METHODE + PROGRAMMES

ANALYSE CONJOINTE DE TABLEAUX QUANTITATIFS

Ch. LAVIT

Préface du Pr. Y. ESCOUFIER

MASSON

MÉTHODE+PROGRAMMES

Collection coordonnée par J.-F. Phélizon

APPLICATIONS INDUSTRIELLES ET SCIENTIFIQUES

GRAPHISME SCIENTIFIQUE SUR MICRO-ORDINATEUR. De la 2e à la 3e génération, 50 applications résolues en Basic, par R. Dony.

CALCUL DES PARTIES CACHÉES. Approximation des courbes par la méthode de Bézier et des B-splines, par R. Dony.

ABC DE CRYPTOGRAPHIE. Avec programmes en Basic, par R. Roubaty.

CALCUL SYMBOLIQUE ET INFORMATIQUE. Du calcul numérique au calcul littéral, programmes en Basic, par A. Deshayes.

SYSTEMES FLEXIBLES DE PRODUCTION. Conception préliminaires, par J.-M. Proth, H. Quentin de Gromard.

MÉTHODES DE CALCUL EN ÉLECTRONIQUE. Programmes en L.S.E. et en Basic, par A. Billès.

TEST DES CIRCUITS NUMÉRIQUES. Aide à la programmation, par J.-C. Couchot.

INTELLIGENCE ARTIFICIELLE

EXERCICES PROGRAMMÉS D'INTELLIGENCE ARTIFICIELLE, par H. Farreny.

INTELLIGENCE ARTIFICIELLE EN MÉDECINE. Des systèmes experts, par M. Fieschi.

COMPRÉHENSION AUTOMATIQUE DU LANGAGE NATUREL.
Le cas du groupe nominal en français, par J.-H. Jayez.
Le cas de l'interrogation simple en français, par A. Gal, J.-H. Jayez, P. Levasseur, M. Liscouet et J.M. Walle.

THÉORIE DES POSSIBILITÉS. Applications à la représentation des connaissances en informatique, par D. Dubois et H. Prade.

INFORMATIQUE POUR L'ENTREPRISE

CONCEPTION ASSISTÉE DES APPLICATIONS INFORMATIQUES,
tome 1.-Étude d'opportunité et analyse conceptuelle, par F. Bodart et Y. Pigneur.
tome 2.-Conception de la base de données, par J.-L. Hainaut.

BASES DE DONNÉES. Conception, réalisation et implantation sur micro-ordinateurs, par J.-L. Tomas.

INFORMATIQUE DOCUMENTAIRE, par A. Dewèze.

PROCÉDURES DE TRI. Programmation en Basic et en Pascal, par R. Guihur.

Suite de la liste en page 3 de couverture.

ANALYSE CONJOINTE DE TABLEAUX QUANTITATIFS

CHEZ LE MÊME ÉDITEUR

Dans la même collection :

Voir page 2 de couverture.

Autres ouvrages :

LA RÉGRESSION. Nouveaux regards sur une ancienne méthode statistique, par R. TOMASSONE, E. LESQUOY et C. MILLIER. *Actualités scientifiques et agronomiques de l'I.N.R.A., n° 13.* 1983, 184 pages.

DISCRIMINATION ET CLASSEMENT, par R. TOMASSONE, M. DANZART, J.-J. DAUDIN et J.-P. MASSON. *Techniques Stochastiques.* 1988, 184 pages.

STATISTIQUE APPLIQUÉE A L'EXPLOITATION DES MESURES. Exposé pratique illustré d'exemples concrets, par la commission d'établissement des méthodes d'analyses du commissariat à l'énergie atomique (CETAMA). 1986, 2e édition revue et corrigée, 576 pages.

INTRODUCTION AUX ANALYSES STATISTIQUES MULTIDIMENSIONNELLES, par J. LEFEBVRE. 1983, 3e édition augmentée, 296 pages.

ANALYSE TYPOLOGIQUE. Théories et applications, par J.-L. CHANDON et S. PINSON. 1981, 264 pages.

HÉTÉROGÉNÉITÉ, ÉCHANTILLONNAGE, HOMOGÉNÉISATION. Ensemble cohérent de théories, par P. GY. *Mesures Physiques.* 1988, 624 pages.

LES ARBRES ET LES REPRÉSENTATIONS DES PROXIMITÉS, par J.-P. BARTHÉLEMY et A. GUÉNOCHE. *Méthode + Programmes* 1988, 256 pages.

METHODE + PROGRAMMES

Collection coordonnée par J.-F. PHELIZON

ANALYSE CONJOINTE DE TABLEAUX QUANTITATIFS

Christine LAVIT

Maître de Conférences
Université des Sciences et Techniques du Languedoc

avec la collaboration de Marie-Claire BERNARD, Carlos PEREZ-HUGALDE et Marie-Odile PERNIN

Préface du Pr. Y. ESCOUFIER

MASSON Paris Milan Barcelone Mexico **1988**

Tous droits de traduction, d'adaptation et de reproduction par tous procédés, réservés pour tous pays.

Toute reproduction ou représentation intégrale ou partielle, par quelque procédé que ce soit, des pages publiées dans le présent ouvrage, faite sans l'autorisation de l'éditeur est illicite et constitue une contrefaçon. Seules sont autorisées, d'une part, les reproductions strictement réservées à l'usage privé du copiste et non destinées à une utilisation collective, et d'autre part, les courtes citations justifiées par le caractère scientifique ou d'information de l'œuvre dans laquelle elles sont incorporées (loi du 11 mars 1957 art. 40 et 41 et Code Pénal art. 425).

Des photocopies payantes peuvent être réalisées avec l'accord de l'éditeur. S'adresser au : Centre Français du Copyright, 6, bis, rue Gabriel-Laumain, 75010 Paris, tél. : 48.24.98.30.

© *Masson, Paris, 1988*

ISBN : 2-225-81478-3
ISSN : 0757-2093

MASSON	120, bd Saint-Germain, 75280 Paris Cedex 06
MASSON ITALIA EDITORI	Via Statuto 2, 20121 Milano
MASSON S.A.	Balmes 151, 08008 Barcelona
MASSON EDITORES	Dakota 383, Colonia Napoles, 03810 Mexico D.F.

Préface

En proposant un ouvrage sur la méthode Statis, Christine LAVIT mène à son aboutissement un travail collectif amorcé dès 1972 au laboratoire de statistique de l'Université des Sciences et Techniques du Languedoc à Montpellier puis poursuivi au sein de l'Unité de Biométrie dans laquelle, depuis 1982, ce laboratoire collabore avec la chaire de mathématiques et informatique de l'Ecole Nationale Supérieure Agronomique de Montpellier et avec le département de biométrie de l'Institut National de la Recherche Agronomique.

Transmettre les acquis d'un groupe restreint de chercheurs à un large public n'est pas chose simple. C'est le mérite de ce livre de le faire dans un texte qui pourra satisfaire des lecteurs aux intérêts divergents.

Le mathématicien y trouvera une présentation élégante et complète des fondements mathématiques de la méthode ainsi que de ses ressemblances et différences avec des méthodes alternatives, proposées par d'autres équipes françaises ou étrangères.

Le praticien de l'analyse des données reconnaîtra dans la diversité des exemples réels longuement étudiés les analogies avec ses propres problématiques ; il comprendra la nature des résultats qu'il peut obtenir et apprendra les régles de lecture de ses résultats.

La mise en oeuvre d'une méthode comme Statis n'est pas envisageable sans le recours à l'ordinateur. La longue expérience acquise par Christine LAVIT dans l'utilisation de la méthode lui a permis de définir des logiciels d'utilisation simple.

L'un des logiciels disponible dans la bibliothèque Modulad nécessite l'emploi d'un gros ordinateur. Un autre, disponible auprès de l'auteur, est utilisable sur micro-ordinateur. Le nombre d'instructions de ces logiciels justifie de réserver leur diffusion à des supports magnétiques. Il reste que, pour aider l'informaticien qui voudrait développer son propre outil, l'ouvrage fournit une description détaillée des étapes à programmer. L'évolution de l'informatique pourra rendre obsolètes les logiciels disponibles aujourd'hui. On trouvera alors dans cette description la base de programmes nouveaux.

De même que les années soixante ont été en France et à l'étranger une période de foisonnement pour les méthodes d'analyse d'un tableau de données, les années quatre-vingt sont celles de l'analyse conjointe de plusieurs tableaux. Cet ouvrage s'insère parmi les nombreuses propositions qui sont faites. Par la rigueur de ses justifications théoriques et le détail des applications qu'il présente, il deviendra à coup sûr une référence précieuse pour les chercheurs de toutes les disciplines que leur investigations conduisent à l'analyse conjointe de plusieurs tableaux de données.

Yves ESCOUFIER
Professeur à l'USTL,
Directeur du Laboratoire de Biométrie
de l'INRA, associé à la Chaire de Mathématiques
et Informatique de l'ENSA MONTPELLIER.

Table des matières

PREFACE 5

CHAPITRE I. GENERALITES SUR LA METHODE STATIS 11

1. **Clé de lecture de l'ouvrage** 11

2. **Exemples d'utilisation de la méthode STATIS** 13

2.1. Phénomènes évoluant dans le temps 13
2.2. Composition du latex le long du tronc d'un hévéa 13
2.3. Suivi de patients 15

3. **Description rapide de la méthode** 17

4. **Paternité de la méthode** 19

5. **Différentes techniques d'analyse conjointe de tableaux de données** 21

CHAPITRE II. APPLICATIONS LINEAIRES ENTRE ESPACES VECTORIELS EUCLIDIENS 23

1. **Espaces euclidiens** 23

1.1. Espace vectoriel euclidien 23
1.2. Dual d'un espace vectoriel euclidien 25
1.3. Espace affine euclidien 28

2. **Applications linéaires entre espaces vectoriels euclidiens** 29

2.1. Transposée et adjointe d'une application linéaire 29
2.2. Opérateurs auto-adjoints 31

3. **Produit scalaire sur l'espace vectoriel des applications linéaires** 37

3.1. Trace d'un opérateur 37
3.2. Produit scalaire entre applications linéaires 38

4. **Valeurs singulières d'une application linéaire** 39

4.1. Décomposition d'une application linéaire 39
4.2. Approximation d'une application linéaire 40

5. **Semi-produit scalaire sur un espace vectoriel** 43

5.1. Espace vectoriel semi-normé 43
5.2. Propriétés d'un semi-produit scalaire sur un espace vectoriel euclidien 44

CHAPITRE III. ANALYSE EN COMPOSANTES PRINCIPALES NON CENTREE D'UN NUAGE DE POINTS PESANTS 47

1. **Objectif de l'analyse en composantes principales non centrée** 47

2. **Définition d'une image euclidienne** 49

3. **Recherche d'une image euclidienne des objets, associée aux produits scalaires** 51

3.1. Forme bilinéaire, symétrique, positive, associée au tableau des produits scalaires 51
3.2. Image euclidienne obtenue à partir d'une forme bilinéaire, symétrique, positive, sur un espace vectoriel euclidien 52
3.3. Représentation d'objets supplémentaires dans cette image euclidienne 54

4. **Recherche d'une image euclidienne approchée dans un espace de dimension fixée** 56

4.1. Solution obtenue par approximation des produits scalaires 56
4.2. Expression de l'approximation en termes d'inertie 58

5. **Cas particulier de produits scalaires centrés** 62

5.1. Centrage du tableau des produits scalaires 62
5.2. Conséquences du centrage 63

6. **Analyse en composantes principales d'un tableau croisant individus et variables** 66

6.1. Produits scalaires entre individus 67
6.2. Analyse en composantes principales 69
6.3. Distances euclidiennes usuelles entre individus 75

CHAPITRE IV. DESCRIPTION THEORIQUE DE LA METHODE STATIS 77

1. **Principe de la méthode** 77

1.1. Quel genre de données peut-on traiter ? 77
1.2. La notion d'étude (X, q, d) 78
1.3. Stratégie de la méthode STATIS 81

2. **Coefficients d'association entre études, proposés par Y. ESCOUFIER** 83

2.1. Produits scalaires entre objets w_k ou v_k 84
2.2. Propriétés de ces produits scalaires 85

3. **Image euclidienne des études, associée aux produits scalaires entre objets** 89

3.1. Choix de l'objet représentant l'étude 89
3.2. Construction de l'image euclidienne des études 90
3.3. Conséquence de la positivité des produits scalaires entre objets 91

4. **Le compromis, résumé majoritaire des objets** 93

4.1. Définition et propriétés du compromis 93
4.2. Interprétation du compromis 96

5. **Représentation des individus, dans le cas d'études (X_k, q_k, d)** 98

5.1. Image euclidienne compromis des individus 98
5.2. Trajectoires des individus 99
5.3. Interprétation du sens de parcours d'une trajectoire, en fonction des variables 102

6. Représentation des variables, dans le cas d'études (X_k, r, d_k) 106

6.1. Image euclidienne compromis des variables 106
6.2. Trajectoires des variables 108

CHAPITRE V . PROGRAMMATION DE LA METHODE STATIS 110

1. Données recueillies sur les mêmes individus 111

1.1. Eléments de l'étude (X_k, Q_k, D) 111
1.2. Calcul des produits scalaires entre W_k 112

2. Image euclidienne des études, associée aux produits scalaires entre W_k 113

2.1. Coordonnées des points 113
2.2. Image euclidienne centrée 114
2.3. Etude supplémentaire 115

3. Positions compromis et trajectoires des individus 116

3.1. Positions compromis des individus 116
3.2. Explication des axes du compromis par les variables 116
3.3. Trajectoires des individus 117

4. Résumé des calculs de base 117

5. Données concernant les mêmes variables 119

6. Description du logiciel STATIS 120

6.1. Sur quel matériel peut-on l'installer ? 120
6.2. Choix proposés pour l'utilisation de la méthode 121
6.3. Modalités d'utilisation du logiciel 122
6.4. Résultats fournis par le logiciel 123

CHAPITRE VI. EXPLICATION DETAILLEE DES RESULTATS FOURNIS PAR LA METHODE STATIS, APPLIQUEE A DES DONNEES DE CROISSANCE 125

1. Description des données 125

1.1. Présentation rapide du fichier auxologique français 125
1.2. Extrait du fichier exploité par la méthode STATIS et objectif poursuivi 127
1.3. Choix effectués pour utiliser la méthode STATIS 127

2. Etude des corrélations entre variables 128

2.1. Interstructure sur les objets v_k 128
2.2. Evolution des corrélations entre variables 134
2.3. Conclusions de l'étude des variables 137

3. Etude des évolutions individuelles 139

3.1. Interstructure sur les objets w_k 139
3.2. Positions compromis des individus 141
3.3. Trajectoires individuelles 143
3.4. Conclusions de l'étude des individus 149

Fichier des données 150

CHAPITRE VII. EVOLUTION ECONOMIQUE DES PROVINCES ESPAGNOLES ENTRE 1960 ET 1979 154

1. **Description des données** 154

1.1. Evolution globale de l'économie espagnole entre 1960 et 1979 154
1.2. Description des données et de l'objectif poursuivi 155

2. **Interstructure et détermination du compromis** 158

3. **Signification économique des axes du compromis** 158

4. **Positions compromis des provinces** 164

5. **Trajectoires des provinces** 166

6. **Conclusions** 169

CHAPITRE VIII. TYPOLOGIE DES CANTONS RURAUX DE L'HERAULT, DECRITS PAR L'EVOLUTION DE LEUR POPULATION ACTIVE ENTRE 1954 ET 1982 171

1. **Description des données** 171

1.1. Le changement social dans les campagnes languedociennes 171
1.2. Les cantons ruraux de l'Hérault 172
1.3. Choix effectués pour l'utilisation de la méthode 175

2. **Analyse de l'interstructure** 175

3. **Typologie des cantons, d'après leurs positions compromis** 177

3.1. Analyse des effectifs 177
3.2. Analyse des profils 181

4. **Typologie des cantons, d'après leurs trajectoires** 181

4.1. Analyse des effectifs 182
4.2. Analyse des profils 185

5. **Conclusions** 186

Fichier des données 188

REFERENCES BIBLIOGRAPHIQUES 193

ANNEXE : PROGRAMMES EN TURBO PASCAL 203

NOTATIONS 249

INDEX ALPHABETIQUE 253

I. Généralités sur la méthode STATIS

Cet ouvrage décrit la méthode STATIS (Structuration des Tableaux A Trois Indices de la Statistique) permettant l'exploration simultanée de plusieurs tableaux de données. Pour mettre en oeuvre cette méthode, nous proposons un logiciel d'utilisation simple, exécutable sur micro-ordinateur IBM PC (*).

Notre but étant d'expliquer les possibilités et les limites de la méthode, nous avons rassemblé dans cette monographie les fondements théoriques et des exemples d'application sur des données réelles. Cependant la présentation de l'ouvrage est organisée de telle sorte que le lecteur puisse le parcourir selon ses préoccupations, et le temps dont il dispose.

1. CLE DE LECTURE DE L'OUVRAGE

Sans entrer dans la théorie. Pour avoir un aperçu des résultats fournis par la méthode STATIS, on peut se reporter aux exemples d'application, décrits dans le paragraphe 2 de ce chapitre, et aux chapitres VII, VIII et enfin VI. La lecture du paragraphe 3 ci-après donnera ensuite une idée générale de la méthode.

(*) Les fonctions du logiciel, ainsi que les conditions d'acquisition, sont décrites p. 120 .

Pour avoir une description complète de la méthode. Les trois chapitres théoriques II, III et IV sont très détaillés et forment un tout. Pour la lecture de ces chapitres, on ne suppose connues que les notions classiques d'algèbre linéaire : espaces vectoriels, bases et applications linéaires ; et les notions élémentaires de statistique : centrage, réduction d'une variable et calcul d'un coefficient de corrélation linéaire.

La méthode STATIS est décrite en détail dans le chapitre IV. Cette méthode est basée sur l'analyse en composantes principales, non centrée sur le centre de gravité des points, à laquelle nous avons consacré le chapitre III. Remarquons que "l'analyse en composantes principales" familière à l'utilisateur de l'Analyse des Données, en est un cas particulier : centrée pour les individus, non centrée pour les variables. Le chapitre II détaille les théorèmes d'algèbre linéaire utilisés par la suite.

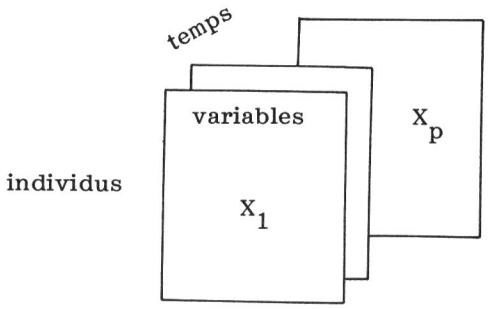
Sans entrer dans les démonstrations. Nous avons donné systématiquement un titre à chaque définition, théorème et paragraphe, pour permettre au lecteur de se repérer facilement, et de suivre l'enchaînement du raisonnement sans rentrer dans les détails. Les notations introduites dans le chapitre II sont regroupées à la fin de l'ouvrage. Dans la suite du texte, les propriétés auxquelles on fait référence sont signalées par un report à la page dans laquelle elles ont été démontrées. L'index alphabétique renvoie à la définition de l'expression utilisée.

