

PRESENTATION DE LA METHODE STATIS PERMETTANT L'ANALYSE CONJOINTE DE PLUSIEURS TABLEAUX DE DONNEES QUANTITATIVES

Christine LAVIT*

RESUME

Supposons qu'on veuille analyser des données se présentant sous la forme de plusieurs tableaux de mesures recueillies sur les mêmes individus dans des situations différentes. La méthode STATIS permet de résumer l'information contenue dans ces données.

Cette méthode, basée sur le principe de l'Analyse en Composantes Principales, répond aux objectifs suivants :

- déceler quels sont les tableaux qui se ressemblent,*
- fournir un tableau «résumé» de l'ensemble,*
- décrire les différences entre tableaux par rapport à ce tableau résumé : sont-elles dues aux individus ou aux variables ?*

Pour mettre en œuvre cette méthode, nous proposons un logiciel d'utilisation simple, exécutable sur micro-ordinateur du type IBM PC, dont il existe également une version pour gros système.

MOTS-CLES

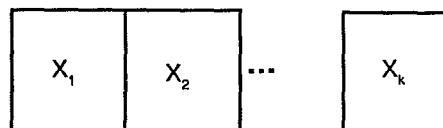
Méthode - Analyse de Données - Traitement des données - Logiciel

INTRODUCTION

Pour explorer des données qui se présentent sous la forme d'un tableau contenant les valeurs de caractères (ou variables) prises par un ensemble d'individus, le chercheur dispose de techniques d'analyse de données, comme l'Analyse en Composantes Principales (ACP), développées il y a une quinzaine d'années, et largement utilisées depuis.

Parfois le protocole de l'expérimentation conduit à considérer que les données recueillies forment, en fait, un ensemble de tableaux (une succession de tableaux lorsqu'ils sont mesurés à des dates différentes, par exemple).

Si ce sont les mêmes variables (en colonnes) qui sont mesurées sur les mêmes individus (correspondant aux lignes), on peut juxtaposer tableaux X_1, \dots, X_k pour obtenir le tableau suivant :



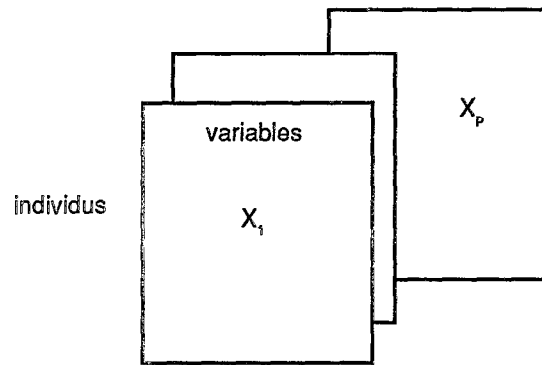
Mais l'ACP de ce grand tableau n'est pas satisfaisante car les variables sont le plus souvent fortement corrélées d'un tableau à l'autre, et on ne peut pas suivre l'évolution des individus. On peut également empiler les tableaux pour en faire l'ACP, mais alors les variables n'interviennent que par leurs corrélations «intra».

* Unité de Biométrie, ENSA.M — INRA — USTL

Différentes techniques d'analyse conjointe de tableaux de données sont actuellement proposées. Certaines cherchent à reconstruire le «cube» de données à partir d'un petit nombre d'individus types, de variables latentes et de conditions prototypes. D'autres supposent l'existence d'un modèle sous-jacent, ou d'une structure commune aux tableaux. D'autres enfin tiennent compte de la structure ordonnée du temps pour analyser une succession de tableaux. Peu de programmes sont diffusés à l'heure actuelle : INDSCAL, CANDECOMP - PARAFAC, TUCKER... dans les logiciels de langue anglaise (consulter la monographie de KROONENBERG à ce sujet) ; l'Analyse Factorielle Multiple, l'Analyse Canonique Généralisée, STATIS... dans les logiciels français.

I — DESCRIPTION RAPIDE DE LA METHODE STATIS

La méthode STATIS est une méthode exploratoire d'Analyse de Données, qui s'applique à des données quantitatives : p tableaux de mesures X_k ont été recueillies en différentes occasions sur les mêmes individus.



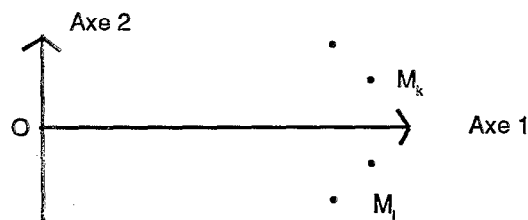
L'idée essentielle de la méthode est la recherche d'une structure commune aux tableaux, qu'on appelle infrastructure. Pour le tableau X_k , cette structure est décrite par les distances mutuelles entre individus, déduites du tableau de produits scalaires $W_k = X_k X_k'$.

1. Analyse globale des relations entre tableaux.

analyse globale des
ressemblances entre
tableaux

On compare les tableaux au moyen des «objets» W_k . Par opposition au terme infrastructure qui décrit la structure des individus à l'intérieur d'un tableau, on appelle interstructure les relations entre tableaux, décrites par les distances entre W_k . Ces distances sont déduites du produit scalaire de Hilbert-Schmidt entre applications linéaires.

A partir de ces produits scalaires, on construit une image euclidienne plane des tableaux. Soit M_1, \dots, M_p le nuage des points-tableaux :



Le cosinus de l'angle entre les vecteurs OM_k et OM_1 est l'approximation du produit scalaire normé entre W_k et W_1 appelé coefficient RV. Un coefficient RV proche de 1 signifie qu'on a la même structure des individus à l'intérieur des tableaux X_k et X_1 et que les positions mutuelles des individus sont stables.

2. Positions compromis des individus.

on obtient un graphique des individus tels qu'ils sont décrits par l'ensemble des tableaux,

A partir de l'image euclidienne des tableaux, on construit un objet compromis W qui peut être considéré comme un tableau de produits scalaires entre individus.

