(S + S^{-1} \right) - p$$

En faisant apparaître les valeurs propres de S :

$$J(\sigma^2 I,S) = \frac{\sigma^2}{2} \left(\sum_{i=1}^{p} \lambda_i + \sum_{i=1}^{p} \frac{1}{\lambda_i} \right) - p$$

Si les inerties totales théoriques sont égales sous les hypothèses H_1 et H_2, on a :

$$\sum_{i=1}^{p} \lambda_i = p\sigma^2$$

Le seul terme variable de $J(\sigma^2 I,S)$ est : $\sum_{i=1}^{p} \frac{1}{\lambda_i}$

On voit que la divergence entre les deux hypothèse sera grande si certaines valeurs propres de S sont voisines de 0. Une valeur propre de S infiniment petite jouera un rôle beaucoup plus discriminant que, par exemple, trois valeurs propres expliquant 80% de l'inertie totale...

Ici encore, la mesure proposée est tout à fait étrangère à nos préoccupations, car le bas du spectre ne nous intéresse pas. L'allongement préférentiel du nuage suivant certaines directions est moins important, dans ce formalisme, que l'appartenance éventuelle à une variété linéaire à $p-1$ dimensions ($\lambda_p = 0$).

Si au lieu de comparer l'hypothèse d'indépendance ($S_1 = \sigma^2 I$) au cas d'une matrice générale $S_2 = S$, définie positive quelconque, on se restreint aux matrices S^* dont on sait a priori que seule la première valeur propre σ_1^2 est supérieure aux autres (qui sont alors supposées toutes égales à $(p\sigma^2 - \sigma_1^2)/p-1$, de façon à assurer une trace $t = p\sigma^2$, identique à celle de S_1), la divergence s'écrit alors :

$$J(\sigma^2 I, S^*) = \frac{\sigma^2}{2}\left(p\sigma^2 + \frac{1}{\sigma_1^2} + \frac{(p-1)^2}{p\sigma^2 - \sigma_1^2}\right) - p$$

Prenons le cas d'une matrice des corrélations pour laquelle $p\sigma^2=1$, et posons $x = \sigma_1^2$ (qui est maintenant une part d'inertie). On a alors :

$$J(I, S^*) = \frac{1}{2p}\left(1 + \frac{1}{x} + \frac{(p-1)^2}{1-x}\right) - p$$

Cette quantité devient infiniment grande lorsque la part d'inertie x tend vers 1 (J est une fonction croissante de x dès que $x > \frac{1}{p}$)

Un résultat analogue peut être établi pour la part d'inertie d'un sous-espace de dimension supérieure à 1, en supposant toujours que les valeurs propres correspondant au sous-espace supplémentaire sont bornées inférieurement par une quantité positive. Ainsi en écartant résolument la possibilité de l'existence d'une relation linéaire entre les comportements (la variance de la "meilleure" relation linéaire est la plus petite valeur propre), taux d'inertie et divergence varient dans le même sens.

II-2. Taux d'inertie et information mutuelle (dans le cas de l'analyse des correspondances).

Rappelons brièvement quelques définitions [réf 1, T1, B5].

Si (i,j) désigne un élément aléatoire de l'ensemble produit I x J, P_{IJ} désignera sa loi, et P_I et P_J les lois marginales correspondantes.

$$P_{IJ} = \left\{ p_{ij} \mid i \in I,\ j \in I \right\} \qquad P_I = \left\{ p_{i.} \mid i \in I \right\}$$

$$P_J = \left\{ p_{.j} \mid j \in J \right\}, \left(p_{i.} = \sum_j p_{ij} \mid i \in I \right), \left(p_{.j} = \sum_i p_{ij} \mid j \in J \right)$$

On appelle degré d'indétermination, ou information sur I (respectivement sur J) la quantité $H(P_I) = -\left\{ \Sigma\ p_{i.} \log_2 p_{i.} \mid i \in I \right\}$

$$\left(\text{respectivement } H(P_J) = -\left\{ \Sigma\ p_{.j} \log_2 p_{.j} \mid j \in J \right\} \right)$$

L'information mutuelle entre I et J sera la quantité positive [cf. réf 1]

$$H\ (P_{IJ}\ ;\ P_I P_J) = H(P_I) + H(P_J) - H(P_{IJ})$$

(Ce n'est autre que l'information $I(P_I P_J\ ;\ P_{IJ})$ de *SHANNON WIENER*)

Cette quantité s'écrit :

$$H\ (P_{IJ},\ P_I P_J) = \sum_i \sum_j p_{ij} \operatorname{Log}_2 \frac{p_{ij}}{p_{i.} p_{.j}}$$

Au voisinage de l'hypothèse d'indépendance, on a l'approximation :

$$H\ (P_{IJ},\ P_I P_J) \approx \frac{1}{2 \operatorname{Log}_2} \left\{ \sum_i \sum_j \frac{(p_{ij} - p_{i.} p_{.j})^2}{p_{i.} p_{.j}} \right\}$$

Ainsi, la somme des valeurs propres, non triviale en analyse des correspondances, représente une approximation de l'information mutuelle entre les ligneset les colonnes du tableau, à condition toutefois de ne pas trop s'éloigner de l'hypothèse d'indépendance.

En principe, les probabilités théoriques P_{IJ}, P_I, P_J étant inconnues, on calcule seulement une estimation de cette information, variable aléatoire dont la loi asymptotique est connue.

Les résultats du chapitre 1, §-2, nous montrent, dans le cas de l'hypothèse d'indépendance, les taux d'inertie sont des variables aléatoires indépendantes de l'estimation (approchée) de l'information mutuelle. Cette propriété, d'une portée pratique limitée car l'on est rarement au voisinage de l'hypothèse d'indépendance, illustre elle aussi les difficultés d'application de la notion d'information mutuelle dès qu'il s'agit d'évaluer les résultats d'une analyse.

La part d'inertie d'un sous-espace peut être significative, sans que l'on puisse considérer que l'information mutuelle globale le soit.

III - VALEUR PRATIQUE DE L'INFORMATION

Les quatre situations pratiques examinées au paragraphe 1 doivent, selon nous, par leur valeur de contre-exemple, décourager l'utilisation des pourcentages d'inertie comme unique critère d'évaluation des résultats des analyses factorielles, qu'il s'agisse d'analyse en composantes principales ou d'analyse des correspondances. Les pourcentages d'inertie ne peuvent que donner une vue pessimiste, une borne inférieure de l'information transmise. Il est également certain que le concept d'information que nous sommes contraints d'utiliser recouvre lui-même des notions assez diverses.

III-1. *Information et Forme.*

La plupart des mathématiciens, physiciens ou statisticiens qui ont élaboré ou simplement étudié la théorie de l'information (fondée principalement par *Cl. SHANNON*) ont été conscients des limites de cette théorie qui ne s'intéresse qu'à la conservation, la transmission, la transformation de l'information, sans aborder le problème (évidemment complexe) de la valeur pratique de l'information. *L. BRILLOUIN* [réf 2], lors d'un survol des problèmes "dépassant la théorie actuelle" rappelle à propos des machines à calculer, que "le mécanisme de calcul n'ajoute aucune information nouvelle, il ne fait que la reproduire dans un langage différent et probablement avec des pertes. Mais le calcul, comme le décodage ou la traduction, ajoute certainement à la valeur pratique de l'information et la rend pleine de signification pour l'utilisateur - ... tout critère relatif à la valeur aura pour corollaire une évaluation de l'information reçue. Ceci revient à sélectionner l'information suivant une certaine loi de mérite. Certaines fractions de l'information pourront être considérées comme sans valeur et seront écartées. Ce problème présente une grande ressemblance avec celui du filtrage. Un filtre est un élément bien connu de la théorie générale et est caractérisé par deux importantes propriétés : il est irréversible et il diminue la quantité totale d'information. Par analogie, nous pouvons énoncer le résultat suivant : la valeur relative de l'informa-

tion pour un utilisateur donné est au plus égale à l'information absolue. Nous avons là un point de départ dont la validité est générale et le problème se pose de définir les différentes méthodes de filtrage que l'on peut utiliser pour évaluer la valeur relative".

L'analogie avec le filtre est très parlante pour nous ; il est assez tentant de considérer un programme d'analyse des données comme un filtre très particulier qui permettrait d'isoler un certain signal (par exemple un réseau d'interrelations que l'on pourrait difficilement imputer au hasard) d'un bruit de fond. Cependant l'analogie ne peut aller très loin, car le signal, dans notre cas, n'a que peu de chose à voir avec un message, et nous ne sommes pas dans la situation d'un processus de communication. Nous devons, pour espérer aller plus loin dans l'étude de l'information en analyse des données, tenir compte de la dérive du concept, en étudiant notamment, d'après *R. THOM* [réf 11, 12] les glissements sémantiques du mot "Information". Nous ne ferons en fait que résumer brièvement l'argumentation de cet auteur qui explique l'instabilité sémantique de ce mot par la complexité des relations que recouvre le concept : "il y a - sous-jacent à l'idée d'information - l'existence d'un demandeur à qui apprendre cette information est profitable ; et d'un donneur qui donne, en général de son plein gré, l'information au demandeur. Dans tous les emplois du mot où l'on est dans l'incapacité d'identifier ces quatre éléments : demandeur-demande-donneur - avantage apporté au demandeur par la connaissance de l'information, on doit suspecter dans l'emploi du mot une certaine malhonnêteté...". On peut alors distinguer plusieurs sens selon les mutilations que subit le schéma ; l'information aux sens judiciaire et journalistique, où il y a effacement du demandeur, et même parfois de la demande ; information au sens des techniques de publicité, où "l'abus est cette fois flagrant : l'information est imposée au destinataire qui ne l'a pas demandée, et son contenu est notoirement plus avantageux au donneur qu'au destinataire" ; l'information au sens de la théorie de *SHANNON* où pratiquement tout est effacé à l'exception du canal de transmission entre le demandeur et le donneur. En biologie, des expressions telles que "information génétique" permettent de masquer une partie de l'ignorance de certains mécanismes "... tout en apportant par la connotation d'intentionnalité du mot information une caution implicite au

finalisme qui sous-tend toute pensée biologique. En ce sens, l'information, c'est la forme obscure de la causalité".

Nous nous rapprochons de nos préoccupations avec le sens du mot information dans la phrase suivante : "l'information fournie par les rayons X après traversée du corps du patient", pour laquelle, selon R. THOM, "... il n'y a ambiguïté ni sur le donneur (le corps du patient) ni sur le demandeur-récepteur (l'observateur) ni sur la finalité globale du processus (interprétation diagnostique). La seule ambiguïté concerne le contenu signifié de mot information que... je réduirais à la forme du message reçu sur la plaque sensible, et interprété par l'observateur". Ceci fait dire à l'auteur que l'on gagnerait beaucoup à remplacer dans certains cas, le mot information par le mot *forme*. L'analogie entre un appareil radiographique et les instruments d'observation que constituent les techniques d'analyse factorielle est grande. Nous l'avons développée ailleurs en comparant notamment les deux obstacles à l'observation directe constitués par l'opacité des tissus dans un cas, le caractère multidimensionnel des données dans l'autre. Ce caractère multidimensionnel ne fait qu'ajouter à l'inadaptation de la notion classique d'information comme le souligne plus loin R. THOM :

> "En réalité, le rêve de la théorie de l'information a été d'établir un algorithme permettant d'évaluer la complexité d'un système, en même temps que son degré d'organisation, l'écart qu'il présente par rapport à une structure *totalement désordonnée*. Un tel algorithme n'existe que pour les morphologies de dimension un : les suites finies de lettres extraites d'un alphabet fini. Pour les morphologies naturelles - pluridimensionnelles - un tel algorithme n'existe pas et les concepts mêmes de complexité, d'ordre, d'organisation, de désordre n'y sont pas définis".

Il est maintenant bien connu des statisticiens que les généralisations multidimensionnelles ne se font pas à n'importe quel prix : le problème de la séparation de populations qui se pose souvent à propos de reconnaissance des formes, est relativement simple dans le cas unidimensionnel (un point sépare deux demi-droites). Dans le

cas d'échantillons multidimensionnels, on est obligé de spécifier la forme des cloisons (hyperplans, hypersurfaces de degré donné, etc...) et par conséquent d'introduire des hypothèses supplémentaires qu'un traitement combinatoire ou en apparence non-paramétrique ne doit pas camoufler. Cependant, s'il n'existe pas d'algorithme universel d'évaluation de la complexité ou de l'organisation d'un donné multidimensionnel, il existe de nombreux algorithmes particuliers - comme ceux que nous utilisons en analyse des données - qui permettent de mettre en évidence certains aspects de cette organisation : un nuage est-il approximativement hypersphérique ? Est-il concentré sur une sous-variété particulière ? Est-il formé de deux sous-nuages que l'on peut considérer comme distincts ? Existe-t-il des axes principaux d'inertie correspondant à des moments prédominants ? Ces questions très partielles ne peuvent être posées qu'après avoir donné de l'objet observé une première représentation sous forme de nuage dans un espace euclidien de grande dimension, étape préliminaire qui suppose déjà une approximation, ou peut-être dans certains cas, toute une théorie.

Le statisticien qui dépouille un fichier d'enquête socio-économique, même aidé d'efficaces instruments de visualisation, est en fait souvent submergé par l'information pourtant partielle qu'il en extrait (ou par les formes qu'il reconnait ou qu'il découvre, si l'on préfère cette terminologie). Il a surtout besoin, c'est du moins notre propos et notre ambition de l'aider dans cette tâche, de mieux comprendre et d'évaluer le peu qu'il voit.

Nous retiendrons surtout de ces considérations sur la notion d'information qu'il sera parfois utile, en analyse des données, de parler de forme plutôt que d'information - (ce qui nous conduira à vérifier la *stabilité d'une forme* et non pas à mesurer une *quantité d'information*, pour valider une représentation) ; ces techniques fonctionnent un peu comme des filtres (très généraux) : elles tentent d'accroître la valeur pratique de l'information, au prix d'une perte d'information brute qui peut être considérable. Dans cette optique, il apparaît peu pertinent de considérer la quantité d'information brute initiale comme une mesure de référence.

En statistique, la théorie de l'information n'est pas totalement inopérante lorsqu'il s'agit d'étudier certains types de modèles multidimensionnels, comme le suggère le paragraphe 2-1. Les résultats disponibles sont cependant assez minces, et ne peuvent être développés semble-t-il que dans des cadres trop restrictifs.

III-2. Description et relation.

Nous voudrions insister ici sur certaines particularités des analyses factorielles qui les distingue fondamentalement des techniques de régression et des méthodes apparentées (analyses canoniques, analyse discriminantes) (1).

On sait que si l'on cherche la meilleure relation linéaire liant p variables, pour lesquelles nous disposons de n observations, la solution (l'hyperplan de régression orthogonale) est fournie par le vecteur de la matrice des covariances expérimentale de ces p variables correspondant à la plus petite valeur propre. La régression linéaire classique est un cas limite de régression orthogonale, dans une métrique - limite qui accorderait un poids infiniment grand à une des variables [cf. par exemple réf 8]. Inversement, on peut dire que si l'on cherche la relation linéaire la plus mauvaise... c'est-à-dire la plus dispersée sur l'ensemble des observations, on obtient la première composante principale, qui est le vecteur propre de la matrice des covariances correspondant à la plus grande valeur propre.

Les coefficients de cette première composante sont les cosinus directeurs de l'axe principal d'inertie du nuage des n observations dans $\mathbb{R}^p$. Il nous fournissent également la meilleure description à une dimension des proximités entre variables (préciser plus le sens de "meilleure" et de "proximité" exigerait un exposé théorique pour lequel nous renvoyons aux manuels existants).

