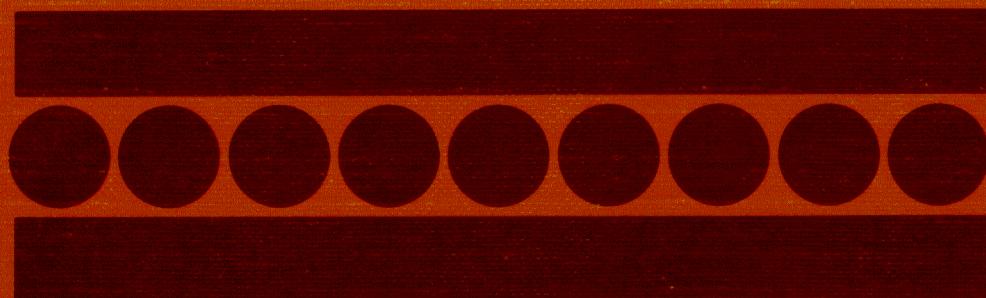


NAZIOARTEKO ESTATISTIKA
MINTEGIA EUSKADIN

1983

SEMINARIO INTERNACIONAL
DE ESTADISTICA EN EUSKADI



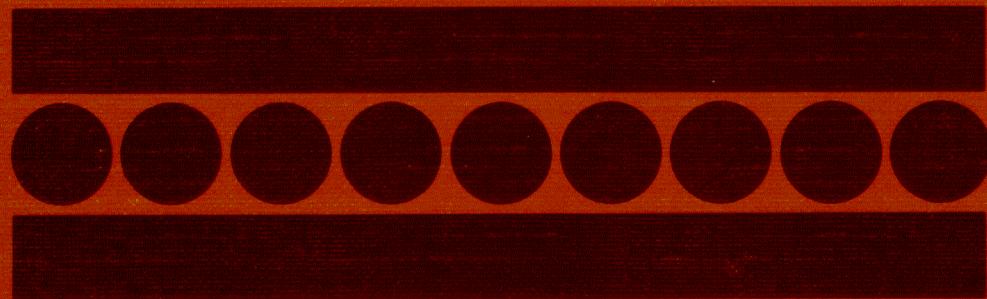
DATU-ANALISA

ANALISIS DE DATOS

ANALYSE DES DONNEES

DATA ANALYSIS

P. CLAPIER



NAZIOARTEKO ESTATISTIKA
MINTEGIA EUSKADIN

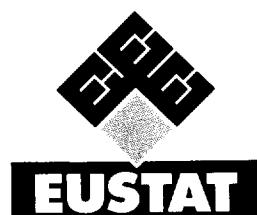
1983

SEMINARIO INTERNACIONAL
DE ESTADISTICA EN EUSKADI

DATU-ANALISA
•
ANALISIS DE DATOS
•
ANALYSE DES DONNEES
•
DATA ANALYSIS

P. CLAPIER

CUADERNO **4** Koadernoa



Zuzendaritza/Dirección: Anjeles Iztueta Azkue
José Pérez Vilaplana

© Eusko Jaurlaritza - Gobierno Vasco

Depósito Legal: S.S. 610/92
ISBN: 84-7542-127-10. Obra Completa
ISBN: 84-7542-157-1. Tomo II.
Impreso en Itxaropena, S.A. - Araba kalea, 45 - Zarautz

AURKEZPENA

Estatistikako Mintegi Internazionalak sustatzean, hainbat xederekin bete nahi luke Eusko Jaurlaritzaren Estatistika Zuzendaritzak, hala nola:

- Unibertsitatearekiko eta, bereziki, Estatistika Sailarekiko lankidetza bultzatu.
- Funtzionari, irakasle, ikasle eta estatistikaren alorrean interesaturik leudekeen guztien birziklapen profesionala erraztu.
- Estatistikako alorrean eta mundu-mailan irakasle prestu eta abangoardiako ikerlari diren pertsonaiak Euskadira ekarri, guzti horrek zuzeneko harremanei eta esperientzien ezagupenei dago-kienez soposatzen duen ondorio positiboarekin.

Iharduketa osagarri bezala eta interesaturik leudekeen ahalik eta pertsona eta Erakunde gehienetara iristearren, Ikastaro hauetako txostenak argitaratzea erabaki da, beti ere txostenemaiaren jatorrizko hizkuntza errespetatuz, horrela gure Herrian gai honi buruzko ezagutza zabaltzen laguntzeko asmoarekin.

PRESENTACION

Al promover los Seminarios Internacionales de Estadística, la Dirección de Estadística del Gobierno Vasco pretende cubrir varios objetivos:

- Fomentar la colaboración con la Universidad y en especial con los Departamentos de Estadística.
- Facilitar el reciclaje profesional de funcionarios, profesores, alumnos y cuantos puedan estar interesados en el campo estadístico.
- Traer a Euskadi a ilustres profesores e investigadores de vanguardia en materia estadística, a nivel mundial, con el consiguiente efecto positivo en cuanto a relación directa y conocimiento de experiencias.

Como actuación complementaria y para llegar al mayor número posible de personas e Instituciones interesadas, se ha decidido publicar las ponencias de estos Cursos, respetando en todo caso la lengua original del ponente, para contribuir así a acrecentar el conocimiento sobre esta materia en nuestro País.

PRESENTATION

La Direction de Statistique du Gouvernement Basque se propose d'atteindre plusieurs objectifs par la promotion des Séminaires Internationaux de Statistique:

- Encourager la collaboration avec l'université et spécialement avec les départements de statistique.
- Faciliter le recyclage professionnel des fonctionnaires, professeurs, élèves, et tous ceux qui pourraient être intéressés par la statistique.
- Inviter en Euskadi des professeurs mondialement renommés et des chercheurs de premier ordre en matière de Statistique avec tout ce que cela pourrait entraîner comme avantage dans les rapports et l'échange d'expériences.

En outre, il a été décidé de publier les exposés de ces rencontres afin d'atteindre le plus grand nombre de personnes et d'institutions intéressées, et pour contribuer ainsi à développer dans notre pays les connaissances sur cette matière. Dans chaque cas la langue d'origine du conférencier sera respectée.

PRESENTATION

In promoting the International Seminars on Statistics, the Statistics Office of the Basque Government is attempting to achieve a number of objectives:

- Encourage joint working with the Basque University and, in particular, with its Department of Statistics.
- Facilitate the in-training of civil servants, teachers and students and of all those interested in the field of statistics.
- Bring to Euskadi distinguished academics and researchers in the front line of statistics work, at a world-wide level, with all the benefits that this will bring through direct contacts and the interchange of experiences and ideas.

As an additional step this year, it has been decided to publish in advance the papers to be presented at these courses, respecting the native language of the speaker, in each case. This is in order that as many interested people and institutions as possible are made aware. In this way we hope to contribute to the growth and awareness concerning this topic in our country.

Vitoria-Gazteiz, Diciembre 1984 Abendua

JOSE IGNACIO GARCIA RAMOS
Estatistikako Zuzendaria
Director de Estadística

SARRERA

Koaderno honek, Eusko Jaurlaritzaren Estatistika-Zuzendaritzak eta Euskal Herriko Unibertsitatearen Matematika Aplikatuko Departamentuak antolaturik, P. CLAPIERek Euskadiko Estatistikako I. Nazioarteko Mintegien barruan "Analyse des Données" gaiari buruz eman duen ikastaroa laburbiltzen du. I. Mintegi honek kontatzen du baita USAko Pittsburgh-eko Unibertsitateko C.R. RAO Jaunaren partaidetzarekin ere, "Lineal Statistical Inference" gaiari buruzko ikastaro batekin, Txileko Santiagoko Unibertsitateko E. CANSADO Jaunaren partaidetzarekin, "Muestreo y Aplicaciones" gaiari buruzko ikastaroarekin, eta orobat Sheffield-eko Unibertsitateko V. BARNETT Jaunaren partaidetzarekin, "Statistical Education" buruzko ikastaroarekin

INTRODUCCION

Este cuaderno resume el curso que sobre "Analyse des Données" ha impartido P. CLAPIER dentro del I Seminario Internacional de Estadística en Euskadi, organizado por la Dirección de Estadística del Gobierno Vasco y el Departamento de Matemática Aplicada de la Universidad del País Vasco. Este I Seminario cuenta con la participación de C.R. RAO de la Universidad de Pittsburgh, USA, con un curso sobre "Lineal Statistical Inference", E. CANSADO de la Universidad de Santiago de Chile con un curso sobre "Muestreo y Aplicaciones" y V. BARNETT de la Universidad de Sheffield con un curso sobre "Statistical Education".

INTRODUCTION

Ce cahier résume le cours que P. CLAPIER a donné sur "Analyse des Données" au I Séminaire International de Statistique d'Euskadi, organisé par le Conseil de Statistique du Gouvernement Basque et par le Département de Mathématique Appliquée de l'Université du Pays Basque. Ce I Séminaire compte avec la participation de C.R. RAO de l'Université de Pittsburgh (USA) avec un cours sur "Lineal Statistical Inference", ainsi que celle de E. CANSADO de l'Université de Santiago de Chile avec un cours sur "Muestreo y Aplicaciones" et celle de V. BARNETT de l'Université de Sheffield avec un cours sur "Statistical Education".

INTRODUCTION

This paper summarises the series of lectures on "Data Analysis" submitted by P. CLAPIER for the I International Statistic Seminar in Euskadi, organised by the Statistycs Office of the Basque Government and by the Department of Applied Mathematics at the University of the Basque Country. This I Seminar included the participation of C.R. RAO of Pittsburgh University (USA) with a course on "Lineal Statistical Inference" of E. CANSADO of Santiago University in Chile with a series of lectures on "Muestreo y Aplicaciones", and of V. BARNETT of Sheffield University with a course on "Statistical Education".

BIOGRAFIA

Patrick Clapier, 1974etik hona (CREDOC) Centre National de Recherche Scientifique-ean Ikerketako Injineru izaki, Datuen Analisiaren irakaskuntzari emana dago Pariseko Institut de Statistique des Universités Ikastetxearen 1976tik hona.

Bere "Maitrise de Mathematiques et Applications Fondamentales" titulua 1972an lortu zuen, Pariseko XI. Barrutiko (ORSAYko) Unibertsitatean eta bere "Certificat de Statistiques Appliquées" berriz, aipatu I.S.U.P. Ikastetxearen, 1976an.

Hainbat artikulu ere argitara eman izan du Datuen Analisiari buruz "Les Cahiers de l'Analyse des Données" eta "Consommations" aldizkarietan, eta, esan dezagun baita zenbait CREDOC-eko txostenen eta INSEEren Oharren autore ere badela.

Bestalde, parte hartu izan du Datuen Analisiari buruzko Bilera Zientifikoetan ere Latin Amerikan, eta gai berari buruz prestakuntzako ikastaroak zuzendatu Caracaseko Unibertsitate Zentralean eta Meridako Unibertsitatean, Venezuelan bi-biak.

Patrick Clapier, Ingeniero de Investigación del Centre National de la Recherche Scientifique (CREDOC) desde 1974, está dedicado a la enseñanza del Análisis de Datos en el Institut de Statistique des Universités de París desde 1976.

Obtuvo su Maîtrise de Mathématiques et Applications Fondamentales en la Université París XI (ORSAY) en 1972 y el Certificat de Statistique Appliquées del I.S.U.P. en 1976.

Ha publicado diversos artículos sobre Análisis de Datos en "Les Cahiers de l'Analyse des Données" y en "Consommations", siendo autor asimismo de diversos informes de CREDOC y Notas del INSEE.

Ha participado en diversas Reuniones Científicas sobre Análisis de Datos en América Latina, y dirigido cursos de formación en el tema en la Universidad Central de Caracas y en la Universidad de Mérida, ambas de Venezuela.

Patrick Clapier Ingénieur de Recherches au Centre National de la Recherche Scientifique (CREDOC) puis 1974, se consacre depuis 1976 à l'enseignement de l'Analyse de données à l'Institut de Statistique de l'Université de Paris.

Il obtient sa Maîtrise de Mathématiques et Applications Fondamentales à l'Université de Paris XI (ORSAY) en 1972 et le Certificat de Statistique Appliquée de l'I.S.U.P. en 1976.

Il a publié divers articles sur l'Analyse des données dans "Les Cahiers de l'Analyse des Données" et dans "Consommations", et il est également l'auteur de divers rapports du CREDOC et de Notices de l'INSEE.

Il a participé à diverses Réunions Scientifiques sur l'Analyse des Données en Amérique Latine et a dirigé des cours de formation sur ce même sujet à la Universidad Central de Caracas et à celle Mérida, toutes les deux au Vénézuéla.

Patrick Clapier, research engineer at the National Centre for Scientific Research (CREDOC) since 1974, is the teaching of data analysis at the Statistics Institute of Paris University and has been since 1976.

He obtained his Mister's degree in Mathematics and Fundamental Applications from the Université Paris XI (ORSAY) in 1972 and the Certificate in Applied Statistics from I.S.U.P. in 1976.

He has published various articles on data analysis in "Les Cahiers de l'Analyse des Données" and in "Consommations", being the author of several CREDOC reports and on INSEE memoranda.

He has participated in various scientific meetings on data analysis in Latin America and he has directed training courses on the theme at the Central University of Caracas and at Mérida University, both in Venezuela.



AURKIBIDEA / INDICE

Aurkezpena/Presentación/Presentation	
Presentation	5
Sarrera/Introducción/Introduction/Introduction	
.....	7
Biografía	9
DATU ANALISA	13
Laburpena	13
ANALYSE DES DONNEES	15
1. A quoi sert l'analyse des Données	15
1.1. Utilité et Place spécifique des méthodes	15
1.2. Les avantages des représentations obtenues	17
1.3. Le gain de productivité dans les dérouilllements d'enquêtes	21
2. A quoi s'applique l'analyse des données.	21
2.1. Domaine d'application	21
2.2. Méthodes adaptées aux grands tableaux	21
2.3. Les qualités d'un tableau	21
3. En quoi consiste l'analyse des données .	23
3.1. Bibliographie	23
3.2. Les différentes méthodes factorielles et les différents types de tableaux	24
3.2.A. Les trois principaux types de tableaux	24
3.2.B. Les diverses méthodes factorielles	24
3.3. Principes des méthodes factorielles et exemples illustratifs	25
3.3.1. Analyse générale: (AG)	25
3.3.2. Analyse des correspondances (AC)	25
3.3.3. Analyse des correspondances multiples (CM)	31
4. Mise en oeuvre pratique des méthodes.	40
ANALISIS DE DATOS	60
1. Para qué sirve el análisis de datos	60
1.1. Utilidad y lugar específico de los métodos	60
1.2. Ventajas de las representaciones obtenidas	62
1.3. Ganancia de productividad en los escrutinios de las encuestas	63
2. A qué se aplica el análisis de datos	63
2.1. Campo de aplicación	63
2.2. Métodos adaptados a tablas grandes	63
2.3. Las cualidades de una tabla	63
3. En qué consiste el análisis de datos	64
3.1. Bibliografía	64
3.2. Diferentes métodos factoriales y diferentes tipos de tablas	65
3.2.A. Los tres principales tipos de tablas	65
3.2.B. Los diversos métodos factoriales	65
3.3. Principios de los métodos factoriales y ejemplos ilustrativos	66
3.3.1. Análisis general: (AG)	66
3.3.2. Análisis en componentes Principales	66
3.3.3. Análisis de Correspondencias (AC)	72
3.3.4. Análisis de Correspondencias múltiples (CM)	79
4. Puesta en práctica de los métodos	81

DATU-ANALISIA

Patrick Clapier
Centre National de la Recherche Scientifique
PARIS

LABURPENA

Sarrera labur baten ondoren, Datu-Analisi izeneko metodoek zertarako balio duen azaltzen da, me todoen baliagarritasuna eta leku espezifikoa azpimarratuz eta lortutako errepresentazioen abantaila itzelak eta mota guztiak inkesten azterken produktibilitatean eta efektibotasunean lortutako aurrekada adieraziz.

Era berean, metodo hauen aplikazio-barrutiak ere aipatzen dira. Ondoren, Datu-Analisiko metodo / nagusiak sailkatzen dira, jarraian analisi-gai diren taula-mota nagusiak sailkatuaz.

Gero, Datu-Analisiko metodo nagusietako batzuen printzipoak tratzen dira, Osagai Nagusizko Ana lisia, Korrespondentzi Analisia eta Korrespondentzi Anizkoitzen Analisia, adibide ilustragarri egoki ekin batera.

Azkenik, datu - multzo baten analisi osoa egiten da, SPAD programa - paketeaz baliatuz.

ANALYSE DES DONNEES

Patrick Clapier
Centre National de la Recherche Scientifique
PARIS

INTRODUCTION

Toutes les méthodes de statistique multidimensionnelle ont le même but : il s'agit à la fois de décrire, de classer et de clarifier les données.

Quel tableau va-t-on soumettre à l'analyse? Quels/codages et recodages va-t-on effectuer sur les données? Quelle est la meilleure mesure des liaisons entre les objets à décrire? Quelle méthode utiliser? Les étapes / sont nombreuses et l'interprétation des résultats en dé-

pend étroitement. Ce n'est donc pas seulement un ensemble de techniques que nous voulons exposer mais aussi / une manière d'aborder les problèmes de dépouillement statistique.

De grands tableaux de données peuvent dans une certaine mesure être traités de façon globale et donner lieu à des synthèses, on peut étudier un phénomène sur plusieurs dimensions à la fois.

1. A QUOI SERT L'ANALYSE DES DONNEES

Nous ferons ici un tour d'horizon sur l'utilité, la place spécifique par rapport à la statistique, les avantages ou particularités de ces méthodes.

1.1 UTILITE ET PLACE SPECIFIQUE DES METHODES

En statistique certaines méthodes sont difficiles / et ne sont pas toujours utiles. Les techniques d'Analyse de données sont relativement faciles puisqu'elles / font essentiellement appel à l'algèbre linéaire et / elles sont très utiles en fournissant des représenta---

tions synthétiques de vastes ensembles de données.

Les méthodes d'Analyse de données appartiennent à la statistique descriptive et jouent le rôle d'instruments d'observation.

Divers logiciels sont complets ou puissants sur l'un des deux domaines suivants:

A) Analyse exploratoire, description uni ou bivariée (Histogrammes, tableaux croisés, ...)

B) Modèles statistiques multivariés (Régression, discrimination, analyse canonique ...)

Cependant ils sont très incomplets dans le domaine / intermédiaire où nous situons l'analyse de données: méthodes à la fois exploratoires et multidimensionnelles ; techniques permettant de découvrir la structure, éventuellement compliquée, d'un grand tableau de nombres à plusieurs dimensions et la traduire par une autre structure plus simple et qui la résume au mieux.

On peut diviser les techniques d'analyse de données en deux grandes familles:

A) Les techniques factorielles:

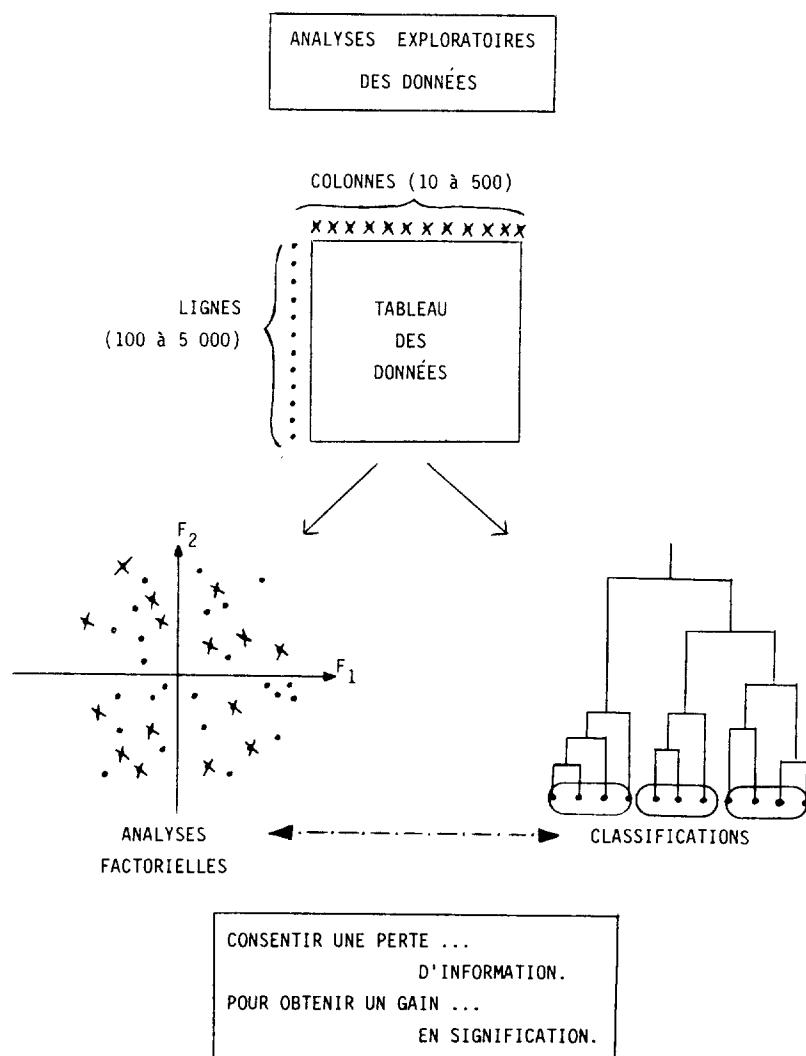
Elles s'apparentent dans leur principe aux techniques d'analyse factorielle proposées et mises au point /

au début du siècle par les psychologues. Elles utilisent des calculs d'ajustement qui font essentiellement appel à l'algèbre linéaire. L'utilisateur obtient des cartes / structurant toutes les positions relatives à la fois de l'ensemble des lignes et de l'ensemble de colonnes du tableau qu'il veut décrire. Cette structuration est repérée par ce qu'on appelle des axes factoriels . Nous / apprendrons à proposer une interprétation des axes, ou / simplement des proximités observées.

B) Les techniques de classification automatique

Elles sont plus récentes, mettent en jeu des calculs algorithmiques, et produisent des hiérarchies permettant de grouper et de ranger les objets à décrire. L'utilisateur obtient des représentations sous la forme / d'arbres.

Ces deux types de résultats sont schématisés à la suite:



Ces deux familles sont plus complémentaires que concurrentes et peuvent avec profit être utilisées conjointement sur un même jeu de données. En pratique on commence souvent par utiliser les méthodes factorielles / pour positionner les objets à décrire les uns par rapport aux autres. On peut ensuite chercher à les regrouper avec les techniques de classification et vérifier / s'il existe des regroupements dont la procédure précédente n'aurait pas rendu compte. C'est ce que nous ferons pendant les deux dernières conférence sur l'enquête des 111 Fincas de MARACAY: Premièrement nous examineront les relations entre l'ensemble des variables caractérisant la production des fermes, deuxièmement une / classification nous permettra de découvrir une catégorie très caractéristique, celle des producteurs de canne à sucre.

On rencontre beaucoup de méthodes dans chacune de / ces deux grandes familles, nous consacreront des développements plus détaillés aux méthodes factorielles.

Des champs d'observation, autrefois jugés trop vastes et trop complexes peuvent maintenant être étudiés. On peut procéder à des relevés de mesures extensives / sans réductions *a priori* du champ que l'on veut étudier. Ceci suscite une nouvelle attitude des chercheurs.

1.2 LES AVANTAGES DES REPRESENTATIONS OBTENUES

Elles donnent une vue synthétique des relations entre les lignes et les colonnes de grands tableaux.

Elles deviennent un nouveau moyen de communication.

Elles permettent de comparer des recherches.

Elles permettent de retrouver des erreurs en vérifiant la qualité de l'information.

Illustrons leur fonction iconographique par un exemple: nous ne ferons ici qu'un très bref commentaire de certains graphiques extraits d'un article des Cahiers / de l'Analyse des Données. (Vol. IV - 1979 - N° 4, pag./ 443-463). Structure des budgets familiaux et des impôts indirects (BUDG. FAM.) par P. CLAPIER et J.L. MARE.

Les données résultent des enquêtes permanentes sur les conditions de vie des ménages effectuées par l'INSEE (cumul des données relatives à trois années successives 1969-70-71). Le tableau de budgets des ménages croise / les 126 postes de consommation avec les 37 catégories-socio-professionnelles (CSP) du chef de Menage. Il - / s'agit d'un tableau de contingence dont l'élément x_{ij}

est la consommation totale des ménages de la CSP i pour le poste j.

Contrairement à l'article on se limite ici au plan / des deux premiers facteurs. L'axe F₁ (voir graphique n°1 et n° 2) est un axe de ruralité, il met en lumière la / spécificité des budgets des ménages agricoles (exploitants, salariés et anciens agriculteurs). Cette spécificité tient principalement à l'importance de l'autoconsommation dans le domaine alimentaire. Les postes d'autoconsommation (symbole *) sont d'ailleurs régulièrement échelonnés sur cet axe (voir graphique n° 1) à partir de / l'origine en fonction de la spécialisation nécessaire à leur production: fruits et légumes, basse-cour, puis élevage bovin ou porcin (productions les plus spécialisées)

On s'est demandé si cet axe manifestait seulement / l'opposition entre les ménages agricoles et le reste de la population, ou s'il exprimait une échelle plus continue fonction de la zone d'habitat, (taille de la commune ou de l'agglomération de résidence). L'analyse d'un fichier portant à peu près sur les mêmes enquêtes et croisant avec les 126 postes de consommation un croisement / entre la C.S.P. (en 13 postes seulement) et la catégorie de commune de résidence (code C.C.D. de l'I.N.S.E.E.) a soutenu cette dernière hypothèse puisqu'il a produit / dans l'espace des 126 postes un premier plan factoriel à peu près identique à celui fourni par l'analyse du fichier des 37 C.S.P., mais avec de plus l'échelonnement / de chaque C.S.P. à peu près régulier sur le premier axe en fonction de la catégorie de commune. (Cf. graphique / n° 3).

Le graphique n° 3 visualise cet échelonnement parallèlement à l'axe F₁ pour chaque C.S.P.

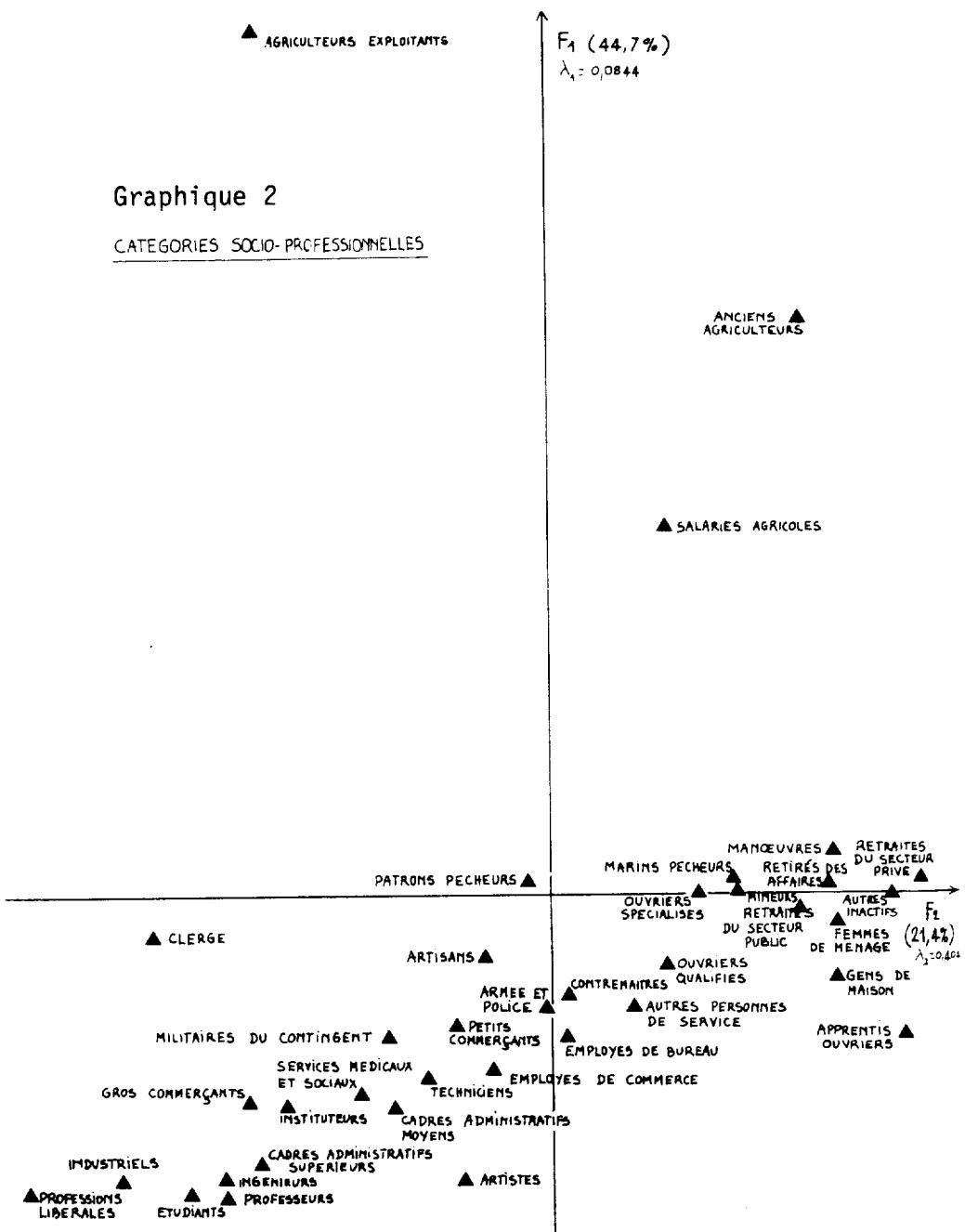
L'axe F₂ définit une échelle de statut des C.S.P. non agricoles: à l'extrême gauche se projettent les professions libérales, les industriels, les gros commerçants.. .; à l'opposé les catégories d'ouvriers, de manœuvres, de retraités.

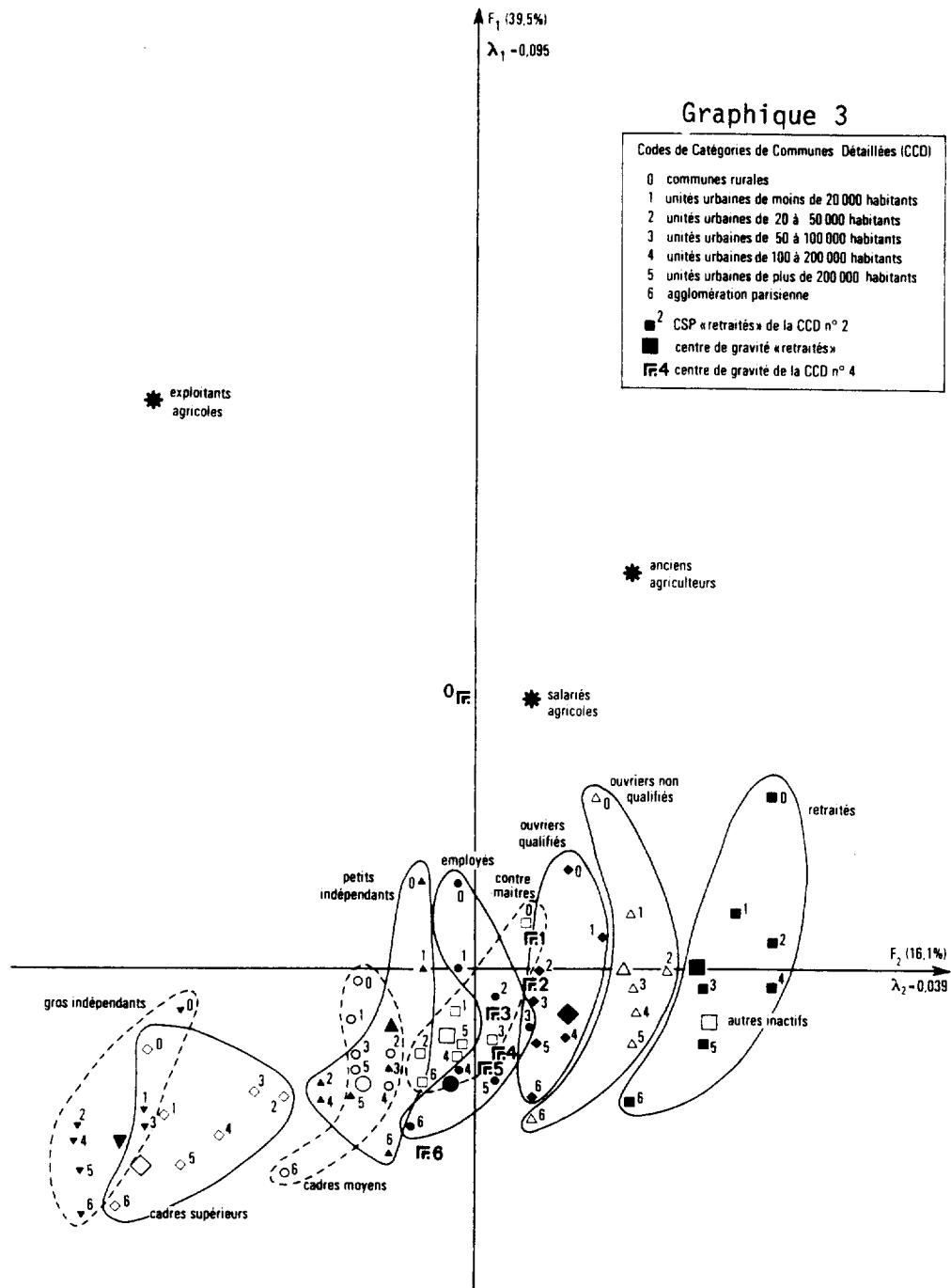
Dans l'espace J des consommations cet axe oppose les dépenses de alimentaires (symbole •) aux dépenses de vacances, d'employé de maison et de culture et loisirs / (symbole □), exception faite pour la télévision. Les dépenses alimentaires sont échelonnées entre les dépenses de première nécessité (pommes de terre, lait...) du côté des plus bas statuts et des produits plus élaborés situés vers le centre de gravité. Nous reprendrons cet exemple après avoir exposé l'Analyse des correspondances.

Graphique 1



Graphique 2
CATEGORIES SOCIO-PROFESSIONNELLES





1.3 LE GAIN DE PRODUCTIVITE DANS LES DÉPOUILLEMENTS D'ENQUÊTES

En nous appuyant sur l'exemple du livre (1): *Techniques de la description statistique* L. LEBART, A. MORINEAU, N. TABARD (pages 143-148), nous abordons deux phases importantes dans l'exploitation des données d'enquêtes:

- a) le choix d'un ensemble de variables actives et / d'un ensemble de variables supplémentaires
- b) en quoi la lecture de la grille socio-administrative nous aide à sélectionner les tableaux croisés les plus significatifs.

Remarques sur l'usage des ordinateurs

L'usage des ordinateurs a permis le développement et la diversification des méthodes mais il n'est pas sans inconvénient:

- Il fournit toujours un résultat même si on a fait un mauvais choix de recodage de données ou un mauvais choix de méthode.
- Il peut donner lieu à des emplois abusifs.
- La facilité d'emploi ne doit pas nous faire oublier/les règles d'interprétation.

2. A QUOI S'APPLIQUE L'ANALYSE DE DONNÉES

2.1 DOMAINES D'APPLICATION

Le domaine d'application est très vaste: socio-économie, démographie, agriculture, médecine, marketing ... / comme nous le verrons lors des exemples exposés. Ces méthodes sont maintenant très répandues dans les services publics et dans les organismes privés français.

En France, de nombreux développements théoriques / sont faits dans l'environnement du Professeur BENZECRI, promoteur de l'Analyse factorielle des correspondances / (voir la revue "Les cahiers de l'Analyse des Données" BP 113; 93103 MONTREUIL CEDEX, FRANCE).

Elles sont de plus en plus utilisées dans de nombreux autres pays. Parmi les congrès récents sur ce thème, citons:

- Les troisièmes journées internationales Analyse de Données et informatique du 4 au 7 d'octobre 1983 à VERSAILLES (FRANCE) organisées par / l'INRIA (Institut national de recherche en informatique et en automatique).
- LAS JORNADAS DE ANALISIS DE DATOS, ENCUENTRO FRANCIA-AMERICA LATINA de 11 al 22 de JULIO de 1983 en CARACAS (VENEZUELA)

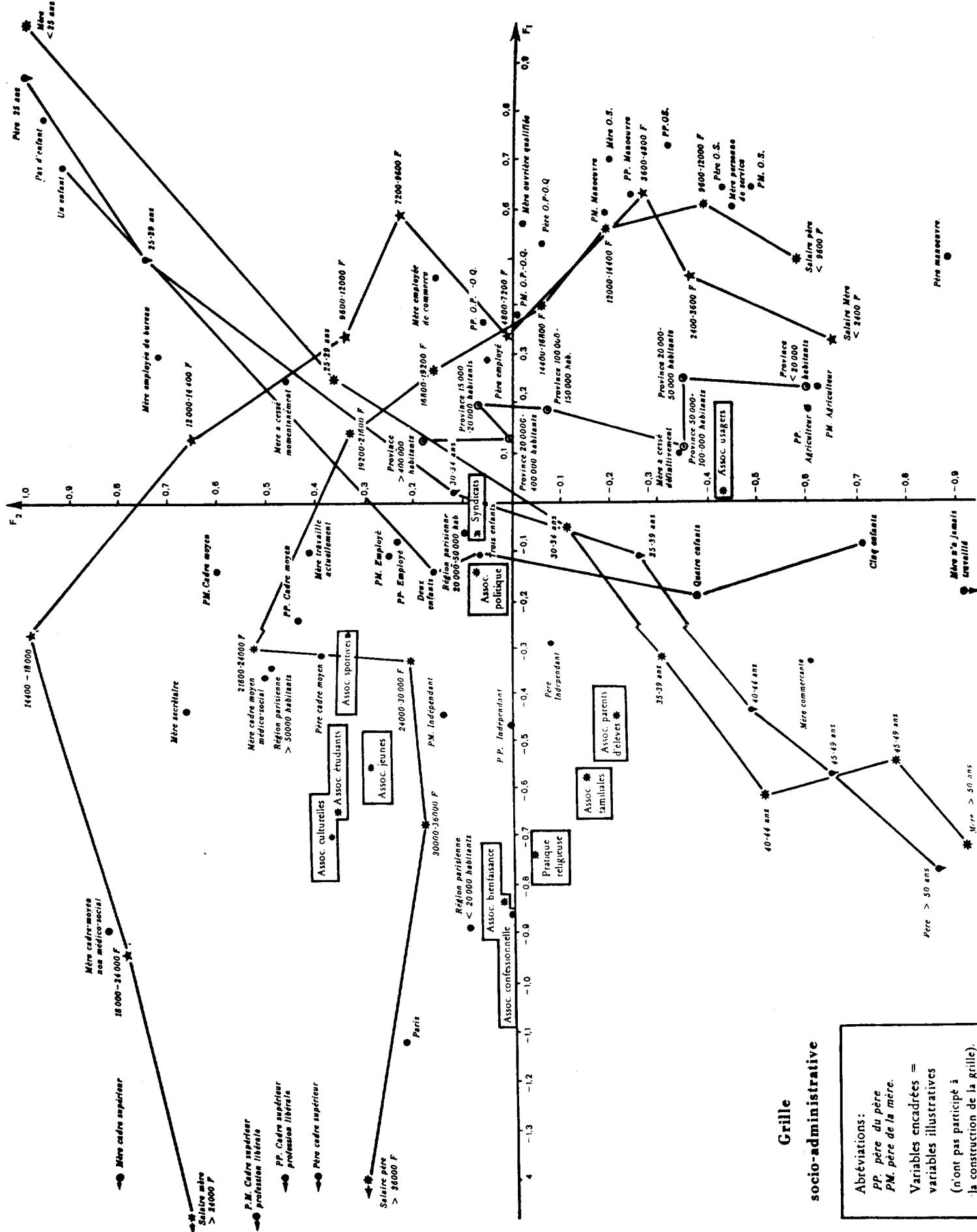
2.2 MÉTHODES ADAPTÉES AUX GRANDS TABLEAUX

Ces méthodes ne sont vraiment utiles que si les dimensions du tableau à analyser sont un obstacle à sa lecture et à son assimilation directe. Au CREDOC l'expérience montre qu'il s'agit le plus souvent d'obtenir une description de nature statistique pour un certain phénomène qui a donné lieu au recueil de mesures ou observations / trop nombreuses et dépendantes les unes des autres pour être interprétables en première lecture. Les fichiers / d'enquêtes socio-économiques sont des exemples types. / Pour les traiter le recours à l'analyse de données est particulièrement profitable.

2.3 LES QUALITÉS D'UN TABLEAU

Il y a trois qualités de base:

- *La pertinence:* Ceci signifie que l'objet du relevé est bien défini et que le problème a un sens.
- *L'exhaustivité:* Les différentes zones du domaine / d'investigation sont bien recouvertes par le relevé.
- *L'homogénéité:* Il faut s'interroger sur les sens/ des opérations pratiquées sur le tableau: en lignes et en colonnes. Il faut s'imaginer ce que représente l'ensemble des termes d'une ligne pour choisir un ensemble / homogène de variables actives (celles qui contribueront à la construction des axes principaux).



3. EN QUOI CONSISTE L'ANALYSE DES DONNEES

3.1 BIBLIOGRAPHIE

Pour les livres les plus cités on conviendra d'apporter dans le texte les sigles suivants (le nom de l'éditeur figure entre parenthèses):

L. LEBART; A. MORINEAU; N. TABARD

"Techniques de la description statistique. Méthodes et logiciels pour l'analyse des grands tableaux"
(DUNOD) Paris, 1977

L. LEBART; A. MORINEAU; J.P. FENELON

"Traitement des données statistiques"
(DUNOD) Paris, 1982, 2^e édition

* Notons que le livre a été traduit en espagnol et / qu'il sera prochainement édité.

L. LEBART; A. MORINEAU

SPAD

"Système portable pour l'analyse des données"

P. PLEUVRET; E. BRIAN; T. ALUYA; L. LEBART, A. MORINEAU
SPAD (TOME II)
(CESIA) 82 Rue de SEVRES
75007, PARIS

BENZECRI et autres

"L'Analyse des données"

Tome 1: "La Taxinomie"

Tome 2: "L'Analyse des correspondances"

(DUNOD) Paris, 1976

BENZECRI, F. BENZECRI

"La pratique de l'Analyse des Données" 3 vol.

(DUNOD) Paris, 1980

L. LEBART; A. MORINEAU; WARWICK

"Multivariate descriptive statistical analysis"

(WILEY) New York, 1983

R. CEHESSAT

"Exercices commentés de statistique et informatique appliquées"

(DUNOD) Paris, 2^e édition, 1981

J.P. MAKACHE; A. CHEVALIER; V. MORICE

"Exercices commentés de mathématique pour l'analyse statistique des données"
(DUNOD) Paris, 1981

J.P. FENELON

"Qu'est-ce que l'analyse des données?"
LEFONEN édition, 26 Rue des Cordelières,
75013 Paris, 1981

LAGARDE

"Initiation à l'analyse des données"
(DUNOD) 1983

J.M. BOUROCHE; G. SAPORTA

"L'analyse des données" PUF, Collection "Que sais-je?"
Paris

F. CAILLEZ; J.P. PAGES

"Introduction à l'analyse des données" SMASH, 9^e Rue Duban
75016 Paris, 1976

G. SAPORTA

"Théorie et méthodes de la statistique"
(TECHNIP) Paris, 1978

CUADRAS

"Análisis multivariante"
Barcelona

VIC BARNETT (ed.)

"Interpreting Multivariate Data"
(WILEY) London

E. DIDAY et autres

"Eléments d'analyse des données"
(DUNOD) Paris, 1983

M. JAMBU; M.O. LEBEAUX

"Classification automatique pour l'analyse des données"

Tome 1: "Méthodes et algorithmes"

Tome 2: "Logiciels"

(DUNOD) Paris, 1978

BERTIER, BOUROCHE

"Analyse des données multidimensionnelles"
(MASSON) Paris

3.2 LES DIFFERENTES MÉTHODES FACTORIELLES ET LES DIFFERENTS TYPES DE TABLEAUX

3.2.A Les trois principaux types de tableaux sont les suivants:

3.2.A.1. Les tableaux de mesures:

Pour chacun des individus d'un ensemble I on a observé un certain nombre de mesures J :

- La taille en centimètres
- Le poids en kilogrammes
- L'âge
- etc.

On désigne par *observations* les lignes I du tableau, et par *variables* les colonnes J. Ces mesures sont des valeurs numériques continues, dans la suite on les appellera *Variables continues*.

On peut recoder ces mesures en classes; par exemple pour l'âge choisir sept classes:

- 1) Moins de 25 ans
- 2) de 25 à 29 ans
- 3) de 30 à 34 ans
- 4) de 35 à 39 ans
- 5) de 40 à 44 ans
- 6) de 45 à 49 ans
- 7) 50 ans et plus

On l'appellera alors *Variable nominale ayant 7 modalités*.