L'image euclidienne des individus, associée à ces produits scalaires, représente les positions mutuelles compromis des individus tels qu'ils sont décrits par l'ensemble des tableaux.

Lorsque les distances entre objets W_k déterminées dans l'interstructure, sont faibles, on peut affirmer qu'il existe bien une structure des individus, commune aux tableaux. Cette structure est alors décrite par les distances compromis entre individus.

qu'on interprète à l'aide des variables de chaque tableau.

Corrélations des variables avec les axes du compromis. On peut considérer que les coordonnées des individus sur un axe sont les valeurs d'une variable fictive, appelée «composante principale». Pour interpréter les positions des individus le long de l'axe, on calcule les corrélations de la composante principale avec les variables des différents tableaux, ou avec des variables exogènes dont on connaîtrait les valeurs sur les individus.

3. Trajectoires des individus.

Dans ce graphique, on trace la trajectoire de chaque individu tel qu'il est décrit par les différents tableaux.

Dans l'image euclidienne compromis des individus, on trace la trajectoire de chaque individu en utilisant la technique des points supplémentaires. L'interstructure a mis en évidence, sans les expliquer, les écarts entre tableaux. Les trajectoires permettent de déceler quels sont les individus responsables de ces écarts.

Lorsque, sur le graphique des corrélations des variables avec les axes du compromis, les points se regroupent nettement par variable, on peut donner un nom aux axes et interpréter le sens de parcours des trajectoires. Ce cas est fréquent, car les variables sont souvent fortement auto-corrélées dans les différentes études.

A Noter :

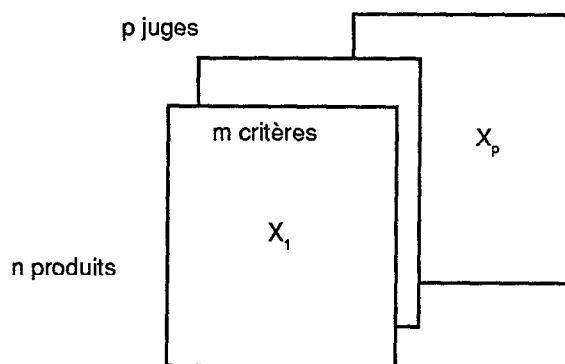
Pour une analyse d'un phénomène évolutif, la méthode donnerait les mêmes résultats si on intervertissait l'ordre des tableaux. Par conséquent, lorsque les tableaux sont indicés par le temps, la structure ordonnée du temps n'intervient qu'implicitement dans l'interprétation des trajectoires.

II — UTILISATIONS POSSIBLES

La méthode STATIS est basée sur le principe de l'Analyse en Composantes Principales, et sur les notions de moyenne et de corrélations entre variables : les ressemblances entre tableaux s'interprètent en termes de corrélations, les trajectoires décrivent l'évolution de chaque individu par rapport à l'évolution de l'individu moyen. Il faut donc que ces notions aient un sens.

La méthode a été conçue pour des données quantitatives, mais on peut l'appliquer à des données «qualitatives en échelle», l'interprétation se fait alors en termes de corrélations de rang.

1. Premier exemple : chaque tableau correspond à l'opinion d'un juge



L'interstructure permet de repérer quels sont les juges dont les réponses sont voisines, et quels sont ceux dont les opinions divergent. Le compromis fournit une répartition des produits qui résume l'opinion «majoritaire» qui ressort de l'enquête. Enfin, l'examen des trajectoires, qu'il vaudrait mieux dans ce cas nommer «écarts» entre les juges, donne une idée des produits qui ont fait l'unanimité, et de ceux pour lesquels il y a désaccord entre les juges.

2. Deuxième exemple : Les tableaux de mesures décrivent la composition du latex lorsqu'il est prélevé à différentes hauteurs du tronc.

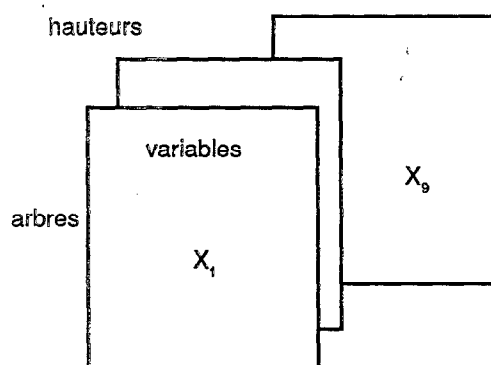
optimiser la production de latex

Objectif de l'étude. Le caoutchouc naturel est extrait du latex qui est un cytoplasme cellulaire situé en particulier dans l'écorce de l'Hévéa. Le latex est prélevé par saignée, c'est à dire par écoulement le long d'une encoche découpée dans l'écorce de l'arbre. Une bonne compréhension des mécanismes biologiques permet d'optimiser la production sans épuiser les réserves de l'arbre, ce qui pourrait nuire à sa rentabilité future. On sait par exemple, que la fréquence de saignée et la stimulation de l'arbre par application de produits chimiques sur l'encoche, influencent la quantité de latex produite.

Description des données. L'expérimentation que nous allons décrire a pour but d'étudier l'évolution de la composition du latex lorsqu'il est prélevé à différentes hauteurs sur le tronc de l'arbre, et plus particulièrement l'influence de la greffe de couronne sur cette évolution.

conduit à étudier l'influence de la greffe de couronne et de la hauteur de prélèvement sur sa composition.

On a donc considéré un lot d'arbres témoins et un lot d'arbres dont la couronne foliaire résulte d'une greffe. Le latex a été prélevé à neuf hauteurs différentes s'échelonnant le long du tronc. Le premier prélèvement est effectué sous l'encoche de saignée, tandis que les trois derniers sont situés, dans le cas des arbres greffés, au dessus de la greffe. Sur le latex récolté, on a mesuré les taux de magnésium, de phosphore, de thiols, de glucides et d'extrait sec. Les valeurs de ces variables reflètent l'état physiologique de l'arbre.



résultats globaux

Résultats. Lorsqu'on applique la méthode STATIS à ces données, un premier résultat global met en évidence les écarts entre tableaux, c'est-à-dire entre hauteurs de prélèvement. En particulier, les prélèvements situés autour de l'encoche qui correspondent à des zones très activées, se distinguent des autres.

