ANNALES DE L'INSEE. — N°s 22-23/1976

Opérateur associé à un tableau de données

par **Yves ESCOUFIER**

A un tableau de données du type individus-caractères, on associe un opérateur de l'espace des caractères en lui-même, ce qui permet de comparer quantitativement différents tableaux. On précise la signification mathématique de ces comparaisons dans différents contextes. On étudie sur un exemple la nature des réponses obtenues par ce procédé.

1 Opérateur associé à un vecteur aléatoire

Soit $(\Omega, \mathcal{A}, P)$ un espace de probabilité. Notons $L_2(\Omega, \mathcal{A}, P)$ l'ensemble des variables aléatoires centrées de variance finie sur $(\Omega, \mathcal{A}, P)$. Dans un travail précédent [2] l'auteur a proposé d'associer à tout vecteur aléatoire X, tel que $X' = (X_1, \ldots, X_2)$, dont les composantes appartiennent à $L_2(\Omega, \mathcal{A}, P)$, un opérateur $U_{\underset{\sim}{X}}$ de $L_2(\Omega, \mathcal{A}, P)$ en lui-même défini par :
pour tout $Y \in L_2(\Omega, \mathcal{A}, P)$

$$U_{\underset{\sim}{X}}(Y) = \sum_{i=1}^{k} E(X_i Y) X_i$$
$$= [E(X Y)]' \underset{\sim}{X}$$

L'intérêt de cette association réside dans les propriétés suivantes [2 et 3].

a. Les vecteurs propres de $U_{\underset{\sim}{X}}$ sont les composantes principales de X.

De manière plus précise si Σ est la matrice des variances et covariances de X, pour tout vecteur $Z \in \mathbf{R}^k$ tel que $\underset{\sim}{\Sigma} Z = \lambda Z$, la variable aléatoire $Y = Z'X$ vérifie $U_{\underset{\sim}{X}}(Y) = \lambda Y$.

Inversement, toute variable Y telle que $U_{\underset{\sim}{X}}(Y) = \lambda Y$ est de la forme $Y = Z'X$ avec $\underset{\sim}{\Sigma} Z = \lambda Z$.

Cette première propriété montre que $U_{\underset{\sim}{X}}$ est caractéristique de X au sens où les valeurs propres et les vecteurs propres de Σ le sont. C'est une caractérisation très forte puisqu'elle définit non seulement les directions des composantes principales mais aussi leurs variances.

b. L'opérateur $U_{\underset{\sim}{X}}$ appartient à la classe des opérateurs de Hilbert-Schmidt (*i. e.* la somme des carrés des valeurs propres est finie) et sur cette classe on peut définir un produit scalaire qui la munit d'une structure d'espace de Hilbert.

On sait que toutes les méthodes d'analyse multivariable reposent sur la structure d'espace de Hilbert de $L_2(\Omega, \mathcal{A}, P)$ pour le produit scalaire défini par la covariance. L'intérêt de la propriété ci-dessus est de permettre de retrouver cette structure mathématique pour des analyses *multi-vectorielles*. A la difficulté près de devoir substituer les opérateurs aux vecteurs, toutes les méthodes d'analyse multivariable sont donc transposables aux analyses multivectorielles.

c. $$< U_{\underset{\sim}{X_1}}, U_{\underset{\sim}{X_2}} > = \mathrm{Tr}(\underset{\sim}{\Sigma_{12}}\, \underset{\sim}{\Sigma_{21}})$$

où :

$$\underset{\sim}{\Sigma_{12}} = \underset{\sim}{\Sigma'_{21}} = E(\underset{\sim}{X_1}\, \underset{\sim}{X'_2})$$

Cette dernière propriété rend applicable l'approche envisagée puisque le produit scalaire entre deux opérateurs y apparaît comme une quantité facile à calculer.

2 Opérateur associé à un tableau de données

Reprenant ce problème dans le contexte de l'analyse des données, J. M. BRAUN [1] et J.P. PAGES [7] considèrent une matrice X, $p \times n$, de p caractères quantitatifs observés sur n individus. Des poids :

$$P_i \left(P_i > 0, \sum_{i=1}^{n} P_i = 1 \right)$$

étant affectés aux individus, la matrice X est supposée centrée. X_i^j est l'élément commun à la ligne i et à la colonne j de X.