3.2.A.2. Les tableaux de contingence

Il existe une seule population. Par exemple les - /

31079 sujets d'une enquête. On peut les classer d'un premier point de vue en dix Catégories socioprofessionnelles. On peut aussi les classer d'un second point de vue J en huit modes d'hébergement en vacances. On peut alors procéder à des comptages: le nombre de sujets possédant à la fois la caractéristique de loger en village de vacances et d'appartenir à la catégorie des employés: il y en a 178. C'est le nombre qui se trouve à l'intersection de la ligne *employés* et de la colonne *village de vacances*. (Voir le tableau croisé 1 de 3.3.3 où à un couple (i,j) correspond un effectif d'individus K_{ij})

3.2.A.3 Un fichier d'enquête

Dans une enquête la plupart des questions sont précodées sous cette forme. On parle de forme disjonctive/complète, c'est à dire que les diverses modalités de réponse à une question s'excluent mutuellement et une seule modalité est obligatoirement choisie par l'enquête. Exemple de question:

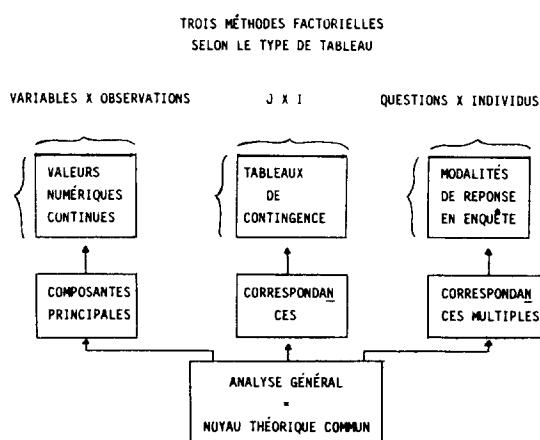
Possédez vous au moins une voiture?

- 1) OUI
- 2) NON
- 3) Sans réponse

Pour un tel fichier on parlera d'individus I ayant répondu à un ensemble de questions.

3.2.B Les diverses méthodes factorielles

Nous schématisons le choix d'une des trois méthodes/ selon le type de tableau à la suite:



3.3 PRINCIPES DES MÉTHODES FACTORIELLES ET EXEMPLES ILLUSTRATIFS

Le schéma du paragraphe précédent résume l'ordre / dans lequel on a exposé les méthodes factorielles, à / l'aide de transparents faits par le professeur MORINEAU.

Tous les points essentiels ont été expliqués car il est indispensable de connaître et de comprendre les opérations effectuées sur les données pour pouvoir interpréter les résultats. Cependant, toutes les démonstrations/ n'ont pas été faites lors des conférences, dans ce cas nous donnons les références bibliographiques nécessaires, principalement celles du livre (2) de L. LEBART, A. MORINEAU et J.P. FENELON "Traitement des données statistiques" (éditeur DUNOD). Il sera prochainement édité en espagnol.

3.3.1 Analyse générale: (AG)

C'est le noyau théorique commun aux trois méthodes / factorielles, on résoud le problème d'approximation numérique suivant: étant donné un tableau rectangulaire de / valeurs numériques représenté par une matrice X à n lignes et p colonnes, de terme général x_{ij} est-il possible de reconstituer les $n.p$ valeurs x_{ij} à partir d'un plus petit nombre de valeurs numériques.

Le tableau X donne lieu à deux représentations géométriques: les n lignes de X peuvent être considérées comme les coordonnées de n points dans un espace à p dimensions \mathbb{R}^p ; où les p colonnes de X peuvent représenter les coordonnées de p points dans un espace à n dimensions \mathbb{R}^n . On cherche à reconstituer approximativement les positions des points d'un nuage de \mathbb{R}^p (ou de \mathbb{R}^n) à partir des coordonnées de ces points sur la base d'un sous espace qui ajuste au mieux ce nuage au sens des Moindres-Carrés. On cherche un ajustement dans \mathbb{R}^p , puis dans \mathbb{R}^n ; et les relations entre les sous-espaces d'ajustement. On obtient la formule de reconstitution exacte du tableau / des données numériques X , puis en se limitant à une reconstitution approchée on mesure la qualité de l'approximation.

Les méthodes d'analyse en composantes principales et d'analyse des correspondances sont présentées à partir / du formalisme de l'analyse générale. Mais lorsqu'il - / s'agit d'analyse statistique (et non plus d'approximation numérique) on dispose le plus souvent d'information complémentaires sur la nature des données; la prise en

compte de ces informations conduit à procéder à des - / transformations préalables sur le tableau de départ.

3.3.2 Analyse en composantes principales (CP)

On l'utilise lorsqu'il s'agit de décrire un tableau R de valeurs Numériques continues du type Individus & variables. On présentera en fait toutes les étapes de l'analyse sur un tableau de dimensions réduites, en gardant à l'esprit le fait que la méthode / de description n'est vraiment utile que si les dimensions du tableau sont précisément un obstacle à sa lecture et à son assimilation directe. Le tableau R qui servira à illustrer les différentes phases de calcul donne les dépenses annuelles de $n = 12$ catégories de ménages (ces données sont des moyennes par catégories) pour $p = 7$ postes de consommation. Les 12 identificateurs en 3 caractères nous informent sur le nombre d'enfants du ménage (2, 3, 4 ou 5) et sur les caractéristiques professionnelles sommaires du chef de ménage (MA = Travailleur manual, EM = Employé non manual, CA = Cadre). Le niveau général des dépenses est très différent selon le poste de consommation, autrement dit, les variables sont très hétérogènes quant à leurs moyennes.

Analyse dans \mathbb{R}^p

Ce n'est pas la position du nuage par rapport à l'origine qui nous intéresse, mais la forme de ce nuage. Ceci nous conduit à prendre comme nouvelle origine dans \mathbb{R}^p le centre de gravité G (ou point moyen) du nuage / dont les p composantes sont les p moyennes arithmétiques \bar{r}_j . L'analyse en composantes principales sera un cas particulier de l'analyse générale du tableau Y après / la transformation suivante du tableau R de départ:

$$R \longrightarrow Y = (Y_{ij}) \\ Y_{ij} = \frac{1}{\sqrt{n}} (r_{ij} - \bar{r}_j)$$

L'influence du niveau général de chacune des variables est ainsi éliminée. Le coefficient $1/\sqrt{n}$ n'a pour effet que de faire coincider la matrice à diagonaliser YY' avec la matrice des covariances expérimentales.

Une modification supplémentaire du tableau de départ

peut également être nécessaire si les dispersions typiques des variables sont très différentes; elle conduira à l'analyse en composantes principales normée, c'est à dire à l'analyse du tableau transformé:

$$R \longrightarrow X = (x_{ij})$$

$$x_{ij} = \frac{1}{\sqrt{n}} \frac{r_{ij} - \bar{r}_j}{s_j}$$

$$\text{avec } s_j^2 = \frac{1}{n} \sum_i (r_{ij} - \bar{r}_j)^2$$

En effet dans le calcul du carré de la distance entre deux individus i et i'

$$d^2(i, i') = \sum_{j=1}^p \left[\frac{1}{\sqrt{n}} \frac{r_{ij} - \bar{r}_j}{s_j} - \frac{1}{\sqrt{n}} \frac{r_{i'j} - \bar{r}_j}{s_j} \right]^2$$

la division de chaque terme de la somme par la variance s_j correspondante réussira à réduire l'effet des variables très dispersées sur la valeur de la distance. Ainsi chaque variable aura une contribution analogue dans la détermination des proximités : la matrice à diagona-

liser $X'X=C$ coincide avec la matrice des corrélations C entre les variables.

Analyse dans \mathbb{R}^n :

L'analyse générale du tableau transformé X dans \mathbb{R}^p induit comme on l'a vu une analyse dans \mathbb{R}^n . Cependant / les indices i et j ne jouant pas des rôles similaires, dans cette transformation les interprétations associées à cette transformation seront radicalement différentes. La distance entre deux points variables j et j' s'écrit: $d^2(j, j') = 2(1-\text{cor}(j, j'))$, où $\text{cor}(j, j')$ est le coefficient de corrélation entre les variables j et j' . Ainsi, les proximités entre points-variables pourront s'interpréter en termes de corrélations:

- Si leur corrélation est fortement positive alors les points sont très voisins:

$$(\text{cor}(j, j') \approx 1) \Rightarrow (d^2(j, j') \approx 0)$$

- Si leur corrélation est fortement négative alors les points sont éloignés:

$$(\text{cor}(j, j') \approx -1) \Rightarrow (d^2(j, j') \approx 4)$$

L'exemple numérique et règles d'interprétation

Tableau R:

Dépenses annuelles de 12 catégories de ménages pour 7 postes de consommation.

	Pain	Légumes	Fruits	Viandes	Volailles	Lait	Vin
MA2(*)	332	428	354	1 437	526	247	427
EM2	293	559	388	1 527	567	239	258
CA2	372	767	562	1 948	927	235	433
MA3	406	563	341	1 507	544	324	407
EM3	386	608	396	1 501	558	319	363
CA3	438	843	689	2 345	1 148	243	341
MA4	534	660	367	1 620	638	414	407
EM4	460	699	484	1 856	762	400	416
CA4	385	789	621	2 366	1 149	304	282
MA5	655	776	423	1 848	759	495	486
EM5	584	995	548	2 056	893	518	319
CA5	515	1 097	887	2 630	1 167	561	284
Moyennes	446,7	737,8	505,0	1 886,7	803,2	358,2	368,6
Ecart-types	102,6	172,2	158,1	378,9	238,9	112,1	68,7

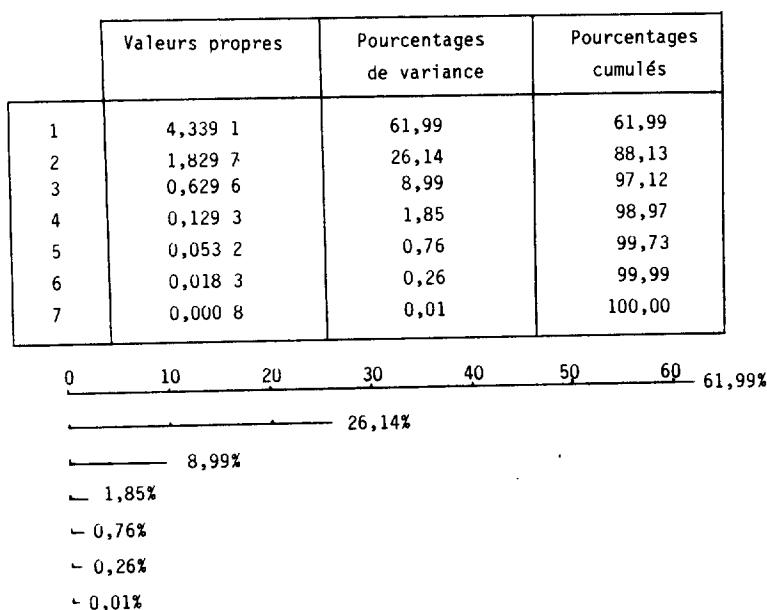
(*) Les identificateurs en 3 caractères nous informent sur le nombre d'enfants du ménage (2,3,4, ou 5) et sur les caractéristiques professionnelles sommaires de chef de ménage (MA = travailleur manuel, EM = employé/non manuel, CA = cadre). En réalité ces données sont elles-mêmes des moyennes par catégories (cf. N. Tabard et al., 1967)

L'analyse nous conduit à diagonaliser la matrice C des corrélations:

	Pain	Légumes	Fruits	Viandes	Volailles	Lait	Vin
Pain	1,00						
Légumes	0,59	1,00					
Fruits	0,20	0,87	1,00				
Viandes	0,32	0,89	0,96	1,00			
Volailles	0,25	0,83	0,93	0,98	1,00		
Lait	0,86	0,66	0,33	0,37	0,23	1,00	
Vin	0,30	-0,35	-0,49	-0,44	-0,40	0,01	1,00

La seule lecture de cette matrice (7x7) nous donne une idée du réseau d'interrelations existant entre les variables. On conçoit facilement que dans le cas d'une matrice (100 x 100) le recours à l'analyse en composantes principales soit alors indispensable.

Le premier résultat numérique interprétable est constitué par la liste des valeurs propres et des pourcentages de variance.



La somme des valeurs propres vaut $p = 7$, trace de la matrice des corrélations. Les premiers pourcentages sont ici exceptionnellement élevés parce qu'il existe une concentration assez nette du nuage dans un sous-espace à deux dimensions (88% de variance), mais aussi parce que les dimensions du tableau sont faibles. Pour savoir si ces pourcentages sont exceptionnellement élevés, une solution empirique consisterait à simuler, puis analyser des tableaux pseudo-aléatoires d'ordre (7 x 12) (cf. = validité / des résultats libre (2) de la bibliographie).

On se limitera ici aux trois premières dimensions pour le calcul des / coordonnées des variables et des individus sur les axes factoriels. Les résultats sont ici:

Variables	Facteur 1 $u_1 \sqrt{\lambda_1}$	Facteur 2 $u_2 \sqrt{\lambda_2}$	Facteur 3 $u_3 \sqrt{\lambda_3}$
Pain	-0,497	0,841	-0,012
Légumes	-0,972	0,131	-0,037
Fruits	-0,931	-0,277	0,112
Viandes	-0,963	-0,190	0,161
Volailles	-0,912	-0,265	0,280
Lait	-0,584	0,707	-0,357
Vin	0,425	0,649	0,620
Observations (ou individus)	cXu_1	cXu_2	cXu_3
MA2	1,078(*)	-0,133	0,154
EM2	0,761	-0,711	-0,512
CA2	0,050	-0,287	0,565
MA3	0,820	0,124	-0,039
EM3	0,681	-0,067	-0,199
CA3	-0,669	-0,534	0,393
MA4	0,378	0,539	-0,109
EM4	0,107	0,248	0,108
CA4	-0,629	-0,686	0,037
MA5	-0,083	1,096	0,222
EM5	-0,777	0,448	-0,390
CA5	-1,718	-0,037	-0,230

(*) Les coordonnées sont calculées à une homothétie près, définie par la constante c (cf. remarque 5).

Remarque 1.- Les coordonnées des points variables sont inférieures à 1 en valeur absolue; en effet ces points sont à la distance / 1 de l'origine dans \mathbb{R}^n , et l'opération de projection est / contractante; elle ne peut que diminuer les distances.

Remarque 2.- Les coordonnées sont définies au signe près sur chaque axe. L'orientation des axes est arbitraire et dépend de l'algorithme de diagonalisation utilisé.

Remarque 3.- On sait que l'analyse du nuage des individus se fait par rapport à leur centre de gravité. Il n'est pas de même pour les points variables; la simple consultation de leurs coordonnées sur le premier axe nous montre que (à l'exception du vin) tous les produits sont d'un même côté de l'origine. Une telle disposition traduit le fait que la plupart des variables sont en fait corrélées positivement entre elles. Si pour un individu une variable prend une valeur forte, toutes les autres variables prennent une valeur forte. Cette caractéristique

téristique de nombreux tableaux apparaît le plus souvent sur le premier axe, qu'on appellera *facteur de taille*. Ici le facteur de taille oppose les gros consommateurs aux petits consommateurs.

Remarque 4.- Un point de la sphère de rayon unité dans \mathbb{R}^n représente une variable (dont les n composantes sont normalisées). La valeur du produit scalaire des vecteurs joignant l'origine à deux points de cette sphère n'est autre que le coefficient de corrélation des variables correspondantes (c'est également le cosinus de l'angle des deux vecteurs). Ainsi, les coordonnées des points variables sur le premier axe ne sont autres que les coefficients de corrélation entre les diverses variables et le premier facteur, considéré lui-même comme variable artificielle (c'est une combinaison linéaire des variables initiales). Par exemple, la variable pain a un coefficient de corrélation de -0,497 avec la variable: facteur 1 (dont les 12 valeurs figurent au bas de la première colonne), ou, ce qui revient au même, fait un angle dont le cosinus vaut -0,497 avec le premier axe. On pourra chercher à interpréter chaque axe en fonction de la combinaison des variables les plus corrélées au facteur correspondant.

Remarque 5.- Si l'échelle des coordonnées des points variables a une interprétation en termes de corrélations, il n'en est pas de même pour les points individus. On applique à leurs coordonnées X_{α} un coefficient dont la valeur $c = \sqrt{n}/p$ est déterminée pour assurer un positionnement dans le plan compatible avec la répartition des points-variables, et permettre ainsi une représentation simultanée des deux nuages.

La figure 1 donne une représentation graphique des points variables dans le premier plan factoriel (axes 1 et 2). La plupart des points sont très proches du cercle d'intersection de la sphère de \mathbb{R}^{12} avec le plan factoriel. La représentation correspondante pour les individus est donnée par la figure 2.

Sur cette figure, on a joint d'une part les points correspondant à des ménages ayant même nombre d'enfants (2,3,4, ou 5 et plus), et d'autre part les points correspondant aux individus ayant même activité (Manuels, Employés autres que Manuels, Cadres). La grille ainsi constituée (déformée du côté des cadres) montre que le facteur de taille mis en évidence sur la figure 1 est lié d'une part à la taille du ménage, d'autre part à son statut socio-professionnel : la dépense globale augmente avec le nombre des enfants et le statut social.

La superposition des figures 1 et 2 avec des précautions d'interprétation rend plus vivante la visualisation, en suggérant quelles variables sont responsables des proximités. On y verrait par exemple que le montant des dépenses en pain, lait et surtout vin rapproche les employés manuels et non manuels, et les éloignent des cadres.

Analyse des données . méthodes factorielles

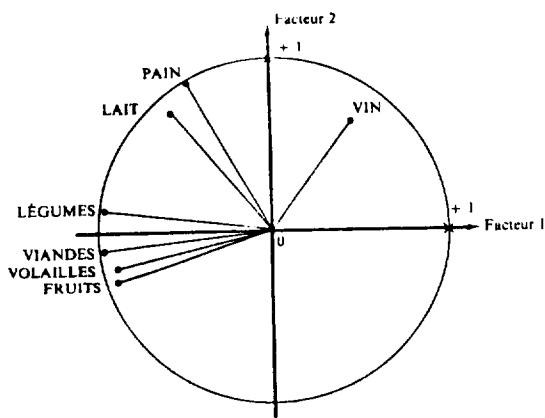


Fig. 1

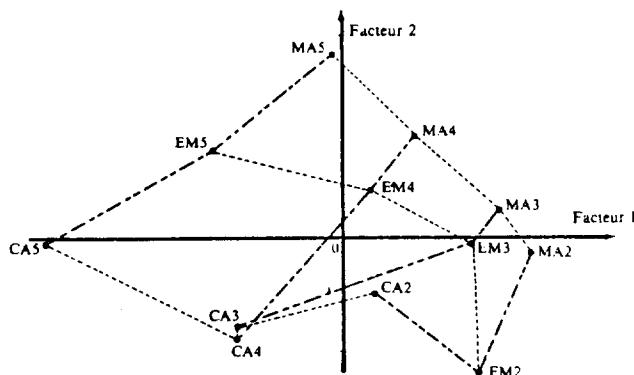


Fig. 2

INTERPRETATION DES AXES

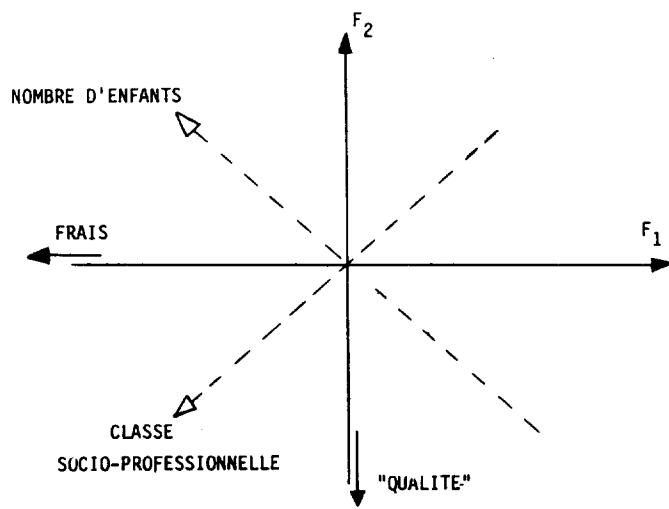


Fig. 3

3.3.3 Analyse des correspondances (AC)

Cette méthode est dévolue plus particulièrement à l'analyse des tableaux de contingence. On prendra comme exemple le tableau 1 qui croise 10 catégories socio-professionnelles avec 8 modes d'hébergement en vacances. A l'intersection de la ligne i et de la colonne j se trouve le nombre k_{ij} d'individus appartenant à la classe correspondante. Ici encore cet exemple de dimensions réduites (10×8) illustrera l'exposé, alors que la méthode de description est surtout utile pour des tableaux plus vastes.

Dans ce tableau, les lignes et les colonnes qui représentent deux partitions d'une même population jouent des rôles analogues. Pour que les distances entre points lignes et points-colonnes aient un sens, il sera utile de faire intervenir les profils des lignes et des colonnes, c'est à dire les fréquences conditionnelles. On utilise les notations suivantes:

$$K = \sum_i \sum_j K_{ij} \text{ effectif total dans le tableau}$$

$$f_{ij} = K_{ij}/K \text{ fréquences relatives}$$

$$\begin{aligned} f_{i.} &= \sum_j f_{ij} \\ f_{.j} &= \sum_i f_{ij} \end{aligned} \quad \text{fréquences relatives marginales}$$

A le tableau 2 figurent les fréquences conditionnelles en pourcentage ($100 \times f_{ij}/f_{i.}$) ou profils-lignes. A le tableau 3 figurent les quantités ($100 \times f_{ij}/f_{.j}$) ou profils-colonnes. Ces quantités définissent les coordonnées dans les deux espaces (à ceci près que l'on utilisera des proportions et non des pourcentages).

Construction des nuages de points:

Dans l'espace \mathbb{R}^p on construira un nuage de n points, chaque point i ayant pour coordonnées

$$\left\{ \frac{f_{ij}}{f_{i.}} \right\}_{j=1 \dots p}$$

Dans l'espace \mathbb{R}^n on construira un nuage de points, chaque point j ayant pour coordonnées:

$$\left\{ \frac{f_{ij}}{f_{.j}} \right\}_{i=1 \dots n}$$

Les proximités entre points s'interpréteront alors

en termes de proximités entre profils: A le tableau 2 on compare les différences de comportement des 10 catégories socio-professionnelles vis à vis de leur mode d'hébergement en vacances. A le tableau 3 on compare les modes d'hébergement selon leurs profils socio-professionnelles.

Remarque: Les n points de \mathbb{R}^p sont situés dans un - sous espace à $p-1$ dimensions puisque pour tout point i les coordonnées vérifient la relation

$$\sum_j \left\{ \frac{f_{ij}}{f_{i.}} \right\} = 1$$

Les p points de \mathbb{R}^n sont également situés dans le - sous-espace à $n-1$ dimensions puisque pour tout point j les coordonnées vérifient la relation:

$$\sum_i \left\{ \frac{f_{ij}}{f_{.j}} \right\} = 1$$

Choix des distances:

Le fait de travailler sur des profils dans les deux espaces \mathbb{R}^p et \mathbb{R}^n nous incite à choisir la distance du "CHI-2": la distance entre deux catégories i et i' sera

$$d^2(i, i') = \sum_{j=1}^p \left\{ \frac{1}{f_{.j}} \left(\frac{f_{ij}}{f_{i.}} - \frac{f_{i'j}}{f_{i'.}} \right)^2 \right\}$$

De même la distance entre deux modes d'hébergement j et j' sera:

$$d^2(j, j') = \sum_{i=1}^n \left\{ \frac{1}{f_{i.}} \left(\frac{f_{ij}}{f_{.j}} - \frac{f_{ij'}}{f_{.j'}} \right)^2 \right\}$$

Cette distance pondérée (le point i est affecté de la masse $f_{i.}$, le point j de masse $f_{.j}$) a surtout - l'avantage de vérifier le principe d'équivalence distributionnelle: si deux points lignes i_1 et i_2 sont confondus dans \mathbb{R}^p et si on les considère comme un seul point affecté de la somme des masses de i_1 et de i_2 (i_1 et i_2 seront remplacés par i_0) alors les distances entre tous les couples de points dans \mathbb{R}^p et dans \mathbb{R}^n restent inchangées. Il en est de même pour deux points-colonnes j_1 et j_2 de \mathbb{R}^n ayant les mêmes propriétés.

Cette propriété est très importante car elle pourra

avoir une influence stabilisatrice sur les résultats. Ainsi, l'agrégation de postes ayant des profils voisins (en lignes ou en colonnes) ne bouleversera pas en général les résultats obtenus. Ceci nous garantit dans une certaine mesure contre l'arbitraire de toute nomenclature.

Analyse dans \mathbb{R}^p

La distance du "CHI-2" écrite précédemment n'étant pas une somme de carrés, le problème n'entre pas dans le cadre de l'analyse en composantes principales. On se ramène cependant à ce cas en modifiant l'échelle des axes. Si on choisit de prendre pour les p coordonnées du point i les quantités:

$$\frac{f_{ij}}{\sqrt{f_{.j}}}$$

alors la distance euclidienne usuelle de deux points i et i' vaut:

$$d^2(i, i') = \sum_{j=1}^p \left\{ \frac{f_{ij}}{\sqrt{f_{.i}}} - \frac{f_{i'j}}{\sqrt{f_{.j}}} \right\}^2$$

et on vérifie aisément que cette distance coincide avec la distance du "CHI-2". On est donc ramené à une analyse simple en prenant comme coordonnées des points du nuage les quantités $f_{ij}/f_{.j}, \sqrt{f_{.j}}$. Le nuage est maintenant dans un hyperplan d'équation:

$$\sum_{j=1}^p \sqrt{f_{.j}} y_j = 1$$

Nous devons maintenant chercher la première composante principale du nuage de points ainsi construit, puisque nous nous intéressons à la forme de ce nuage, et non à sa position par rapport à l'origine.

Calculons les coordonnées du centre de gravité du nuage dans \mathbb{R}^p . Puisque chaque point i est affecté de la masse $f_{.i}$, la j -ième composante du centre de gravité G s'écrit:

$$g_j = \sum_{i=1}^n f_{.i} (f_{ij}/f_{.i} \cdot \sqrt{f_{.j}}) = \sqrt{f_{.j}}$$

Après translation de l'origine au centre de gravité les coordonnées du point i s'écrivent $(f_{ij}/f_{.i} \cdot \sqrt{f_{.j}} - \sqrt{f_{.j}})$. Ce point i est muni de la masse $f_{.i}$, qui intervient dans le critère d'ajustement du sous-espace porté par le vecteur unitaire u . Si l'on désigne par

$\hat{\psi}_i$, la projection du point i sur cet axe u on aura :

$$\hat{\psi}_i = \sum_j (f_{ij}/f_{.i} \cdot \sqrt{f_{.j}} - \sqrt{f_{.j}}) u_j$$

La quantité à rendre maximale pour la détermination de u sera : $\sum_i f_{.i} \hat{\psi}_i^2$. Ainsi on est conduit à diagonaliser la matrice T des covariances, dont le terme général s'écrit :

$$t_{jj'} = \sum_{i=1}^p f_{.i} (f_{ij}/f_{.i} \cdot \sqrt{f_{.j}} - \sqrt{f_{.j}}) (f_{ij'}/f_{.i} \cdot \sqrt{f_{.j'}} - \sqrt{f_{.j'}})$$

Il est possible de donner à cette matrice T une forme simple. Posons en effet:

$$x_{ij} = (f_{ij} - f_{.i} f_{.j}) / \sqrt{f_{.i} f_{.j}}$$

Alors la matrice T à diagonaliser s'exprime en fonction du tableau (n,p) noté X :

$$T = X^T X$$

L'analyse des correspondances se ramène à une analyse du Tableau X , remarquons que la transformation des données fait intervenir de façon symétrique les indices i et j . La transformation utilisée pour l'analyse en composantes principales était dissymétrique.

Analyse dans \mathbb{R}^n

Les ensembles mis en correspondances dans le tableau des fréquences jouent des rôles analogues: l'analyse dans \mathbb{R}^n peut donc se déduire de celle menée dans \mathbb{R}^p par permutation des rôles des indices i et j .

Ainsi, les coordonnées du point j seront maintenant les quantités $f_{ij}/f_{.j} \sqrt{f_{.i}}$. Ce point j sera muni de la masse $f_{.j}$. La i -ième coordonnée du centre de gravité H du nuage des p points s'écrit:

$$h_i = \sqrt{f_{.i}}$$

La matrice de covariance à diagonaliser aura pour terme général:

$$w_{ii'} = \sum_{j=1}^p f_{.j} (f_{ij}/f_{.j} \sqrt{f_{.i}} - \sqrt{f_{.i}}) (f_{i'j}/f_{.j} \sqrt{f_{.i'}} - \sqrt{f_{.i'}})$$

La même transformation que précédemment nous ramène à une analyse générale du tableau X de terme général

$$x_{ij} = (f_{ij} - f_{i.} f_{.j}) / \sqrt{f_{i.} f_{.j}}$$

Alors la matrice W s'exprime en fonction du tableau X:

$$W = X \cdot X'$$

Relation entre les deux espaces \mathbb{R}^p et \mathbb{R}^n

On se rapportera à la bibliographie pour la démonstration des relations de double transition entre les coordonnées des points - lignes et des points-colonnes sur un axe α :

$$\hat{\psi}_{\alpha i} = \frac{1}{\sqrt{\lambda \alpha}} \sum_{j=1}^p (f_{ij}/f_{i.}) \hat{\varphi}_{\alpha j}$$

Dans cette relation la matrice de terme général $(f_{ij} / f_{i.})$, permettant de calculer les coordonnées d'un point i à partir des coordonnées de tous les points j, n'est autre que la tableau des profils - lignes (divisés par 100) du tableau 2. Ainsi, dans notre exemple, la coordonnée d'une profession s'obtient comme produit par $(1/\sqrt{\lambda \alpha})$ de l'abscisse sur l'axe α du barycentre de tous les modes d'hébergement, les poids étant les éléments du profil des hébergements de cette profession.

De la même façon la relation suivante:

$$\hat{\varphi}_{\alpha j} = \frac{1}{\sqrt{\lambda \alpha}} \sum_{i=1}^n (f_{ij}/f_{.j}) \hat{\psi}_{\alpha i}$$

montre que la coordonnée d'un point mode d'hébergement est l'homothétique des points-professions, avec pour poids les éléments du profil socio-professionnel de ce mode d'hébergement, c'est-à-dire les colonnes du tableau 3.

Remarques. Ces deux relations ne sont pas des cas particuliers de celles établies lors de l'analyse générale:

$$u_{\alpha} = \frac{1}{\sqrt{\lambda \alpha}} X' U_{\alpha}$$

$$U_{\alpha} = \frac{1}{\sqrt{\lambda \alpha}} X u_{\alpha}$$

En effet, ici les matrices de passage ne sont pas transposées l'une de l'autre.

Par ailleurs ces relations impliquent que pour tout α , on a $\lambda \alpha \leq 1$: toutes les valeurs propres sont inférieures à 1. En effet, s'il existait α tel que $\lambda \alpha > 1$ alors on aurait $1/\sqrt{\lambda \alpha} < 1$; la deuxième relation exprimerait que les $\hat{\varphi}_{\alpha j}$ sont à l'intérieur (bornes exclues) de l'intervalle couvert par les $\hat{\psi}_{\alpha i}$, alors que la première relation exprimerait l'inclusion stricte inverse, ce qui est impossible.

Autre présentation de l'analyse des correspondances:

Elle permet de montrer le caractère optimal de l'analyse des correspondances en matière de représentations simultanées. C'est là recherche de la meilleure représentation simultanée de deux ensembles lignes et colonnes.

Aides à l'interprétation:

Pour interpréter les axes déterminés lors d'une analyse des correspondances, nous calculerons deux séries de coefficients pour chacun des éléments des deux ensembles mis en correspondance:

1) les contributions absolues, qui expriment la part prise par un élément donné dans la variance expliquée par un facteur;

2) les contributions relatives, ou corrélations élément-facteur, qui expriment la part prise par un facteur dans l'explication de la dispersion d'un élément.

Alors que les contributions absolues permettront de savoir quelles variables sont responsables de la construction d'un facteur, les contributions relatives exposeront celles qui sont des caractéristiques exclusives de ce facteur.

L'utilisation de ces coefficients sera précisée à l'occasion de l'exemple numérique.

Exemple numérique

Lors de l'analyse du tableau k_{ij}

Tabl. 1

Catégorie socio-professionnelle	Mode d'hébergement									TOTAL
	Hôtel	Pension de famille	Maison louée chez l'habitant	Maison en propriété	Chez des Parents	Chez des Amis	Tente. Caravane	Village de vacances	Divers	
Agriculteurs	160	28	0	321	36	141	45	65	796	
Salariés agricoles	35	34	1	178	8	0	4	0	260	
Patrons ind. et co.	700	354	229	959	185	292	119	140	2 978	
Cadres sup. Professions libérales	961	471	633	1 580	305	360	162	148	4 620	
Cadres moyens	572	537	279	1 689	206	748	155	112	4 298	
Employés	441	404	166	1 079	178	434	178	92	2 972	
Ouvriers	783	1 114	387	4 052	497	1 464	525	387	9 209	
Personnels de serv.	65	43	21	294	79	57	18	6	583	
Autres actifs	77	60	189	839	53	124	28	53	1 423	
Non actifs	741	332	327	1 789	311	236	102	102	3 940	
TOTAL	4 535	3 377	2 232	12 780	1 858	3 856	1 336	1 105	31 079	

Source: M. Goguel (1965)

Tabl. 2

Profils-lignes

$$\frac{f_{ij}}{f_i} \times 100$$

	Hôtel	Pension de famille	Maison louée chez l'habitant	Maison en propriété	Chez des Parents	Chez des Amis	Tente. Caravane	Village de vacances	Divers	TOTAL
Agriculteurs	20,1	3,5	0,0	40,4	4,5	17,7	5,7	8,1	100,00	
Salariés agricoles	13,4	13,1	0,4	68,5	3,1	0,0	1,5	0,0	100,00	
Patrons ind. et co.	23,5	11,9	7,7	32,2	6,2	9,8	4,0	4,7	100,00	
Cadres sup. Professions libérales	20,8	10,2	13,7	34,2	6,6	7,8	3,5	3,2	100,00	
Cadres moyens	13,3	12,5	6,5	39,3	4,8	17,4	3,6	2,6	100,00	
Employés	14,8	13,6	5,6	36,3	6,0	14,6	6,0	3,1	100,00	
Ouvriers	8,5	12,1	4,2	44,0	5,4	15,9	5,7	4,2	100,00	
Personnels de serv.	11,1	7,4	3,6	50,4	13,6	9,8	3,1	1,0	100,00	
Autres actifs	5,4	4,2	13,3	59,0	3,7	8,7	2,0	3,7	100,00	
Non actifs	18,8	8,4	8,3	45,4	7,9	6,0	2,6	2,6	100,00	
TOTAL	14,6	10,9	7,2	41,0	6,0	12,4	4,3	3,6	100,00	

Profils-colonnes

$$\frac{f_{ij}}{f_{.j}} \times 100$$

Agriculteurs	3,5	0,8	0,0	2,5	1,9	3,6	3,4	5,9	2,6
Salariés agricoles	0,8	1,0	0,0	1,4	0,4	0,0	0,3	0,0	0,8
Patrons ind. et co.	15,5	10,5	10,3	7,5	10,0	7,6	8,9	12,7	9,6
Cadres sup., Professions libérales	21,2	13,9	28,4	12,4	16,4	9,3	12,2	13,4	14,9
Cadres moyens	12,6	15,9	12,6	13,2	11,1	19,4	11,6	10,1	13,8
Employés	9,7	12,0	7,4	8,4	9,6	11,3	13,3	8,3	9,6
Ouvriers	17,3	33,0	17,3	31,7	26,8	38,0	39,3	35,0	29,5
Personnels de serv.	1,4	1,3	0,9	2,3	4,2	1,5	1,3	0,5	1,9
Autres actifs	1,7	1,8	8,5	6,6	2,9	3,2	2,1	4,8	4,6
Non actifs	16,3	9,8	14,6	14,0	16,7	6,1	7,6	9,3	12,7
TOTAL	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00

Les premiers résultats à consulter sont la trace (qui vaut 0,0839), et les valeurs propres et pourcentages d'inertie.

	Valeurs propres	Pourcentages de variance	Pourcentages cumulés
1	0,044 3	53,01	53,01
2	0,020 2	24,22	77,23
3	0,008 8	10,53	87,76
4	0,005 4	6,54	94,30
5	0,002 3	2,85	97,15
6	0,002 1	2,57	99,71
7	0,000 2	0,29	100,00

On remarque tout d'abord que le support du nuage est bien un espace à $8 - 1 = 7$ dimensions. Contrairement à l'analyse en composantes principales, la trace ici n'est pas constante; d'autre part les valeurs propres par elles-mêmes ont une interprétation: elles interviennent par le biais du coefficient $1/\sqrt{\lambda_\alpha}$ pour caractériser la qualité de la représentation simultanée.

Le tableau 4 donne les valeurs des masses relatives, des coordonnées, contributions absolues et relatives pour les trois premiers facteurs de l'analyse.

Tableau 4 Points-colonnes (modes d'hébergement)

Intitulés	Masses $f_{.j}$	Coordonnées			Contributions absolues			Contributions relatives		
		$\hat{\varphi}_{^1 j}$	$\hat{\varphi}_{^2 j}$	$\hat{\varphi}_{^3 j}$	$ca_1(j)$	$ca_2(j)$	$ca_3(j)$	$cr_1(j)$	$cr_2(j)$	$cr_3(j)$
Hôtel	0,146	0,33	0,17	0,09	35,7	22,0	12,5	0,73	0,20	0,05
Maison louée	0,109	-0,08	0,14	-0,06	1,5	10,2	4,4	0,12	0,37	0,07
Propriété	0,072	0,41	-0,13	-0,26	26,9	6,1	56,3	0,65	0,07	0,27
Parents	0,411	-0,06	-0,14	0,04	3,8	41,7	6,9	0,15	0,78	0,06
Amis	0,060	0,11	-0,01	0,12	1,7	0,0	9,1	0,20	0,00	0,21
Tente	0,124	-0,30	0,14	-0,09	25,4	12,1	10,3	0,72	0,16	0,06
Village	0,043	-0,21	0,16	-0,02	4,4	5,6	0,1	0,47	0,27	0,00
Divers	0,036	-0,08	0,11	0,03	0,5	2,2	0,5	0,07	0,12	0,01

Points-lignes (catégories socio-professionnelles)

Intitulés	Masses f_i	Coordonnées			Contributions absolues			Contributions relatives		
		$\hat{\psi}_{^1 i}$	$\hat{\psi}_{^2 i}$	$\hat{\psi}_{^3 i}$	$ca_1(i)$	$ca_2(i)$	$ca_3(i)$	$cr_1(i)$	$cr_2(i)$	$cr_3(i)$
Agriculteurs	0,026	-0,14	0,18	0,24	1,1	3,9	17,3	0,08	0,13	0,25
Salariés	0,008	-0,04	-0,39	0,35	0,0	6,1	11,4	0,00	0,34	0,27
Agricoles										
Patrons	0,096	0,21	0,18	0,06	9,5	16,0	3,5	0,52	0,40	0,04
Cadres sup.	0,149	0,33	0,02	-0,10	35,6	0,3	16,7	0,90	0,00	0,08
Cadres Moyens	0,138	-0,10	0,06	-0,07	3,2	2,4	8,3	0,33	0,11	0,17
Employés	0,096	-0,07	0,13	-0,01	1,0	8,0	0,2	0,18	0,62	0,01
Ouvriers	0,296	-0,24	-0,01	-0,01	37,3	0,1	0,2	0,97	0,00	0,00
Personnel										
Service	0,019	-0,04	-0,20	0,24	0,1	3,7	12,2	0,01	0,22	0,31
Autres actifs	0,046	0,01	-0,47	-0,13	0,0	50,7	8,8	0,00	0,86	0,07
Non actifs	0,127	0,21	-0,12	0,12	12,2	8,6	21,5	0,59	0,19	0,21

Les proximités observables sur le premier plan factoriel sont schématisées sur la figure 1. Il est clair que cette représentation est plus vivante que les tableaux 1, 2 et 3. Cependant dans le cas de recueils de données de dimensions aussi modestes, l'analyse des correspondances aura surtout une fonction illustrative.

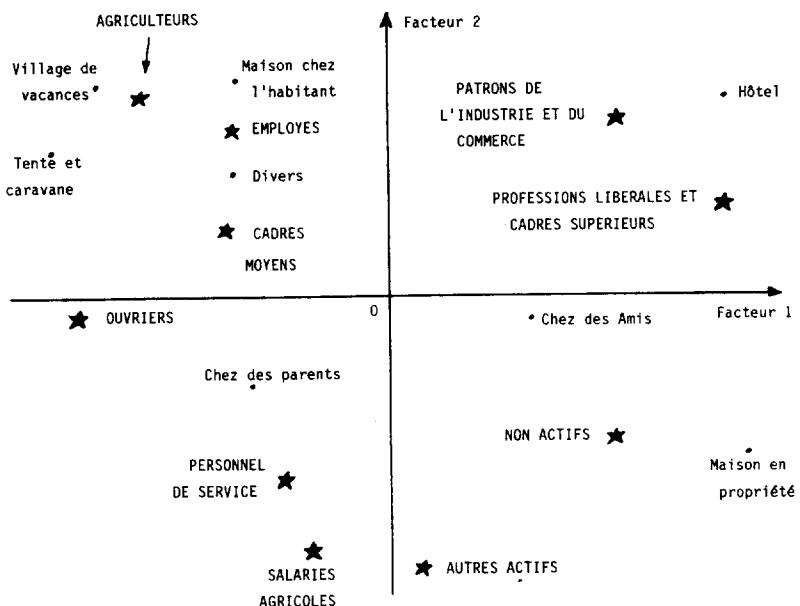


Fig. 1. Proximités entre modes d'hébergement en vacances et catégories socio-professionnelles

Nous allons, à propos de cet exemple, commenter les règles de lecture et d'interprétation de la méthode.

La proximité entre deux points-catégories socio-professionnelles signifiera une / similitude des profils d'hébergement (lignes du tableau 2). Il en est ainsi pour les Patrons et les Professions libérales, par exemple.

La proximité entre deux points-modes d'hébergement exprimera une similitude des profils socio-professionnels (colonnes du tableau 3). C'est ce que l'on constate pour par exemple: Village de vacances et Tentes et caravanes.

Enfin, la représentation simultanée des points-lignes et des points-colonnes permet d'identifier les variables responsables de certaines proximités. Ainsi, la proximité entre les profils d'hébergement des Patrons et des Professions Libérales s'explique notamment par une importance supérieure à la moyenne dans le mode Hôtel, et inférieure à la moyenne dans le mode Tente et caravane. La comparaison des deux profils- / lignes correspondants au profil moyen (ligne "total" du tableau 2) nous confirme cette observation.

De la même façon, la proximité entre Village de vacances et Tente et caravane / s'expliquera par une proportion plus forte que la moyenne d'ouvriers, ce que nous confirmé la comparaison à la colonne "total" des deux colonnes correspondantes du ta- / bleau 3.

On notera que les proximités entre profils s'interprètent toujours par référence au profil moyen. Puisque les distances à l'origine des axes (centre de gravité) s'interprètent en termes d'écart au profil moyen les proximités seront plus aisées à interpréter à la périphérie du graphique.

La lecture des contributions absolues va permettre de savoir quelles sont les variables responsables de la construction des axes. Pour le premier axe. *Hôtel. Propriétés, et tente-caravane* expliquent à elles trois 88% de la variance. Le point *village de vacances* ne participe que très peu à la construction de cet axe, contrairement à ce que pouvait laisser suggérer son abscisse: sa masse est en effet trois fois plus faible que *Tente-caravane*.

Sur le deuxième axe, *Maison en propriété et Parents* ont des abscisses voisines, / alors que la contribution de ce dernier point (41.7) dépasse largement celle du premier (6.1).

Pour les catégories socio-professionnelles, ce sont surtout les cadres supérieurs et les ouvriers qui contribuent au premier axe (resp. 35.6 et 37.3) tout en s'opposant. La lecture des contributions relatives (resp. 0.90 et 0.97) nous apprend que ces deux catégories caractérisent exclusivement ce premier axe.

On voit comment la consultation des contributions permet de nuancer la première lecture du graphique, en précisant l'effet des différences de masses (contributions absolues), et en faisant apparaître les cosinus des angles, dans l'espace, entre variables et axes (contributions relatives).

3.3.4 Analyse des correspondances multiples. (CM)

Dans une enquête la plupart des questions sont pré-codées sous forme disjonctive complète, c'est à dire que les diverses modalités de réponse à une question s'excluent mutuellement et une seule modalité est obligatoirement choisie par l'enquête (cf. le paragraphe 3.2).

Notations:

L'ensemble des questions est désigné par Q . Une question q consiste en un ensemble J_q de modalités de réponse.

On désignera par $R(n, Q)$ le tableau de codage réduit. L'élément r_{iq} contient le numéro de la modalité de la question q choisie par le sujet i .

On désignera par $Z(n, J)$ le tableau disjointif complet décrivant les réponses des n individus par un codage binaire.

On désignera par B le tableau de contingence de BURT associé au tableau des réponses Z .

Le tableau B est formé de Q^2 blocs. Le $q^{\text{ème}}$ bloc / diagonal $Z_q^T Z_q$ est une matrice diagonale d'ordre J_q^2 (puisque deux modalités de réponse à une même question ne peuvent pas être choisies simultanément).

Le bloc indicé par (q, q') , d'ordre $J_q \times J_{q'}$, est le tableau de contingence croisant les réponses aux deux questions q et q' .

Désignons par D , d'ordre $J \times J$, la matrice diagonale ayant les mêmes éléments diagonaux que B (ces éléments diagonaux sont les effectifs correspondant à chacune des modalités).

La matrice D peut également être considérée comme format Q^2 blocs (seuls les Q blocs diagonaux sont des matrices non nulles, le $q^{\text{ème}}$ bloc diagonal :

$D_q = Z_q^T Z_q$, dont les termes de la diagonale sont les effectifs correspondant aux diverses modalités de réponse à la question q).

Cas de deux questions:

Le tableau des réponses Z s'écrit $Z = Z_1/Z_2$. Il est alors équivalent au point de vue de la description des associations entre modalités, d'effectuer:

- (1) L'analyse des correspondances du tableau Z d'ordre ($n \times J$).
- (2) L'analyse des correspondances du tableau B d'ordre ($J \times J$).
- (3) L'analyse des correspondances du tableau K = $Z_1^T Z_2$ d'ordre ($J_1 \times J_2$).