Sur un deuxième graphique, chaque arbre est représenté par un point «compromis». Les arbres témoins se séparent nettement des arbres greffés, ce qui signifie que l'influence de la greffe semble se faire sentir de façon globale sur l'arbre.

visualisation de l'influence de la greffe à l'aide des trajectoires d'arbres "moyen"

Trajectoires de points moyens. La dernière étape de la méthode trace la «trajectoire» de chaque arbre, c'est-à-dire les différentes positions de l'arbre décrit par chaque tableau. Ces trajectoires individuelles ne nous intéressent pas particulièrement. Par contre, on visualise la différence entre les deux traitements en créant un arbre «moyen» fictif pour chaque traitement, et en traçant la trajectoire de ces deux arbres supplémentaires. Les deux trajectoires se séparent très nettement,

(*) expérimentation réalisée par l'Institut de Recherche sur le caoutchouc, CIRAD.

et montrent que l'évolution de la composition physiologique du latex le long du tronc diffère suivant le traitement. De plus, le sens de parcours des trajectoires s'interprète en fonction de la teneur en extrait sec et en glucides, et donne une explication de la différence constatée entre les traitements.

3. Troisième exemple : suivi de patients

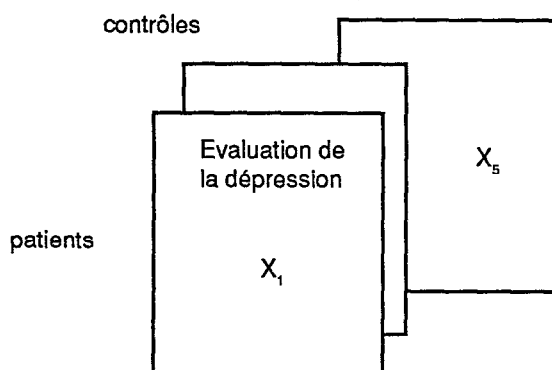
Le logiciel STATIS permet de trier les individus selon un critère qualitatif, et de ne tracer que les trajectoires du groupe d'individus sélectionnés. Au vu des trajectoires, on peut caractériser ce groupe, mais aussi le comparer à une autre groupe puisque les graphiques sont réalisés avec la même échelle.

Prenons comme exemple le fichier proposé par l'Association pour la Méthodologie de la Recherche en Psychiatrie, comme base commune pour la confrontation de différentes méthodes d'analyse, aux Journées de Statistique 1985 et 1986.

Description des données. Le fichier décrit le suivi, sur huit mois, de 230 patients atteints de dépression nerveuse. Pendant la première période s'étalant sur deux mois, ces patients sont traités par un anti-dépresseur. Pendant la deuxième période de six mois, les patients sont divisés en trois groupes. Le premier groupe continue à être traité par l'anti-dépresseur, le second groupe est traité par l'anti-dépresseur pendant deux mois puis reçoit un placebo, le troisième groupe reçoit un placebo pendant toute la période.

Un contrôle est effectué au début et à la fin de la première période, et tous les deux mois pour la deuxième période. Les données recueillies sont de plusieurs types. La situation familiale, professionnelle, sociale, psychologique du patient est décrite par un certain nombre de variables qualitatives. Lors de chaque contrôle, la dépression est évaluée à l'aide d'une vingtaine de scores. Enfin, on dispose des réponses de chaque patient à un questionnaire sur les événements intercurrents, survenus entre deux contrôles.

étude de l'effet d'un anti-dépresseur



le sens de la trajectoire renseigne sur l'évolution de l'état du patient

Tri des trajectoires. Dans cet exemple, le sens de parcours de la trajectoire d'un patient renseigne sur l'évolution de sa dépression. Une trajectoire dirigée vers le bas et la gauche du graphique reflète une aggravation de la dépression. A l'opposé, une trajectoire évoluant vers le haut et la droite du graphique indique une amélioration de l'état du patient. Le nombre de patients étant très élevé, la sélection des trajectoires selon un ou plusieurs critères qualitatifs s'avère très utile pour infirmer ou confirmer des hypothèses.

On peut s'intéresser, par exemple, à la comparaison des trois traitements, sur les femmes entre 40 et 50 ans souffrant d'un type particulier de dépression. Ou juger de l'impact de certains événements intercurrents sur l'état du patient, etc.

Contrairement à l'exemple précédent, nous connaissons un grand nombre d'informations sur chaque patient. La trajectoire moyenne du groupe des femmes de 40 à 50 ans, souffrant d'une dépression réactionnelle apporterait peu d'informations. Par contre, l'examen des trajectoires individuelles de ces patientes donne une idée de la disparité des évolutions, et incite à en rechercher l'explication au moyen d'autres critères.

données manquantes

Suivi des patients sortis de l'étude prématurément. Les patients qui n'ont pas subi les derniers contrôles sont traités en éléments supplémentaires. Le logiciel offre la possibilité de ne tracer que la partie connue de la trajectoire, qui s'interprète avec les mêmes règles que la trajectoire complète. On peut ensuite, de la même façon que précédemment, trier les patients suivant la raison de sortie de l'étude : refus d'un traitement prolongé, effets secondaires nécessitant l'arrêt du traitement, patients perdus de vue, résultat thérapeutique jugé insuffisant, etc.