Si $(\underline{e}_1, \ldots, \underline{e}_p)$ désigne la base canonique de $E = \mathbf{R}^p$, à l'individu i est associé le vecteur :

$$E : \underline{X}_i = \sum_{k=1}^{p} X_i^k \underline{e}_k$$

De la même manière, si $(\underline{f}_1, \ldots, \underline{f}_n)$ est la base canonique de $F = \mathbf{R}^n$, au caractère j est associé le vecteur de :

$$F : \underline{X}^j = \sum_{k=1}^{n} X_k^j \underline{f}_k$$

A l'objet i est également associé le vecteur $\underline{f}_i^*$ de F*, dual de F, défini par :

$$\underline{f}_i^* (\underline{f}_j) = \begin{cases} 1 & \text{si } j = i \\ 0 & \text{sinon} \end{cases}$$

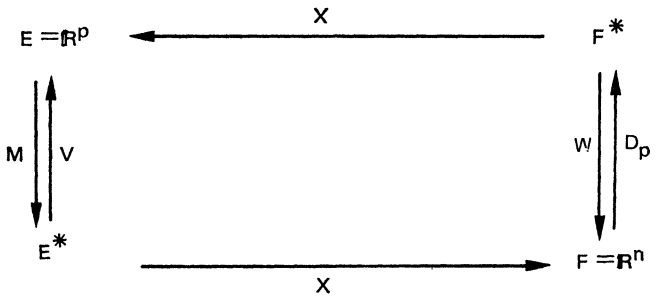
Il apparaît alors que la transformation linéaire de F* dans E qui à $\underline{f}_i^*$ fait correspondre $\underline{X}_i$ a pour matrice associée X lorsqu'on rapporte E à la base $(\underline{e}_1, \ldots, \underline{e}_p)$ et F* à la base $(\underline{f}_1^*, \ldots, \underline{f}_n^*)$.

De façon analogue, au caractère j est associé le vecteur $\underline{e}_j^*$ de E*, dual de E, défini par :

$$\underline{e}_j^* (\underline{e}_i) = \begin{cases} 1 & \text{si } i = j \\ 0 & \text{sinon} \end{cases}$$

et X′ est la matrice associée à la transformation linéaire qui fait correspondre $\underline{e}_j^*$ de E* à $\underline{X}^j$ de F lorsqu'on rapporte F à la base $(\underline{f}_1, \ldots, \underline{f}_n)$ et E* à la base $(\underline{e}^*_1, \ldots, \underline{e}^*_p)$.

Si on choisit pour mesurer la ressemblance des individus la métrique euclidienne M sur E et pour mesurer la ressemblance entre variables, la métrique euclidienne diagonale D_p sur F ($D_p(e_i, e_i) = P_i$), on est conduit au schéma de dualité :



$W = X' \circ M \circ X$ est la métrique euclidienne qui s'impose sur F* si on veut que les distances entre objets soient les mêmes calculées dans F* et dans E.

a. $U = W \circ D_p$ est un élément de L(F, F), ensemble des applications linéaires de F, espace des caractères, en lui-même. Par définition les composantes principales sont les caractères $\underline{C}^j$ qui vérifient :

$$U\,\underline{C}^j = \lambda_j\,\underline{C}^j$$

et :

$$D_p\,(\underline{C}^j, \underline{C}^j) = \lambda_j$$

L'application U est entièrement déterminée par la donnée du triplet (X, M, D_p). La manière dont W a été construit fait que U est aussi caractéristique du couple (D, D_p) où D est la matrice des distances entre objets.

b. U appartient à l'ensemble des applications linéaires, D_p symétriques, positives de F dans F, ensemble qui peut être muni d'un produit scalaire calculé à partir d'une base quelconque ($\xi_1, \ldots, \xi_n$) de F par :

$$< U_1, U_2 > = \sum_{i=1}^{n} < U_1(\xi_i), U_2(\xi_i) >^2$$

c. Soient (X_1, M_1, D_p) et (X_2, M_2, D_p) les triplets définissant deux études faites sur les mêmes individus. Posons :

$$V_{12} = X_1 \circ D_p \circ X_2' = V_{21}'$$

On montre que [7] :

$$\begin{aligned} < U_1, U_2 > &= \mathrm{Tr}\,(V_{21} \circ M_1 \circ V_{12} \circ M_2) \\ &= \mathrm{Tr}\,(X_2 \circ D_p \circ X_1' \circ M_1 \circ X_1 \circ D_p \circ X_2' \circ M_2) \\ &= \mathrm{Tr}\,(X_1' \circ M_1 \circ X_1 \circ D_p \circ X_2' \circ M_2 \circ X_2 \circ D_p) \\ &= \mathrm{Tr}\,(U_1 \circ U_2) \end{aligned}$$