---

(1) Le fait que l'analyse des correspondances des tableaux de contingence puisse être considérée comme un cas particulier d'analyse discriminante (et donc d'analyse canonique) appliquée à des codages spéciaux n'est pas en contradiction avec la différence fondamentale dont nous allons parler, précisément parce que ces codages spéciaux (disjonctifs) induisent des propriétés très particulières.

Le bas du spectre nous fournit donc des *relations* linéaires permettant éventuellement de *prévoir*, pour un individu, la valeur d'une variable connaissant les valeurs de toutes les autres. Nous allons voir que ces relations font intervenir des coefficients qui sont *instables* et *vulnérables*, et d'autant plus que les variables sont nombreuses.

Au contraire, le haut du spectre nous fournit des combinaisons linéaires qui nous permettent de *discriminer* les individus et qui *décrivent* les associations entre variables. Les coefficients en sont généralement *stables* et le sont d'autant plus que les variables sont nombreuses.

Stabilité le long du spectre.

Supposons que l'on ait réalisé l'analyse factorielle d'un tableau X de valeurs numériques quelconques, donnant lieu à une séquence de valeurs propres distinctes.

Les premiers facteurs de l'analyse seront beaucoup plus stables, vis-à-vis de petites perturbations apportées au tableau X, que des facteurs correspondant aux plus petites valeurs propres. Ce résultat bien connu de la théorie de la perturbation [cf. réf 5, 13] n'était pas ignoré non plus des statisticiens.

Nous pouvons citer le théorème établi par *J.C. DEVILLE*, dont la démonstration est facilement accessible [cf. réf 3], et qui considère de façon plus générale un opérateur compact hermitien positif C d'un espace de *HILBERT E.* (de tels opérateurs sont classiquement munis de la norme : $\|C\| = \sup_{\|x\| = 1} \|Cx\|$

les k premières valeurs propres de C sont supposées simples. A la valeur propre λ_i correspond le vecteur propre unitaire x_i .

On note : $\Delta = \inf_{i = 1,\dots k} (\lambda_i - \lambda_{i+1})$

Le résultat est le suivant : si C' désigne un opérateur compact hermitien positif de E, tel que $\|C-C'\| \leqslant \varepsilon < \Delta/2$, on a

alors $|\lambda_i - \lambda'_i| < \varepsilon/2$ pour $i \leqslant k$; de plus, à chaque λ'_i (valeur propre de C') on peut associer un vecteur propre unitaire x'_i de C' tel que :

$$\|x_i - x'_i\| \leqslant 2\sqrt{2}\,\varepsilon/\Delta$$

Ainsi, la perturbation subie par le vecteur propre sera évidemment d'autant plus importante que la matrice d'inertie sera modifiée (coefficient ε), mais sera surtout sensible à l'écart Δ entre deux valeurs propres consécutives qui devra être le plus grand possible. L'écart Δ ne peut être grand que pour le haut du spectre, dont le résultat ci-dessus montre qu'il correspondra en général à des sous-espaces stables.

Les contre-exemples ne manquent pas pour mettre en évidence l'instabilité des vecteurs propres correspondant à des valeurs propres dont l'écart absolu est faible, et donc l'instabilité des coefficients intervenant dans une relation (valeurs propres très faibles) approximativement vérifiée par les variables.

En fait la recherche de relations n'est justifiée que si un problème de prévision se pose à propos d'un petit nombre de variables. Qu'il s'agisse de régression linéaire ou de régression orthogonale, les variables sont *mises en concurrence*, ce qui hypothèque lourdement l'interprétation des résultats, sensibles aux colinéarités et aux substitutions.

Puisque l'analogie entre "analyse factorielle" et "filtre" a été évoquée au paragraphe précédent, on peut déjà préciser comment ce filtre agit : il isole la partie la plus stable du réseau d'interrelations existant entre les lignes et les colonnes d'un tableau de valeurs numériques. Bien entendu, la définition des interrelations dépend de la nature du tableau et de la technique particulière d'analyse factorielle employée (similitudes entre profils pour l'analyse des correspondances, liaisons linéaires entre variables pour l'analyse en composantes principales) ; la définition de la stabilité dépend de la classe de perturbations que l'utilisateur entend as-

socier à son recueil de données.

III-3. Stabilité des formes.

Les calculs par simulations peuvent être utilisés, dans les épreuves de validation en analyse des données, de deux façons radicalement différentes : pour valider une hypothèse relative à un paramètre, ou pour vérifier la stabilité d'un résultat [cf. réf 1, chap. B.9].

III-3.1. Validation d'une hypothèse.

Dans ce cas, les simulations ne sont qu'un substitut de techniques statistiques classiques ; celles-ci ne peuvent pas toujours s'appliquer, puisque les hypothèses donnant lieu à des formulations analytiques et à des tabulations ne forment qu'un éventail restreint. On simule alors la loi du (ou des) paramètres en respectant plusieurs pseudo-réalisations de ce paramètre, que l'on compare à la valeur réellement observée. Un programme-test permet alors de remédier à l'absence de table et d'estimer, de façon assez grossière en général, des seuils de signification.

Les simulations de ce type sont très coûteuses (puisqu'elles exigent de répéter des calculs qui sont en général complexes : par exemple diagonalisation de matrice d'ordre élevé). Les résultats, du point de vue de l'analyse des données, sont souvent assez minces, et sont parfois difficiles à interpréter de façon rigoureuse. Le cas le plus fréquent est sans doute celui-ci : l'utilisateur cherche le nombre de facteurs à retenir lors d'une analyse factorielle.

S'il ne s'agit pas d'analyse de correspondances sur vraies tables de contingence, pour laquelle des tables approchées existent [cf. chapitre 1], il va procéder à une vingtaine de simulations pour atteindre l'équivalent d'un seuil de 5% (qui n'est évidemment pas comparable à un seuil exact comme celui que fourniraient des tables, du point de vue des risques encourus). La détermination simultanée de plusieurs seuils pose quelques problèmes : la loi jointe des va-

leurs propres et des pourcentages d'inertie n'est évidemment pas connue lorsque l'hypothèse d'indépendance des lignes et des colonnes du tableau n'est pas vérifiée [cf. annexe 1]

Ces procédures en général coûteuses ne permettent donc pas de formuler des conclusions de façon rigoureuse. Elles sont cependant le seul recours lorsqu'il s'agit de juger la signification d'un paramètre issu de calculs complexes.

III-3.2. Stabilité d'un résultat.

Il s'agit ici d'une application beaucoup plus pertinente et beaucoup moins coûteuse des procédures de simulation. Elle consiste simplement en une vérification de la stabilité des configurations obtenues à la suite de modifications du tableau initial. Une seule simulation peut suffire dans certains cas.

Avant d'examiner les différents types de perturbations que subiront les données soumises à l'analyse, il convient de rappeler quels sont les différents facteurs qui peuvent conditionner la qualité des résultats d'une analyse factorielle.

Nous en distinguerons quatre :

- les erreurs de mesure,
- le choix et le poids des variables,
- le codage des variables
- les aléas d'échantillonnage (ex : le choix des individus ou observations),

chacune de ces quatre sources de perturbation initiale donnera lieu à des modifications du tableau de données, qui ne devront pas affecter les configurations supposées stables.

La nature et l'amplitude des modifications, l'appréciation de la stabilité des résultats se font lors de "colloques singuliers" réunissant l'utilisateur et le statisticien.

a) *Les erreurs de mesure* : l'ordre de grandeur de ces erreurs, ainsi que leur distribution approximative dans la population doivent être spécifiées par l'utilisateur en fonction de sa propre connaissance du domaine d'étude. Dans le cas classique des réponses ordonnées du type : "pas du tout d'accord, pas très d'accord, assez d'accord, tout à fait d'accord", on peut supposer par exemple que l'individu enquêté a une chance sur deux d'avoir exprimé exactement ce qu'il ressentait, une chance sur quatre (sauf aux extrêmités) de répondre en fait une modalité immédiatement contigüe. Les programmes permettront en général de simuler une grande variété de situations introduisibles analytiquement.

b) *Le choix et le poids des variables :* ce problème se pose lorsque le statisticien a la possibilité d'échantillonner dans l'espace des variables, ce qui n'est pas toujours le cas. Les classiques critères d'homogénéité et d'exhaustivité ne fournissent qu'un cadre général. On pourra donc effectuer des "ponctions aléatoires" dans l'ensemble des variables, afin d'éprouver la sensibilité des résultats vis-à-vis de la composition de cet ensemble. Ce procédé a été utilisé pour vérifier la stabilité de la typologie des départements français décrits par environ 180 variables (1). Le fait de supprimer 50 d'entre elles ne modifiait pratiquement pas la typologie plane des départements répartis d'abord suivant un axe d'urbanisation, puis un axe de latitude. Le problème du poids des variables se pose surtout en analyse en composantes principales (ou en analyse des correspondances s'il s'agit de tableaux de notes ou de mesures, et non de comptages). On peut par exemple, pour éprouver une éventuelle invariance, imposer à chaque variable une variance comprise entre 1 et 2, en procédant par exemple à un tirage de série pseudo-aléatoire uniforme entre ces deux valeurs, et diagonaliser la matrice des covariances ainsi construite.

---

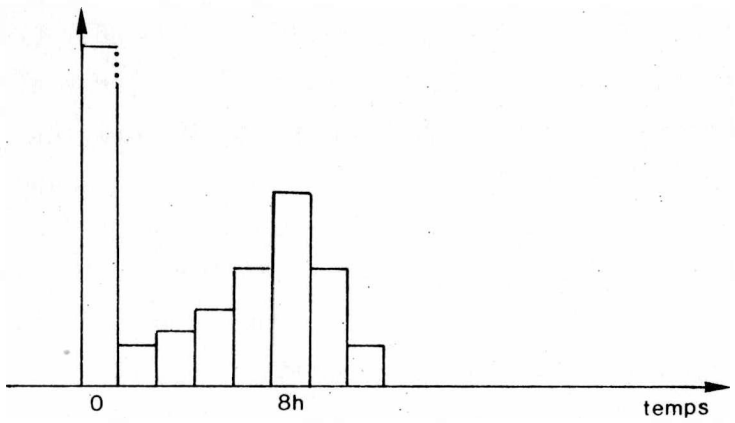
(1) L. LEBART, F. TONNELLIER, S. SANDIER - Aspects géographiques du système des soins médicaux - Consommation n°4 - 1974.

c) *Le codage des variables :* nous désignons ici par codage la transformation préliminaire et contrôlée des données brutes qui précède l'analyse multidimensionnelle. Il s'agit selon nous d'une opération qui est fondamentalement empirique (ce qui n'implique évidemment pas que l'école philosophique de l'utilisateur soit l'empirisme !) parce que liée indissolublement au contenu signifié de l'information (terminologie de *R. THOM*). Le codage a, comme l'analyse des données, pour raison d'être l'augmentation de la valeur pratique de l'information. Il ne s'agit pas dans la phase de codage de rendre celle-ci plus facilement assimilable à l'homme, mais plus aisément utilisable par l'algorithme, qu'il s'agisse d'analyse factorielle ou de classification.

Prenons un exemple pratique qui met en évidence ces différents aspects du codage. Cet exemple sera emprunté à l'enquête CNAF-CREDOC de 1971 sur les "Besoins et aspirations des familles et des jeunes", dont un premier compte-rendu a été publié par *N. TABARD* (1). Une grille socio-administrative a été construite, à partir de variables telles que "profession", "revenus", "nombre d'enfants", "âge", etc... Le mode de construction et d'utilisation de cette grille est précisé en annexe 2. Le problème de codage que nous allons examiner concerne une variable supplémentaire (ou illustrative). Il se poserait de la même façon pour une variable participant effectivement à l'analyse. Cette variable est "le temps de travail à l'extérieur" de la mère de famille, mesuré pour 2003 personnes. L'histogramme de cette variable est fortement bimodal puisque, dans cet échantillon particulier, près de la moitié des femmesne travaille pas à l'extérieur.

---

(1) N. TABARD - "Besoins et aspirations des familles et des jeunes" - Coll. Etudes C.A.F. n°16 - 1974 - 514 pages.

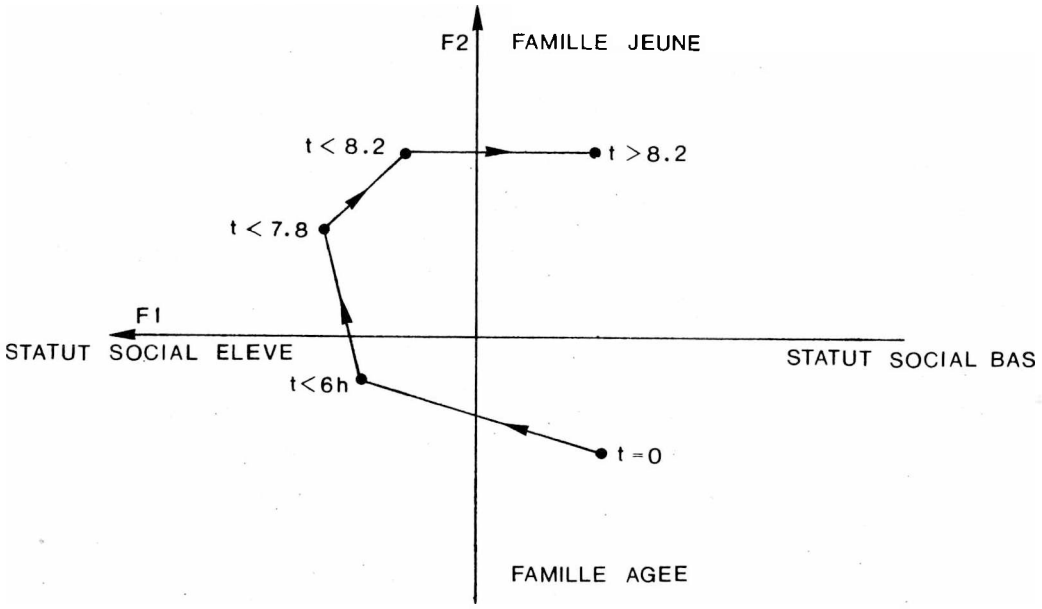


Esquisse de l'histogramme

Le problème du codage se résume ici à un choix de partitions, le nombre de classes et les limites de classes restant à fixer. Il est clair que la variable numérique "temps de travail" est inutilisable telle quelle en analyse factorielle. Ces calculs ne faisant intervenir que les moments d'ordre 2 sont surtout intéressants pour les variables ayant des distributions marginales unimodales et relativement symétriques, à défaut de variables normales ; ces calculs privilégient également les relations linéaires entre variables, ce qui est une contrainte difficilement acceptable.

Il paraît a priori souhaitable d'isoler à l'intérieur d'une même classe les femmes au foyer (t = 0), d'isoler également les femmes ayant un temps de travail exceptionnellement long. Aucun algorithme aveugle, à notre connaissance, ne peut produire une partition qui satisfasse ces deux exigences à partir du seul histogramme. Un algorithme tel que celui de *W.D. FISHER (On grouping for maximum homogeneity J.A.S.A. n°53, 1958)* qui fournit pourtant des partitions en k classes optimales exactes ne pourra même pas, si l'on demande moins de cinq classes, isoler exactement dans une classe les femmes au foyer.

Nous avons, en pratique, construit cinq classes à partir d'un histogramme détaillé, et obtenu le résultat suivant, dans le plan des deux premiers facteurs représenté plus en détail en Annexe 2.



Trajectoire des temps de travail
dans le plan F1,F2

La liaison avec le premier facteur est non linéaire, et n'aurait pas été mise en évidence si le temps de travail avait été positionné comme une variable numérique, sans division en classes, ou même si l'on s'était contenté de deux classes. C'est donc au prix d'une perte d'information brute (division en classes) avant toute analyse, et à partir de considérations d'ordre sémantique sur les limites de classe que l'algorithme fournit les indications les plus fines sur le phénomène étudié.