Pour programmer la méthode. La méthode STATIS est simple à mettre en oeuvre. Les seuls calculs nécessaires sont la multiplication de matrices, le calcul de la trace d'une matrice carrée et la diagonalisation d'une matrice symétrique. La programmation de la méthode est extrêmement rapide si on dispose d'un langage matriciel. Les calculs matriciels nécessaires sont détaillés dans le chapitre V et programmés en Turbo Pascal en annexe.

2. EXEMPLES D'UTILISATION DE LA METHODE STATIS

2.1. PHENOMENES EVOLUANT DANS LE TEMPS

Le plus souvent, comme pour les exemples étudiés dans les chapitres VI et VII, les tableaux de données $X_1,\ldots,X_p$ décrivent un phénomène évolutif dans le temps, les mêmes variables étant mesurées sur les mêmes individus.



Cependant, il n'est pas nécessaire d'avoir toujours les mêmes variables pour utiliser la méthode STATIS. Dans le chapitre VIII par exemple, les données sont extraites des cinq derniers recensements de population, alors que la définition des catégories socioprofessionnelles a été modifiée entre les deux derniers recensements. En médecine, il arrive que les examens ne soient pas toujours réalisés avec le même matériel au cours du temps, ou qu'accidentellement un dosage n'ait pu être effectué. Le suivi des patients pourra quand même être analysé par la méthode STATIS, mais il est bien évident qu'il faudra tenir compte de ces perturbations dans l'interprétation des résultats.

Nous présentons les différentes possibilités de la méthode sur des exemples concrets.

2.2. COMPOSITION DU LATEX LE LONG DU TRONC D'UN HEVEA

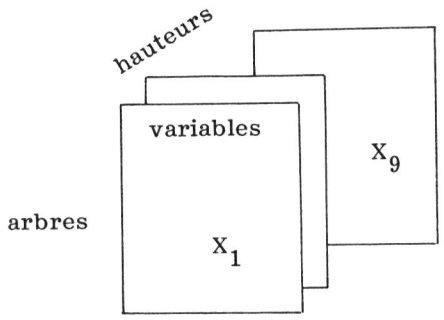
Nous utilisons comme exemple illustratif, une expérimentation réalisée par l'Institut de Recherche sur le Caoutchouc, dont l'analyse statistique a été effectuée par P. TRAISSAC (*).

(*) Assistant, Ecole Nationale Supérieure Agronomique de Montpellier.

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

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

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



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

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

Trajectoires de points moyens. La dernière étape de la méthode trace la "trajectoire" de chaque arbre, c'est-à-dire les différentes positions de l'arbre décrit par chaque tableau. Ces trajectoires individuelles ne nous intéressent pas particulièrement. Par contre, on visualise la différence entre les deux traitements en créant un arbre "moyen" fictif pour chaque traitement, et en traçant la trajectoire de ces deux arbres supplémentaires. Les deux trajectoires se séparent très nettement, et montrent que l'évolution de la composition physiologique du latex le long du tronc diffère suivant le traitement. De plus, le sens de parcours des trajectoires s'interprète en fonction de la teneur en extrait sec et en glucides, et donne une explication de la différence constatée entre les traitements.

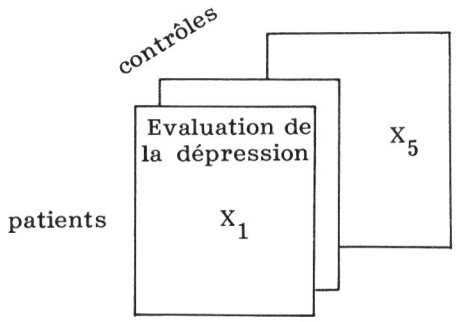
2.3. SUIVI DE PATIENTS

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

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

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

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



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

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

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

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

3. DESCRIPTION RAPIDE DE LA METHODE

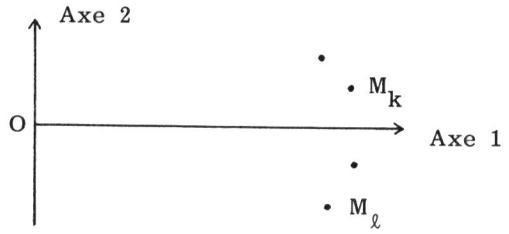
La méthode STATIS est une méthode exploratoire d'Analyse de Données, qui s'applique à des données quantitatives : p tableaux de mesures X_k ont été recueillies en différentes occasions sur les *mêmes* individus.

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

Analyse globale des relations entre tableaux. On compare les tableaux au moyen des "objets" W_k.

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

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



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

Positions compromis des individus. A partir de l'image euclidienne des tableaux, on construit un objet *compromis* W qui peut être considéré comme un tableau de produits scalaires entre individus. L'image euclidienne des individus, associée à ces produits scalaires, représente les positions mutuelles compromis des individus tels qu'ils sont décrits par l'ensemble des tableaux.

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

Corrélations des variables avec les axes du compromis. On peut considérer que les coordonnées des individus sur un axe sont les valeurs d'une variable fictive, appelée "composante principale".

Pour interpréter les positions des individus le long de l'axe, on calcule les corrélations de la composante principale avec les variables des différents tableaux, ou avec des variables exogènes dont on connaîtrait les valeurs sur les individus.

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

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

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

4. PATERNITE DE LA METHODE

La méthode STATIS est l'aboutissement d'un travail collectif de l'équipe dirigée par Monsieur le Professeur Y. ESCOUFIER, dans le cadre du laboratoire de statistique de l'Université des Sciences et Techniques du Languedoc, puis de l'unité de biométrie dans laquelle ce laboratoire collabore avec la chaire de mathématique et informatique de l'Ecole Nationale Supérieure Agronomique de Montpellier, et avec le département de biométrie de l'Institut National de la Recherche Agronomique.

Premiers travaux. L'interstructure exposée dans PAGES - ESCOUFIER - CAZES (1976) et CAILLIEZ - PAGES (1976) est basée sur les coefficients d'association entre tableaux proposés par Y. ESCOUFIER (1970).

Développée par H. L'HERMIER DES PLANTES dans sa thèse de 3e cycle en 1976, la méthode STATIS a été comparée aux techniques proposées par CARROLL - CHANG (1970), TUCKER (1966), et GOWER (1975) par F. GLAÇON dans sa thèse de 3e cycle en 1981.

Premières applications. La version programmée par M.Cl. PLACE en 1980, puis Ch. LAVIT et C. ROUX (*) en 1983, a été implantée dans la bibliothèque de programmes d'Analyse de Données ANADO du Centre National Universitaire Sud de Calcul. Ce premier logiciel a permis de mettre la méthode à l'épreuve sur des exemples réels, et d'en améliorer la technique.

Par exemple, si l'interstructure n'a jamais été controversée, le calcul des trajectoires a posé quelques problèmes et suscité des propositions diverses. Les chercheurs qui ont collaboré à cet ouvrage, ont fait partie des premiers utilisateurs et nous ont aidé à choisir la solution actuelle qui, bien que non optimale, semble refléter mieux que les autres, les écarts individuels.

Programme actuel. Nous appuyant sur l'expérience de l'utilisation de cette première version, nous avons allégé les calculs et les sorties, en supprimant toutes les informations redondantes.

La version actuelle (LAVIT 1986) a été réalisée avec l'aide du club MODULAD, qui a fourni des normes strictes de programmation pour que le logiciel soit portable, et qui l'a testé sur différents matériels, et les conseils de G. CARAUX (**) pour la programmation de la table traçante BENSON. Cette version a été adaptée en 1987 par G. FRANCILLON (***), pour IBM PC.

(*) INRA, Unité de Biométrie, MONTPELLIER.

(**) Maître Assistant, Ecole Nationale Supérieure Agronomique de MONTPELLIER.

(***) Ingénieur informaticien, Centre de Coopération Internationale en Recherche Agronomique pour le Développement.

La présentation des chapitres théoriques II, III et IV est inspirée des travaux de R. SABATIER (1987). Sa grande compétence nous a été très utile.

5. DIFFERENTES TECHNIQUES D'ANALYSE CONJOINTE DE TABLEAUX DE DONNEES

La bibliographie présentée p. 196 et suivantes, n'a pas fait l'objet d'une recherche systématique, et n'est absolument pas représentative de l'ensemble du travail effectué dans ce domaine, notamment hors de France. On peut cependant classer les méthodes selon l'aspect des données qu'elles privilégient, et l'information particulière qu'elles en extraient.

Approche symétrique du "cube" de données. Le modèle désigné sous le nom de TUCKER-3 par KROONENBERG (1983) - éventuellement généralisé à quatre (LASTOVICKA (1981)) ou à n dimensions (POLIT (1986)) - cherche à réduire l'information contenue dans un cube de données en fonction d'un petit nombre d'individus types, de variables latentes et de conditions prototypes. L'objectif de cette méthode est la reconstruction du cube de données à partir de ces éléments.

Les données sont considérées comme un ensemble de tableaux. La plupart des méthodes se classent dans cette catégorie, celles qui empilent les tableaux, comme celles qui les juxtaposent (double analyse en composantes principales, analyse canonique généralisée, méthode LONGI ...).

Certaines méthodes supposent l'existence d'un modèle sous-jacent comme INDSCAL (CARROLL - CHANG (1970)), ou TUCKER-2. D'autres cherchent à mettre en évidence une structure commune aux différents tableaux : par exemple, des composantes principales communes (FLURY (1983) - KRZANOWSKI (1984)). Parmi ces dernières méthodes, nombreuses sont celles qui sont basées sur l'emploi de coefficients d'association entre tableaux (FOUCART (1984), COPPI (1986), BOLASCO (1986), D'AMBRA et MARCHETTI (1986)), ou sur l'utilisation d'un tableau moyen servant de référence.

D'ALESSIO (1986) propose une reconstruction des données à partir de plusieurs compromis.

Nous classons également dans cette catégorie, les méthodes qui ajustent un tableau à un autre (comme la méthode PROCUSTE), ou à un modèle (ESCOFIER (1984) - HEIJDEN et de LEEUW (1985)).

Les données sont considérées comme une succession de tableaux. Très souvent, les tableaux décrivent un phénomène évolutif, tout en étant trop espacés dans le temps, ou trop peu nombreux, pour qu'on puisse se placer dans le cadre de la théorie des processus, ou des séries chronologiques.

On tient compte de la structure ordonnée du temps lorsqu'on considère le premier tableau comme tableau de référence, ou qu'on analyse les différences entre deux tableaux successifs. L'analyse des évolutions (LE FOLL (1972) - CARLIER (1985) - PICARD (1987)) privilégie les distances entre deux points successifs d'une trajectoire. Enfin, certaines méthodes ajustent les données d'un tableau à un modèle faisant intervenir le temps (KROONENBERG (1983)), ou établissent un modèle de prévision (FOUCART (1981)).

II. Applications linéaires entre espaces vectoriels euclidiens

Pour la lecture de ce chapitre, on suppose connues les notions classiques d'algèbre linéaire : espaces vectoriels, bases, applications linéaires, valeurs propres et vecteurs propres d'une application linéaire diagonalisable. On introduit la notion d'espace euclidien et on se limite à l'énoncé des propriétés des applications linéaires entre ces espaces, utiles pour la suite.

1. ESPACES EUCLIDIENS

1.1. ESPACE VECTORIEL EUCLIDIEN

Produit scalaire. On appelle produit scalaire une forme bilinéaire, symétrique, définie, positive sur un espace vectoriel réel E. C'est-à-dire une application q de ExE dans R qui fait correspondre au couple (x,y) le nombre $(x|y)_q$ tel que :

$(x|y)_q = (y|x)_q$ — q est symétrique,

$(\lambda x|y)_q = \lambda\,(x|y)_q \qquad \lambda \in R$ — q est linéaire en x

$(x+z|y)_q = (x|y)_q + (z|y)_q$ — donc en y par symétrie,

$(x|x)_q = 0 \iff x = 0$ — q est définie,

$(x|x)_q \geqslant 0$ — q est positive.

Espace vectoriel euclidien. On appelle espace vectoriel euclidien le couple (E, q) formé d'un espace vectoriel réel E de dimension finie n, et d'un produit scalaire q sur E.

Orthogonalité. Deux vecteurs x et y de E sont dits q-orthogonaux si $(x|y)_q = 0$.

F étant un sous-espace vectoriel de E, l'ensemble des vecteurs q-orthogonaux à tout vecteur de F forme un sous-espace noté $F^{\perp}$. Il est facile de vérifier que E est la somme directe de F et $F^{\perp}$.

Inégalité de Cauchy-Schwarz. x et y étant deux vecteurs quelconques de E, $(x|y)_q^2 \leqslant (x|x)_q \cdot (y|y)_q$. L'égalité est réalisée si et seulement si x et y sont linéairement dépendants.

□ *Démonstration.* Il suffit d'écrire que $(x+ty|x+ty)_q$ est un trinôme en t, positif ou nul quel que soit t □

Norme euclidienne. L'application de E dans R^+ qui associe à tout vecteur x, le nombre réel $\| x \|_q = (x|x)_q^{1/2}$ est une norme. Elle vérifie les propriétés d'une norme :

$$\| x \|_q \geqslant 0$$

$$\| x \|_q = 0 \iff x = 0$$

$$\| \lambda x \|_q = |\lambda| \cdot \| x \|_q$$

$$\| x+y \|_q \leqslant \| x \|_q + \| y \|_q$$ (découle de l'inégalité de Cauchy-Schwarz),

mais aussi des propriétés spécifiquement euclidiennes comme le théorème de Pythagore :

$$x \text{ et } y \ q\text{-orthogonaux} \iff \| x+y \|_q^2 = \| x \|_q^2 + \| y \|_q^2 ,$$

et l'égalité du parallélogramme :

$$\| x+y \|_q^2 + \| x-y \|_q^2 = 2(\| x \|_q^2 + \| y \|_q^2).$$

Distance euclidienne. L'espace vectoriel euclidien (E, q) est aussi un espace métrique. En effet, si on définit la distance entre deux vecteurs x et y de E par :

$$d_{xy} = \| x-y \|_q = (\| x \|_q^2 + \| y \|_q^2 - 2(x|y)_q)^{1/2} ,$$

d_{xy} vérifie les propriétés d'une distance (qui découlent dans ce cas des propriétés de la norme).

$$d_{xy} = d_{yx}$$
$$d_{xy} \geqslant 0$$
$$d_{xy} = 0 \quad \Longleftrightarrow \quad x = y$$
$$d_{xy} \leqslant d_{xz} + d_{zy} \qquad \text{(inégalité triangulaire).}$$

Remarque. Le produit scalaire q est appelé par abus "métrique".

Sous-espace. La restriction de q à un sous-espace vectoriel de E est encore un produit scalaire. Tout sous-espace vectoriel de E est donc un sous-espace euclidien de (E, q).

Base q-orthonormée. Lorsque les vecteurs $e_1,\ldots,e_n$ sont q-orthogonaux deux à deux, la base $(e_1,\ldots,e_n)$ est dite q-orthogonale. Si, en plus, chaque norme $\| e_i \|_q$ est égale à 1, la base $(e_1,\ldots,e_n)$ est dite q-orthonormée.

Sur une base q-orthonormée, la décomposition d'un vecteur x s'écrit :

$$x = \sum_{i=1}^{n} (x|e_i)_q \; e_i \quad .$$

Un espace euclidien (E, q) possède une base q-orthonormée.

□ *Démonstration* par récurrence sur la dimension de E. q étant définie, pour tout vecteur x non nul, $e_1 = \dfrac{x}{\| x \|_q}$ est une base du sous-espace F engendré par x. D'après l'hypothèse de récurrence, le sous-espace $F^{\perp}$ possède une base $(e_2,\ldots,e_n)$ orthonormée pour la restriction de q à ce sous-espace, et il est clair que $(e_1,\ldots,e_n)$ forme une base q-orthonormée de E □

1.2. DUAL D'UN ESPACE VECTORIEL EUCLIDIEN

1.2.1. DEFINITION DE L'ESPACE DUAL

Espace dual. Le dual E^* d'un espace vectoriel euclidien (E, q) est l'espace vectoriel $\mathcal{L}(E, R)$ des applications linéaires de E dans R, appelées formes linéaires sur E.

Notation d'une forme linéaire. On note $\langle x|u\rangle_{E^*}$ la valeur de la forme linéaire u appliquée au vecteur x. Si l'espace n'a pas besoin d'être précisé, on écrira simplement $\langle x|u\rangle$.

Base duale. Soit $(e_1,\ldots,e_n)$ une base de E. Les éléments e_i^* de E^*, définis par leurs valeurs sur $e_1,\ldots,e_n$ de la manière suivante :

$$\forall\, i \in \{1,\ldots,n\} \qquad \langle e_i|e_i^*\rangle = 1$$

$$\forall\, i, \forall\, j \in \{1,\ldots,n\} \qquad \langle e_j|e_i^*\rangle = 0 \quad \text{pour } i \neq j\ ,$$

forment une base de E^*, appelée base duale de $(e_1,\ldots,e_n)$.

□ *Démonstration.* Soit x un vecteur de E, et $\sum_{i=1}^{n} x_i\, e_i$ sa décomposition dans la base $(e_1,\ldots,e_n)$. Alors, d'après la définition de e_i^* : $\langle x|e_i^*\rangle = x_i$. Il en résulte que toute forme linéaire u de E est une combinaison linéaire des e_i^* :

$$\langle x|u\rangle = \langle \sum_i \langle x|e_i^*\rangle\, e_i|u\rangle = \sum_i \langle x|e_i^*\rangle \langle e_i|u\rangle = \langle x \,|\, \sum_i \langle e_i|u\rangle\, e_i^*\rangle$$

qui entraîne que $u = \sum_i \langle e_i|u\rangle\, e_i^*$ □

Remarque. Lorsque $(e_1,\ldots,e_n)$ est une base q-orthonormée : $\langle x|e_i^*\rangle = (x|e_i)_q$.

1.2.2. EXPRESSION DU PRODUIT SCALAIRE COMME ISOMORPHISME DE L'ESPACE SUR SON DUAL

Associons au produit scalaire q, l'application linéaire q qui fait correspondre à un vecteur quelconque y de E, la forme linéaire q(y) définie par :

$$\forall\, x \in E \qquad \langle x|q(y)\rangle = (x|y)_q\ .$$

Propriété. q est un isomorphisme de E sur E^*.

□ *Démonstration.* Pour toute forme linéaire u, il existe un vecteur y tel que : $\forall\, x \in E \quad \langle x|u\rangle = (x|y)_q$. En effet, si on se place dans une base q-orthonormée $(e_1,\ldots,e_n)$ de E :

$$\langle x|u\rangle = \langle x|\sum_i \langle e_i|u\rangle\, e_i^*\rangle = \sum_i \langle e_i|u\rangle\, (x|e_i)_q = (x|\sum_i \langle e_i|u\rangle\, e_i)_q\ .$$

Supposons qu'il existe deux vecteurs y_1 et y_2 tels que les formes linéaires $q(y_1)$ et $q(y_2)$ soient égales. Appliquées à un vecteur quelconque x, en particulier $x = y_1 - y_2$, elles donnent le même résultat : $(x|y_1)_q = (x|y_2)_q$ entraîne $(x|y_1 - y_2)_q = 0$, soit $\| y_1 - y_2 \|_q = 0$ qui implique $y_1 = y_2$ □

1.2.3. PRODUIT SCALAIRE SUR L'ESPACE DUAL

Produit scalaire dual. L'application q^* de $E^* \times E^*$ dans R, définie de la manière suivante :

$$\forall u,\ \forall v \in E^* \qquad (u|v)_{q^*} = \langle q^{-1}(u)|v \rangle,$$

est un produit scalaire sur E^*, appelé produit scalaire dual de q.