III — INTERPRETATION DES RESULTATS FOURNIS PAR LA METHODE STATIS SUR DES DONNEES ECONOMIQUES DECRIVANT L'EVOLUTION DES PROVINCES ESPAGNOLES ENTRE 1960 ET 1979

1. Evolution globale de l'économie espagnole entre 1960 et 1979

Une croissance rapide, suivie de la crise de 1973

Le « plan de estabilizacion » de 1959 a mis fin à la période d'autarcie et de protectionnisme, qui caractérisait l'économie de l'Espagne depuis la fin de la guerre civile. Favorisée par le contexte de prospérité de l'Europe, l'Espagne rentre alors dans une phase de croissance rapide qui va fortement altérer sa structure socio-économique. Puis l'économie subit les conséquences de la crise internationale, déclenchée par le premier choc pétrolier en 1973, et la récession et le chômage s'installent.

a entraîné un déséquilibre.

L'agriculture subit, elle aussi, de profondes transformations. Le travail de PEREZ-HULGADE (1987) étudie plus particulièrement la relation entre le développement économique et les résultats du secteur agricole. La croissance rapide de l'économie, au début des années 60, a entraîné une situation de déséquilibre pour les exploitations agricoles, qui persiste encore de nos jours. L'augmentation du revenu par tête a produit, non seulement une augmentation de la demande agricole, mais aussi un changement dans les habitudes de consommation alimentaire, que le système productif traditionnel n'a plus pu satisfaire.

particulièrement dans le secteur agricole.

De plus, l'industrialisation de l'Espagne, et de l'Europe, offrait des possibilités d'emplois qui ont complètement bouleversé le marché du travail dans les campagnes.

Dans ces conditions, les exploitations agricoles ont dû se transformer pour maintenir leur revenu, et survivre. Quelques chiffres mettent en évidence cet effort de modernisation : pendant la période considérée, l'agriculture a perdu la moitié des emplois, a multiplié par cinq l'achat aux autres branches, et presque doublé sa production en termes réels. Les résultats économiques traduisent le degré de réussite de ce processus d'adaptation.

qu'il faut analyser au niveau des provinces

Or, que ce soit du point de vue de l'économie globale, ou de l'agriculture, les changements décrits sont la résultante d'évolutions régionales très hétérogènes. De plus, la polarisation géographique de l'activité économique a provoqué un courant migratoire, sans précédent dans l'histoire de l'Espagne, vers les régions à plus haut revenu et vers l'étranger. Environ deux millions d'habitants abandonnent, en termes nets, les 34 provinces d'émigration vers les 16 autres. Parallèlement, le système financier a contribué à drainer les ressources des zones pauvres vers les riches. Il est, par conséquent, nécessaire d'analyser les changements économiques au niveau de la province, pour comprendre l'évolution économique de l'Espagne.

Le choix des indicateurs économiques.

Pour caractériser l'économie provinciale, nous avons utilisé des indicateurs concernant l'économie globale (densité de population et productivité moyenne du travail) ; la structure sectorielle, mesurée par la contribution de chacun des trois secteurs, agricole, industriel et tertiaire, à l'emploi et au Produit Intérieur Brut provincial en pourcentage ; les résultats économiques des secteurs, définis par la productivité du travail (valeur ajoutée par rapport au nombre d'emplois) et le salaire moyen. Pour le secteur agricole, on a considéré en plus le revenu moyen par actif non salarié. Le choix, parfois délicat, de ces indicateurs est expliqué dans PEREZ-HUGALDE (1987).

La méthode STATIS donne une première description qu'il faut affiner à l'aide d'autres informations.

Dans le cas de données aussi riches et aussi complexes, la méthode STATIS doit être considérée comme une approche descriptive préliminaire. En effet, on est en présence de séries chronologiques multivariées, dont STATIS ne tient pas compte puisqu'elle ignore l'aspect temporel des données. Mais de plus, dans cet exemple, les mouvements migratoires importants entre provinces pendant cette période, introduisent une information liée à la contiguïté géographique que nous n'avons pas prise en compte, et qui devrait être exploitée pour comprendre l'évolution de certaines provinces.

2. Interprétation de l'interstructure

Les coefficients RV entre les années sont particulièrement élevés. Ce qui signifie que les positions mutuelles des provinces évoluent peu d'une année à l'autre, principalement entre 1960 et 1967. Par contre, les normes des W_k évoluent. La diminution régulière de la norme, à partir du maximum atteint en 1967, indique que les corrélations entre les variables d'une même année sont de moins en moins fortes. Nous verrons que cela se traduit par une spécialisation, de plus en plus marquée, des provinces.

évolution globale

Figure 1 — Image Euclidienne des années, associée à l'approximation d'ordre 2 des produits scalaires ($W_{k1} | W_{k2}$)

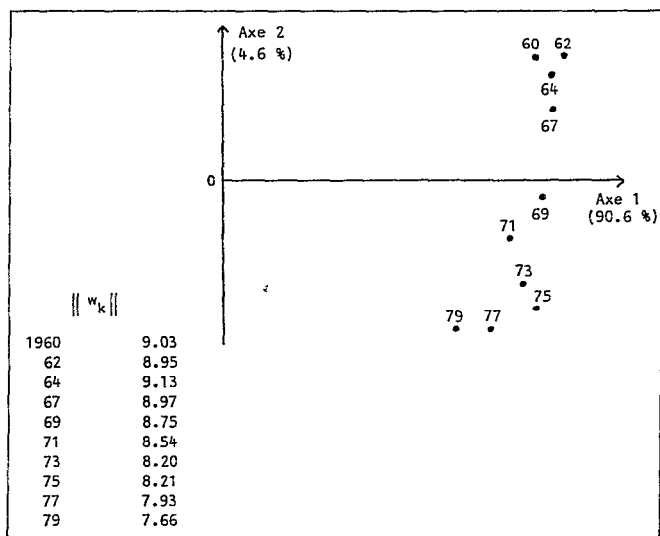
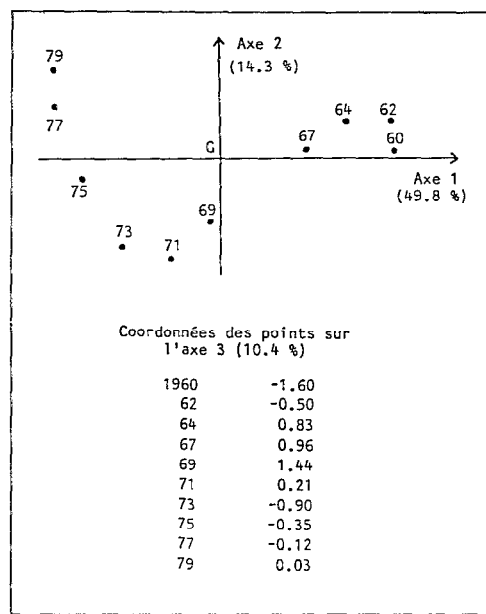