Ainsi le codage n'est donc pas une simple formalité avant l'analyse : il suppose une connaissance simultanée des données (en général à partir de statistiques descriptives élémentaires) de la méthode et de ses exigences. Le codage peut surtout être considéré comme source de perturbation éventuelle des résultats dans le cas des notes, des échelles ou des classements (en analyse des rangs ou des préférences, par exemple). Il est alors utile de vérifier que les configurations obtenues résistent à des changements de variables monotones très déformants (logarithme exponentiel, etc...) afin de s'assurer que l'ordre des notes est plus important que les propriétés métriques particulières de l'échelle utilisée.

Enfin, dans tous les cas, il est intéressant de mettre en évidence un "codage minimal" qui est la forme de codage la plus frustre susceptible de conserver les configurations observées. Citons-en deux exemples : une analyse factorielle fut réalisée sur un tableau de dépenses individuelles de consommation (d'après l'enquête CREDOC-UNCAF de 1963) et produisit une certaine typologie des postes de consommation ; cette analyse fut refaite (1) en codant simplement "1" les dépenses strictement positives, quels que soient leurs montants et donna alors une typologie des postes très voisine de la précédente. On conçoit que l'interprétation de la première analyse soit modifiée par ce dernier résultat, qui souligne l'importance de l'accès à certains types de consommation, indépendamment de l'intensité de ces consommations.

Un résultat analogue a été établi à propos d'une typologie des activités réalisée à partir de budget-temps, qui n'a pas été bouleversée lorsque les durées positives des activités ont été remplacées par des "1", les durées nulles étant toujours codées par des "0".

La simple mention d'une activité (lecture, promenade, soins aux enfants, ...) jouait donc un rôle déterminant.

d) *Les fluctuations d'échantillonnage :* deux types de calculs de stabilité peuvent être exécutés comme dans le cas *b)* ci-dessus : des modificaitons des pondérations des individus ou observations, des ponctions ou des fractionnements de l'échantillon. Les deux opérations donnent une idée de la stabilité des résultats. Bien entendu, un échantillon où certains aspects de la population parente ne sont pas du tout représentés ne pourra fournir des résultats extrapolables, même si les configurations obtenues sont stables. Toutefois, les typologies obtenues par analyse factorielle n'exigent pas une représentativité de l'échantillon aussi stricte que les estimations de pourcentages ou de moments d'ordre 1. Cette relative stabilité est surtout un fait d'expérience.

---

(1) B. JOUSSELLIN - Les choix de consommation et les budgets des ménages - Consommation n°1 - 1972.

On peut néanmoins en donner une explication très partielle en rappelant que les sous-espaces correspondant au haut du spectre [cf. paragraphe III-2.] sont les plus stables vis-à-vis des éventuelles perturbations de la matrice à diagonaliser ; et que cette matrice elle-même (par exemple matrice des corrélations expérimentales en analyse en composantes principales) est moins sensible que les moments d'ordre 1 (moyennes, pourcentages) aux fluctuations d'échantillonnages.

La grille socio-administrative de l'annexe 2 est à cet égard assez démonstrative. Le plan de sondage particulier de cette enquête destinée à étudier des couches de la population très spécifiques prévoyait notamment une sur-représentation des mères de famille exerçant une activité extérieure - a chaque famille est ainsi affecté un coefficient de redressement (les disparités de ces coefficients sont considérables). La typologie des variables socio-économiques de l'annexe 2 fait preuve d'une stabilité étonnante, puisqu'elle est la même que l'échantillon soit redressé ou que l'analyse soit faite directement sur les données brutes. De même qu'il était utile de rechercher un codage "minimal" susceptible de reconstituer les configurations, on peut s'intéresser à la taille minimale de l'échantillon qui conserve la typologie des variables. Ceci permet de hiérarchiser les facteurs à partir de leur "assise" dans l'échantillon (i.e. : l'analyse de 200 individus pris au hasard parmi les 2000 suffit à isoler l'axe 1 (dit de statut social), alors qu'il en faut au moins 500 pour faire apparaître les axes 1 et 2, etc...).

Dans l'étude déjà citée sur la répartition géographique du système des soins médicaux, le plan des deux premiers facteurs restait pratiquement identique à lui-même, que les départements soient considérés comme des unités statistiques de même poids (aspect géographique privilégié) ou au contraire soient affectés d'un poids proportionnel à leur population (aspect démographique privilégié). Ce résultat est d'autant plus rassurant quant à la stabilité de ces facteurs que ces poids varient presque de 1 à 100.

REFERENCES BIBLIOGRAPHIQUES DU CHAPITRE II

[1] BENZECRI J.P. (1973) Analyse des données - Tomes 1 et 2 - Dunod - Paris.

[2] BRILLOUIN L. (1959) La science et la théorie de l'information - Masson - Paris.

[3] DEVILLE J.C. (1974) Méthodes statistiques et numériques de l'analyse harmonique - Annales de l'INSEE n° 15.

[4] JEFFREYS H. (1946) An invariant form for the prior Probability in Estimation Problems - Proc. Roy. Soc. (London) A. Vol. 186 - p. 453-461.

[5] KATO (1966) Perturbation Theory for Linear Operator - Springer Verlag - Berlin.

[6] KULLBACK S. (1959) Information Theory and Statistics - Wiley - New-York.

[7] LEBART L., TABARD N. (1970) : Application des méthodes d'analyse des données à la préparation des enquête auprès des ménages - J. Soc. Stat. - Paris - n°4 - 1970.

[8] MALINVAUD E. (1964) Méthodes statistiques de l'économétrie - Dunod.

[9] RENYI A. (1966) Calcul des probabilités - Dunod - Paris.

[10] SHÜTZENBERGER M.P. (1954) : Contribution aux applications statistiques de la théorie de l'information - Publ. de l'ISUP - 3 - 1954 - P. 1 - 117.

[11] THOM R. (1968) Topologie et signification - Age de la science n°4.

[12] THOM R. (1974) Modèles mathématiques de la morphogenèse - 10/18 - n° 887.

[13] WILKINSON J.H. (1965) The Algebraïc eigenvalue Problem - Clarendon Press - Oxford.

Chapitre III

ETUDE CRITIQUE DE L'UTILISATION DES METHODES.

I - Généralités.

II - Analyse de quelques critiques et réserves.

III - Nouvelles possibilités offertes par l'analyse des données.

IV - Références bibliographiques.

I - GENERALITES

Dans ce paragraphe d'introduction, nous délimitons l'objet de notre analyse critique. Dans l'éventail des méthodes disponibles, nous étudierons plus spécifiquement les techniques d'analyse factorielle descriptive, les plus utilisées, et selon nous les plus utiles ; et nous préciserons les matériaux statistiques (tableaux ayant certaines propriétés et qualités) sur lesquels ces techniques s'appliquent avec profit. Auparavant, nous dirons quelques mots des circonstances de l'apparition d'un nouvel instrument d'observation, en nous référant à un exemple emprunté à l'histoire de la biologie.

I-1. Apparition d'un nouvel instrument d'observation.

Pour pouvoir apprécier la validité de représentations issues des méthodes d'analyse des données, il est nécessaire de comprendre quelle est la nature de ces résultats : aucune analogie, aucun exemple emprunté à d'autres techniques d'investigation ne permet jusqu'à présent de caractériser de façon pleinement satisfaisante ces méthodes et les représentations auxquelles elles conduisent.

On peut, en première analyse, parler *d'instrument d'observation de la réalité multidimensionnelle* ; définition peu précise qui suscite aussitôt toute une série d'analogies avec les instruments d'observation de "l'infiniment éloigné" que constituent les lunettes astronomiques et les télescopes, de "l'infiniment petit" que permettent d'explorer les divers types de microscopes, de "l'opaque" partiellement révélé par les appareils radiographiques et les scintillographes.

L'apparition et la diffusion de techniques nouvelles se faisant toujours à l'intérieur de groupes sociaux particuliers, dans des cadres institutionnels spécifiques, on peut penser que les obstacles qui surgissent, les incompréhensions, les illusions éventuelles vis-à-vis des techniques de statistiques multidimensionnelles sont similaires aux réactions qui accompagnèrent, en d'autres temps, l'apparition de nouveaux instruments d'observation (plus précisément des instru-

ments permettant d'observer des aspects de la réalité jusque là inexplorés).

L'exemple de l'apparition, puis de la diffusion du microscope dans la communauté scientifique de l'époque est, selon nous, assez riche d'enseignement. Comme le note F. JACOB [réf 16] *"Même quand un instrument vient soudain accroître le pouvoir de résolution des sens il ne présente jamais que l'application pratique d'une conception abstraite. Le microscope, c'est la réutilisation des théories physiques sur la lumière. Et il ne suffit pas de "voir" un corps jusque là invisible pour le transformer en objet d'analyse. Quand LEEUWENHOECK contemple pour la première fois une goutte d'eau au microscope, il y trouve un monde inconnu, des formes qui grouillent, des êtres qui vivent ; toute une faune imprévisible que l'instrument, soudain, rend accessible à l'observation. Mais la pensée n'a alors que faire de tout ce monde... Cette découverte permet seulement d'alimenter les conversations... Sujet de distraction aussi, pour les cours et les salons qui s'adonnent à la science amusante..."*. Le champ d'observation de l'analyse des données n'est évidemment aucunement comparable à celui des microscopes, et l'analogie ne pourra aller très loin. Il n'y aurait cependant que quelques mots à changer pour évoquer l'apparition de certaines techniques d'analyse factorielle et les réactions qu'elles suscitent. Même quand F. JACOB nous dit plus loin que LEEUWENHOECK, à force d'observer des animalcules dans la semence mâle, déclare dans une lettre à LEIBNIZ : *"les animalcules diffèrent par le sexe, et l'on distingue des mâles et des femelles"*. Il s'agit incontestablement d'une projection des idées régnantes à l'époque qui considèrent la femelle comme un simple réceptacle ; les observations erronées (qui fourmillent dans l'histoire de la biologie) ne mettent évidemment pas en cause l'instrument ; elles nous rappellent simplement que le récepteur ultime de l'information est l'homme, et qu'aux erreurs instrumentales s'ajoutent toujours les erreurs de l'observateur. Nous aurons l'occasion de reparler de ces problèmes qui touchent de très près les procédures de validation des résultats plus bas.

I-2. Les méthodes.

Pour l'essentiel, les méthodes d'analyse des données recouvrent les méthodes d'analyse factorielle au sens large et les techniques de classification. Par une ironie terminologique assez largement imposée par l'usage, les méthodes d'analyse factorielle au sens strict (analyse factorielle des psychologues, ou encore, analyse factorielle en facteurs communs et spécifiques) ne font pas partie des techniques d'analyse des données proprement dites, puisqu'elles ne se limitent pas au rôle d'instrument d'observation et stipulent un modèle statistique particulier. (Les factorialistes eux-mêmes étaient cependant assez divisés sur ce sujet ; *Cyril BURT* et ses disciples ne rejetaient pas systématiquement l'utilisation de la méthode à des fins purement exploratoires).

Les deux principales méthodes d'analyse factorielle de données amorphe (*) sont alors l'analyse en composantes principales et l'analyse des correspondances, cette dernière s'appliquant avec succès à une classe de codage sensiblement plus large. Les méthodes de classifications automatiques sont d'un usage moins courant, bien qu'indispensables pour des opérations spécifiques comme les constructions de nomenclatures par agrégation. Elles peuvent cependant compléter et enrichir les résultats des analyses factorielles. La multiplicité des techniques existantes et l'effervescence qui règne autour de ce domaine nouveau de recherche (plus de 1000 publications par an selon *R.M. CORMACK* [réf 10]) nous incitent à une certaine prudence méthodologique.

L'utilité des algorithmes d'agrégation autour de centres mobiles [réf 1, 11, 3 : T.1] est indicutable lors du traitement de fichiers volumineux ; certaines méthodes de classification ascendante hiérarchiques ont également fait leurs preuves [cf. réf 3 : T1] , mais l'utilisateur est perplexe devant la multitude des variantes proposées, au point qu'on peut dire [cf. réf 10] "qu'il est presque plus facile

---

(*) Rappelons que nous désignons par recueil de données amorphes un recueil sur lequel on n'a pas d'information "a priori" du type "division des variables ou individus en deux ou plusieurs groupes", "relations d'ordre sur les individus" etc...

pour un chercheur d'inventer à propos d'un recueil de données un nouvel algorithme que de tirer quelque chose de ces données...". A peine nées, les méthodes de classification seraient-elles menacées par le développement incontrôlé de métastases théoriques, source de diversions et de confusions ? Ce ne serait pas un phénomène nouveau en mathématique appliquée ni même en statistique. Il sera de toute façon limité si l'on s'astreint à juger des *applications* d'après les normes et les critères de la discipline d'où sont issues les données.

D'autre part, si l'outil qu'est la classification est fondamentalement différent de l'analyse factorielle, le champ d'observation et la méthode de travail sont relativement proches. C'est pourquoi nous ne pensons pas être trop restrictifs en nous limitant, dans les propos qui vont suivre, à des considérations sur les deux méthodes d'analyse factorielle précitées.

I-3. Les matériaux observables.

Les domaines auxquels les méthodes d'analyse des données peuvent s'appliquer sont évidemment très variés, mais il est bien connu que le recueil de données soumis à l'analyse doit posséder certaines qualités. Les deux premières qualités sont l'homogénéité et l'exhaustivité [cf. réf 3 : T.A2 n°2]. L'homogénéité est habituellement comprise comme une homogénéité de texture du tableau analysé (le codage doit permettre une certaine comparabilité entre lignes ou entre colonnes ; on ne doit pas mélanger des quantités exprimées en grammes et en mètres : d'où une conversion en codage par classe ou par rang afin d'atteindre cette homogénéité). Cependant, il est important, pour la clarté de l'interprétation, que les matériaux analysés simultanément aient également une certaine homogénéité de substance, ou, si l'on préfère de contenu, respectant ainsi le principe de pertinence, recommandé par les linguistes [cf. réf 23] qui consiste à ne retenir dans la masse hétérogène des faits que ceux qui se rapportent à un seul point de vue. Cette condition supplémentaire permet souvent

de clarifier et de rendre plus aisées les interprétations : en pratique, cela reviendra à distinguer plusieurs familles de variables, certaines d'entre elles jouant un rôle actif dans la construction des typologies, les autres n'intervenant que comme variables illustratives [cf. §-III.1].

Le critère d'exhaustivité est plus difficile à expliciter ; il s'agit plus, en fait, d'un conseil que d'une règle stricte : il implique que toutes les situations ou tous les aspects d'un phénomène soient représentés, sans exiger cependant une représentativité au sens de la théorie des sondages ; car l'exhaustivité s'applique aussi bien aux variables qu'aux individus, et l'on serait bien en peine de définir "la population parente" d'où est extraite l'"échantillon" de variables qui nous intéresse.

A ces deux exigences, nous ajouterons une première condition assez évidente, mais parfois oubliée : le tableau de données doit être vaste. En statistique, l'infini est parfois de l'ordre de 30...
Il est cependant extrèmement embarrassant de dire à partir de quelle taille de tableau l'analyse factorielle mérite d'être utilisée car ce choix dépend de la nature même du tableau : une table de contingence 10 x 10 peut être intéressante à analyser, alors qu'un tableau binaire (notes de présence-absence) de même taille ne le sera probablement pas. L'apport des méthodes d'analyse factorielle sera en général d'autant plus important que le tableau analysé est vaste ; la visualisation s'avèrera en effet beaucoup plus nécessaire, les résultats seront plus stables, l'interprétation plus riche.