6 671646 6 42 6 A

La dernière égalité présente l'intérêt de permettre le calcul du produit scalaire entre opérateurs (et donc de la distance entre opérateurs) à partir des opérateurs eux-mêmes. Ceci est très utile pour toutes les applications dans lesquelles les données initiales sont constituées par une matrice de similarité : en effet, une matrice de similarité, sous réserve d'être semi-définie positive, peut jouer le rôle de W; une matrice de dissimilarité peut, par la méthode de TORGERSON [8], fournir une matrice de similarité qui jouera le rôle de W sous la réserve d'être semi-définie positive. Ces questions ont été discutées en [5].

3 Application à la comparaison de deux études portant sur les mêmes individus

Nous disposons de deux études représentées par les triplets (X_1, M_1, D_p) et (X_2, M_2, D_p) et donc de leurs opérateurs associés U_1 et U_2.

a. Pour $k = 1, \ldots, p_1$, appelons $\underline{V}^k$ et λ_k les vecteurs et valeurs propres de U_1. Pour $r = 1, \ldots, p_2$, S^r et μ_r sont les vecteurs et valeurs propres de U_2. Un calcul simple [1] montre que :

$$d^2(U_1, U_2) = \sum_{k=1}^{p_1} \lambda_k^2 + \sum_{r=1}^{p_2} \mu_r^2 - 2 \sum_{k=1}^{p_1} \sum_{r=1}^{p_2} \lambda_k \mu_r D_p^2 (\underline{V}^k, \underline{S}^r)$$

Il s'ensuit que deux opérateurs qui ont les mêmes vecteurs propres sont confondus. En particulier si C est la matrice $p \times n$ des composantes principales d'une étude (X, M, D_p) les opérateurs associés à (X, M, D_p) et à (C, I, D_p) sont confondus.

Inversement, si deux opérateurs sont à distances nulles, leurs vecteurs propres sont les mêmes dans la mesure où toutes les valeurs propres sont différentes pour une étude. Si ce n'est pas le cas, les vecteurs propres des opérateurs associés à des valeurs propres égales ne sont définis qu'à une rotation près.

Il découle de la définition des métriques W_1 et W_2 que deux études ont des opérateurs associés confondus si et seulement si les distances entre individus fournies par les deux études sont les mêmes.

b. Supposons que les n individus soient répartis en k classes.

On sait que pour réaliser une analyse discriminante on effectue en fait l'analyse en composantes principales des points moyens des classes affectées du poids des individus de leur classe, $E = \mathbf{R}^p$ étant muni de la métrique $M = D^{-1}$ où D est la métrique des variances et covariances à l'intérieur des classes.

Ce qui précède peut être appliqué à la comparaison de deux analyses discriminantes portant sur les mêmes individus répartis dans les mêmes classes. Les analyses seront équivalentes dès lors qu'elles définissent les mêmes distances entre les points moyens des classes.

c. Si V est régulière, on peut prendre dans E la métrique V^{-1}. Alors U est idempotent et D_p symétrique; c'est l'opérateur de projection D_p orthogonale sur le sous-espace de $\mathbf{R}^n$ engendré par les caractères $\underline{X^j}$.

Pour calculer les distances entre les opérateurs associés à deux études :

$$(X_1, V_1^{-1}, D_p) \quad \text{et} \quad (X_2, V_2^{-1}, D_p)$$

on peut prendre pour bases les variables canoniques. L'idempotence des opérateurs donne alors :

$$d^2(U_1, U_2) = \text{rang}(U_1) + \text{rang}(U_2) - 2\sum_{i=1}^{\nu} r_i^2$$

où $\nu = \min[\text{rang}(U_1), \text{rang}(U_2)]$ et r_i est le $i^{\text{ème}}$ coefficient de corrélation canonique.