Les valeurs propres ne sont pas les mêmes:

Tableau Analysé	Valeur propre
$K = Z_1^T Z_2$	λ_K
$Z = (Z_1; Z_2)$	$\mu_Z = \frac{1 + \sqrt{\lambda_K}}{2}$
$B = Z^T Z$	$\mu_B = \mu_z^2$

On peut faire deux remarques:

Remarque 1:

Dans l'analyse du grand tableau disjonctif Z, les points représentant les diverses modalités de réponse / aux deux questions sont des éléments appartenant à un même ensemble (l'ensemble des colonnes de Z). Par contre dans l'analyse du tableau de contingence $Z_1^T Z_2$, les éléments se divisent en points-lignes et points-colonnes. Le fait que les typologies obtenues dans l'espace des premiers facteurs soient identiques à un coefficient près (les valeurs propres n'étant pas les mêmes) nous prouve que la représentation simultanée des points lignes et des points-colonnes dans l'analyse des correspondances d'un tableau de contingence n'est pas uniquement un artifice graphique.

Remarque 2:

Ces trois analyses, effectuées sur la même information brute, donnent des résultats similaires, mais avec des valeurs propres différentes et donc des taux d'inertie différents. Les relations entre ces taux d'inertie montrent que les taux sont toujours beaucoup plus élevés dans l'analyse du tableau de contingence $Z_1^T Z_2$ que dans celle du tableau Z. D'une manière générale, l'analyse des tableaux sous forme disjonctive donne toujours

des taux d'inertie faibles, ceux ci donnent une idée / trop pessimiste de la part d'information extraite.

On généralise ensuite au cas de Q questions. Les propriétés de la méthode sont très intéressantes.

Les Q sous-nuages de points correspondant au J_q modalités d'une question q ont le même centre de gravité, celui du nuage global. Les coordonnées des modalités de réponse à chaque question sont donc centrées sur chaque axe factoriel.

L'inertie totale vaut $(J/Q) - 1$, elle vaut 1 dans le cas particulier où toutes les questions ont deux modalités de réponse ($J = 2Q$).

La part d'inertie due à une modalité de réponse est d'autant plus grande que l'effectif dans cette modalité est faible:

$$C_{(j)} = \frac{1}{Q} \left(1 - \frac{d_{ij}}{n} \right)$$

il faut donc éviter les modalités à effectif faible.

La part d'inertie due à une question est fonction / croissante du nombre de modalités de réponse

$$C_q = \frac{1}{Q} (J_q - 1)$$

Il faut donc équilibrer le nombre de modalités dans l'ensemble des questions actives. En pratique si une question a un nombre de modalités trop grand on peut effectuer un regroupement en un nombre de modalités raisonnable qui préserve le sens de la question, c'est ce regroupement qu'on mettra en variable active, on peut d'autre part projeter les modalités de la question d'origine en variable illustrative.

Le rang de Z est $J-Q+1$. Dans l'analyse par rapport / au centre de gravité G on a $J-Q$ valeurs propres non nulles, le rang de la matrice à diagonaliser est $J-Q$.

Zone de confiance d'une modalité supplémentaire. Les valeurs-test éditées par les programmes (pour les modalités actives ou supplémentaires) permettent de comparer / les significations respectives de plusieurs modalités / supplémentaires comme nous le verrons dans l'exemple traité au paragraphe suivant.

4. MISE EN OEUVRE PRACTIQUE DES MÉTHODES:

L'option choisie pour le séminaire est de faire une analyse complète d'un exemple en utilisant le système / SPAD. Le paquete SPAD (système portable pour l'analyse des données) a été implanté et des listings ont été sortis au centre de calcul de l'Université de BILBAO durant la même semaine que ce séminaire. La description, l'interprétation et l'analyse des résultats de l'enquête sur les 111 FINCAS DE MARACAY (VENEZUELA) ont été faites en utilisant les copies de listings que les participants pouvaient annoter. Tout en commentant ces listings à l'aide de transparents nous montrions comment se servir des différentes étapes de SPAD à l'aide des notices d'utilisation.

Certes nous avons exposé d'autres exemples à titre illustratif mais ils ne peuvent remplacer l'expérience pratique que chaque chercheur peut acquérir en mettant en oeuvre sa propre chaîne de travail. Cette expérience est maintenant possible puisque l'outil a été rendu accessible.

L'enquête que nous analysons ici porte sur 111 Finca de Maracay. Les données ont été recueillies par Rany Courbon et ont fait l'objet d'un des ateliers de travail sur ordinateur des Jornadas de Análisis de Datos de Informática, encuentro Francia-América Latina (Julio de 1983 à Caracas, Venezuela. D'autres chercheurs travaillant pour le ministère de l'agriculture du Venezuela préparaient des enquêtes similaires dans d'autres villes. L'exemple n'est pas de grande taille mais illustre très bien ce que nous appelons une analyse descriptive exploratoire : premièrement nous examinerons les relations entre l'ensemble des variables caractérisant

la production des fermes, deuxièmement nous emploieront/ deux méthodes de classification adaptées à ce type de tableau pour résumer la composition de ces fermes par des classes bien typées.

Avec SPAD les traitements constituant une analyse sont découplés en étapes: étape d'organisation des données, étape des calculs proprements dits, étape des éditions graphiques, étapes de classification, etc. Une analyse complète est donc une succession d'étapes. Chacune d'elles est commandée par quelques paramètres, et la communication des résultats entre étapes s'effectue automatiquement par le moyen de fichiers externes pouvant être conservés pour des usages différents. Les livres (2) et (3) de SPAD cités en bibliographie (cf. le paragraphe 3 . 1) contiennent le catalogue des étapes, la liste d'analyses classiques avec les assemblages des étapes qui la composent, les principes pour l'exécution d'une analyse. Le chapitre II rassemble les notices d'utilisation de chaque étape. On y trouvera la liste et les définitions des paramètres de commande cas par cas.

Exemple commenté:

Nous enchainons dix étapes dans une même exécution, leur assemblage est simple, chaque étape étant simplement appelée par un mot clé : LISTP, DONNE, LILEX, MULTC, GRAPH, SEMIS, TAMIS, GRAFK. Le mot clé STOP marque la fin d'une analyse.

L'étape LISTP effectue le listing de l'ensemble des cartes paramètres. Ici, on obtient la page de listing suivante:

LISTAGE DES PARAMETRES DE COMMANDE

```

-----+---1---+---2---+---3---+---4---+---5---+---6---+---7---+---8
1 LISTP
2 DONNE
3 23 111 1 1 17
4 (A4,23F1.0)
5 LILEX
6 111 FINCAS DE MARACAY(REMÝ) CON 23 VARIABLES . EL 15.07.83(CARACAS).....
7 24 78 23 111 0 1 0 0 1 1 -1
8 1 0
9 1121211112121222222222
10 MULTC
11 6 0 0 2 1 30 0 2
12 GRAPH
13 1 1 58 0 1 1 58 0 0 0 0 0 0 0 0 0 0
14 1 2 2 3
15 SEMIS
16 2 10 5 15 6
17 TAMIS
18 1 30 10 1 1 0 3
19 GRAFK
20 1 1 58 1 1 58
21 1 2 2 3
22 TAMIS
23 0 30 10 1 1 0 5
24 GRAFK
25 1 1 58 1 1 58
26 1 2 2 3
27 STOP FIN DE LOS PARAMETROS DE FINCP2
-----+---1---+---2---+---3---+---4---+---5---+---6---+---7---+---8

```

FIN DU FICHIER DES COMMANDES

Dans cet exemple on utilise ensuite l'étape DONNE:

- le fichier initial compte 23 variables (NQEXA), / 111 fincas (nombre de lignes du tableau de données, - / IEXA = 111), un identificateur de finca compté en nombre de A4 (ici NVIDI = 1, un A4).

- ce fichier est lu avec le format fortran suivant (A4, 23F1.0), un enregistrement est utilisé (LFMT = 1) pour introduire ce format.

- le numéro logique du fichier (NDON2 = 17) est introduit en paramètre.

- l'étape DONNE lit le fichier initial des données/ et crée un fichier archive (NDONA = 2) utilisable en entrée par l'étape LILEX. On suppose ici que le fichier / initial est un fichier propre qu'il suffit d'archiver ;

les opérations de nettoyage, de recodage, de fusion, etc ... ont été effectuées préalablement.

L'étape LILEX réalise le prélèvement de l'information utile sur les fichiers archives, et permet de choisir les éléments actifs et illustratifs des analyses qui suivront: ici nous gardons les 111 fincas comme lignes actives, attribuons le poids 1 à chaque finca (NVPOI=0), mais nous faisons une sélection parmi les 23 variables, les 10 variables caractérisant la production de fermes sont actives (type 1, variable nominale active), les 13 autres sont illustratives (type 2, variable nominale --/ illustrative). Cette sélection est très importante (voir le paragraphe 2.3 sur les qualités d'un tableau) c'est ici qu'on fait cette sélection et qu'on introduit le dictionnaire des variables, celles ci sont réordonnées / comme on le voit sur cette et la page de listing suivante:

```

ETAPE ** LILEX **

TITRE DE L-ANALYSE          111 FINCAS DE MARACAY(REMY) CON 23 VARIABLES . EL 15.07.83(CARACAS).....
-----
```

CARTE DES PARAMETRES POUR LILEX

```

NOUTI= 24      JUTI= 78      NQEXA= 23      IEXA= 111      MOFIL= 0      LEX= 1      MEDIT= 1      MDICO= -1
NVPOI= 0      NVIDI= 1      NVFIL= 0
```

MODE DE SELECTION (COLONNES ET LIGNES)

```

MODIU= 1      MODIG= 0
..... FICHIER EN ENTREE = 1 (NDICA)
.....           2 (NDONA)
..... FICHIER EN ECRITURE = 8 (NDIC)
.....           9 (NDON)
..... UTILISATION DE MEMOIRES          VOUS AVEZ RESERVE 10000          VOUS AVEZ BESOIN DE 1376
```

MISE EN PLACE DES COLONNES OU VARIABLES

SELECTION DES ELEMENTS SELON LE MODE 1

BILAN DE LA SELECTION

```

TYPE 1      NOMBRE DE VARIABLES      10
TYPE 2      NOMBRE DE VARIABLES      13
VECTEUR INDICATEUR DES      23 ELEMENTS PAR GROUPE DE 10
1121211111 2121222222 222
```

DICTIONNAIRE DES VARIABLES ET DES MODALITES DE REPONSE

VARIABLES DU TYPE 1 10
 MODALITES CORRESPONDANTES 33

1 UNIDAD AGROECOLOGICA	/ UAE1=ALUVIAL1	/ UAE2=ALUVIAL2	/ UAE3=ALUVIAL3	(5 MODALITES)	/ UAE4=LAGUNERO	/
	/ UAE5=ASOCIACION					
2 SUPERFICIE UTILIZADA	/ SUP1=MENOR DE 5 HA	/ SUP2=DE 5.1 A 10 HA	/ SUP3=MAYOR DE 10.1 HA	(3 MODALITES)		
	/					
4 RUBRO	/ RUB1=FRUTAL	/ RUB2=FRUTAL Y CULTIVOS	/ RUB3=CANA DE AZUCAR	(4 MODALITES)	/ RUB4=CULT.DE CICLO CORTO	/
6 RIEGO	/ RIE1=SIN RIEGO	/ RIE2=CON RIEGO		(2 MODALITES)		
7 TECNOLOGIA	/ TEC1=ALTA	/ TEC2=MEDIANA ALTA	/ TEC3=MEDIANA BAJA	(4 MODALITES)	/ TEC4=BAJA	/
8 MAQUINARIA	/ MAQ1=NO USA	/ MAQ2=USA ALQUILADA	/ MAQ3=USA PROPIA	(3 MODALITES)		
9 TRABAJO	/ TRA1=EXCLUSIV FAMILIAR	/ TRA2=PRINCIP FAMILIAR	/ TRA3=ASALARIADO	(3 MODALITES)		/
10 NACIONALIDAD	/ NAC1=VENEZOLANO	/ NAC2=CANARIO		(2 MODALITES)		
12 VIVE DE LA FINCA	/ VIF1=EXCLUSIVAMENTE	/ VIF2=PRINCIPALMENTE	/ VIF3=NO VIVE DE LA FINCA	(3 MODALITES)		
14 CREDITO	/ CRE1=NO RECIBE SIN PROB	/ CRE2=NO RECIBE CON PROB	/ CRE3=SI RECIBE SIN PROB	(4 MODALITES)	/ CRE4=SI RECIBE CON PROB	/

VARIABLES DU TYPE 2 13
 MODALITES CORRESPONDANTES 44

3 SUPERFICIE EN FRUTAL	/ SUF1=SIN FRUTAL	/ SUF2=DE 0.1 A 2 HA	/ SUF3=DE 2.1 A 5 HA	(4 MODALITES)	/ SUF4=MAYOR DE 5 HA	/
	/ CAN1=LO MISMO	/ CAN2=FRUTAL	/ CAN3=CULT DE CICLO CORTO	(3 MODALITES)		
5 CULTIVO ANTERIOR						
11 EDAD	/ EDA1=20 A 40 ANOS	/ EDA2=41 A 50 ANOS	/ EDA3=51 A 60 ANOS	(4 MODALITES)	/ EDA4=MAYOR DE 61 ANOS	/
13 DESTINO DE LA PRODUCCION	/ DES1=AGROINDUSTRIA	/ DES2=CONSUMO FAMILIAR	/ DES3=INTERMEDIARIO	(4 MODALITES)	/ DES4=AGRO/INTER	/
15 ASISTENCIA TECNICA	/ ATE1=NO RECIBE	/ ATE2=SI RECIBE		(2 MODALITES)		
16 NIVEL DE VIDA	/ NIV1=ALTO	/ NIV2=MEDIO	/ NIV3=BAJO	(3 MODALITES)		
17 COMPORTAMIENTO DEL INGRESO	/ ING1=HA BAJADO	/ ING2=SE HA MANTENIDO	/ ING3=HA SUBIDO	(3 MODALITES)		
18 PROBLEMAS PLAGAS Y ENFERMEDADES	/ PLE1=NO TIENE	/ PLE2=TRISTEZA/STENOMA	/ PLE3=OTROS	(4 MODALITES)	/ PLE4=AMBOS	/
19 PROBLEMAS DE MALEZAS	/ PMA1=NO TIENE	/ PMA2=SI TIENE		(2 MODALITES)		
20 PROBLEMA DE AGUA	/ PAG1=NO TIENE	/ PAG2=DRENAJE	/ PAG3=RIEGO	(3 MODALITES)		
21 PRINCIPAL PROBLEMA	/ PRO1=NO TIENE	/ PRO2=AGUA	/ PRO3=CREDITO	(7 MODALITES)	/ PRO4=ROBO	/
	/ PRO5=PLAGAS	/ PRO6=CRECIMIENTO DEL LAGO	/ PRO7=OTROS			
22 LABOR QUE REALIZA EL PRODUCTOR	/ LAB1=TODO	/ LAB2=ALGUNAS COSAS	/ LAB3=SUPERVISA	(3 MODALITES)		
23 SE SIENTE SATISFECHO DE LA FINCA	/ SAT1=NO	/ SAT2=SI		(2 MODALITES)		

I EFFECTIF TOTAL DES LIGNES 111 I

FIN DE L-ETAPE ** LILEX **

Nous avons choisi un ensemble homogène de variables actives caractérisant les conditions de production des fermes: *unidad agro-ecológica, superficie utilizada, rubro, riego, tecnología, maquinaria, trabajo, nacionalidad, vive de la finca, crédito*, au total 33 modalités de réponse figurant sur les pages précédentes, ce sont --/ celles qui contribueront à la construction des axes - / principaux dans l'analyse des correspondances multiples (étape MULTC). Il est clair que la question *Se siente sa tisfecho de la finca* (question 23 avec les réponses NO ou SI) est une question d'opinion, nous avons donc choisi de la mettre en illustrative. *El nivel de vida* (question 16 avec les trois modalités de réponse ALTO, MEDIO BAJO) est plutôt un résultat de la production de la finca et non une condition de production, on la met donc en illustrative.

L'étape MULTC gère les calculs et effectue les éditions pour une analyse des correspondances multiples, et ceci selon les paramètres demandés. La partie théorique de cette méthode figure au paragraphe 3.3.4-(CM). Commençons brièvement le choix des paramètres fait ici: nous calculons et sauvegardons six axes factoriels (NFAC = 6)

TABLEAU DE BURT

	UAE1	UAE2	UAE3	UAE4	UAE5	SUP1	SUP2	SUP3	RUB1	RUB2	RUB3	RUB4	RIE1	RIE2	TEC1	TEC2	TEC3	TEC4	MAQ1	MAQ2	MAQ3	TRA1	TRA2	TRA3
UAE1	I	9.																						
UAE2	I	0.	19.																					
UAE3	I	0.	0.	9.																				
UAE4	I	0.	0.	0.	57.																			
UAE5	I	0.	0.	0.	0.	17.																		
SUP1	I	0.	11.	6.	30.	4.	51.																	
SUP2	I	3.	4.	0.	18.	2.	0.	27.																
SUP3	I	6.	4.	3.	9.	11.	0.	0.	33.															
RUB1	I	9.	4.	1.	15.	5.	7.	12.	15.	15.	34.													
RUB2	I	0.	1.	0.	40.	12.	31.	12.	14.	0.	57.													
RUB3	I	0.	9.	7.	0.	0.	11.	2.	3.	0.	0.	0.	16.											
RUB4	I	0.	2.	0.	2.	0.	2.	1.	1.	0.	0.	0.	0.	4.										
RIE1	I	0.	3.	1.	40.	5.	32.	14.	3.	8.	39.	0.	2.	49.										
RIE2	I	9.	16.	B.	17.	12.	19.	13.	30.	26.	18.	16.	2.	0.	62.									
TEC1	I	6.	16.	8.	18.	10.	20.	15.	23.	23.	17.	16.	2.	10.	48.	58.								
TEC2	I	2.	0.	0.	9.	3.	5.	3.	6.	6.	8.	0.	0.	5.	9.	0.	14.							
TEC3	I	1.	0.	1.	17.	2.	11.	8.	4.	4.	17.	0.	0.	0.	17.	4.	0.	0.	21.					
TEC4	I	0.	3.	0.	13.	2.	15.	3.	0.	1.	15.	0.	2.	17.	1.	0.	0.	0.	18.					
MAQ1	I	0.	1.	0.	15.	2.	16.	2.	0.	1.	16.	0.	1.	18.	0.	1.	3.	2.	12.	18.				
MAQ2	I	0.	11.	7.	17.	2.	27.	7.	3.	3.	20.	13.	1.	18.	19.	20.	3.	10.	4.	0.	37.			
MAQ3	I	9.	7.	2.	25.	13.	8.	18.	30.	30.	21.	3.	2.	13.	43.	37.	8.	9.	2.	0.	0.	56.		
TRA1	I	1.	2.	2.	7.	0.	9.	3.	0.	5.	5.	0.	2.	7.	5.	4.	1.	2.	5.	5.	4.	3.	12.	
TRA2	I	2.	5.	1.	32.	12.	29.	13.	10.	10.	35.	5.	2.	30.	22.	22.	7.	12.	11.	11.	18.	23.	0.	52.
TRA3	I	6.	12.	6.	18.	5.	13.	11.	23.	19.	17.	11.	0.	12.	35.	32.	6.	7.	2.	2.	15.	30.	0.	47.
NAC1	I	2.	16.	9.	49.	8.	49.	20.	15.	15.	49.	16.	4.	46.	38.	42.	6.	19.	17.	18.	36.	30.	12.	41.
NAC2	I	7.	3.	0.	8.	9.	2.	7.	18.	19.	8.	0.	0.	3.	24.	16.	8.	2.	1.	0.	1.	26.	0.	11.
VIF1	I	6.	5.	4.	29.	13.	20.	15.	22.	26.	26.	5.	0.	22.	35.	30.	10.	10.	7.	5.	19.	33.	5.	24.
VIF2	I	3.	10.	4.	22.	2.	23.	11.	7.	7.	21.	10.	3.	22.	19.	21.	4.	8.	8.	10.	14.	17.	4.	24.
VIF3	I	0.	4.	1.	6.	2.	8.	1.	4.	1.	10.	1.	1.	5.	8.	7.	0.	3.	3.	3.	4.	6.	3.	4.
CRE1	I	5.	1.	0.	16.	4.	12.	6.	8.	9.	16.	0.	1.	12.	14.	9.	5.	6.	5.	6.	15.	2.	15.	9.
CRE2	I	2.	9.	3.	35.	11.	29.	14.	17.	20.	35.	2.	3.	35.	25.	29.	6.	13.	12.	13.	18.	29.	9.	31.
CRE3	I	2.	4.	1.	6.	1.	5.	5.	4.	5.	4.	0.	2.	12.	9.	3.	2.	0.	0.	6.	8.	1.	2.	11.
CRE4	I	0.	5.	5.	0.	1.	5.	2.	4.	0.	2.	9.	0.	0.	11.	11.	0.	0.	0.	7.	4.	0.	4.	7.
NAC1	I	84.																						
NAC2	I	0.	27.																					
VIF1	I	35.	22.	57.																				
VIF2	I	37.	4.	0.	41.																			
VIF3	I	12.	1.	0.	0.	13.																		
CRE1	I	17.	9.	14.	8.	4.	26.																	
CRE2	I	47.	13.	31.	23.	6.	0.	60.																
CRE3	I	10.	4.	9.	3.	2.	0.	0.	14.															
CRE4	I	10.	1.	3.	7.	1.	0.	0.	0.	11.														

NAC1 NAC2 VIF1 VIF2 VIF3 CRE1 CRE2 CRE3 CRE4

on ne se sort pas de la possibilité de transférer une modalité en illustrative si l'effectif est inférieur ou / égal au paramètre NMIN (ici NMIN = 0), on ne demande pas d'édition concernant les fincas (LIST3 = 0). Pour deux facteurs successifs (NTEXT = 2) on demande une édition / des rangements des modalités (actives et illustratives) sur un page (NPAGT = 1) et pour les 30 (NLIGT) modalités à plus forte inertie sur le facteur. On ne demande pas d'édition des corrélations des variables continues avec les facteurs (NCOR = 0) puisque on n'a pas de variables continues dans cet exemple. Pour l'édition du tableau du BURT, croissant les questions actives, on a demandé l'édition des effectifs et des profils (NTAB = 2).

On ne peut faire figurer ici que les transparents ou pages de listing les plus importants, tout d'abord le tableau de BURT et l'histogramme des premières valeurs pro pres. Il y a deux facteurs largement dominants, notre / attention portera donc surtout sur le plan (F_1, F_2) des deux premiers facteurs. Pour les modalités actives les masses, les distances à l'origine, les coordonnées et les contributions, nous permettent de mettre en évidence les principales oppositions.

PROFILS HORIZONTAUX (EN POUR 1000) DU TABLEAU DE BURT
LE TRI-A-PLAT DE CHAQUE QUESTION (EN POUR 1000) FIGURE SUR LA DIAGONALE CORRESPONDANTE

	UAE1	UAE2	UAE3	UAE4	UAE5	SUP1	SUP2	SUP3	RUB1	RUB2	RUB3	RUB4	RIE1	RIE2	TEC1	TEC2	TEC3	TEC4	MAQ1	MAQ2	MAQ3	TRA1	TRA2	TRA3
UAE1 I	81.	0.	0.	0.	0.	333.	667.	1000.	0.	0.	0.	0.	1000.	667.	222.	111.	0.	0.	0.	1000.	111.	222.	667.	
UAE2 I	0.	171.	0.	0.	0.	579.	211.	211.	211.	474.	105.	158.	642.	842.	0.	0.	158.	53.	579.	368.	105.	263.	632.	
UAE3 I	0.	0.	81.	0.	0.	657.	0.	333.	111.	111.	778.	0.	111.	889.	889.	0.	0.	0.	0.	778.	222.	222.	111.	667.
UAE4 I	0.	0.	0.	514.	0.	526.	316.	158.	263.	702.	0.	35.	702.	298.	316.	158.	298.	228.	263.	298.	439.	123.	561.	316.
UAE5 I	0.	0.	0.	0.	153.	235.	118.	647.	294.	706.	0.	0.	294.	706.	588.	176.	118.	118.	118.	765.	0.	706.	294.	
SUP1 I	0.	216.	118.	588.	78.	459.	0.	0.	137.	608.	216.	39.	627.	373.	392.	98.	216.	294.	314.	529.	157.	176.	569.	255.
SUP2 I	111.	148.	0.	667.	74.	0.	243.	0.	444.	444.	74.	37.	519.	481.	556.	111.	222.	111.	74.	259.	667.	111.	481.	407.
SUP3 I	182.	121.	91.	273.	333.	0.	0.	297.	455.	424.	91.	30.	91.	909.	697.	182.	121.	0.	0.	91.	909.	0.	303.	697.
RUB1 I	265.	118.	29.	441.	147.	206.	353.	441.	306.	0.	0.	0.	235.	765.	676.	176.	118.	29.	29.	88.	882.	147.	294.	559.
RUB2 I	0.	70.	18.	702.	211.	544.	211.	248.	0.	514.	0.	0.	684.	316.	298.	140.	263.	281.	351.	368.	88.	614.	298.	559.
RUB3 I	0.	563.	438.	0.	0.	688.	125.	188.	0.	0.	144.	0.	0.	1000.	1000.	0.	0.	0.	0.	813.	188.	0.	313.	688.
RUB4 I	0.	500.	0.	500.	250.	250.	0.	0.	0.	36.	500.	500.	500.	0.	0.	500.	250.	500.	500.	500.	500.	500.	0.	
RIE1 I	0.	61.	20.	816.	102.	653.	286.	61.	163.	796.	0.	41.	441.	0.	204.	102.	347.	347.	367.	367.	265.	143.	612.	245.
RIE2 I	145.	258.	129.	274.	194.	306.	210.	484.	419.	290.	258.	32.	0.	559.	774.	145.	65.	16.	0.	306.	694.	81.	355.	565.
TEC1 I	103.	276.	138.	310.	172.	345.	259.	397.	397.	293.	276.	34.	172.	828.	523.	0.	0.	0.	17.	345.	638.	69.	379.	552.
TEC2 I	143.	0.	643.	214.	357.	214.	429.	429.	571.	0.	0.	357.	643.	0.	128.	0.	0.	214.	214.	571.	71.	500.	429.	
TEC3 I	48.	0.	48.	810.	95.	524.	286.	190.	190.	810.	0.	0.	810.	190.	0.	0.	189.	0.	95.	476.	429.	95.	571.	333.
TEC4 I	0.	167.	0.	722.	111.	833.	167.	0.	56.	833.	0.	111.	944.	56.	0.	0.	162.	667.	222.	111.	278.	611.	111.	
MAQ1 I	0.	56.	0.	833.	111.	889.	111.	0.	56.	889.	0.	56.	1000.	0.	56.	167.	111.	667.	162.	0.	0.	278.	611.	111.
MAQ2 I	0.	297.	189.	459.	54.	730.	189.	81.	81.	541.	351.	27.	486.	514.	541.	81.	270.	108.	0.	333.	0.	108.	486.	405.
MAQ3 I	161.	125.	36.	446.	232.	143.	321.	536.	536.	375.	54.	36.	232.	768.	661.	143.	161.	36.	0.	0.	505.	54.	411.	536.
TRA1 I	83.	167.	167.	583.	0.	750.	250.	0.	417.	417.	0.	167.	583.	417.	333.	83.	167.	417.	417.	333.	250.	108.	0.	0.
TRA2 I	38.	96.	19.	615.	231.	558.	250.	192.	192.	673.	96.	38.	577.	423.	423.	135.	231.	212.	212.	346.	442.	0.	468.	0.
TRA3 I	128.	255.	128.	383.	106.	277.	234.	489.	404.	362.	234.	0.	255.	745.	681.	128.	149.	43.	43.	319.	638.	0.	0.	423.
NAC1 I	24.	190.	107.	583.	95.	583.	238.	179.	179.	583.	190.	48.	548.	452.	500.	71.	226.	202.	214.	429.	357.	143.	488.	369.
NAC2 I	259.	111.	0.	296.	333.	74.	259.	667.	704.	296.	0.	0.	111.	889.	593.	296.	74.	37.	0.	37.	963.	0.	407.	593.
VIF1 I	105.	88.	70.	509.	228.	351.	263.	386.	456.	456.	88.	0.	386.	614.	526.	175.	175.	123.	88.	333.	579.	88.	421.	491.
VIF2 I	73.	244.	98.	537.	49.	561.	268.	171.	171.	512.	244.	73.	537.	463.	512.	98.	195.	195.	244.	341.	415.	98.	585.	317.
VIF3 I	0.	308.	77.	462.	154.	615.	77.	308.	77.	769.	77.	77.	385.	615.	538.	0.	231.	231.	231.	308.	462.	231.	308.	462.
CRE1 I	192.	38.	0.	615.	154.	462.	231.	308.	346.	615.	0.	38.	462.	538.	346.	192.	231.	231.	192.	577.	77.	577.	346.	
CRE2 I	33.	150.	50.	583.	183.	483.	233.	283.	333.	583.	33.	50.	583.	417.	483.	100.	217.	200.	217.	300.	483.	150.	517.	333.
CRE3 I	143.	286.	71.	429.	71.	357.	286.	357.	286.	357.	0.	143.	857.	643.	214.	143.	0.	0.	429.	571.	71.	143.	766.	
CRE4 I	0.	455.	455.	0.	91.	455.	182.	364.	0.	182.	618.	0.	0.	1000.	1000.	0.	0.	0.	0.	636.	364.	0.	364.	636.

	NAC1	NAC2	VIF1	VIF2	VIF3	CRE1	CRE2	CRE3	CRE4
UAE1 I	222.	778.	667.	333.	0.	556.	222.	222.	0.
UAE2 I	842.	158.	263.	526.	211.	53.	474.	211.	263.
UAE3 I	1000.	0.	444.	444.	111.	0.	333.	111.	556.
UAE4 I	860.	140.	509.	386.	105.	281.	614.	105.	0.
UAE5 I	471.	529.	765.	118.	118.	235.	647.	59.	59.
SUP1 I	961.	39.	392.	451.	157.	235.	569.	98.	98.
SUP2 I	741.	259.	556.	407.	37.	222.	519.	185.	74.
SUP3 I	455.	545.	667.	212.	121.	242.	515.	121.	121.
RUB1 I	441.	559.	765.	206.	29.	265.	588.	147.	0.
RUB2 I	860.	140.	456.	368.	175.	281.	614.	70.	35.
RUB3 I	1000.	0.	313.	625.	63.	0.	125.	313.	563.
RUB4 I	1000.	0.	750.	250.	250.	750.	0.	0.	0.
RIE1 I	939.	61.	449.	449.	102.	245.	714.	41.	0.
RIE2 I	613.	387.	565.	306.	129.	226.	403.	194.	177.
TEC1 I	724.	276.	517.	362.	121.	155.	500.	155.	190.
TEC2 I	429.	571.	714.	286.	0.	357.	429.	214.	0.
TEC3 I	905.	95.	476.	381.	143.	286.	619.	95.	0.
TEC4 I	944.	56.	389.	444.	167.	333.	667.	0.	0.
MAQ1 I	1000.	0.	278.	556.	167.	278.	722.	0.	0.
MAQ2 I	973.	27.	514.	378.	108.	162.	486.	162.	189.
MAQ3 I	536.	464.	589.	304.	107.	268.	518.	143.	71.
TRA1 I	1000.	0.	417.	333.	250.	167.	750.	83.	0.
TRA2 I	788.	212.	462.	462.	77.	288.	596.	38.	77.
TRA3 I	660.	340.	596.	277.	128.	191.	426.	234.	149.
NAC1 I	757.	0.	417.	440.	143.	202.	560.	119.	119.
NAC2 I	0.	243.	815.	148.	37.	333.	481.	148.	37.
VIF1 I	614.	386.	514.	0.	0.	246.	544.	158.	53.
VIF2 I	902.	98.	0.	369.	0.	195.	561.	73.	171.
VIF3 I	923.	77.	0.	0.	117.	308.	462.	154.	77.
CRE1 I	654.	346.	538.	308.	154.	234.	0.	0.	0.
CRE2 I	783.	217.	517.	383.	100.	0.	541.	0.	0.
CRE3 I	714.	286.	643.	214.	143.	0.	0.	126.	0.
CRE4 I	909.	91.	273.	636.	91.	0.	0.	0.	99.

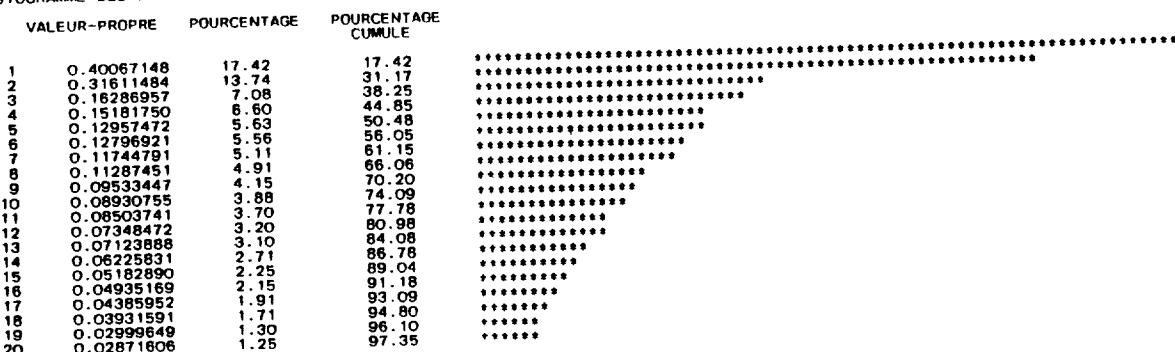
SOMME DES POIDS = 111.000

EDITION DES VALEURS-PROPRIÉTÉS

SOMME DES VALEURS-PROPRIÉTÉS

2.29995728

HISTOGRAMME DES PREMIERES VALEURS-PROPRIÉTÉS



EDITION SOMMAIRE DES VALEURS-PROPRIÉTÉS DE 21 A 23

0.02370763 0.02216966 0.01502028

Sur le premier axe : les modalités con riego, nacionalidad = canario, les fermes de surface les plus grandes superficie utilizada mayor de 10.1 Ha, Rubro = Frutal, Maquinaria = usa propia, Trabajo = asalariado, Tecnología = alta ont de fortes contributions absolues et des coordonnées négatives. Elles s'opposent aux modalités actives suivantes : sin riego, no usa maquinaria, tecnología = baja, les fermes de surface la plus petite superficie utilizada menor de 5 Ha, unidad agro-ecológica = lagunero, nacionalidad = venezolano, et avec une contribution plus faible trabajo principalmente familiar.

Sur le deuxième axe : Caña de azúcar, usa alquilada la maquinaria, tecnología alta s'opposent à Frutal, usa propia la maquinaria et canario.

Les positions respectives de l'ensemble des modalités dans le plan (F_1 , F_2) des deux premiers facteurs peuvent être obtenues avec l'étape GRAPH, c'est ce que nous faisons ici. Il est cependant préférable de travailler / sur deux graphiques séparés: l'un pour les modalités actives (figure 1), l'autre pour les modalités illustratives (figure 2).

De plus il est important:

- pour le premier de tenir compte des contributions comme nous l'avons fait pour décrire les principales oppositions entre les modalités actives, c'est à dire entre les points qui contribuent le plus à la construction de ces deux premiers axes dominants (contributions abso-

lues fortes) et qui sont par conséquent bien représentés dans ce plan (contributions relatives fortes). Au contraire une modalité comme cultivo de ciclo corto (RUB4 / de la question RUBRO) est très mal représentée en projection dans ce plan (contributions relatives très faibles, respectivement 0.02 et 0.00 en F_1 et F_2). Ce point est correlé au troisième axe (contribution relative 0.33), il contribue très largement à sa construction (contribution absolue de 13.6% à F_3). De plus remarquons que seulement 4 fincas sur 111 ont choisi cette modalité.

- Pour le second de tenir compte des valeurs-test / qui permettent de comparer les significations respectives des modalités illustratives. Remarquons qu'ici elles sont nombreuses à être faites (supérieures à 2 en valeur absolue) sur les deux axes fortement dominants. A titre d'exemple les modalités des deux questions Nivel de vida comportamiento del ingreso sont significatives dans ce plan factorial. Le nivel de vida est très corrélé à la question active superficie utilizada. Les deux modalités SI et NO de la question Se siente satisfecho de la finca sont significatives sur le deuxième et le troisième axe, il est amusant de remarquer que les satisfaits ne se projettent pas du côté de ceux qui ont un niveau de vie / plus élevé.

En bref, ce réseau d'interrelations nous aide à retourner aux données, au tableau de BURT; à refaire des recodages et à extraire les tableaux les plus intéressants pour mieux expliciter les résultats de / l'analyse.

POUR LES MODALITES ACTIVES
 EDITION DES COORDONNEES ET DES CONTRIBUTIONS

NOMS MASSES DISTO	COORDONNEES	CONTRIBUTIONS ABSOLUES						CONTRIBUTIONS RELATIVES											
		F1	F2	F3	F4	F5	F6	F1	F2	F3	F4	F5	F6	F1	F2	F3	F4	F5	F6
***** QUESTION 1 UNIDAD AGROECOLÓGICA																			
UAE1 0.008	11.33	-1.46	0.97	0.83	0.80	0.41	1.11	4.3	2.4	3.4	3.4	1.1	7.8	0.19	0.08	0.06	0.06	0.01	0.11
UAE2 0.017	4.84	-0.31	-1.14	0.77	-0.02	0.04	-0.61	0.4	7.1	6.3	0.0	0.0	5.0	0.02	0.27	0.12	0.00	0.00	0.06
UAE3 0.008	11.33	-0.48	-1.95	-0.44	-0.21	-0.30	1.04	0.5	9.8	1.0	0.2	0.8	6.9	0.02	0.34	0.02	0.00	0.01	0.10
UAE4 0.051	0.95	0.54	0.34	-0.16	0.37	-0.01	-0.01	3.8	1.9	0.8	4.6	0.0	0.0	0.31	0.12	0.03	0.14	0.00	0.00
UAE5 0.015	5.53	-0.45	0.67	-0.54	-1.52	-0.08	-0.43	0.8	2.2	2.8	23.4	0.1	2.2	0.04	0.08	0.05	0.42	0.00	0.03
		CONTRIBUTION CUMULEE						9.7	23.3	14.3	31.6	1.7	21.9						
***** QUESTION 2 SUPERFICIE UTILIZADA																			
SUP1 0.046	1.18	0.66	-0.39	0.02	-0.11	-0.18	0.32	5.0	2.2	0.0	0.4	1.1	3.7	0.37	0.13	0.00	0.01	0.03	0.09
SUP2 0.024	3.11	-0.08	0.28	0.04	0.99	0.59	-0.47	0.0	0.6	0.0	15.8	6.5	4.3	0.00	0.03	0.00	0.32	0.11	0.07
SUP3 0.030	2.36	-0.95	0.37	-0.07	-0.64	-0.20	-0.11	6.8	1.3	0.1	8.1	1.0	0.3	0.39	0.06	0.00	0.18	0.02	0.00
		CONTRIBUTION CUMULEE						11.8	4.0	0.1	24.3	8.5	8.2						
***** QUESTION 4 RUBRO																			
RUB1 0.031	2.26	-0.80	0.63	0.40	0.46	0.10	0.12	4.9	3.8	3.1	4.3	0.2	0.3	0.28	0.17	0.07	0.08	0.00	0.01
RUB2 0.051	0.95	0.58	0.24	-0.39	-0.25	-0.18	-0.09	4.4	0.9	4.7	2.1	1.3	0.3	0.36	0.06	0.16	0.07	0.03	0.01
RUB3 0.014	5.94	-0.57	-2.11	-0.23	-0.01	0.25	0.43	1.2	20.3	0.5	0.0	0.7	2.1	0.05	0.75	0.01	0.00	0.01	0.03
RUB4 0.004	26.75	0.75	-0.31	2.97	-0.35	0.68	-1.42	0.5	0.1	19.8	0.3	1.3	5.7	0.02	0.00	0.33	0.00	0.02	0.08
		CONTRIBUTION CUMULEE						10.9	25.1	27.8	6.7	3.5	8.4						
***** QUESTION 6 RIEGO																			
RIE1 0.044	1.27	0.91	0.26	-0.16	0.18	0.05	-0.03	9.2	0.9	0.7	0.8	0.1	0.0	0.66	0.05	0.02	0.02	0.00	0.00
RIE2 0.056	0.79	-0.72	-0.20	0.12	-0.13	-0.04	0.02	7.3	0.7	0.5	0.6	0.1	0.0	0.66	0.05	0.02	0.02	0.00	0.00
		CONTRIBUTION CUMULEE						16.5	1.6	1.2	1.4	0.2	0.0						
***** QUESTION 7 TECNOLOGIA																			
TEC1 0.052	0.91	-0.56	-0.38	0.14	-0.08	0.11	-0.27	4.1	2.4	0.8	0.2	0.5	2.9	0.34	0.16	0.02	0.01	0.01	0.06
TEC2 0.013	6.93	-0.31	0.84	-0.31	-0.02	0.01	1.28	0.3	2.8	0.8	0.0	0.0	16.2	0.01	0.10	0.01	0.00	0.00	0.24
TEC3 0.019	4.29	0.60	0.30	-0.98	0.89	-0.33	-0.55	1.7	0.5	10.8	5.9	1.6	4.4	0.06	0.02	0.21	0.11	0.03	0.07
TEC4 0.016	5.17	1.35	0.21	0.90	-0.54	0.02	0.50	7.4	0.2	8.1	3.1	0.0	3.2	0.35	0.01	0.16	0.06	0.00	0.05
		CONTRIBUTION CUMULEE						13.5	6.0	20.1	9.2	2.1	26.7						
***** QUESTION 8 MAQUINARIA																			
MAQ1 0.018	5.17	1.45	0.32	0.65	-0.51	0.04	0.76	8.6	0.5	4.3	2.8	0.0	7.4	0.41	0.02	0.06	0.05	0.00	0.11
MAQ2 0.033	2.00	0.26	-0.83	-0.47	0.29	-0.22	0.04	0.6	7.2	4.5	1.9	1.2	0.0	0.03	0.34	0.11	0.04	0.02	0.00
MAQ3 0.050	0.98	-0.64	0.44	0.10	-0.03	0.13	-0.27	5.1	3.1	0.3	0.0	0.8	2.9	0.41	0.20	0.01	0.00	0.02	0.08
		CONTRIBUTION CUMULEE						14.2	10.9	9.1	4.7	1.9	10.3						
***** QUESTION 9 TRABAJO																			
TRA1 0.011	8.25	0.78	-0.07	1.65	0.42	-0.80	0.32	1.6	0.0	18.0	1.3	5.4	0.8	0.07	0.00	0.33	0.02	0.06	0.01
TRA2 0.047	1.13	0.39	0.18	-0.35	-0.29	0.52	-0.09	1.8	0.5	3.4	2.6	9.9	0.3	0.13	0.03	0.11	0.07	0.24	0.01
TRA3 0.042	1.36	-0.63	-0.18	-0.04	0.21	-0.37	0.02	4.2	0.5	0.0	1.2	4.8	0.0	0.29	0.02	0.00	0.03	0.10	0.00
		CONTRIBUTION CUMULEE						7.6	1.0	21.5	5.1	19.8	1.2						
***** QUESTION 10 NACIONALIDAD																			
NAC1 0.076	0.32	0.35	-0.28	-0.02	0.08	-0.04	-0.08	2.3	1.9	0.0	0.4	0.1	0.3	0.38	0.24	0.00	0.02	0.01	0.02
NAC2 0.024	3.11	-1.09	0.87	0.05	-0.26	0.13	0.24	7.2	5.8	0.0	1.1	0.3	1.1	0.38	0.24	0.00	0.02	0.01	0.02
		CONTRIBUTION CUMULEE						9.5	7.7	0.0	1.5	0.4	1.4						
***** QUESTION 12 VIVE DE LA FINCA																			
VIF1 0.051	0.95	-0.31	0.34	-0.25	0.08	-0.22	0.18	1.3	1.8	2.0	0.2	1.9	1.1	0.10	0.12	0.07	0.01	0.05	0.03
VIF2 0.037	1.71	0.33	-0.38	0.16	0.09	0.84	-0.00	1.0	1.7	0.6	0.2	20.2	0.0	0.07	0.09	0.01	0.00	0.41	0.00
VIF3 0.012	7.54	0.32	-0.26	0.59	-0.61	-1.68	-0.70	0.3	0.2	2.5	2.8	25.6	4.5	0.01	0.01	0.05	0.05	0.38	0.07
		CONTRIBUTION CUMULEE						2.6	3.8	5.0	3.2	47.7	5.6						
***** QUESTION 14 CREDITO																			
CRE1 0.023	3.27	0.06	0.59	-0.01	-0.08	0.06	0.66	0.0	2.5	0.0	0.1	0.1	8.0	0.00	0.11	0.00	0.00	0.00	0.13
CRE2 0.054	0.85	0.26	0.20	0.08	-0.10	0.06	-0.40	0.9	0.7	0.2	0.3	0.1	6.8	0.08	0.05	0.01	0.01	0.00	0.19
CRE3 0.013	8.93	-0.71	-0.35	-0.03	1.06	-0.98	0.19	1.6	0.5	0.0	9.4	9.3	0.4	0.07	0.02	0.00	0.16	0.14	0.01
CRE4 0.010	9.09	-0.68	-2.02	-0.37	-0.63	0.77	0.37	1.2	12.8	0.8	2.6	4.6	1.1	0.05	0.45	0.01	0.04	0.07	0.02
		CONTRIBUTION CUMULEE						3.7	16.5	1.0	12.4	14.1	16.3						