Une seconde condition d'obtention de résultats aisément interprétables est d'exiger du recueil de données qu'il soit amorphe. Il convient en effet d'utiliser tout ce que l'on sait (si cela est possible), pour arriver facilement à en savoir plus. Des exemples de tableaux homogènes (en texture et en substance) et non-amorphes sont fournis par les recueils de données traduisant des évolutions chronologiques [e.g. évolution du prix de l'ensemble des produits alimentaires sur 60 trimestres]. Il existe dans ce cas un facteur prédominant et objectif : le temps ; une possibilité de description aus-

si efficace qu'élémentaire : les diagrammes chronologiques ; des techniques numériques élémentaires d'analyse (lissage, désaisonnalisation) ; enfin des techniques plus élaborées (analyse des spectres et cospectres, etc..., faisant intervenir des hypothèses d'ordre statistique du type : stationnarité des fluctuations autour d'un trend, etc...)

On voit que, même en se limitant aux techniques descriptives élémentaires, on peut espérer atteindre des résultats assez fins, ce qui diminue l'urgence du recours à l'analyse des données brutes. Par contre, l'analyse du tableau des accroissements correspondant à des observations consécutives (qui, lui, sera en général amorphe) pourra être d'un grand intérêt.

II - ANALYSE DE QUELQUES CRITIQUES ET RESERVES.

On étudie dans ce paragraphe les circonstances de l'utilisation effective de l'analyse des données et les critiques et réserves que peuvent susciter ces méthodes de la part des utilisateurs dans les disciplines relevant des sciences humaines et aussi de la part des statisticiens familiers de méthodes plus classiques.

On évoque tout d'abord le problème des utilisations inconsidérées, qui relève probablement plus d'une sociologie des institutions scientifiques que de notre discipline, et qui ne concerne pas spécifiquement l'analyse des données.

Puis on analyse une série de critiques : certaines paraissent peu fondées, parce que suscitées principalement par des applications maladroites ou irrelevantes ("les résultats sont des évidences") ; d'autres méritent plus d'attention : l'ordinateur introduit une division du travail qui peut s'accompagner d'une déqualification du statisticien, d'une perte de contact avec le matériel brut.

On rappelle enfin que le mot "hypothèse" peut désigner des notions fort différentes lors des applications statistiques, ce qui est à l'origine de certaines confusions,et que l'on ne travaille que sur des tableaux ayant (éventuellement) une certaine structure, et non directement sur des structures.

II-1. Us et abus de l'analyse des données en sciences humaines.

"Donnez un marteau à un enfant, et vous verrez que tout lui paraîtra mériter le coup de marteau". A. KAPLAN [réf 17, in réf 8].

Le statisticien spécialiste des méthodes d'analyse des données doit lutter sur deux fronts : celui des monomaniaques de la technique, dont il faut réfréner l'ardeur dévastatrice ; celui des réticents dont les motivations mériteront aussi d'être examinées en détail.

Qu'une technique nouvelle donne lieu à des usages inconsidérés n'a rien de surprenant, surtout si le champ d'application de la technique est vaste, si son utilisation fait intervenir cet outil encore mythologique qu'est l'ordinateur. Les techniques n'ayant aucun intérêt pratique risquent évidemment moins d'être galvaudées. Déjà, en 1952, le psychologue *H.J. EYSENCK* analysait *"The uses and abuses of Factor Analysis"* [cf. réf 13], dans un article qui nous a inspiré le titre de ce paragraphe. Il ne s'agissait alors que de l'analyse factorielle des psychologues, et du mauvais usage qui était fait de l'interprétation des facteurs et du modèle sous-jacent.

Il arrive malheureusement que la méthodologie statistique ne soit pour un chercheur en sciences sociales, qu'un recours devant une situation désespérée, ou même un refuge où il sera à l'abri des critiques de ses pairs. Quelquefois, il s'agit d'une obligation quasi-administrative : des étudiants en sciences économiques ont ainsi pu demander d'étoffer leurs thèses avec "quelques analyses" destinées à impressionner un éventuel jury, en proposant de soumettre à l'analyse des correspondances des tableaux petits et hétérogènes qui ne méritaient nullement un tel traitement ; ou encore un statisticien qui vend ses services dans un bureau d'étude voit un riche client demander qu'une analyse factorielle soit faite sur des données exécrables. Comme un médecin à qui est demandé un certificat de complaisance, doit-il éconduire ce client qui trouvera satisfaction auprès d'un confrère moins scupuleux ? Certes, si la situation financière du bureau le permet...

Recours, refuge, imposture, les méthodes sont aussi des exutoires. Comme le déplorait déjà le statisticien britannique *Cyril BURT* en 1955, *"Beaucoup de chercheurs semblent supposer que l'on peut partir de n'importe quel recueil de données qui vous tombent sous les mains, et ils attendent que l'analyse statistique en extraie les facteurs appropriés. Quand elle n'y parvient pas ils se plaignent des méthodes"*. [cf. réf 9]

Les méthodes ne doivent pas être jugées d'après les circonstances sociales de leur utilisation. *J.P. BENZECRI* dans son *"Histoire de l'analyse des données"* [réf 4] cite à ce propos (*) la phrase de *STUART MILL "On peut peser du cuivre et le donner pour or, la balance reste sans reproche"*. Les excès et les erreurs que nous dénonçons ici sont dus pour une large part à l'inexpérience et aux illusions des utilisateurs en sciences sociales ; ceci peut nous laisser espérer qu'il s'agit d'une phase transitoire d'adaptation.

Les chercheurs qui sont réticents vis-à-vis de l'utilisation des techniques d'analyse des données (alors que leur activité leur permettrait parfois d'user avec profit de ces techniques) sont pour la plupart des statisticiens de formation classique, souvent extrêmement compétents dans leur spécialité. C'est pourquoi nous examinerons avec attention certaines de leurs objections, auxquelles nous souscrirons d'ailleurs partiellement, avant d'exposer en détail les avantages que l'on peut attendre de ces méthodes, qui sont parfois ignorés.

Une première objection, à laquelle nous ne répondrons pratiquement pas, consiste à évoquer les utilisations abusives ou maladroites des techniques, auxquelles il a été fait allusion au début de ce paragraphe.

Répétons qu'il s'agit là d'un phénomène social lié à la division du travail, à la parcellisation des tâches, à ce que *R. THOM* [réf 27] appelle la sociologisation de la science. Il est d'ailleurs évident que la statistique classique n'a pas été épargnée : les applications licites et pertinentes des techniques relevant du modèle linéaire général (régression et analyse de la variance) ne sont qu'une part infime des applications effectivement réalisées. L'utilisation de l'ordinateur introduit incontestablement une dispersion et des diversions peu favorables à une utilisation critique des outils statis-

---

(*) Le contexte est cependant fort différent, puisqu'il s'agit de la part de *STUART MILL* de critiquer les travaux sur la théorie des jugements de *CONDORCET, LAPLACE, POISSON*.

tiques, et nécessite par conséquent un surcroît de vigilance de la part de l'utilisateur [cf. §-II.3]. Une autre objection concerne la nature des résultats auxquels on reproche de traduire surtout des évidences, et auxquels on dénie le caractère de scientificité [cf. §-II.2].

Nous commencerons par examiner ce dernier point qui révèle en fait une connaissance très partielle des possibilités réelles des méthodes, et quelquefois des illusions sur celles des autres méthodes.

II-2. Evidence et scientificité.

Les chercheurs n'ont pas attendu les techniques d'analyse de données pour voir taxer certains de leurs résultats d'évidence ; ceci est particulièrement vrai dans les sciences humaines où, comme beaucoup d'assertions sont possibles, les résultats sont toujours évidents a posteriori. Il s'agit en fait d'une évidence parmi d'autres... L'exemple célèbre de LAZARSFELD [réf 18, cité dans réf 8] permet de répondre à ceux qui pensent que *"les enquêtes ne font qu'exprimer d'une manière compliquée des observations qui étaient déjà évidentes pour tout le monde"*. Il pose six affirmations, dont le libellé est assorti de commentaire, et dont la réponse semble tellement évidente, que l'on peut se demander *"pourquoi dépenser tant d'argent et tant d'énergie à établir de telles découvertes"*.

Nous ne donnerons que les deux affirmations les plus courtes :

- *"Pendant leur service militaire, les ruraux ont, d'ordinaire, meilleur moral que les citadins (après tout ils sont habitués à une vie plus dure)"*.

- *"Les simples soldats de race blanche sont davantage portés à devenir sous-officiers que les soldats de race noire (le manque d'ambition des noirs est presque proverbial)"*.

En fait, chacune des six propositions énonce exactement le con-

traire de résultats réels. LAZARSFELD conclut : *"si nous avions mentionné au début les résultats de l'enquête, le lecteur les aurait également qualifiés d'évidents. Ce qui est évident, c'est que quelque chose ne va pas dans tout ce raisonnement sur "l'évidence". En réalité, il faudrait le retourner : puisque toute espèce de réaction humaine est concevable, il est d'une grande importance de savoir quelles réactions se produisent, en fait, le plus fréquemment et dans quelles conditions, alors, seulement, la science sociale pourra aller plus loin".*

Une règle simple, impliquant une petite discipline, permet au statisticien de s'armer contre ces reproches : elle consiste à demander avant toute analyse aux utilisateurs qui font appel à sa compétence de tracer graphiquement (ou au moins d'esquisser) la typologie qu'ils pressentent comme la plus probable. Il est vraisemblable que, par exemple, le premier axe puisse être à peu près prévu par l'utilisateur ; même dans ce cas, le résultat permettra de déceler quelques inversions dans l'importance des contributions, de nuancer la première esquisse. Cette démarche a également l'avantage de susciter une analyse préalable des données (consultation des moyennes, des histogrammes, des quantiles, des coefficients de variations) beaucoup trop souvent négligée.

Indépendamment de ces évidences inhérentes aux sciences sociales, il en existe d'autres, qui proviennent d'une application maladroite des techniques, ou simplement d'une insuffisante analyse des résultats. Parmi les maladresses les plus souvent rencontrées, on peut noter le cas des données n'ayant pas les qualités requises pour être analysées :

- *Le tableau de données est trop petit :* un examen visuel suffit à en déceler les traits pertinents.

- *Le tableau de données n'est pas amorphe :* bien qu'étant homogène (condition préalable nécessaire), le tableau peut-être doté d'une structure connue a priori ; c'est le cas des tableaux décrivant l'évolution d'un vecteur aléatoire dans le temps. Il n'est pas étonnant de retrouver cette structure, qui n'est pas à proprement

parler une évidence, mais une donnée exogène.

Comme le montrera le paragraphe III-1, les techniques d'analyse factorielle n'ont pas pour seule fin la contemplation d'une image "bligodimensionnelle" du tableau analysé ; cette image, même si elle ne bouleverse pas les conceptions antérieures du chercheur, sera pour lui une trame destinée à recevoir diverses variables illustratives qui l'aideront à éprouver certaines hypothèses : cette démarche *est souvent la phase la plus riche d'enseignement de l'analyse.*

Habitué à des assertions (e.g. : on rejette l'hypothèse d'homogénéité, telle liaison ne peut être imputée au hasard), le statisticien est parfois décontenancé par les descriptions, qui paraissent moins "scientifiques", le critère de scientificité étant souvent une ressemblance formelle avec les sciences exactes, et plus particulièrement la physique.

H.J. EYSENCK [réf 14] déclare notamment *"si nous regardons l'analyse factorielle purement comme une méthode commode de description, permettant de réduire des faits complexes à un niveau plus économique et plus simple, mais n'ayant pas d'implication causale, alors elle n'a pas d'intérêt scientifique, bien qu'elle puisse présenter une utilité pratique".* Nous sommes assez peu sensibles à cette distinction, et ne voyons pas d'antinomie entre ces deux qualités. Nous ne partageons toujours pas les vues de cet auteur lorsqu'il poursuit *"dans ce cas, il n'existe aucun problème de validité, car la validité implique quelque référence dépassant la simple description, fut-elle économique, de séries d'observations".* Nous n'aurions pas cité ce texte s'il n'était révélateur d'une double illusion fréquente chez les utilisateurs (principalement en psychologie et plus généralement dans les sciences de l'homme) concernant d'une part la capacité des méthodes à mettre en évidence des relations causales, d'autre part, la qualité des descriptions qu'elles peuvent fournir. Le premier point est classique : nous n'observons que des covariations, qui sont des présomptions d'autant plus fortes de causalité qu'elles sont calculées dans un contexte laissant, autant que faire se peut, "toutes les choses égales par ailleurs" (par sélection de sous-échantillons, ou, avec une extrême prudence, utilisation de

corrélations partielles).

Le second point révèle l'ambiguïté du terme "description", qui désigne à la fois les résultats et la description de ceux-ci. Les techniques d'analyses factorielles, telles que nous les utilisons, nous donnent en fait des images dont nous pouvons retenir, en vue d'une description, la partie stable, c'est à dire la partie commune aux images issues de certains voisinages des données.

En fait il est utile de distinguer la méthode de l'outil proprement dit (celui-ci pouvant donner lieu à applications insignifiantes sans celle-là) comme le suggère *J.P. BENZECRI* dans sa conclusion de l'étude historique citée en [réf 4]. Mais avant de procéder à une tentative de "défense et illustration de la méthode" (§-III), on examinera une critique plus nuancée qui concerne les conditions techniques et pratiques d'utilisation de l'analyse des données (qui concerne en fait toute les applications statistiques nécessitant le recours à l'ordinateur).

II-3. L'usage de l'ordinateur.

Les résultats sont trop facilement acquis : leur interprétation détaillée est négligée par les statisticiens "nantis". *La disparition de l'effet dissuasif du calcul n'est pas sans inconvénient :* il y a seulement une quinzaine d'années, avant d'aborder l'épreuve du calcul qui le détournait souvent de son champ d'intérêt propre le statisticien étudiait longuement ses hypothèses, envisageait méticuleusement toutes les conséquences de ses calculs, éventuellement faisait des essais sur maquette (sous-échantillon).

Au cours de la phase de calcul, il était en contact quasi-physique avec les données statistiques ; les anomalies ou incohérences pouvaient apparaître de façon progressive ; la critique toujours possible au cours de l'exécution de certaines opérations arithmétiques ou des diverses manipulations permettait éventuellement d'infléchir le déroulement des calculs, ou de les abandonner prématurément. Mais aussi le temps consacré à l'interprétation des résultats était à la

mesure du travail nécessaire à leur établissement. Brièvement, disons que toute une série d'opérations qui n'était pas sans influence sur la qualité scientifique des travaux ne relève plus de l'activité du statisticien.

Le travail des taxinomistes polonais ("WROCLAW TAXONOMY", comprenant notamment l'algorithme de recherche de l'arbre de longueur minimale de FLOREK (1952) redécouvert plus tard par les statisticiens occidentaux) est un exemple d'analyse "manuelle" des données donnant lieu à une connaissance très fine du matériel statistique. Les représentations obtenues sont beaucoup moins riches d'information que celles issues de l'analyse des correspondances, et les défauts et la vulnérabilité de ces algorithmes sont connus. Cependant, le processus d'établissement des résultats est lui-même un mode de connaissance.

Un exemple beaucoup plus actuel est fourni en France par les utilisateurs de la technique manuelle de traitement de tableau de J. BERTIN [cf. réf 6, 7] qui préfèrent se contenter d'une représentation très partielle (ne faisant intervenir que des modifications de l'ordre des lignes et des colonnes du tableau) mais dont le processus de formation est contrôlable visuellement et aisément explicable aux utilisateurs les plus variés. Cependant, ces manipulations manuelles sont limitées aux petits tableaux (les fichiers d'enquêtes les plus courants donnent lieu à des tableaux de l'ordre de (200x1500) qui nécessitent évidemment le recours à l'ordinateur). Il est d'ailleurs possible, après une analyse des correspondances, de faire éditer le tableau initial en réordonnant ses lignes et ses colonnes selon l'ordre des premiers facteurs. On démontre d'ailleurs que l'ordre obtenu a des propriétés optimales pour certains types de tableaux.