Cette distance correspond bien à l'intuition : deux ensembles de caractères sont équivalents du point de vue de la projection D_p orthogonale dans la mesure où ils engendrent les mêmes sous-espaces de $\mathbf{R}^n$.

d. Soit D une matrice de distances entre objets résultant d'une classification hiérarchique ou d'une partition. On sait que l'application de la méthode de TORGERSON [8] à une telle matrice fournit une métrique semi-définie positive sur $\mathbf{R}^n$. On peut donc envisager de comparer deux classifications hiérarchiques, ou deux partitions, ou une classification hiérarchique et une matrice de distances initiales à partir de la quantité

$$d^2(U_1, U_2)$$

Des travaux sont en cours pour comparer cette approche aux approches habituelles.

e. Tous les résultats de ce troisième paragraphe peuvent être utilisés pour répondre à la question suivante : quelle métrique M_2 faut-il choisir pour une étude portant sur un tableau X_2, afin que les distances entre objets pour l'étude (X_2, M_2, D_p) soient les plus proches possibles de celles fournies par une étude connue sur les mêmes objets (X_1, M_1, D_p)?

Dans le cas où M_2 est diagonale, ce problème a été résolu [4]. La solution unique est obtenue par la résolution d'un problème de programmation quadratique convexe.

Une extension de ce problème est en cours qui s'intéresse au problème suivant : quel est le triplet (X_2, M_2, D_p) où X_2 est une sous-matrice, $k \times n$, de X_1 et M_2 une métrique diagonale qui définit entre les objets les distances les plus voisines de celles fournies par l'étude de (X_1, M_1, D_p)? La solution est obtenue par une méthode pas à pas qui ne fournit qu'un optimum local.

4 Application à l'analyse d'un tableau de données à trois dimensions : un exemple

Les données considérées proviennent du laboratoire d'hydrométéorologie de l'Université des Sciences et Techniques du Languedoc (Montpellier). Elles concernent les hauteurs de pluies annuelles observées en 120 stations réparties dans la région indiquée par la figure 1. La région est découpée en 36 zones correspondant aux départements français. On notera que pour la clarté des figures, ont été regroupés sous le nom de Région de Paris (Code 75), les sept départements suivants :