□ *Démonstration.* En remarquant que, q étant bijective, il existe un couple (x,y) de ExE tel que :

$$\langle q^{-1}(u)|v\rangle = \langle x|q(y)\rangle = (x|y)_q = (y|x)_q = \langle y|q(x)\rangle = \langle q^{-1}(v)|u\rangle ,$$

il est facile d'établir que q^* est un produit scalaire □

Base duale d'une base q-orthonormée. Si $(e_1,\ldots,e_n)$ est une base q-orthonormée de E, $(q(e_1),\ldots,q(e_n))$ est la base de E^*, duale de $(e_1,\ldots,e_n)$, q^*-orthonormée.

□ *Démonstration.* Des égalités :

$$\langle e_j|q(e_j)\rangle = (e_j|e_j)_q = 1$$

$$\langle e_i|q(e_j)\rangle = (e_i|e_j)_q = 0 \quad \text{pour } i \neq j ,$$

on en déduit que $(q(e_1),\ldots,q(e_n))$ est la base duale de $(e_1,\ldots,e_n)$. Cette base est q^*-orthonormée car :

$$(q(e_i)|q(e_j))_{q^*} = \langle q^{-1}q(e_i)|q(e_j)\rangle = \langle e_i|q(e_j)\rangle = (e_i|e_j)_q \quad \square$$

1.2.4. IDENTIFICATION D'UN ESPACE VECTORIEL AVEC SON BIDUAL

Bidual. L'ensemble des formes linéaires sur E^* est appelé bidual de E et noté E^{**}.

Isomorphisme de E sur E^{**}. L'application de E dans E^{**} qui, à tout vecteur x de E, fait correspondre la forme linéaire f(x) sur E^*, définie par : $\langle u | f(x) \rangle_{E^{**}} = \langle x | u \rangle_{E^*}$, est un isomorphisme de E sur E^{**}. On l'appelle isomorphisme canonique de E sur E^{**}, et il permet d'identifier l'espace avec son bidual. Ainsi, on écrira par abus de notation $\langle x | u \rangle_{E^*} = \langle u | x \rangle_{E^{**}}$.

□ *Démonstration.* L'application de $E \times E^*$ dans R qui, au couple (x,u) fait correspondre $\langle x | u \rangle$, est linéaire en x et en u. Il en résulte que l'application f de E dans E^{**} est linéaire, et que f(x) est une forme linéaire sur E^*.

Soient $(e_1,\ldots,e_n)$ une base de E et $(e_1^*,\ldots,e_n^*)$ la base de E^*, duale de $(e_1,\ldots,e_n)$. La relation $\langle e_j^* | f(e_i) \rangle = \langle e_i | e_j^* \rangle$ montre que $(f(e_1),\ldots,f(e_n))$ est la base duale de $(e_1^*,\ldots,e_n^*)$. L'application f, mettant en correspondance bijective les bases de E et de E^{**}, est un isomorphisme □

1.3. ESPACE AFFINE EUCLIDIEN

Espace affine. Un ensemble non vide H est un espace affine, associé à un espace vectoriel E, s'il existe une application de HxE dans H, notée $(M,x) \longrightarrow M + x$, telle que :

1. $\forall M \in H, \ \forall x, \forall y \in E \quad M+0 = M$
$M+(x+y) = (M+x)+y.$

2. Pour deux éléments M_i et M_j quelconques de H, il existe un vecteur x unique de E tel que $M_i+x = M_j$. On désigne ce vecteur par M_iM_j. Les éléments de H sont appelés points.

Dimension de l'espace affine. C'est la dimension de l'espace vectoriel sous-jacent.

Repère affine. On appelle repère affine le couple formé d'un point O de H, appelé origine du repère, et d'une base de E. Dans ce repère, les coordonnées du point M sont les composantes du vecteur OM dans la base.

Espace affine euclidien. Lorsque l'espace vectoriel sous-jacent (E, q) est euclidien, H est appelé espace affine euclidien. On peut alors définir une distance euclidienne entre points.

Distance euclidienne. O étant un point quelconque de H, la distance entre deux points M_i et M_j est définie par :

$$d_{M_iM_j} = \| OM_i - OM_j \|_q ,$$

et ne dépend évidemment pas du point O choisi.

2. APPLICATIONS LINEAIRES ENTRE ESPACES VECTORIELS EUCLIDIENS

Soient (E, q) et (F, r) deux espaces vectoriels euclidiens, de dimensions respectives n et m. $\mathcal{L}$ (E, F) désigne l'espace vectoriel des applications linéaires de E dans F.

2.1. TRANSPOSEE ET ADJOINTE D'UNE APPLICATION LINEAIRE

Transposée. On appelle transposée de l'application linéaire a de $\mathcal{L}$ (E, F), l'application linéaire a' de $\mathcal{L}$ (F^*, E^*), définie de façon unique par :

$$\forall x \in E,\ \forall v \in F^* \qquad \langle x | a'(v) \rangle_{E^*} = \langle a(x) | v \rangle_{F^*} .$$

Propriétés. Soit (G, s) un autre espace vectoriel euclidien. a et b étant deux éléments de $\mathcal{L}$ (E, F) et c un élément de $\mathcal{L}$ (F, G) :

$(a+b)' = a'+b'$ $\qquad (\lambda\ a)' = \lambda\ a'$ $\qquad (ac)' = c'a'$

$(a^{-1})' = (a')^{-1}$ si a est bijective,

$(a')' = a$ après identification des espaces E et F avec leurs biduaux.

□ *Démonstration.* Ces formules s'obtiennent sans difficulté à partir de la définition de la transposée, en utilisant la linéarité de $\langle a(x) | v \rangle$ par rapport à x, v et a □

Transposé du produit scalaire. L'espace vectoriel euclidien (E, q) étant identifié avec son bidual, le produit scalaire q, considéré comme un élément de $\mathcal{L}(E, E^*)$, est égal à son transposé.

□ *Démonstration.* Il faut démontrer que, pour tout vecteur x de E, q(x) et q'(x) désignent la même forme linéaire, c'est-à-dire qu'appliquées à un vecteur quelconque y de E, elles donnent le même résultat.

$$\langle y|q'(x)\rangle_{E^*} = \langle q(y)|x\rangle_{E^{**}} = \langle x|q(y)\rangle_{E^*}$$
$$= (x|y)_q = (y|x)_q = \langle y|q(x)\rangle_{E^*} \quad \square$$

Adjointe. On appelle adjointe de l'application linéaire a de $\mathcal{L}(E, F)$, l'application linéaire a^* de $\mathcal{L}(F, E)$ définie de façon unique par l'égalité des produits scalaires :

$$\forall x \in E, \forall y \in F \qquad (x|a^*(y))_q = (a(x)|y)_r .$$

Ceci entraîne que $a^* = q^{-1}\, a'\, r$.

□ *Démonstration* de l'égalité $a^* = q^{-1}\, a'r$.

$$\forall x \in E, \forall y \in F \qquad (a(x)|y)_r = \langle a(x)|r(y)\rangle_{F^*} = \langle x|a'r(y)\rangle_{E^*}$$
$$(a(x)|y)_r = (x|a^*(y))_q = \langle x|q\ a^*(y)\rangle_{E^*}.$$

On a donc $a'r = q\ a^*$ et $a^* = q^{-1}\, a'r$ □

Propriétés. a et b étant deux éléments de $\mathcal{L}(E, F)$, et c un élément de $\mathcal{L}(F, G)$:

1. Ker(a) est q-orthogonal à $\mathrm{Im}(a^*)$, et $\mathrm{Ker}(a^*)$ est r-orthogonal à Im(a),
2. $(a+b)^* = a^* + b^*$
3. $(\lambda a)^* = \lambda\, a^*$
4. $(a^*)^* = a$
5. $(a^*)' = (a')^*$
6. $(a\ c)^* = c^*\ a^*$.

□ *Démonstration.* 1. découle de l'égalité $(x|a^*(y))_q = (a(x)|y)_r$. La linéarité de la transposition entraîne 2 et 3.

$(a^*)^* = r^{-1}\ (a^*)'q = r^{-1}(q^{-1}a'\ r)'\ q = a$.

$(a^*)' = (q^{-1}\ a'\ r)' = r\ a\ q^{-1} = (a')^*$.

Enfin si s est le produit scalaire sur G,

$(ac)^* = s^{-1}(ac)'q = s^{-1}\ c'a'\ q = s^{-1}\ c'\ r\ r^{-1}\ a'q = c^*\ a^*$ □

2.2. OPERATEURS AUTO-ADJOINTS

2.2.1. DEFINITIONS

L'espace vectoriel $\mathcal{L}(E, E)$ des applications linéaires de E dans lui-même est noté simplement $\mathcal{L}(E)$. Ses éléments sont appelés opérateurs de E.

Opérateur auto-adjoint. Un opérateur de (E, q) est auto-adjoint lorsqu'il est égal à son adjoint. On dit aussi q-symétrique car $a = a^* \iff qa = (qa)' = a'q$.

Opérateur orthogonal. Un opérateur a de (E, q) est orthogonal lorsque $a^{-1} = a^*$.

***q*-projecteur.** Un q-projecteur a est un opérateur auto-adjoint idempotent : $a = a^* = a^2$.

Des propriétés de l'adjoint, il résulte que Ker(a) est q-orthogonal à Im(a), et que E est la somme directe de ces deux sous-espaces. Inversement, à partir d'un sous-espace E_k de E, engendré par la base q-orthonormée $(e_1,\ldots,e_k)$, on définit le q-projecteur a sur E_k par :

$$\forall\, x \in E \qquad a(x) = \sum_{i=1}^{k} (x|e_i)_q \; e_i \; .$$

2.2.2. DECOMPOSITION SPECTRALE D'UN OPERATEUR AUTO-ADJOINT

Eléments propres. Soit a un opérateur de E. Un vecteur x non nul de E est vecteur propre de a, s'il existe un nombre réel λ tel que $a(x) = \lambda x$. λ est appelé valeur propre de a. L'ensemble des vecteurs propres, associés à la même valeur propre, forme un sous-espace vectoriel de E, appelé espace propre.

Propriété. Un opérateur auto-adjoint est diagonalisable, c'est-à-dire qu'il existe une base q-orthonormée de vecteurs propres.

□ *Démonstration.*

Première étape : un opérateur auto-adjoint admet au moins une valeur propre. La recherche des éléments propres d'un opérateur a, revient à trouver un couple $(\lambda , x) \in R \times E$, solution de l'équation $a(x) = \lambda x$. Pour cela, on se place dans une base $(e_1,\ldots,e_n)$ de E. L'équation $a(x) = \lambda x$ qui s'écrit $x_1 a(e_1) + \ldots x_n a(e_n) = \lambda (x_1 e_1 + \ldots + x_n e_n)$, se décompose en un système de n équations à n+1 inconnues $\lambda , x_1,\ldots,x_n$, que l'on résout sous la contrainte $\| x \|_q^2 = 1$: $\forall i \in \{1,\ldots,n\} \quad \sum_{j=1}^{n} x_j \langle a(e_j) | e_i^* \rangle = \lambda x_i$.
R n'étant pas algébriquement clos, ce système n'est pas forcément résoluble. Mais si on se place dans le corps C des nombres complexes, on sait qu'il existe des solutions. Soit $(\lambda = \alpha + i \beta , x = y + iz)$ l'une d'entre elles.

Les relations numériques existant entre les éléments $\langle a(e_j) | e_i^* \rangle$ et les coordonnées de x dans la base $(e_1,\ldots,e_n)$, se décomposent en deux égalités, correspondant à la partie réelle et à la partie imaginaire que l'on écrit de façon abusive : $a(y) = \alpha y - \beta z$ et $a(z) = \alpha z + \beta y$. Faisons alors le produit scalaire de z avec les deux membres de la première égalité, puis celui de y avec les deux membres de la seconde, et soustrayons les deux égalités. a étant auto-adjoint, les produits scalaires se simplifient et on obtient $\beta (\| y \|_q^2 + \| z \|_q^2) = 0$, qui entraîne $\beta = 0$ puisque le vecteur x n'est pas nul. z peut être choisi égal à 0, et l'équation $a(x) = \lambda x$ admet le couple (α , y) pour solution.

Deuxième étape. a étant un opérateur de E, si le sous-espace F de (E, q) est stable pour a, alors le sous-espace $F^{\perp}$, q-orthogonal à F, est stable pour a^*. En effet, pour $x \in F$ et $y \in F^{\perp}$, la relation $(x | a^*(y))_q = (a(x) | y)_q = 0$ entraîne $a^*(y) \in F^{\perp}$.

Troisième étape. On raisonne par récurrence sur la dimension n de E. Pour n=1, tout opérateur de E est diagonalisable. Pour n quelconque, soit x un vecteur propre, q-normé à 1, de l'opérateur auto-adjoint a de E. Soit F le sous-espace q-orthogonal au sous-espace engendré par x. a étant auto-adjoint, est stable sur F et sa restriction a_{n-1} à F est auto-adjointe. D'après l'hypothèse de récurrence, a_{n-1} est diagonalisable. En complétant par x la base de F, formée de vecteurs propres de a_{n-1} (qui sont vecteurs propres de a),

on obtient une base q-orthonormée de E, formée de vecteurs propres de a □

Remarques. Il est facile de démontrer que les espaces propres associés à deux valeurs propres distinctes sont q-orthogonaux.

Si on compte une valeur propre, racine multiple du système d'équations précédent, avec son ordre de multiplicité, le nombre de valeurs propres est égal à n, et le nombre de valeurs propres non nulles est égal au rang de l'opérateur.

Décomposition spectrale. a étant un opérateur auto-adjoint de E, appelons $\lambda_1 \geqslant \ldots \geqslant \lambda_n$ ses valeurs propres ordonnées, qu'elles soient distinctes ou non, et $(e_1,\ldots,e_n)$ une base q-orthonormée de vecteurs propres associés. Alors, ρ étant le rang de a :

$$\forall x \in E \qquad a(x) = \sum_{i=1}^{\rho} \lambda_i (x|e_i)_q \, e_i \ .$$

□ *Démonstration.* Soit x un vecteur quelconque de E. Décomposons-le dans la base $(e_1,\ldots,e_n)$, alors :

$$a(x) = \sum_{i=1}^{n} (x|e_i)_q \, a(e_i) = \sum_{i=1}^{n} (x|e_i)_q \, \lambda_i e_i \ ,$$

qui se réduit aux ρ premiers termes, puisque le nombre de valeurs propres non nulles est égal au rang de l'opérateur □

2.2.3. PROPRIETES DES VALEURS PROPRES

Quotient de Rayleigh. Le quotient de Rayleigh d'un opérateur a de E est l'application de $E - \{0\}$ dans R, qui fait correspondre à un vecteur x non nul de E, le nombre réel

$$R_a(x) = \frac{(a(x)|x)_q}{(x|x)_q} \ .$$

Caractérisation d'une valeur propre. Soient $\lambda_1 \geqslant \lambda_2 \geqslant \ldots \geqslant \lambda_n$ les valeurs propres d'un opérateur a auto-adjoint. E_k désignant le sous-espace engendré par un ensemble de vecteurs propres de a, associés aux k premières valeurs propres,

toute valeur propre λ_k peut être caractérisée par

$$\lambda_k = \sup_{x \perp E_{k-1}} R_a(x) = \inf_{x \in E_k} R_a(x),$$

qui entraîne $\lambda_1 = \sup_{x \in E} R_a(x)$ et $\lambda_n = \inf_{x \in E} R_a(x)$.

□ *Démonstration.* Soit $(e_1,\ldots,e_n)$ une base q-orthonormée de E, formée de vecteurs propres de a, respectivement associés aux valeurs propres $\lambda_1 \geqslant \ldots \geqslant \lambda_n$. En écrivant la décomposition d'un vecteur quelconque x de E dans cette base, on obtient :

$(a(x)|x)_q = (\sum_{i=1}^{n} (x|e_i)_q \, a(e_i) \mid \sum_{j=1}^{n} (x|e_j)_q \, e_j)_q$ qui se simplifie,

puisque la base $(e_1,\ldots,e_n)$ est q-orthonormée et que $a(e_i) = \lambda_i e_i$ pour tout i, en $\sum_i \lambda_i (x|e_i)_q^2$. Dans cette base, le carré de la norme de x est égal à $\sum_{i=1}^{n} (x|e_i)_q^2$.

Pour un vecteur x orthogonal à E_{k-1}, $(a(x)|x)_q$ se réduit à $\sum_{i=k}^{n} \lambda_i (x|e_i)_q^2$, et est majoré par $\lambda_k (\sum_{i=k}^{n} (x|e_i)_q^2) = \lambda_k (x|x)_q$.

Pour un vecteur x de E_k, $(a(x)|x)_q$ réduit à $\sum_{i=1}^{k} \lambda_i (x|e_i)_q^2$, est minoré par $\lambda_k (x|x)_q$. Dans les deux cas, la borne λ_k est atteinte lorsque $x = e_k$ □

Propriété. La somme des valeurs du quotient de Rayleigh, pour k vecteurs $x_1,\ldots,x_k$ q-orthogonaux deux à deux, est toujours inférieure à $\lambda_1 + \ldots + \lambda_k$. L'égalité est réalisée lorsque chaque vecteur x_j est un vecteur propre associé à λ_j.

□ *Démonstration.* Sans perte de généralité, la démonstration peut se limiter au cas de k vecteurs $x_1,\ldots,x_k$ q-orthonormés. Dans la base q-orthonormée $(e_1,\ldots,e_n)$ formée de vecteurs propres, respectivement associés aux valeurs propres $\lambda_1 \geqslant \ldots \geqslant \lambda_n$:

$$(a(x_j)|x_j)_q = \sum_{i=1}^{n} \lambda_i (x_j|e_i)_q^2 \qquad \text{et}$$

$$\sum_{j=1}^{k} (a(x_j)|x_j)_q = \lambda_1 (\sum_j (x_j|e_1)_q^2) + \ldots + \lambda_n (\sum_j (x_j|e_n)_q^2).$$

La somme des coefficients c_i des λ_i, somme des carrés des normes des x_i, est égale à k. En remarquant que $x_1,\ldots,x_k$, q-orthonormés, peuvent être complétés pour former une base q-orthonormée de E, on en déduit que c_i est borné par $(e_i|e_i)_q^2$ égal à 1. Alors, la somme des k valeurs du quotient de Rayleigh est majorée par :

$$\sum_{i=1}^{k-1} \lambda_i c_i + \lambda_k (k - \sum_{i=1}^{k-1} c_i) = \sum_{i=1}^{k-1} (\lambda_i - \lambda_k) c_i + k \lambda_k$$

$$\leqslant \sum_{i=1}^{k-1} (\lambda_i - \lambda_k) + k \lambda_k = \sum_{i=1}^{k} \lambda_i ,$$

et l'égalité est réalisée lorsque les k premiers coefficients sont égaux à 1 et les autres nuls □

2.2.4. TRANSPOSE D'UN OPERATEUR AUTO-ADJOINT

Propriété. Si a est un opérateur auto-adjoint de (E, q), son transposé a' est un opérateur auto-adjoint de (E^*, q^*).