COORDONNEES ET VALEURS-TESTS POUR TOUTES LES MODALITES

POIDS		COORDONNEES										VALEURS-TESTS CORRIGEES							
** VARIABLE 1		UNIDAD AGROECOLOGICA										-4.53	3.01	2.59	2.48	1.28	3.46	*	
UAE1	9.	-1.46	0.97	0.83	0.80	0.41	1.11	*				-1.48	-5.46	3.69	-0.09	0.17	-2.93	*	
UAE2	19.	-0.31	-1.14	0.77	-0.02	0.04	-0.61	*				-1.49	-6.09	-1.38	-0.67	-0.92	3.24	*	
UAE3	9.	-0.48	-1.95	-0.44	-0.21	-0.30	1.04	*				5.86	3.85	-1.70	3.97	-0.06	-0.07	*	
UAE4	57.	0.54	0.34	-0.16	0.37	-0.01	-0.01	*				-2.01	2.98	-2.42	-6.79	-0.37	-1.92	*	
UAE5	17.	-0.45	0.67	0.54	-1.52	-0.08	-0.43	*											
** VARIABLE 2		SUPERFICIE UTILIZADA										6.38	-3.74	0.22	-1.05	-1.73	3.09	*	
SUP1	51.	* 0.66	-0.39	0.02	-0.11	-0.18	0.32	*				-0.47	1.67	0.22	5.90	3.49	-2.82	*	
SUP2	27.	-0.08	0.28	0.04	0.99	0.59	-0.47	*				6.51	2.51	-0.45	-4.40	-1.39	-0.73	*	
SUP3	33.	-0.95	0.37	-0.07	-0.64	-0.20	-0.11	*											
** VARIABLE 4		RUBRO										-5.57	4.37	2.81	3.22	0.70	0.81	*	
RUB1	34.	* -0.80	0.63	0.40	0.46	0.10	0.12	*				6.30	2.58	-4.16	-2.68	-1.93	-0.98	*	
RUB2	57.	* 0.58	0.24	-0.39	-0.25	-0.18	-0.09	*				-2.46	-9.07	-0.97	-0.04	1.10	1.86	*	
RUB3	16.	* -0.57	-2.11	-0.23	-0.01	0.25	0.43	*				1.53	-0.63	6.03	-0.71	1.38	-2.87	*	
RUB4	4.	* 0.75	-0.31	2.97	-0.35	0.68	-1.42	*											
** VARIABLE 6		RIEGO										8.52	2.39	-1.45	1.54	0.49	-0.25	*	
RIE1	49.	* 0.91	0.26	-0.16	0.16	0.05	-0.03	*				-8.52	-2.39	1.45	-1.54	-0.49	0.25	*	
RIE2	62.	* -0.72	-0.20	0.12	-0.13	-0.04	0.02	*											
** VARIABLE 7		TECNOLOGIA										-6.15	-4.15	1.56	-0.84	1.23	-2.93	*	
TEC1	58.	* -0.56	-0.38	0.14	-0.08	0.11	-0.27	*				-1.23	3.36	-1.25	-0.09	0.04	5.10	*	
TEC2	14.	* 0.31	0.84	-0.31	-0.02	0.01	1.28	*				3.02	1.51	-4.84	3.48	-1.69	-2.76	*	
TEC3	21.	* 0.60	0.30	-0.96	0.69	-0.33	-0.55	*				6.23	0.99	4.15	-2.48	0.09	2.31	*	
TEC4	18.	* 1.35	0.21	0.90	-0.54	0.02	0.50	*											
** VARIABLE 8		MAQUINARIA										8.71	1.49	3.02	-2.36	0.20	3.51	*	
MAQ1	18.	* 1.45	0.32	0.65	-0.51	0.04	0.76	*				1.92	-6.14	-3.47	2.16	-1.60	0.30	*	
MAQ2	37.	* 0.26	-0.83	-0.47	0.29	-0.22	0.04	*				-6.76	4.69	1.05	-0.30	1.36	-2.87	*	
MAQ3	56.	* -0.64	0.44	0.10	-0.03	0.13	-0.27	*											
** VARIABLE 9		TRABAJO										2.85	-0.25	6.01	1.53	-2.94	1.15	*	
TRA1	12.	* 0.78	-0.07	1.65	0.42	-0.80	0.32	*				3.83	1.79	-3.41	-2.83	5.15	-0.93	*	
TRA2	52.	* 0.39	0.18	-0.35	-0.29	0.52	-0.09	*				-5.66	-1.65	-0.34	1.90	-3.36	0.22	*	
TRA3	47.	* -0.63	-0.18	-0.04	0.21	-0.37	0.02	*											
** VARIABLE 10		NACIONALIDAD										6.46	-5.18	-0.29	1.57	-0.79	-1.41	*	
NAC1	84.	* 0.35	-0.28	-0.02	0.08	-0.04	-0.08	*				-6.46	5.18	0.29	-1.57	0.79	1.41	*	
NAC2	27.	* -1.09	0.87	0.05	-0.26	0.13	0.24	*											
** VARIABLE 12		VIVE DE LA FINCA										-3.38	3.61	-2.68	0.82	-2.39	1.76	*	
VIF1	57.	* -0.31	0.34	-0.25	0.08	-0.22	0.16	*				2.69	-3.08	1.28	0.59	6.75	-0.04	*	
VIF2	41.	* 0.33	-0.38	0.16	0.09	0.84	-0.00	*				1.22	-0.99	2.24	-2.31	-6.43	-2.68	*	
VIF3	13.	* 0.32	-0.26	0.59	-0.61	-1.68	0.70	*											
** VARIABLE 14		CREDITO										0.35	3.40	-0.03	-0.46	0.35	3.85	*	
CRE1	26.	* 0.06	0.59	-0.01	-0.08	0.06	0.66	*				3.01	2.27	0.86	-1.11	0.66	-4.56	*	
CRE2	60.	* 0.26	0.20	0.08	-0.10	0.06	-0.40	*				-2.82	-1.41	-0.10	4.24	-3.89	0.78	*	
CRE3	14.	* -0.71	-0.35	-0.03	1.05	-0.98	0.19	*				-2.38	-7.03	1.27	-2.19	2.69	1.29	*	
CRE4	17.	* -0.68	-0.22	-0.37	0.63	0.77	0.37	*											
** VARIABLE 3		SUPERFICIE EN FRUTAL										-1.50	-8.60	2.04	-0.37	1.67	0.31	*	
SUF1	20.	* -0.31	-1.75	0.41	-0.08	0.34	0.06	*				6.36	0.38	-1.32	-0.12	-1.24	1.56	*	
SUF2	25.	* 1.12	0.07	-0.23	-0.02	-0.22	0.28	*				2.52	1.26	-0.49	-0.33	-0.49	0.06	*	
SUF3	24.	* 0.46	-2.03	-0.09	-0.06	-0.09	0.01	*											
SUF4	42.	* -0.78	0.66	-0.01	0.08	0.02	-0.20	*				-6.42	5.42	-0.06	0.68	0.15	-1.64	*	
** VARIABLE 5		CULTIVO ANTERIOR										1.39	2.05	0.60	-2.05	-0.71	-1.50	*	
CAN1	51.	* 0.14	0.21	0.06	-0.21	-0.07	-0.15	*				-1.46	1.18	0.03	0.13	-0.41	-1.02	*	
CAN2	11.	* -0.42	0.33	0.01	0.04	-0.12	0.29	*				-0.52	-2.75	-0.62	1.98	0.98	2.12	*	
CAN3	49.	* -0.06	-0.30	-0.07	0.21	0.10	0.23	*											
** VARIABLE 11		EDAD										-1.93	3.59	1.54	0.56	-0.61	0.86	*	
EDA1	42.	* -0.24	0.44	0.19	0.07	-0.07	0.11	*				-0.01	-1.72	-0.54	-0.09	-0.48	-0.26	*	
EDA2	30.	* -0.00	-0.27	-0.08	-0.01	-0.07	-0.04	*				-1.19	-1.71	-1.72	-1.45	-0.37	0.12	*	
EDA3	14.	* -0.30	0.43	-0.43	-0.36	-0.09	0.03	*				3.20	-0.98	0.15	0.60	1.49	-0.82	*	
EDA4	25.	* 0.57	-0.17	0.03	0.11	0.26	-0.14	*											
** VARIABLE 13		DESTINO DE LA PRODUCCION										-3.07	-8.62	-0.88	-1.42	0.33	0.61	*	
DES1	20.	* -0.62	-1.75	-0.18	-0.29	0.07	0.12	*				6.38	1.05	0.39	0.25	-0.52	1.38	*	
DES2	26.	* 1.10	0.18	0.07	0.04	-0.09	0.24	*				-0.10	2.85	0.05	-0.35	1.15	-3.23	*	
DES3	39.	* -0.01	0.37	0.01	-0.05	0.15	-0.42	*				-3.49	3.55	0.36	1.43	-1.08	1.71	*	
DES4	26.	* -0.60	0.61	0.06	0.25	-0.18	0.29	*											
** VARIABLE 15		ASISTENCIA TECNICA										5.55	4.04	-0.11	0.60	-0.26	-0.79	*	
ATE1	76.	* 0.38	0.26	-0.01	0.04	-0.02	-0.05	*				-5.26	-4.25	-0.10	-0.65	0.09	0.55	*	
ATE2	34.	* -0.76	-0.61	-0.01	-0.09	0.01	0.08	*											
** VARIABLE 16		NIVEL DE VIDA										-4.40	2.12	1.11	-1.31	-0.03	0.96	*	
NIV1	17.	* -0.99	0.48	0.25	-0.29	-0.01	0.22	*				-1.65	2.86	0.63	0.48	1.05	-0.76	*	
NIV2	45.	* -0.19	0.33	-0.07	0.05	0.12	-0.09	*				4.51	-4.19	-1.23	0.52	-0.84	0.27	*	
NIV3	50.	* 0.48	-0.44	-0.13	0.05	-0.09	0.03	*											
** VARIABLE 17		COMPORTAMIENTO DEL INGRESO										3.13	2.78	-0.72	1.40	2.20	0.51	*	
ING1	71.	* 0.22	0.20	-0.05	0.10	0.16	0.04	*				-1.46	0.61	1.03	-2.05	-2.53	-0.88	*	
ING2	18.	* -0.34	0.14	0.24	-0.48	-0.59	-0.20	*				-2.41	-3.76	-0.04	0.11	-0.40	0.16	*	
ING3	24.	* -0.44	-0.68	-0.01	0.02	-0.07	0.03	*											
** VARIABLE 18		PROBLEMAS PLAGAS Y ENFERMEDADES										-1.74	-5.02	0.22	-0.44	-0.17	-0.50	*	
PLE1	14.	* -0.44	-1.26	0.06	-0.11	-0.04	-0.13	*				-1.84	4.32	1.24	1.25	-0.21	2.41	*	
PLE2	34.	* 0.26	0.62	0.18	0.18	-0.03	0.35	*				1.48	-4.25	2.36	-0.81	0.00	-0.12	*	
PLE3	24.	* 0.27	-0.77	0.43	-0.15	-0.00	-0.02	*				1.71	2.98	-3.38	-0.20	0.32	-1.88	*	
PLE4	39.	* 0.22	0.39	-0.44	-0.03	0.04	-0.24	*				-1.02	-1.35	0.46	-0.19	-0.39	0.31	*	
** VARIABLE 19		PROBLEMAS DE MALEZAS										-0.32	-1.55	-1.33	0.78	-1.82	1.59	*	
PMA1	66.	* -0.02	-0.12	-0.10	0.06	-0.14	0.13	*				0.32							

0.872	NAC2	-----	TEC2	-----		
0.834	I	PRO5				
0.797	I					
0.759	I					
0.722	UAE1					
0.695	I	SUF4	UAE5			
0.647	I	RUB1	DES4	PLE2		
0.610	I				CRE1	
0.573	I					
0.535	I					
0.498	NIV1					
0.460	I	MAQ3	EDA1		PRO6	
0.423	I				PLE4	
0.386	I	SUP3	CAN2	VIF1NIV2	DES3	
0.348	I			SUP2		
0.311	I					MAQ1
0.274	I					
0.236	I					
0.199	I					
0.161	I					
0.124	I					
0.087	I					
0.049	I					
0.012	I	LAB3				
-0.026	I					
-0.063	I					
-0.100	I		PRO1			
-0.138	I			PMA1		
-0.175	I		RIE2	TRA3		
-0.212	I				PRO2	EDA4
-0.250	I					
-0.287	I			PAG2		
-0.325	I				EDO2SAT2	VIF3NAC1
-0.362	I		CRE3		CAN3	
-0.399	I		TEC1			RUB4
-0.437	I			PRO7EDA3		
-0.474	I					
-0.511	I					
-0.549	I					
-0.586	I					
-0.624	I		ATE2			
-0.661	I			PRO3		
-0.698	I					
-0.735	I				PLE3	
-0.773	I					
-0.810	I				MAQ2	
-0.848	I					
-0.885	I					
-0.923	I					
-0.960	I					
-0.997	I					
-1.035	I					
-1.072	I					
-1.110	I			UAE2		
-1.147	I					
-1.184	I					
-1.222	I					
-1.259	I					
-----CRE4RUB3UAE3SUF1-----						
-1.086		-0.599	-0.111	0.378	0.864	1.351

POINTS MULTIPLES (100 AU MAXIMUM)

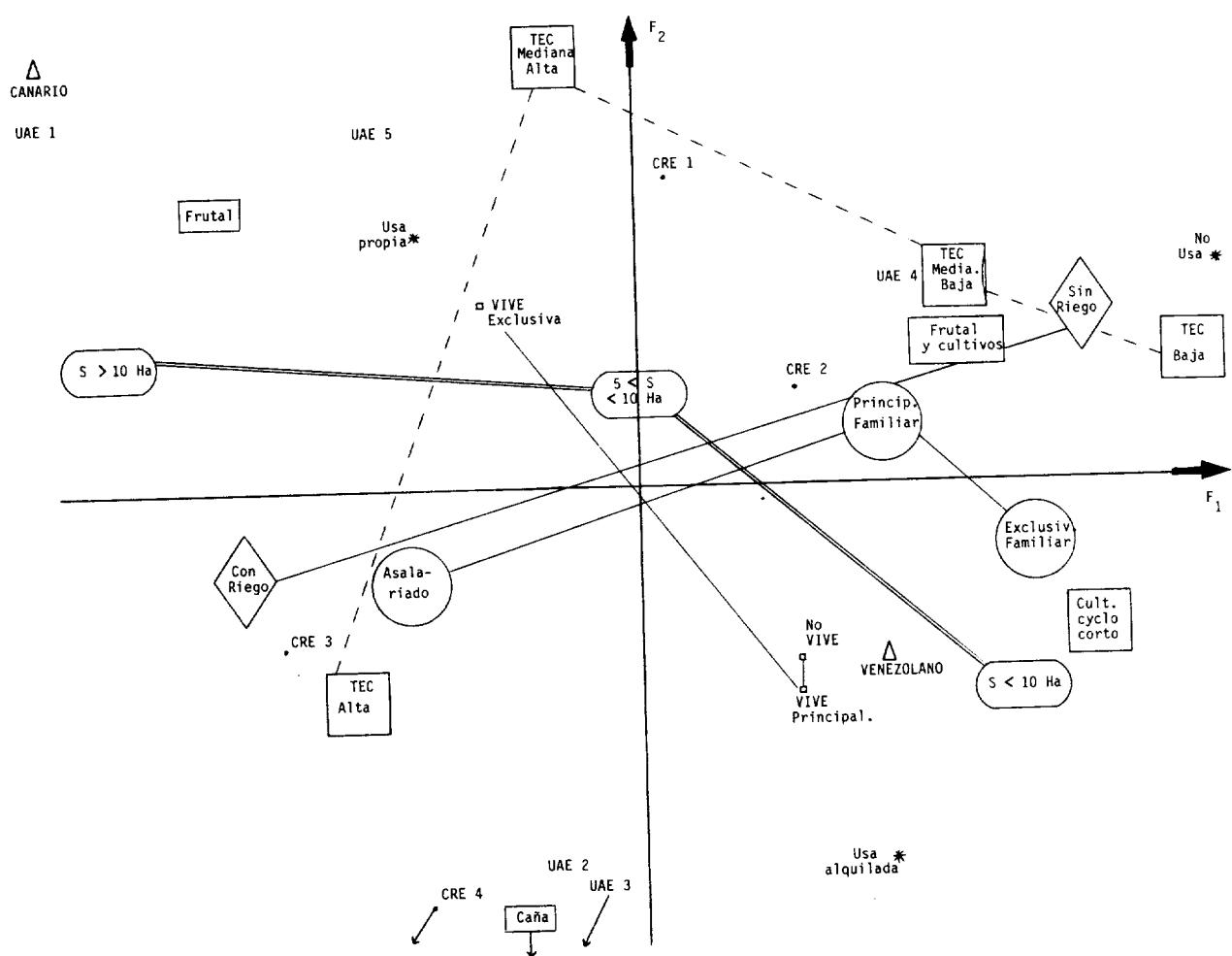


Figura 1: VARIABLES ACTIVAS

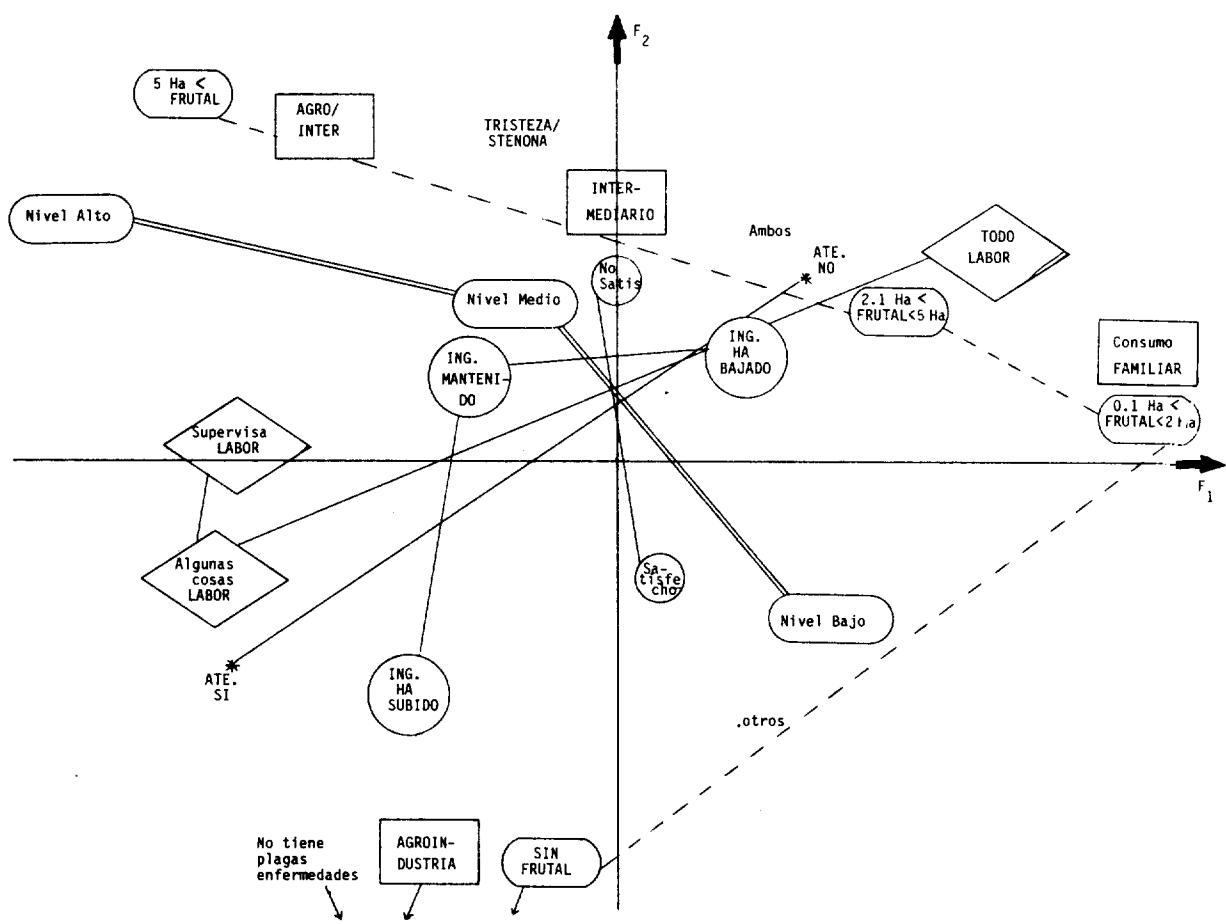


Figura 2: VARIABLES ILUSTRATIVAS

Mise en oeuvre de deux techniques de classification avec SPAD

On peut utiliser ces techniques pour vérifier s'il existe des regroupements dont l'analyse précédente n'aurait pas rendu compte. On peut aussi vouloir directement définir des classes de *fincas*. Ici nous utiliserons deux méthodes de classification hiérarchique: la classification mixte (étape SEMIS) et les voisins réciproques (étape RECIF). Faisons tout d'abord deux remarques valables pour les deux méthodes.

Remarque 1: les deux algorithmes travaillent sur les résultats d'une analyse factorielle, ici les coordonnées / des *fincas* sur les six premiers facteurs.

Remarque 2: le critère de distance retenu dans SEMIS et dans RECIF est celui de la variance:

$$d(a,b) = \frac{m_a m_b}{m_a + m_b} \left[\delta(g_a, g_b) \right]^2$$

où $d(a,b)$ est la distance entre deux parties a et b , m_a et m_b les poids de ces parties, et g_a et g_b leurs centres de gravité et δ la distance Euclidienne. Il existe d'autres méthodes utilisant d'autres critères d'agrégation.

L'étape SEMIS:

a) Idée directrice:

La première étape consiste à obtenir à faible coût une partition de N objets en un nombre k de classes homogènes, ce nombre k étant très supérieur au nombre de classes réelles dans la population observée. Dans une seconde étape on effectue une classification ascendante hiérarchique dans laquelle les éléments terminaux de l'arbre sont précisément les classes de la partition. L'étape finale consiste à effectuer une inspection visuelle de l'arbre, pour le couper au niveau qui soit le plus approprié. On obtient ainsi une partition définitive des N individus en k_0 classes ($k_0 < k < N$) que l'on décrit ensuite.

b) La méthode et le vocabulaire utilisés:

Préalinaire: Il s'agit d'obtenir rapidement un / grand nombre de petits groupes, très homogènes. La procédure travaillant sur la totalité des individus devra / être économique en espace mémoire utilisé. On utilisera ici la partition définie par les groupements stables, résul-

tant du croisement d'un certain nombre de partitions de base. Chaque partition de base est calculée par l'algorithme des centres mobiles réalisé par lecture directe des données, celle-ci assure une faible occupation de / mémoire. En général les calculs seront effectués sur les coordonnées des individus sur les premiers axes principaux d'une analyse, en utilisant la distance euclidienne usuelle. Cette distance est donc équivalente à la distance utilisée dans le système d'axes initial, quelle que soit cette distance (distance euclidienne classique entre valeurs, réduites ou pas ; distance du χ^2)

Agrégation hiérarchique des groupements stables.

Certains des groupements stables peuvent être très proches les uns des autres, correspondant à une classe artificiellement coupée par l'étape précédente. D'autre part la procédure va créer en général certaines classes de petite taille, parfois réduits à un seul élément. La tâche de la phase d'agrégation hiérarchique est justement de construire les classes que se sont éclatées en de nombreux morceaux et d'agréger à leurs pôles d'origine les éléments qui paraissent dispersés.

L'arbre sera construit avec le critère d'agrégation de WARD qui présente l'avantage de prendre en compte / l'importance des éléments à classer (sous la forme de poids ou de pondération dans le calcul de la perte / d'inertie par agrégation). Si l'effectif à classer est faible, on peut utiliser l'algorithme de base moins rapide, l'algorithme de recherche en chaîne des voisins réciproques.

La coupe du dendrogramme définira la partition finale des individus. Les choix du niveau de cette coupe / et du nombre de classes de la partition se fait au vu / de l'arbre: il faut couper au dessus des groupements / bas (qui réunissent les éléments les plus proches), et en dessous des groupements hauts qui ne font que réunir en un seul ensemble les diverses classes de la population observée.

L'exemple des "fincas":

les cinq paramètres de l'étape SEMIS ont été fixés de la manière suivante:

NBASE = 2, le nombre de partitions de base effectuées.
NITER = 10, le nombre d'itérations pour déterminer chaque partition de base.
NCLAS = 5, le nombre de classes dans chaque partition de base.

NKLA = 15, le nombre de groupements stables demandés, le croisement fournit (NCLAS)^{NBASE} classes dont certaines peuvent être vides, les NKLA - 1) premiers groupements / sont les classes à plus fort effectif et la dernière classe regroupe les éléments restants. Ici le croisement pourrait fournir $5^2=25$ classes mais il y en a seulement 9 qui ne sont pas vides, on n'a que 111 fincas. La sortie

listing précise les cinq individus source tirés au hasard pour chacune des deux partitions de base, fournit une description des noeuds de la hiérarchie, une description des classes et un dendrogramme. C'est en voyant cet arbre qu'on peut décider du nombre de classes de la partition finale, ici nous choisirons trois classes, quatre ne se justifierait pas, mais on aurait pu en choisir cinq.

ALGORITHME DES CENTRES-MOBILES ET CONSTRUCTION DES CLASSES STABLES
PARTITION DES 111 INDIVIDUS CARACTERISES PAR 6 COORDONNEES CARTESIENNES

LA PARTITION CONTIENDRA 15 CLASSES
LES 14 PREMIERES CONTIENNENT LES INDIVIDUS LES PLUS STABLES DANS LES 2 PARTITIONS DE BASE.
CHAQUE PARTITION DE BASE EST FORMEE EN 10 ITERATIONS AUTOUR DE 5 INDIVIDUS-SOURCE TIRES AU HASARD

APPEL D'UNE PARTITION, CONSTRUISTE SUR LES INDIVIDUS-SOURCE
POIDS DES CLASSES APRES 10 ITERATIONS 44 26 70 100 78
 18. 19. 28. 25. 21.

APPEL D'UNE PARTITION, CONSTRUISTE SUR LES INDIVIDUS-SOURCE
POIDS DES CLASSES APRES 10 ITERATIONS 105 90 11 73 84
 20. 30. 20. 23. 18.

EFFECTIFS DECROISSANT NON-NULS DANS LES 9 CLASSES DE LA PARTITION CROISEE
SUIVIS DES POURCENTAGES CUMULES

27	20	19	18	18	3	3	2	1
24.3	42.3	59.5	75.7	91.9	94.6	97.3	99.1	100.0

DESCRIPTION DES NOEUDS DE LA HIERARCHIE INDICE = (INDICE REEL)/(INDICE MAX) INDICE MAX = 94.43124

NO	AINE	BENJ	NBRE	POIDS	INDICE
10	8	9	2	3.00	0.005365 **
11	2	7	2	23.00	0.012820 ****
12	1	6	2	30.00	0.028708 *****
13	10	11	4	26.00	0.035096 *****
14	3	13	5	45.00	0.101935 *****
15	4	12	3	48.00	0.136791 *****
16	5	14	6	63.00	0.295638 *****
17	15	16	9	111.00	0.383846 *****

NOEUD	INDICE	AINE	BENJ	EFFECTIF	DESCRIPTION DES CLASSES DE LA HIERARCHIE INDICE = (INDICE REEL)/(INDICE MAX) INDICE MAX = 94.43124										
10	0.005365	8	9	2	8	9									
11	0.012820	2	7	2	2	7									
12	0.028708	1	6	2	1	6									
13	0.035096	10	11	4	8	9	2	7							
14	0.101935	3	13	5	3	8	9	2	7						
15	0.136791	4	12	3	4	1	6								
16	0.295638	5	14	6	5	3	8	9	2	7					
17	0.383646	15	16	9	4	1	6	5	3	8	9	2	7		

POIDS	INDICE	DENDROGRAMME (ECHELLE 0.00537 0.38365) INDICE =(IND.REEL)/(IND.MAX) IND.MAX = 94.43124													
3.000	0.013	7												
20.000	0.035	2												
1.000	0.005	9	...												
2.000	0.102	8												
19.000	0.296	3												
18.000	0.384	5												
3.000	0.029	6												
27.000	0.137	1												
18.000	-----	4												

FIN DE L-ETAPE ** SEMIS **

Couper par ici pour une
partition en 3 classes

Description des trois classes: étape TAMIS.

Cette étape de SPAD permet d'illustrer et d'interpréter une classification d'éléments, que les classes - / soient issues d'un algorithme de classification comme celle que nous venons d'effectuer, ou que les classes -- / soient les modalités d'une variable nominale du fichier des données. On peut obtenir successivement les descriptions de plusieurs partitions (ou peut demander jusqu'à 5 nombres de classes différentes) ici nous ne demandons que la description des trois classes (KPAR = 3). Pour les autres paramètres nous choisissons: KEDIT = 1, pour éditer tous les tableaux croisant la classification avec/toutes les variables nominales, cette sortie est longue nous ne faisons ici figurer que les deux premières pages correspondant aux trois premières variables nominales / (unidad agroecológica, superficie utilizada et Rubro),

JPROX = 30, pour caractériser les 30 modalités les plus typiques de chaque classe-

MINOX = 10, on ne retient que les modalités dont l'effectif est supérieur à 10:

KVAX = 0, pas d'édition pour la caractérisation des classes par les variables continues.

NVMRO = 1, pour l'édition des numéros de classes auxquelles appartiennent les individus.

NVTRI = 0, l'étape TAMIS décrit une classification faite précédemment (par SEMIS) et stockés sur un fichier de travail (NGRI).

DISTRIBUTION DES 111 INDIVIDUS EN 3 CLASSES
EFFECTIFS ET PROFILS PAR QUESTION

UNIDAD AGROECOLOGICA

EFFECTIFS	* 1*	* 2*	* 3*
ALUVIAL 1 *UAE1*	9*	0	9
ALUVIAL 2 *UAE2*	19*	2	10
ALUVIAL 3 *UAE3*	9*	1	0
LAGUNERO *UAE4*	57*	40	17
ASOCIACION *UAE5*	17*	5	12
MARGES -----	111*	48	45

CHI2 = 96.327 , DEG.LIB.= 8 , NOMBRE DES EFF.THEORIQUES INF 5 = 8 , PROB (CHI2 SUP 96.327) ENVIRON 0.000

PROFIL (TOTAL DES COLONNES=1000) * 1* * 2* * 3*

ALUVIAL 1 *UAE1*	81*	0	200	0
ALUVIAL 2 *UAE2*	171*	42	156	556
ALUVIAL 3 *UAE3*	81*	21	0	444
LAGUNERO *UAE4*	514*	833	378	0
ASOCIACION *UAE5*	153*	104	267	0
MARGES -----	1000*	1000	1000	1000

PROFIL (TOTAL DES LIGNES=1000) * 1* * 2* * 3*

ALUVIAL 1 *UAE1*	1000*	0	1000	0
ALUVIAL 2 *UAE2*	1000*	105	368	526
ALUVIAL 3 *UAE3*	1000*	111	0	889
LAGUNERO *UAE4*	1000*	702	298	0
ASOCIACION *UAE5*	1000*	294	706	0
MARGES -----	1000*	432	405	162

SUPERFICIE UTILIZADA

EFFECTIFS	* 1*	* 2*	* 3*
MENOR DE 5 HA *SUP1*	51*	34	4
DE 5.1 A 10 HA *SUP2*	27*	12	13
MAYOR DE 10.1 HA *SUP3*	33*	2	28
MARGES -----	111*	48	45

CHI2 = 51.876 , DEG.LIB.= 4 , NOMBRE DES EFF.THEORIQUES INF 5 = 1 , PROB (CHI2 SUP 51.876) ENVIRON 0.000

PROFIL (TOTAL DES COLONNES=1000) * 1* * 2* * 3*

MENOR DE 5 HA *SUP1*	459*	708	89	722
DE 5.1 A 10 HA *SUP2*	243*	250	289	111
MAYOR DE 10.1 HA *SUP3*	297*	42	622	167
MARGES -----	1000*	1000	1000	1000

PROFIL (TOTAL DES LIGNES=1000) * 1* * 2* * 3*

MENOR DE 5 HA *SUP1*	1000*	667	78	255
DE 5.1 A 10 HA *SUP2*	1000*	444	481	74
MAYOR DE 10.1 HA *SUP3*	1000*	61	848	91
MARGES -----	1000*	432	405	162

RUBRO

EFFECTIFS	* 1*	* 2*	* 3*
FRUTAL *RUB1*	34*	3	29
FRUTAL Y CULTIVOS *RUB2*	57*	43	14
CANA DE AZUCAR *RUB3*	16*	0	0
CULT.DE CICLO CORTO *RUB4*	4*	2	2
MARGES -----	111*	48	45

CHI2 = 138.293 , DEG.LIB.= 6 , NOMBRE DES EFF.THEORIQUES INF 5 = 4 , PROB (CHI2 SUP 138.293) ENVIRON 0.000

PROFIL (TOTAL DES COLONNES=1000) * 1* * 2* * 3*

FRUTAL *RUB1*	306*	63	644	111
FRUTAL Y CULTIVOS *RUB2*	514*	896	311	0
CANA DE AZUCAR *RUB3*	144*	0	0	889
CULT.DE CICLO CORTO *RUB4*	36*	42	44	0
MARGES -----	1000*	1000	1000	1000

PROFIL (TOTAL DES LIGNES=1000) * 1* * 2* * 3*

FRUTAL *RUB1*	1000*	88	853	59
FRUTAL Y CULTIVOS *RUB2*	1000*	754	246	0
CANA DE AZUCAR *RUB3*	1000*	0	0	1000
CULT.DE CICLO CORTO *RUB4*	1000*	500	500	0
MARGES -----	1000*	432	405	162

RIEGO

EFFECTIFS * 1* * 2* * 3*

Commentons brièvement la description des trois classes par les modalités les plus typiques, sur les deux pages qui suivent: 6 valeurs édites:

1) **Critère:** c'est approximativement une variable de LAPLACE-GAUSS centrée réduite mesurant la distance entre / le pourcentage de la modalité dans la classe et son pourcentage dans la population totale, compte tenu des effectifs de la modalité et de la classe. Cette valeur / sert à ranger les modalités par ordre de pouvoir descriptif de la classe.

2) **Proba:** c'est la probabilité que la valeur du coefficient précédent soit dépassée, dans l'hypothèse d'une modalité constituée par tirage aléatoire exhaustif (loi hypergéométrique, voir la note sur la caractérisation statistique d'une classe et les valeurs-test, de A. MORINEAU, dans le Bulletin technique du centre de statistique et informatique appliquées, CESIA, volume 2, 1984, / Numéros 1-2). De façon pratique, on se garde d'interpréter la présence d'une modalité dont le critère est inférieur à 2 ou la PROBA es supérieure à 0,025.

3) **Poids:** il s'agit du nombre pondéré d'individus de la population totale qui ont formulé la réponse, ici c'est le nombre de fincas.

4) **Pourcentage global:** pourcentage associé a "POIDS".

5) **Pourcentage MOD/CLA :** pourcentage des individus de la classe ayant choisi cette modalité de réponse. **Exemple :** Dans la classe 3 il y a 18 fincas, 16 ont répondu caña de azúcar à la question Rubro: MOD/CLA = 88,9 (voir pages de listing).

6) **Pourcentages CLA/MOD :** pourcentage dans la classe des individus ayant choisi cette modalité de réponse. **Exemple** Dans la classe 3 se trouvent les 16 fincas de caña de / azúcar, c'est à dire, toutes: CLA/MOD = 100,0.

Dans notre exemple on voit que nos trois classes sont très typées (voir les deux pages Description des classes) nous ne pouvons pas développer cette description ici.

Etape GRAFK: elle fournit deux types de graphiques

- une projection des centres des classes sur les premiers plans factoriels, choisis par l'utilisateur. Nous ne faisons pas figurer ce graphique ici.

- une projection de la totalité des individus, affectés du numéro de leur classe d'appartenance, voir la page de listing ci-jointe: nos 3 classes sont très nettement disjointes dans l'espace des deux premiers facteurs avec ce jeu de données on obtient 3 classes bien typées/ mais souvenons nous que les deux premiers facteurs sont largement dominants.

DESCRIPTION DES CLASSES

CLASSE	1
RIEGO	
RUBRO	
NACIONALIDAD	
DESTINO DE LA PRODUCCION	
ASISTENCIA TECNICA	
MAQUINARIA	
SUPERFICIE EN FRUTAL	
LABOR QUE REALIZA EL PRODUCTOR	
UNIDAD AGROECOLOGICA	
TECNOLOGIA	
SUPERFICIE UTILIZADA	
TECNOLOGIA	
NIVEL DE VIDA	
TRABAJO	
COMPORTAMIENTO DEL INGRESO	
PROBLEMAS PLAGAS Y ENFERMEDADES	
EDAD	
CREDITO	
SUPERFICIE EN FRUTAL	
PROBLEMA DE AGUA	
PRINCIPAL PROBLEMA	
CULTIVO ANTERIOR	
MAQUINARIA	
PRINCIPAL PROBLEMA	
VIVE DE LA FINCA	
PRINCIPAL PROBLEMA	
TRABAJO	
SE SIENTE SATISFECHO DE LA FINCA	

IDENT CRITERE PROBA POIDS POURCENTAGES GLOBAL MOD/CLA CLA/MOD

	(*) 1*)	48	43.2
-SIN RIEGO	(RIE1) 0.97 0.0	49	44.1
-FRUTAL Y CULTIVOS.	(RUB2) 7.35 0.0	57	51.4
-VENEZOLANO	(NAC1) 7.06 0.0	84	75.7
-CONSUMO FAMILIAR	(DES2) 6.80 0.0	28	23.4
-NO RECIBE	(ATE1) 6.78 0.0	78	68.5
-NO USA	(MAQ1) 6.65 0.0	18	16.2
-DE 0.1 A 2 HA	(SUF2) 6.63 0.0	25	22.5
-TODO	(LAB1) 6.20 0.0	56	50.5
-LAQUERO	(UAE4) 6.15 0.0	57	51.4
-BAJA	(TEG4) 5.36 0.000	18	16.2
-MENOR DE 5 HA	(SUP1) 4.79 0.000	51	45.9
-MEDIANA BAJA	(TEC3) 4.68 0.000	21	18.9
-BAJO	(NIV3) 4.19 0.000	50	45.0
-PRINCIP FAMILIAR	(TRA2) 3.86 0.000	52	46.8
-HA BAJADO	(ING1) 3.59 0.000	71	64.0
-AMBOS	(PLE4) 3.08 0.001	39	35.1
-MAYOR DE 61 ANOS	(EDA4) 3.05 0.001	25	22.5
-NO RECIBE CON PROB	(CRE2) 2.93 0.002	60	54.1
-DE 2.1 A 5 HA	(SUF3) 2.83 0.002	24	21.6
-RIEGO	(PAG3) 2.59 0.005	31	27.9
-AGUA	(PRO2) 2.58 0.005	29	26.1
-LO MISMO	(CAN1) 2.49 0.006	51	45.9
-USA ALQUILADA	(MAC2) 2.23 0.013	37	33.3
-ROBO	(PRO4) 1.97 0.024	14	12.8
-PRINCIPALMENTE	(VIF2) 1.90 0.029	41	36.9
-CREENCIERTE DEL LAGO	(PRO6) 1.42 0.078	20	18.0
-EXCLUSIV FAMILIAR	(TRA1) 1.41 0.079	12	10.8
-SI	(SAT2) 1.37 0.085	60	54.1

DESCRIPTION DES CLASSES

IDENT CRITERE PROBA POIDS POURCENTAGES
GLOBAL MOD/CLA CLA/MOD

CLASSE	2	(* 2 *)	45	40.5
MAQUINARIA	=USA PROPIA (MAQ3)	8.83 0.0	56	50.5 97.8 78.6
NACIONALIDAD	=CANARIO (NAC2)	8.26 0.0	27	24.3 80.0 100.0
SUPERFICIE EN FRUTAL	=MAYOR DE 5 HA (SUF4)	7.66 0.0	42	37.8 80.0 85.7
RUBRO	=FRUTAL (RUB1)	6.52 0.0	34	30.6 64.4 85.3
SUPERFICIE UTILIZADA	=MAYOR DE 10.1 HA (SUP3)	6.33 0.0	33	29.7 62.2 84.8
RIESGO	=CON RIEGO (RIE2)	6.19 0.0	62	55.9 88.9 64.5
NIVEL DE VIDA	=ALTO (NIV1)	4.67 0.000	17	15.3 33.3 88.2
EDAD	=20 A 40 ANOS (EDA1)	4.14 0.000	42	37.8 60.0 64.3
DESTINO DE LA PRODUCCION	=AGRO/INTER (DES4)	4.03 0.000	28	23.4 42.2 73.1
TECNOLOGIA	=ALTA (TEC1)	3.90 0.000	58	52.3 73.3 56.9
TRABAJO	=ASALARIADO (TRA3)	3.30 0.000	47	42.3 60.0 57.4
VIVE DE LA FINCA	=EXCLUSIVAMENTE (VIF1)	3.27 0.001	57	51.4 68.9 54.4
PRINCIPAL PROBLEMA	=PLAGAS (PRO5)	3.24 0.001	11	9.9 20.0 81.8
NIVEL DE VIDA	=MEDIO (NIV2)	3.24 0.001	45	40.5 57.8 57.8
PROBLEMAS PLAGAS Y ENFERMEDADES	=TRISTEZA/STENOMA (PLE2)	3.20 0.001	34	30.6 46.7 61.6
SE SIENTE SATISFECHO DE LA FINCA	=NO (SAT1)	3.04 0.001	51	45.9 62.2 54.9
UNIDAD AGROECOLOGICA	=ASOCIACION (UAE5)	2.98 0.002	17	15.3 26.7 70.6
CULTIVO ANTERIOR	=FRUTAL (CAN2)	2.57 0.005	11	9.9 17.8 72.7
PROBLEMA DE AGUA	=NO TIENE (PAG1)	2.49 0.006	52	46.8 60.0 51.9
DESTINO DE LA PRODUCCION	=INTERMEDIARIO (DES3)	2.30 0.011	39	35.1 46.7 53.8
COMPORTAMIENTO DEL INGRESO	=SE HA MANTENIDO (ING2)	2.18 0.015	16	14.4 22.2 62.5
ASISTENCIA TECNICA	=SI RECIBE (ATE2)	1.97 0.024	34	30.6 40.0 52.9
LABOR QUE REALIZA EL PRODUCTOR	=ALGUNAS COSAS (LAB2)	1.97 0.024	48	43.2 53.3 50.0
TECNOLOGIA	=MEDIANA ALTA (TEC2)	1.63 0.052	14	12.6 17.8 57.1
CREDITO	=NO RECIBE SIN PROB (CRE1)	1.35 0.088	26	23.4 28.9 50.0