Le procès que nous faisons ainsi à la trop grande disponibilité des méthodes et à l'usage parfois inconsidéré qui en résulte n'est évidemment qu'une pièce à ajouter au procès beaucoup plus général des applications indésirables du développement technologique : celui-ci introduit en général une division du travail (utilisateur, statisticien, informaticien dans notre cas) qui est à l'origine de beaucoup de difficultés, d'incompréhension, de gaspillage.

Le bilan de l'automatisation est cependant loin d'être négatif ; l'effet dissuasif du calcul n'a pas toujours été salutaire : il a parfois entraîné une modélisation excessive, un éloignement des faits et des observations, des théories pléthoriques n'ayant que peu de rapport avec la réalité. Mais l'apport le plus fondamental ne concerne évidemment pas la vitesse de résolution des problèmes anciens, mais la possibilité de traiter simultanément des masses d'informations qui seraient dénaturées ou mutilées par une parcellisation ; également la possibilité d'expérimenter sur ces traitements. Nous étudions plus précisément ce que nous estimons être les principaux apports de l'analyse des données aux sciences humaines au paragraphe III.

II-4. Présupposés et hypothèses.

L'étude de la validité des images ou des représentations issues d'un instrument d'observation suppose la mise en évidence d'erreurs instrumentales, mais aussi d'erreurs de l'observateur, et enfin d'erreurs ou confusions relatives au matériel observé lui-même.

Dans certaines disciplines, notamment dans les sciences sociales, une utilisation hâtive et inadaptée de l'analyse des données a engendré quelques déceptions à la mesure de certaines illusions : les méthodes n'utiliseraient aucune sorte d'hypothèse, n'auraient aucune exigence en ce qui concerne la préparation des données, et fourniraient néanmoins des images ayant une valeur intrinsèque.

Une confusion concernant les différentes acceptions du mot hypothèse, et les différents types d'hypothèses, pourrait être à l'origine d'un véritable malentendu. Tout d'abord, nous réserverons le nom de *présupposés* aux hypothèses, quelles qu'elles soient, lorsqu'elles sont *non explicitement formulées.*

A propos de ceux-ci et des discussions souvent stériles auxquels ils donnent lieu, nous partageons un peu l'agacement d'A. REGNIER qui écrit notamment [cf. ref 24, p.394] :

"Des présupposés, il y en a toujours, la question est simplement de savoir s'ils vont introduire ou non des erreurs dans l'analyse qu'on se propose de faire. On n'a pas forcément besoin de savoir quels sont exactement ces présupposés pour savoir qu'ils n'introduiront pas d'erreur appréciable. En particulier, on n'a pas besoin de savoir toujours ce qu'on néglige pour savoir que c'est négligeable. Ainsi, dans une mesure, il n'y a pas à savoir quelle est exactement l'erreur qu'on commet, il suffit de savoir qu'elle est inférieure à un certain seuil. Si tel n'était pas le cas, la science, au lieu d'avancer, aurait passé son temps sur place, à regarder ses pieds".

Il n'est évidemment pas question de nier l'existence des présupposés : lorsque cela est possible, on cherchera à les expliciter, et donc à leur donner le statut d'hypothèses. Dans les autres cas, on s'efforcera d'être vigilants, sans pour autant se sentir paralysés. En se limitant au domaine des applications de la statistique, on peut distinguer deux grandes familles d'hypothèses :

1) *Les hypothèses relatives au contenu même de la recherche*, sans lesquelles il n'y aurait même pas de recueil de données (que nous appellerons : hypothèses générales).

Par exemple, dans un modèle économétrique classique : l'hypothèse selon laquelle la consommation de viande d'un ménage dépend du revenu et de la taille de ce ménage.

2) *Les hypothèses nécessaires à l'utilisation d'une technique statistique particulière* (hypothèses techniques (*)). Pour l'exemple précité, pour fixer les idées, hypothèses que les variables exogènes sont mesurées sans erreurs trop importantes, que les résidus vérifient la condition d'homoscedasticité et éventuellement soient distribués suivant une loi normale, etc...

(*) M.G. MORLAT a proposé de désigner ces hypothèses par "hypothèses de commodité". Intervention aux Journées de méthodologie de la recherche en psychiatrie - Marseille - Avril 1975.

Cette distinction se fait assez indépendamment de ce que l'on désigne habituellement par "hypothèse de travail". Toujours selon A. REGNIER [op. cit. p.398] *"un travail qui s'appuie sur une hypothèse de travail donnée doit être mené de telle façon que si elle est fausse, on s'en aperçoive, la différence avec un travail de confirmation d'une hypothèse étant qu'ici, confirmer l'hypothèse de travail n'est pas le tout"*.

Des hypothèses de travail peuvent intervenir au niveau général comme au niveau technique.

Dans beaucoup d'applications statistiques, l'hypothèse technique la plus importante est l'hypothèse de normalité.

Et quand un statisticien dit qu'il ne fait pas ou peu d'hypothèses, cela signifie en général qu'il va utiliser une procédure robuste ou non-paramétrique, donc qu'il compte s'affranchir de certaines hypothèses techniques, et non qu'il adhère à un quelconque "dogme de l'immaculée perception" (l'expression est de F. NIETZSCHE ; [cf. réf 8, p.64]).

Les techniques d'analyse de données figurent précisément, en statistique, parmi les techniques qui stipulent les hypothèses techniques les plus faibles (*).

Il est d'ailleurs à peine exagéré de dire dans ce cas "que l'on ne fait pas d'hypothèse", si l'on pense à ce qu'est l'hypothèse de multinormalité (toutes les lois marginales et conditionnelles sont normales, toutes les surfaces de régression des hyperplans, etc...)

Un débat entre statisticiens se tient actuellement au niveau des hypothèses techniques, et ce serait ajouter à une certaine confusion que d'introduire pêle-mêle dans ce débat les hypothèses générales et les présupposés des observations.

(*) Il serait donc paradoxal de leur reprocher précisément ces hypothèses, à moins de contester a priori l'utilité même des vastes recueils de données codées [cf. réf 12].

En tout état de cause, une étude critique d'une technique particulière d'analyse de données doit se fonder, selon nous, sur une comparaison avec toute autre méthode disponible appliquée au même corpus.

Structure et codage.

Les méthodes d'analyse factorielle descriptive permettent d'obtenir des images d'un tableau de données. Contrairement aux images de la statistique descriptive élémentaires, ces images ne sont pas d'une lecture évidente, elles méritent la connaissance de *règles d'interprétation* strictes, dont l'expérience montre qu'elles sont en général assez mal connues des utilisateurs.

Ces règles d'interprétation permettent de procéder à des inductions concernant la structure du tableau de données ; d'où l'idée de procéder à un *étalonnage* de l'instrument d'observation, en soumettant à l'analyse des tableaux ayant des structures connues, comme les matrices associées à des graphes symétriques. Nous avons ainsi comparé (*) les images obtenues (dans les plans des deux premiers facteurs) en appliquant respectivement l'analyse en composantes principales et l'analyse des correspondances à la matrice associée à un réseau à maille carrée. Un travail également empirique de *H. LEBRAS* [réf 22] s'adresse, semble-t-il, à ceux qui entretiennent de dangereuses illusions sur la capacité qu'auraient certaines méthodes d'analyse de mettre en évidence des structures abstraites indépendamment du codage et de la "mise en tableau".

Si l'on change le codage, les images changent (en général) *mais les règles d'interprétation changent aussi*, le seul invariant est le couple (image, règle). C'est pourquoi nous estimons risqué de diffuser dans des revues non spécialisées des plans factoriels sans les accompagner de minutieuses recommandations à l'intention des lecteurs.

(*) L. LEBART, Introduction à l'analyse des données - Consommation n°4 - 1969 p.65-87, Dunod, éd.

Il existe cependant des codages meilleurs que d'autres : ce sont ceux qui conduisent aux règles d'interprétation les plus simples et les plus claires.

Dans le cas des réseaux à mailles carrées, *J.P. BENZECRI* a montré qu'un calcul analytique pouvait éviter le recours à l'ordinateur [cf. réf 3, T.II-B, n°10, sur l'analyse de la correspondance définie par un graphe, ex.4 p.256]. En fait il suffit de choisir un graphe plus simple (par exemple *le cycle* étudié au chapitre II, §.I-2) pour étudier *analytiquement* l'effet des variations de codage de la relation binaire sur les résultats.

Nous avons déjà souligné l'importance de la notion de codage au chapitre II, §.III-2.c, qui est une préparation du matériel à observer nécessitant obligatoirement la collaboration du statisticien et de l'utilisateur, dont les exigences peuvent être contradictoires. L'une des premières fonctions du codage est de donner un *sens* aux distances entre lignes et colonnes du tableau *avant toute analyse*, de façon à permettre une induction aisée au moment de la lecture des résultats. Cette exigence limite considérablement l'éventail des codages possibles.

III - NOUVELLES POSSIBILITES OFFERTES PAR L'ANALYSE DES DONNEES.

L'utilisation licite et pertinente des méthodes d'analyse des données aux sciences humaines a un impact qu'il est encore difficile d'apprécier. On peut, sans prétendre être exhaustif, distinguer plusieurs types de contributions que nous allons expliciter et commenter. Ces conséquences heureuses de l'utilisation des méthodes ont été observées lors d'applications au domaine socio-économique ; il n'est donc pas impossible que des apports décisifs remarqués à propos d'autres domaines (psychologie, histoire, etc...) viennent s'ajouter à ceux que nous avons relevés.

Nous distinguerons, de façon peut-être artificielle, des *apports d'ordre technique* (gain de productivité dans certaines phases des dépouillements d'enquête, possibilités nouvelles d'éprouver la cohérence d'informations complexes, de détecter des erreurs de natures diverses, construction d'indices synthétiques), et des apports fondamentaux que nous regroupons dans le paragraphe intitulé : *De nouvelles voies méthodologiques*.

Nous évoquerons les perspectives offertes par l'extension des champs d'observation, les possibilités de découverte de faits ou phénomènes jusque là inaccessibles, enfin l'enrichissement conceptuel qui résulte des possibilités de visualisation.

III-1. *Apports d'ordre technique.*

a) *Gain de productivité.*

Ce paragraphe traitera principalement de certaines phases des dépouillements d'enquêtes, pour lesquels l'analyse des données est devenue un outil selon nous indispensable : ces techniques permettent de procéder à un ordonnancement rationnel des tâches dans le temps, d'éviter certains piétinement, de contrôler par des représentations visuelles la plupart des étapes de travail, enfin d'accéder à des informations autrefois inaccessibles. Elles permettent également de procéder à des tests de cohérence et à des détec-

tions d'erreurs, mais ce dernier point fera l'objet du paragraphe 2 ci-dessous. L'ensemble des opérations implique en fait simultanément un gain de productivité et un gain dans la qualité des résultats. L'aide apportée par l'analyse des données (en fait par une variante de l'analyse des dorrespondances) est résumée dans l'article cité en [ref 20]; l'exemple d'application contenu dans cet article figure en annexe 2.

Indiquons-en les traits essentiels.

Classiquement, le dépouillement d'enquête consiste en l'établissement d'une série de tabulations, dont la plupart sont prévues à l'avance lors d'un *plan de tabulation* établi au moment de la conception du questionnaire. Si l'enquête porte sur un sujet nouveau, le plan de tabulation risque d'être incomplet ou inadapté (une tabulation exhaustive est en général impossible lorsque le questionnaire n'est pas particulièrement réduit).

Ce plan de tabulation est fait à partir de variables que l'on peut qualifier de socio-administratives (âge, profession, nombre d'enfants, catégories de communes, niveau d'instruction, etc...). Il ne tiendra pas compte, en général, des interrelations existant entre ces variables (profession et niveau d'instruction par exemple) et la lecture des tableaux sera riche en redondances. La construction d'une grille socio-administrative (cf. : annexe 2) permet au contraire de synthétiser ce réseau d'interrelations, et la technique de projection de variables illustratives permet de remplacer la lecture de plusieurs tableaux par une seule lecture graphique. Si la visualisation se fait à partir des deux dimensions principales de la grille, celle-ci peut cependant avoir plus de deux dimensions. Il est alors possible de procéder à un tri automatique de toutes les variables relatives au contenu de l'enquête, en croisant au besoin celles-ci deux à deux, et de retenir les plus dépendantes de la grille, c'est-à-dire de l'ensemble des variables socio-administratives. Il s'agit en quelque sorte d'une sélection des tabulations les plus pertinentes opérée dans un ensemble beaucoup trop vaste pour que l'opération puisse être faite manuellement. (Il faudrait alors consulter, par exemple, des milliers de tableaux).

Il s'agit donc ici d'application d'analyse des données, non pas à un travail de recherche, mais à une activité presque routinière pour un statisticien qui dispose maintenant d'outils à la mesure des recueils de données qu'il a pour charge d'analyser. L'ère de la sous-exploitation des données d'enquêtes socio-économiques (par ailleurs extrémement coûteuses à fabriquer) est peut-être révolue...

b) Epreuves de cohérence des données et détection d'erreurs.

La détection des valeurs aberrantes ou erronées est une péripétie familière aux statisticiens qui utilisent les techniques d'analyses factorielles.

Toujours dans le domaine des dépouillements de fichiers d'enquêtes socio-économiques, on peut procéder à une *véritable critique des données* en faisant apparaître des "variables techniques" sur la grille socio-administrative [cf. annexe 2, et supra]. De même que l'on procède à une illustration de cette grille par des caractéristiques des ménages relatifs au contenu même de l'enquête (ménages ayant répondu positivement à telle question, ménages appartenant à une association particulière, à une certaine tranche de revenu, etc...), on peut aussi illustrer cette grille par des variables caractérisant la fabrication de l'information (ménages enquêtés par une même personne, heure de l'interview, éventuellement appréciation ou remarques de l'enquêteur codées convenablement, etc...).

On peut également procéder à une analyse critique du questionnaire lui même, en étudiant les positions des "non-réponses" à certaines questions par rapport aux positions des réponses effectivement exprimées [cf. réf 26]. Enfin, a propos de n'importe quel recueil de données, on peut détecter des erreurs de mesure, ou des anomalies tenant à la méthode de mesure.

Ainsi un récent travail sur les notations de 1 à 20 (analyse de tableaux de contingence croisant les 20 notes et les matières

ou objets auxquels ces mots s'appliquaient, lors de deux recueils concernant respectivement des matières scolaires et des appréciations sur des qualités de magasins, [cf. réf 24]) a montré que sur le premier facteur, l'ordre des notes était respecté, avec dans les deux cas une interversion, celle des notes 18 et 19. Ceci permet de conjecturer un biais inhérent à l'acte de notation, imputable sans doute à une perception particulière de l'échelle des notes.

Un autre exemple est fourni par un travail de géologie [réf 15], qui a consisté, en particulier, en l'analyse d'un ensemble d'échantillons de sables caractérisés par la proportion des différents minéraux lourds qu'ils contenaient. Le premier axe factoriel extrait a classé ces minéraux dans l'ordre de leurs densités, avec cependant une exception, qui révèle probablement une confusion (due à une ressemblance malheureuse) dans la détermination (optique) des différents minéraux.

Il s'agit dans ce dernier cas d'une anomalie qui aurait été indécelable sans un traitement global des matériaux disponibles.

c) Construction d'indices synthétiques.

Le premier facteur est la combinaison linéaire des variables ayant variance maximale, et donc le meilleur indice discriminatoire entre individus. Dans beaucoup d'application, cet indice synthétique a un grand pouvoir descriptif et possède une interprétation intéressante : c'est par exemple le facteur général d'aptitude de l'analyse des psychologues ; d'autres facteurs peuvent également avoir des interprétation intéressantes et constituer des variables artificielles. On a pu ainsi constituer un indicateur socio-culturel [réf 26] à partir de variables telles que : profession, niveau d'instruction, profession des ascendants, branche d'activité, etc... ayant un grand pouvoir explicatif de certaines attitudes.