□ *Démonstration.* $a = a^* \iff a' = (a^*)' = (a')^*$ □

Eléments propres du transposé. Si $(e_1,\ldots,e_n)$ est une base q-orthonormée, formée de vecteurs propres de a associés aux valeurs propres $\lambda_1 \geqslant \ldots \geqslant \lambda_n$, la base duale $(e_1^*,\ldots,e_n^*)$ est un système de vecteurs propres de a', associés aux mêmes valeurs propres.

□ *Démonstration.* q étant un isomorphisme de E sur E^*, soient u et v deux formes linéaires quelconques sur E, $x = q^{-1}(u)$ et $y = q^{-1}(v)$ les vecteurs de E correspondants, alors :

$$(a'(u)|u)_{q^*} = (u|a'(u))_{q^*} = \langle q^{-1}(u)|a'(u)\rangle$$

$$= \langle x|a'q(x)\rangle = \langle a(x)|q(x)\rangle = (a(x)|x)_q$$

et $(u|v)_{q^*} = \langle q^{-1}(u)|v\rangle = \langle x|q(y)\rangle = (x|y)_q$.

Il en résulte que, pour tout vecteur x de E, $R_a(x) = R_{a'}(q(x))$ et :

$$\lambda_1(a) = \sup_{x \in E} R_a(x) = \sup_{u \in E^*} R_{a'}(u) = \lambda_1(a') .$$

La borne est atteinte lorsque $x = e_1$ dans E et $u = q(e_1) = e_1^*$ dans E^*.

On reprend ce raisonnement pour les autres valeurs propres, après avoir remarqué que, si x est q-orthogonal au sous-espace E_k de E engendré par $(e_1,\ldots,e_k)$, $u = q(x)$ est q^*-orthogonal au sous-espace E_k^* de E^*, engendré par $(e_1^*,\ldots,e_k^*)$ □

2.2.5. OPERATEURS AUTO-ADJOINTS POSITIFS

Définition. Un opérateur a auto-adjoint de E est positif si, pour tout vecteur x de E : $(a(x)|x)_q \geqslant 0$. Un opérateur auto-adjoint qui a toutes ses valeurs propres positives ou nulles, est positif.

□ *Démonstration.* La caractérisation de la plus petite valeur propre comme borne inférieure du quotient de Rayleigh, établit l'équivalence entre les deux définitions □

Etude des opérateurs a^*a et $a\,a^*$. a étant une application linéaire de E dans F, les opérateurs a^*a de E, et $a\,a^*$ de F, ont les propriétés suivantes :

1. a^*a et $a\,a^*$ sont auto-adjoints et positifs.
2. $\mathrm{Ker}(a^*a) = \mathrm{Ker}(a)$.
3. a^*a et $a\,a^*$ ont les mêmes valeurs propres. A un vecteur propre x q-normé de a^*a, correspond le vecteur propre $\lambda^{-1/2}\, a(x)$, r-normé, de $a\,a^*$.
4. a^*a, $a\,a^*$, a et a^* ont le même rang.

□ *Démonstration.* 1. a^*a est un opérateur auto-adjoint car $(a^*a)^* = a^*a$. Soit x un vecteur quelconque de E. Utilisant la symétrie de q et la définition de l'adjointe :

$$(a^*a(x)|x)_q = (x|a^*a(x))_q = (a(x)|a(x))_r \geqslant 0.$$

On démontrerait de la même façon que $a\,a^*$ est auto-adjoint et positif.

2. Si x est un vecteur de $\mathrm{Ker}(a^*a)$, $a^*a(x) = 0$ et $(a(x)|a(x))_r = (x|a^*a(x))_q = 0$, c'est-à-dire $\| a(x) \|_r = 0$, qui implique $a(x) = 0$. L'implication inverse est immédiate. On aurait de même $\mathrm{Ker}(a\,a^*) = \mathrm{Ker}(a^*)$.

3. Soit λ une valeur propre, et x un vecteur propre q-normé de a^*a, associé à λ. Alors $a^*a(x) = \lambda\, x$ entraîne a $a^*(ax) = \lambda(ax)$. a^*a et a ayant même noyau, a(x) n'est pas nul, c'est donc un vecteur propre de a a^*, associé à la valeur propre λ, mais dont la r-norme est $\lambda^{1/2}$. En effet $(a(x)|\,a(x))_r = (x|\,a^*a(x))_q = \lambda(x|x)_q = \lambda$.

4. a^*a et a a^*, ayant les mêmes valeurs propres, ont le même rang. a^*a et a, ayant le même noyau, ont le même rang. Il en est de même pour a a^* et a^* □

3. PRODUIT SCALAIRE SUR L'ESPACE VECTORIEL DES APPLICATIONS LINEAIRES

3.1. TRACE D'UN OPERATEUR

Définition de la trace. Par rapport à une base $(e_1,\ldots,e_n)$ de E, la trace d'un opérateur a de E est égale à

$$\mathrm{tr}(a) = \sum_{i=1}^{n} \langle a(e_i)|e_i^* \rangle .$$

Propriété. La trace ne dépend pas de la base dans laquelle on exprime l'opérateur. En particulier, la trace d'un opérateur diagonalisable est égale à la somme de ses valeurs propres.

□ *Démonstration.* Soient $(e_1,\ldots,e_n)$ et $(\varepsilon_1,\ldots,\varepsilon_n)$ deux bases quelconques de E.

$$\begin{aligned}
\mathrm{Tr}(a) &= \sum_i \langle a(e_i)|e_i^* \rangle = \sum_i \langle a\Big(\sum_j \langle e_i|\varepsilon_j^* \rangle\, \varepsilon_j\Big)|e_i^* \rangle \\
&= \sum_i \sum_j \langle e_i|\varepsilon_j^* \rangle \langle a(\varepsilon_j)|e_i^* \rangle = \sum_j \Big\langle a(\varepsilon_j)\Big|\sum_i \langle e_i|\varepsilon_j^* \rangle\, e_i^* \Big\rangle \\
&= \sum_j \langle a(\varepsilon_j)|\varepsilon_j^* \rangle, \quad \text{car } \langle e_i|\,\varepsilon_j^* \rangle \text{ est la coordonnée de } \varepsilon_j^* \text{ sur } e_i^* .
\end{aligned}$$

Dans une base q-orthonormée de vecteurs propres,

$$\mathrm{Tr}(a) = \sum_i (a(e_i)\,|\,e_i)_q = \sum_i \lambda_i \quad \square$$

Propriétés. L'application qui, à un opérateur de E, fait correspondre sa trace, est une forme linéaire de $\mathcal{L}(E)$. a étant un opérateur de E, b un élément de $\mathcal{L}(E, F)$, c un élément de $\mathcal{L}(F, E)$: $Tr(a) = Tr(a')$ et $Tr(bc) = Tr(cb)$.

□ *Démonstration.* La linéarité de la trace découle de sa définition.

Soient $(e_1,\ldots,e_n)$ une base de E et $(f_1,\ldots,f_m)$ une base de F : $Tr(a) = \sum_i < a(e_i) \,|\, e_i^* > = \sum_i < e_i | a'(e_i^*) > = Tr(a')$,

car $< e_i | a'(e_i^*) >$ est la coordonnée de $a'(e_i^*)$ sur e_i^*.

$$Tr(cb) = \sum_i < cb(e_i) | e_i^* >_{E^*} = \sum_i < b(e_i) | c'(e_i^*) >_{F^*}$$

$$= \sum_i < b(e_i) \,|\, \sum_j < f_j \,|\, c'(e_i^*) >_{F^*} f_j^* >_{F^*}$$

$$= \sum_j \sum_i < f_j \,|\, c'(e_i^*) >_{F^*} < b(e_i) | f_j^* >_{F^*}$$

$$= \sum_j < b(\sum_i < c(f_j) \,|\, e_i^* >_{E^*} e_i) \,|\, f_j^* >_{F^*}$$

$$= \sum_j < bc(f_j) \,|\, f_j^* >_{F^*} = Tr(bc) \quad \square$$

3.2. PRODUIT SCALAIRE ENTRE APPLICATIONS LINEAIRES

Produit scalaire de Hilbert-Schmidt. L'application qui, à deux éléments a et b de $\mathcal{L}(E, F)$, fait correspondre $Tr(a^*b)$ est un produit scalaire, appelé produit scalaire de Hilbert-Schmidt. Ce produit scalaire munit $\mathcal{L}(E, F)$ de la structure d'espace vectoriel euclidien. $Tr(a^*b)$ sera noté $(a|b)_t$.

□ *Démonstration.*

$$Tr(a^*b) = Tr(q^{-1} a' r b) = Tr(q^{-1} a' r b)' = Tr(b' r a q^{-1})$$

$$= Tr(q^{-1} b' r a) = Tr(b^*a).$$

La trace étant une forme linéaire, $Tr(a^*b)$ est linéaire par rapport à a, et linéaire par rapport à b par symétrie.

a^*a étant auto-adjoint et positif, ses valeurs propres sont positives ou nulles. Le nombre $Tr(a^*a)$, somme des valeurs propres de a^*a, est positif ou nul. Supposons qu'il soit nul, cela signifie que toutes les valeurs propres de a^*a sont nulles. Ce qui entraîne $Ker(a^*a) = E = Ker(a)$, c'est-à-dire $a = 0$ □

L'ensemble des opérateurs auto-adjoints de E est un sous-espace vectoriel euclidien de $\mathcal{L}(E)$, muni du produit scalaire de Hilbert-Schmidt. Parmi ces opérateurs, ceux qui sont positifs forment un sous-ensemble stable pour l'addition et la multiplication par un nombre réel positif. On considère la restriction du produit scalaire de Hilbert-Schmidt à ce sous-ensemble.

□ **Propriété.** Si a et b sont deux opérateurs de E auto-adjoints et positifs, Tr(ab) est positif ou nul.

Démonstration. b étant auto-adjoint et positif, il existe une base q-orthonormée $(e_1,\ldots,e_n)$ de vecteurs propres de b, tels que, pour tout i : $b(e_i) = \lambda_i e_i$, avec $\lambda_i \geqslant 0$. La trace ne dépendant pas de la base choisie, plaçons-nous dans cette base :

$$\mathrm{Tr}(ab) = \sum_i (ab(e_i)|e_i)_q = \sum_i \lambda_i (a(e_i)|e_i)_q \geqslant 0 ,$$

car a étant un opérateur auto-adjoint positif, $(a(e_i)|e_i)_q$ est positif ou nul, pour tout i □

4. VALEURS SINGULIERES D'UNE APPLICATION LINEAIRE

4.1. DECOMPOSITION D'UNE APPLICATION LINEAIRE

Valeurs singulieres. Soit a un élément de $\mathcal{L}(E, F)$. L'opérateur a^*a de E est auto-adjoint et positif. Ses valeurs propres sont positives ou nulles. On peut donc définir les racines carrées des valeurs propres de a^*a, que l'on appelle valeurs singulières de a et que l'on note $\sigma_1(a)$, ..., $\sigma_n(a)$.

Décomposition en valeurs singulières. a est un élément de $\mathcal{L}(E, F)$ de rang ρ. Soient $\sigma_1(a) \geqslant \sigma_2(a) \ldots \geqslant \sigma_\rho(a)$ ses valeurs singulières non nulles, $(e_1,\ldots,e_\rho)$ des vecteurs propres q-orthonormés de a^*a, $(f_1,\ldots,f_\rho)$ des vecteurs propres r-orthonormés de $a\,a^*$, chacun des e_i, f_i étant associé à la valeur propre $\sigma_i^2(a)$. Alors,

$$\forall\, x \in E \qquad a(x) = \sum_{i=1}^{\rho} \sigma_i(a)\, (x|e_i)_q\, f_i .$$

□ *Démonstration.* On utilise les propriétés, déjà étudiées p. 36, de a a^* et a^*a. A savoir, a a^* et a^*a ont les mêmes valeurs propres ; si f_i, r-normé à 1, est vecteur propre de a a^*, alors $a^*(f_i)$ est vecteur propre de a^*a associé à la même valeur propre $\sigma_i^2(a)$, mais de q-norme $\sigma_i(a)$; a a^* et a^* ont le même noyau, a a^* et a ont le même rang.

Soit $(f_1,\ldots,f_m)$ une base r-orthonormée de F, formée de vecteurs propres de a a^* associés aux valeurs propres $\sigma_1^2(a),\ldots,\sigma_m^2(a)$. Pour un vecteur x quelconque de E, la décomposition de a(x) dans cette base s'écrit :

$$a(x) = \sum_{i=1}^{m} (a(x)|f_i)_r \, f_i = \sum_{i=1}^{m} (x|a^*(f_i))_q \, f_i \ ,$$

qui se réduit à $\quad a(x) = \sum_{i=1}^{\rho} (x|a^*(f_i))_q \, f_i \ ,$

car pour i supérieur au rang ρ de a, donc de a a^*, f_i appartient à $\mathrm{Ker}(a\,a^*) = \mathrm{Ker}(a^*)$. Et on obtient le résultat en écrivant $a^*(f_i)$, vecteur propre de a^*a, q-normé à $\sigma_i(a)$, sous la forme $\sigma_i(a)\, e_i$ □

Remarque. Si a est un opérateur auto-adjoint et positif, sa décomposition en valeurs singulières n'est autre que sa décomposition spectrale.

4.2. APPROXIMATION D'UNE APPLICATION LINEAIRE

Propriété. Soient a une application linéaire de $\mathcal{L}(E, F)$ de rang ρ, et b une application linéaire de $\mathcal{L}(E, F)$ de rang h, inférieur ou égal à ρ .

Alors les valeurs singulières de a et de a-b, rangées par ordre décroissant, vérifient :

$$\sigma_i(a-b) \geqslant \sigma_{i+h}(a) \qquad \text{pour } i+h \leqslant \rho \ ,$$

$$\sigma_i(a-b) \geqslant 0 \qquad \text{pour } i+h > \rho \ .$$

Les égalités sont atteintes lorsque b est l'application a_h, définie par la somme des h premiers éléments de la décomposition en valeurs singulières de a :

$$\forall\, x \in E \qquad a_h(x) = \sum_{i=1}^{h} \sigma_i(a)\, (x|e_i)_q \, f_i \ .$$

Bien que la définition de a_h dépende du choix des vecteurs propres, a_h est unique sauf lorsque $\sigma_h(a) = \sigma_{h+1}(a)$.

□ *Démonstration.*

Notation : $E_i(c)$ désigne le sous-espace de dimension i, engendré par les vecteurs propres associés aux i plus grandes valeurs propres $\lambda_1(c) \geqslant \ldots \geqslant \lambda_i(c)$ de l'opérateur c.

Soit p le q-projecteur sur Ker(b). Nous allons démontrer que, pour $i \leqslant \rho - h$:

$$\sigma_i^2(a-b) = \lambda_i((a-b)^*(a-b)) \geqslant \lambda_i(p(a-b)^*(a-b)\ p)$$
$$= \lambda_i(pa^*ap) \geqslant \lambda_{i+h}(a^*a) = \sigma_{i+h}^2(a).$$

Tout d'abord, Im(p) étant égal à Ker(b), $bp = 0$ et, puisque p est auto-adjoint, $pb^* = 0$. Il en résulte que $p(a-b)^*(a-b)\ p$ se réduit à pa^*ap.

Ensuite, pour établir ces inégalités, on utilise la caractérisation des valeurs propres énoncée p. 33 , et les deux résultats suivants. D'une part, F et G étant deux sous-espaces de E, pour vérifier que leur intersection ne se réduit pas au vecteur nul, on calcule la dimension de $F \cap G$ par la formule :
$\dim(F \cup G) = \dim F + \dim G - \dim(F \cap G)$, et en majorant $\dim(F \cup G)$ par n, on minore $\dim(F \cap G)$ par $\dim F + \dim G - n$. D'autre part, p étant auto-adjoint, pour un opérateur c quelconque de E et un vecteur x de Im(p) :

$$(c(x)\,|\,x)_q = (cp(x)\,|\,p(x))_q = (pcp(x)\,|\,x)_q\ .$$

Alors, après avoir vérifié que l'intersection des sous-espaces utilisés ne se réduit pas au vecteur nul, on peut écrire :

$$\lambda_{i+h}(a^*a) = \inf_{x \in E_{i+h}(a^*a)} R_{a^*a}(x) \leqslant \inf_{x \in E_{i+h}(a^*a) \cap \mathrm{Im}(p)} R_{a^*a}(x)$$

$$\leqslant \inf_{\substack{x \in E_{i+h}(a^*a) \cap \mathrm{Im}(p) \\ x \perp E_{i-1}(pa^*ap)}} R_{pa^*ap}(x) \qquad \text{puisque } R_{a^*a}(x) = R_{pa^*ap}(x), \text{ pour } x \in \mathrm{Im}(p),$$

$$\leqslant \sup_{x \perp E_{i-1}(pa^*ap)} R_{pa^*ap}(x) = \lambda_i(pa^*ap)\ .$$

Le rang de a^*a étant égal au rang de a, il découle de cette inégalité que, pour $i+h \leqslant \rho$, les valeurs propres $\lambda_i(pa^*ap)$ sont différentes de zéro. Le rang de $pa^*ap = p(a-b)^*(a-b)p$ est donc supérieur ou égal à $\rho-h$. Cela entraîne que, pour $i \leqslant \rho-h$:

$$E_i(p(a-b)^*(a-b)p) \subset \mathrm{Im}(p(a-b)^*(a-b)p) \subset \mathrm{Im}(p).$$

Alors, si on note F l'intersection des sous-espaces $E_{i-1}^{\perp}((a-b)^*(a-b))$ et $E_i(p(a-b)^*(a-b)p)$, F est inclus dans Im(p) et :

$$\lambda_i((a-b)^*(a-b)) = \sup_{x \perp E_{i-1}((a-b)^*(a-b))} R_{(a-b)^*(a-b)}(x)$$

$$\geqslant \sup_{x \in F} R_{p(a-b)^*(a-b)p}(x) \geqslant \lambda_i(p(a-b)^*(a-b)p) .$$

Il y a égalité lorsque $b = a_h$. En effet, $\mathrm{Im}(a-a_h)$ est le sous-espace de F engendré par $f_{h+1},\ldots,f_\rho$. On a donc $\mathrm{rg}((a-a_h)^*(a-a_h)) = \mathrm{rg}(a-a_h) = \rho - h$ et $\sigma_i(a-a_h) = 0$ pour $i > \rho-h$. D'autre part, pour $i \leqslant \rho-h$, $\sigma_i(a-a_h) = \sigma_{i+h}(a)$. Ce dernier résultat s'établit en démontrant que pour j appartenant à $\{h+1,\ldots,\rho\}$, $\sigma_j(a)$ est valeur singulière de $(a-a_h)$:

$$(a-a_h)(e_j) = \sum_{k=h+1}^{\rho} \sigma_k(a)(e_j|e_k)_q \; f_k = \sigma_j(a)\, f_j ,$$

$$\text{et } (a-a_h)^*(a-a_h)(e_j) = \sum_{k=1}^{n} ((a-a_h)^*(a-a_h)(e_j)|e_k)_q \; e_k$$

$$= \sum_k ((a-a_h)(e_j)|(a-a_h)(e_k))_r e_k = \sigma_j^2(a)\, e_j .$$

Pour établir l'unicité de a_h, remarquons que si $\sigma_h(a)$ est différent de $\sigma_{h+1}(a)$, le sous-espace de F engendré par $f_1,\ldots,f_h$ est une réunion de sous-espaces propres de $a\,a^*$. Si on appelle p le q-projecteur sur ce sous-espace, alors $a_h = pa$ est défini de manière unique. Lorsque $\sigma_h(a)$ est égal à $\sigma_{h+1}(a)$, le sous-espace propre correspondant n'est pas pris dans sa totalité, et a_h dépend du choix des vecteurs e_h et f_h.