DESCRIPTION DES CLASSES

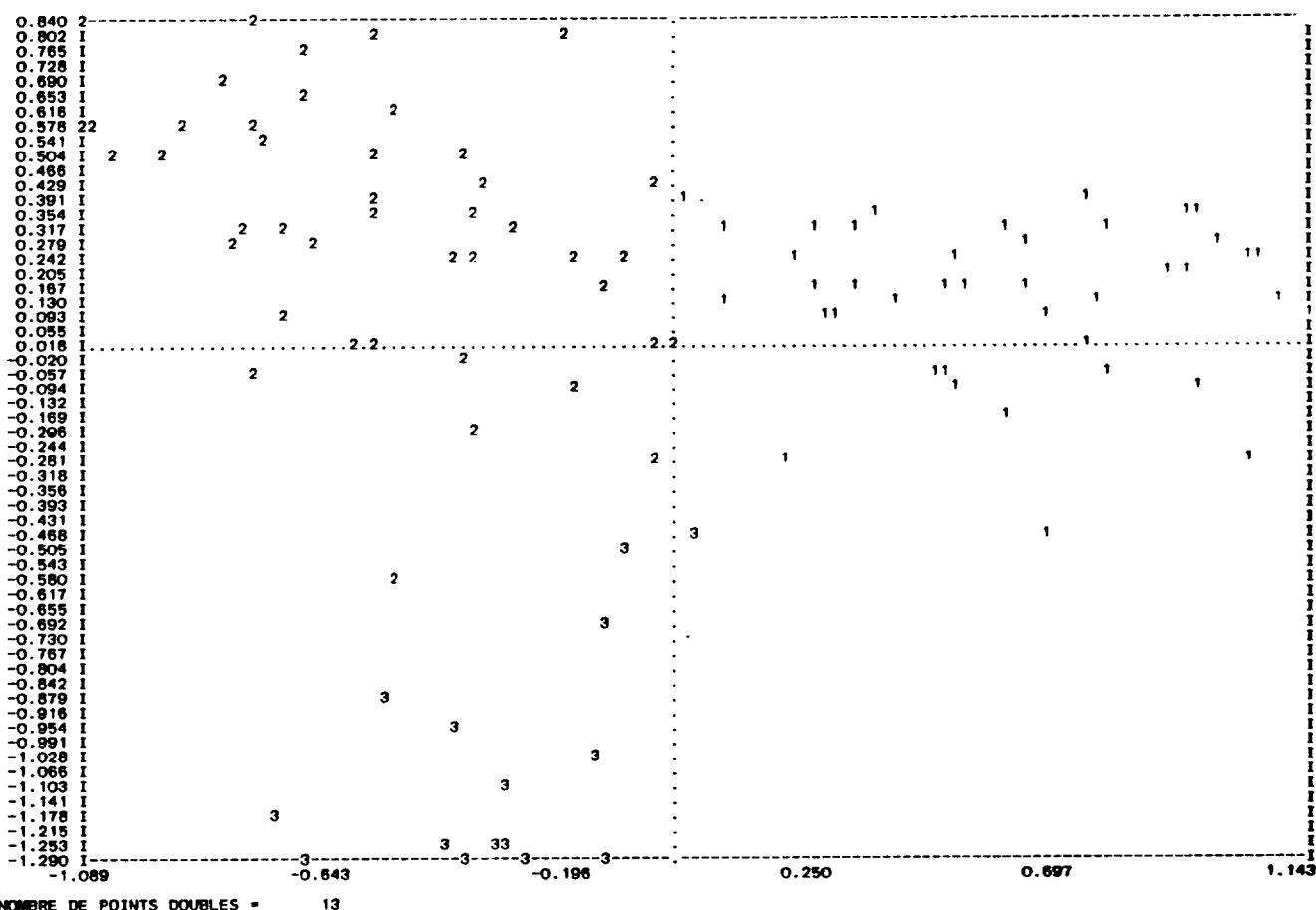
IDENT CRITERE PROBA POIDS POURCENTAGES
GLOBAL MOD/CLA CLA/MOD

CLASSE	3	(* 3 *)	18	16.2
RUBRO	=CANA DE AZUCAR (RUB3)	7.62 0.0	16	14.4 88.9 100.0
DESTINO DE LA PRODUCCION	=AGROINDUSTRIA (DES1)	6.92 0.0	20	18.0 88.9 80.0
SUPERFICIE EN FRUTAL	=SIN FRUTAL (SUF1)	6.92 0.0	20	18.0 88.9 80.0
TECNOLOGIA	=ALTA (TEC1)	6.11 0.0	58	52.3 100.0 31.0
RIESGO	=CON RIEGO (RIE2)	5.88 0.0	62	55.9 100.0 29.0
LABOR QUE REALIZA EL PRODUCTOR	=ALGUNAS COSAS (LAB2)	5.36 0.000	48	43.2 94.4 35.4
ASISTENCIA TECNICA	=SI RECIBE (ATE2)	5.17 0.000	34	30.6 83.3 44.1
CREDITO	=SI RECIBE CON PROB (CRE4)	4.97 0.000	11	9.9 50.0 81.8
MAQUINARIA	=USA ALQUILADA (MAQ2)	4.92 0.000	37	33.3 83.3 40.5
NACIONALIDAD	=VENEZOLANO (NAC1)	4.40 0.000	84	75.7 100.0 21.4
UNIDAD AGROECOLOGICA	=ALUVIAL2 (UAE2)	4.21 0.000	19	17.1 55.6 52.6
COMPORTAMIENTO DEL INGRESO	=HA SUBIDO (ING3)	4.13 0.000	24	21.6 61.1 45.8
CULTIVO ANTERIOR	=CULT DE CICLO CORTO (CAN3)	3.98 0.000	49	44.1 83.3 30.6
PROBLEMAS PLAGAS Y ENFERMEDADES	=NO TIENE (PLE1)	3.89 0.000	14	12.6 44.4 57.1
PROBLEMAS PLAGAS Y ENFERMEDADES	=OTROS (PLE3)	3.14 0.001	24	21.6 50.0 37.5
PRINCIPAL PROBLEMA	=CREDITO (PRO3)	2.86 0.002	22	19.8 44.4 38.4
NIVEL DE VIDA	=BAJO (NIV3)	2.79 0.003	50	45.0 72.2 26.0
PROBLEMA DE AGUA	=DRENAJE (PAQ2)	2.73 0.003	28	25.2 50.0 32.1
SUPERFICIE UTILIZADA	=MENOR DE 5 HA (SUP1)	2.71 0.003	51	45.9 72.2 25.5
SE SIENTE SATISFECHO DE LA FINCA	=SI (SAT2)	2.53 0.006	60	54.1 77.8 23.3
CREDITO	=SI RECIBE SIN PROB (CRE3)	2.24 0.012	14	12.6 27.8 35.7
EDAD	=51 A 60 ANOS (EDA3)	2.24 0.012	14	12.6 27.8 35.7
PROBLEMAS DE MALEZAS	=NO TIENE (PM41)	2.04 0.021	66	59.5 77.8 21.2
EDAD	=41 A 50 ANOS (EDA2)	2.03 0.021	30	27.0 44.4 26.7
VIVE DE LA FINCA	=PRINCIPALMENTE (VIF2)	2.02 0.022	41	36.9 55.6 24.4
TRABAJO	=ASALARIADO (TRA3)	2.00 0.023	47	42.3 61.1 23.4

CLASSIFICATION DES INDIVIDUS

A01 / 1 A02 / 2 A03 / 2 A04 / 1 A05 / 1 A06 / 3 A07 / 1 A08 / 1 A09 / 1 A10 / 2 A11 / 2 A12 / 2 A13 / 1 A14 / 2 A15 / 1 A16 / 2 A17 / 2 A18 / 2 A19 / 3 A20 / 2 A21 / 1 A22 / 1 A23 / 1 A24 / 2 A25 / 1 A26 / 2 A27 / 1 A28 / 2 A29 / 1 A30 / 2 A31 / 3 A32 / 3 A33 / 3 A34 / 2 A35 / 2 A36 / 2 A37 / 3 A38 / 2 A39 / 2 A40 / 2 A41 / 3 A42 / 1 A44 / 1 A45 / 1 A46 / 2 A47 / 1 A48 / 2 A49 / 2 A50 / 1 A51 / 3 A52 / 2 A53 / 2 A54 / 2 A55 / 1 A56 / 2 A57 / 2 A58 / 1 A59 / 1 A60 / 3 A61 / 2 A62 / 3 A63 / 2 A64 / 3 A65 / 1 A66 / 2 A67 / 1 A68 / 1 A69 / 2 A70 / 3 A71 / 1 A72 / 1 A73 / 2 A74 / 2 A75 / 1 A76 / 2 A77 / 3 A78 / 3 A79 / 3 A80 / 2 A81 / 1 A82 / 3 A83 / 2 A84 / 1 A85 / 2 A86 / 2 A87 / 3 A88 / 1 A89 / 1 A90 / 1 A91 / 2 A92 / 2 A93 / 1 A94 / 3 A95 / 1 A96 / 1 A97 / 1 A98 / 2 A99 / 2 B01 / 1 B02 / 1 B03 / 1 B04 / 1 B05 / 2 B06 / 1 B07 / 1 B08 / 1 B09 / 1 B10 / 1 B11 / 1 B12 / 1 B13 / 1 B14 / 2

FIN DE L-ETAPE ** TAMIS **



L'étape RECIP: elle effectue la classification hiérarchique selon la méthode des voisins réciproques avec recherche en chaîne, selon les principes énoncés par J. P. BENZCRI (1982) et MAC QUITTY (1966) et l'algorithme de J. JUAN (1982). C'est un algorithme rapide dont le temps de calcul est toujours en n^2 (si n est le nombre d'individus) et non en n^3 comme dans la procédure classique de LANCE et WILLIAMS. Ici on effectue cette classification sur les 111 fincas à partir de leurs coordonnées sur les 6 premiers axes factoriels (NFAC = 6). On effectue ensuite une coupure du dendrogramme issu de la classification au niveau spécifié par le paramètre NKLA ici NKLA = 28. Nous faisons figurer ici la description/ de la partie de l'arbre après cette coupure en 28 classes, mais pas la description de cette coupure. Ici aussi on peut choisir de couper en 3 classes et demander là description de ces trois classes avec une autre étape TAMIS et une autre étape GRAFK. Nous ne faisons fi-

gurer ici que le graphique obtenu avec GRAFK. Sur ce même jeu de données, entre cette méthode des voisins réciproques (RECIP) et la méthode mixte (SEMIS) les numéros des classes 1 et 2 ont été permutés, ceci n'est pas important pour interpréter, mais remarquons surtout que les trois classes ne sont pas aussi disjointes dans le plan des deux premiers facteurs que les trois classes obtenues avec l'autre méthode de classification.

Avec RECIP on ne choisit que le nombre de facteurs / sur lesquels on veut travailler puis le niveau de coupure de l'arbre. Le résultat est ensuite toujours le même.

Pour l'étape SEMIS le résultat est influencé par le choix des centres (bien que cette influence - / puisse être réduite en augmentant le nombre / d'itérations) d'une part, et par la valeur donnée aux autres paramètres d'autre part.

CLASIFICATION ASCENDANTE HIERARCHIQUE

DESCRIPTION DES NIVEAUX DE LA HIERARCHIE

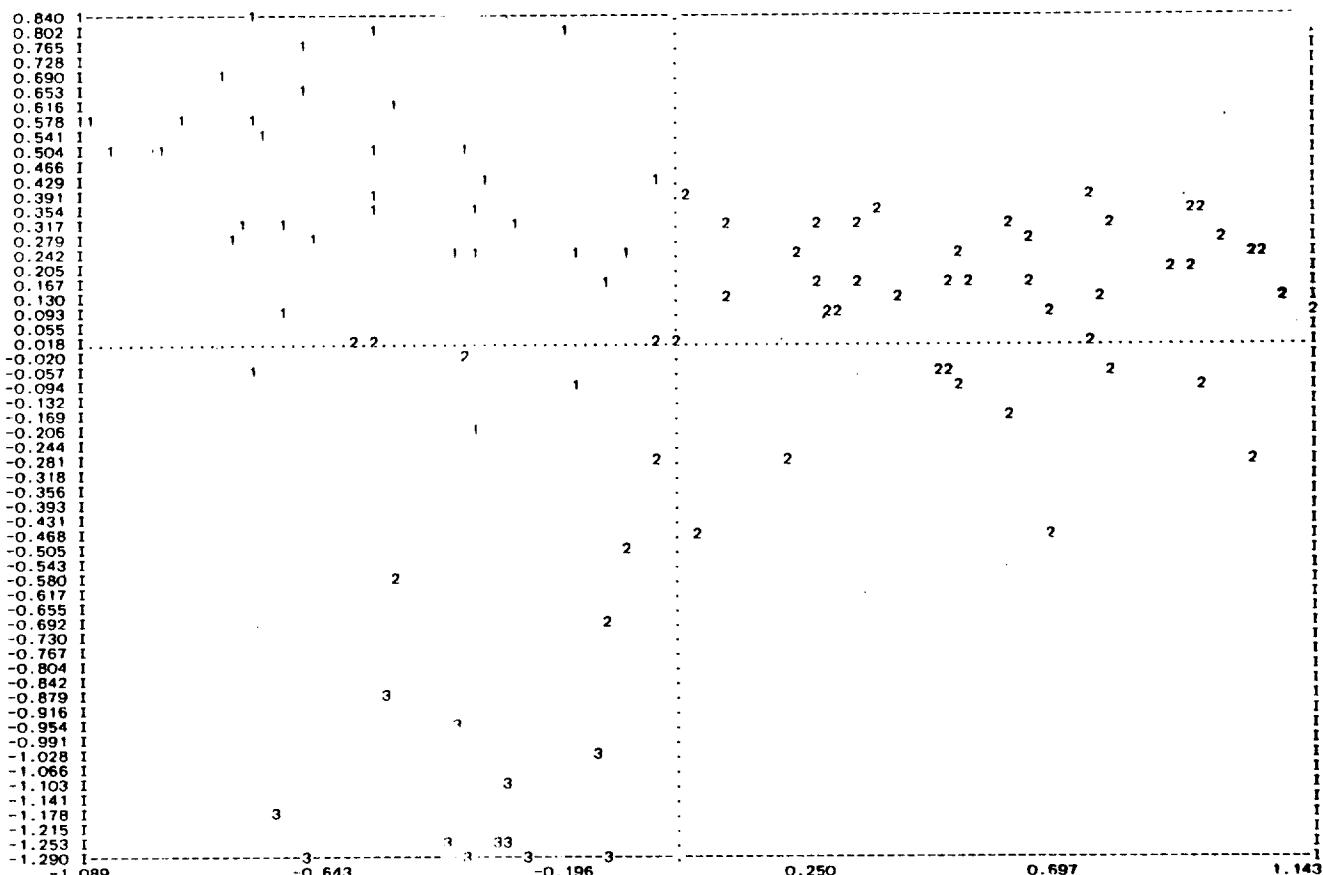
NO	AINE	BENJ	NBRE	POIDS	INDICE	MOTIF
112	99	1	2	2	0.0	
113	45	108	2	2	0.0	
114	113	89	3	3	0.0	
115	106	104	2	2	0.0	
116	115	53	3	3	0.0	
117	95	101	2	2	0.0	
118	39	29	2	2	0.0	
119	36	62	2	2	0.0	
120	38	54	2	2	0.0	
121	75	32	2	2	0.0	
122	31	77	2	2	0.0	
123	42	41	2	2	0.0003818	
124	108	112	3	3	0.0004189	
125	94	56	2	2	0.0004210	
126	25	7	2	2	0.0004265	
127	13	66	2	2	0.0004265	
128	67	16	2	2	0.0004380	
129	10	14	2	2	0.0005033	
130	84	12	2	2	0.0005033	
131	37	50	2	2	0.0005033	
132	107	88	2	2	0.0005153	
133	52	83	2	2	0.0005174	
134	58	76	2	2	0.0005203	
135	92	80	2	2	0.0005257	
136	21	15	2	2	0.0005697	
137	100	9	2	2	0.0005916	
138	33	47	2	2	0.0006013	
139	114	88	4	4	0.0006398	
140	90	46	2	2	0.0007313	
141	117	8	3	3	0.0007596	
142	105	73	2	2	0.0007839	
143	4	110	2	2	0.0008414	
144	87	48	2	2	0.0009095	
145	118	131	4	4	0.0009202	
146	102	78	2	2	0.0009519	
147	43	70	2	2	0.0009628	
148	140	89	3	3	0.0010239	
149	6	61	2	2	0.0010793	
150	20	96	2	2	0.0011152	
151	51	44	2	2	0.0011209	
152	60	68	2	2	0.0011430	
153	135	40	3	3	0.0011890	
154	138	35	3	3	0.0011928	
155	97	146	3	3	0.0012027	
156	19	5	2	2	0.0012076	
157	22	125	3	3	0.0012804	
158	81	71	2	2	0.0012924	
159	141	57	4	4	0.0013215	
160	103	132	3	3	0.0013392	
161	27	147	3	3	0.0014528	
162	144	127	4	4	0.0015531	
163	137	123	4	4	0.0015599	
164	129	120	4	4	0.0016565	
165	139	65	5	5	0.0016643	
166	133	11	3	3	0.0016856	
167	72	55	2	2	0.0016872	
168	34	111	2	2	0.0016877	
169	165	126	7	7	0.0017817	
170	23	82	2	2	0.0018437	
171	30	85	2	2	0.0019382	
172	91	116	4	4	0.0019387	
173	128	26	3	3	0.0019746	
174	17	28	2	2	0.0020335	
175	18	158	3	3	0.0020613	
176	134	119	4	4	0.0020771	
177	130	145	6	6	0.0022923	
178	169	136	9	9	0.0023570	
179	155	59	4	4	0.0026214	
180	121	122	4	4	0.0026302	
181	161	143	5	5	0.0026313	
182	156	3	3	3	0.0026923	
183	177	74	7	7	0.0030849	
184	176	171	6	6	0.0031024	
185	173	150	5	5	0.0031110	
186	153	152	5	5	0.0031569	
187	183	24	8	8	0.0037804	
188	86	63	2	2	0.0038193	
189	93	124	4	4	0.0038668	
190	160	157	6	6	0.0038918	
191	79	170	3	3	0.0041545	
192	168	166	5	5	0.0042115	
193	164	64	5	5	0.0042541	
194	172	162	8	8	0.0043994	
195	142	163	6	6	0.0050684	
196	149	49	3	3	0.0053803	
197	2	185	6	6	0.0054679	
198	151	193	7	7	0.0057711	
199	192	174	7	7	0.0068758	
200	159	194	12	12	0.0078211	
201	196	182	6	6	0.0086910	
202	148	179	7	7	0.0101075	
203	187	154	11	11	0.0104817	
204	181	178	14	14	0.0107084	
205	180	186	9	9	0.0115655	
206	175	201	9	9	0.0119599	
207	197	199	13	13	0.0126691	
208	190	189	10	10	0.0136511	
209	204	208	24	24	0.0156032	
210	198	202	14	14	0.0170313	
211	200	195	18	18	0.0191904	
212	206	191	12	12	0.0194796	
213	205	184	15	15	0.0200641	
214	188	167	4	4	0.0252591	
215	210	207	27	27	0.0384030	
216	212	214	16	16	0.0593368	
217	215	203	38	38	0.0747892	
218	211	209	42	42	0.090324	
219	216	218	58	58	0.1215211	
220	213	219	73	73	0.2632939	
221	217	220	111	111	0.2762651	

SOMME DES INDICES DE NIVEAU = 1.28901

POIDS INDICE DENDROGRAMME (Echelle 0.00433 0.23623)
 INDICE = (IND.REEL)/(IND.MAX) IND.MAX = 1.16949

P. CLAPIER

4.000	0.012	189
6.000	0.013	190
9.000	0.009	178
5.000	0.000	181
4.000	0.004	163	...
2.000	0.016	142
8.000	0.007	194
4.000	0.104	159
2.000	0.022	167
2.000	0.051	188
3.000	0.017	191
3.000	0.007	182
1.000	0.005	49
2.000	0.010	149
3.000	0.225	175
6.000	0.017	184
5.000	0.010	186
4.000	0.236	180
3.000	0.009	154
8.000	0.064	187
2.000	0.006	174
5.000	0.011	192
5.000	0.005	185
1.000	0.033	2
4.000	0.009	179
3.000	0.015	148
5.000	0.005	193
2.000	-----	151



NOMBRE DE POINTS DOUBLES = 13

ANALISIS DE DATOS

Patrick Clapier
Centre National de la Recherche Scientifique
PARIS

INTRODUCCION

Todos los métodos de estadística multidimensional / tienen el mismo fin: tratan a la vez de describir, clasificar y clarificar los datos.

¿Qué tabla va a someterse al análisis? ¿Qué codificaciones y recodificaciones van a efectuarse sobre los datos? ¿Cuál es la mejor medida de las relaciones entre los objetos a describir? ¿Qué método utilizar? Las etapas son numerosas y la interpretación de los re-

sultados depende estrechamente de ellas. Por consiguiente, no se trata de exponer únicamente un conjunto de técnicas, sino también una manera de abordar problemas de escrutinio estadístico.

Las grandes tablas de datos pueden en cierta medida tratarse globalmente dando lugar a síntesis, pudiéndose / también estudiar un fenómeno en varias dimensiones a la vez.

1. PARA QUE SIRVE EL ANALISIS DE DATOS

Haremos aquí un breve repaso sobre la utilidad , el lugar que ocupa dentro de la estadística y las ventajas o particularidades de estos métodos.

1.1 UTILIDAD Y LUGAR ESPECIFICO DE LOS METODOS

En estadística algunos métodos son difíciles y no siempre son útiles. Las técnicas del Análisis de datos son relativamente sencillos puesto que utilizan esencialmente Algebra lineal y son muy útiles al suministrar /

representaciones sintéticas de grandes conjuntos de datos.

Los métodos del Análisis de datos están dentro de la Estadística Descriptiva y juegan el papel de instrumentos de observación.

Existen algoritmos que son completos o potentes en/ uno de los 2 dominios siguientes:

A) Análisis exploratorio, descripción unio bivariada

- (Histogramas, tablas cruzadas, ...)
- B) Modelos estadísticos multivariados. (Regresión, / discriminación, análisis canónico, ...)

Sin embargo, son muy incompletos en el campo intermedio donde nosotros situamos el análisis de datos: métodos a la vez exploratorios y multidimensionales, técnicas que permiten descubrir la estructura, ocasionalmente complicada, de una gran tabla de números de varias dimensiones, y traducirla por otra estructura más sencilla y que la resume mejor.

Se pueden dividir las técnicas del Análisis de datos en 2 grandes familias:

A) Las técnicas factoriales:

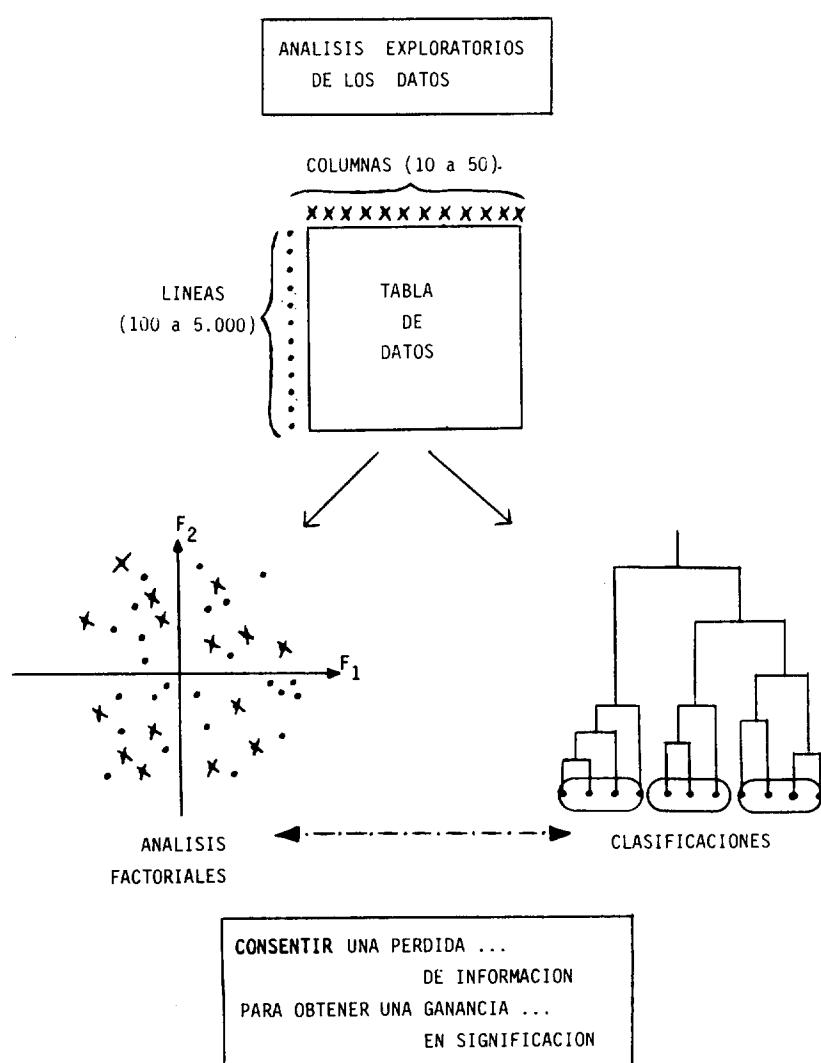
Se entroncan en su origen con las técnicas del análisis factorial propuestas y desarrolladas a comien-

zos de siglo por los sicólogos. Utilizan cálculos de ajuste basados esencialmente en el Algebra lineal. El usuario obtiene mapas estructurando a la vez todas las posiciones relativas del conjunto de filas y columnas de la tabla que quiere describir. Esta estructuración es referenciada en lo que se llaman ejes factoriales. Nosotros enseñaremos a proponer una interpretación de los ejes, o simplemente de las proximidades observadas.

B) Las técnicas de clasificación automática:

Son más recientes, utilizan cálculos algorítmicos, y dan lugar a jerarquías que permiten agrupar y ordenar los objetos a describir. El investigador obtiene representaciones bajo la forma de árboles.

Estos dos tipos de resultados se esquematizan a continuación:



Estas dos familias son más complementarias que concurrentes y pueden utilizarse conjuntamente sobre un mismo conjunto de datos. En la práctica se empieza a menudo por utilizar métodos factoriales para posicionar los objetos a describir unos en relación con otros. // Después, se puede intentar buscar agrupaciones con las técnicas de clasificación y verificar si existen agrupaciones en las que no se habría caído en cuenta con el procedimiento anterior. Esto es lo que tenemos durante las 2 últimas conferencias en la encuesta de 111 Fincas de MARACAY: Primero examinaremos las relaciones entre el conjunto de variables que caracterizan la producción de las granjas, después una clasificación nos permitirá descubrir una categoría muy típica, la de los productores de caña de azúcar.

Se encuentran multitud de métodos en cada una de estas dos grandes familias, desarrollaremos con más detalle los métodos factoriales.

Algunos campos de observación, juzgados en otro tiempo como demasiado grandes y demasiado complejos, pueden ser estudiados ahora. Se puede proceder al estudio de medidas extensivas sin reducciones a priori del campo que se quiera estudiar. Esto suscita una nueva actitud en los investigadores.

1.2. VENTAJAS DE LAS REPRESENTACIONES OBTENIDAS

Dan una visión sintética de las relaciones entre las líneas y las columnas de grandes tablas.

- Se convierten en un nuevo medio de comunicación.
- Permiten comparar los resultados.
- Permiten detectar errores verificando la calidad de la información.

Ilustrémos su función iconográfica con un ejemplo: no haremos aquí más que un comentario muy breve de algunos extractos gráficos de un artículo de Cahiers de l'Analyse des Données (Vol. IV-1979-Nº 4 - Pag. 443-463) Structures des budgets familiaux et des impôts/indirects (BUDG.FAM), por P. CLAPIER y J.L. MADRE.

Los datos provienen de las encuestas permanentes sobre las condiciones de vida de las familias, efectuadas por el INSEE (acumulación de datos relativos a 3 años sucesivos 1969-70-71). La tabla de presupuestos de las familias cruza los 126 tipos de consumo con las 37 categorías-socio-profesionales (CSP) del cabeza de familia. Se trata de una tabla de contingencia cuyo elemento x_{ij} es el consumo total de las familias de la

CPS i para el tipo j.

Contrariamente al artículo, aquí nos limitaremos al plano de los 2 primeros factores. El eje F_1 , (ver gráfico nº 1 y nº 2 en el texto francés) es un eje de ruralidad, resalta lo específico en los presupuestos de las familias agrícolas (explotadores de tierras, asalariados y agricultores jubilados). Esta especificidad tienen principalmente relevancia en el autoconsumo en el terreno alimenticio. Los tipos de autoconsumo (símbolo *) están por otra parte regularmente escalonados sobre este eje (Ver gráfico 1 del texto francés) a partir del origen, en función de la especialización necesaria para su producción: frutos y legumbres, aves de corral, después/cría de bovino o porcino (producciones más especializadas).

Uno se pregunta si este eje manifiesta solamente la oposición entre las familias agrícolas y el resto de la población, o si expresa una escala más continua en función de la zona de hábitat (tamaño del municipio o aglomeración de residencia). El análisis de un fichero basado aproximadamente sobre las mismas encuestas y que cruza los 126 tipos de consumo y la CSP (13 categorías solamente) con la categoría del municipio de residencia (código C.C.D. en el I.N.S.E.E.) ha mantenido esta última hipótesis puesto que ha mostrado en el espacio de los 126 tipos un primer plano factorial aproximadamente igual al previsto para el análisis del fichero de las 37 C.S.P., pero con un escalonamiento de cada C.S.P. más o menos regular sobre el primer eje en función de la categoría de municipio (v. Gráfica nº 3 del texto francés).

La gráfica nº 3 visualiza este escalonamiento paralelo al eje F_1 para cada C.S.P.

El eje F_2 define una escala de estatus de las C.S.P. no agrícolas: en el extremo izquierdo se proyectan las profesiones liberales, las industriales, los grandes comerciantes ..., en el opuesto las categorías de obreros, jornaleros, jubilados.

En el espacio J de los consumos este eje opone los gastos alimenticios (Símbolo .) a los gastos de vacaciones, de empleada de hogar y de cultura y ocios (Símbolo /) excepción hecha para la televisión. Los gastos alimenticios están escalonados entre los gastos de la necesidad (patatas, leche, ...) de lado de los status más bajos y los de productos más elaborados situados cerca del centro de gravedad. Retornaremos a este ejemplo/ después de haber expuesto el Análisis de Correspondencias

1.3 GANANCIA DE PRODUCTIVIDAD EN LOS ESCRUTINIOS DE ENCUESTAS

Apoyándonos en el ejemplo del libro (1): *Techniques de la description statistique*, de L. LEBART, A. MORINEAU M. TABARD (pags. 143-148), abordamos dos fases importantes en la explotación de datos de encuestas:

- a) elección de un conjunto de variables activas y de un conjunto de variables suplementarias.
- b) qué ayuda nos da para la elección de las tablas cruzadas más significativas la lectura de la red socio-administrativa.

Notas sobre el uso de los ordenadores

El uso de los ordenadores ha permitido el desarrollo y la diversificación de los métodos pero con algunos inconvenientes:

- Proporciona siempre un resultado, incluso si se ha hecho una mala elección de codificación de datos o una / mala elección del método.
- Puede dar lugar a empleos abusivos.
- La facilidad de empleo no debe hacernos olvidar / las reglas de interpretación.

2. A QUE SE APLICA EL ANALISIS DE DATOS

2.1 CAMPO DE APLICACION

El campo de aplicación es muy amplio: socio-economía, demografía, agricultura, medicina, marketing ... como se verá con los ejemplos expuestos. En la actualidad estos métodos están muy difundidos tanto en los servicios públicos como en los organismos privados franceses.

En Francia, se hacen numerosos desarrollos teóricos alrededor del profesor BENZECRI, promotor del Análisis Factorial de Correspondencias. (Ver la revista *Les cahiers de l'analyse des données*. BP113; 93103 MONTREUIL/CEDEX; FRANCE).

Estos métodos se utilizan cada vez más en numerosos países. Entre los congresos recientes sobre este tema citamos:

- Les troisièmes journées internationales Analyse de Données et informatique, del 4 al 7 de octubre de 1983 en Versailles (FRANCIA), organizadas por el INRIA (Institut national de recherche en informatique et en automatique).
- LAS JORNADAS DE ANALISIS DE DATOS, encuentro Francia-América Latina, del 11 al 22 de Julio de 1983 en Caracas (Venezuela).

2.2 METODOS ADAPTADOS A TABLAS GRANDES

Estos métodos realmente sólo son útiles cuando las dimensiones de la tabla a analizar son un obstáculo para su lectura y para su asimilación directa. En CREDOC la experiencia demuestra que a menudo se trata de obtener / una descripción de naturaleza estadística para un cierto fenómeno, que ha dado lugar a la recogida de medidas u / observaciones demasiado numerosas e interdependientes // para ser interpretables en una primera lectura. Los ficheros de encuestas socio-económicas son los ejemplos / típicos. Para tratarlos, el recurso del Análisis de datos es particularmente útil.

2.3 LAS CUALIDADES DE UNA TABLA

Hay tres cualidades básicas:

- La pertinencia: esto significa que el objeto en cuestión está bien definido y que el problema tiene sentido.
- La exhaustividad: las diferentes zonas del campo de investigación están bien cubiertas por el estudio.
- La homogeneidad: es preciso preguntarse sobre el / sentido de las operaciones realizadas sobre la tabla: en líneas y columnas. Es preciso imaginarse lo que representa el conjunto de los términos de una línea para escoger un conjunto homogéneo de variables activas (las cuales contribuirán a la construcción de los ejes principales).

3. EN QUE CONSISTE EL ANALISIS DE DATOS

3.1 BIBLIOGRAFIA

Entre los libros más citados se convendrá adoptar / en el texto las siglas siguientes (el nombre del editor figura entre paréntesis):

L. LEBART; A. MORINEAU; N. TABARD [1]

"Techniques de la description statistique. Méthodes et logiciels pour l'analyse des grands tableaux"
(DUNOD) París, 1977

L. LEBART; A. MORINEAU; J.P. FENELON [2]

"Traitement des données statistiques"
(DUNOD) París, 1982, 2^a edición

* Recordemos que el libro ha sido traducido en español y será próximamente editado.

L. LEBART; A. MORINEAU

SPAD

"Système portable pour l'analyse des données"

P. PLEUVRET; E. BRIAN; T. ALUYA; L. LEBART; A. MORINEAU
SPAD (TOME II)

(CESIA) 82 Rue de SEVRES
75007, PARIS

BENZECRI y otros

"L'Analyse des données"

Tome 1: "La Taxinomie"

Tome 2: "L'Analyse des correspondances"

(DUNOD) París, 1976

BENZECRI

"La pratique de l'Analyse des Données" 3 vol.

(DUNOD) París, 1980

L. LEBART; A. MORINEAU; WARWICK

"Multivariate descriptive statistical analysis"

(WILEY) New York, 1983

R. CEHESSAT

"Exercices commentés de statistique et informatique appliquées"

(DUNOD) París, 2^a edición, 1981

J.P. NAKACHE; A. CHEVALIER; V. MORICE

"Exercices commentés de mathématique pour l'analyse statistique des données"
(DUNOD) París, 1981

J.P. FENELON

"Qu'est-ce que l'analyse des données?"
LEFONEN édition, 26 Rue des Cordelières,
75013 Paris, 1981

LAGARDE

"Initiation à l'analyse des données"
(DUNOD) 1983

J.M. BOUROCHE; G. SAPORTA

"L'analyse des données" PUF, Collection "Que sais-je",
París

F. CAILLEZ; J.P. PAGES

"Introduction à l'analyse des données" SMASH,9 Rue Duban
75016 París, 1976

G. SAPORTA

"Théorie et méthodes de la statistique"
(TECHNIP) París, 1978

CUADRAS

"Análisis multivariante"
Barcelona

VIC BARNETT (ed.)

"Interpreting Multivariate Data"
(WILEY) London

E. DIDAY y otros

"Eléments d'analyse des données"
(DUNOD) París, 1983

M. JAMBU; M.O. LEBEAUX

"Classification automatique pour l'analyse des données"

Tome 1: "Méthodes et algorithmes"

Tome 2: "Logiciels"

(DUNOD) París, 1978

BERTIER; BOUROCHE

"Analyse des données multidimensionnelles"

(MASSON) París

3.2 DIFERENTES MÉTODOS FACTORIALES Y DIFERENTES TIPOS TABLAS

3.2.A Los tres principales tipos de tablas son los siguientes:

3.2.A.1 Tablas de medidas:

Para cada individuo de un conjunto I se observan un cierto número de medidas J:

- La altura en centímetros.
- El peso en Kg.
- La edad.
- Etc.

Llamaremos observaciones a las líneas I de la tabla, y variables a las columnas J. Estas medidas son valores numéricos continuos, en lo sucesivo se les llamará variables continuas.

Podemos clasificar estas medidas en clases: por ejemplo para la edad escoger 7 clases:

- 1) Menos de 25 años
- 2) de 25 a 29 años
- 3) de 30 a 34 años
- 4) de 35 a 39 años
- 5) de 40 a 44 años
- 6) de 45 a 49 años
- 7) 50 años y más

Se le llamará entonces variable nominal teniendo 7 modalidades.

3.2.A.2 Las tablas de contingencia

Existe una única población. Por ejemplo los 31079 sujetos de una encuesta. Podemos clasificarlos desde un

1er. punto de vista en 10 categorías socio-profesionales. Podemos clasificarlos también desde un segundo punto de vista J en 8 formas de alojamiento en vacaciones. Se puede entonces proceder al recuento: número de sujetos que poseen a la vez la característica de residir en un pueblo en vacaciones y pertenecer a la categoría de empleados: hay 178. Este es el número que se encuentra en la intersección de la línea empleados y de la columna pueblo de vacaciones. (Ver la tabla cruzada 1 del apartado 3.3.3 donde a un par (i,j) corresponde un efectivo de k_{ij} individuos)

3.2.A.3 Ficheros de encuesta

En una encuesta la mayor parte de las preguntas están pre-codificadas en forma disyuntiva completa, es decir, que las diversas modalidades de respuesta a una pregunta se excluyen mutuamente y una única modalidad es escogida obligatoriamente por el encuestado. Ejemplo de pregunta:

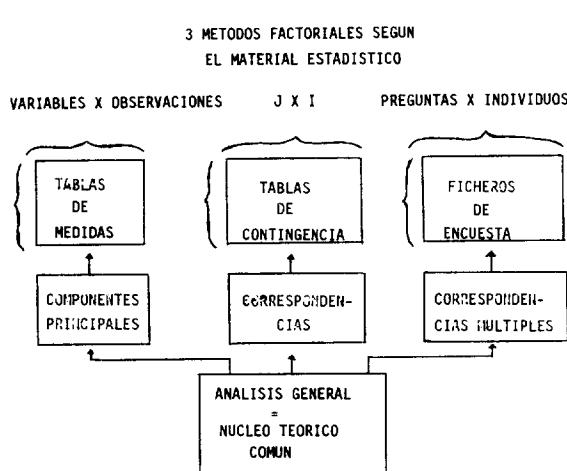
¿ Posee Ud. al menos un coche?

- Tres modalidades de respuesta:
- 1) SI
 - 2) NO
 - 3) Sin responder

Para un fichero de este tipo, se hablaría de los individuos I habiendo respondido a un conjunto de preguntas dado Q.

3.2.B Los diversos métodos factoriales

A continuación esquematizamos la elección de uno de los 3 métodos según el tipo de tabla:



3.3 PRINCIPIOS DE LOS MÉTODOS FACTORIALES Y EJEMPLOS ILUSTRATIVOS

El esquema del parágrafo anterior resume el orden / en el cual se han expuesto los métodos factoriales, con la ayuda de las transparencias hechas por el profesor / MORINEAU.

Todos los puntos esenciales han sido explicados / puesto que es indispensable conocer y comprender las / operaciones efectuadas sobre los datos para poder in-/ terpretar los resultados. Sin embargo, no se han hecho todas las demostraciones durante las conferencias, en ese caso dábamos las referencias bibliográficas necesarias, principalmente las del libro (2) de L. LEBART, A. MORINEAU y J.P. FNELON, "Traitment des données statistiques".

3.3.1 Análisis General (A.G.)

Es el núcleo teórico común a los 3 métodos factoriales, se resuelve el siguiente problema de aproximación/numérica: siendo dada una tabla rectangular de valores/numéricos representada por una matriz X con n líneas y p columnas, de término general x_{ij} es posible reconstruir los np valores x_{ij} a partir de un número menor / de valores numéricos.

La tabla X da lugar a 2 representaciones geométricas: las n líneas de X pueden considerarse como las coordenadas de n puntos en un espacio de p dimensiones \mathbb{R}^p ; donde las p columnas de X pueden representar las coordenadas de p puntos en un espacio/ de n dimensiones \mathbb{R}^n . Se intenta reconstruir aproximadamente las posiciones de los puntos de una nube de \mathbb{R}^p (o de \mathbb{R}^n) a partir de las coordenadas de estos puntos en la base del subespacio que mejor se ajusta a esta / nube en el sentido de los mínimos-cuadrados. Se intenta un ajuste en \mathbb{R}^p , después en \mathbb{R}^n ; y las relaciones entre los 2 subespacios de ajuste. Se obtiene la fórmula de reconstrucción exacta de la tabla de datos numéricos X , después nos limitaremos a una reconstrucción aproximada midiéndose la calidad de la aproximación.

Los métodos de análisis en componentes principales/ y de análisis de correspondencias están presentados a partir de la formulación del análisis general. Pero / cuando se trata de análisis estadístico (y no de aproximación numérica) se dispone más a menudo de informaciones complementarias sobre la naturaleza de los datos el tener en cuenta estas informaciones nos lleva al

realizar algunas transformaciones previas sobre la tabla de partida.

3.3.2 Análisis en Componentes Principales (C.P.)

Se utiliza cuando se trata de describir una tabla R de valores numéricos continuos del tipo Individuos & Variables. Se presentará de hecho todas las etapas del análisis sobre una tabla de dimensiones reducidas, sin // olvidar que el método de descripción no es útil más / cuando las dimensiones de la tabla son un obstáculo para su lectura y su asimilación directa. La tabla R que / servirá para ilustrar las diferentes fases de cálculo da los gastos anuales de $n = 12$ categorías de familias (esos datos son algunos términos medios por categorías) para $p = 7$ tipos de consumo. Los 12 identificadores en 3 caracteres nos informan sobre el nº de niños de la familia (2, 3, 4 o 5) y sobre las características profesionales principales del cabeza de familia (MA = trabajador / manual, EM = empleado no manual, CA = directivos). El / nivel general de los gastos es muy diferente según el tipo de consumo, es decir, las variables son muy heterogéneas en cuanto a sus medias.

Análisis en \mathbb{R}^p

No es la posición de la nube con relación al origen lo que nos interesa, sino la forma de esta nube. Esto / nos lleva a tomar como nuevo origen en \mathbb{R}^p el centro de gravedad G (o punto medio) de la nube cuyas p componentes son las p medias aritméticas \bar{r}_j . El análisis en componentes principales será un caso particular del análisis general de la tabla Y después de la transformación siguiente de la tabla R de partida:

$$R \xrightarrow{} Y = (Y_{ij}) \\ Y_{ij} = \frac{1}{\sqrt{n}} (r_{ij} - \bar{r}_j)$$

La influencia del nivel general de cada una de las variables es así eliminada. El coeficiente $1/\sqrt{n}$ no es más que para hacer coincidir la matriz a diagonalizar / $Y'Y$ con la matriz de covarianzas experimentales.

Una modificación suplementaria de la tabla de partida puede ser igualmente necesaria si las desviaciones / típicas de las variables son muy diferentes; ella nos conducirá al análisis en componentes principales normalizado.

es decir, al análisis de la tabla transformada:

$$R \longrightarrow X = (x_{ij}) \\ x_{ij} = \frac{1}{\sqrt{n}} \frac{r_{ij} - \bar{r}_j}{s_j}$$

$$\text{con } s_j^2 = \frac{1}{n} \sum_i (r_{ij} - \bar{r}_j)^2$$

En efecto en el cálculo del cuadrado de la distancia entre 2 individuos i e i'

$$d^2(i, i') = \sum_{j=1}^p \left\{ \frac{1}{\sqrt{n}} \frac{r_{ij} - \bar{r}_j}{s_j} - \frac{1}{\sqrt{n}} \frac{r_{i'j} - \bar{r}_j}{s_j} \right\}^2$$

la división de cada término de la suma por la varianza s_j correspondiente conseguirá reducir el efecto de las variables muy dispersas sobre el valor de la distancia. Así cada variable tendrá una contribución análoga en la determinación de las proximidades: en este último caso la matriz a diagonalizar $X'X = C$ coincide con la matriz de las correlaciones C entre las variables.

Análisis en \mathbb{R}^n

El análisis general de la tabla transformada X en \mathbb{R}^p induce como se ha visto un análisis en \mathbb{R}^n . Sin embargo, los índices i y j no juegan papeles similares, en esta transformación serán radicalmente diferentes. La distancia entre dos puntos variables i y j' se escribe: $d^2(j, j') = 2(1 - \text{cor}(j, j'))$ donde $\text{cor}(j, j')$ es el coeficiente de correlación entre las variables j y j' . Así, las proximidades entre puntos-variables podrán interpretarse en términos de correlaciones:

- Si su correlación es fuertemente positiva entonces los puntos están muy cercanos:

$$(\text{cor}(j, j') \approx 1) \Rightarrow (d^2(j, j') \approx 0)$$

- Si su correlación es fuertemente negativa entonces los puntos están alejados:

$$(\text{cor}(j, j') \approx -1) \Rightarrow (d^2(j, j') \approx 4)$$

Ejemplo numérico y reglas de interpretación

Tabla R:
Gastos anuales de 12 categorías de familias en 7 tipos de consumo.

	Pan	Legumbres	Frutas	Carne	Aves	Leche	Vino
MA2(*)	332	428	354	1 437	526	247	427
EM2	293	559	388	1 527	567	239	258
CA2	372	767	562	1 948	927	235	433
MA3	406	563	341	1 507	544	324	407
EM3	386	608	396	1 501	558	319	363
CA3	438	843	689	2 345	1 148	243	341
MA4	534	660	367	1 620	638	414	407
EM4	460	699	484	1 856	762	400	416
CA4	385	789	621	2 366	1 149	304	282
MA5	655	776	423	1 848	759	495	486
EM5	584	995	548	2 056	893	518	319
CA5	515	1 097	887	2 630	1 167	561	284
Medias	446,7	737,8	505,0	1 886,7	803,2	358,2	368,6
Desv. Típicas	102,6	172,2	158,1	378,9	238,9	112,1	68,7

* Los identificadores en 3 caracteres nos informan sobre el número de niños de la familia (2, 3, 4 o 5) y sobre las características profesionales principales del cabeza de familia (MA = trabajador manual, EM = empleado no manual, CA = directivo). En realidad estos datos son las medias por categorías (Ver M. Tabard y al., 1967)

El Análisis nos conduce a diagonalizar la matriz C de las correlaciones

	Pan	Legumbres	Frutas	Carnes	Aves	Leche	Vino
Pan	1,00						
Legumbres	0,59	1,00					
Frutas	0,20	0,87	1,00				
Carnes	0,32	0,89	0,96	1,00			
Aves	0,25	0,83	0,93	0,98	1,00		
Leche	0,86	0,66	0,33	0,37	0,23	1,00	
Vino	0,30	-0,35	-0,49	-0,44	-0,40	0,01	1,00

La simple lectura de esta matriz (7x7) nos da una idea de la red de interrelaciones existente entre las variables. Se concibe fácilmente que en el caso de una matriz (100 x 100), el recurso al análisis en componentes principales sea entonces indispensable.

El primer resultado numérico interpretable está constituido por la lista de los valores propios y de los porcentajes de varianza.

	Valores Propios	Porcentajes de Varianza	Porcentajes Acumulados
1	4,339 1	61,99	61,99
2	1,829 7	26,14	88,13
3	0,629 6	8,99	97,12
4	0,129 3	1,85	98,97
5	0,053 2	0,76	99,73
6	0,018 3	0,26	99,99
7	0,000 8	0,01	100,00

0 10 20 30 40 50 60 61,99%
 ──
 26,14%
 ──
 8,99%
 ──
 1,85%
 ──
 0,76%
 ──
 0,26%
 ──────────────────────────────────
 0,01%

La suma de los valores propios vale $p = 7$, traza de la matriz de correlaciones. Los primeros porcentajes son aquí excepcionalmente elevados puesto que existe una concentración bastante clara de la nube en un subespacio / de dos dimensiones (88% de varianza), pero también porque las dimensiones de la tabla son débiles. Para saber si estos porcentajes son excepcionalmente / elevados, una solución empírica consistirá en simular, para después analizar las tablas *seudo-aleatorias* de orden (7 X 12) (Vse. "Validité des résultats" libro (2) de la bibliograffa).