De façon analogue a été construit un indicateur de la forme de la mortalité [réf 19] à partir des composantes des profils de

mortalité départementale qui s'est révélé fortement corrélé au taux net de mortalité départementale

III-2. De nouvelles voies méthodologiques.

a) Accès à de nouveaux champs d'observation.

La probabilité de traiter simultanément de nombreuses informations permet de procéder à des recueils de données ne *réduisant pas a priori le champ de l'observable*. Les statisticiens ont pendant longtemps été préoccupés par les observations nombreuses de variables elles-mêmes peu nombreuses, qu'il s'agisse de valider un modèle particulier de dépendance, ou simplement de tester une hypothèse. L'infini du statisticien concerne presque toujours la dimension répétitive "individu ou observation", et exceptionnellement le nombre des variables. (La convergence de certaines estimations en analyse factorielle, lorsque le nombre des variables augmente indéfiniment a été évoquée par HOTELLING dès 1933, puis par d'autres factorialistes). L'obstacle du calcul étant levé, on peut maintenant procéder à un *échantillonnage sur deux dimensions* que l'analyse des correspondances, destinée initialement aux tables de contingence, traite d'ailleurs de façon symétrique. Cette possibilité d'exploration de la dimension "variable" est une innovation dont les premières conséquences ne sont qu'entrevues.

Traitant de l'économie, J.P. BENZECRI [réf 5] doute que les approches réductionnistes (explication du complexe par le simple) y soient possibles comme en physique car dans cette discipline encore, comme dans d'autres branches des sciences humaines *"l'ordre du composé vaut plus que les propiétés élémentaires des composants"*.

Ceci permet de définir, sinon un champ d'observation précis, du moins une certaine optique, un certain regard sur le "composé", le "complexe", mots vagues qui se préciseront lors de chaque application, car ces entités se matérialiseront sous forme de "corpus" ou recueil de données ayant certaines qualités (homogénéité et

exhaustivité, principalement, cf. supra). Définir les limites d'un champ d'observation nous procure le même embarras que définir l'exhaustivité : à vrai dire les limites n'ont pas besoin d'être précisées, pourvu qu'elles ne soient pas un obstacle à la compréhension du phénomène à un certain niveau.

L'analyse des données apparaît ainsi comme une méthode de description préalable à une approche "compréhensive ou structurale", par opposition à l'approche "explicative ou réductionniste" (terminologie de R. THOM (*) [réf 27].

Nous verrons plus bas que l'analyse d'un corpus dans sa totalité apporte des informations originales . De nombreuses réponses à des questions d'opinions relatives à une même thème permettent de déceler et d'analyser les contradictions, les incompréhensions, les réticences des enquêtés.

Les analyses des villes de France caractérisées par leurs profils socio-professionnels [réf 21] ou par leur profil de mortalité [réf 19] permettent de dégager des axes stables et font apparaître des phénomènes que l'on aurait pu croire étrangers au corpus. Les résultats qui seront évoqués au paragraphe *b)* ci-dessous sont précisément à imputer au traitement *simultané* de nombreuses variables.

b) Mise en évidence de phénomènes ou de faits méconnus.

Arrive-t-il que l'on découvre des phénomènes alors qu'on ne s'y attendait pas lors d'une analyse ? La méthode de dépouillement d'enquête évoquée au paragraphe 1 nous a montré que l'on pouvait trouver plus vite et de façon plus systématique des résultats quand même accessibles par des méthodes plus classiques au prix d'un effort considérable et souvent dirimant. Mais il arrive que la seule des-

---

(*) "L'approche structurale considère la morphologie empirique telle qu'elle apparaît, et elle ne cherchera pas, - à moins d'y être contrainte - à introduire une théorie causale externe au champ empirique donné, ni, non plus, à considérer des "atomes" appartenant à un niveau d'organisation plus petit" (réf 27, op.cit.).

cription d'un corpus dégage des résultats parfois surprenants. Nous choisirons deux exemples à titre d'illustration schématique en renvoyant le lecteur aux travaux cités pour plus de détails.

b.1) Le profil de mortalité des villes [réf 19 (pages 134-150)]

Une typologie des villes de plus de 50.000 habitants, et des zones départementales complémentaires a été faite en 1969 ; chacune de ces villes ou départements était caractérisée par 17 causes de décès relatifs à toutes les classes d'âge. Il s'agissait donc de données extrêmement grossières, d'une précision toute relative. Les données concernaient la période 1962-1965. (L'analyse a été faite pour chaque année, chaque ville était décrite par son profil de mortalité dont les composantes sont les parts de chacune des causes de décès dans le total des décès ; la population n'intervenant pas). La typologie obtenue s'est révélée stable au cours du temps.

Le plan des deux premiers facteurs (cf. extraits fig.1) permet de voir les villes se séparer des zones départementales complémentaires et des regroupements régionaux. Mais même à l'intérieur de la région parisienne, on observe une séparation entre Boulogne, Neuilly, Asnières, Saint-Maur, Vincennes, Rueil, Versailles, d'une part, Saint-Ouen, Saint-Denis, Aubervilliers et d'autres communes ouvrières d'autre part.

Sans faire de commentaires plus approfondis, nous dirons que les résultats sont aussi fins que les données étaient grossières ; mais le tableau était vaste, et, pour caractériser une ville, 17 informations fragiles peuvent permettre d'observer deux ou trois dimensions stables.

b.2) Contradiction et approbation dans les questionnaires.

Un ensemble de réponses à un questionnaire d'opinion (17 questions totalisant 60 modalités de réponses, posées à 2003 enquêtés [cf. réf 26]) décrivant les attitudes face au travail fé-

LA TYPOLOGIE DES VILLES ET DES DEPARTEMENTS D'APRES LES PROFILS DE MORTALITES

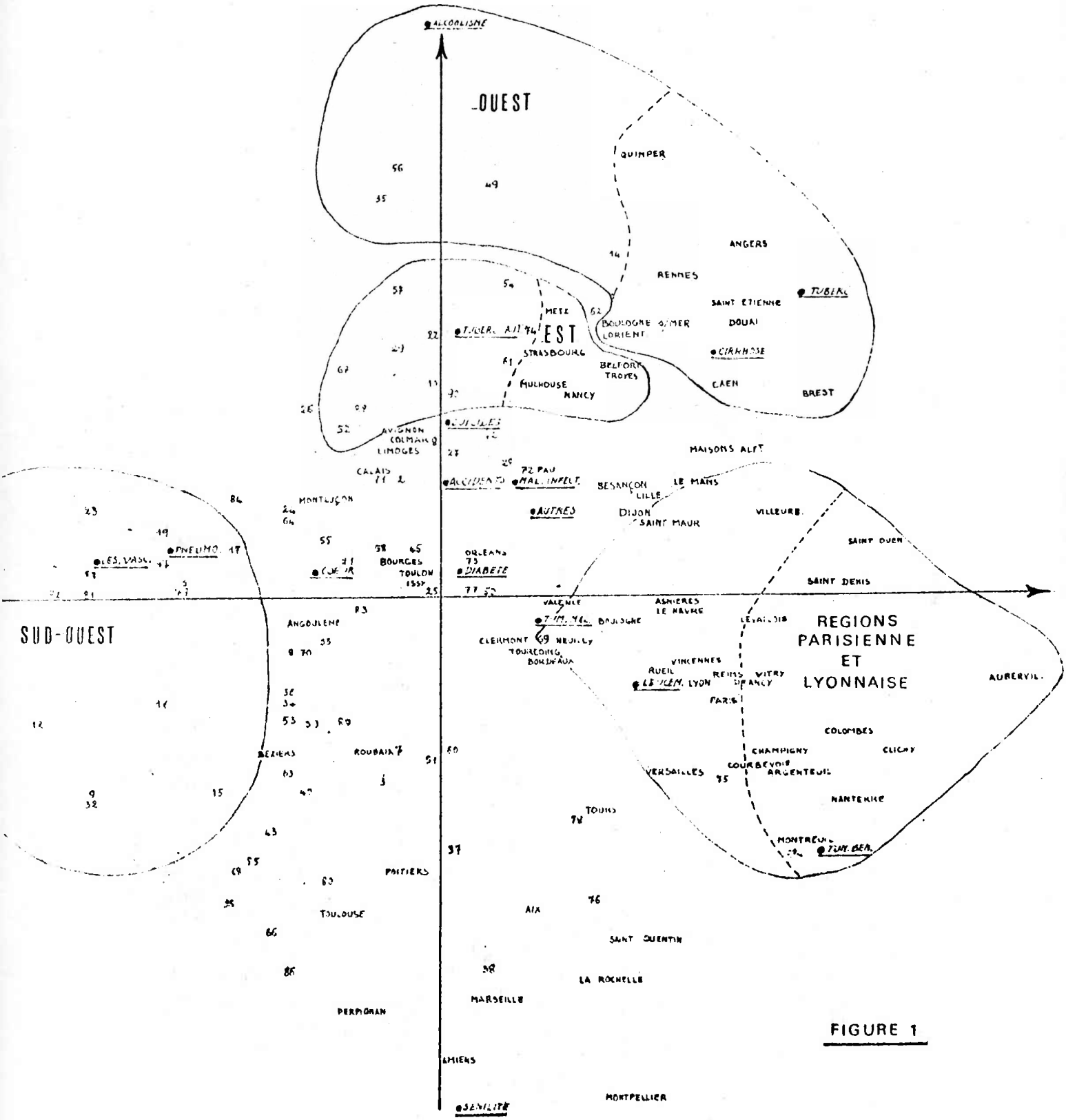


FIGURE 1

minin a été soumis à une analyse des correspondances. Le troisième facteur extrait devait réserver une surprise : alors que le premier facteur discriminait entre les attitudes favorables et défavorables au travail féminin, le second opposait les réponses modérées aux deux types de réponses extrêmes ; le troisième opposait les réponses positives aux réponses négatives, indépendamment du contenu des questions. Les individus enquêtés les plus discriminés par ce dernier axe sont donc des individus qui se contredisent, soit pour approuver systématiquement, soit au contraire (mais ils sont moins nombreux) pour répondre négativement.

La mise en évidence d'un tel phénomène est incontestablement d'une grande importance pour comprendre la validité des enquêtes d'opinion, puisqu'elle révèle notamment une absence de symétrie entre les réponses positives et négatives. La technique d'analyse a fait ici apparaître la *non-neutralité des données*. On se reportera, pour plus d'information, à l'article de *N. TABARD* (*).

A l'issue de ces deux exemples, il convient néanmoins de rappeler que ce n'est pas toujours la phase de description qui est la plus riche d'enseignement, mais au contraire la phase d'illustration de la représentation obtenue pas des données n'ayant pas participé à l'analyse.

c) Nouveaux matériaux et nouveaux concepts.

La présentation sous forme de graphiques des résultats (graphiques dont les règles de lecture sont extrèmement délicates, contrairement à ce que leur caractère suggestif laisserait prévoir) constitue une innovation méthodologique en soi. En effet, le langage usuel, de par son caractère linéaire et séquentiel, permet d'exprimer facilement des liaisons non-symétriques telles que les

---

(*) N. TABARD - Sondages : oui, non, non-réponses.

implications. Une telle relation causale est plus aisée à traduire qu'une relation de *covariation* ; ceci explique en partie pourquoi les chercheurs sont parfois fascinés (et rendus perplexes) par les figures bidimensionnelles issues de l'analyse factorielle. Une description en langage clair de l'ensemble des covariations observées est une tâche décourageante. Que l'on songe au nombre de phrases nécessaires à la description de la typologie des variables de l'annexe 2, si l'on désire pouvoir reconstituer, même très grossièrement, l'allure générale de la figure,(sans faire intervenir, évidemment, les valeurs numériques des coordonnées...). Par où commencer la description, puisqu'il n'existe pas de *relation d'ordre* naturelle dans un espace à deux dimensions ? L'exercice de la description mérite d'être fait. On s'aperçoit très rapidement que l'agencement des diverses variables peut être décrit de façon plus *économique* en faisant intervenir des catégories d'un ordre différent de celui des variables (statut social, cycle de vie, ruralité, aisance, dénuement, privilèges, ... etc... dans le cas de la figure de l'annexe 2) et la *description* devient progressivement *interprétation*. (En sciences humaines, un simple résumé contient presque toujours une part d'interprétation). Pour résumer une partie de la typologie de la figure 1 ci-dessus, nous pouvons par exemple dire : *"à l'intérieur de la région parisienne, qui se distingue elle-même du reste du pays, on peut distinguer deux zones relativement connexes : les agglomérations à caractère plutôt résidentiel et les agglomérations plutôt ouvrières, qui occupent une position plus excentrées"*. Il va de soi qu'il ne s'agit pas d'une *explication* du phénomène observé (relation entre forme de la mortalité et situation géographique) mais d'une assertion interprétative, puisque certains éléments extérieurs au corpus ont été nommés.

L'image bidimensionnelle antérieure au choix d'un langage de description, avec le système de connotation que véhicule celui-ci, constitue selon nous un document de travail, un objet de communication d'une grande densité.

Le chercheur en sciences humaines qui désire observer peut maintenant rassembler ses données sur une base plus large. Il dispose de plus d'une *méthode* pour analyser ses données, qui, parce qu'elle est systématique, suppose un apprentissage, permet des comparaisons, facilite la communication.

Il peut espérer observer des faits qui échappent à tout examen visuel direct des données et dispose, sous formes de cartes munies de leurs règles de lecture, de nouveaux matériaux, pas plus neutres que ses données, mais pas moins, s'il maîtrise l'instrument ; plus neutres en tout cas que toute description en langage clair, et cependant plus proches de la pensée que les données brutes.

REFERENCES BIBLIOGRAPHIQUES DU CHAPITRE III

[1] BALL G.H, HALL D.J. — A Clustering Technique for summarizing Multivariate - data, Behavioral Sciences n°12 - 1967 - pp. 153-155.

[2] BENZECRI J.P. (1973) — L'âme au bout du rasoir - in "Logique et Biologie" Publié sous la direction de B. RYBAK.

[3] BENZECRI J.P. (1973) — L'analyse des données - T.1 - La taxinomie. T.2 - L'analyse des correspondances - Dunod.

[4] BENZECRI J.P. (1975) — Histoire et Préhistoire de l'analyse des données - 150 p. Multigraphié (L.S.M. Université Pierre et Marie Curie - T.45-55)

[5] BENZECRI J.P. (1974) — "La place de l'a priori" - in Encycl. Univers. Organum .

[6] BERTIN J. (1967) — Sémiologie graphique - Paris.

[7] BERTIN J. (1973) — Article GRAPHIQUE (représentation) - Encycl. Univ.

[8] BOURDIEU P., PASSERON J.C., CHAMBOREDON J.C. (1968) - Le métier de sociologue - Mouton.

[9] BURT C. (1955) — L'analyse factorielle - Méthodes et résultats - in Colloques internationaux du C.N.R.S. (1955) "L'analyse factorielle et ses applications".

[10] CORMACK R.M. (1971) — A Review of Classification - Journal of the Royal Statistical Society - Série A - vol.134 - Part.3.

[11] DIDAY E. (1970) — La méthode des nuées dynamiques - R.S.A. vol.19, n°2.

[12] DREYFUS J. (1975) — Implications ou neutralité des méthodes statistiques appliquées aux sciences humaines - Rapport CREDOC-DGRST - 96 pages.

[13] EYSENCK H.J. (1952) — The uses and abuses of Factor Analysis - J. Appl. Stat. 1, p.45-49.

[14] EYSENCK H.J. (1955) — L'analyse factorielle et le problème de la validité in : Colloques internationaux du C.N.R.S. : (1955) l'analyse factorielle et ses applications (CNRS).