RAO (1980) établit la réciproque, à savoir que parmi les applications linéaires de rang h, seule a_h vérifie ces égalités □

Approximations. Pour la norme de Hilbert-Schmidt, a_h est la meilleure approximation de rang h de l'application linéaire a, puisqu'elle réalise :

$$\min_{\substack{b \in \mathcal{L}(E, F) \\ \text{rang } b \leqslant h}} \| a-b \|_t^2 = \| a-a_h \|_t^2 .$$

Ce minimum est égal à :

$$\| a-a_h \|_t^2 = \mathrm{Tr}((a-a_h)^*(a-a_h)) = \sum_{i=1}^{\rho} \sigma_i^2 (a-a_h) = \sum_{i=1}^{\rho-h} \sigma_i^2 (a-a_h) = \sum_{i=h+1}^{\rho} \sigma_i^2 (a).$$

RAO (1980) énonce ce résultat pour toute norme unitairement invariante, et SABATIER (1987) propose un théorème plus général, caractérisant les fonctionnelles f des valeurs singulières qui réalisent :

$$\min_{\substack{b \in \mathcal{L}(E, F) \\ \text{rang } b \leqslant h}} f(a-b) = f(a-a_h) .$$

5. SEMI-PRODUIT SCALAIRE SUR UN ESPACE VECTORIEL

5.1. ESPACE VECTORIEL SEMI-NORME

Semi-produit scalaire. On appelle semi-produit scalaire sur l'espace vectoriel réel E, une forme bilinéaire, symétrique, semi-définie positive *s*. C'est-à-dire une application de ExE dans R qui vérifie les propriétés de symétrie, bilinéarité et positivité, données dans la définition du produit scalaire, mais si x = 0 entraîne évidemment $(x|x)_s = 0$, la réciproque n'est pas vraie.

Espace vectoriel semi-normé. Le couple (E, *s*) formé d'un espace vectoriel réel de dimension finie E, et d'un semi-produit scalaire *s* sur E, est appelé espace vectoriel semi-normé.

Semi-distance. Le nombre $\| x \|_s = (x|x)_s^{1/2}$ est appelé semi-norme de x, et l'application de ExE dans R qui, à deux vecteurs x et y, fait correspondre $d_{xy} = \| x-y \|_s$ est une semi-distance : elle est symétrique, positive et vérifie l'inégalité triangulaire.

Mais deux vecteurs distincts x et y de E peuvent avoir une semi-distance nulle. Ils sont alors également éloignés de tout autre vecteur z. Par contre, le théorème de Pythagore et l'égalité du parallélogramme sont encore vrais.

Application linéaire associée au semi-produit scalaire. Au semi-produit scalaire s, on associe l'application linéaire s de E dans E^*, définie par :

$$\forall x,\ \forall y \in E \qquad < x | s(y) > = (x|y)_s \ .$$

5.2. PROPRIETES D'UN SEMI-PRODUIT SCALAIRE SUR UN ESPACE VECTORIEL EUCLIDIEN

Soit s un semi-produit scalaire sur un espace euclidien (E, q).

Propriété. sq^{-1} est un opérateur auto-adjoint et positif de (E^*, q^*), dont le rang est égal au rang de s.

□ *Démonstration.* La symétrie de s entraîne que sq^{-1} est auto-adjoint et que $rg(sq^{-1}) = rg(s)$. En effet, si on identifie E avec son bidual, l'application linéaire s est égale à sa transposée puisque la démonstration, établie pour un produit scalaire p. 30, n'utilise pas la bijectivité de l'application linéaire associée au produit scalaire. Le produit scalaire q^{-1} sur E^* étant égal à son transposé, cela entraîne que $(q^{-1}sq^{-1})' = q^{-1}(sq^{-1})$. sq^{-1} est donc auto-adjoint. s^* étant égal à $q^{-1}sq^{-1}$, $rg(sq^{-1}) = rg(sq^{-1})(sq^{-1})^* = rg(ss^*) = rg(s)$.

La positivité de s entraîne que sq^{-1} est positif. En effet, soit u une forme linéaire quelconque de E^* :

$$(sq^{-1}(u)|u)_{q^*} = (u \mid sq^{-1}(u))_{q^*} = < q^{-1}(u) | sq^{-1}(u) >_{E^*}$$
$$= (q^{-1}(u)| q^{-1}(u))_s \geqslant 0 \quad \square$$

Décomposition en valeurs singulières de s. Appelons ρ le rang de s. La décomposition en valeurs singulières de s s'exprime en fonction des valeurs propres non nulles $\lambda_1(sq^{-1}) \geqslant \ldots \geqslant \lambda_\rho(sq^{-1})$, et des vecteurs propres associés $\varepsilon_1^*, \ldots, \varepsilon_\rho^*$, q^*-orthonormés, de l'opérateur sq^{-1} :

$$\forall\, x \in E \qquad s(x) = \sum_{i=1}^{\rho} \lambda_i(sq^{-1}) \langle x|\varepsilon_i^*\rangle \varepsilon_i^* .$$

□ *Démonstration.* Rappelons que la décomposition en valeurs singulières de s s'exprime en fonction des vecteurs propres de ss^* et de s^*s. Or s^* étant égal à $q^{-1}s\,q^{-1}$, les vecteurs propres de ss^* et de s^*s se déduisent des vecteurs propres $\varepsilon_1^*, \ldots, \varepsilon_\rho^*$ de sq^{-1}, de la manière suivante.

En multipliant les deux membres de l'égalité $sq^{-1}(\varepsilon_i^*) = \lambda_i(sq^{-1})\,\varepsilon_i^*$, par sq^{-1}, on obtient :

$$ss^*(\varepsilon_i^*) = \lambda_i(sq^{-1})\; sq^{-1}(\varepsilon_i^*) = \lambda_i^2\,(sq^{-1})\,\varepsilon_i^* .$$

ε_i^* est donc vecteur propre de ss^*, associé à $\lambda_i^2\,(sq^{-1})$. sq^{-1} étant positif, ses valeurs propres sont positives. On sait d'autre part que $rg(sq^{-1}) = rg(s) = rg(ss^*)$. Il en résulte que $\sigma_i(s) = \lambda_i(sq^{-1})$, et que $(\varepsilon_1^*, \ldots, \varepsilon_\rho^*)$ forme un ensemble de vecteurs propres de ss^*, associés aux valeurs propres non nulles.

Par un raisonnement analogue, on établit que les vecteurs propres $\varepsilon_1 = q^{-1}(\varepsilon_1^*), \ldots, \varepsilon_\rho = q^{-1}(\varepsilon_\rho^*)$ du transposé $q^{-1}s$ de sq^{-1} sont vecteurs propres de s^*s.

La décomposition en valeurs singulières de s s'écrit donc :

$$s(x) = \sum_{i=1}^{\rho} \lambda_i(sq^{-1})(x|\varepsilon_i)_q\, \varepsilon_i^* ,$$

avec $(x|\varepsilon_i)_q = \langle x|q(\varepsilon_i)\rangle = \langle x|\varepsilon_i^*\rangle$ □

Relation entre *s* et le produit scalaire *q* sur E. Soit b un opérateur de (E, q). La forme bilinéaire s définie par :

$$\forall x, \forall y \in E \qquad (x|y)_s = (b(x)|b(y))_q \qquad (1)$$

est un semi-produit scalaire sur E. s est un produit scalaire si et seulement si b est injectif. $s = q$ lorsque b est un opérateur orthogonal, c'est-à-dire lorsque $b^{-1} = b^*$.

Réciproquement, si s est un semi-produit scalaire sur E, il existe un opérateur b de (E, q), dont le rang est égal au rang de s, qui vérifie (1). Cet opérateur n'est pas unique, il est défini à la multiplication par un opérateur orthogonal près.

□ *Démonstration.* La forme bilinéaire s définie par (1) à partir de l'opérateur b est, par définition, symétrique et positive. Si b est injective, Ker(b) se réduit à 0 :

$$(x|x)_s = 0 \iff b(x) = 0 \iff x = 0 ,$$

et s est un produit scalaire. La condition (1) s'écrit encore :

$$\forall x, \forall y \in E \qquad \langle x|s(y)\rangle = \langle b(x)|qb(y)\rangle = \langle x|b'qb(y)\rangle ,$$

c'est-à-dire $s = b'q\,b$. Il en résulte que :

$$b^* = q^{-1}\,b'\,q = b^{-1} \implies s = q .$$

Pour établir la réciproque, on considère une base $(\varepsilon_1, \ldots, \varepsilon_n)$ q-orthonormée, de vecteurs propres de s^*s. L'opérateur b de (E, q) défini par :

$$b(\varepsilon_i) = \sigma_i^{1/2}(s)\,\varepsilon_i \qquad \text{pour } i \leqslant \rho ,$$

$$b(\varepsilon_i) = 0 \qquad \text{pour } i > \rho ,$$

vérifie la condition (1). En effet, s et s^*s ayant le même noyau, $s(\varepsilon_i) = 0$ pour $i > \rho$. De la décomposition en valeurs singulières de s, on déduit que $s(\varepsilon_i) = \sigma_i(s)\varepsilon_i^*$ pour $i \leqslant \rho$. Il en résulte que :

$$\forall i, \forall j \in \{1,\ldots,n\} \qquad (\varepsilon_i|\varepsilon_j)_s = (b(\varepsilon_i)|b(\varepsilon_j))_q = \sigma_i(s) \quad \text{pour } i = j,$$

$$= 0 \quad \text{pour } i \neq j.$$

La condition (1) entraîne que $s = b'qb$, et que

$$rg(s) = rg(sq^{-1}) = rg(b'(b')^*) = rg(b') = rg(b).$$

a étant un opérateur orthogonal quelconque de (E, q), l'opérateur ab vérifie lui aussi la condition (1) puisque :

$$(ab(x)|ab(y))_q = (b(x)|a^*ab(y))_q = (b(x)|b(y))_q = (x|y)_s \quad \square$$

III. Analyse en composantes principales non centrée d'un nuage de points pesants

1. OBJECTIF DE L'ANALYSE EN COMPOSANTES PRINCIPALES NON CENTREE

Objectif : L'appellation "analyse en composantes principales" évoque habituellement une méthode permettant de traiter un tableau de mesures X, contenant les valeurs de m variables quantitatives observées sur n individus.

Nous sommes confrontés à des données différentes, que nous allons traiter par une analyse en composantes principales particulière. A partir des tableaux de mesures X_k, la méthode STATIS calcule des produits scalaires : produits scalaires entre les objets $X_k X'_k$ dans l'interstructure, puis produits scalaires "compromis" entre individus. Dans le cas des objets $X_k X'_k$, l'espace vectoriel dans lequel ces produits scalaires ont été calculés ne nous intéresse pas. Et dans le cas du compromis, on a en réalité pour seule donnée, une forme bilinéaire, symétrique, positive, sur un espace vectoriel euclidien.

On se pose donc le problème suivant. Disposant uniquement d'un tableau de produits scalaires entre objets, peut-on représenter ces objets par des points d'un espace affine euclidien de faible dimension, de telle sorte que les produits scalaires soient le plus fidèlement restitués ?

Méthode proposée. Un produit scalaire entre deux objets est, en quelque sorte, une mesure d'association entre ces objets. Néanmoins, le tableau de produits scalaires ne peut pas être considéré comme un tableau de similarités, sauf dans le cas de produits scalaires normés (corrélations entre variables par exemple), car les éléments de la diagonale ne sont pas égaux.

L'étude des distances déduites de ces produits scalaires par "l'analyse factorielle d'un tableau de distances" ne résout pas non plus le problème. En effet, dans la mesure où on s'intéresse à la restitution des produits scalaires, centrer la représentation des objets (lorsqu'ils sont affectés d'un poids) sur leur centre de gravité ne se justifie pas.

Nous sommes donc amenés à présenter l'analyse en composantes principales non centrée d'un nuage de points. Nous adopterons le point de vue du positionnement multidimensionnel, tel qu'il est décrit simplement par KRUSKAL-WISH (1978). En particulier, lorsqu'on applique cette méthode au tableau XX' des produits scalaires entre individus, on retrouve les résultats classiques de l'analyse en composantes principales, avec une optique différente. On ne peut plus, en effet, considérer que les individus et les variables jouent des rôles symétriques. En fait, une variable n'interviendra que par sa corrélation avec un axe factoriel, pour expliquer la position des individus le long d'un axe.

Présentation algébrique des calculs. Contrairement à la plupart des auteurs qui présentent les calculs nécessaires à l'Analyse des Données sous forme matricielle, nous préférons l'utilisation des applications linéaires entre espaces vectoriels euclidiens. Nous disposons ainsi d'un outil algébrique plus précis, encore valable dans un espace de dimension infinie.

En Analyse des Données, on travaille dans des espaces vectoriels E et F de dimension finie. Après avoir fait le choix d'une base dans E et d'une base dans F, il y a correspondance bijective entre les applications linéaires et leurs matrices dans ces bases. Les tableaux de données étant considérés comme des matrices, le passage d'une application linéaire à sa matrice sera quelquefois indispensable.

Précisons, dès maintenant les notations utilisées dans la suite du texte. Soit a une application linéaire de E dans F. On note par la lettre majuscule A, sa matrice dans les bases $(e_1,\ldots,e_n)$ de E et $(f_1,\ldots,f_m)$ de F. L'élément A_{ij}, correspondant à l'intersection de la ligne i et de la colonne j de la matrice A, est égal à la coordonnée de $a(e_j)$ sur f_i. On l'écrira sous la forme $< a(e_j) | f_i^* >$, ou $(a(e_j)|f_i)_r$ si $(f_1,\ldots,f_m)$ est une base r-orthonormée de l'espace vectoriel euclidien (F, r).

2. DEFINITION D'UNE IMAGE EUCLIDIENNE

La représentation des objets, par des points d'un espace affine euclidien, fait appel à la notion d'image euclidienne que nous allons préciser. Si on peut définir la distance entre points d'un espace affine euclidien (p. 29), la notion de produit scalaire est propre aux espaces vectoriels, et ne peut être transposée qu'à des couples de points de même origine. A la définition classique d'image euclidienne, associée aux distances, nous rajoutons la définition d'une image euclidienne, associée aux produits scalaires.

Image euclidienne des objets, associée aux distances. Supposons connues les distances mutuelles d_{ij} entre n objets. Soient H un espace affine euclidien, et (F, r) l'espace vectoriel euclidien sous-jacent.

On appelle image euclidienne des objets, associée aux distances, un nuage composé de n points $M_1,\ldots,M_n$ de H qui vérifient :

$$\forall\ i,\ \forall\ j\ \in\ \{1,\ldots,n\} \qquad d_{M_iM_j} = d_{ij}\ .$$

Remarque. Les objets sont associés aux points de l'espace affine.

Image euclidienne des objets, associée aux produits scalaires. Soit S un tableau symétrique de dimension n, dont l'élément S_{ij} est un produit scalaire entre les objets i et j. Soit H un espace affine euclidien, d'espace vectoriel sous-jacent (F, r).

On appelle image euclidienne des objets, associée aux produits scalaires, un nuage composé des n points $M_1,\ldots,M_n$ et d'un point O de H, qui restitue les produits scalaires S_{ij} de la manière suivante :

$$\forall\ i,\ \forall\ j\ \in\ \{1,\ldots,n\} \qquad (OM_i \mid OM_j)_r = S_{ij}\ .$$

Remarques.

1. Les objets sont associés aux points M_i de l'espace affine, et aux vecteurs OM_i de l'espace vectoriel.

2. Les points $M_1,\ldots,M_n$ forment une image euclidienne des objets, associée aux distances déduites des produits scalaires : $d_{ij} = (S_{ii} + S_{jj} - 2\,S_{ij})^{1/2}$.

Points pesants. Supposons que les objets soient affectés des poids $p_1,\ldots,p_n$, strictement positifs, qui vérifient $p_1+\ldots+p_n = 1$. Les points de l'espace affine sont alors affectés des mêmes poids.

Equivalence entre images euclidiennes des objets, associées aux produits scalaires. Il résulte de la définition qu'il existe une infinité d'images. En effet, les choix de l'espace affine et du point O dans cet espace étant faits, l'image n'est déterminée qu'à une isométrie près.

Deux images O, $M_1,\ldots,M_n$ et O', $M'_1,\ldots, M'_n$ (pouvant appartenir à deux espaces affines différents, dont les espaces vectoriels sous-jacents sont munis des produits scalaires r et r') sont équivalentes si elles restituent les mêmes produits scalaires :

$$\forall\, i,\ \forall\, j \in \{1,\ldots,n\} \qquad (OM_i \mid OM_j)_r = (O'M'_i \mid O'M'_j)_{r'} .$$

Construction d'une image euclidienne des objets, associée aux produits scalaires. En pratique, on choisit le point O dans H, et une base $\mathcal{B}$ r-orthonormée de F, pour définir un repère affine $(O, \mathcal{B})$ dont l'origine est le point O. Les points $M_1,\ldots,M_n$ sont définis par leurs coordonnées dans ce repère.

Tracé du nuage. Lorsque la dimension de l'espace vectoriel F est égale à deux, il est possible de dessiner le nuage de points.

On représente, par convention, les vecteurs de la base $\mathcal{B}$ r-orthonormée par des vecteurs de même longueur, formant un angle droit. Dans ce système d'axes orthonormés, un point M_i est repéré par son abscisse x_i et son ordonnée y_i, qui sont les coordonnées du vecteur OM_i dans la base $\mathcal{B}$. La base $\mathcal{B}$ étant r-orthonormée, le produit scalaire $(OM_i \mid OM_j)_r$ a l'expression classique $x_ix_j + y_iy_j$, et la distance $d_{M_iM_j}$ est égale à $((x_i-x_j)^2 + (y_i-y_j)^2)^{1/2}$.

Nécessité d'une image approchée. La dimension minimale de l'espace vectoriel F, nécessaire pour créer une image euclidienne, sera étudiée ultérieurement. Toutefois, dès que cette dimension est supérieure à deux, la représentation visuelle du nuage de points n'est plus possible, et on cherche une image approchée dans un espace de dimension plus petite.