75 Ville de Paris;
78 Yvelines;
91 Essonne;
92 Hauts-de-Seine;
93 Seine-Saint-Denis;
94 Val-de-Marne;
95 Val-d'Oise.

FIGURE 1

La région intéressée par l'étude

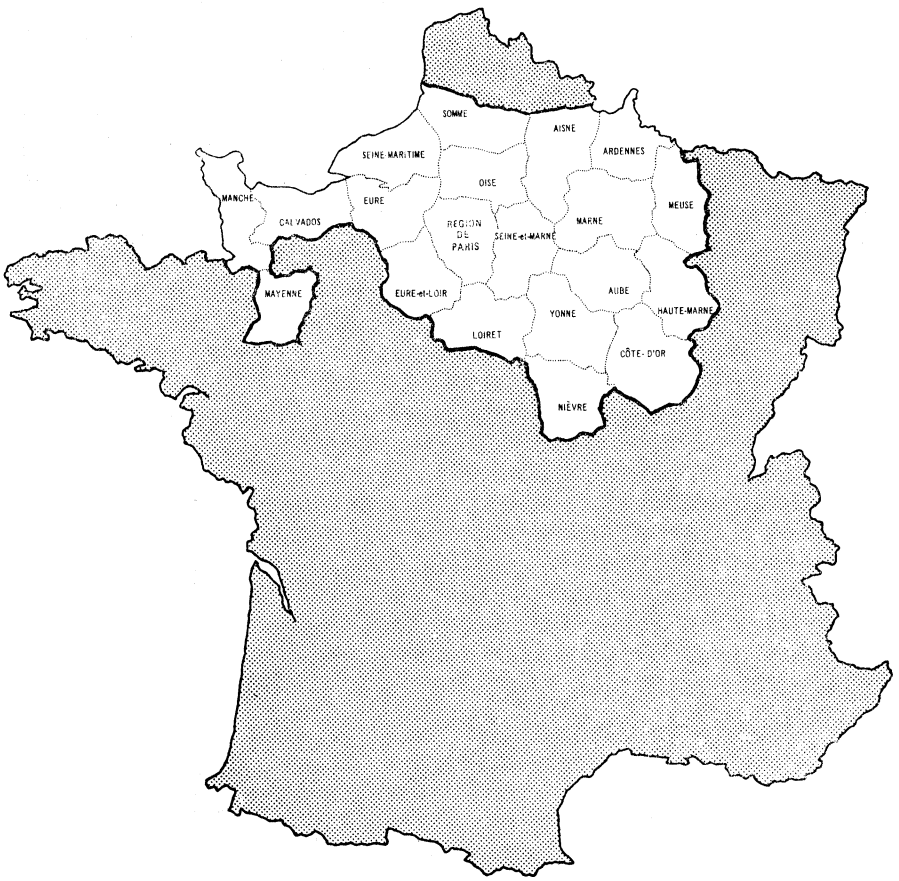


TABLEAU 1

Nombre de stations et hauteur moyenne des pluies par département

Département		Nombre de stations	Hauteur moyenne (cm)
Nom	Code		
Aisne	2	5	88,76
Ardennes	8	3	88,28
Aube	10	3	71,46
Calvados	14	6	90,16
Côte-d'Or	21	11	81,37
Eure	27	1	70,08
Eure-et-Loir	28	6	58,63
Loiret	45	3	64,35
Manche	50	1	103,73
Marne	51	12	69,51
Haute-Marne	52	7	90,40
Mayenne	53	1	110,95
Meuse	55	4	93,73
Nièvre	58	6	117,33
Oise	60	9	65,66
Ville de Paris	75	11	50,90
Seine-Maritime	76	3	93,24
Seine-et-Marne	77	10	61,64
Yvelines	78	1	51,90
Somme	80	1	53,22
Yonne	89	8	78,00
Essonne	91	2	60,61
Hauts-de-Seine	92	1	47,81
Seine-Saint-Denis	93	1	57,31
Val-de-Marne	94	3	51,99
Val-d'Oise	95	1	64,14

Ce regroupement n'est fait qu'au niveau de la représentation graphique; les départements sont isolés au niveau des calculs.

Le tableau 1 donne pour chaque département son nom, son code administratif, le nombre de stations d'observation qu'il contient et la hauteur annuelle moyenne des pluies observées dans les différentes stations qu'il contient pour les vingt dernières années.

Le premier travail consiste à calculer la matrice 120×120 des variances et covariances pour les 120 stations. En fait, l'étude utilise la matrice des

TABLEAU 2

Matrice départementale

0,903																										
0,868	0,956																									
0,635	0,690	0,819																								
0,373	0,629	0,596	0,834																							
0,680	0,549	0,578	0,264	0,820																						
0,612	0,781	0,812	0,686	0,521	1,000																					
0,520	0,524	0,561	0,543	0,395	0,669	0,847																				
0,776	0,729	0,676	0,467	0,714	0,676	0,669	0,950																			
0,325	0,349	0,416	0,419	0,251	0,489	0,618	0,403	1,000																		
0,744	0,818	0,792	0,508	0,621	0,720	0,594	0,757	0,409	0,830																	
0,739	0,687	0,703	0,347	0,751	0,616	0,474	0,748	0,310	0,740	0,821																
0,286	0,442	0,450	0,447	0,231	0,375	0,422	0,372	0,705	0,432	0,337	1,000															
0,674	0,734	0,749	0,454	0,574	0,738	0,561	0,699	0,370	0,750	0,713	0,355	0,799														
0,674	0,632	0,651	0,333	0,793	0,551	0,421	0,757	0,271	0,665	0,762	0,325	0,615	0,860													
0,595	0,846	0,801	0,646	0,485	0,807	0,616	0,720	0,448	0,732	0,583	0,499	0,658	0,569	0,902												
0,642	0,681	0,653	0,559	0,467	0,728	0,770	0,756	0,561	0,712	0,559	0,458	0,648	0,513	0,765	0,931											
0,429	0,716	0,679	0,735	0,305	0,711	0,509	0,492	0,285	0,584	0,423	0,358	0,533	0,370	0,708	0,576	0,810										
0,751	0,706	0,702	0,466	0,626	0,678	0,692	0,818	0,514	0,771	0,698	0,422	0,697	0,646	0,704	0,805	0,498	0,859									
0,638	0,662	0,651	0,535	0,484	0,785	0,752	0,723	0,589	0,699	0,532	0,399	0,621	0,517	0,767	0,887	0,542	0,769	1,000								
0,597	0,854	0,757	0,665	0,425	0,717	0,510	0,622	0,277	0,715	0,580	0,414	0,625	0,525	0,772	0,659	0,744	0,610	0,608	1,000							
0,729	0,660	0,649	0,391	0,737	0,616	0,526	0,805	0,342	0,699	0,741	0,341	0,611	0,760	0,624	0,614	0,424	0,725	0,611	0,564	0,766						
0,748	0,738	0,738	0,511	0,618	0,757	0,746	0,832	0,479	0,788	0,713	0,381	0,717	0,648	0,734	0,849	0,561	0,857	0,808	0,725	0,722	0,943					
0,537	0,592	0,581	0,545	0,437	0,778	0,777	0,723	0,534	0,604	0,464	0,338	0,558	0,465	0,732	0,831	0,514	0,693	0,919	0,574	0,570	0,772	1,000				
0,680	0,773	0,740	0,625	0,473	0,796	0,762	0,760	0,575	0,779	0,604	0,495	0,728	0,537	0,806	0,938	0,648	0,832	0,883	0,744	0,631	0,876	0,802	1,000			
0,705	0,731	0,732	0,531	0,530	0,754	0,750	0,756	0,529	0,774	0,653	0,417	0,707	0,558	0,739	0,872	0,583	0,842	0,815	0,723	0,654	0,913	0,757	0,919	0,938		
0,494	0,719	0,747	0,631	0,428	0,784	0,701	0,675	0,619	0,635	0,507	0,514	0,616	0,505	0,830	0,763	0,611	0,676	0,788	0,647	0,561	0,726	0,793	0,777	0,712	1,000	