Figure 2 — Image Euclidienne centrée



$$\text{Produits Scalaires Normes : } Rv(k1, k2) = \frac{(w_{k1} | w_{k2})}{\|w_{k1}\| \|w_{k2}\|}$$

	1960	62	64	67	69	71	73	75	77
1960									
62	.954								
64	.934	.966							
67	.933	.945	.963						
69	.891	.900	.928	.953					
71	.875	.881	.893	.926	.957				
73	.839	.839	.850	.885	.911	.954			
75	.840	.841	.859	.895	.924	.934	.934		
77	.803	.816	.837	.876	.898	.905	.906	.955	
79	.784	.807	.828	.865	.880	.884	.890	.910	.938

3. Signification économique des axes du compromis

Le graphique des corrélations des variables de chaque tableau avec les axes du compromis, permet de comparer l'évolution des différents secteurs, de donner une signification économique aux axes, et d'interpréter la situation compromise et les trajectoires des provinces (Fig. n°3 et n°4).

Figure 3 — Signification économique des axes du compromis.

L'axe 1 indique le degré de développement, l'axe 2 oppose les secteurs industriel et tertiaire

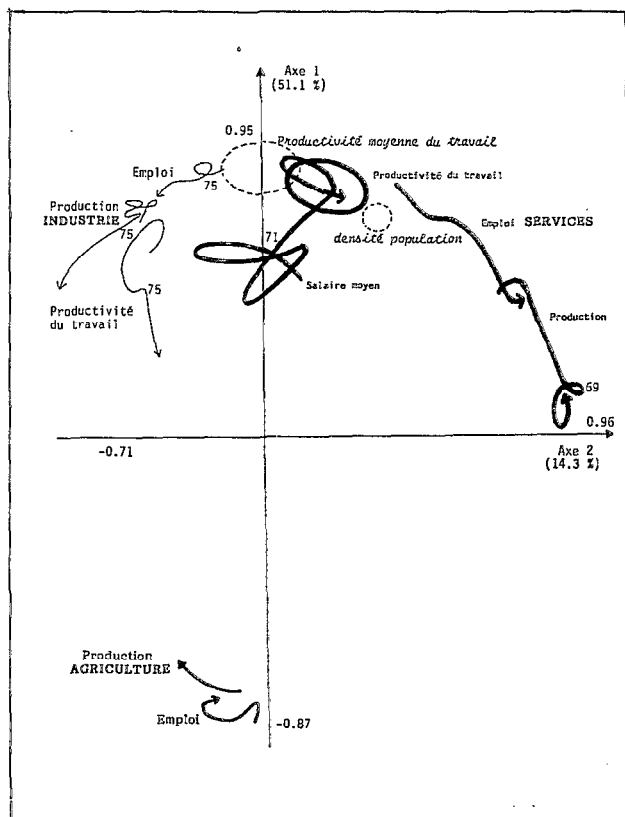
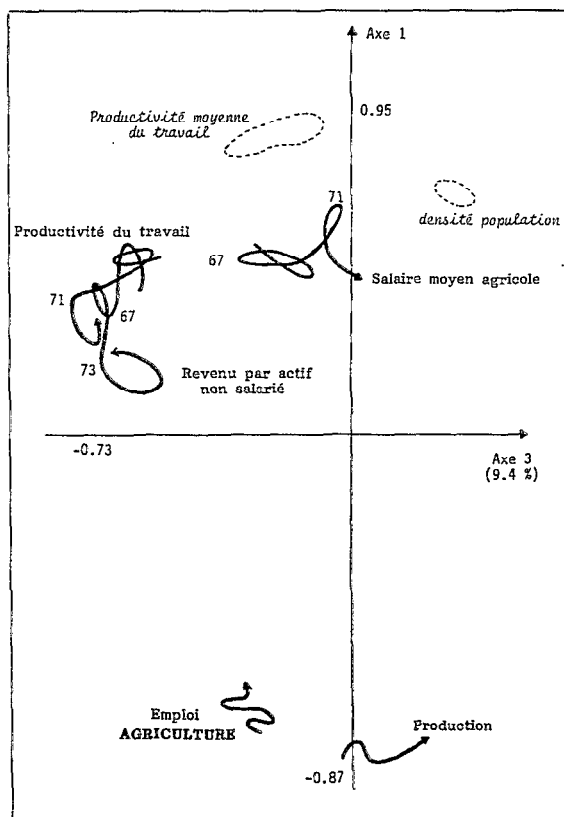


Figure 4 — L'axe 3 est lié au secteur agricole : productivité du travail et revenu par actif non salarié



opposition entre les provinces industrialisées, et les provinces agricoles,

Degré de développement. Toutes les variables, sauf celles qui représentent le poids du secteur agricole dans l'ensemble économique (production et emploi dans l'agriculture), sont corrélées avec le premier axe du compromis. Cet axe, représentant un fort pourcentage de l'inertie, discrimine les provinces selon leur degré de développement. Il oppose les provinces industrialisées, à forte densité de population, qui ont une productivité moyenne du travail élevée, et de bons résultats économiques dans les secteurs non agricoles, à celles qui ont les caractéristiques contraires parmi lesquelles on remarque, en particulier, les provinces qui ont un secteur prépondérant.

entre le secteur industriel et le secteur tertiaire

Opposition industrielle, tertiaire. Le deuxième axe oppose le secteur tertiaire au secteur industriel. En effet, tout au long de la période, les variables production et emploi du secteur tertiaire sont de plus en plus corrélées avec l'axe 2 et, simultanément, de moins en moins corrélées avec l'axe 1. Par contre la productivité du travail de ce secteur reste corrélée avec le premier axe, et peut être considérée comme une caractéristique significative du degré de développement.