[15] HEIN P. (1974) — Méthodes statistiques nouvelles et sédimentologie - Thèse 3ème cycle - Géologie quantitative - Univers. Paris VI.

[16] JACOB F. (1970) — La logique du vivant - Gallimard

[17] KAPLAN A. (1954) — The Conduct of Inquiry - Chandler Pub. Comp. - San Francisco.

[18] LAZARSFELD P.F. (1949) — "The American Soldier : an Expository review" Publ. Opinion Quart. vol. 13, n°3 Fall.

[19] LEBART L. (1970) Recherche sur le coût de protection de la vie humaine dans le domaine médical - Rapport DGRST (CREDOC) - 162 pages.

[20] LEBART L. (1975) L'orientation du dépouillement de certaines enquêtes par l'analyse des correspondances multiples. Consommation n°2 - Dunod.

[21] LEBART L., TABARD N. (1971) - La morphologie sociale des communes urbaines Consommation n°2 - 1971. - Dunod.

[22] LEBRAS H. (1974) Vingt analyses multivariée d'une structure connue. Mathématique et sciences humaines - n°47, p.37-55.

[23] MARTINET A. (1960) Eléments de linguistique générale - A. Colin.

[24] PAGES J. (1975) Notation et classement : deux méthodes de recueil de données - Etude critique à partir d'un échantillon restreint - Consommation n°4, 1975. - Dunod.

[25] REGNIER A. (1974) La crise du langage scientifique - Ed. Anthropos.

[26] TABARD N. (1974) Besoins et aspirations des familles et des jeunes - Coll. Etudes CAF - n°16 - 514 pages.

[27] THOM R. (1974) "La science, malgré tout..." - Encycl. Univers. (Organum).

ANNEXES

Annexe I

EXTRAITS DE CARACTERISTIQUES DES LOIS MARGINALES ET JOINTES DES VALEURS PROPRES ET TAUX D'INERTIE (Histogrammes et matrice des corrélations)

Cette annexe donne, à titre d'illustration, les distributions marginales empiriques des taux d'inertie correspondant aux quatre premières valeurs propres d'un tableau de contingence (8 x 8).

Ces distributions empiriques sont établies à partir de 1000 simulations selon le schéma multinomial exact, qui pouvait être utilisé ici compte tenu des petites dimensions du tableau. (L'effectif total du tableau pour chaque simulation, est k=1000 ; on sait cependant que la loi des taux d'inertie ne dépend pas de cet effectif total, pourvu qu'il soit suffisamment élevé pour permettre l'approximation normale de la loi multinomiale, ce qui est le cas ici) .

Les matrices des corrélations donnent une première idée de la loi jointe des cinq premières valeurs propres, des taux correspondants, de la trace, pour des tableaux (8 x 8), (10 x 10), (12 x 12). Il existe une nette corrélation positive entre valeurs propres successives, et entre celles-ci et la trace ; par contre, les corrélations entre pourcentages d'inertie sont en général négatives.

Rappelons que, pour 1000 observations, la variance d'un coefficient de corrélation est $1/\sqrt{999} \simeq 0{,}031$ (variance lorsque la population parente est normale, mais aussi variance de la loi de permutation).

Nous pourrons vérifier, au niveau des coefficients de corrélation, la propriété d'indépendance des taux d'inertie et de la trace démontrée au chapitre I.

On remarquera également que l'intensité de la corrélation entre valeurs propres successives s'accroît lorsque les dimensions du tableau augmentent. Au contraire, la corrélation négative entre les premiers taux d'inertie s'atténue lorsque ces dimensions augmentent. Parallèlement, le coefficient de corrélation entre une valeur propre et le taux d'inertie correspondant augmente (diminue en valeur absolue, s'il est négatif).

Histogramme des taux d'inertie (1000 pseudo-réalisations, tableau 8 x 8)

VARIABLE NUMERO 1 --- DESCRIPTION DE LA DISTRIBUTION EN 35 CLASSES

Valeur propre 1

MOYENNE DE LA CLASSE	LIMITE INFERIEURE	EFFECTIF DE LA CLASSE	
25.932	25.932	1	**X
28.115	27.109	1	**X
29.388	28.286	2	**XX
30.237	29.463	10	**XXXXXXXXXX
31.330	30.640	15	**XXXXXXXXXXXXXXX
32.362	31.817	17	**XXXXXXXXXXXXXXXXX
33.555	32.994	34	**XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
34.870	34.171	43	**XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
35.976	35.348	53	**XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
37.082	36.525	62	**XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
38.278	37.702	74	**XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
39.459	38.879	67	**XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
40.647	40.057	80	**XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
41.807	41.234	74	**XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
42.962	42.411	72	**XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
44.154	43.588	57	**XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
45.352	44.765	71	**XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
46.490	45.942	50	**XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
47.707	47.119	52	**XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
48.806	48.296	29	**XXXXXXXXXXXXXXXXXXXXXXXXXXXXX
49.989	49.473	31	**XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
51.193	50.650	23	**XXXXXXXXXXXXXXXXXXXXXXX
52.355	51.827	28	**XXXXXXXXXXXXXXXXXXXXXXXXXXXX
53.671	53.004	11	**XXXXXXXXXXX
54.678	54.181	16	**XXXXXXXXXXXXXXXX
55.581	55.358	7	**XXXXXXX
57.120	56.536	4	**XXXX
57.786	57.713	3	**XXX
59.032	58.890	2	**XX
60.478	60.067	7	**XXXXXXX
62.344	61.244	1	**X
62.929	62.421	1	**X
0.0	63.598	0	**
64.907	64.775	1	**X
65.952	65.952	1	**X

VARIABLE NUMERO 2 --- DESCRIPTION DE LA DISTRIBUTION EN 35 CLASSES

Valeur propre 2

MOYENNE DE LA CLASSE	LIMITE INFERIEURE	EFFECTIF DE LA CLASSE	
14.524	14.216	3	**XXX
15.364	14.962	1	**X
16.212	15.707	5	**XXXXX
16.739	16.452	3	**XXX
17.721	17.197	5	**XXXXX
18.435	17.942	2	**XX
19.086	18.688	17	**XXXXXXXXXXXXXXXXX
19.795	19.433	27	**XXXXXXXXXXXXXXXXXXXXXXXXXXX
20.572	20.178	28	**XXXXXXXXXXXXXXXXXXXXXXXXXXXX
21.359	20.923	41	**XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
22.004	21.668	34	**XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
22.824	22.414	54	**XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
23.539	23.159	57	**XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
24.236	23.904	59	**XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
24.981	24.649	66	**XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
25.815	25.394	80	**XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
26.534	26.140	82	**XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
27.273	26.885	77	**XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
27.956	27.630	71	**XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
28.729	28.375	61	**XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
29.497	29.120	45	**XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
30.217	29.865	44	**XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
30.918	30.611	29	**XXXXXXXXXXXXXXXXXXXXXXXXXXXXX
31.750	31.356	29	**XXXXXXXXXXXXXXXXXXXXXXXXXXXXX
32.510	32.101	24	**XXXXXXXXXXXXXXXXXXXXXXXX
33.201	32.846	15	**XXXXXXXXXXXXXXX
33.904	33.591	19	**XXXXXXXXXXXXXXXXXXX
34.803	34.337	10	**XXXXXXXXXX
35.553	35.082	4	**XXXX
35.955	35.827	4	**XXXX
36.620	36.572	1	**X
37.851	37.317	1	**X
38.527	38.063	1	**X
0.0	38.808	0	**
39.553	39.553	1	**X

Histogramme des taux d'inertie (1000 pseudo-réalisations, tableau 8 x 8)

VARIABLE NUMERO 3--- DESCRIPTION DE LA DISTRIBUTION EN 35 CLASSES

Valeur propre 3

MOYENNE DE LA CLASSE	LIMITE INFERIEURE DE LA CLASSE	EFFECTIF	
7.084	6.773	3	**XXX
7.563	7.305	3	**XXX
8.066	7.836	1	**X
8.697	8.367	5	**XXXXX
9.215	8.899	7	**XXXXXXX
9.753	9.430	8	**XXXXXXXX
10.211	9.962	14	**XXXXXXXXXXXXXX
10.773	10.493	26	**XXXXXXXXXXXXXXXXXXXXXXXXX
11.287	11.025	19	**XXXXXXXXXXXXXXXXXX
11.873	11.556	38	**XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
12.342	12.087	33	**XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
12.866	12.619	41	**XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
13.394	13.150	49	**XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
13.970	13.682	46	**XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
14.471	14.213	63	**XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
15.021	14.745	55	**XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
15.530	15.276	65	**XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
16.068	15.807	57	**XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
16.620	16.339	67	**XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
17.135	16.870	60	**XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
17.669	17.402	44	**XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
18.205	17.933	56	**XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
18.686	18.464	52	**XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
19.236	18.996	33	**XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
19.813	19.527	25	**XXXXXXXXXXXXXXXXXXXXXXX
20.323	20.059	28	**XXXXXXXXXXXXXXXXXXXXXXXXXX
20.863	20.590	20	**XXXXXXXXXXXXXXXXXX
21.371	21.122	30	**XXXXXXXXXXXXXXXXXXXXXXXXXXXXX
21.876	21.653	19	**XXXXXXXXXXXXXXXXX
22.433	22.184	15	**XXXXXXXXXXXXXX
22.992	22.716	5	**XXXXX
23.553	23.247	4	**XXXX
24.084	23.779	6	**XXXXXX
24.384	24.310	2	**XX
24.842	24.842	1	**X

VARIABLE NUMERO 4 --- DESCRIPTION DE LA DISTRIBUTION EN 35 CLASSES

Valeur propre 4

MOYENNE DE LA CLASSE	LIMITE INFERIEURE DE LA CLASSE	EFFECTIF	
2.799	2.799	1	**X
3.350	3.218	3	**XXX
3.861	3.636	9	**XXXXXXXXX
4.258	4.055	14	**XXXXXXXXXXXXX
4.671	4.473	15	**XXXXXXXXXXXXXX
5.124	4.892	25	**XXXXXXXXXXXXXXXXXXXXXXX
5.508	5.310	40	**XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
5.932	5.729	36	**XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
6.379	6.147	36	**XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
6.766	6.566	47	**XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
7.188	6.984	61	**XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
7.589	7.403	58	**XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
8.014	7.821	59	**XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
8.420	8.240	56	**XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
8.883	8.658	59	**XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
9.275	9.077	64	**XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
9.682	9.495	58	**XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
10.124	9.914	53	**XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
10.511	10.332	43	**XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
10.916	10.751	45	**XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
11.374	11.169	39	**XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
11.794	11.588	28	**XXXXXXXXXXXXXXXXXXXXXXXX
12.185	12.007	30	**XXXXXXXXXXXXXXXXXXXXXXXXX
12.606	12.425	28	**XXXXXXXXXXXXXXXXXXXXXXXX
13.022	12.844	29	**XXXXXXXXXXXXXXXXXXXXXXXXX
13.450	13.262	12	**XXXXXXXXXXXX
13.854	13.681	11	**XXXXXXXXXXX
14.310	14.099	14	**XXXXXXXXXXXXX
14.695	14.518	10	**XXXXXXXXXX
15.169	14.936	6	**XXXXXX
15.565	15.355	4	**XXXX
15.888	15.773	1	**X
16.432	16.192	5	**XXXXX
0.0	16.610	0	**
17.029	17.029	1	**X

Lois jointes des cinq premières valeurs propres, des taux d'inertie correspondants, et de la trace.
(1000 réalisations pseudo-aléatoires)

MATRICE DE CORRELATION

1 - Tableau 8 x 8

	VP1	VP2	VP3	VP4	VP5	PC1	PC2	PC3	PC4	PC5	TRA
VP1	1.00	0.48	0.26	0.14	0.10	0.60	-0.23	-0.40	-0.34	-0.25	0.80
VP2	0.48	1.00	0.47	0.30	0.23	-0.23	0.61	-0.11	-0.15	-0.11	0.78
VP3	0.26	0.47	1.00	0.51	0.36	-0.43	-0.08	0.70	0.17	0.09	0.66
VP4	0.14	0.30	0.51	1.00	0.57	-0.46	-0.20	0.17	0.82	0.37	0.54
VP5	0.10	0.23	0.36	0.57	1.00	-0.40	-0.19	0.05	0.37	0.89	0.44
PC1	0.60	-0.23	-0.43	-0.46	-0.40	1.00	-0.40	-0.60	-0.56	-0.45	0.01
PC2	-0.23	0.61	-0.08	-0.20	-0.19	-0.40	1.00	-0.10	-0.25	-0.21	-0.01
PC3	-0.40	-0.11	0.70	0.17	0.05	-0.60	-0.10	1.00	0.24	0.08	-0.06
PC4	-0.34	-0.15	0.17	0.82	0.37	-0.56	-0.25	0.24	1.00	0.42	-0.01
PC5	-0.25	-0.11	0.09	0.37	0.89	-0.45	-0.21	0.08	0.42	1.00	0.03
TRA	0.80	0.78	0.66	0.54	0.44	0.01	-0.01	-0.06	-0.01	0.03	1.00

MATRICE DE CORRELATION

2 - Tableau 10 x 10

	VP1	VP2	VP3	VP4	VP5	PC1	PC2	PC3	PC4	PC5	TRA
VP1	1.00	0.51	0.31	0.23	0.18	0.65	-0.16	-0.35	-0.35	-0.33	0.78
VP2	0.51	1.00	0.55	0.38	0.31	-0.13	0.61	-0.02	-0.17	-0.19	0.79
VP3	0.31	0.55	1.00	0.61	0.41	-0.36	-0.02	0.69	0.17	-0.02	0.72
VP4	0.23	0.38	0.61	1.00	0.59	-0.39	-0.20	0.20	0.75	0.24	0.65
VP5	0.18	0.31	0.41	0.59	1.00	-0.37	-0.21	0.02	0.29	0.80	0.56
PC1	0.65	-0.13	-0.36	-0.39	-0.37	1.00	-0.25	-0.57	-0.55	-0.48	0.03
PC2	-0.16	0.61	-0.02	-0.20	-0.21	-0.25	1.00	-0.04	-0.27	-0.26	0.01
PC3	-0.35	-0.02	0.69	0.20	0.02	-0.57	-0.04	1.00	0.26	0.02	0.01
PC4	-0.35	-0.17	0.17	0.75	0.29	-0.55	-0.27	0.26	1.00	0.35	-0.00
PC5	-0.33	-0.19	-0.02	0.24	0.80	-0.48	-0.26	0.02	0.35	1.00	-0.03
TRA	0.78	0.79	0.72	0.65	0.56	0.03	0.01	0.01	-0.00	-0.03	1.00

Légende :

VPi = $i^{ème}$ valeur propre

PCi = $i^{ème}$ taux d'inertie

TRA = Trace

MATRICE DE CORRELATION

3 - Tableau 12 x 12

	VP1	VP2	VP3	VP4	VP5	PC1	PC2	PC3	PC4	PC5	TRA
VP1	1.00	0.56	0.41	0.35	0.30	0.68	-0.11	-0.26	-0.33	-0.31	0.79
VP2	0.56	1.00	0.61	0.44	0.34	-0.02	0.60	0.04	-0.20	-0.26	0.79
VP3	0.41	0.61	1.00	0.63	0.43	-0.21	0.00	0.66	0.08	-0.12	0.77
VP4	0.35	0.44	0.63	1.00	0.60	-0.25	-0.20	0.15	0.66	0.14	0.71
VP5	0.30	0.34	0.43	0.60	1.00	-0.24	-0.26	-0.06	0.18	0.72	0.63
PC1	0.68	-0.02	-0.21	-0.25	-0.24	1.00	-0.16	-0.45	-0.48	-0.40	0.11
PC2	-0.11	0.60	0.00	-0.20	-0.26	-0.16	1.00	0.02	-0.26	-0.33	-0.01
PC3	-0.26	0.04	0.66	0.15	-0.06	-0.45	0.02	1.00	0.19	-0.11	0.03
PC4	-0.33	-0.20	0.08	0.66	0.18	-0.48	-0.26	0.19	1.00	0.28	-0.05
PC5	-0.31	-0.26	-0.12	0.14	0.72	-0.40	-0.33	-0.11	0.28	1.00	-0.07
TRA	0.79	0.79	0.77	0.71	0.63	0.11	-0.01	0.03	-0.05	-0.07	1.00

Annexe II

On trouvera dans cette annexe un exemple (1) d'analyse des correspondances multiples auquel nous nous sommes référés plusieurs fois dans le cours du texte (notamment au chapitre II, §-3, à propos des problèmes de codage et de stabilité vis-à-vis des fluctuations d'échantillonnage, et au chapitre III).