corrélations, ce qui introduit un changement d'échelle dans les observations, mais ne change rien aux interprétations que nous pourrons faire.

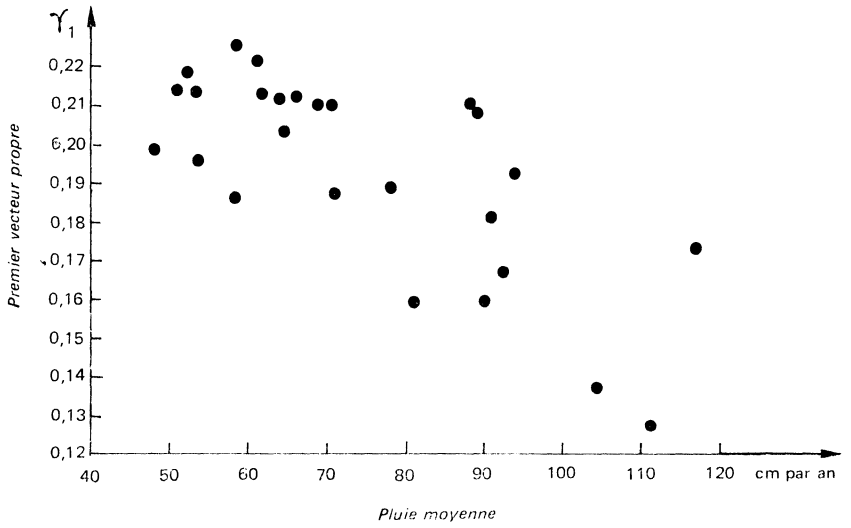
La seconde étape est constituée par le calcul de la matrice 26 × 26 dont l'élément (i, j) est la quantité $< U_i, U_j >$ divisée par $k_i \times k_j$ où k_i est le nombre de stations dans le département i. Cette matrice est donnée par le tableau 2.

Le tableau 3 donne les quatre premières valeurs propres de cette matrice et les vecteurs propres qui leur sont associés. On remarquera que le rapport entre la somme de ces quatre valeurs propres et la trace de la matrice est assez grand pour qu'on puisse se limiter à une étude sur ces quatre vecteurs.

Ainsi qu'il est fréquent dans ce genre d'étude, tous les départements ont à peu près la même coordonnée sur le premier vecteur propre qui peut être regardé comme déterminant la valeur moyenne des éléments de la matrice (composante de taille). La figure 2 montre curieusement que ces coordonnées sont d'autant plus faibles que la hauteur moyenne des pluies du département auquel elles correspondent est élevée.

FIGURE 2

Représentation des différents départements



Les figures 3, 4 et 5 ont été obtenues grâce à deux simplifications. D'une part nous avons donné à la zone « Région de Paris » la valeur moyenne des 7 départements qui la composent, d'autre part, pour permettre une lecture plus aisée, nous avons simplement distingué entre départements à coordonnées positives et départements à coordonnées négatives. Il est tentant de dire que le second vecteur oppose les départements soumis à des pluies continentales à ceux qui sont soumis à des pluies océaniques (fig. 3) tandis que le troisième résume l'influence des pluies venues du Nord (fig. 4). La signification du quatrième vecteur s'éclaire si l'on superpose la carte qu'il donne (fig. 5) à la carte des précipitations du mois de juillet : alors ce vecteur oppose assez bien les départements les plus pluvieux aux moins pluvieux.