Le déplacement des variables production et emploi du secteur industriel, est comparable à celui du secteur tertiaire, mais de signe contraire sur le deuxième axe, plus tardif, et de moindre intensité. Contrairement au secteur tertiaire, la productivité du travail industriel est de moins en moins corrélée avec le premier axe, et n'est plus un indicateur du degré de développement à la fin de la période.

Agriculture. La productivité du travail agricole, et le revenu par actif non salarié dans ce secteur, sont très corrélés (la corrélation oscille entre 0.64 et 0.77 durant la période étudiée), et définissent le troisième axe. Parallèlement, ces variables sont de moins en moins corrélées avec le premier axe. Cela signifie que les agriculteurs riches ne se trouvent pas nécessairement dans les régions à fort degré de développement.

Par contre, la variable salaire moyen du secteur agricole est davantage corrélée avec le degré de développement de 67 à 75. Les travailleurs agricoles perçoivent donc des salaires dont le niveau semble dépendre plus des caractéristiques économiques générales, que de l'efficacité du secteur.

4. Positions compromis des provinces

forte disparité des provinces,

selon leur dominante économique.

Le développement et les résultats agricoles ne sont pas liés

Une simple observation des points compromis sur la figure n°5 met en évidence la forte disparité des provinces espagnoles, quant à leur développement. La situation tout à fait exceptionnelle de Barcelone, du Pays Basque (Vizcaya, Guipuzcoa et Alava), et de Madrid dénote l'existence d'un progrès économique non comparable à celui du reste de l'Espagne.

Nous pouvons remarquer d'autre part, que les provinces moyennement développées se séparent sur l'axe 2 selon leur dominante économique. Ainsi nous retrouvons, éloignées à droite, de nombreuses provinces de la côte méditerranéenne (Malaga, Granada, Cadiz, Valencia, Murcia, Almería), les Baléares et les îles Canaries, pour lesquels le secteur services est prépondérant. Sur la gauche se trouvent des provinces dont la structure économique est plus équilibrée, ou nettement industrialisée comme Oviedo.

Les provinces présentant le plus faible niveau de développement ne se séparent pas sur le deuxième axe, parce que leur structure économique est dominée par la composante agricole. Dans les provinces d'Orense et de Lugo, situées en bas du graphique, l'emploi agricole représente en moyenne sur la période, 57 % du total.

La position compromise des provinces dans le plan 1-3 (figure n°6) illustre le peu de liaison constatée entre les résultats de l'agriculture, et le développement global. En effet, les provinces les plus développées ont des résultats moyens, ou faibles, dans l'agriculture. Par contre, de nombreuses provinces à faible niveau de développement ont un secteur agricole efficace, ou des exploitations bien rémunérées. Nous observons que les provinces qui enregistrent, en moyenne, les résultats les plus défavorables sont celles de la Galice (Orense, Lugo, Pontevedra et la Coruña) et de la côte cantabrique (Santander, Oviedo, Vizcaya et Guipuzcoa). Bien que ces provinces aient des niveaux de développement différents, elles ont la caractéristique commune d'être orientées vers l'élevage bovin.

Figure 5 — Position compromise et trajectoires des provinces dans le plan 1.2.

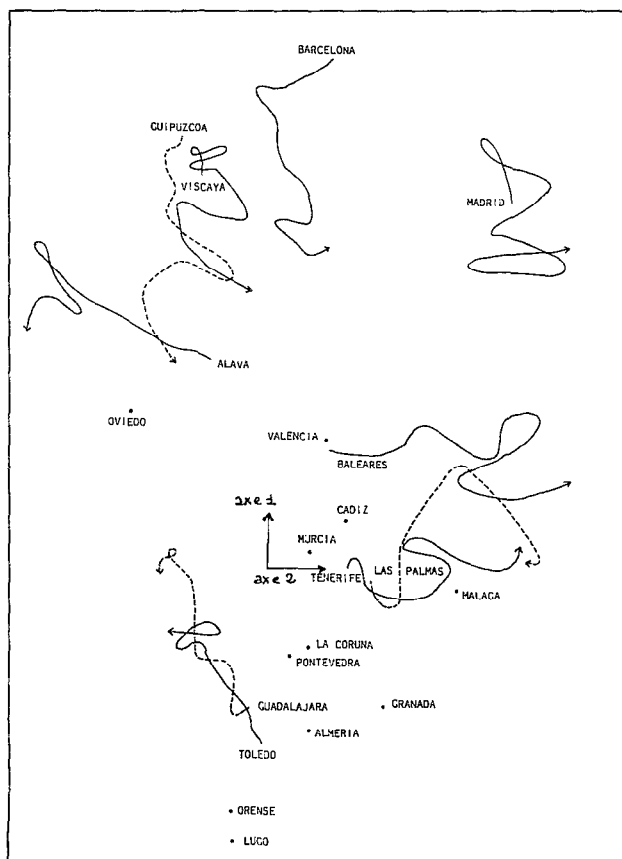
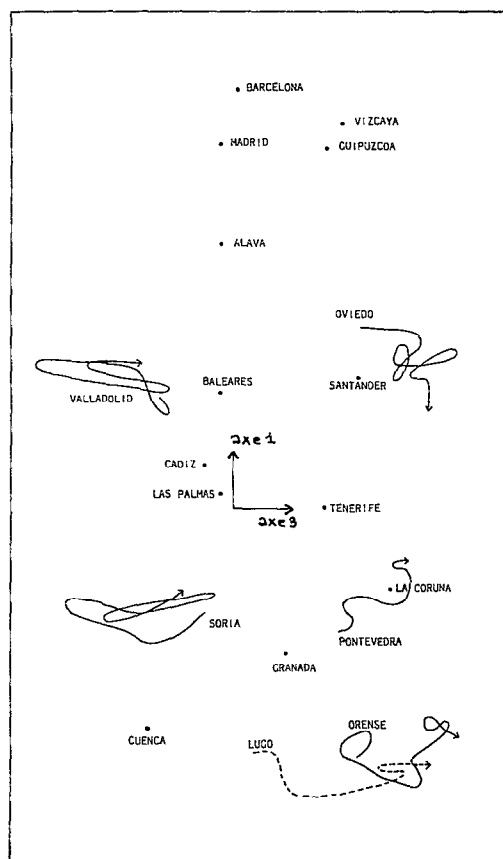