3. RECHERCHE D'UNE IMAGE EUCLIDIENNE DES OBJETS, ASSOCIEE AUX PRODUITS SCALAIRES

3.1. FORME BILINEAIRE, SYMETRIQUE, POSITIVE, ASSOCIEE AU TABLEAU DES PRODUITS SCALAIRES

Propriété. Soit E un espace vectoriel dont la dimension est égale au nombre n d'objets. On choisit une base $(e_1,\ldots,e_n)$ dans E.

A partir du tableau S des produits scalaires entre objets, on définit la forme bilinéaire s par ses valeurs sur les vecteurs de base :

$$\forall\, i,\ \forall\, j \in \{1,\ldots,n\} \qquad (e_i|e_j)_s = S_{ij} \ .$$

s , symétrique et positive, est un semi-produit scalaire sur E.

□ *Démonstration.* La forme bilinéaire s est symétrique par construction. Montrons qu'elle est positive. Le fait d'avoir calculé des produits scalaires entre objets suppose que ces objets sont des vecteurs x^i d'un espace vectoriel euclidien (G, q). Ainsi $S_{ij} = (x^i|x^j)_q$. Soit x un vecteur de E, alors :

$$(x|x)_s = \Big(\sum_i \langle x|e_i^*\rangle\, e_i \Big|\, \sum_j \langle x|e_j^*\rangle\, e_j\Big)_s = \sum_i \langle x|e_i^*\rangle \langle x|e_j^*\rangle\, (e_i|e_j)_s$$

$$= \sum_i \langle x|e_i^*\rangle \langle x|e_j^*\rangle\, (x^i|x^j)_q = \Big(\sum_i \langle x|e_i^*\rangle\, x^i \Big|\, \sum_j \langle x|e_j^*\rangle\, x^j\Big)_q \ ,$$

égal au carré de la q-norme du vecteur $\sum_i \langle x|e_i^*\rangle\, x^i$, est positif ou nul. En général, s n'est pas définie, sauf si les objets x^i forment une base de G □

Structure euclidienne sur E, définie par les poids des objets. Supposons que les objets soient affectés des poids $p_1,\ldots,p_n$, strictement positifs. Il est alors logique d'imposer que, dans l'image euclidienne approchée, le produit scalaire S_{ij} entre les objets i et j soit d'autant mieux représenté que les poids p_i et p_j sont forts. La mesure de la qualité de l'approximation (p. 57) s'exprimant en fonction du produit scalaire sur l'espace dual E^*, il faut munir E^* du produit scalaire d, défini à partir des poids des objets de la manière suivante.

Définition du produit scalaire d. La forme bilinéaire d, définie par ses valeurs sur la base $(e_1^*,\ldots,e_n^*)$, duale de la base $(e_1,\ldots,e_n)$:

$$\forall\, i \in \{1,\ldots,n\} \qquad (e_i^* \mid e_i^*)_d = p_i$$

$$\forall\, i,\ \forall\, j \in \{1,\ldots,n\},\ i \neq j \qquad (e_i^* \mid e_j^*)_d = 0\ ,$$

est un produit scalaire sur E^*.

□ *Démonstration.* d est symétrique par construction. Soit u un élément quelconque de E^* :

$$(u|u)_d = (\sum_i < e_i|u > e_i^* \mid \sum_j < e_j|u > e_j^*)_d = \sum_i < e_i|u >^2 p_i$$

est positif, et ne s'annule que lorsque chacune des coordonnées $< e_i|u >$ de u dans la base $(e_1^*,\ldots,e_n^*)$ est nulle, c'est-à-dire lorsque $u = 0$ □

Remarque. La base $(e_1^*,\ldots,e_n^*)$ est d-orthogonale mais non d-normée. Au produit scalaire d, on associe l'application linéaire d de E^* dans E (après avoir identifié E avec son bidual). Alors :

$$d(e_i^*) = \sum_j < d(e_i^*)|e_j^* > e_j = \sum_j (e_i^* \mid e_j^*)_d\ e_j = p_i\ e_i\ .$$

3.2. IMAGE EUCLIDIENNE OBTENUE A PARTIR D'UNE FORME BILINEAIRE, SYMETRIQUE, POSITIVE, SUR UN ESPACE VECTORIEL EUCLIDIEN

Soit s l'application linéaire associée à la forme bilinéaire s sur E. L'espace vectoriel E, après avoir été identifié à son bidual, est muni du produit scalaire d^*, dual du produit scalaire d.

La *structure euclidienne* de E permet de décomposer s en valeurs singulières, et d'en déduire une image euclidienne des objets, associée aux produits scalaires du tableau S.

Décomposition en valeurs singulières de s. Les propriétés d'une forme bilinéaire s, symétrique, positive, sur un espace vectoriel euclidien, ont été étudiées p. 44. En particulier, la décomposition en valeurs singulières de l'application linéaire s s'exprime en fonction des valeurs propres et vecteurs propres de l'opérateur auto-adjoint et positif sd de E^*.

ρ désigne le rang de s. $\varepsilon_1^*,\ldots,\varepsilon_\rho^*$ sont des vecteurs propres d-orthonormés, associés aux valeurs propres non nulles de sd :

$$\forall\ x \in E \qquad s(x) = \sum_{\ell=1}^{\rho} \lambda_\ell\,(sd) < x | \varepsilon_\ell^* > \varepsilon_\ell^* \ .$$

Construction d'une image euclidienne. Soit (F, r) un espace vectoriel euclidien de dimension ρ, égale au rang de s, dans lequel on choisit une base $(f_1,\ldots,f_\rho)$ r-orthonormée. Soit H un espace affine euclidien associé à (F, r), dans lequel on choisit un point O.

A partir de la décomposition en valeurs singulières de s, on construit l'image euclidienne suivante. Soient $M_1,\ldots,M_n$ les points de H dont les coordonnées dans le repère $(O, (f_1,\ldots,f_\rho))$ sont égales à :

$$\forall\ \ell \in \{1,\ldots,\rho\} \qquad (OM_i | f_\ell)_r = \lambda_\ell^{1/2} < e_i | \varepsilon_\ell^* > = \lambda_\ell^{-1/2} (s(e_i) | \varepsilon_\ell^*)_d \ .$$

Ces points forment, avec le point O, une image euclidienne des objets, associée aux produits scalaires du tableau S.

□ *Démonstration.* Puisque sur la base $(\varepsilon_1^*,\ldots,\varepsilon_n^*)$ d-orthonormée de E^*, formée des vecteurs propres de sd associés aux valeurs propres $\lambda_1 \geqslant \ldots \geqslant \lambda_n$, $s(e_i) = \sum_{\ell=1}^{\rho} \lambda_\ell < e_i | \varepsilon_\ell^* > \varepsilon_\ell^*$, on en conclut que Im(s) est inclus dans le sous-espace de E^*, engendré par $(\varepsilon_1^*,\ldots,\varepsilon_\rho^*)$, et que la coordonnée de $s(e_i)$ sur ε_ℓ^*, pour $\ell \leqslant \rho$, est égale à :

$$(s(e_i) | \varepsilon_\ell^*)_d = \lambda_\ell < e_i | \varepsilon_\ell^* > \ .$$

L'image euclidienne ainsi construite restitue les produits scalaires du tableau S, car pour i et j appartenant à $\{1,\ldots,n\}$:

$$S_{ij} = (e_i|e_j)_s = \langle e_i|s(e_j)\rangle = \langle e_i \mid \sum_{\ell=1}^{\rho} \lambda_\ell \langle e_j|\varepsilon_\ell^*\rangle \varepsilon_\ell^*\rangle$$

$$= \sum_{\ell=1}^{\rho} \lambda_\ell \langle e_j|\varepsilon_\ell^*\rangle \langle e_i|\varepsilon_\ell^*\rangle ,$$

qui est l'expression de $(OM_i|OM_j)_r$ dans la base r-orthonormée $(f_1,\ldots,f_\rho)$ □

3.3. REPRESENTATION D'OBJETS SUPPLEMENTAIRES DANS CETTE IMAGE EUCLIDIENNE

Définition d'un objet "supplémentaire". Un objet supplémentaire est un objet auquel on a affecté un poids nul.

On connaît les produits scalaires de cet objet avec les n objets déjà étudiés, que l'on qualifie par opposition "d'actifs". Mais on ne l'a pas pris en compte dans la construction de l'image euclidienne précédente. Pour déterminer malgré cela sa position dans cette image euclidienne, on doit modifier la formalisation du problème.

Forme bilinéaire associée au tableau de produits scalaires. Supposons qu'on ait n_1 objets supplémentaires, en plus des n objets actifs. Le tableau S_1 des produits scalaires entre objets est maintenant de dimension $n + n_1$. On se place alors dans un espace vectoriel E_1 de dimension $n + n_1$, dans lequel on choisit une base $(e_1,\ldots,e_n,e_{n+1},\ldots,e_{n+n_1})$. Le tableau S_1 définit un semi-produit scalaire s_1 sur E_1 :

$$\forall\, i,\ \forall\, j \in \{1,\ldots,n+n_1\} \qquad (e_i|e_j)_{s_1} = (S_1)_{ij} \quad .$$

La restriction de s_1 au sous-espace E de E_1, engendré par $(e_1,\ldots,e_n)$, est égale à s. Ce qui entraîne, en particulier, que Ker(s) est inclus dans $\mathrm{Ker}(s_1)$.

Semi-produit scalaire associé au poids des objets. Sur l'espace dual E_1^*, on définit le semi-produit scalaire d_1 :

$$\forall\, i,\ \forall\, j \in \{1,\ldots,n+n_1\}\ ,\ i \neq j \qquad (e_i^* \mid e_j^*)_{d_1} = 0 \quad ,$$

$$(e_i^* \mid e_i^*)_{d_1} = p_i \quad \text{pour } i \leqslant n \quad , \quad (e_i^* \mid e_i^*)_{d_1} = 0 \quad \text{pour } i > n \ .$$

La restriction de d_1 au sous-espace E^* de E_1^*, engendré par $(e_1^*,\ldots,e_n^*)$, est égale au produit scalaire d.

Représentation des objets supplémentaires. Dans l'image euclidienne des objets actifs, on représente l'objet supplémentaire n+j par le point M_{n+j}, dont les coordonnées dans le repère $(O, (f_1,\ldots,f_\rho))$ sont égales à :

$$\forall\ \ell \in \{1,\ldots,\rho\} \qquad (OM_{n+j} \mid f_\ell)_r = \lambda_\ell^{-1/2} (s_1(e_{n+j}) \mid \varepsilon_\ell^*)_{d_1} \ .$$

Restitution des produits scalaires entre objets actifs et objets supplémentaires.

$$\forall\ i \in \{1,\ldots,n\}\ ,\ \forall\ j \in \{1,\ldots,n_1\} \qquad (OM_i \mid OM_{n+j})_r = (S_1)_{i,n+j} \ .$$

□ *Démonstration.* Montrons tout d'abord que la $\ell^{\text{ième}}$ coordonnée du point M_{n+j} dans le repère $(O, (f_1,\ldots,f_\rho))$ est égale à $\lambda_\ell^{-1/2} < \varepsilon_\ell \mid s_1(e_{n+j}) >$. En effet, pour $\ell \leqslant \rho$, ε_ℓ^* appartient à E^*, ce qui entraîne $d_1(\varepsilon_\ell^*) = d(\varepsilon_\ell^*)$. De plus, $(\varepsilon_1^*,\ldots,\varepsilon_n^*)$ formant une base d-orthonormée de E^*, $d(\varepsilon_\ell^*) = \varepsilon_\ell$. On a donc :

$$(s_1(e_{n+j}) \mid \varepsilon_\ell^*)_{d_1} = < s_1(e_{n+j}) \mid \varepsilon_\ell >_{E^{**}} = < \varepsilon_\ell \mid s_1(e_{n+j}) >_{E^*} \ .$$

Pour $\ell \in \{\rho+1,\ldots,n\}$, ε_ℓ appartient à $\mathrm{Ker}(s)$, donc à $\mathrm{Ker}(s_1)$, ce qui entraîne :

$$< \varepsilon_\ell \mid s_1(e_{n+j}) > = (\varepsilon_\ell \mid e_{n+j})_{s_1} = < e_{n+j} \mid s_1(\varepsilon_\ell) > = 0 \ .$$

Alors le produit scalaire $(OM_i \mid OM_{n+j})_r$, égal à $\sum_{\ell=1}^{\rho} (s_1(e_{n+j}) \mid \varepsilon_\ell^*)_{d_1} < e_i \mid \varepsilon_\ell^* >$, peut s'écrire :

$$\sum_{\ell=1}^{n} < \varepsilon_\ell \mid s_1(e_{n+j}) > < e_i \mid \varepsilon_\ell^* > = < \sum_{\ell} < e_i \mid \varepsilon_\ell^* > \varepsilon_\ell \mid s_1(e_{n+j}) >$$

$$= < e_i \mid s_1(e_{n+j}) > = (S_1)_{i,n+j} \quad \square$$

4. RECHERCHE D'UNE IMAGE EUCLIDIENNE APPROCHEE DANS UN ESPACE DE DIMENSION FIXEE

4.1. SOLUTION OBTENUE PAR APPROXIMATION DES PRODUITS SCALAIRES

Approximation de l'application linéaire s (p.40). Parmi les éléments de $\mathcal{L}(E,E^*)$ de rang h fixé, inférieur à $\rho = rg(s)$, la meilleure approximation de s, pour la norme de Hilbert-Schmidt, est l'application linéaire s_h égale à la somme des h premiers éléments de la décomposition en valeurs singulières de s :

$$\forall\, x \in E \qquad s_h(x) = \sum_{\ell=1}^{h} \lambda_\ell(sd) \, \langle x | \varepsilon_\ell^* \rangle \, \varepsilon_\ell^* \quad .$$

L'application linéaire s_h est unique, sauf lorsque $\lambda_h(sd) = \lambda_{h+1}(sd)$. Mais, dans tous les cas :

$$\| s - s_h \|_t^2 = \min_{\substack{c \in \mathcal{L}(E,E^*) \\ rg(c) = h}} \| s - c \|_t^2 = \sum_{\ell=h+1}^{\rho} \lambda_\ell^2 (sd).$$

Image euclidienne approchée. Soient H un espace affine euclidien de dimension h, (F, r) l'espace vectoriel euclidien sous-jacent, $\mathcal{B}$ une base r-orthonormée de F et O un point quelconque de H.

Soient $M_1, \ldots, M_n$ les points dont les coordonnées dans le repère $(O, \mathcal{B})$ sont égales à :

$$\forall \ell \in \{1, \ldots, h\} \qquad \lambda_\ell^{-1/2} \, (s(e_i) | \, \varepsilon_\ell^*)_d = \lambda_\ell^{1/2} \, \langle e_i | \, \varepsilon_\ell^* \rangle \quad .$$

Ces points forment, avec le point O, une image euclidienne des objets, associée à l'approximation d'ordre h des produits scalaires.

□ *Démonstration.*

$$(OM_i | OM_j)_r = \sum_{\ell=1}^{h} \lambda_\ell \langle e_i | \varepsilon_\ell^* \rangle \langle e_j | \varepsilon_\ell^* \rangle$$

$$= \langle e_j \mid \sum_{\ell=1}^{h} \lambda_\ell \langle e_i | \varepsilon_\ell^* \rangle \varepsilon_\ell^* \rangle$$

$$= \langle e_j \mid s_h(e_i) \rangle = (e_i | e_j)_{s_h} \quad \square$$

Représentation d'un objet supplémentaire. Contrairement à l'image euclidienne des objets, associée aux produits scalaires qui, non seulement restitue les produits scalaires et les distances entre objets actifs, mais aussi entre objets actifs et objets supplémentaires, l'approximation ne concerne que les objets actifs.

On peut toutefois représenter l'objet supplémentaire n+j par le point M_{n+j} dont la $\ell^{\text{ième}}$ coordonnée dans le repère $(O, \mathcal{B})$ est égale à $\lambda_\ell^{-1/2} (s_1(e_{n+j}) | \varepsilon_\ell^*)_{d_1}$.

Mesure de la qualité de l'approximation. La base $(e_1^*,\ldots,e_n^*)$ étant d-orthogonale, $\| s - s_h \|_t^2$ se décompose en la somme :

$$\| s - s_h \|_t^2 = \sum_{j=1}^{n} p_j \| s(e_j) - s_h(e_j) \|_d^2 .$$

Une mesure de l'écart entre le tableau des produits scalaires S, et son approximation S_h d'ordre h est donnée par :

$$\sum_{i=1}^{n} \sum_{j=1}^{n} p_i p_j (S - S_h)_{ij}^2 = \| s - s_h \|_t^2 = \sum_{\ell=h+1}^{\rho} \lambda_\ell^2 (sd) .$$

□ *Démonstration.* Rappelons tout d'abord que la base $(e_1^*,\ldots,e_n^*)$ étant d-orthogonale mais non d-normée, $d(e_j^*) = p_j e_j$ (remarque p. 52). E étant identifié à son bidual, s, s_h et donc $(s - s_h)^*$ sont égaux à leurs transposés. On a alors :

$$\| s-s_h \|_t^2 = \mathrm{Tr}((s-s_h)^*(s-s_h)) = \sum_{j=1}^{n} \langle (s-s_h)^*(s-s_h)(e_j) | e_j^* \rangle_{E^*}$$

$$= \sum_j \langle (s-s_h)(e_j) | (s-s_h)^*(e_j^*) \rangle_{E^{**}} = \sum_j \langle (s-s_h)(e_j) | d(s-s_h)d(e_j^*) \rangle_{E^{**}}$$

$$= \sum_j p_j ((s-s_h)(e_j) | (s-s_h)(e_j))_d = \sum_j p_j \| s(e_j)-s_h(e_j) \|_d^2$$

$$= \sum_j p_j \sum_{i=1}^{n} \langle e_i | (s-s_h)(e_j) \rangle_{E^*}^2 \quad (e_i^* | e_i^*)_d$$

$$= \sum_i \sum_j p_i p_j (\langle e_i | s(e_j) \rangle - \langle e_i | s_h(e_j) \rangle)^2 = \sum_i \sum_j p_i p_j (S - S_h)_{ij}^2 \quad \square$$

4.2. EXPRESSION DE L'APPROXIMATION EN TERMES D'INERTIE

Une présentation classique de l'analyse en composantes principales consiste à caractériser une image euclidienne approchée, comme étant un nuage projeté d'inertie maximum. Cette caractérisation est encore valable dans le cas de l'analyse en composantes principales non centrée.

4.2.1. INERTIE D'UN NUAGE DE POINTS PESANTS

Définition de l'inertie. Soit H un espace affine euclidien, d'espace vectoriel euclidien sous-jacent (F, r). On considère un nuage de points $M_1,\ldots,M_n$ de H, affectés des poids $p_1,\ldots,p_n$.

Par analogie avec un système de points pesants en mécanique, on appelle inertie du nuage par rapport à un point O quelconque de H, la quantité $\sum_{i=1}^{n} p_i \, (OM_i|OM_i)_r$.