Nos limitaremos aquí a las tres primeras dimensiones para el cálculo de las coordenadas de las variables y de los individuos sobre los ejes factoriales. Los resultados están expuestos a continuación:

Variables	Facteur 1 $u_1 \sqrt{\lambda_1}$	Facteur 2 $u_2 \sqrt{\lambda_2}$	Facteur 3 $u_3 \sqrt{\lambda_3}$
Pan	-0,497	0,841	-0,012
Legumbres	-0,972	0,131	-0,037
Frutas	-0,931	-0,277	0,112
Carnes	-0,963	-0,190	0,161
Aves	-0,912	-0,265	0,280
Leche	-0,584	0,707	-0,357
Vino	0,425	0,649	0,620
Observaciones (o individuos)	$c \bar{x} u_1$	$c \bar{x} u_2$	$c \bar{x} u_3$
MA2	1,078(*)	-0,133	0,154
EM2	0,761	-0,711	-0,512
CA2	0,050	-0,287	0,565
MA3	0,820	0,124	-0,039
EM3	0,681	-0,067	-0,199
CA3	-0,669	-0,534	0,393
MA4	0,378	0,539	-0,109
EM4	0,107	0,248	0,108
CA4	-0,629	-0,686	0,037
MA5	-0,083	1,096	0,222
EM5	-0,777	0,448	-0,390
CA5	-1,718	-0,037	-0,230

(*) Las coordenadas están calculadas por una homotecia, definida por la constante c (ver nota 5)

Nota 1.- Las coordenadas de los puntos variables son inferiores a 1 en valor absoluto: en efecto estos puntos están a la distancia 1 del origen en \mathbb{R}^n y la operación de proyección es contractante: no puede más que disminuir las distancias.

Nota 2.- Las coordenadas están definidas salvo el signo sobre cada eje. La orientación de los ejes es arbitraria y depende del algoritmo de diagonalización utilizado.

Nota 3.- Se sabe que el análisis de la nube de los individuos se hace respecto al centro de gravedad. No es lo mismo para los puntos-variable: la simple consulta de sus coordenadas sobre el primer eje nos muestra que (a excepción del vino) todos los productos están a un mismo lado del origen. Tal disposición traduce el hecho de que la mayor parte de las variables están de hecho correlacionadas positivamente entre ellas. Si para un individuo una variable toma un valor grande, las otras variables toman también un valor grande. Esta característica de numerosas

tablas aparece más a menudo sobre el primer eje, que se llamará factor de talla. Aquí el factor de talla opone los grandes consumidores a los pequeños consumidores.

Nota 4.- Un punto de la esfera de radio unidad en \mathbb{R}^n representa una variable (en la que las n componentes están normalizadas). El valor del producto escalar de dos vectores que unen el origen a 2 puntos de esta esfera no es otro que el coeficiente de correlación de las variables correspondientes (esto es igual al coseno del ángulo de los 2 vectores). Así, las coordenadas de los / puntos variables sobre el primer eje no son otras que los coeficientes de correlación entre las diversas variables y el primer factor, considerado él mismo como variable artificial (esto es una combinación lineal de las variables iniciales). Por ejemplo la variable pan tiene un coeficiente de correlación de -0,497 / con la variable: Factor 1 (cuyos 12 valores figuran en la base de la 1^a columna), o, lo que es lo mismo, forma un ángulo cuyo coseno vale -0,497 con el primer eje. Se podrá buscar interpretar cada eje en función de la combinación de las variables más correlacionadas al factor correspondiente.

Nota 5.- Si la escala de las coordenadas de los puntos variables tiene / una interpretación en términos de correlación, no es lo mismo / para los puntos-individuos. Se aplica a sus coordenadas Xu_α un coeficiente cuyo valor $c = \sqrt{n/p}$ está determinado para asegurar un posicionamiento en el plano compatible con la repartición de los puntos-variables, y permitir así una representación simultánea de las dos nubes.

La figura 1 da una representación gráfica de los puntos variables en el primer plano factorial (ejes 1 y 2). La mayor parte de los puntos están próximos al círculo intersección de la esfera de \mathbb{R}^{12} con el plano factorial. La representación correspondiente para los individuos está dada por la figura 2.

Sobre esta figura, se ha unido por una parte los puntos correspondientes a las familias con un mismo número de hijos (2,3,4 o 5 y más), y por otra parte los puntos correspondientes a los individuos con una misma actividad (Manuales, Empleados, Directivos). La rejilla así construida (deformada del lado de los directivos) muestra que el factor de talla puesto en evidencia sobre la figura (pag. CP11) está ligado por una parte el tamaño de la familia, por otra parte a su status socio-profesional: el gasto total aumenta con el número de los niños y el status social.

La superposición de las figuras 1 y 2 con precauciones de interpretación, representa mejor la visualización, sugiriendo qué variables son responsables de las proximidades. Se vería allí por ejemplo que el importe de los gastos en pan, leche y sobre todo vino acerca a los empleados manuales y no manuales, y los aleja de los directivos.

Análisis de datos - métodos factoriales

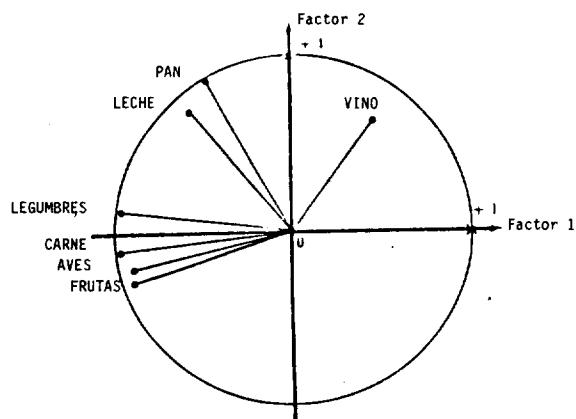


Fig. 1

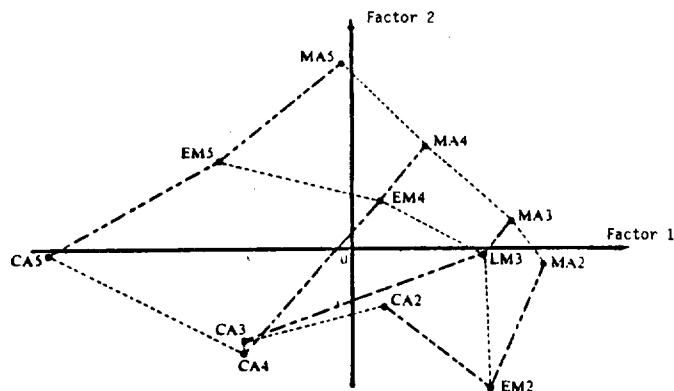


Fig. 2

INTERPRETACION DE LOS EJES

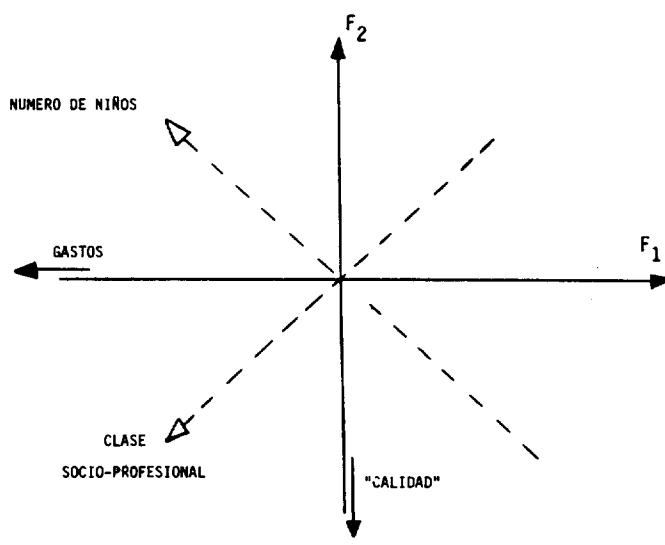


Fig. 3

3.3.3 Análisis de correspondencias (AC)

Este método es indicado para el análisis de las tablas de contingencia.

Se tomará como ejemplo la tabla 1 que cruza 10 categorías socio-profesionales con 8 tipos de alojamiento en vacaciones. En la intersección de la línea i y de la columna j se encuentra el número k_{ij} de individuos pertenecientes a la clase correspondiente. Aquí, este ejemplo de dimensiones reducidas (10×8) ilustrará la exposición, aún cuando el método de descripción es útil sobre todo para tablas más grandes.

En esta tabla, las líneas y las columnas que representan dos particiones de una misma población juegan papeles análogos. Para que las distancias entre puntos-líneas y puntos-columnas tengan un sentido, será útil hacer intervenir los perfiles de las líneas y las columnas es decir, las frecuencias condicionales. Se utilizan las notaciones siguientes:

$$K = \sum_i \sum_j k_{ij} \text{ efectivo total de la tabla}$$

$$f_{ij} = \frac{k_{ij}}{K} \text{ frecuencias relativas}$$

$$\begin{aligned} f_{i.} &= \sum_j f_{ij} \\ f_{.j} &= \sum_i f_{ij} \end{aligned} \quad \left. \begin{array}{l} \text{frecuencias relativas marginales} \\ \text{frecuencias relativas marginales} \end{array} \right\}$$

En la tabla 2 figuran las frecuencias condicionales en porcentaje ($100 \times f_{ij}/f_{i.}$) o perfiles-líneas. En la tabla 3 figuran las cantidades ($100 \times f_{ij}/f_{.j}$) o perfiles-columnas. Estas cantidades definen las coordenadas en los 2 espacios (salvo que se utilizaran proporciones y no porcentajes).

Construcción de nubes de puntos:

En el espacio \mathbb{R}^p se construirá una nube de n puntos, cada punto i tiene por coordenadas:

$$\left\{ \frac{f_{ij}}{f_{i.}} \right\} \quad j = 1 \dots p$$

En el espacio \mathbb{R}^n se construirá una nube de p puntos, cada punto j tiene por coordenadas:

$$\left\{ \frac{f_{ij}}{f_{.j}} \right\} \quad i = 1 \dots n$$

Las proximidades entre puntos se interpretan enton-

ces en términos de proximidades entre perfiles: En la tabla 2 se comparan las diferencias de comportamiento de las 10 categorías socio-profesionales respecto a su tipo de alojamiento en vacaciones. En la tabla 3 se compara los tipos de alojamiento según sus perfiles socio profesionales.

Nota: Los n puntos de \mathbb{R}^p están situados en un subespacio de $p-1$ dimensiones puesto que para todo i las coordenadas verifican la relación:

$$\sum_j \left\{ \frac{f_{ij}}{f_{i.}} \right\} = 1$$

Los p puntos de \mathbb{R}^n están igualmente situados en el subespacio de $n-1$ dimensiones puesto que para todo punto j las coordenadas verifican la relación

$$\sum_i \left\{ \frac{f_{ij}}{f_{.j}} \right\} = 1$$

Elección de las distancias:

El hecho de trabajar sobre los perfiles en los 2 espacios \mathbb{R}^p y \mathbb{R}^n nos incita a escoger la distancia del "CHI-2":

La distancia entre dos categorías i e i' será:

$$d^2(i, i') = \sum_{j=1}^p \left\{ \frac{1}{f_{.j}} \left(\frac{f_{ij}}{f_{i.}} - \frac{f_{i'j}}{f_{i'.}} \right)^2 \right\}$$

Igualmente la distancia entre dos tipos de alojamiento j y j' será:

$$d^2(j, j') = \sum_{i=1}^n \left\{ \frac{1}{f_{i.}} \left(\frac{f_{ij}}{f_{.j}} - \frac{f_{ij'}}{f_{.j'}} \right)^2 \right\}$$

Esta distancia ponderada (el punto i está afectado por la masa $f_{i.}$, el punto j por la masa $f_{.j}$) tiene sobre todo la ventaja de verificar el principio de equivalencia distribucional: si dos puntos líneas i_1 e i_2 son confundidos en \mathbb{R}^p y si se les considera como un sólo punto afectado de la suma de las masas de i_1 y de i_2 (i_1 e i_2 serán reemplazados por i_0), entonces las distancias entre todos los pares de puntos en \mathbb{R}^p y en \mathbb{R}^n no se alteran. Lo mismo para dos puntos-columna j_1 y j_2 de \mathbb{R}^n teniendo las mismas propiedades.

Esta propiedad es muy importante porque podrá tener una influencia estabilizadora sobre los resultados. Así, la agregación de tipos que tienen perfiles vedados

(en líneas o columnas) no influirá en general en los resultados obtenidos. Esto nos garantiza en cierta medida contra la arbitrariedad de toda cuantificación.

Análisis en \mathbb{R}^p

La distancia del "CHI-2" escrita anteriormente no es una suma de cuadrados, el problema no se presentaba en el análisis en componentes principales. Volveremos / sin embargo al supuesto modificando la escala de los ejes. Si se escoge tomar para las p coordenadas del punto i las cantidades:

$$\frac{f_{ij}}{f_i \cdot \sqrt{f_{..j}}}$$

entonces la distancia euclídea usual de dos puntos i e i' vale:

$$d^2(i, i') = \sum_{j=1}^p \left\{ \frac{f_{ij}}{f_i \cdot \sqrt{f_{..j}}} - \frac{f_{i'j}}{f_{i'} \cdot \sqrt{f_{..j}}} \right\}^2$$

y se verifica fácilmente que esta distancia coincide con la distancia del CHI-2. Luego se vuelve de nuevo a un análisis simple tomando como coordenadas de los puntos de la nube las cantidades $f_{ij}/f_i \cdot \sqrt{f_{..j}}$. La nube está ahora en un hipoplano de ecuación:

$$\sum_{i=j}^p \sqrt{f_{..j}} y_j = 1$$

Debemos ahora buscar la primera componente principal de la nube de puntos así construida, puesto que nos interesamos por la forma de esta nube, y no por su posición respecto al origen.

Calculamos las coordenadas del centro de gravedad / de la nube en \mathbb{R}^p . Puesto que cada punto i está afectado de la masa f_{ij} , la j-ésima componente del centro de gravedad G se escribe:

$$g_j = \sum_{i=1}^n f_{ij} \left(\frac{f_{ij}}{f_i \cdot \sqrt{f_{..j}}} \right) = \sqrt{f_{..j}}$$

Después de la traslación del origen al centro de gravedad, las coordenadas del punto i se escriben $(f_{ij}/f_i \cdot \sqrt{f_{..j}}, \sqrt{f_{..j}})$. Este punto i está afectado de la masa f_{ij} , que interviene en el criterio de ajuste del subespacio dirigido por el vector unitario u. Si se designa por $\hat{\psi}_i$ la proyección del punto i sobre este eje u se tendrá:

$$\hat{\psi}_i = \sum_j (f_{ij}/f_i \cdot \sqrt{f_{..j}} - \sqrt{f_{..i}}) u_j$$

La cantidad a maximizar para la determinación de u será: $\sum_i \hat{\psi}_i^2$. Así se trata de diagonalizar la matriz T de las covarianzas, cuyo término general se escribe:

$$t_{jj} = \sum_{i=1}^n f_{ij} \left(\frac{f_{ij}}{f_i \cdot \sqrt{f_{..j}}} - \sqrt{f_{..j}} \right) \left(\frac{f_{ij}}{f_i \cdot \sqrt{f_{..j}}} - \sqrt{f_{..j}} \right)^T$$

Es posible dar a esta matriz T una forma simple. Ponemos en efecto:

$$x_{ij} = (f_{ij} - f_i \cdot f_{..j}) / \sqrt{f_i \cdot f_{..j}}$$

Entonces la matriz T a diagonalizar se expresa en función de la tabla (n,p) notada X:

$$T = X^T X$$

El análisis de correspondencias conduce a un análisis general de la tabla X, notemos que la transformación de los datos hace intervenir de manera simétrica los índices i y j. La transformación utilizada para el análisis en componentes principales era disimétrica.

Análisis en \mathbb{R}^n

Los conjuntos puestos en correspondencia en la tabla de las frecuencias juegan papeles análogos: el análisis en \mathbb{R}^n puede deducirse entonces del de \mathbb{R}^p por permutación de los índices i y j.

Así, las coordenadas del punto j serán ahora las cantidades $f_{ij}/f_{..j} \sqrt{f_{..i}}$. Este punto j estará provisto de la masa $f_{..j}$. La i-ésima coordenada del centro de gravedad H de la nube de los p puntos se escribe:

$$h_i = \sqrt{f_{..i}}$$

La matriz de covarianza a diagonalizar tendrá término general:

$$w_{ii} = \sum_{j=1}^p f_{..j} \left(\frac{f_{ij}}{f_{..j}} \sqrt{f_{..i}} - \sqrt{f_{..i}} \right) \left(\frac{f_{ij}}{f_{..j}} \sqrt{f_{..i}} - \sqrt{f_{..i}} \right)^T$$

La misma transformación anterior nos conduce a un análisis general de la tabla X de término general

$$x_{ij} = (f_{ij} - f_i \cdot f_{..j}) / \sqrt{f_i \cdot f_{..j}}$$

Entonces la matriz W se expresa en función de la tabla / X :

$$W = X X'$$

Relación entre los dos espacios \mathbb{R}^p y \mathbb{R}^n

Se hará referencia a la bibliografía para la demostración de las relaciones de doble transición entre las coordenadas de los puntos-líneas y de los puntos-columnas sobre un eje α :

$$\hat{\psi}_{\alpha i} = \frac{1}{\sqrt{\lambda_{\alpha}}} \sum_{j=1}^p (f_{ij}/f_{.j}) \hat{\varphi}_{aj}$$

En esta relación, la matriz de término general - ///($f_{ij}/f_{.j}$), que permite calcular las coordenadas de un // punto i a partir de las coordenadas de todos los puntos j, no es otra que la tabla de los perfiles- líneas// (divididos por 100) de la tabla 2. Así en nuestro // ejemplo, la coordenada de una profesión se ob- // tiene como producto por ($1/\sqrt{\lambda_{\alpha}}$) de la abscisa sobre el eje α del baricentro de todos los tipos de alojamiento, siendo los pesos los elementos del perfil de los alojamientos de esta profesión.

De la misma manera la relación siguiente:

$$\hat{\varphi}_{aj} = \frac{1}{\sqrt{\lambda_{\alpha}}} \sum_{i=1}^n (f_{ij}/f_{.i}) \hat{\psi}_{\alpha i}$$

muestra que la coordenada de un punto-tipo de alojamiento es el homotético del baricentro de los puntos-profesiones, con pesos los elementos del perfil socio-profesional de ese tipo de alojamiento, es decir las columnas de la tabla 3.

Notas: Estas dos relaciones no son casos particulares de las establecidas en el caso del análisis general:

$$U_{\alpha} = \frac{1}{\sqrt{\lambda_{\alpha}}} X' U_{\alpha}$$

$$U_{\alpha} = \frac{1}{\sqrt{\lambda_{\alpha}}} X U_{\alpha}$$

En efecto, aquí las matrices de paso no son traspuestas la una de la otra.

Por otra parte estas relaciones implican que para todo α , se tiene $\lambda_{\alpha} \leq 1$: todos los valores propios son inferiores a 1. En efecto, si existiera α tal que $\lambda_{\alpha} > 1$, entonces se tendría $1/\sqrt{\lambda_{\alpha}} < 1$; la segunda relación expresaría que las $\hat{\varphi}_{aj}$ están en el interior (límites excluidos) del intervalo cubierto por las $\hat{\psi}_{\alpha i}$, cuando la primera relación expresaría la inclusión estricta inversa, lo que es imposible.

Otra presentación del análisis de correspondencias permite mostrar el carácter óptimo del análisis de correspondencias en materia de representaciones simultáneas, esto es la búsqueda de la mejor presentación simultánea de // dos conjuntos líneas y columnas.

Ayudas para la interpretación:

Para interpretar los ejes determinados desde un análisis de correspondencias, calcularemos dos series de / coeficientes para cada uno de los elementos de los dos conjuntos puestos en correspondencia:

1) las contribuciones absolutas, que expresan la parte tomada por un elemento dado en la varianza explicada por un factor;

2) las contribuciones relativas, o correlaciones elemento-factor, que expresan la parte tomada por un factor en la explicación de la dispersión de un elemento.

Aún cuando las contribuciones absolutas permitirán / saber qué variables son responsables de la construcción/ de un factor, las contribuciones relativas exhibirán las que son características exclusivas de este factor.

La utilización de estos coeficientes será precisada/ con ocasión del ejemplo numérico.

Ejemplo numéricoDesde el análisis de la tabla K_{ij}

Tabl. 1

Categoría Socio-Profesional	Tipo de Albergue									TOTAL
	Hotel Pensión	Casa de Alquiler	Casa propia	Casa de los padres	Casa de amigos	Camping Caravana	Pueblo de Vacaciones	Varios		
Agricultores	160	28	0	321	36	141	45	65	796	
Asalariados Agrícolas	35	34	1	178	8	0	4	0	260	
Patrones Ind. y Com.	700	354	229	959	185	292	119	140	2 978	
Cuadros Sup. Profesiones Liberales	961	471	633	1 580	305	360	162	148	4 620	
Cuadros medios	572	537	279	1 689	206	748	155	112	4 298	
Empleados	441	404	166	1 079	178	434	178	92	2 972	
Obreros	783	1 114	387	4 052	497	1 464	525	387	9 209	
Personal de Servicio	65	43	21	294	79	57	18	6	583	
Otros Activos	77	60	189	839	53	124	28	53	1 423	
Inactivos	741	332	327	1 789	311	236	102	102	3 940	
TOTAL	4 535	3 377	2 232	12 780	1 858	3 856	1 336	1 105	31 079	

Tabl. 2

Perfiles-línea

$$\frac{f_{ij}}{f_j} \times 100$$

	Hotel Pensión	Casa de Alquiler	Casa propia	Casa de los padres	Casa de amigos	Camping Caravana	Pueblo de Vacaciones	Varios	TOTAL
Agricultores	20,1	3,5	0,0	40,4	4,5	17,7	5,7	8,1	100,00
Asalariados Agrícolas	13,4	13,1	0,4	68,5	3,1	0,0	1,5	0,0	100,00
Patrones Ind. y Com.	23,5	11,9	7,7	32,2	6,2	9,8	4,0	4,7	100,00
Cuadros Sup. Profesiones Liberales	20,8	10,2	13,7	34,2	6,6	7,8	3,5	3,2	100,00
Cuadros medios	13,3	12,5	6,5	39,3	4,8	17,4	3,6	2,6	100,00
Empleados	14,8	13,6	5,6	36,3	6,0	14,6	6,0	3,1	100,00
Obreros	8,5	12,1	4,2	44,0	5,4	15,9	5,7	4,2	100,00
Personal de Servicio	11,1	7,4	3,6	50,4	13,6	9,8	3,1	1,0	100,00
Otros Activos	5,4	4,2	13,3	59,0	3,7	8,7	2,0	3,7	100,00
No activos	18,8	8,4	8,3	45,4	7,9	6,0	2,6	2,6	100,00
TOTAL	14,6	10,9	7,2	41,0	6,0	12,4	4,3	3,6	100,00

Tabla 3.

Perfiles-columnas

$$\frac{f_{ij}}{f_j} \times 100$$

Agricultores	3,5	0,8	0,0	2,5	1,9	3,6	3,4	5,9	2,6
Asalariados Agrícolas	0,8	1,0	0,0	1,4	0,4	0,0	0,3	0,0	0,8
Patrones Ind. y Com.	15,5	10,5	10,3	7,5	10,0	7,6	8,9	12,7	9,6
Cuadros sup. Profesiones Liberales	21,2	13,9	28,4	12,4	16,4	9,3	12,2	13,4	14,9
Cuadros medios	12,6	15,9	12,6	13,2	11,1	19,4	11,6	10,1	13,8
Empleados	9,7	12,0	7,4	8,4	9,6	11,3	13,3	8,3	9,6
Obreros	17,3	33,0	17,3	31,7	26,8	38,0	39,3	35,0	29,5
Personal de Servicio	1,4	1,3	0,9	2,3	4,2	1,5	1,3	0,5	1,9
Otros Activos	1,7	1,8	8,5	6,6	2,9	3,2	2,1	4,8	4,6
Inactivos	16,3	9,8	14,6	14,0	16,7	6,1	7,6	9,3	12,7
TOTAL	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00

Los primeros resultados a consultar son la traza, que vale 0,0839, y los valores propios y porcentajes de inercia.

	Valores propios	Porcentajes de Varianza		Porcentajes Acumulados
1	0,044 3	53,01		53,01
2	0,020 2	24,22		77,23
3	0,008 8	10,53		87,76
4	0,005 4	6,54		94,30
5	0,002 3	2,85		97,15
6	0,002 1	2,57		99,71
7	0,000 2	0,29		100,00

Se observa primero que el soporte de la nube es un espacio con $8-1 = 7$ dimensiones. Contrariamente al análisis en componentes principales, la traza aquí no es constante; por otra parte los valores propios en sí mismos tienen una interpretación: intervienen en el sesgo / del coeficiente $1/\sqrt{\lambda_i}$ para caracterizar la calidad de la representación simultánea.

La tabla 4 da los valores de las masas relativas de las coordenadas, contribuciones absolutas y relativas para los tres primeros factores del análisis.

Tabla 4

Puntos-columnas (tipo de albergue)

	Masas $f_{.j}$	Coordenadas			Contribuciones absolutas			Contribuciones relativas		
		$\hat{\varphi}_{1j}$	$\hat{\varphi}_{2j}$	$\hat{\varphi}_{3j}$	$ca_1(j)$	$ca_2(j)$	$ca_3(j)$	$cr_1(j)$	$cr_2(j)$	$cr_3(j)$
Hotel	0,146	0,33	0,17	0,09	35,7	22,0	12,5	0,73	0,20	0,05
Alquiler	0,109	-0,08	0,14	-0,06	1,5	10,2	4,4	0,12	0,37	0,07
Propiedad	0,072	0,41	-0,13	-0,26	26,9	6,1	56,3	0,65	0,07	0,27
Parientes	0,411	-0,06	-0,14	0,04	3,8	41,7	6,9	0,15	0,78	0,06
Amigos	0,060	0,11	-0,01	0,12	1,7	0,0	9,1	0,20	0,00	0,21
Camping	0,124	-0,30	0,14	-0,09	25,4	12,1	10,3	0,72	0,16	0,06
Aldeas	0,043	-0,21	0,16	-0,02	4,4	5,6	0,1	0,47	0,27	0,00
Diversos	0,036	-0,08	0,11	0,03	0,5	2,2	0,5	0,07	0,12	0,01

Puntos-líneas (categoría socio-profesional)

	Masas $f_{i.}$	Coordenadas			Contribuciones absolutas			Contribuciones relativas		
		$\hat{\psi}_{1i}$	$\hat{\psi}_{2i}$	$\hat{\psi}_{3i}$	$ca_1(i)$	$ca_2(i)$	$ca_3(i)$	$cr_1(i)$	$cr_2(i)$	$cr_3(i)$
Agricultores	0,026	-0,14	0,18	0,24	1,1	3,9	17,3	0,08	0,13	0,25
Asalar. Agrícolas	0,008	-0,04	-0,39	0,35	0,0	6,1	11,4	0,00	0,34	0,27
Patrones	0,096	0,21	0,18	0,06	9,5	16,0	3,5	0,52	0,40	0,04
Cuadros sup.	0,149	0,33	0,02	-0,10	35,6	0,3	16,7	0,90	0,00	0,08
Cuadros medios	0,138	-0,10	0,06	-0,07	3,2	2,4	8,3	0,33	0,11	0,17
Empleados	0,096	-0,07	0,13	-0,01	1,0	8,0	0,2	0,18	0,62	0,01
Obreros	0,296	-0,24	-0,01	-0,01	37,3	0,1	0,2	0,97	0,00	0,00
Pers. de Servicio	0,019	-0,04	-0,20	0,24	0,1	3,7	12,2	0,01	0,22	0,31
Otros	0,046	0,01	-0,47	-0,13	0,0	50,7	8,8	0,00	0,86	0,07
Inactivos	0,127	0,21	-0,12	0,12	12,2	8,6	21,5	0,59	0,19	0,21

Las proximidades observables sobre el primer plano factorial están esquematizadas sobre la figura 1. Está claro que esta representación es más evidente que las tablas 1, 2 y 3. Sin embargo en el caso de recogidas de datos de dimensiones tan modestas, el análisis de correspondencias tendrá sobre todo función ilustrativa.

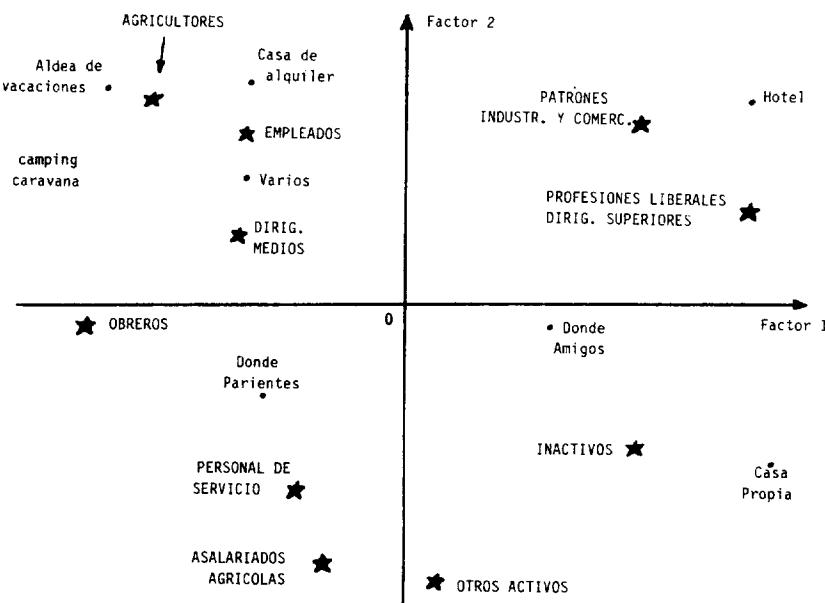


Fig. 1. Proximidades entre tipos de albergue en vacaciones
y categorías socio-profesionales

Vamos, a propósito de este ejemplo, a comentar las reglas de lectura y la interpretación del todo:

La proximidad entre dos puntos-categorías socio-profesionales significará una similitud de los / perfiles de alojamiento (líneas de la tabla 2). Se constata para los Patrones y las Profesiones Li-/ berales, por ejemplo.

La proximidad entre dos puntos-tipos de alojamiento expresará una similitud de los perfiles socio-profesionales (columnas de la tabla 3). Esto es lo que se constata por ejemplo para: *pueblo de vacaciones* y *camping caravana*.

En fin, la representación simultánea de los puntos-línea y de los puntos-línea y de los puntos-columna permite identificar las variables responsables de ciertas proximidades. Así, la proximidad entre los perfiles de albergue de los Patrones y de las Profesiones liberales se explica especialmente / por una importancia superior a la media en el tipo *hotel* e inferior a la media en el tipo *camping-ca-/ ravana*. La comparación de los perfiles-línea respecto al perfil medio (línea total de la tabla 2) nos confirma esta observación.

De la misma manera, la proximidad entre *pueblo de vacaciones* y *tienda-caravana* se explicará por / una proporción más importante que la media de *obreros*, lo que nos confirma la comparación a la columna total de las dos columnas correspondientes de la tabla 3.

Se notará que las proximidades entre perfiles se interpretan siempre respecto al perfil medio. Puesto que las distancias al origen de los ejes (centro de gravedad) se interpretan en términos de desviación al perfil medio, las proximidades serán más fáciles de interpretar en la periferia de la gráfica.

La lectura de las contribuciones absolutas va a permitir saber cuáles son las variables responsables de la construcción de los ejes. Para el primer eje, *Hotel, Propiedad y camping-caravana* explican entre las tres 88% de la varianza. El punto pueblo de vacaciones no participa más que muy poco en la construcción de este eje, contrariamente a lo que podía dejar sugerir su abscisa: su masa es en efecto 3 veces / más débil que *camping-caravana*.

Sobre el segundo eje, *Casa en propiedad y Padres* tienen abscisas cercanas, aún cuando la contribución de este último punto (41.7) excede largamente la del primero (6.1).

Para las categorías socio-profesionales, son sobre todo los *Cuadros superiores y Obreros* los que contribuyen al primer eje (resp. 35,6 y 37,3) ambos en oposición. La lectura de las contribuciones relativas (resp. 0,90 y 0,97) nos enseña que estas dos categorías caracterizan exclusivamente este primer eje.

Se ve de qué manera la consulta de las contribuciones permite matizar la 1^a lectura del gráfico, precisando el efecto de las diferentes masas (contribuciones absolutas), y haciendo aparecer los cosenos de los ángulos en el espacio entre variables y ejes (contribuciones relativas).

3.3.4 Análisis de Correspondencias Múltiples (CM)

En una encuesta la mayor parte de las preguntas están pre-codificadas bajo la forma *disyuntiva completa*, es decir que las diversas modalidades de respuesta a una pregunta se excluyen mutuamente y una sola modalidad es obligatoriamente escogida por el encuestado. (Ver parágrafo 3.2)

Notas:

El conjunto de las cuestiones es designado por Q. Una cuestión q consiste en un conjunto J_q de modalidades de respuesta $J = \sum_q J_q$

Se designará por $R(n,Q)$ la tabla de código reducida, el elemento r_{iq} contiene el número de la modalidad de la pregunta q escogida por el sujeto i.

Se designará por $Z(n,J)$ la tabla *disyuntiva completa* describiendo las respuestas de los n individuos por un código binario.

Se designará por B la tabla de contingencias de BURT asociada a la tabla de respuestas Z.

La tabla B está formada de Q^2 bloques. El q-ésimo bloque diagonal $Z_q' Z_q$ es una matriz diagonal de orden J_q^2 (puesto que dos modalidades de respuesta a una misma pregunta no pueden ser escogidas simultáneamente).

El bloque indicado por (q,q') , de orden $J_q \times J_{q'}$, no es otra cosa sino la tabla de contingencia que cruza las respuestas a las dos preguntas q y q'.

Designaremos por D, de orden $J \times J$, la matriz diagonal que tiene los mismos elementos diagonales que B (estos elementos diagonales son los efectivos correspondientes a cada una de las modalidades).

La matriz D puede ser igualmente considerada como formada de Q^2 bloques (solamente los bloques 0 diagonales / son matrices no nulas, para el q-ésimo bloque diagonal: $D_q = Z_q' Z_q$, matriz diagonal cuyos términos diagonales / son los efectivos correspondientes a las diversas modalidades de respuestas a la pregunta q).

Caso de dos preguntas:

La tabla de respuestas Z se escribe $Z = (Z_1 / Z_2)$. Es entonces equivalente desde el punto de vista de la descripción de las asociaciones entre modalidades, el efectuar:

- (1) El análisis de las correspondencias de la tabla Z de orden ($n \times J$).
- (2) El análisis de las correspondencias de la tabla B de orden ($J \times J$).
- (3) El análisis de las correspondencias de la tabla K = $Z_1^T Z_2$ de orden ($J_1 \times J_2$).

Los valores propios no son los mismos:

Tabla Analizada	Valor propio
$K = Z_1^T Z_2$	λ_K
$Z = (Z_1 ; Z_2)$	$\mu_Z = \frac{1 + \sqrt{\lambda_K}}{2}$
$B = Z^T Z$	$\mu_B = \mu_Z^2$

Se pueden hacer así dos comentarios:

Comentario 1:

En el análisis de la gran tabla disyuntiva Z, los puntos representando las diversas modalidades de respuesta a las dos preguntas son elementos de un mismo conjunto (el conjunto de las columnas de Z). Mientras que en el análisis de la tabla de contingencia $Z_1^T Z_2$, los elementos se dividen en puntos-líneas y puntos-columnas. El hecho de que las tipologías obtenidas en el espacio de los primeros factores sean idénticas salvo un coeficiente aproximado (debido al hecho de que los valores propios no son los mismos) nos prueba que la representación simultánea de los puntos-líneas y de los puntos-columnas en el análisis de las correspondencias de una tabla de contingencia no es únicamente un artificio gráfico.

Comentario 2:

Estos tres análisis, efectuados sobre la misma información bruta, dan resultados similares, pero con valores propios diferentes, luego tasas de inercia diferentes. Las relaciones entre estas tasas de inercia nos muestran que estas son siempre mucho más elevadas en el análisis de la tabla de contingencia $Z_1^T Z_2$ que en el de la tabla Z. De una manera general, el análisis de las

tablas bajo forma disyuntiva da siempre tasas de inercia débiles, que dan una idea demasiado pesimista de la parte de información extraída.

Se generaliza después al caso de 0 preguntas. Las propiedades del método son muy interesantes.

Las Q sub-nubes de puntos correspondientes a las J_q modalidades de una pregunta q tienen el mismo centro de gravedad, el de la nube global. Las coordenadas de las modalidades de respuesta a cada pregunta están así centradas sobre cada eje factorial.

La inercia total vale $(J/Q) - 1$, vale 1 en el caso particular donde todas las preguntas tienen dos modalidades de respuesta ($J = 2Q$)

La parte de inercia debida a una modalidad de respuesta es tanto más grande cuanto menor sea el efectivo de esta modalidad.

$$C_{ij} = \frac{1}{Q} (1 - \frac{d_{ij}}{n})$$

Hay que evitar las modalidades efectivo débil.

La parte de inercia debida a una cuestión es función creciente del número de modalidades de respuesta.

$$C_q = \frac{1}{Q} (J_q - 1)$$

Es preciso entonces equilibrar el nº de modalidades en el conjunto de las preguntas activas. En la práctica si una pregunta tiene un número de modalidades demasiado grande se puede efectuar un reagrupamiento en un número de modalidades razonable que conserve el sentido de la pregunta, éste es el reagrupamiento que se pondrá en variable activa, se puede por otra parte proyectar las modalidades de la pregunta de origen en variable ilustrativa.

El rango de Z es $J-Q+1$. En el análisis con respecto al centro de gravedad G se tiene $J-Q$ valores propios no nulos, el rango de la matriz a diagonalizar es $J-Q$.

Zona de confianza de una modalidad suplementaria. Los valores-test editados por los programas (para las modalidades activas o suplementarias) permiten comparar las significaciones respectivas de varias modalidades suplementarias como veremos en el ejemplo tratado en el párrafo siguiente.

4. PUESTA EN PRACTICA DE LOS METODOS

La opción escogida para el seminario es hacer un análisis completo de un ejemplo, utilizando el sistema SPAD. El paquete SPAD (système portable pour l'analyse des données) ha sido implantado y los listados han sido sacados en el Centro de Cálculo de la Universidad de Bilbao durante la misma semana del seminario. La descripción, la interpretación y el análisis de los resultados de la encuesta sobre las 111 Fincas de MARACAY (VENEZUELA) se han hecho utilizando las copias de los listados que los participantes podían anotar. Comentando estos listados con la ayuda de transparencias enseñábamos como servirnos de las diferentes etapas de SPAD con la ayuda de reseñas de utilización.

Ciertamente hemos expuesto otros ejemplos a título / ilustrativo pero no pueden reemplazar la experiencia práctica que cada investigador puede adquirir trabajando sus propios datos. Esta experiencia es posible ahora puesto que la herramienta se ha vuelto/ accesible.

La encuesta que analizamos aquí se apoya sobre 111 / fincas de Maracay, los datos han sido recogidos por - / Rainy Courbon y han sido el objeto de uno de los talleres de trabajo sobre el ordenador de las Jornadas de Análisis de Datos e Informática, encuentro Francia- América Latina (Julio de 1983 en Caracas, Venezuela). Otros investigadores trabajando para el ministerio de agricultura de Venezuela preparan encuestas similares en otras ciudades. El ejemplo no es de gran tamaño pero ilustra muy bien lo que nosotros llamamos un análisis descriptivo exploratorio: primeramente examinaremos las relaciones entre el conjunto de las variables que caracterizan la/

producción de las granjas, seguidamente nosotros emplearemos dos métodos de clasificación adaptados a este tipo de tabla para resumir la composición de estas granjas por / clases bien tipificadas.

Con SPAD los tratamientos que constituyen un análisis se descomponen en etapas: etapa de organización de los datos, etapa de los cálculos propiamente dichos, etapa de / las ediciones gráficas, etapas de clasificación, etc. Un análisis completo es pues una sucesión de etapas. Cada // una de ellas se determina por algunos parámetros, y la comunicación de los resultados entre etapas se efectúa automáticamente por medio de ficheros externos que pueden ser conservados para usos posteriores. Los libros de SPAD citados en la bibliografía (Ver el parágrafo 3.1) contienen el catálogo de las etapas, las listas de los análisis clásicos con los ensamblajes de las etapas que las componen, los principios para la ejecución de un análisis. / El capítulo II reúne las reseñas de utilización de cada / etapa. Se encontrará allí la lista y las definiciones de los parámetros de comando caso por caso.

Ejemplo comentado:

Nosotros encadenamos diez etapas en una misma ejecución, su ensamblaje es simple, cada etapa es simplemente llamada por una palabra clave: LISTP, DONNE, LILEX, MULTGRAPH, SEMIS, TAMIS, GRAFK. La palabra clave STOP marca / el fin de un análisis.

La etapa LISTP efectúa el listado del conjunto de las tarjetas de parámetros. Aquí, se obtiene la página de listado siguiente:

LISTAGE DES PARAMETRES DE COMMANDE

1	-----+-----1-----+-----2-----+-----3-----+-----4-----+-----5-----+-----6-----+-----7-----+-----8-----
2	LISTP
3	DONNE
4	23 111 1 1 17
5	(AA,23F1.0)
6	LILEX
7	111 FINCAS DE MARACAY(REMY) CON 23 VARIABLES . EL 15.07.83(CARACAS).....
8	24 78 23 111 0 1 0 0 1 1 -1
9	1 0
10	112121111212122222222
11	MULTC
12	6 0 0 2 1 30 0 2
13	GRAPH
14	1 1 58 0 1 1 58 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
15	SEMIS
16	2 10 5 15 6
17	TAMIS
18	1 30 10 1 1 0 3
19	GRAFK
20	1 1 58 1 1 58
21	1 2 2 3
22	TAMIS
23	0 30 10 1 1 0 5
24	GRAFK
25	1 1 58 1 1 58
26	1 2 2 3
27	STOP FIN DE LOS PARAMETROS DE FINCP2
	-----+-----1-----+-----2-----+-----3-----+-----4-----+-----5-----+-----6-----+-----7-----+-----8-----

En este ejemplo se utiliza después la etapa DONNE :

- el fichero inicial cuenta con 23 variables (NQEXA), 111 fincas (número de líneas de la tabla de datos, IEXA = 111), un identificador de finca A4(aquí NVIDI = 1, un A4).

- este fichero es leído con el formato fortran siguiente (A4, 23 F 1.0). Se utiliza un registro (LFMT = 1) para introducir este formato.

- el número lógico del fichero (NDON 2=17) es introducido como parámetro.

- la etapa DONNE lee el fichero inicial de los datos y crea un fichero archivo (NDONA = 2) utilizable en entrada por la etapa LILEX. Se supone aquí que el fichero inicial es un fichero limpio que basta para archivar; las /

operaciones de depuración, de recodificación, de fusión, etc.... se han efectuado previamente.