Figure 6 — Positions compromise et trajectoires des provinces dans le plan 1.3.



5. Trajectoires des provinces

Provinces dont le développement économique est en progression.

Nous allons expliquer l'évolution des provinces dont les déplacements sont importants, sur les figures n°5 et n°6. Les tableaux permettent un retour aux données, pour les années extrêmes 1960 et 1979.

En général, les provinces qui ont une trajectoire ascendante sur la figure n°5, et qui améliorent par conséquent leur niveau de développement, sont celles pour lesquelles on constate une progression du poids de l'industrie, et où la productivité du travail a évolué plus rapidement que la moyenne. C'est le cas, par exemple, de Guadalajara, de Toledo, et d'Alava. On peut observer cependant que la capacité d'entraînement économique du secteur industriel, est moins forte pendant les années 70 que dans la décennie précédente : les trajectoires tournent sur elles-mêmes à partir de cette date.

Tableau 1 — Données concernant les provinces qui ont connu une progression ou une régression importante de leur niveau de développement. La moyenne de chaque variable pour une année a été ramenée à 100

	GUADALAJARA		TOLEDO		GUIPUZCOA		VIZCAYA	
	1960	1979	1960	1979	1960	1979	1960	1979
EMPLOI								
AGRICULTURE	126.7	90.7	127.4	116.8	22.9	24.1	26.8	14.8
INDUSTRIE	73.6	116.4	72.6	107.0	202.3	164.3	201.6	164.4
SERVICES	82.5	95.8	82.2	85.7	130.7	96.8	129.1	105.8
PRODUCTION								
AGRICULTURE	130.8	112.6	164.5	150.8	14.0	16.9	19.4	13.8
INDUSTRIE	87.3	115.4	56.1	105.2	176.9	145.3	177.9	137.5
SERVICES	87.9	89.0	86.8	86.7	92.8	88.3	98.5	95.7
PRODUCTIVITE DU TRAVAIL								
GLOBALE	80.4	100.9	79.6	91.1	182.6	117.2	157.9	122.0
AGRICULTURE	87.5	123.7	108.2	116.0	118.3	80.9	120.3	112.2
INDUSTRIE	94.7	98.9	61.0	88.6	158.5	102.5	138.3	100.8
SERVICES	86.8	94.3	85.2	92.7	131.4	107.5	122.1	110.9

Tableau 2 — Provinces en forte progression dans le secteur services

	BALEARES		LAS PALMAS		S.C. DE TENERIFE	
	1960	1979	1960	1979	1960	1979
DISTRIBUTION DE L'EMPLOI						
AGRICULTURE	78.5	54.7	102.5	59.6	123.6	79.0
INDUSTRIE	116.0	82.0	78.2	69.9	67.7	71.0
SERVICES	122.4	143.5	107.6	142.7	89.5	135.0
DISTRIBUTION DE LA VALEUR AJOUTEE						
AGRICULTURE	68.0	30.3	104.6	44.7	90.5	57.1
INDUSTRIE	104.9	58.1	64.1	58.4	88.3	66.4
SERVICES	124.1	143.3	121.0	133.4	120.0	132.1
PRODUCTIVITE DU TRAVAIL						
GLOBALE	113.2	120.4	100.2	106.9	91.4	104.7
AGRICULTURE	103.3	65.9	107.6	79.2	70.5	74.8
INDUSTRIE	101.6	84.3	81.6	88.2	118.4	96.7
SERVICES	116.4	120.8	114.2	100.5	124.1	103.0

Tableau 3 — Résultats du secteur agricole pour les provinces à orientation bovine.

	LA CORUNA	LUGO	ORENSE	PONTEVEDRA	OVIEDO	SANTANDER	VIZCAYA	GUIPUZCOA
PRODUCTIVITE DU TRAVAIL								
1960	65.7	84.4	48.5	63.5	82.9	96.0	120.3	118.3
1979	53.2	46.4	24.1	35.2	46.2	57.3	112.0	80.9
REVENU PAR ACTIF NON SALARIE								
1960	54.5	75.6	38.2	53.2	65.9	89.0	94.2	93.3
1979	42.7	34.9	18.9	29.0	36.5	45.9	98.7	69.1

en régression.

Provinces
spécialisées.

Provinces
agricoles.

Les régions de vieille tradition industrielle comme le Pays Basque (Vizcaya, Guipuzcoa), Asturias (Oviedo), et Cantabrie (Santander) n'ont plus un niveau de développement relatif aussi fort à la fin de la période. Cette tendance est illustrée sur la figure n°5 par les trajectoires de Vizcaya et de Guipuzcoa, dont la «dégringolade» le long de l'axe 1 s'accélère à partir de la fin des années 60. Les observations que nous venons de faire expliquent le comportement des variables significatives du poids de l'industrie, dont la corrélation avec le premier axe diminue. En particulier, à partir de 1973, la productivité du travail industriel évolue moins rapidement dans les provinces industrielles du Pays Basque, Asturias, et Cantabrie, que la moyenne.