Il n'est pas dans notre propos d'entrer plus avant dans une interprétation climatologique. Nous avons simplement voulu montrer que notre démarche des paragraphes précédents permet d'étendre sans difficulté une technique classique d'analyse multivariable à des données ayant une structure de famille de variables vectorielles.

TABLEAU 3

Valeurs propres de la matrice du tableau 2 et vecteurs propres associés

	1re valeur	2e valeur	3e valeur	4e valeur
Valeur propre	16,836 8	1,965 2	1,333 0	1,029 1
Pourcentage	0,713 5	0,083 2	0,056 4	0,043 6
Cumulé	0,713 5	0,796 7	0,853 1	0,896 7
Vecteur propre / Composante	Y_1	Y_2	Y_3	Y_4
Aisne	0,209 3	0,046 9	0,230 0	0,117 5
Ardennes	0,212 6	0,055 2	0,304 1	0,068 3
Aube	0,189 7	0,249 0	— 0,072 6	— 0,019 1
Calvados	0,160 1	— 0,246 4	0,305 5	0,058 5
Côte-d'Or	0,161 2	0,374 4	— 0,114 6	0,113 8
Eure	0,213 4	— 0,090 5	0,153 8	— 0,091 4
Eure-et-Loir	0,187 0	— 0,186 7	— 0,218 1	— 0,109 5
Loiret	0,212 8	0,176 9	— 0,136 1	— 0,039 5
Manche	0,138 6	— 0,330 7	— 0,369 5	0,411 2
Marne	0,210 3	0,118 7	0,060 4	0,053 7
Haute-Marne	0,185 5	0,310 8	— 0,016 6	0,139 4
Mayenne	0,127 2	— 0,247 9	— 0,080 4	0,699 3
Meuse	0,194 7	0,121 6	0,059 6	0,024 8
Nièvre	0,174 7	0,339 3	— 0,045 9	0,169 5
Oise	0,212 7	— 0,095 8	0,173 7	0,004 1
Ville de Paris	0,214 7	— 0,138 8	— 0,155 3	— 0,169 3
Seine-Maritime	0,168 5	— 0,151 4	0,402 7	— 0,033 8
Seine-et-Marne	0,212 7	0,072 9	— 0,170 0	— 0,030 2
Yvelines	0,212 3	— 0,151 5	— 0,193 6	— 0,226 0
Somme	0,196 1	— 0,027 7	0,383 4	— 0,019 8
Yonne	0,189 1	0,248 5	— 0,086 8	0,065 3
Essonne	0,222 9	0,047 0	— 0,115 4	— 0,141 0
Hauts-de-Seine	0,199 5	— 0,188 9	— 0,193 0	— 0,304 3
Seine-Saint-Denis	0,226 2	— 0,133 7	— 0,056 3	— 0,121 0
Val-de-Marne	0,219 6	— 0,040 2	— 0,097 2	— 0,138 3
Val-d'Oise	0,204 8	— 0,215 6	0,004 9	0,022 6

Trace de $\Sigma = 23{,}5971$.