UN EXEMPLE DE DEPOUILLEMENT PARTIEL D'ENQUETE.

Cette annexe comprend trois parties. Nous allons tout d'abord montrer comment l'analyse des correspondances multiples permet de dresser une *grille socio-administrative* prête à recevoir des informations relatives à certains thèmes de l'enquête, en nous appuyant sur un exemple. Puis, nous rappellerons comment la technique de projection de variables illustratives s'apparente à la théorie de la régression multiple, et nous montrerons comment elle permet de sélectionner certains tableaux. Enfin, nous donnerons un exemple d'illustration de la grille, et de sélection de tableaux. Bien entendu, dans le cadre de cet exposé méthodologique, nous raisonnerons en dimensions réduites (pour des raisons d'encombrement graphique, en particulier) ; la technique ne prend tout son sens que dans le cas d'une application à grande échelle.

---

(1) Pour un exposé plus complet de la méthode utilisée, on pourra se reporter à l'article :
"L'orientation du dépouillement de certaines enquêtes par l'analyse des correspondances multiples" - L. LEBART - Consommation n°2 - 1975 - Dunod.

1 - Un exemple de grille socio-administrative

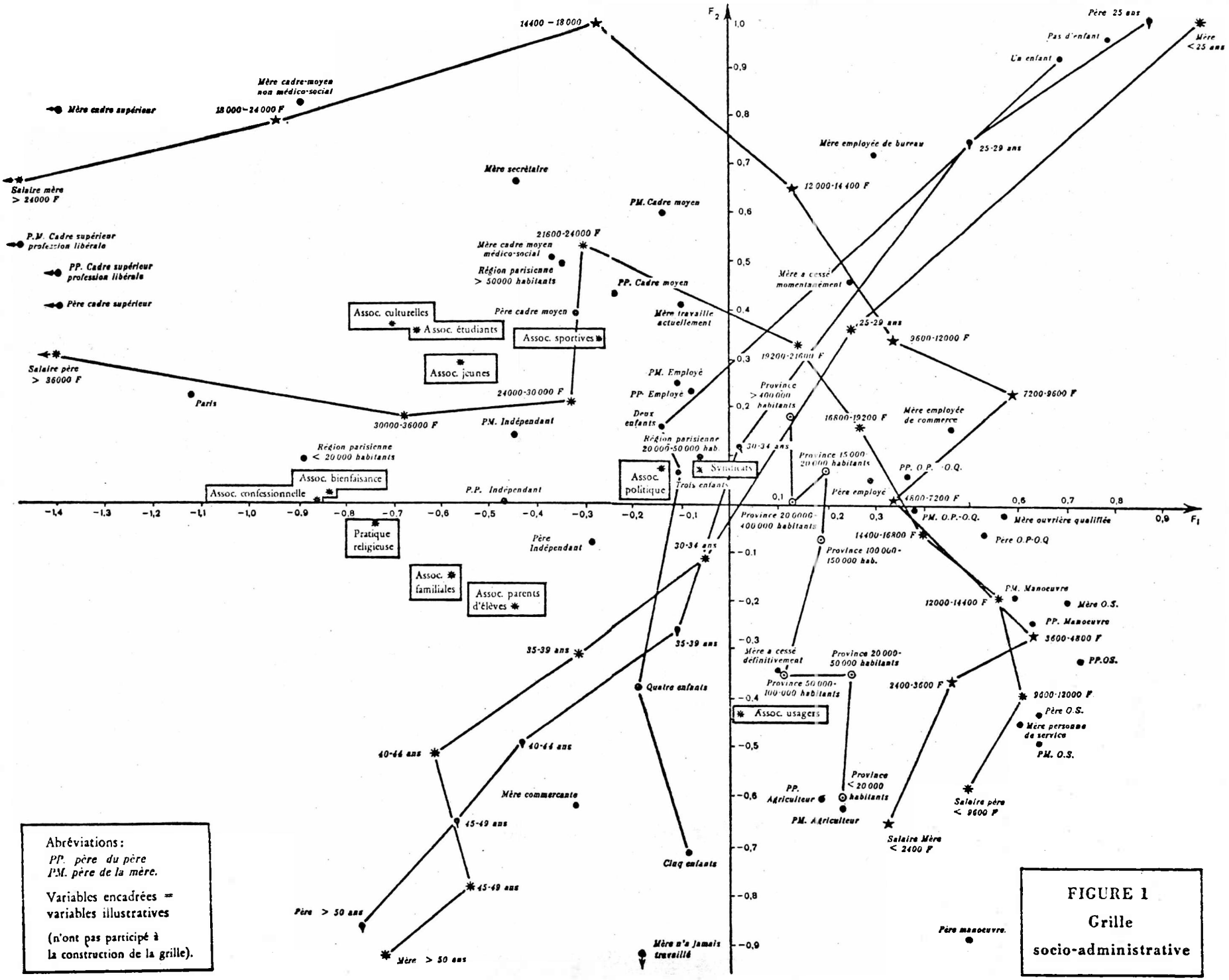
Nous prendrons l'exemple de l'enquête CNAF-CREDOC réalisée en France en 1971 auprès de 2003 familles dont le chef ou son conjoint sont salariés. Le plan de sondage assez particulier prévoyait une sur-représentation des familles nombreuses, des femmes exerçant une activité rémunérée, et prenait également en compte l'âge de l'aîné des enfants et certaines caractéristiques de la commune de résidence.

La figure 1 ci-après résume la première phase du dépouillement : l'établissement d'une grille (ou d'une carte) socio-administrative. Nous nous limitons à deux facteurs ici pour insister sur la partie de l'opération qui peut être visualisée. Mais la partie automatique du traitement, notamment la sélection des variables illustratives les plus importantes, peut évidemment se faire dans l'espace des **k** facteurs jugés significatifs après une procédure de simulation.

La figure 1 nous donne une typologie plane de 98 modalités de réponses relatives à 11 variables principales :

- profession du père
- profession de la mère
- salaire du père
- salaire de la mère (lorsqu'il existe)
- âge du père
- âge de la mère
- nombre d'enfants
- catégorie de communes
- profession des ascendants paternels et maternels (2 variables)
- activité de la mère

Il est clair qu'il existe une part d'arbitraire dans le choix des variables destinées à construire la grille ; il est d'ailleurs possible, si le questionnaire s'y prête, d'envisager l'élaboration de plusieurs types de grilles privilégiant chacune certains aspects des caractéristiques des familles. Il est cependant souhaitable d'obtenir une configuration qui soit la plus stable possible ; dans l'exemple que nous présentons, la grille n'a pas été modifiée de façon notable par la



Abréviations :
PP. père du père
PM. père de la mère.

Variables encadrées = variables illustratives

(n'ont pas participé à la construction de la grille).

FIGURE 1
Grille socio-administrative

suppression d'un tiers de l'échantillon. De plus, malgré la stratification extrêmement déformante de l'échantillon, le plan des deux premiers facteurs est sensiblement le même que l'analyse soit faite sur les données brutes ou sur les données redressées. Cette stabilité n'est pas très surprenante, car cette typologie décrit des associations et non des niveaux, qui, eux, sont très sensibles aux systèmes de pondérations. Non seulement les moments d'ordre 2 qui décrivent les associations sont beaucoup plus indépendants des systèmes de pondération que les moments d'ordre 1, mais de plus, le sous-espace engendré par les premières directions propres des matrices de moments d'ordre 2 est lui-même assez stable vis-à-vis des éventuelles fluctuations de ces moments. (Cf. Chapitre II, §-3).

Interprétation générale de la figure 1

Certaines des variables analysées ont des modalités ordonnées de façon naturelle (par exemple, la variable "âge de la mère" comprend huit modalités ; si l'on excepte la modalité "non-réponse, sans objet", il existe un ordre naturel des différentes classes d'âge). Sur le graphique, ces modalités ordonnées sont jointes par un trait, afin de permettre de suivre facilement l'évolution du phénomène continu sous-jacent.

Les variables encadrées n'ont pas participé à l'analyse. Pour l'interprétation de leur position, on se reportera au paragraphe 3.

Cet éparpillement de variables, apparemment assez confus, s'ordonne en réalité autour de deux grands thèmes : l'âge de la famille et son statut social.

Suivons en effet les diverses classes d'âge du père, depuis le haut droit du graphique jusqu'au bas gauche, suivant une diagonale assez rectiligne : le long de cet axe, on trouve également les classes d'âge croissantes de la mère, moins dispersées toutefois que celles de leur mari ; on trouve également le nombre d'enfants croissant progressivement, avec un léger infléchissement sur la droite, correspondant au fait que les familles les plus nombreuses se trouvent surtout dans les milieux modestes ; on trouve également, le long de ce même axe,

les différentes classes d'âge de l'aîné des enfants, également rangées par ordre croissant dans la même direction.

Assez perpendiculairement à cette direction se trouvent les lignes brisées joignant les différentes classes de salaire du père et de celui de la mère lorsqu'elle travaille. Ces deux lignes brisées sont repliées à leurs extrémités, et tournent leur concavité vers le haut. Les extrémités correspondent aux statuts sociaux précisément les plus extrêmes, ce repliement vers le haut nous montre que ce sont les familles plutôt âgées qui occupent les situations sociales les plus divergentes : aisance ou dénuement.

Le long de ces lignes brisées, les catégories socio-professionnelles du père et de la mère viennent illustrer et confirmer ce trajet le long de l'échelle sociale : aux manoeuvres, gens de maison et ouvriers de diverses catégories à droite s'opposent sur la gauche les professions libérales et les cadres supérieurs.

III-2 - Régression visualisée et variables illustratives

Sur la figure 1 auraient également pu figurer les 2003 familles (qui sont les lignes du tableau Z analysé). Avec ce type de codage, les relations dites de transition ont une interprétation simple : une famille est caractérisée par 11 réponses ; pour obtenir sa position, il suffit de prendre le centre de gravité des 11 points-réponses correspondants sur la figure 1, dont on dilatera les coordonnées (en faisant agir les coefficients "$1/\sqrt{\lambda}$" relatifs à chaque axe).

De façon analogue, projeter les réponses à une question supplémentaire revient à calculer les centres de gravité des familles concernées par chacune des réponses, et à effectuer sur ces centres la dilatation précédente.

Il est clair que cette procédure est apparentée à la régression multiple dont elle constitue une variante descriptive. On peut en effet considérer les réponses aux variables socio-administratives comme des variables exogènes, engendrant une certaine variété linéaire. Toutefois, au lieu de projeter les réponses supplémentaires

(variables endogènes) directement sur cette variété, on commencera par ajuster celle-ci par un sous-espace de plus faibles dimensions (deux pour la figure 1). De plus, au lieu de nous intéresser aux coefficients de régression (coordonnées des projections dans la base des variables socio-administratives), on observe simplement les positions relatives des différentes réponses.

La figure 1 constitue donc une trame prête à accueillir différents tissages selon les thèmes de l'enquête.

Précisons que, du point de vue des calculs, il est beaucoup moins onéreux de projeter les modalités de réponse à une question supplémentaire que de croiser cette question avec l'ensemble des variables socio-administratives. De plus, la lecture du graphique est beaucoup plus aisée et suggestive que celle des 11 tableaux croisés correspondants.

Enfin, il n'est pas nécessaire d'observer toutes les réponses illustratives, car celles-ci peuvent en effet être très nombreuses (plusieurs milliers dans le cas de l'enquête citée). Il est en effet possible de sélectionner automatiquement celles qui occupent les positions les plus significatives (sur les k premiers facteurs par exemple).

Supposons qu'une réponse concerne n familles. L'hypothèse nulle sera : n points sont pris au hasard parmi les 2003 points repérés par leurs coordonnées dans l'espace des k premiers facteurs. Il est facile, dans ces conditions, de construire des seuils de signification pour la distance de la réponse supplémentaire à l'origine. On peut alors, soit retenir les réponses correspondant à un certain seuil de signification, soit classer ces réponses par ordre de signification croissante.

Bien entendu, ce filtrage automatique étant assez peu coûteux, il est possible de croiser les questions supplémentaires, ce qui permet de détecter certains types d'interactions.

Une extension de la notion de "contribution absolue"

La variance de l'abscisse f_i d'une réponse supplémentaire sur un axe factoriel relatif à la valeur propre λ est : $1/n_i$, si n_i est

l'effectif concerné par la réponse : en effet, la variance de la population est λ, et la variance de la moyenne de n_i individus pris au hasard est donc λ/n_i. Comme l'abscisse sur l'axe est obtenue en multipliant la moyenne par $1/\sqrt{\lambda}$, on trouve bien le résultat annoncé.

Ainsi, les quantités $f_i\sqrt{n_i}$ sont, dans l'hypothèse nulle, des réalisations de variables aléatoires ayant même variance, et sont donc comparables entre elles du point de vue de leur signification. Ces mêmes quantités calculées pour les variables ayant participé à l'analyse sont proportionnelles aux racines carrées des classiques contributions absolues. Ce sont elles qui nous permettront de classer et de sélectionner les variables illustratives.

3 - Un exemple d'illustration de la grille.

Sur la figure 1, sont mises en évidence les positions des individus appartenant à des associations, appartenance mesurant ici le degré d'insertion sociale de la famille (variables encadrées). Il s'agit de variables *illustratives*, projetées après réalisation de la trame sous-jacente. Comme on le voit, à l'exception des associations syndicales, d'usagers et politiques, l'appartenance aux divers autres groupes ne semble pas indépendante du statut social tel que celui-ci est décrit par le premier facteur. Le test précédent nous montre que les positions de ces derniers groupes sont toutes significatives d'une répartition non aléatoire dans le plan.

Ainsi, pour prendre un exemple, les associations culturelles concernent 118 familles. L'écart-type de l'abscisse du point représentatif est donc de l'ordre de 0.09, alors que l'abscisse du point observé est de -0.70 ; celui-ci est donc à plus de 7 écart-types de l'origine qui est la moyenne théorique.

La figure 1 nous suggère alors de prendre comme variable explicative (explication au sens statistique) le salaire du père, qui sera par exemple divisé en deux classes, compte tenu des faibles effectifs correspondant à certaines associations. On obtient ainsi le tableau suivant (qui ne concerne que les familles dont le père est salarié, et dont le salaire a effectivement été déclaré) :

Tableau 1

Appartenance à des associations selon le salaire du père

Pourcentage* d'adhérents	Salaire annuel du père < 30000 F.	Salaire annuel du père ⩾ 30000 F.
Association familiale	4.3	7.9
Association parents d'élèves	28.8	51.1
Association syndicale	18.3	18.3
Association de bienfaisance	3.4	8.1
Association politique	2.8	3.7
Association confessionnelle	5.0	13.4
Association culturelle	5.8	13.1
Effectif de l'échantillon	1306	255

* Pourcentage après redressement, des familles où le père ou la mère appartiennent à une association.

1ex. nº 1

30 NOV 1982

6 JAN 1983