La etapa LILEX muestra la información útil sobre // los ficheros archivos, y permite escoger los elementos / activos e ilustrativos de análisis posteriores: aquí nosotros guardamos las 111 fincas como líneas activas, /// atribuimos el peso 1 a cada finca (NVPOI = 0), pero hacemos una selección entre las 23 variables, las 10 variables que caracterizan la producción de las granjas son / activas (tipo 1, variable nominal activa), las otras 13 son ilustrativas (tipo 2, variable nominal ilustrativa). Esta selección es muy importante (Ver el párrafo 2.3 / sobre las cualidades de una tabla). Es aquí donde se hace esta selección que introduce el diccionario de las variables, estas se reordenan como se ve en ésta y en la / siguiente página de listado:

```

ETAPE ** LILEX **

TITRE DE L-ANALYSE           111 FINCAS DE MARACAY(REMY) CON 23 VARIABLES . EL 15.07.83(CARACAS).....
-----  

CARTE DES PARAMETRES POUR LILEX
-----  

NOUTI= 24      JUTI = 78      NQEXA= 23      IEXA = 111
NVPOI= 0        NVIDI= 1       NVFIL= 0       MOFIL= 0      LEX = 1      MEDIT= 1      MDICO= -1  

-----  

MODE DE SELECTION (COLONNES ET LIGNES)
-----  

MODIU= 1      MODIG= 0
..... FICHIER EN ENTREE = 1 (NOICA)
..... FICHIER EN ENTREE = 2 (NOONA)
..... FICHIER EN ECRITURE = 8 (NOIC)
..... FICHIER EN ECRITURE = 9 (NOON)
..... UTILISATION DE MEMOIRES          VOUS AVEZ RESERVE 10000          VOUS AVEZ BESOIN DE 1376
-----  

MISE EN PLACE DES COLONNES OU VARIABLES
-----  

SELECTION DES ELEMENTS SELON LE MODE 1
-----  

BILAN DE LA SELECTION
-----  

TYPE 1      NOMBRE DE VARIABLES    10
TYPE 2      NOMBRE DE VARIABLES    13
-----  

VECTEUR INDICATEUR DES 23 ELEMENTS PAR GROUPE DE 10
1121211111 2121222222 222
-----
```

MISE EN PLACE DU FICHIER

DICTIONNAIRE DES VARIABLES ET DES MODALITES DE REPONSE

VARIABLES DU TYPE 1 10
 MODALITES CORRESPONDANTES 33

1 UNIDAD AGROECOLOGICA			(5 MODALITES)		
/ UAE1=ALUVIAL 1	/ UAE2=ALUVIAL 2	/ UAE3=ALUVIAL 3	/ UAE4=LAGUNERO		/
/ UAE5=ASOCIACION					
2 SUPERFICIE UTILIZADA		(3 MODALITES)			
/ SUP1=MENOR DE 5 HA	/ SUP2=DE 5.1 A 10 HA	/ SUP3=MAYOR DE 10.1 HA			
4 RUBRO		(4 MODALITES)			
/ RUB1=FRUTAL	/ RUB2=FRUTAL Y CULTIVOS	/ RUB3=CANA DE AZUCAR	/ RUB4=CULT.DE CICLO CORTO		/
6 RIEGO		(2 MODALITES)			
/ RIE1=SIN RIEGO	/ RIE2=CON RIEGO				
7 TECNOLOGIA		(4 MODALITES)			
/ TEC1=ALTA	/ TEC2=MEDIANA ALTA	/ TEC3=MEDIANA BAJA	/ TEC4=BAJA		/
8 MAQUINARIA		(3 MODALITES)			
/ MAQ1=NO USA	/ MAQ2=USA ALQUILADA	/ MAQ3=USA PROPIA			
9 TRABAJO		(3 MODALITES)			
/ TRA1=EXCLUSIV FAMILIAR	/ TRA2=PRINCIP FAMILIAR	/ TRA3=ASALARIADO			
10 NACIONALIDAD		(2 MODALITES)			
/ NAC1=VENEZOLANO	/ NAC2=CANARIO				
12 VIVE DE LA FINCA		(3 MODALITES)			
/ VIF1=EXCLUSIVAMENTE	/ VIF2=PRINCIPALMENTE	/ VIF3=NO VIVE DE LA FINCA			
14 CREDITO		(4 MODALITES)			
/ CRE1=NO RECIBE SIN PROB	/ CRE2=NO RECIBE CON PROB	/ CRE3=SI RECIBE SIN PROB	/ CRE4=SI RECIBE CON PROB		/

VARIABLES DU TYPE 2 13
 MODALITES CORRESPONDANTES 44

3 SUPERFICIE EN FRUTAL		(4 MODALITES)			
/ SUF1=SIN FRUTAL	/ SUF2=DE 0.1 A 2 HA	/ SUF3=DE 2.1 A 5 HA	/ SUF4=MAYOR DE 5 HA		/
5 CULTIVO ANTERIOR		(3 MODALITES)			
/ CAN1=LO MISMO	/ CAN2=FRUTAL	/ CAN3=CULT DE CICLO CORTO			
11 EDAO		(4 MODALITES)			
/ EDA1=20 A 40 ANOS	/ EDA2=41 A 50 ANOS	/ EDA3=51 A 60 ANOS	/ EDA4=MAYOR DE 61 ANOS		/
13 DESTINO DE LA PRODUCCION		(4 MODALITES)			
/ DES1=AGROINDUSTRIA	/ DES2=CONSUMO FAMILIAR	/ DES3=INTERMEDIARIO	/ DES4=AGRO/INTER		/
15 ASISTENCIA TECNICA		(2 MODALITES)			
/ ATE1=NO RECIBE	/ ATE2=SI RECIBE				
16 NIVEL DE VIDA		(3 MODALITES)			
/ NIV1=ALTO	/ NIV2=MEDIO	/ NIV3=BAJO			
17 COMPORTAMIENTO DEL INGRESO		(3 MODALITES)			
/ ING1=HA BAJADO	/ ING2=SE HA MANTENIDO	/ ING3=HA SUBIDO			
18 PROBLEMAS PLAGAS Y ENFERMEDADES		(4 MODALITES)			
/ PLE1=NO TIENE	/ PLE2=TRISTEZA/STENOMA	/ PLE3=OTROS	/ PLE4=AMBOS		/
19 PROBLEMAS DE MALEZAS		(2 MODALITES)			
/ PMA1=NO TIENE	/ PMA2=SI TIENE				
20 PROBLEMA DE AGUA		(3 MODALITES)			
/ PAG1=NO TIENE	/ PAG2=DRENAGE	/ PAG3=RIEGO			
21 PRINCIPAL PROBLEMA		(7 MODALITES)			
/ PRO1=NO TIENE	/ PRO2=AGUA	/ PRO3=CREDITO	/ PRO4=ROBO		/
/ PRO5=PLAGAS	/ PRO6=CRECIMIENTO DEL LAGO	/ PRO7=OTROS			
22 LABOR QUE REALIZA EL PRODUCTOR		(3 MODALITES)			
/ LAB1=TODO	/ LAB2=ALGUNAS COSAS	/ LAB3=SUPERVISA			
23 SE SIENTE SATISFECHO DE LA FINCA		(2 MODALITES)			
/ SAT1=NO	/ SAT2=SI				

I EFFECTIF TOTAL DES LIGNES 111 I

FIN DE L-ETAPE ** LILEX **

Nosotros hemos escogido un conjunto homogéneo de variables activas que caracterizan las condiciones de producción de las granjas: unidad agroecológica, superficie utilizada, rubro, riego, tecnología, maquinaria, trabajo, nacionalidad, vive de la finca, crédito, en total 33 modalidades de respuesta figuran sobre las páginas precedentes; éstas son las que contribuirán a la construcción de los ejes principales en el análisis de correspondencias múltiples (etapa MULTC). Está claro que la pregunta se siente satisfecho de la finca (pregunta 23 con respuestas NO o SI) es una pregunta de opinión, luego hemos decidido ponerla como ilustrativa. El nivel de vida (pregunta 16 con las tres modalidades de respuesta ALTO, MEDIO, BAJO) es más bien un resultado de la producción de la finca y no una condición de producción, se la pone entonces como ilustrativa.

La etapa MULTC realiza los cálculos y efectúa las / ediciones para un análisis de correspondencias múltiples y esto según los parámetros solicitados. La parte teórica de este método figura en el párrafo 3.3.4. Comentemos brevemente la elección que se ha hecho de los / parámetros: nosotros calculamos y salvaguardamos seis e-

jes factoriales (NFAC = 6), no se utiliza la posibilidad de transferir una modalidad a ilustrativa si el efecto es inferior o igual al parámetro NMIN (aquel NMIN = 0), no se solicita la edición que concierne a las fincas (LIST=3=0). Para dos factores sucesivos (NTEXT = 2) se solicita una / edición de los rangos de las modalidades (activas e ilustrativas) sobre una página (NPAGT=1) y también para las / 30 (NLIGT) modalidades con inercia más fuerte sobre el // factor. No se solicita edición de las correlaciones de / las variables continuas con los factores (NCOR=0) puesto/ que no se tienen variables continuas en este ejemplo. Pa-//ra la edición de la tabla de BURT, cruzando las preguntas activas, se ha solicitado la edición de los efectivos y / de los perfiles (NTAB = 2).

No se pueden exponer aquí las trasparencias o páginas de listado más importantes: la tabla de BURT y el histograma de los primeros valores propios. Hay dos factores ampliamente dominantes, nuestra atención se centrará sobre todo el plano (F_1 , F_2) de los dos primeros factores. Para las modalidades activas, las masa, las distancias al origen, las coordenadas y las contribuciones, nos permiten poner en evidencia las principales oposiciones.

TABLEAU DE BURT

PROFILS HORIZONTAUX (EN POUR 1000) DU TANIEAU DE BURT
LE TRI-A-PLAT DE CHAQUE QUESTION (EN POUR 1000) FIGURE SUR LA DIAGONALE CORRESPONDANTE

	UAE1	UAE2	UAE3	UAE4	UAE5	SUP1	SUP2	SUP3	RUB1	RUB2	RUB3	RUB4	RIE1	RIE2	TEC1	TEC2	TEC3	TEC4	MAQ1	MAQ2	MAQ3	TRA1	TRA2	TRA3
UAE1 I	81.	0.	0.	0.	0.	333.	667.	1000.	0.	0.	0.	0.	1000.	667.	222.	111.	0.	0.	0.	1000.	111.	222.	667.	
UAE2 I	0.	171.	0.	0.	0.	579.	211.	211.	211.	474.	105.	158.	842.	842.	0.	0.	158.	53.	579.	368.	105.	263.	632.	
UAE3 I	0.	0.	81.	0.	0.	667.	0.	333.	111.	778.	0.	111.	889.	889.	0.	111.	0.	0.	778.	222.	222.	111.	667.	
UAE4 I	0.	0.	0.	514.	0.	526.	316.	158.	263.	702.	0.	35.	702.	298.	316.	158.	298.	228.	263.	298.	439.	123.	561.	316.
UAE5 I	0.	0.	0.	0.	153.	235.	118.	647.	294.	706.	0.	0.	294.	706.	588.	176.	118.	118.	118.	765.	0.	706.	294.	
SUP1 I	0.	216.	118.	588.	78.	459.	0.	0.	137.	608.	216.	39.	627.	373.	392.	98.	216.	294.	314.	529.	157.	176.	569.	255.
SUP2 I	111.	148.	0.	667.	74.	0.	243.	0.	444.	444.	74.	37.	519.	481.	556.	111.	222.	111.	74.	259.	667.	111.	481.	407.
SUP3 I	182.	121.	91.	273.	333.	0.	0.	297.	455.	424.	91.	30.	91.	909.	697.	182.	121.	0.	0.	91.	909.	0.	303.	697.
RUB1 I	265.	118.	29.	441.	147.	206.	353.	441.	306.	0.	0.	0.	235.	765.	676.	176.	118.	29.	29.	88.	882.	147.	294.	559.
RUB2 I	0.	70.	18.	702.	211.	544.	211.	246.	0.	514.	0.	0.	684.	316.	298.	140.	298.	263.	281.	351.	368.	88.	614.	298.
RUB3 I	0.	563.	438.	0.	0.	688.	125.	188.	0.	0.	144.	0.	0.	1000.	1000.	0.	0.	0.	0.	813.	188.	0.	313.	688.
RUB4 I	0.	500.	0.	500.	0.	250.	250.	250.	0.	0.	0.	36.	500.	500.	500.	0.	0.	500.	250.	250.	500.	500.	500.	0.
RIE1 I	0.	61.	20.	816.	102.	653.	286.	61.	163.	796.	0.	41.	441.	0.	204.	102.	347.	347.	367.	367.	265.	143.	612.	245.
RIE2 I	145.	258.	129.	274.	194.	306.	210.	484.	419.	290.	258.	32.	0.	559.	774.	145.	65.	16.	0.	306.	694.	81.	355.	565.
TEC1 I	103.	276.	138.	310.	172.	345.	259.	397.	397.	293.	276.	34.	172.	828.	523.	0.	0.	0.	17.	345.	638.	69.	379.	552.
TEC2 I	143.	0.	0.	643.	214.	357.	214.	429.	429.	571.	0.	0.	357.	643.	0.	126.	0.	0.	214.	214.	571.	71.	500.	429.
TEC3 I	48.	0.	48.	810.	95.	524.	286.	190.	190.	810.	0.	0.	810.	190.	0.	0.	189.	0.	95.	476.	429.	95.	571.	333.
TEC4 I	0.	167.	0.	722.	111.	833.	167.	0.	56.	833.	0.	111.	944.	56.	0.	0.	162.	667.	222.	111.	278.	611.	111.	
MAQ1 I	0.	56.	0.	833.	111.	889.	111.	0.	56.	889.	0.	56.	1000.	0.	56.	167.	111.	667.	162.	0.	0.	278.	611.	111.
MAQ2 I	0.	297.	189.	459.	54.	730.	189.	81.	81.	541.	351.	27.	486.	514.	541.	81.	270.	108.	0.	333.	0.	108.	486.	405.
MAQ3 I	161.	125.	36.	446.	232.	143.	321.	536.	536.	375.	54.	36.	232.	768.	661.	143.	161.	36.	0.	0.	508.	54.	411.	536.
TRA1 I	83.	167.	167.	583.	0.	750.	250.	0.	417.	417.	0.	167.	583.	417.	333.	83.	167.	417.	417.	333.	250.	108.	0.	0.
TRA2 I	38.	98.	19.	615.	231.	558.	250.	192.	192.	673.	96.	38.	577.	423.	423.	135.	231.	212.	212.	346.	442.	0.	468.	0.
TRA3 I	128.	255.	128.	383.	106.	277.	234.	489.	404.	362.	234.	0.	255.	745.	681.	128.	149.	43.	43.	319.	638.	0.	0.	423.
NAC1 I	24.	190.	107.	583.	95.	583.	238.	179.	179.	583.	190.	48.	548.	452.	500.	71.	226.	202.	214.	429.	357.	143.	488.	369.
NAC2 I	259.	111.	0.	296.	333.	74.	259.	667.	704.	296.	0.	0.	111.	889.	593.	296.	74.	37.	0.	37.	963.	0.	407.	593.
VIF1 I	105.	88.	70.	509.	228.	351.	263.	386.	456.	456.	88.	0.	386.	614.	526.	175.	175.	123.	88.	333.	579.	88.	421.	491.
VIF2 I	73.	244.	98.	537.	49.	561.	268.	171.	171.	512.	244.	73.	537.	463.	512.	98.	195.	195.	244.	341.	415.	98.	585.	317.
VIF3 I	0.	308.	77.	462.	154.	615.	77.	308.	77.	769.	77.	77.	385.	615.	538.	0.	231.	231.	308.	462.	231.	308.	462.	
CRE1 I	192.	38.	0.	615.	154.	462.	231.	308.	316.	615.	0.	38.	462.	538.	346.	192.	231.	231.	231.	577.	77.	577.	346.	
CRE2 I	33.	150.	50.	583.	183.	483.	233.	283.	333.	583.	33.	50.	583.	417.	483.	100.	217.	200.	217.	300.	483.	150.	517.	333.
CRE3 I	143.	286.	71.	429.	71.	357.	357.	286.	357.	286.	357.	0.	143.	857.	643.	214.	143.	0.	0.	429.	571.	71.	143.	786.
CRE4 I	0.	455.	455.	0.	91.	455.	182.	364.	0.	182.	818.	0.	0.	1000.	1000.	0.	0.	0.	0.	636.	364.	0.	364.	636.

UAE1 UAE2 UAE3 UAE4 UAE5 SUP1 SUP2 SUP3 RUB1 RUB2 RUB3 RUB4 RIE1 RIE2 TEC1 TEC2 TEC3 TEC4 MAQ1 MAQ2 MAQ3 TRA1 TRA2 TRA3

	NAC1	NAC2	VIF1	VIF2	VIF3	CRE1	CRE2	CRE3	CRE4
UAE1 I	222.	778.	667.	333.	0.	556.	222.	222.	0.
UAE2 I	842.	158.	263.	526.	211.	53.	474.	211.	263.
UAE3 I	1000.	0.	444.	444.	111.	0.	333.	111.	556.
UAE4 I	860.	140.	509.	386.	105.	281.	614.	105.	0.
UAE5 I	471.	529.	765.	118.	118.	235.	647.	59.	59.
SUP1 I	961.	39.	392.	451.	157.	235.	569.	98.	98.
SUP2 I	741.	259.	556.	407.	37.	222.	519.	185.	74.
SUP3 I	455.	545.	667.	212.	121.	242.	515.	121.	121.
RUB1 I	441.	559.	765.	206.	29.	265.	588.	147.	0.
RUB2 I	860.	140.	456.	368.	175.	281.	614.	70.	35.
RUB3 I	1000.	0.	313.	625.	63.	0.	125.	313.	563.
RUB4 I	1000.	0.	0.	750.	250.	250.	750.	0.	0.
RIE1 I	939.	61.	449.	449.	102.	245.	714.	41.	0.
RIE2 I	613.	387.	565.	306.	129.	226.	403.	194.	177.
TEC1 I	724.	276.	517.	362.	121.	155.	500.	155.	190.
TEC2 I	429.	571.	714.	286.	0.	357.	429.	214.	0.
TEC3 I	905.	95.	476.	381.	143.	286.	619.	95.	0.
TEC4 I	944.	56.	389.	444.	167.	333.	667.	0.	0.
MAQ1 I	1000.	0.	278.	556.	167.	278.	722.	0.	0.
MAQ2 I	973.	27.	514.	378.	108.	162.	486.	162.	189.
MAQ3 I	536.	464.	589.	304.	107.	268.	518.	143.	71.
TRA1 I	1000.	0.	417.	333.	250.	167.	750.	83.	0.
TRA2 I	788.	212.	462.	462.	77.	288.	596.	38.	77.
TRA3 I	660.	340.	596.	277.	128.	191.	426.	234.	149.
NAC1 I	757.	0.	417.	440.	143.	202.	560.	119.	119.
NAC2 I	0.	243.	815.	148.	37.	333.	481.	148.	37.
VIF1 I	614.	386.	514.	0.	0.	246.	544.	158.	53.
VIF2 I	902.	98.	0.	369.	0.	195.	561.	73.	171.
VIF3 I	923.	77.	0.	0.	117.	308.	462.	154.	77.
CRE1 I	654.	346.	538.	308.	154.	234.	0.	0.	0.
CRE2 I	783.	217.	517.	383.	100.	0.	541.	0.	0.
CRE3 I	714.	286.	643.	214.	143.	0.	0.	126.	0.
CRE4 I	909.	91.	273.	636.	91.	0.	0.	0.	99.

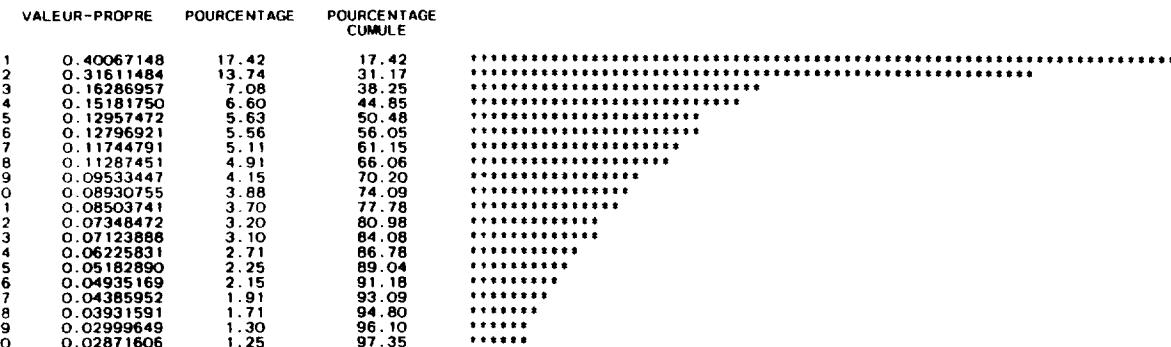
NAC1 NAC2 VIF1 VIF2 VIF3 CRE1 CRE2 CRE3 CRE4

SOMME DES POIDS = 111.000

EDITION DES VALEURS-PROPS

SOMME DES VALEURS-PROPS 2.29995728

HISTOGRAMME DES PREMIERES VALEURS-PROPS



EDITION SOMMAIRE DES VALEURS-PROPS DE 21 A 23

0.02370763 0.02216968 0.01502028

Sobre el primer eje: las modalidades con riego, nacionalidad = canario, superficie utilizada mayor de 10.1 Ha., Rubro = Frutal, Maquinaria = usa propia, Trabajo = asalariado, tecnología = alta tienen fuertes contribuciones absolutas y coordenadas negativas. Se oponen a las modalidades activas siguientes: Sin riego, No usa maquinaria, Tecnología = baja, superficie utilizada menor de 5 Ha., unidad agroecológica = lagúnero, nacionalidad = venezolano, y con una contribución más débil // trabajo principalmente familiar.

Sobre el segundo eje: Caña de azúcar, usa alquilada la maquinaria, tecnología alta, se oponen a Frutal, usa propia maquinaria y canario.

Las posiciones respectivas del conjunto de las modalidades en el plano (F_1 , F_2) de los dos primeros factores pueden ser obtenidas con la etapa GRAPH, esto es lo que hacemos aquí. Es, sin embargo preferible trabajar sobre dos gráficas separadas: la una para las modalidades activas (figura 1), la otra para las modalidades ilustrativas (figura 2).

Además es importante:

- primero, tener en cuenta las contribuciones como lo hemos hecho para describir las principales oposiciones entre las modalidades activas, es decir, entre los puntos que contribuyen más a la construcción de estos /

dos primeros ejes dominantes (contribuciones absolutas / fuertes) y las cuales están en consecuencia bien representadas en este plano (contribuciones relativas fuertes). Al contrario una modalidad como cultivo de ciclo corto / (RUB4 de la pregunta RUBRO) está muy mal representada como proyección en este plano (contribuciones relativas muy débiles, respectivamente 0,02 y 0,00 sobre F_1 y F_2). Este punto está situado en el tercer eje (contribución relativa 0,33), y contribuye ampliamente a su construcción / (contribución absoluta de 19,6% a F_3). Además observamos que sólo 4 Fincas sobre 111 han escogido esta modalidad.

- Segundo, tener en cuenta los valores-tests que permiten comparar las significaciones respectivas de las modalidades ilustrativas. Observamos que aquí son numerosas / las fuertes (superiores a dos en valor absoluto) sobre los dos ejes fuertemente dominantes. A título de ejemplo las modalidades de las dos preguntas Nivel de vida, comportamiento del ingreso son significativas en este plano factorial. El nivel de vida está muy correlacionado a la pregunta activa superficie utilizada. Las dos modalidades SI y NO de la pregunta se siente satisfecho de la finca son significativas sobre el segundo y tercer eje, es divertido observar que los satisfechos no se proyectan del lado de los que tienen un nivel de vida más elevado.

En suma, esta red de interrelaciones nos ayuda a volver a los datos, a la tabla de BURT, a rehacer los códigos y a extraer las tablas más interesantes para mejorar explicitar los resultados del análisis.

POUR LES MOYALITES ACTIVES
 EDITION DES COORDONNEES ET DES CONTRIBUTIONS

NOMS	MASSES	DISTO	COORDONNEES						CONTRIBUTIONS ABSOLUES						CONTRIBUTIONS RELATIVES							
			F1	F2	F3	F4	F5	F6	F1	F2	F3	F4	F5	F6	F1	F2	F3	F4	F5	F6		
***** QUESTION 1 UNIDAD AGROECOLOGICA																						
UAE1 0.008 11.33 -1.46 0.97 0.83 0.80 0.41 1.11 + 4.3 2.4 3.4 3.4 1.1 7.8 * 0.19 0.08 0.06 0.06 0.01 0.11 *																						
UAE2 0.017 4.84 -0.31 -1.14 0.77 -0.02 0.04 -0.61 + 0.4 7.1 6.3 0.0 0.0 5.0 * 0.02 0.27 0.12 0.00 0.00 0.08 *																						
UAE3 0.008 11.33 -0.48 -1.95 -0.44 -0.21 -0.30 1.04 + 0.5 9.8 1.0 0.2 0.6 6.9 * 0.02 0.34 0.02 0.00 0.01 0.10 *																						
UAE4 0.051 0.95 0.54 0.34 -0.16 0.37 -0.01 -0.01 + 3.8 1.9 0.8 4.6 0.0 0.0 0.31 0.12 0.03 0.14 0.00 0.00 0.00 *																						
UAE5 0.015 5.53 -0.45 0.67 -0.54 -1.52 -0.08 -0.43 + 0.8 2.2 2.8 23.4 0.1 2.2 * 0.04 0.08 0.05 0.42 0.00 0.03 *																						
CONTRIBUCTION CUMULEE- 9.7 23.3 14.3 31.6 1.7 21.9 *																						
***** QUESTION 2 SUPERFICIE UTILIZADA																						
SUP1 0.046 1.18 0.66 -0.39 0.02 -0.11 -0.18 0.32 + 5.0 2.2 0.0 0.4 1.1 3.7 * 0.37 0.13 0.00 0.01 0.03 0.09 *																						
SUP2 0.024 3.11 -0.08 0.28 0.04 0.99 0.59 -0.47 + 0.0 0.6 0.0 15.8 6.5 4.3 * 0.00 0.03 0.00 0.32 0.11 0.07 *																						
SUP3 0.030 2.36 -0.95 0.37 -0.07 -0.64 -0.20 -0.11 + 6.8 1.3 0.1 8.1 1.0 0.3 * 0.39 0.06 0.00 0.18 0.02 0.00 *																						
CONTRIBUCTION CUMULEE- 11.8 4.0 0.1 24.3 8.5 8.2 *																						
***** QUESTION 4 RUBRO																						
RUB1 0.031 2.26 -0.80 0.63 0.40 0.46 0.10 0.12 + 4.9 3.8 3.1 4.3 0.2 0.3 * 0.28 0.17 0.07 0.09 0.00 0.01 *																						
RUB2 0.051 0.95 0.58 0.24 -0.39 -0.25 -0.18 -0.09 + 4.4 0.9 4.7 2.1 1.3 0.3 * 0.36 0.06 0.16 0.07 0.03 0.01 *																						
RUB3 0.014 5.94 -0.57 -2.11 -0.23 -0.01 0.25 0.43 + 1.2 20.3 0.5 0.0 0.7 2.1 * 0.05 0.75 0.01 0.00 0.01 0.03 *																						
RUB4 0.004 26.75 0.75 -0.31 2.97 -0.35 0.68 -1.42 + 0.5 0.1 19.6 0.3 1.3 5.7 * 0.02 0.00 0.33 0.00 0.02 0.08 *																						
CONTRIBUCTION CUMULEE- 10.9 25.1 27.8 6.7 3.5 8.4 *																						
***** QUESTION 6 RIEGO																						
RIE1 0.044 1.27 0.91 0.26 -0.16 0.16 0.05 -0.03 + 9.2 0.9 0.7 0.8 0.1 0.0 * 0.66 0.05 0.02 0.02 0.00 0.00 *																						
RIE2 0.056 0.79 -0.72 0.20 0.12 -0.13 -0.04 0.02 + 7.3 0.7 0.5 0.6 0.1 0.0 * 0.66 0.05 0.02 0.02 0.00 0.00 *																						
CONTRIBUCTION CUMULEE- 16.5 1.6 1.2 1.4 0.2 0.0 *																						
***** QUESTION 7 TECNOLOGIA																						
TEC1 0.052 0.91 -0.56 -0.38 0.14 -0.08 0.11 -0.27 + 4.1 2.4 0.6 0.2 0.5 2.9 * 0.34 0.16 0.02 0.01 0.01 0.08 *																						
TEC2 0.013 6.93 -0.31 0.84 -0.31 -0.02 0.01 1.28 + 0.3 2.8 0.8 0.0 0.0 16.2 * 0.01 0.10 0.01 0.00 0.00 0.24 *																						
TEC3 0.019 4.29 0.60 0.30 -0.96 0.69 -0.33 -0.55 + 1.7 0.5 10.6 5.9 1.6 4.4 * 0.08 0.02 0.21 0.11 0.03 0.07 *																						
TEC4 0.016 5.17 1.35 0.21 0.90 -0.54 0.02 0.50 + 7.4 0.2 8.1 3.1 0.0 3.2 * 0.35 0.01 0.16 0.06 0.00 0.05 *																						
CONTRIBUCTION CUMULEE- 13.5 6.0 20.1 9.2 2.1 26.7 *																						
***** QUESTION 8 MAQUINARIA																						
MAQ1 0.016 5.17 1.45 0.32 0.65 -0.51 0.04 0.76 + 8.6 0.5 4.3 2.8 0.0 7.4 * 0.41 0.02 0.08 0.05 0.00 0.11 *																						
MAQ2 0.033 2.00 0.26 -0.83 -0.47 0.29 -0.22 0.04 + 0.6 7.2 4.5 1.9 1.2 0.0 0.3 * 0.03 0.34 0.11 0.04 0.02 0.00 *																						
MAQ3 0.050 0.98 -0.64 0.44 0.10 -0.03 0.13 -0.27 + 5.1 3.1 0.3 0.0 0.6 2.9 * 0.41 0.20 0.01 0.00 0.02 0.08 *																						
CONTRIBUCTION CUMULEE- 14.2 10.9 9.1 4.7 1.9 10.3 *																						
***** QUESTION 9 TRABAJO																						
TRA1 0.011 8.25 0.78 -0.07 1.65 0.42 0.80 0.32 + 1.6 0.0 18.0 1.3 5.4 0.8 * 0.07 0.00 0.33 0.02 0.08 0.01 *																						
TRA2 0.047 1.13 0.39 0.18 -0.35 -0.26 0.52 -0.09 + 1.8 0.5 3.4 2.6 9.9 0.3 * 0.13 0.03 0.11 0.07 0.24 0.01 *																						
TRA3 0.042 1.36 -0.63 -0.18 -0.04 0.21 -0.37 -0.02 + 4.2 0.5 0.0 1.2 4.6 0.0 * 0.29 0.02 0.00 0.03 0.10 0.00 *																						
CONTRIBUCTION CUMULEE- 7.6 1.0 21.5 5.1 19.8 1.2 *																						
***** QUESTION 10 NACIONALIDAD																						
NAC1 0.076 0.32 0.35 -0.28 -0.02 0.08 -0.04 -0.08 + 2.3 1.9 0.0 0.4 0.1 0.3 * 0.38 0.24 0.00 0.02 0.01 0.02 *																						
NAC2 0.024 3.11 -1.09 0.87 0.05 -0.26 0.13 0.24 + 7.2 5.8 0.0 1.1 0.3 1.1 * 0.38 0.24 0.00 0.02 0.01 0.02 *																						
CONTRIBUCTION CUMULEE																						

COORDONNEES ET VALEURS-TESTS POUR TOUTES LES MODALITES

POIDS	*	COORDONNEES	*	VALEURS-TESTS CORRIGEES	*
** VARIABLE 1	1	UNIDAD AGROECOLOGICA	*		
UAE1	9.	* 1.46 0.97 0.83 0.80 0.41 1.11 *	*	-4.53 3.01 2.59 2.48 1.28 3.46 *	
UAE2	19.	* -0.31 -1.14 0.77 -0.02 0.04 -0.61 *	*	-1.48 -5.46 3.69 -0.09 0.17 -2.93 *	
UAE3	9.	* -0.48 -1.95 -0.44 -0.21 -0.30 1.04 *	*	-1.49 -6.09 -1.38 -0.67 -0.92 3.24 *	
UAE4	57.	* 0.54 0.34 -0.16 0.37 -0.01 -0.01 *	*	5.86 3.65 -1.70 3.97 -0.06 -0.07 *	
UAE5	17.	* -0.45 0.67 -0.54 -1.52 -0.08 -0.43 *	*	-2.01 2.98 -2.42 -6.79 -0.37 -1.92 *	
** VARIABLE 2		SUPERFICIE UTILISÉE	*		
SUP1	51.	* 0.66 -0.39 0.02 -0.11 -0.18 0.32 *	*	6.38 -3.74 0.22 -1.05 -1.73 3.09 *	
SUP2	27.	* -0.08 0.28 0.04 0.99 0.59 -0.47 *	*	-0.47 1.67 0.22 5.90 3.49 -2.82 *	
SUP3	33.	* -0.95 0.37 -0.07 -0.64 -0.20 -0.11 *	*	-6.51 2.51 -0.45 -4.40 -1.39 -0.73 *	
** VARIABLE 4		RUBRO	*		
RUB1	34.	* -0.80 0.63 0.40 0.46 0.10 0.12 *	*	-5.57 4.37 2.81 3.22 0.70 0.81 *	
RUB2	57.	* 0.58 0.24 -0.39 -0.25 -0.18 -0.09 *	*	6.30 2.58 -4.16 -2.68 -1.93 -0.98 *	
RUB3	16.	* -0.57 -2.11 -0.23 -0.01 0.25 0.43 *	*	-2.46 -9.07 -0.97 -0.04 1.10 1.86 *	
RUB4	4.	* 0.75 -0.31 2.97 -0.35 0.68 -1.42 *	*	1.53 -0.63 6.03 -0.71 1.38 -2.87 *	
** VARIABLE 6		RIEGO	*		
RIE1	49.	* 0.91 0.26 -0.16 0.16 0.05 -0.03 *	*	8.52 2.39 -1.45 1.54 0.49 -0.25 *	
RIE2	62.	* -0.72 -0.20 0.12 -0.13 -0.04 0.02 *	*	-8.52 -2.39 1.45 -1.54 -0.49 0.25 *	
** VARIABLE 7		TECNOLOGIA	*		
TEC1	58.	* -0.56 -0.38 0.14 -0.08 0.11 -0.27 *	*	-6.15 -4.15 1.56 -0.84 1.23 -2.93 *	
TEC2	14.	* -0.31 0.84 -0.31 -0.02 0.01 1.28 *	*	-1.23 3.36 -1.25 -0.09 0.04 5.10 *	
TEC3	21.	* 0.60 0.30 -0.96 0.69 -0.33 -0.55 *	*	3.02 1.51 -4.84 3.48 -1.69 -2.76 *	
TEC4	18.	* 1.35 0.21 0.90 -0.54 0.02 0.50 *	*	6.23 0.99 4.15 -2.48 0.09 2.31 *	
** VARIABLE 8		MAQUINARIA	*		
MAQ1	18.	* 1.45 0.32 0.65 -0.51 0.04 0.76 *	*	6.71 1.49 3.02 -2.36 0.20 3.51 *	
MAQ2	37.	* 0.26 -0.83 -0.47 0.29 -0.22 0.04 *	*	1.92 -6.14 -3.47 2.16 -1.60 0.30 *	
MAQ3	56.	* -0.64 0.44 0.10 -0.03 0.13 -0.27 *	*	-6.75 4.69 1.05 -0.30 1.36 -2.87 *	
** VARIABLE 9		TRABAJO	*		
TRA1	12.	* 0.78 -0.07 1.65 0.42 -0.80 0.32 *	*	2.85 -0.25 6.01 1.53 -2.94 1.15 *	
TRA2	52.	* 0.39 0.18 -0.35 -0.29 0.52 -0.09 *	*	3.83 1.79 -3.41 -2.83 5.15 -0.93 *	
TRA3	47.	* -0.63 -0.18 -0.04 0.21 -0.37 0.02 *	*	-5.66 -1.65 -0.34 1.90 -3.36 0.22 *	
** VARIABLE 10		NACIONALIDAD	*		
NAC1	84.	* 0.35 -0.28 -0.02 0.08 -0.04 -0.08 *	*	6.46 -5.18 -0.29 1.57 -0.79 -1.41 *	
NAC2	27.	* -1.09 0.87 0.05 -0.26 0.13 0.24 *	*	-6.46 5.18 0.29 -1.57 0.79 1.41 *	
** VARIABLE 12		VIVE DE LA FINCA	*		
VIF1	57.	* -0.31 0.34 -0.25 0.08 -0.22 0.16 *	*	-3.38 3.61 -2.68 0.82 -2.39 1.75 *	
VIF2	41.	* 0.33 -0.38 0.16 0.09 0.84 -0.00 *	*	2.69 -3.08 1.28 0.69 6.75 -0.04 *	
VIF3	13.	* 0.32 -0.26 0.59 -0.61 -1.68 -0.70 *	*	1.22 -0.99 2.24 -2.31 -6.43 -2.68 *	
** VARIABLE 14		CREDITO	*		
CRE1	26.	* 0.06 0.59 -0.01 -0.08 0.06 0.66 *	*	0.35 3.40 -0.03 -0.46 0.35 3.85 *	
CRE2	60.	* 0.26 0.20 -0.08 -0.10 0.06 -0.40 *	*	3.01 2.27 0.86 -1.11 0.68 -4.56 *	
CRE3	14.	* -0.71 -0.35 -0.03 1.06 -0.98 0.19 *	*	-2.82 -1.41 -0.10 4.24 -3.89 0.78 *	
CRE4	11.	* -0.68 -2.02 -0.37 -0.63 0.77 0.37 *	*	-2.38 -7.03 -1.27 -2.19 2.69 1.29 *	
** VARIABLE 3		SUPERFICIE EN FRUTAL	*		
SUF1	20.	* -0.31 -1.75 0.41 -0.08 0.34 0.06 *	*	-1.50 -8.60 2.04 -0.37 1.67 0.31 *	
SUF2	25.	* 1.12 -0.07 -0.23 -0.02 -0.22 0.28 *	*	6.36 0.38 -1.32 -0.12 -1.24 1.56 *	
SUF3	24.	* 0.46 0.23 -0.09 -0.06 -0.09 0.01 *	*	2.52 1.26 -0.49 -0.33 -0.49 0.06 *	
SUF4	42.	* -0.78 0.66 -0.01 0.08 0.02 -0.20 *	*	-6.42 5.42 -0.06 0.68 0.15 -1.64 *	
** VARIABLE 5		CULTIVO ANTERIOR	*		
CAN1	51.	* 0.14 0.21 0.06 -0.21 -0.07 0.15 *	*	1.39 2.05 0.60 -2.05 -0.71 -1.50 *	
CAN2	11.	* -0.42 0.33 0.01 0.04 -0.12 -0.29 *	*	-1.46 1.16 0.03 0.13 -0.41 -1.02 *	
CAN3	49.	* -0.06 -0.30 -0.07 0.21 0.10 0.23 *	*	-0.52 -2.75 -0.62 1.98 0.96 2.12 *	
** VARIABLE 11		EUAD	*		
EDA1	42.	* -0.24 0.44 0.19 0.07 -0.07 0.11 *	*	-1.93 3.59 1.54 0.56 -0.61 0.86 *	
EDA2	30.	* 0.00 -0.27 -0.08 -0.01 -0.07 0.04 *	*	-0.01 -1.72 -0.54 -0.09 -0.46 -0.26 *	
EDA3	14.	* -0.30 -0.43 -0.43 -0.36 -0.09 0.03 *	*	-1.19 -1.71 -1.72 -1.45 -0.37 0.12 *	
EDA4	25.	* 0.57 -0.17 0.03 0.11 0.26 -0.14 *	*	3.20 -0.98 0.15 0.60 1.49 -0.82 *	
** VARIABLE 13		DESTINO DE LA PRODUCTION	*		
DES1	20.	* 0.62 -1.75 -0.18 -0.29 0.07 0.12 *	*	-3.07 -8.62 -0.88 -1.42 0.33 0.61 *	
DES2	26.	* 1.10 0.18 0.07 0.04 -0.09 0.24 *	*	6.38 1.05 0.39 0.25 -0.52 1.38 *	
DES3	39.	* -0.01 0.37 0.01 -0.05 0.15 -0.42 *	*	-0.10 2.85 0.05 -0.35 1.15 -3.23 *	
DES4	26.	* -0.60 0.61 0.06 0.25 -0.19 0.29 *	*	-3.49 3.55 0.36 1.43 -1.08 1.71 *	
** VARIABLE 15		ASISTENCIA TECNICA	*		
ATE1	76.	* 0.36 0.26 -0.01 0.04 -0.02 -0.05 *	*	5.55 4.04 -0.11 0.60 -0.28 -0.79 *	
ATE2	34.	* -0.76 -0.61 -0.01 -0.09 0.01 0.08 *	*	-5.26 -4.25 -0.10 -0.65 0.09 0.55 *	
** VARIABLE 16		NIVEL DE VIDA	*		
NIV1	17.	* -0.99 0.48 0.25 -0.29 -0.01 0.22 *	*	-4.40 2.12 1.11 -1.31 -0.03 0.98 *	
NIV2	45.	* -0.19 0.33 0.07 0.05 0.12 -0.09 *	*	-1.65 2.86 0.63 0.48 1.05 -0.76 *	
NIV3	50.	* 0.48 -0.44 -0.13 0.05 -0.09 0.03 *	*	4.51 -4.19 -1.23 0.52 -0.84 0.27 *	
** VARIABLE 17		COMPORTAMIENTO DEL INGRESO	*		
ING1	71.	* 0.22 0.20 -0.05 0.10 0.16 0.04 *	*	3.13 2.78 -0.72 1.40 2.20 0.51 *	
ING2	16.	* -0.34 0.14 0.24 -0.48 -0.59 0.20 *	*	-1.46 0.61 1.03 -2.05 -2.53 -0.88 *	
ING3	24.	* -0.44 -0.68 -0.01 0.02 -0.07 0.03 *	*	-2.41 -3.76 -0.04 0.11 -0.40 0.16 *	
** VARIABLE 18		PROBLEMAS PLAGAS Y ENFERMEDADES	*		
PLE1	14.	* -0.44 -1.26 0.06 -0.11 -0.04 -0.13 *	*	-1.74 -5.02 0.22 -0.44 -0.17 -0.50 *	
PLE2	34.	* 0.26 0.62 0.18 0.18 -0.03 0.35 *	*	1.84 4.32 1.24 1.25 -0.21 2.41 *	
PLE3	24.	* 0.27 -0.77 0.43 -0.15 -0.00 -0.02 *	*	1.48 -4.25 2.36 -0.81 -0.00 -0.12 *	
PLE4	39.	* 0.22 0.39 -0.44 -0.03 0.04 -0.24 *	*	1.71 2.98 -3.38 -0.20 0.32 -1.88 *	
** VARIABLE 19		PROBLEMAS DE MALEZAS	*		
FMA1	66.	* -0.02 -0.12 -0.10 0.06 -0.14 0.13 *	*	-0.32 -1.55 -1.33 0.78 -1.82 1.59 *	
FMA2	45.	* 0.04 0.18 0.15 -0.09 0.21 -0.18 *	*	0.32 1.55 1.33 -0.78 1.82 -1.59 *	
** VARIABLE 20		PROBLEMA DE AGUA	*		
PAG1	52.	* -0.06 0.36 -0.06 0.10 0.07 0.21 *	*	-0.58 3.58 -0.59 0.94 0.70 2.11 *	
PAG2	28.	* -0.35 -0.27 0.07 -0.38 0.16 0.03 *	*	-2.13 -1.65 0.44 -2.30 1.00 0.21 *	
PAG3	31.	* 0.41 -0.37 0.03 0.18 -0.27 -0.39 *	*	2.71 -2.39 0.23 1.18 -1.75 -2.55 *	
** VARIABLE 21		PRINCIPAL PROBLEMA	*		
PRO1	6.	* -0.59 -0.11 0.74 0.26 -0.51 0.21 *	*	-1.48 -0.29 1.86 0.66 -1.29 0.52 *	
PRO2	29.	* 0.31 -0.14 -0.08 0.20 -0.29 -0.38 *	*	1.94 -0.90 -0.47 1.27 -1.84 -2.38 *	
PRO3	22.	* -0.21 -0.60 0.24 -0.05 0.31 -0.08 *	*	-1.12 -3.15 1.23 -0.26 1.64 -0.41 *	
PRO4	14.	* 0.31 0.24 -0.49 0.06 0.31 0.19 *	*	1.23 0.96 -1.94 0.24 1.23 0.74 *	
PRO5	11.	* -0.80 0.80 -0.25 -0.09 0.06 0.35 *	*	-2.78 2.78 -0.86 -0.30 0.20 1.23 *	
PRO6	20.	* 0.33 0.50 0.04 -0.29 0.04 0.21 *	*	1.63 2.43 0.18 -1.41 0.22 1.01 *	
PRO7	9.	* -0.33 -0.43 0.15 -0.06 -0.13 0.10 *	*	-1.02 -1.35 0.46 -0.19 -0.39 0.31 *	
** VARIABLE 22		LABOR QUE REALIZA EL PRODUCTOR	*		
LAB1	56.	* 0.63 0.28 0.02 0.03 0.07 -0.02 *	*	6.65 2.96 0.16 0.29 0.71 -0.22 *	
LAB2	48.	* -0.64 -0.33 -0.03 0.00 0.10 0.09 *	*	-5.86 -2.99 -0.30 0.03 0.93 0.83 *	
LAB3	7.	* -0.64 0.00 0.10 -0.25 -1.24 -0.45 *	*	-1.74 0.01 0.28 -0.67 -3.36 -1.24 *	
** VARIABLE 23		SE SIENTE SATISFECHO DE LA FINCA	*		
SAT1	51.	* -0.08 0.31 0.32 -0.06 -0.11 0.02 *	*	-0.73 2.99 3.11 -0.60 -1.03 0.15 *	
SAT2	60.	* 0.06 -0.26 -0.27 0.05 0.09 -0.01 *	*	0.73 -2.99 -3.11 0.60 1.03 -0.15 *	

0.872	NAC2	-----	TEC2	-----	I
0.834	I	PRO5			I
0.797	I				I
0.759	I				I
0.722	UAE1				I
0.685	I	SUF4	UAE5		I
0.647	I	RUB1	DES4	PLE2	I
0.610	I				I
0.573	I				I
0.535	I				I
0.498	NIV1				I
0.460	I	MAQ3	EDA1		I
0.423	I				I
0.386	I	SUP3	CAN2	VIF1NIV2	I
0.348	I			SUP2	I
0.311	I			DES3	I
0.274	I				I
0.236	I				I
0.199	I				I
0.161	I				I
0.124	I				I
0.087	I				I
0.049	I				I
0.012	I				I
-0.026	I				I
-0.063	I				I
-0.100	I	PRO1	PMA1		I
-0.138	I				I
-0.175	I				I
-0.212	I				I
-0.250	I				I
-0.287	I				I
-0.325	I				I
-0.362	I	CRE3	TEC1	PAG2	I
-0.399	I			ED4SAT2	I
-0.437	I			CAN3	I
-0.474	I				I
-0.511	I				I
-0.549	I				I
-0.586	I				I
-0.624	I				I
-0.661	I				I
-0.698	I				I
-0.736	I				I
-0.773	I				I
-0.810	I				I
-0.848	I				I
-0.885	I				I
-0.923	I				I
-0.960	I				I
-0.997	I				I
-1.035	I				I
-1.072	I				I
-1.110	I				I
-1.147	I				I
-1.184	I				I
-1.222	I				I
-1.259	I				I
-1.086					
		-0.599	-0.111	0.376	0.864
					1.351

POINTS MULTIPLES (100 AU MAXIMUM)