Décomposition de l'inertie dans une base de F. L'inertie d'un nuage se calcule en fait dans F, à partir des vecteurs OM_i et du produit scalaire. Sur une base $(f_1,\ldots,f_\rho)$ r-orthonormée de F, l'inertie se décompose en :

$$\sum_{i=1}^{n} p_i (OM_i|OM_i)_r = \sum_i p_i \Big(\sum_{k=1}^{\rho} (OM_i|f_k)_r \; f_k \Big| \sum_{\ell=1}^{\rho} (OM_i|f_\ell)_r \; f_\ell \Big)_r$$

$$= \sum_{\ell=1}^{\rho} \sum_{i=1}^{n} p_i \, (OM_i|f_\ell)_r^2 \quad .$$

4.2.2. EXPRESSION DE L'INERTIE COMME SOMME DES VALEURS D'UN QUOTIENT DE RAYLEIGH

Soit O, $M_1,\ldots,M_n$ une image euclidienne des objets, associée aux produits scalaires, dans l'espace affine euclidien H, d'espace vectoriel euclidien sous-jacent (F, r). A partir de cette image euclidienne, on définit l'application linéaire suivante.

Définition d'une application linéaire a de E dans F. A l'objet i, on a associé, d'une part le vecteur de base e_i de E, d'autre part le vecteur OM_i de F.

Pour passer d'un espace vectoriel à l'autre, on définit l'application a de (E, d^*) dans (F, r) par ses valeurs sur les vecteurs de base :

$$\forall\, i \in \{1,\ldots,n\} \qquad a(e_i) = OM_i \ .$$

Relation entre a et s. Rappelons que s est l'application linéaire de E dans E^*, associée au tableau de produits scalaires entre objets. $sd = a'(a')^*$, ce qui entraîne $rg(a) = rg(s)$.

□ *Démonstration.* Pour deux vecteurs e_i et e_j quelconques de la base $(e_1,\ldots,e_n)$ de E :

$$\langle e_i | s(e_j) \rangle = (e_i | e_j)_s = (OM_i | OM_j)_r = (a(e_i) | a(e_j))_r$$
$$= \langle a(e_i) | ra(e_j) \rangle_{F^*} = \langle e_i | a'r\, a(e_j) \rangle_{E^*} \ .$$

Ce qui entraîne $s = a'r\, a$, $sd = a'(r\, a\, d) = a'(a')^*$,
et $rg(a) = rg(a') = rg(a'(a')^*) = rg(sd) = rg(s)$ □

Expression de l'inertie, à l'aide du quotient de Rayleigh associé à $(a')^*a'$. Soit $(f_1,\ldots,f_\rho)$ une base r-orthonormée de F.

1. $\forall\, \ell \in \{1,\ldots,\rho\} \qquad (OM_i | f_\ell)_r = \langle e_i | a'(f_\ell^*) \rangle$.

2. L'inertie du nuage $M_1,\ldots,M_n$ par rapport au point O, est la somme des valeurs du quotient de Rayleigh associé à $(a')^*a'$, sur les vecteurs de base :

$$\sum_{i=1}^{n} p_i\, (OM_i | OM_i)_r = \sum_{\ell=1}^{\rho} R_{(a')^*a'}\, (f_\ell^*) = Tr(sd) \ .$$

□ *Démonstration.*

$$(OM_i | f_\ell)_r = (a(e_i) | f_\ell)_r = \langle a(e_i) | r(f_\ell) \rangle_{F^*}$$
$$= \langle a(e_i) | f_\ell^* \rangle_{F^*} = \langle e_i | a'(f_\ell^*) \rangle_{E^*} \ .$$

$$\sum_{i=1}^{n} p_i (OM_i | f_\ell)_r^2 = \sum_i (e_i^* | e_i^*)_d \ \langle e_i | a'(f_\ell^*) \rangle^2$$
$$= \Big(\sum_i \langle e_i | a'(f_\ell^*) \rangle\, e_i^* \ \Big|\ \sum_j \langle e_j | a'(f_\ell^*) \rangle\, e_j^* \Big)_d$$
$$= (a'(f_\ell^*) | a'(f_\ell^*))_d = ((a')^*\, a'(f_\ell^*) | f_\ell^*)_{r^*}$$
$$= R_{(a')^*a'}\, (f_\ell^*) \ .$$

Il en résulte que l'inertie du nuage, égale à $\sum_{\ell=1}^{\rho} \sum_{i=1}^{n} p_i \, (OM_i | f_\ell)_r^2$, s'écrit :

$$\sum_{\ell=1}^{\rho} R_{(a')^* a'} (f_\ell^*) = \mathrm{Tr}((a')^* a') = \mathrm{Tr}(aa^*) = \| a \|_t^2$$
$$= \mathrm{Tr}(a'(a')^*) = \mathrm{Tr}(sd) \quad \square$$

4.2.3. IMAGE EUCLIDIENNE APPROCHEE OBTENUE PAR PROJECTION

Projection r-orthogonale d'un point. Soient F_h un sous-espace vectoriel de (F, r) de dimension h, et H_h un sous-espace affine de H, associé à F_h.

La projection r-orthogonale d'un point M de H sur H_h est le point M', défini de manière unique par la condition : MM' appartient à $F_h^{\perp}$.

Projection de l'image euclidienne. Les points $O, M_1, \ldots, M_n$ se projettent sur H_h en $O', M'_1, \ldots, M'_n$. Dans F, cela se traduit par la projection des vecteurs OM_i sur les vecteurs $O'M'_i$ de F_h.

Inertie du nuage projeté. Soit $(\phi_1, \ldots, \phi_h)$ une base r-orthonormée de F_h. L'inertie du nuage projeté $M'_1, \ldots, M'_n$ par rapport à O', se décompose en $\sum_{\ell=1}^{h} \sum_{i=1}^{n} p_i (O'M'_i | \phi_\ell)_r^2$. Or, $O'M'_i$ étant la projection de OM_i sur F_h, cela entraîne que $(O'M'_i | \phi_\ell)_r = (OM_i | \phi_\ell)_r$, pour $\ell \leqslant h$, et que l'inertie est égale à :

$$\sum_{\ell=1}^{h} \sum_{i=1}^{n} p_i \, (OM_i | \phi_\ell)_r^2 = \sum_{\ell=1}^{h} R_{(a')^* a'} (\phi_\ell^*) \, .$$

Nuage projeté d'inertie maximum. L'inertie du nuage projeté est maximum, lorsque le sous espace F_h est engendré par un ensemble $\phi_1, \ldots, \phi_h$, de vecteurs propres de a a^*, associés aux h plus grandes valeurs propres.

□ *Démonstration.* h étant fixé, inférieur à ρ, l'inertie du nuage projeté, égale à $\sum_{\ell=1}^{h} R_{(a')^* a'}(\phi_\ell^*)$, est maximum (p. 34) lorsque $\phi_1^*, \ldots, \phi_h^*$ sont des vecteurs propres r^*-orthonormés de $(a')^* a'$, associés aux h plus grandes valeurs propres. Ce maximum est égal à $\sum_{\ell=1}^{h} \sigma_\ell^2(a') = \sum_{\ell=1}^{h} \lambda_\ell(sd)$. Alors les vecteurs $\phi_1 = r^{-1}(\phi_1^*), \ldots, \phi_h = r^{-1}(\phi_h^*)$ sont des vecteurs propres, r-orthonormés, du transposé a a^* de $(a')^* a'$, associés aux mêmes valeurs propres (p. 35) □

Image euclidienne approchée. Lorsque l'inertie du nuage projeté est maximum, les points O', M_1', ..., M_n' forment une image euclidienne des objets, associée à l'approximation d'ordre h des produits scalaires.

□ *Démonstration.* On déduit (p. 36) des vecteurs propres $\phi_1^*, \ldots, \phi_h^*$ de $(a')^* a'$, les vecteurs propres $a'(\phi_1^*), \ldots, a'(\phi_h^*)$ de $a'(a')^* = sd$. Ces vecteurs propres ont pour d-norme la racine carrée de la valeur propre. Alors, puisque pour $\ell \leqslant h$:

$$(O'M_i' \mid \phi_\ell)_r = (OM_i \mid \phi_\ell)_r = \langle e_i \mid a'(\phi_\ell^*) \rangle ,$$

le produit scalaire $(O'M_i' \mid O'M_j')_r$ est égal à :

$$\begin{aligned}(O'M_i' \mid O'M_j')_r &= \sum_{\ell=1}^{h} (O'M_i' \mid \phi_\ell)_r \, (O'M_j' \mid \phi_\ell)_r \\ &= \sum_{\ell} \langle e_i \mid a'(\phi_\ell^*) \rangle \langle e_j \mid a'(\phi_\ell^*) \rangle \\ &= \langle e_j \mid \sum_{\ell} \langle e_i \mid a'(\phi_\ell^*) \rangle a'(\phi_\ell^*) \rangle \\ &= \langle e_j \mid s_h(e_i) \rangle = (e_i \mid e_j)_{s_h} .\end{aligned}$$

Lorsque l'application linéaire s_h est définie de manière unique, c'est-à-dire lorsque $\lambda_h(sd)$ est différent de $\lambda_{h+1}(sd)$, cette image euclidienne est équivalente à celle obtenue p. 56 . Remarquons qu'il découle de la définition de l'équivalence entre images euclidiennes, que deux images équivalentes ont la même inertie □

Mesure de la perte d'inertie. L'inertie d'un nuage projeté est toujours inférieure à l'inertie du nuage initial.

En effet, il suffit de décomposer l'inertie sur une base de F, formée d'une base de F_h et d'une base de $F_h^{\perp}$ pour le démontrer.

La différence entre l'inertie du nuage initial et l'inertie du nuage projeté, atteint sa borne minimum $\sum_{\ell=h+1}^{\rho} \sigma_\ell^2(a') = \sum_{\ell=h+1}^{\rho} \lambda_\ell (sd)$, pour le nuage $O', M'_1, \ldots, M'_n$, ou pour un nuage équivalent.

On mesure habituellement la qualité de l'approximation, par le rapport de l'inertie du nuage projeté sur l'inertie totale.

5. CAS PARTICULIER DE PRODUITS SCALAIRES CENTRES

On peut centrer le tableau S des produits scalaires, afin d'obtenir une image euclidienne des objets, dans laquelle l'origine O est le centre de gravité des points $M_1, \ldots, M_n$. L'image euclidienne approchée, obtenue à partir du tableau des produits scalaires centrés, a l'avantage de faire apparaître les proximités et oppositions entre objets.

5.1. CENTRAGE DU TABLEAU DES PRODUITS SCALAIRES

Définition d'un tableau centré. A partir du tableau symétrique S des produits scalaires entre objets, on définit le tableau symétrique W, centré en lignes et en colonnes par rapport aux poids $p_1, \ldots, p_n$ des objets. Rappelons que les poids vérifient l'égalité $p_1 + \ldots + p_n = 1$.

$$\forall\, i,\ \forall\, j \in \{1, \ldots, n\} \qquad W_{ij} = S_{ij} - S_{i.} - S_{.j} + S_{..} \quad ,$$

avec $S_{i.} = \sum_{j=1}^{n} p_j S_{ij}$ $\qquad S_{.j} = \sum_{i=1}^{n} p_i S_{ij}$

$$S_{..} = \sum_{i=1}^{n} \sum_{j=1}^{n} p_i p_j S_{ij} = \sum_{i=1}^{n} p_i S_{i.} = \sum_{j=1}^{n} p_j S_{.j} \quad .$$

Remarque. W est appelé matrice de TORGERSON (1958).

Caractérisation d'un tableau centré. Pour qu'un tableau S de produits scalaires entre objets, soit centré en lignes et en colonnes par rapport aux poids des objets, il faut et il suffit que :

$$\forall\, j \in \{1,\ldots,n\} \qquad \sum_{i=1}^{n} p_i\, S_{ij} = 0 \quad .$$

□ *Démonstration.* La condition $S_{.j} = \sum_i p_i\, S_{ij} = 0$, pour tout j, entraîne que $S_{..} = 0$ et, de par la symétrie de S, $S_{i.} = 0$ pour tout i. Il en résulte que :

$$\forall\, i,\ \forall\, j \qquad W_{ij} = S_{ij} \quad .$$

Supposons maintenant que le tableau S soit centré en lignes et en colonnes, c'est-à-dire égal au tableau W. Alors :

$$\forall\, j \in \{1,\ldots,n\} \quad \sum_i p_i\, S_{ij} = \sum_i p_i\, S_{ij} - \sum_i p_i\, S_{i.} - \left(\sum_i p_i\right) S_{.j} + \left(\sum_i p_i\right) S_{..}$$

$$= S_{.j} - S_{..} - S_{.j} + S_{..} = 0 \quad \square$$

5.2. CONSEQUENCES DU CENTRAGE

Dans les paragraphes précédents, on n'a fait aucune hypothèse sur le tableau S des produits scalaires entre objets. Les résultats sont, par conséquent, valables lorsque le tableau S est centré. En particulier, dans l'image euclidienne des objets, associée aux produits scalaires centrés, la distance entre les points M_i et M_j est égale à la distance d_{ij} entre les objets i et j, déduite des produits scalaires initiaux. A ces résultats, s'ajoutent des propriétés particulières découlant du centrage. C'est la raison pour laquelle nous distinguons par la notation W un tableau centré, et par w l'application linéaire de E dans E^*, qui lui est associée.

Image euclidienne centrée. Si le tableau des produits scalaires est centré par rapport aux poids des objets, le point O de l'image euclidienne est le centre de gravité G des points $M_1,\ldots,M_n$, et réciproquement.

□ *Démonstration.* S_{ij} étant égal au produit scalaire $(OM_i|\, OM_j)_r$,

$$S_{.j} = \sum_i p_i\, (OM_i|\, OM_j)_r = \left(\sum_i p_i\, OM_i|\, OM_j\right)_r = (OG|\, OM_j)_r \quad .$$

De la même manière, $S_{i.} = (OM_i|OG)_r$ et

$$S_{..} = \sum_i p_i S_{i.} = (\sum_i p_i \ OM_i|OG)_r = (OG|OG)_r \ .$$

Reportons ces termes dans l'expression de W_{ij}. Regroupons S_{ij} et $S_{i.}$ d'une part, $S_{.j}$ et $S_{..}$ d'autre part :

$$W_{ij} = (OM_i|OM_j)_r - (OM_i|OG)_r - (OG|OM_j)_r + (OG|OG)_r$$
$$= (OM_i|GM_j)_r - (OG|GM_j)_r = (GM_i|GM_j)_r \ .$$

Réciproquement, l'égalité $\sum_i p_i \ GM_i = 0$ qui caractérise le centre de gravité, entraîne $\sum_i p_i \ (GM_i|GM_j)_r = 0$ qui caractérise un tableau de produits scalaires centrés.

Cette propriété se conserve par projection □

Inertie d'un nuage centré. L'inertie d'un nuage $M_1,\ldots,M_n$ par rapport à son centre de gravité est égale à

$$\frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} p_i \ p_j \ d^2_{M_iM_j} \ .$$

□ *Démonstration.* La distance entre deux points M_iM_j se déduit des produits scalaires entre les vecteurs GM_i et GM_j :

$$d^2_{M_iM_j} = (GM_i|GM_i)_r + (GM_j|GM_j)_r - 2(GM_i|GM_j)_r \ .$$

On reporte cette égalité dans $\sum_i \sum_j p_i \ p_j \ d^2_{M_iM_j}$. En remarquant que le terme $\sum_i \sum_j p_i \ p_j \ (GM_i|GM_j)_r$ s'annule car les produits scalaires sont centrés, et que chacun des termes $\sum_i p_i \sum_j p_j \ (GM_j|GM_j)_r$ et $\sum_j p_j \sum_i p_i \ (GM_i|GM_i)_r$ est égal à l'inertie du nuage, on obtient le résultat annoncé □

Expression de l'approximation en termes de distances. On traduit les propriétés relatives à l'inertie, en termes de distances entre points. Parmi les projections de l'image euclidienne centrée sur un sous-espace affine de dimension h fixée, le nuage $O',M'_1,\ldots,M'_n$ d'inertie maximum, réalise le maximum de $\sum_i \sum_j p_i \ p_j \ d^2_{M'_i \ M'_j}$.

La différence $\sum_i \sum_j p_i\, p_j\, (d^2_{ij} - d^2_{M'_i\, M'_j})$ est égale à $\sum_{\ell=h+1}^{\rho} 2\lambda_\ell\,(wd)$.

Propriétés de l'opérateur wd.

1. $rg(wd) \leqslant n-1$.

2. Tout vecteur propre ε^*_ℓ, associé à une valeur propre non nulle, est "centré" pour le poids des objets. C'est-à-dire qu'il vérifie :
$\sum_{i=1}^{n} p_i < e_i \,|\varepsilon^*_\ell > = 0$.

□ *Démonstration.* Montrons tout d'abord qu'un élément quelconque de E^* est centré si et seulement si, il est d-orthogonal à l'élément $e^*_1 + \ldots + e^*_n$.

$$\forall\, u \in E^* \qquad (\sum_{i=1}^{n} e^*_i \,|\, u)_d = \sum_i (e^*_i \,|\, \sum_{j=1}^{n} < e_j \,|\, u > e^*_j)_d$$

$$= \sum_i (e^*_i \,|\, e^*_i)_d < e_i | u > = \sum_i p_i < e_i | u > .$$

Montrons ensuite que $wd(e^*_1 + \ldots + e^*_n) = 0$, c'est-à-dire que $e^*_1 + \ldots + e^*_n$ est vecteur propre de wd, associé à la valeur propre nulle. Après avoir identifié E avec son bidual, w est égal à son transposé. Ceci permet de démontrer que la coordonnée de $wd(e^*_1 + \ldots + e^*_n)$ sur n'importe quel vecteur de base e^*_j est nulle :

$$< e_j | wd(e^*_1 + \ldots + e^*_n) > = \sum_{i=1}^{n} < e_j | wd(e^*_i) >_{E^*} = \sum_i < w(e_j) | d(e^*_i) >_{E^{**}}$$

$$= \sum_i < w(e_j) | p_i e_i >_{E^{**}} = \sum_i p_i < e_i | w(e_j) >_{E^*}$$

$$= \sum_i p_i\, W_{ij} = 0 \quad ,$$

car le tableau des produits scalaires W est centré.

Il en résulte que $rg(w) = rg(wd) \leqslant n-1$, et que les vecteurs propres de wd, associés à des valeurs propres non nulles, sont d-orthogonaux à $e^*_1 + \ldots + e^*_n$, donc centrés □

Interprétation de la position des objets le long d'un axe.
Supposons qu'on connaisse les valeurs d'une variable quantitative x sur les n objets.

On peut considérer que les coordonnées des objets, sur un axe de l'image euclidienne, sont les valeurs d'une autre variable quantitative, appelée "composante principale". On calcule la corrélation entre la composante principale et la variable.

Dans une image euclidienne centrée, les points se répartissent tout le long de l'axe. La position des objets ayant une forte coordonnée positive ou négative, peut alors être interprétée en fonction des valeurs de la variable x.

Objets supplémentaires. Rappelons que les objets supplémentaires sont affectés d'un poids nul. La définition du centrage du tableau des produits scalaires entre objets actifs et supplémentaires, reste la même. Mais l'image euclidienne est centrée sur le centre de gravité des objets actifs.

L'inertie, la restitution des distances, et l'explication de la position des objets le long d'un axe par le calcul de la corrélation de la composante principale avec une variable, ne concernent évidemment que les objets actifs.