FIGURE 3. ***Second vecteur propre***

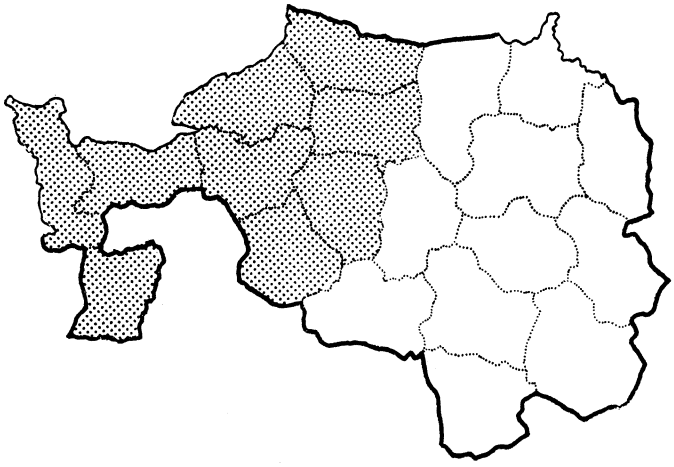


FIGURE 4. ***Troisième vecteur propre***

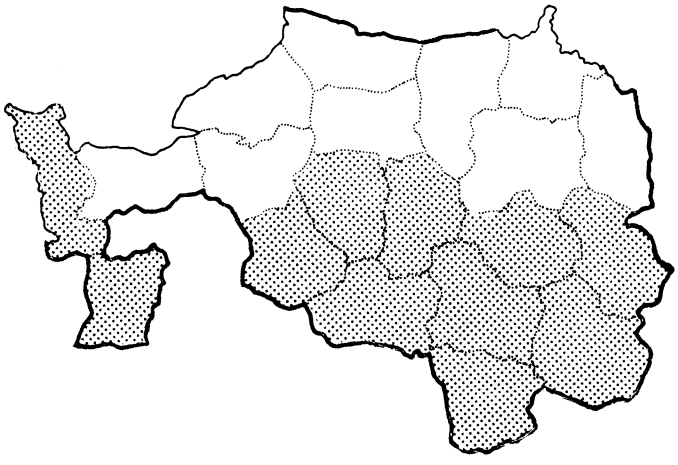
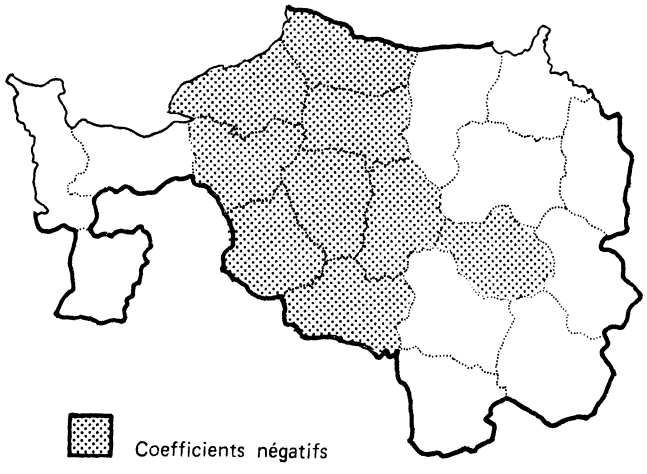


FIGURE 5. ***Quatrième vecteur propre***



● Bibliographie

[1] Braun J.M. — « Séries chronologiques multiples, recherche d'indicateurs », *Revue de statistique appliquée,* vol. XXI, n° 1, 1973, p. 81 à 106.

[2] Escoufier Y. — « Échantillonnage dans une population de variables aléatoires réelles », *Publications de l'Institut universitaire de Paris,* n° 19, fasc. 4, 1970, p. 1 à 47.

[3] Escoufier Y. — « Le traitement des variables vectorielles », *Biometrics,* n° 29, 1973, p. 751 à 760.

[4] Escoufier Y., Robert P., Cambon Y. — *Construction of a Vector Equivalent to a Given Vector from the Point of View of the Analysis of Principal Components* in Compstat 1974, Physica Verlag, Gerhart Bruckmann, Wien, 1974, p. 155 à 164.

[5] Escoufier Y. — « Le positionnement multidimensionnel », *Revue de statistique appliquée,* vol. XXIII, n° 4, 1975, p. 5 à 14.

[6] Holman E.W. — « The Relation between Hierarchical and Euclidean Models for Psycological Distances », *Psychometrika,* vol. XXXVII, n° 4, 1972, p. 417 à 423.

[7] Pages J.P. — *A propos des opérateurs d'Y. Escoufier,* Séminaire IRIA, 1974.

[8] Torgerson W.S. — *Theory and Methods of Scaling,* Wiley and Sons, New York, 1958.

Summary

An Operator Associated with Cross-Classification Tables

by **Yves ESCOUFIER**

An operator mapping the characters space into itself is associated with tables where individual and characters are cross-classified, which provides a quantitative basis for comparing different tables. The mathematical interpretation of these comparisons is given, and the whole method is illustrated by an example.

Reseña

Operador agregado a un cuadro de datos

por **Yves ESCOUFIER**

Un operador del espacio de los carácteres en si-mismo vá agregado a un cuadro de datos de tipo indivíduos-carácteres. Tal agregación facilita elementos cuantitativos para equiparar varios estudios. El significado matemático de esas equiparaciones está determinado en diversos contextos. Un ejemplo permite estudiar la naturaleza de las respuestas obtenidas mediante ese procedimiento.