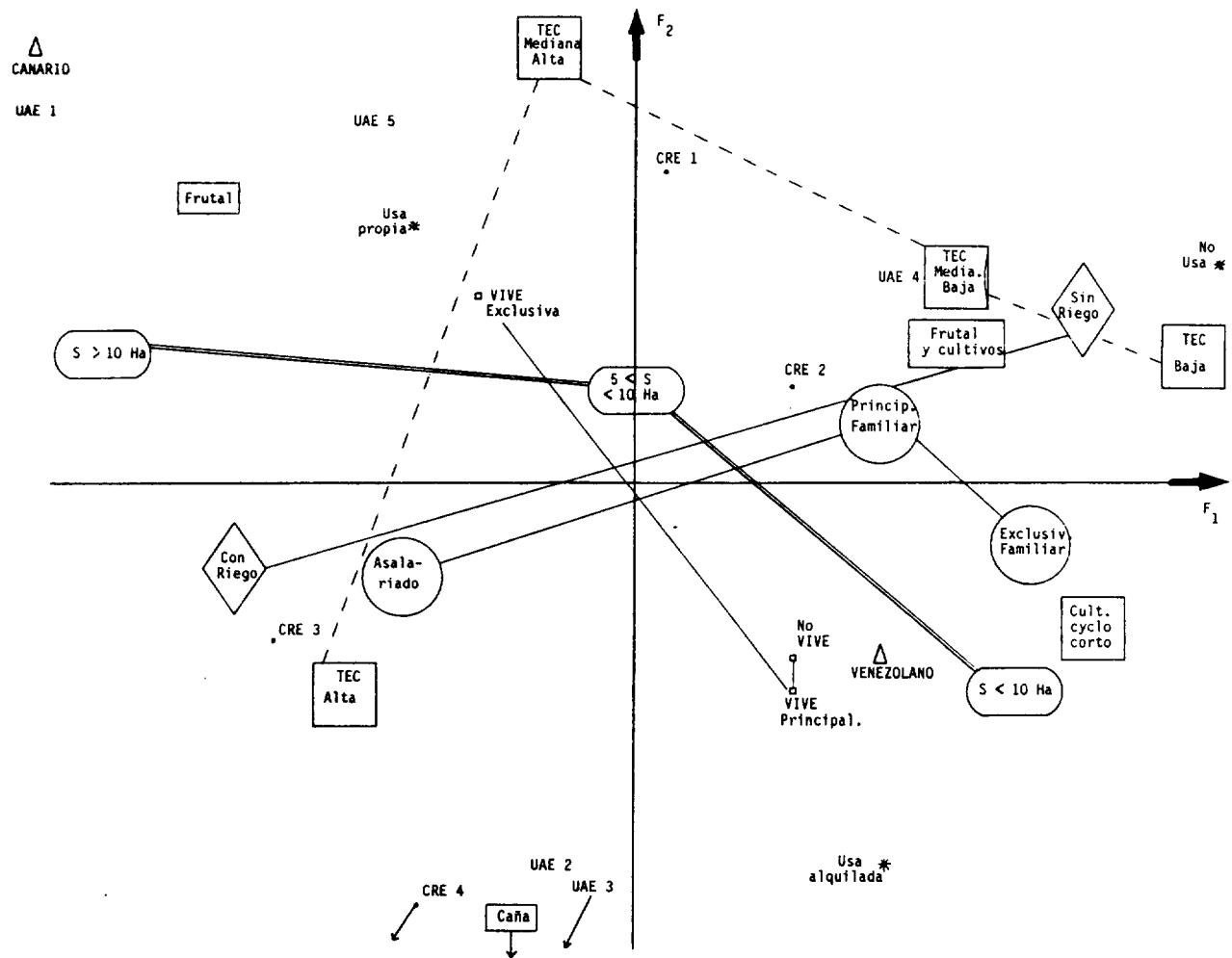


Figura 1: VARIABLES ACTIVAS

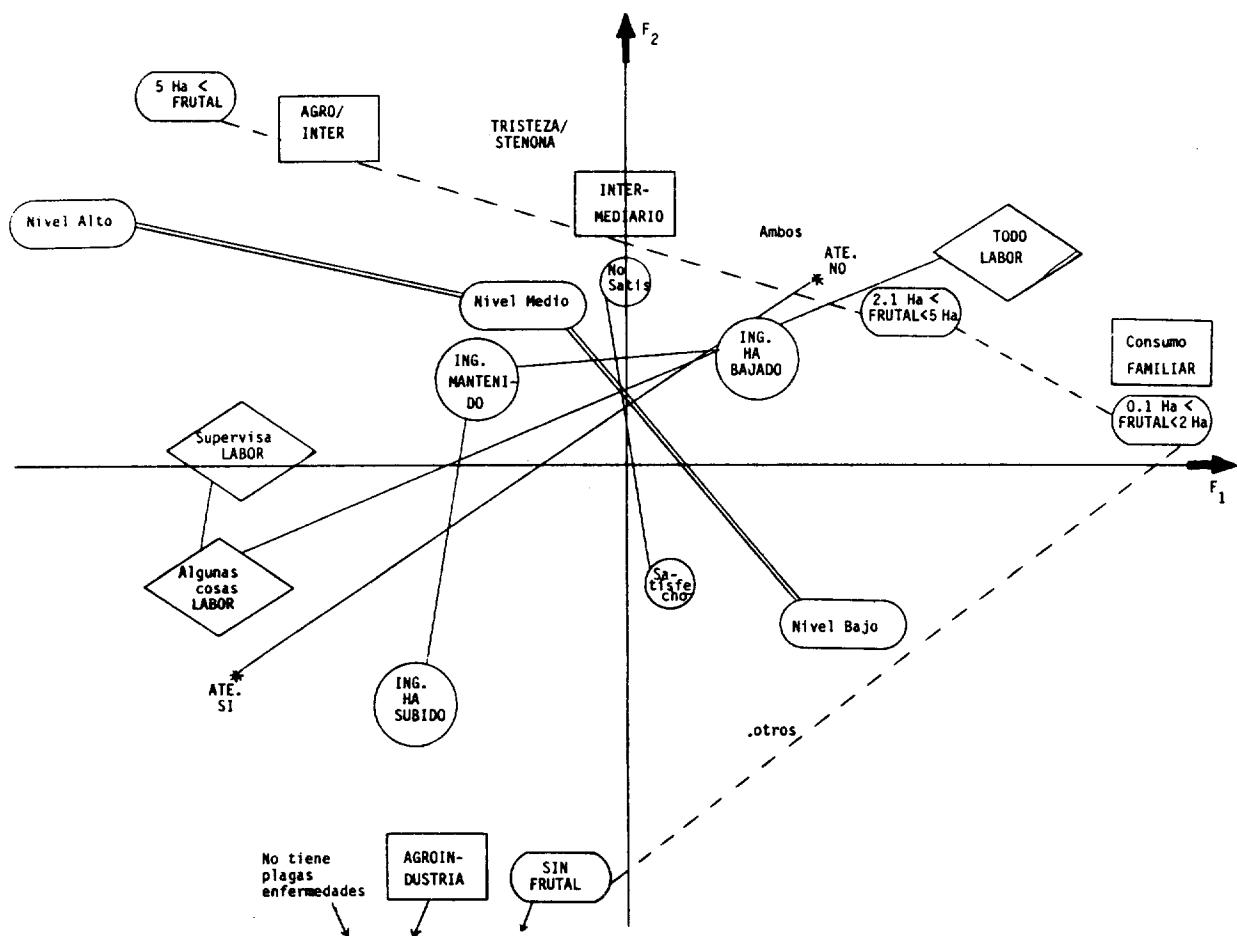


Figura 2: VARIABLES ILUSTRATIVAS

Puesta en práctica de dos técnicas de clasificación con SPAD

Se pueden utilizar estas técnicas para verificar si existen reagrupamientos que en el análisis precedente / no se habrían tenido en cuenta. Se puede también querer definir las clases de fincas. Aquí directamente utilizaremos dos métodos de clasificación jerárquica: la clasificación mixta (etapa SEMIS) y los vecinos recíprocos/ (etapa RECIP). Hacemos primero dos observaciones válidas para los dos métodos.

Observación 1: los dos algoritmos trabajan sobre los resultados de un análisis factorial, aquí las coordenadas de las fincas sobre los seis primeros factores.

Observación 2: el criterio de distancia escogida en SEMIS y en RECIP es el de la varianza:

$$d(a,b) = \frac{m_a m_b}{m_a + m_b} \cdot [(g_a, g_b)]^2$$

donde $d(a,b)$ es la distancia entre dos partes a y b, m_a y m_b el peso de estas partes, g_a y g_b sus centros de gravedad y \cdot la distancia Euclidiana. Existen otros / métodos que utilizan otros criterios de agregación.

La etapa SEMIS:

a) Idea directriz:

Una primera etapa consiste en obtener a un bajo costo una partición de N objetos en un número K de clases homogéneas, ese número K es muy superior al número de clases reales en la población observada. En / una segunda etapa se efectúa una clasificación jerárquica ascendente, donde los elementos terminales del árbol son precisamente las clases de la partición. La etapa / final consiste en efectuar una inspección visual del árbol para cortarlo al nivel que sea más apropiado. Uno obtiene así una partición definitiva de los N individuos en k_0 clases ($k_0 < k < N$) que pasan a describirse.

b) El método y vocabulario utilizados:

Parte Preliminar: Se trata de buscar rápidamente un gran número de pequeños grupos muy homogéneos. El procedimiento trabajando sobre la totalidad de los individuos deberá ser económico en el uso de memoria. Se utilizará aquí la partición definida por las agrupaciones/

estables resultado del cruce de un cierto número de particiones de base. Cada partición de base es calculada / por el algoritmo de centros móviles realizado por lectura directa de los datos, lo que asegura una débil ocupación de memoria. En general los cálculos serán efectuados sobre las coordenadas de los individuos sobre los primeros ejes principales de un análisis, utilizando la distancia euclíadiana usual. Esta distancia es entonces / equivalente a la distancia utilizada en el sistema de ejes inicial, cualquiera que sea esta distancia (distancia euclíadiana clásica entre valores, reducidos o no ; distancia del χ^2).

Agregación jerárquica de las agrupaciones estables: algunas de las agrupaciones estables pueden estar muy / próximas unas de otras, corresponden a una clase que se ha copiado artificialmente en la etapa precedente. Por otra parte el procedimiento va a crear en general algunas clases de tamaño pequeño, algunas veces reducidas / a un solo elemento. La tarea de la fase de agregación / jerárquica es justamente construir las clases que se han roto en muchos pedazos y agregar a sus polos de origen los elementos que aparezcan dispersos.

El árbol será construido con el criterio de agregación de Ward que presenta la ventaja de tener en cuenta la importancia de los elementos a clasificar (bajo la forma de pesos o ponderación en el cálculo de la pérdida de inercia por agregación). Si el efectivo a clasificar es débil, se puede utilizar el algoritmo de base menos rápido, el algoritmo de búsqueda en cadena de los vecinos recíprocos.

El corte del dendograma definirá la partición final de los individuos. La elección del nivel de este corte y del número de clases de la partición se hace a la vista del árbol: es necesario cortar encima de las agrupaciones bajas (que reúnen los elementos más próximos), y por debajo de las agrupaciones altas que no hacen más / que reunir en un solo conjunto las diversas clases de la población observada.

El ejemplo de las "fincas":

Los cinco parámetros de la etapa SEMIS han sido fijados de la manera siguiente:

NBASE = 2, el número de particiones de base efectuadas. NITER = 10, el número de iteraciones para determinar cada partición de base.

NCLAS = 5, el número de clases en cada partición de base.

NKLA = 15, el número de grupos estables pedidos, el cruce compuesto de (NCLAS)^{NBASE} clases de las cuales algunas pueden estar vacías, los (NKLA -1) primeros grupos son las clases de más fuerte efectivo y la última clase reagrupa los elementos restantes. Aquí el cruce podría suministrar $5^2 = 25$ clases pero hay sola -

mente 9 que no están vacías, no se tienen más que 111 / fincas. La salida listada determina los 5 individuos / fuentes extraídos al azar para cada una de las 2 particiones de base, proporciona una descripción de los nudos de la jerarquía, una descripción de las clases y un dendograma. Es viendo este árbol cuando se puede decidir el número de clases de la partición final, aquí nosotros elegimos tres clases, cuatro no se justificaría, pero se habría podido escoger cinco.

ALGORITHME DES CENTRES-MOBILES ET CONSTRUCTION DES CLASSES STABLES

PARTITION DES 111 INDIVIDUS CARACTERISES PAR 6 COORDONNEES CARTESIENNES

LA PARTITION CONTIENDRA 15 CLASSES
LES 14 PREMIERES CONTIENNENT LES INDIVIDUS LES PLUS STABLES DANS LES 2 PARTITIONS DE BASE
CHAQUE PARTITION DE BASE EST FORMEE EN 10 ITERATIONS AUTOUR DE 5 INDIVIDUS-SOURCE TIRES AU HASARD

APPEL D'UNE PARTITION, CONSTRUISTE SUR LES INDIVIDUS-SOURCE
POIDS DES CLASSES APRES 10 ITERATIONS 44 26 70 100 78
18. 19. 28. 25. 21.

APPEL D'UNE PARTITION, CONSTRUISTE SUR LES INDIVIDUS-SOURCE
POIDS DES CLASSES APRES 10 ITERATIONS 105 90 11 73 84
20. 30. 20. 23. 18.

EFFECTIFS DECROISSANT NON-NULS DANS LES 9 CLASSES DE LA PARTITION CROISEE
SUIVIS DES POURCENTAGES CUMULES

27	20	19	18	18	3	3	2	1
24.3	42.3	59.5	75.7	91.9	94.6	97.3	99.1	100.0

DESCRIPTION DES NOEUDS DE LA HIERARCHIE INDICE = (INDICE REEL)/(INDICE MAX) INDICE MAX = 94.43124

NO	AINE	BENJ	NBRE	POIDS	INDICE
----	------	------	------	-------	--------

10	8	9	2	3.00	0.005365 **
11	2	7	2	23.00	0.012820 ***
12	1	6	2	30.00	0.028708 *****
13	10	11	4	26.00	0.035096 *****
14	3	13	5	45.00	0.101935 *****
15	4	12	3	48.00	0.136791 *****
16	5	14	6	63.00	0.295638 *****
17	15	16	9	111.00	0.383646 *****

NOEUD	INDICE	AINE	BENJ	EFFECTIF	DESCRIPTION DES CLASSES DE LA HIERARCHIE INDICE = (INDICE REEL)/(INDICE MAX) INDICE MAX = 94.43124							
10	0.005365	8	9	2	8	9						
11	0.012820	2	7	2	2	7						
12	0.028708	1	6	2	1	6						
13	0.035096	10	11	4	8	9	2	7				
14	0.101935	3	13	5	3	8	9	2	7			
15	0.136791	4	12	3	4	1	6					
16	0.295638	5	14	6	5	3	8	9	2	7		
17	0.383646	15	16	9	4	1	6	5	3	8	9	2

POIDS	INDICE	DENOROGRAMME (ECHELLE 0.00537 0.38365) INDICE = (IND.REEL)/(IND.MAX) IND. MAX = 94.43124						
3.000	0.013	7					
20.000	0.035	2					
1.000	0.005	9					
2.000	0.102	8					
19.000	0.296	3					
18.000	0.384	5					
3.000	0.029	6					
27.000	0.137	1					
18.000	-----	4					

FIN DE L-ETAPE ** SEMIS **

Descripción de las tres clases: etapa TAMIS

Esta etapa de SPAD permite ilustrar e interpretar // una clasificación de elementos , independientemente que/ las clases provengan de un algoritmo de clasificación // como ésta que acabamos de efectuar, o que las clases // sean las modalidades de una variable nominal del fiche- ro de los datos. Se pueden obtener sucesivamente las // descripciones de varias particiones (se pueden solicitar hasta cinco diferentes) aquí nosotros no solicitamos más que la descripción de tres clases (KPAR = 3). Para los / otros parámetros escogemos: KEDIT = 1 para editar todas/ las tablas que cruzan la clasificación con todas las va- riaciones nominales, esta salida es larga nosotros no ha-/ cemos figurar aquí más que las dos primeras páginas que/ corresponden a las tres primeras variables nominales // (unidad agroecológica, superficie utilizada y Rubro).

da y Rubro.)

JPROX = 30, para caracterizar las 30 modalidades más tí- picas de cada clase.

MINOX = 10, no se retienen mas que las modalidades cuyo efectivo es superior a 10.

KVAX = 0, no hay edición para la caracterización de // clases por variables continuas.

NVMRO = 1, para la edición de los nos. de clase a los cuales pertenecen los individuos.

NVTRI = 0, la etapa TAMIS describe una clasificación he-cha anteriormente (por SEMIS) y almacenada en un fichero de trabajo (NGRI).

DISTRIBUCTION DES 111 INDIVIDUS EN 3 CLASSES
EFFECTIFS ET PROFILS PAR QUESTION

UNIDAD AGROECOLOGICA

EFFECTIFS	* 1*	* 2*	* 3*
ALUVIAL1 *UAE1*	9*	0	9
ALUVIAL2 *UAE2*	19*	2	7
ALUVIAL3 *UAE3*	9*	1	0
LAGUNERO *UAE4*	57*	40	17
ASOCIACION *UAE5*	17*	5	12
MARGES -----*	111*	48	45

CHI2 = 96.327 , DEG.LIB.= 8 , NOMBRE DES EFF.THEORIQUES INF 5 = 8 , PROB (CHI2 SUP 96.327) ENVIRON 0.000

PROFIL (TOTAL DES COLONNES=1000) * 1* * 2* * 3*

ALUVIAL1 *UAE1*	81*	0	200	0
ALUVIAL2 *UAE2*	171*	42	156	556
ALUVIAL3 *UAE3*	81*	21	0	444
LAGUNERO *UAE4*	514*	833	378	0
ASOCIACION *UAE5*	153*	104	267	0
MARGES -----*	1000*	1000	1000	1000

PROFIL (TOTAL DES LIGNES=1000) * 1* * 2* * 3*

ALUVIAL1 *UAE1*	1000*	0	1000	0
ALUVIAL2 *UAE2*	1000*	105	368	526
ALUVIAL3 *UAE3*	1000*	111	0	889
LAGUNERO *UAE4*	1000*	702	298	0
ASOCIACION *UAE5*	1000*	294	706	0
MARGES -----*	1000*	432	405	162

SUPERFICIE UTILIZADA

EFFECTIFS	* 1*	* 2*	* 3*
MENOR DE 5 HA *SUP1*	51*	34	4
DE 5.1 A 10 HA *SUP2*	27*	12	13
MAYOR DE 10.1 HA *SUP3*	33*	2	28
MARGES -----*	111*	48	45

CHI2 = 51.876 , DEG.LIB.= 4 , NOMBRE DES EFF.THEORIQUES INF 5 = 1 , PROB (CHI2 SUP 51.876) ENVIRON 0.000

PROFIL (TOTAL DES COLONNES=1000) * 1* * 2* * 3*

MENOR DE 5 HA *SUP1*	459*	708	89	722
DE 5.1 A 10 HA *SUP2*	243*	250	289	111
MAYOR DE 10.1 HA *SUP3*	297*	42	622	167
MARGES -----*	1000*	1000	1000	1000

PROFIL (TOTAL DES LIGNES=1000) * 1* * 2* * 3*

MENOR DE 5 HA *SUP1*	1000*	667	78	255
DE 5.1 A 10 HA *SUP2*	1000*	444	481	74
MAYOR DE 10.1 HA *SUP3*	1000*	61	848	91
MARGES -----*	1000*	432	405	162

RUBRO

EFFECTIFS	* 1*	* 2*	* 3*	
FRUTAL *RUB1*	34*	3	29	2
FRUTAL Y CULTIVOS *RUB2*	57*	43	14	0
CANA DE AZUCAR *RUB3*	16*	0	0	16
CULT.DE CICLO CORTO *RUB4*	4*	2	2	0
MARGES -----*	111*	48	45	18

CHI2 = 138.293 , DEG.LIB.= 6 , NOMBRE DES EFF.THEORIQUES INF 5 = 4 , PROB (CHI2 SUP 138.293) ENVIRON 0.000

PROFIL (TOTAL DES COLONNES=1000) * 1* * 2* * 3*

FRUTAL *RUB1*	306*	63	644	111
FRUTAL Y CULTIVOS *RUB2*	514*	898	311	0
CANA DE AZUCAR *RUB3*	144*	0	0	889
CULT.DE CICLO CORTO *RUB4*	36*	42	44	0
MARGES -----*	1000*	1000	1000	1000

PROFIL (TOTAL DES LIGNES=1000) * 1* * 2* * 3*

FRUTAL *RUB1*	1000*	88	853	59
FRUTAL Y CULTIVOS *RUB2*	1000*	754	248	0
CANA DE AZUCAR *RUB3*	1000*	0	0	1000
CULT.DE CICLO CORTO *RUB4*	1000*	500	500	0
MARGES -----*	1000*	432	405	162

RIEGO

EFFECTIFS	* 1*	* 2*	* 3*
-----------	------	------	------

Comentamos brevemente la descripción de las tres clases para las modalidades más típicas, sobre las dos páginas que siguen seis valores son editados:

1) **Criterio:** es aproximadamente una variable de LAPLACE-GAUSS centrada reducida que mide la distancia entre el / porcentaje de la modalidad en la clase y su porcentaje / en la población total, teniendo en cuenta los efectivos de la modalidad y de la clase. Este valor sirve para ordenar las modalidades por orden de poder descriptivo de la clase.

2) **Proba:** es la probabilidad de que el valor del coeficiente anterior se rebase en la hipótesis de una modalidad constituida por tirada aleatoria exhaustiva (ley hipergeométrica, ver la nota sobre la caracterización estadística de una clase y los valores-tests, de A. MORI - NEAU, en el Bulletin Technique du centre de Statistique/ et informatique appliquées, CESIA, volumen 2, 1984, números 1-2). De manera práctica, no se debe de interpretar la presencia de una modalidad cuyo criterio es inferior a 2 o la PROBA es superior a 0,025.

3) **Poids:** se trata del número ponderado de individuos de la población total que han formulado la respuesta, aquí esto es el número de fincas.

4) **Porcentaje global:** porcentaje asociado a "POIDS".

5) **Porcentaje MOD/CLA:** porcentaje de los individuos de / la clase que ha escogido esta modalidad de respuesta.

Ejemplo: En la clase 3 hay 18 fincas, 16 han respondido caña de azúcar a la pregunta Rubro: MOD/CLA = 88,9 (Ver páginas de listado)

6) **Porcentaje CLA/MOD:** porcentaje en la clase de los individuos que han escogido esta modalidad de respuesta.

Ejemplo: En la clase 3 se encuentran las 16 fincas de caña de azúcar, es decir todas: CLA/MOD = 100,0.

En nuestro ejemplo se ve que nuestras tres clases / son muy típicas .

Etapa GRAFK: ella suministra dos tipos de gráficas:

- una proyección de los centros de las clases sobre los primeros planos factoriales, escogidas por el investigador. Nosotros no hacemos figurar esta gráfica aquí.

- una proyección de la totalidad de los individuos, afectados del número de su clase de pertenencia, ver la pág. de listado adjunto: nuestras tres clases son netamente disjuntas en el espacio de los dos primeros factores; con este juego de datos se obtienen tres clases // bien tipificadas pero recordamos que los dos primeros / factores son ampliamente dominantes.

DESCRIPTION DES CLASSES

CLASSE	1	(*) 1*	48	43.2	IDENT CRITERE PROBA POIDS POURCENTAGES GLOBAL MOD/CLA CLA/MOD	
RIESGO						
RUBRO						
NACIONALIDAD						
DESTINO DE LA PRODUCCION						
ASISTENCIA TECNICA						
MAQUINARIA						
SUPERFICIE EN FRUTAL						
LABOR QUE REALIZA EL PRODUCTOR						
UNIDAD AGROECOLOGICA						
TECNOLOGIA						
SUPERFICIE UTILIZADA						
TECNOLOGIA						
NIVEL DE VIDA						
TRABAJO						
COMPORTAMIENTO DEL INGRESO						
PROBLEMAS PLAGAS Y ENFERMEDADES						
EDAD						
CREDITO						
SUPERFICIE EN FRUTAL						
PROBLEMA DE AGUA						
PRINCIPAL PROBLEMA						
CULTIVO ANTERIOR						
MAQUINARIA						
PRINCIPAL PROBLEMA						
VIVE DE LA FINCA						
PRINCIPAL PROBLEMA						
TRABAJO						
SE SIENTE SATISFECHO DE LA FINCA						
SIN RIESGO	(RIE1)	8.97 0.0	49	44.1	91.7	89.8
-FRUTAL Y CULTIVOS	(RUB2)	7.35 0.0	57	51.4	89.6	75.4
-VENEZOLANO	(NAC1)	7.06 0.0	84	75.7	100.0	57.1
-CONSUMO FAMILIAR	(DES2)	6.80 0.0	26	23.4	52.1	96.2
-NO RECIBE	(ATE1)	6.79 0.0	76	68.5	97.9	61.8
-NO USA	(MAQ1)	6.65 0.0	18	16.2	37.5	100.0
-DE 0.1 A 2 HA	(SUF2)	6.63 0.0	25	22.5	50.0	96.0
-TODO	(LAB1)	6.30 0.0	56	50.5	83.3	71.4
-LAGUNERO	(UAE4)	6.15 0.0	57	51.4	83.3	70.2
-BAJA	(TEG4)	5.38 0.000	18	16.2	35.4	94.4
-MENOR DE 5 HA	(SUP1)	4.79 0.000	51	45.9	70.8	66.7
-MEDIANA BAJA	(TEC3)	4.68 0.000	21	18.9	37.5	85.7
-BAJO	(NIV3)	4.19 0.000	50	45.0	66.7	64.0
-PRINCIP FAMILIAR	(TR42)	3.86 0.000	52	46.8	66.7	61.5
-HA BAJADO	(ING1)	3.59 0.000	71	64.0	81.3	54.9
-AMBOS	(PLE4)	3.06 0.001	39	35.1	50.0	61.5
-MAYOR DE 61 AÑOS	(EDA4)	3.05 0.001	25	22.5	35.4	66.0
-NO RECIBE CON PROB	(CRE2)	2.93 0.002	60	54.1	68.8	55.0
-DE 2.1 A 5 HA	(SUF3)	2.83 0.002	24	21.6	33.3	66.7
-RIESGO	(PAG3)	2.59 0.005	31	27.9	39.6	51.3
-AGUA	(PRO2)	2.59 0.005	29	26.1	37.5	52.1
-LO MISMO	(CAN1)	2.49 0.006	51	45.9	58.3	54.9
-USA ALQUILADA	(MAQ2)	2.23 0.013	37	33.3	43.8	56.8
-ROBO	(PRO4)	1.97 0.024	14	12.6	18.8	64.3
-PRINCIPALMENTE	(VIF2)	1.90 0.029	41	36.9	45.8	53.7
-CRECIMIENTO DEL LAGO	(PRO6)	1.42 0.078	20	18.0	22.9	55.0
-EXCLUSIV FAMILIAR	(TRA1)	1.41 0.079	12	10.8	14.6	58.3
-SI	(SAT2)	1.37 0.085	60	54.1	60.4	48.3

DESCRIPTION DES CLASSES

IDENT CRITERE PROBA POIDS POURCENTAGES
GLOBAL MOD/CLA CLA/MOD

CLASSE	2	(* 2 *)	45	40.5
MAQUINARIA				
NACIONALIDAD				
SUPERFICIE EN FRUTAL				
RUBRO				
SUPERFICIE UTILIZADA				
RIEGO				
NIVEL DE VIDA				
EDAD				
DESTINO DE LA PRODUCCION				
TECNOLOGIA				
TRABAJO				
VIVE DE LA FINCA				
PRINCIPAL PROBLEMA				
NIVEL DE VIDA				
PROBLEMAS PLAGAS Y ENFERMEDADES				
SE SIENTE SATISFECHO DE LA FINCA				
UNIDAD AGROECOLOGICA				
CULTIVO ANTERIOR				
PROBLEMA DE AGUA				
DESTINO DE LA PRODUCCION				
COMPORTAMIENTO DEL INGRESO				
ASISTENCIA TECNICA				
LABOR QUE REALIZA EL PRODUCTOR				
TECNOLOGIA				
CREDITO				

DESCRIPTION DES CLASSES

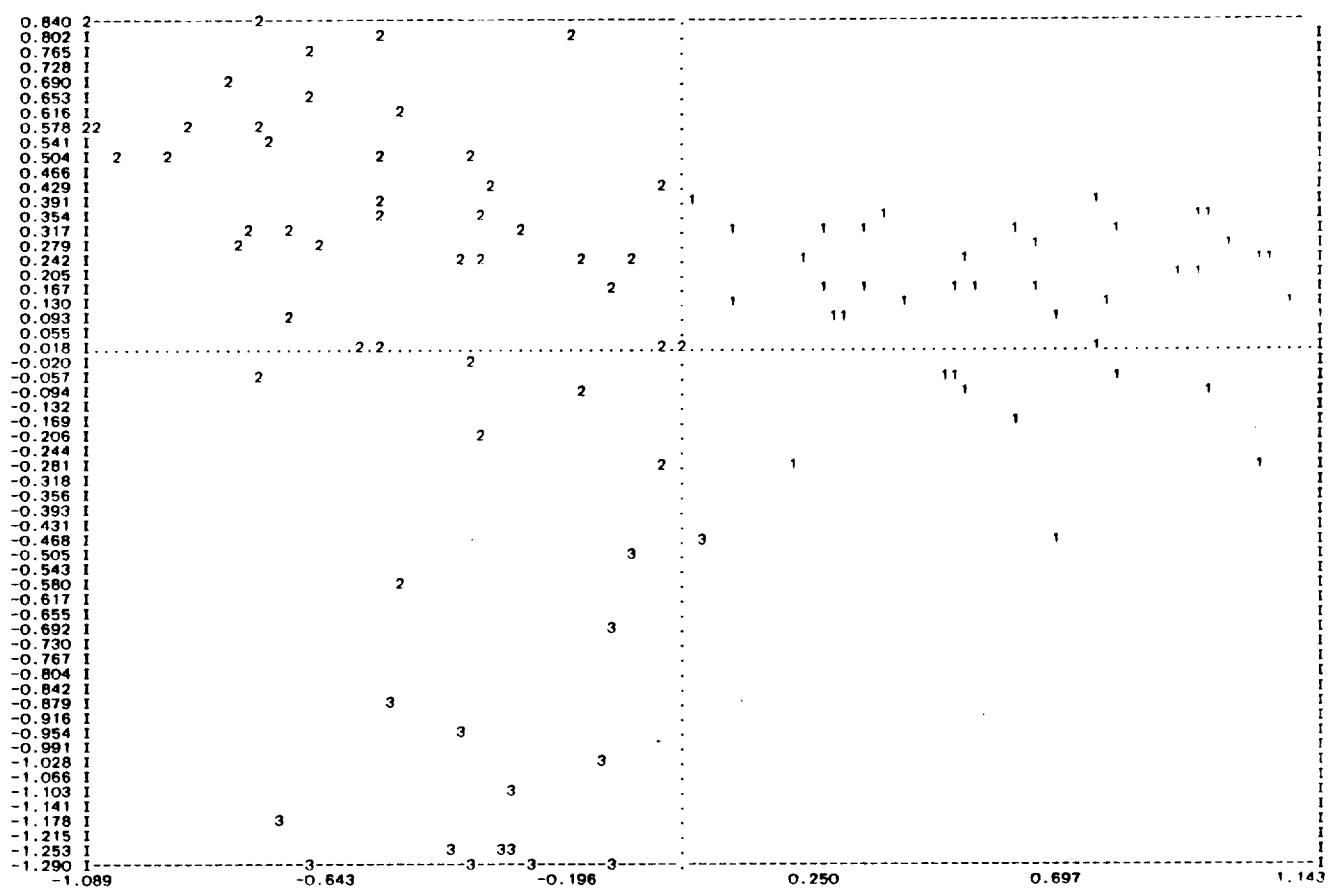
IDENT CRITERE PROBA POIDS POURCENTAGES
GLOBAL MOD/CLA CLA/MOD

CLASSE	3	(* 3 *)	18	16.2
RUBRO				
DESTINO DE LA PRODUCCION				
SUPERFICIE EN FRUTAL				
TECNOLOGIA				
RIEGO				
LABOR QUE REALIZA EL PRODUCTOR				
ASISTENCIA TECNICA				
CREDITO				
MAQUINARIA				
NACIONALIDAD				
UNIDAD AGROECOLOGICA				
COMPORTAMIENTO DEL INGRESO				
CULTIVO ANTERIOR				
PROBLEMAS PLAGAS Y ENFERMEDADES				
PROBLEMAS PLAGAS Y ENFERMEDADES				
PRINCIPAL PROBLEMA				
NIVEL DE VIDA				
PROBLEMA DE AGUA				
SUPERFICIE UTILIZADA				
SE SIENTE SATISFECHO DE LA FINCA				
CREDITO				
EDAD				
PROBLEMAS DE MALEZAS				
EDAD				
VIVE DE LA FINCA				
TRABAJO				

CLASSIFICATION DES INDIVIDUS

A01 / 1 A02 / 2 A03 / 2 A04 / 1 A05 / 1 A06 / 3 A07 / 1 A08 / 1 A09 / 1 A10 / 2 A11 / 2 A12 / 2 A13 / 1 A14 / 2 A15 / 1 A16 / 2 A17 / 2 A18 / 2 A19 / 3 A20 / 2 A21 / 1 A22 / 1 A23 / 1 A24 / 2 A25 / 1 A26 / 2 A27 / 1 A28 / 2 A30 / 2 A31 / 3 A32 / 3 A33 / 3 A34 / 2 A35 / 2 A36 / 2 A37 / 3 A38 / 2 A39 / 2 A40 / 2 A41 / 3 A42 / 1 A44 / 1 A45 / 1 A46 / 2 A47 / 1 A48 / 2 A49 / 2 A50 / 1 A51 / 3 A52 / 2 A53 / 2 A54 / 2 A55 / 1 A56 / 2 A57 / 2 A58 / 1 A59 / 1 A60 / 3 A61 / 2 A62 / 3 A63 / 2 A64 / 3 A65 / 1 A66 / 2 A67 / 1 A68 / 1 A69 / 2 A70 / 3 A71 / 1 A72 / 1 A73 / 2 A74 / 2 A75 / 1 A76 / 2 A77 / 3 A78 / 3 A79 / 3 A80 / 2 A81 / 1 A82 / 3 A83 / 2 A84 / 1 A85 / 2 A86 / 2 A87 / 3 A88 / 1 A89 / 1 A90 / 1 A91 / 2 A92 / 2 A93 / 1 A94 / 3 A95 / 1 A96 / 1 A97 / 1 A98 / 2 A99 / 2 B01 / 1 B02 / 1 B03 / 1 B04 / 1 B05 / 2 B06 / 1 B07 / 1 B08 / 1 B09 / 1 B10 / 1 B11 / 1 B12 / 1 B13 / 1 B14 / 2

FIN DE L-ETAPE ** TAMIS **



NOMBRE DE POINTS DOUBLES = 13

La etapa RECIP: efectúa la clasificación jerárquica según el método de los vecinos recíprocos con búsqueda en cadena, según los principios enunciados por J. P. // BENZECRI (1982) y MAC QUITTY (1966) y el algoritmo de J. JUAN (1982). Esto es un algoritmo rápido cuyo tiempo de cálculo es siempre en n^2 (si n es el nº de individuos) y no en n^3 como en el procedimiento clásico de LANCE y WILLIAMS. Aquí se efectúa esta clasificación sobre las 111 fincas a partir de sus coordenadas / sobre los seis primeros ejes factoriales (NFAC = 6). Se efectúa luego un corte del dendograma sacado de la clasificación al nivel especificado por el parámetro NKLA aquí NKLA = 28. Nosotros hacemos figurar aquí la descripción de la parte del árbol tras este corte en 28 clases, pero no la descripción de este corte. Aquí también se puede escoger cortar en tres clases y solicitar la descripción de estas tres clases con otra etapa / TAMIS y otra etapa GRAFK. Nosotros no hacemos figurar /

aquí mas que la gráfica obtenida con GRAFK. Sobre este / mismo juego de datos, entre este método de los vecinos recíprocos (RECIP) y el método mixto (SEMIS) los números de las clases 1 y 2 han sido permutados, esto no es importante para la interpretación pero subrayamos /// que las tres clases no son tan disjuntas en el plano de los dos primeros factores como las tres clases obtenidas con el otro método de clasificación.

Con RECIP no se escoge mas que el número de factores sobre los cuales se quiere trabajar además del nivel de corte del árbol. El resultado es entonces siempre el mismo.

Para la etapa SEMIS el resultado está influenciado / por la elección de los centros (siempre que esta influencia pueda ser reducida aumentando el número de iteraciones) por una parte, y por el valor dado a los otros parámetros por otra parte.

CLASSEMENT ASCENDANT HIERARCHIQUE

DESCRIPTION DES NIVEAUX DE LA HIERARCHIE

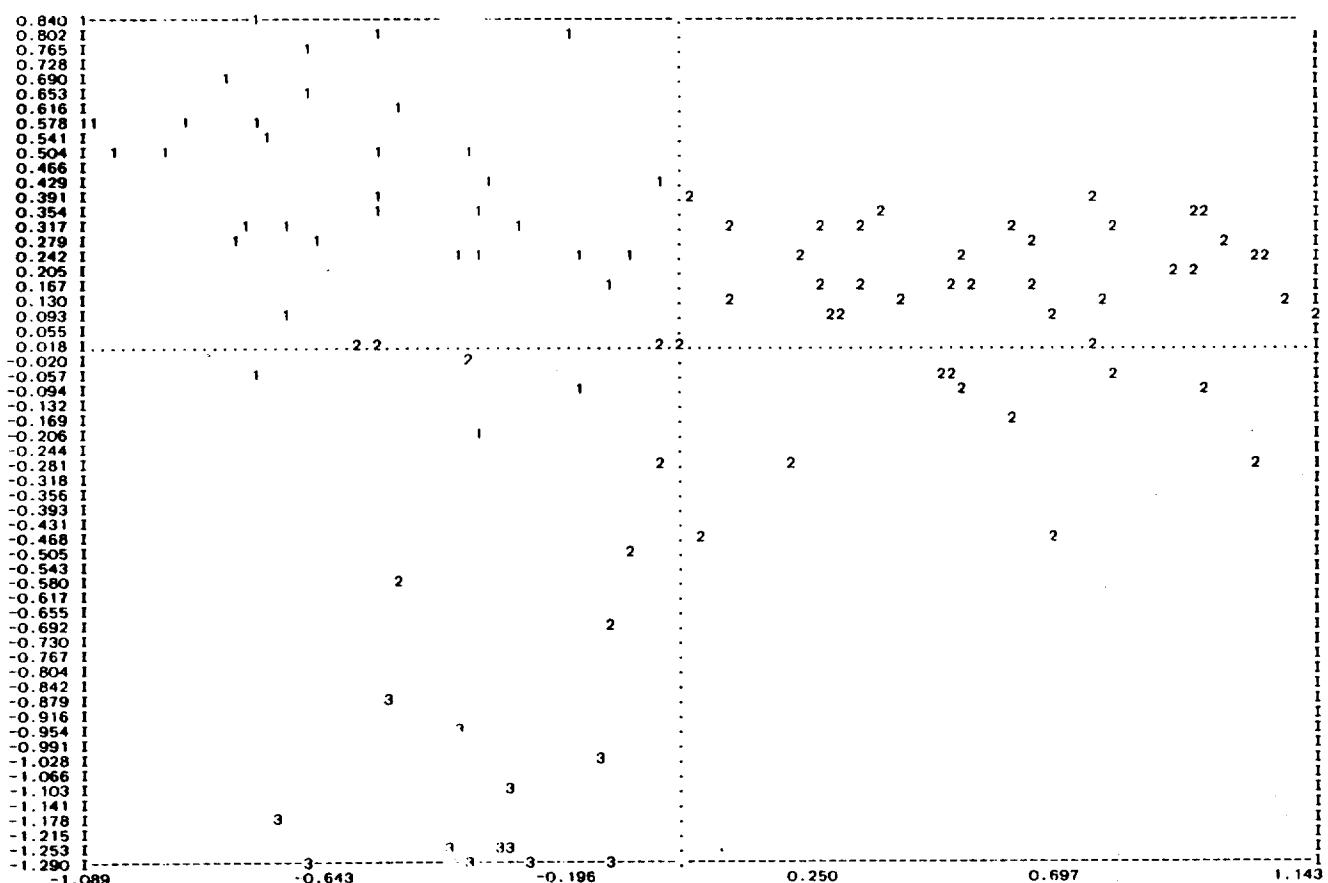
NO	ATME	BENJ	NORE	Poids	Indice	HISTOGRAMME DES INDICES DE NIVEAU
112	99	1	2	2	0.0	
113	45	108	2	2	0.0	
114	113	89	3	3	0.0	
115	106	104	2	2	0.0	
116	115	53	3	3	0.0	
117	95	101	2	2	0.0	
118	39	29	2	2	0.0	
119	36	62	2	2	0.0	
120	38	54	2	2	0.0	
121	75	32	2	2	0.0	
122	31	77	2	2	0.0	
123	42	41	2	2	0.0003818	
124	109	112	3	3	0.004189	
125	94	56	2	2	0.004210	
126	25	7	2	2	0.004265	
127	13	66	2	2	0.004265	
128	67	16	2	2	0.004380	
129	10	14	2	2	0.005033	
130	84	12	2	2	0.005033	
131	37	50	2	2	0.005033	
132	107	98	2	2	0.005153	
133	52	83	2	2	0.005174	
134	58	76	2	2	0.005203	
135	92	80	2	2	0.005257	
136	21	15	2	2	0.005697	
137	100	9	2	2	0.005816	
138	33	47	2	2	0.006013	
139	114	88	4	4	0.006398	
140	90	46	2	2	0.007313	
141	117	8	3	3	0.007596	
142	105	73	2	2	0.007939	
143	4	110	2	2	0.008414	
144	87	48	2	2	0.009095	
145	118	131	4	4	0.009202	
146	102	78	2	2	0.009519	
147	43	70	2	2	0.009628	
148	140	89	3	3	0.0010239	
149	6	61	2	2	0.0010793	
150	20	96	2	2	0.0011152	
151	51	44	2	2	0.0011209	
152	60	68	2	2	0.0011430	
153	135	40	3	3	0.0011890	
154	138	35	3	3	0.0011928	
155	97	146	3	3	0.0012027	
156	19	5	2	2	0.0012076	
157	22	125	3	3	0.0012804	
158	81	71	2	2	0.0012924	
159	141	57	4	4	0.0013215	
160	103	132	3	3	0.0013392	
161	27	147	3	3	0.0014528	
162	144	127	4	4	0.0015531	
163	137	123	4	4	0.0015599	
164	129	120	4	4	0.0016565	
165	139	65	5	5	0.0016643	
166	133	11	3	3	0.0016856	
167	72	55	2	2	0.0016872	
168	34	111	2	2	0.0016877	
169	165	126	7	7	0.0017817	
170	23	82	2	2	0.0018437	
171	30	85	2	2	0.0019382	
172	91	116	4	4	0.0019387	
173	128	26	3	3	0.0019746	
174	17	28	2	2	0.0020335	
175	18	158	3	3	0.0020613	
176	134	119	4	4	0.0020771	
177	130	145	6	6	0.0022923	
178	169	136	9	9	0.0023570	
179	155	59	4	4	0.0026214	
180	121	122	4	4	0.0026302	
181	161	143	5	5	0.0026313	
182	156	3	3	3	0.0026923	
183	177	74	7	7	0.0030849	
184	176	171	6	6	0.0031024	
185	173	150	5	5	0.0031110	
186	153	152	5	5	0.0031569	
187	183	24	8	8	0.0037804	
188	86	63	2	2	0.0038193	
189	93	124	4	4	0.0038668	
190	160	157	6	6	0.0038918	
191	79	170	3	3	0.0041545	
192	168	166	5	5	0.0042115	
193	164	64	5	5	0.0042541	
194	172	162	8	8	0.0043994	
195	142	163	6	6	0.0050684	
196	149	49	3	3	0.0053803	
197	2	185	6	6	0.0054679	
198	151	193	7	7	0.0057711	
199	192	174	7	7	0.0068758	
200	159	194	12	12	0.0078211	
201	196	182	6	6	0.0086910	
202	148	179	7	7	0.0101075	
203	187	154	11	11	0.0104817	
204	181	178	14	14	0.0107084	
205	180	186	9	9	0.0115655	
206	175	201	9	9	0.0119599	
207	197	199	13	13	0.0126691	
208	190	189	10	10	0.0136511	
209	204	208	24	24	0.0156032	
210	198	202	14	14	0.0170313	
211	200	195	18	18	0.0191904	
212	206	191	12	12	0.0194796	
213	205	184	15	15	0.0200641	
214	188	167	4	4	0.0252591	
215	210	207	27	27	0.0384030	
216	212	214	16	16	0.0593368	
217	215	203	38	38	0.0747892	
218	211	209	42	42	0.0930324	
219	216	218	58	58	0.1215211	
220	213	219	73	73	0.2632939	
221	217	220	111	111	0.2762651	

SOMME DES INDICES DE NIVEAU = 1.28901

POIDS INDICE DENDROGRAMME (LIGUELE 0.00433 0.23623)
INDICE = (IND.REEL)/(IND.MAX) IND.MAX = 1.16949

4.000 0.012 189
6.000 0.013 190
9.000 0.009 178
5.000 0.080 181
4.000 0.004 163
2.000 0.016 142
8.000 0.007 194
4.000 0.104 159
2.000 0.022 167
2.000 0.051 188
3.000 0.017 191
3.000 0.007 182
1.000 0.005 49
2.000 0.010 149
3.000 0.225 175
6.000 0.017 184
5.000 0.010 186
4.000 0.236 180
3.000 0.009 154
8.000 0.064 187
2.000 0.006 174
5.000 0.011 192
5.000 0.005 185
1.000 0.033 2
4.000 0.009 179
3.000 0.015 148
5.000 0.005 193
2.000 ----- 151

P. CLAPIER



NOMBRE DE POINTS DOUBLES = 13