Toujours sur la figure n°5, les déplacements latéraux vers la droite sont significatifs d'un processus de spécialisation dans le secteur tertiaire. C'est le cas de provinces qui avaient déjà un secteur service important au début des années 60 (le point de la trajectoire pour l'année 1960 est sur la droite du graphique), comme les Baléares, les îles Canaries ainsi que certaines provinces de la côte Méditerranéenne. Mais cela n'a pas entraîné une progression significative de leur développement économique : leurs trajectoires sont, dans l'ensemble, horizontales.

Sur la figure n°6, les mouvements vers la gauche sont significatifs d'une amélioration des résultats économiques du secteur agricole. Ces mouvements concernent essentiellement des provinces moyennement développées, qui avaient déjà de bons résultats dans ce secteur au début de la période.

Il n'y a pas de trajectoires pratiquement horizontales, sans changement de direction, comme celles des provinces touristiques de la figure n°5. En effet, les trajectoires oscillent le plus souvent, parallèlement au troisième axe avec de grandes amplitudes, comme celles de Valladolid et de Soria, et semblent caractériser des agricultures extensives, dont les résultats sont soumis aux conditions climatiques, ou aux aléas du marché. Par contre, les provinces à orientation bovine comme Orense, Lugo, Oviedo, qui avaient au départ un secteur agricole peu efficace, voient leurs résultats agricoles se dégrader.

CONCLUSION

Cet article présente, de façon très succincte, une méthode d'analyse conjointe de tableaux de données quantitatives, la méthode STATIS, au moyen d'exemples d'application. Le chercheur peut reconnaître ainsi les analogies avec ses propres problématiques, et avoir un aperçu de ce que peut lui apporter une méthode adaptée à ce type de données.

BIBLIOGRAPHIE

KROONENBERG, P., 1983.— Three mode principal component analysis.— PAYS BAS : DSWO Press.

PEREZ HUGALDE C. — Evolucion de las relaciones entre la productividad del trabajo en la agricultura, las estructuras agrarias y el desarrollo economico en Espana (1960-1979). Un analisis mediante metodos estadisticos multivariados, Thèse de doctorat, Université Polytechnique de Madrid, 1987.

La théorie mathématique de la méthode STATIS se trouve dans :

ESCOUFIER, Y., 1977.— Operators related to a data matrix, Recent developments in statistics. — North-Holland, p. 125-131.

ESCOUFIER, Y., 1980. — L'analyse conjointe de plusieurs matrices de données. — in : Biométrie et temps., p. 59-76.

GLAÇON, F., 1981. — Analyse conjointe de plusieurs matrices de données. Comparaison de différentes méthodes. — Th. 3^e cycle, Grenoble.

LAVIT, C., 1988. — Analyse conjointe de tableaux quantitatifs., Paris : MASSON (collection méthode + programmes)

Résultats de différentes techniques d'analyse conjointe de tableaux sur un même jeu de données :

ESCOUFIER Y., BERNARD M.C, LAVIT C., BARRE A., FICHET B., CARLIER A., FOUcart T., LAFAYE J.Y, 1985.— Analyse conjointe de plusieurs tableaux de données.— in : numéro spécial de Statistique et Analyse des Données, n° 10 (1).

SEMPE M., KROONENBERG P.M., LAVIT C., PERNIN M.O., LEWIS P.J, CALOMME G., VAN HOOF J., PONTIER J., MINEO A., 1987. — Multivariate and longitudinal data on growing children, Data analysis : the ins and outs of solving real problems.— London Plenum .— p 7-112.

Comparalson théorique :

CARLIER A., LAVIT C., PAGES M., PERNIN M.O., TURLLOT J.C., (à paraître).— Analysis of data tables indexed by time : a comparative review.— Proceedings of International Meeting on the Analysis of Multiway Data Matrices, 28-30 mars 1988.

Logiciels :

LAVIT C., 1986 : Méthode STATIS. Programme de la bibliothèque MODULAD, utilisable sur IBM TSO - IBM Vm Cms - 368. Multics - CDC. NOS. be - Vax Vms - Mini 6 mode 400 - HP 9000 Unix.
S'adresser à INRIA, domaine de Voluceau - Rocquencourt, BP 105, 78150 LE CHESNAY Cedex.

LAVIT C., FRANCILLON G., 1987 : Méthode STATIS. Exécutable sur IBM PC.
S'adresser à Unité de Biométrie, Place Viala, 34060 MONTPELLIER Cedex.

Presentation of the Statist method for the combined analysis of several tables of quantitative data.- Christine LAVIT

The Statist method makes it possible to summarize information in analysis of data in the form of several tables. The method is based on Principal Component Analysis and meets the following objectives :

- detection of tables that are similar to each other,*
- production of a «summary» table of all the data,*
- description of the differences between tables in relation to the summary table : are these differences caused by individuals or by the variables ?*

A simple program for using this method and which can run on an IBM PC type microcomputer is proposed ; a system for large systems also exists.

Key words : Method - Data analysis - Data processing - Software

Presentación del metodo STATIS que permite el análisis paralelo de varios cuadros de datos cuantitativos - Christine.LAVIT

Se supone que debemos analizar datos que se presentan en varios cuadros de medidas, referentes a los mismos individuos en situaciones diferentes. El método STATIS permite resumir la información contenida en estos datos. Este método, que se basa sobre el principio de Analisis en Componentes Principales, responde a los siguientes objetivos :

- determinar los cuadros que se parecen,*
- realizar un cuadro que resuma el conjunto,*
- describir las diferencias entre los cuadros en relación con el cuadro «resumen» : ¿se deben estas diferencias a los individuos o a las variables?*

Para iniciar este método, proponemos un programa de informatica de utilización simple, que funciona sobre micro-computadores de tipo IBM-PC, y del cual existe tambien una versión para los grandes sistemas de tamaño más importante.

Palabras claves : Método - Analisis de datos - Tratamiento de datos - Programa de informatica.