6. ANALYSE EN COMPOSANTES PRINCIPALES D'UN TABLEAU CROISANT INDIVIDUS ET VARIABLES

On connaît les valeurs de m variables quantitatives observées sur n individus. Les individus sont affectés des poids $p_1,\ldots,p_n$. Les valeurs des variables *centrées* pour les poids des individus, sont rangées dans un tableau X comportant n lignes et m colonnes.

A partir de ces données, on peut calculer un tableau de produits scalaires entre les objets particuliers que sont les individus, et se ramener au cas précédent.

La démarche habituelle est différente car l'espace vectoriel des individus, à partir duquel on calcule les produits scalaires est, cette fois-ci, connu. On peut alors construire une image euclidienne des individus, associée aux produits scalaires, directement à partir du tableau de données X. On obtient une image euclidienne approchée par projection de cette image euclidienne.

6.1. PRODUITS SCALAIRES ENTRE INDIVIDUS

Espace vectoriel euclidien des individus. Soit (F, r) un espace vectoriel euclidien dont la dimension est égale au nombre m de variables. On choisit une base $(f_1,\ldots,f_m)$ r-orthonormée dans F, qui servira de *base de référence*. Les variables étant quantitatives, les valeurs $X_{i1},\ldots,X_{im}$ de la ligne i du tableau X sont des éléments de R. On peut, par conséquent, considérer que ces valeurs sont les coordonnées d'un vecteur x^i de l'espace vectoriel réel F, sur la base $(f_1,\ldots,f_m)$. On associe le vecteur x^i à l'individu i.

Distance euclidienne classique. Si on décide de prendre pour distance entre les individus i et j, la quantité

$d_{ij} = \left(\sum_{k=1}^{m} (X_{ik} - X_{jk})^2\right)^{1/2}$, cela revient à utiliser le produit scalaire r pour calculer les distances entre les vecteurs x^i et x^j de (F, r). En effet, exprimé dans la base $(f_1,\ldots,f_m)$ r-orthonormée, $(x^i|x^j)_r$ se réduit à $X_{i1}\,X_{j1} + \ldots + X_{im}\,X_{jm}$, et $d_{x^i x^j} = d_{ij}$.

Transformation linéaire des données. Supposons qu'à partir des valeurs "brutes" contenues dans le tableau de données X, on crée de nouvelles données par combinaison linéaire des valeurs initiales : la variable j prend la valeur $Y_{ij} = X_{i1}B_{j1} + \ldots + X_{im}\,B_{jm}$ sur l'individu i. Soit b l'opérateur de (F, r) dont la matrice dans la base $(f_1,\ldots,f_m)$ est B. On a alors associé à l'individu i un nouveau vecteur de F : $y^i = b(x^i) = Y_{i1}\,f_1 + \ldots + Y_{im}\,f_m$.

Lorsque cette transformation est injective mais non orthogonale, calculer les produits scalaires entre individus à partir des données transformées, revient à utiliser un produit scalaire q, différent du produit scalaire initial r, tout en gardant l'identification de l'individu i au vecteur $x^i = X_{i1}\,f_1 + \ldots + X_{im}\,f_m$, défini à partir des valeurs brutes.

□ *Démonstration.* Nous avons déjà montré (p. 45) que le semi-produit scalaire q, défini sur F par :

$(x^i|x^j)_q = (b(x^i)|b(x^j))_r = (y^i|y^j)_r$, est un produit scalaire si et seulement si l'opérateur b de E est injectif.

q est différent de r, si et seulement si, b n'est pas orthogonal ($b^* \neq b^{-1}$) □

Choix d'un autre produit scalaire. Calculer les distances entre individus à partir d'un produit scalaire q (différent du produit scalaire initial r) revient à prendre pour données, un nouveau tableau de mesures Y. Ce tableau est défini à une transformation orthogonale près, et se déduit du tableau de données brutes X par une transformation linéaire.

□ *Démonstration.* Il existe un opérateur injectif b de (F, r), défini à la multiplication par un opérateur orthogonal près, qui vérifie :

$$\forall\, i,\ \forall\, j \in \{1,\ldots,n\} \qquad (x^i \mid x^j)_q = (b(x^i) \mid b(x^j))_r \ .$$

On a alors associé à l'individu i, le vecteur $y^i = b(x^i)$, ce qui revient à prendre pour tableau de données, le tableau Y de terme général :

$$Y_{ij} = \langle b(x^i) \mid f_j^* \rangle = \sum_{k=1}^{m} X_{ik} \langle b(f_k) \mid f_j^* \rangle = \sum_k X_{ik} B_{jk} \quad \square$$

Centrage du tableau des produits scalaires. Puisque les variables mesurées sur les individus sont centrées, le tableau des produits scalaires entre individus est centré.

□ *Démonstration.* Les variables étant centrées, $\sum_{i=1}^{n} p_i X_{ij} = 0$, pour tout j. Alors :

$$\sum_{i=1}^{n} p_i x^i = \sum_i \sum_{j=1}^{m} p_i \langle x^i \mid f_j^* \rangle f_j = \sum_j \left(\sum_i p_i X_{ij} \right) f_j = 0 \ .$$

Quel que soit le produit scalaire utilisé entre les individus, on a donc $\sum_i p_i (x^i \mid x^j) = 0$, qui caractérise un tableau de produits scalaires centrés □

6.2. ANALYSE EN COMPOSANTES PRINCIPALES

6.2.1. IMAGE EUCLIDIENNE DES INDIVIDUS, CONSTRUITE A PARTIR DES PRODUITS SCALAIRES

On reprend les résultats du paragraphe 3 (p. 51). Soit E un espace vectoriel dont la dimension est égale au nombre n d'individus, on choisit une base $(e_1,\ldots,e_n)$ dans E. On munit l'espace dual E^* du produit scalaire d, défini à partir des poids des individus :

$$\forall\, i \in \{1,\ldots,n\} \qquad (e_i^* \mid e_i^*)_d = p_i$$
$$\forall\, i, \forall\, j \in \{1,\ldots,n\},\ i \neq j \qquad (e_i^* \mid e_j^*)_d = 0 \ .$$

Image euclidienne. A partir du tableau W des produits scalaires centrés entre individus, on définit le semi-produit scalaire w sur $(E,\ d^*)$:

$$\forall\, i,\ \forall\, j \in \{1,\ldots,n\} \qquad (e_i \mid e_j)_w = W_{ij} \ .$$

On note w, l'application linéaire de E dans E^*, associée au semi-produit scalaire w.

La décomposition en valeurs singulières de w permet de construire une image euclidienne des individus, associée aux produits scalaires centrés, dans un espace affine euclidien de dimension $\rho = rg(w) = rg(wd)$.

Les vecteurs propres de wd associés aux h plus grandes valeurs propres, fournissent une image euclidienne approchée d'ordre h. On appelle "plan principal" l'image euclidienne approchée d'ordre 2.

Interprétation de la position des individus le long d'un axe. Le tableau de produits scalaires entre individus étant centré, on peut interpréter facilement la position des individus le long d'un axe. On calcule les corrélations de la composante principale correspondante avec n'importe quelle variable quantitative mesurée sur les individus, y compris avec des variables n'ayant pas été prises en compte dans le tableau de données X.

6.2.2. IMAGE EUCLIDIENNE DES INDIVIDUS, CONSTRUITE A PARTIR DES DONNEES

Image euclidienne. Dans l'espace vectoriel euclidien (F, q) des individus (q pouvant être éventuellement égal au produit scalaire initial r), on choisit une base $\mathcal{B}$ q-orthonormée. Soit H un espace affine euclidien associé à (F, q), dans lequel on choisit un point G.

On considère les points $M_1,\ldots,M_n$ de H, qui ont pour coordonnées dans le repère $(G, \mathcal{B})$, les composantes des vecteurs x^i dans $\mathcal{B}$. Ces points forment, avec le point G, une image euclidienne centrée des produits scalaires $(x^i|x^j)_q$ entre individus.

Remarque. Puisque cette image euclidienne est équivalente à celle obtenue à partir du tableau des produits scalaires W, les points $M_1,\ldots,M_n$ (ainsi que le centre de gravité G) appartiennent en réalité à un sous-espace affine de H, de dimension $\rho = rg(w)$.

Application linéaire associée au tableau de données. A l'individu i, on a associé, d'une part le vecteur de base e_i de E, d'autre part le vecteur $x^i = GM_i$ de F. On définit l'application linéaire a de (E, d^*) dans (F, q) par ses valeurs sur les vecteurs de base :

$$\forall\ i \in \{1,\ldots,n\} \qquad a(e_i) = x^i \quad .$$

X est la matrice de l'application transposée a' de F^* dans E^*, dans les bases $(f_1^*,\ldots,f_m^*)$ de F^* et $(e_1^*,\ldots,e_n^*)$ de E^*.

□ *Démonstration.*

$$X_{ij} = \langle x^i|f_j^*\rangle = \langle a(e_i)|f_j^*\rangle_{F^*} = \langle e_i|a'(f_j^*)\rangle_{E^*} \qquad \square$$

Semi-produit scalaire associé à la matrice de variances. Puisque $\langle e_i|a'(f_j^*)\rangle = X_{ij}$, on associe à la variable j, l'élément $x_j = a'(f_j^*)$ de E^*, qui a pour coordonnées dans la base $(e_1^*,\ldots,e_n^*)$, les valeurs de cette variable.

En exprimant le produit scalaire $(x_i|x_j)_d$ dans la base $(e_1^*,\ldots,e_n^*)$, il est facile de vérifier que $(x_i|x_j)_d$ est égal à la covariance entre les variables i et j, puisque les variables sont centrées.

Or :

$$(x_i|x_j)_d = (a'(f_i^*) \mid a'(f_j^*))_d = \langle a'(f_i^*) \mid da'(f_j^*) \rangle_{E^{**}}$$
$$= \langle da'(f_j^*) | a'(f_i^*) \rangle_{E^*} = \langle ada'(f_j^*) | f_i^* \rangle_{F^*} \quad .$$

Il en résulte que l'application linéaire ada', de F^* dans F, a pour matrice dans les bases $(f_1^*,\ldots,f_m^*)$ de F^* et $(f_1,\ldots,f_m)$ de F, la matrice de variances V. C'est la raison pour laquelle on désigne ada' par la lettre v. La forme bilinéaire v, définie sur (F^*, q^*) par :

$$\forall\, i,\ \forall\, j \in \{1,\ldots,n\} \qquad (f_i^*|f_j^*)_v = \langle f_i^*|v(f_j^*)\rangle_{F^{**}} = (x_i|x_j)_d \quad ,$$

est symétrique et positive (propriété p. 51).

Relations entre v et w. Reprenons les résultats du paragraphe 4.2.2. (p. 59). Des égalités :

$$(a(e_i)|a(e_j))_q = (x^i|x^j)_q = (e_i|e_j)_w \quad ,$$

on déduit a'qa = w et a'(a')* = wd. v étant égal à ada', cela entraîne que (a')*a' = qv et aa^* = vq .

wd = a'(a')* et qv = (a')* a' sont des opérateurs auto-adjoints et positifs, dont les propriétés ont été étudiées p. 36 . En particulier, ils ont les mêmes valeurs propres. L'opérateur qv et son transposé vq également.

L'inertie du nuage $M_1,\ldots,M_n$ par rapport au centre de gravité G est égale à :

$$\frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} p_i\, p_j\, d^2_{M_iM_j} = \sum_{\ell=1}^{\rho} R_{qv}(f_\ell^*) = Tr(vq) = Tr(wd).$$

Image euclidienne approchée. En projetant les vecteurs x^i de F sur le sous-espace engendré par les vecteurs propres de aa^* = vq, associés aux h plus grandes valeurs propres, on obtient une image euclidienne des individus, associée à l'approximation d'ordre h des produits scalaires.

6.2.3. IMAGE EUCLIDIENNE DES VARIABLES, ASSOCIEE AUX CORRELATIONS

v est une forme bilinéaire, symétrique, positive, sur l'espace vectoriel euclidien (F^*, q^*).

On sait construire (p. 53) une image euclidienne des variables, associée aux produits scalaires $(x_i|x_j)_d$, à partir de la décomposition en valeurs singulières de v.

Soit $O,N_1,\ldots,N_m$ le nuage obtenu. Le produit scalaire q n'étant pas lié à la notion de poids, on ne peut pas parler d'inertie, ni de centrage de ce nuage de points, ni de variables supplémentaires.

Relation entre image euclidienne des variables, et image euclidienne des individus. Lorsque les variables du tableau X sont réduites, les coordonnées des points de l'image euclidienne des variables sont égales aux corrélations entre les variables et les composantes principales.

□ *Démonstration.* Etablissons tout d'abord les "formules de transition" entre les éléments propres de wd et de vq. $wd = a'(a')^*$ et $vq = aa^*$ ont les mêmes valeurs propres, et le même rang ρ . Si ε_ℓ^* est un vecteur propre de wd, d -normé, associé à la valeur propre λ_ℓ , $\varepsilon_\ell = d(\varepsilon_\ell^*)$ est vecteur propre du transposé a^*a de wd. $a(\varepsilon_\ell)$ est vecteur propre de $aa^* = vq$, associé à la même valeur propre λ_ℓ , mais la q -norme de $a(\varepsilon_\ell)$ est égale à la racine carrée de λ_ℓ.

Les variables étant réduites, la corrélation entre une variable x_j et la $\ell^{\text{ième}}$ composante principale, (utile pour expliquer la position des individus le long de l'axe), se réduit à $(x_j|\varepsilon_\ell^*)_d$. Puisque $x_j = a'(f_j^*)$, $(x_j|\varepsilon_\ell^*)_d$ est égal à :

$$< d(\varepsilon_\ell^*)|a'(f_j^*) >_{E^*} = < \varepsilon_\ell | a'(f_j^*) >_{E^*} = < a(\varepsilon_\ell)| f_j^* >_{F^*} ,$$

qui est la $\ell^{\text{ième}}$ coordonnée du point N_j dans l'image euclidienne des variables, associée aux corrélations $(x_i|x_j)_d$ □

Remarque. La norme du vecteur ON_j est égale à la d -norme du vecteur x_j, et cette norme vaut 1 puisque les variables sont réduites. Les points N_j sont donc situés sur une hypersphère (de dimension ρ) de centre O et de rayon 1.

Approximation des corrélations. D'après la propriété précédente, l'image euclidienne des variables, associée aux corrélations, permet d'interpréter la position des individus.

Remarquons, tout d'abord, que l'image euclidienne reconstruit les éléments de la matrice V. Pour reconstruire les corrélations il faut, par conséquent, que les variables du tableau X soient réduites.

Remarquons, ensuite, que la construction de l'image euclidienne approchée, ainsi que la mesure de l'approximation, dépendent du produit scalaire q . Ce nuage peut être très éloigné du nuage qu'on aurait obtenu avec le produit scalaire initial r. Ce qui rend son interprétation délicate.

Par conséquent, ce n'est que dans le cas où q est égal à r (on utilise la distance classique entre individus, calculée sur les variables centrées et réduites du tableau X) que l'image euclidienne d'ordre h restitue la meilleure approximation des corrélations. On appelle "cercle des corrélations" l'image euclidienne d'ordre 2. La mesure de l'approximation s'exprime en fonction des écarts entre la matrice des corrélations V, et son approximation V_h d'ordre h.

$$\| v - v_h \|_t^2 = \sum_{i=1}^{n} \sum_{j=1}^{m} (V - V_h)_{ij}^2 = \sum_{\ell=h+1}^{\rho} \lambda_\ell^2 \quad .$$

6.2.4. APPROXIMATION DU TABLEAU DE DONNEES

Reconstruction du tableau de données. Soient $\varepsilon_1^*, \ldots, \varepsilon_\rho^*$ des vecteurs propres d-orthonormés de wd, et $\phi_1, \ldots, \phi_\rho$ des vecteurs propres q-orthonormés de vq, associés aux mêmes valeurs propres non nulles $\lambda_1 \geqslant \ldots \geqslant \lambda_\rho$. Alors :

$$X_{ij} = \langle e_i | a'(f_j^*) \rangle = \sum_{\ell=1}^{\rho} \lambda_\ell^{1/2} \langle \phi_\ell | f_j^* \rangle_{F^*} \langle e_i | \varepsilon_\ell^* \rangle_{E^*} \quad .$$

□ *Démonstration.* On écrit la décomposition en valeurs singulières de l'application linéaire a' de (F^*, q^*) dans (E^*, d). Les valeurs singulières de a' sont égales aux racines carrées des valeurs propres de $a'(a')^* = wd$. Soit ϕ_ℓ un vecteur propre q-normé de $aa^* = vq$, alors $q(\phi_\ell)$ est vecteur propre q^*-normé du transposé $(a^*)'$ a'. On obtient le résultat en écrivant que :

$$a'(f_j^*) = \sum_{\ell=1}^{\rho} \lambda_\ell^{1/2} (f_j^* | q(\phi_\ell))_{q^*} \varepsilon_\ell^* = \sum_{\ell=1}^{\rho} \lambda_\ell^{1/2} \langle \phi_\ell | f_j^* \rangle_{F^*} \varepsilon_\ell^* \quad \square$$

Approximation du tableau de données. La meilleure approximation, de rang h inférieur à ρ, de l'application linéaire a' pour la norme de Hilbert-Schmidt, est l'application linéaire a'_h définie par la somme des h premiers termes de la décomposition en valeurs singulières de a' :

$$\forall\, j \in \{1,\ldots,m\} \qquad a'_h(f^*_j) = \sum_{\ell=1}^{h} \lambda_\ell^{1/2} \langle \phi_\ell \,|\, f^*_j \rangle \varepsilon^*_\ell \quad .$$

a_h réalise $\min_{\substack{c \in \mathcal{L}(F^*,E^*) \\ rg(c)=h}} \| a' - c \|_t^2 = \sum_{\ell=h+1}^{\rho} \lambda_\ell$.

Mesure de la qualité de l'approximation. Lorsqu'on utilise la distance classique entre les individus, c'est-à-dire le produit scalaire r sur F :

$$\| a'-a'_h \|_t^2 = \sum_{i=1}^{n} \sum_{j=1}^{m} p_i \, (X - X_h)^2_{ij} = \sum_{\ell=h+1}^{\rho} \lambda_\ell \quad .$$

Mais, lorsqu'on utilise un produit scalaire q différent du produit scalaire initial r, cela revient à prendre pour tableau de données (p. 68) un tableau Y déduit de X par une transformation linéaire. Le tableau Y est défini à une transformation orthogonale près.
$\| a'-a'_h \|_t^2$ s'exprime en fonction des écarts entre un quelconque de ces tableaux Y et son approximation d'ordre h :

$$\| a'-a'_h \|_t^2 = \sum_i \sum_j p_i \, (Y - Y_h)^2_{ij} \quad .$$

□ *Démonstration.* Les calculs sont faits dans le cas d'un produit scalaire quelconque q. Soit b un opérateur de F, tel que $q = b'r\,b$, $Y_{ij} = \langle b(x^i) \,|\, f^*_j \rangle_{F^*} = \langle ba(e_i) \,|\, f^*_j \rangle_{F^*} = \langle e_i \,|\, a'b'(f^*_j) \rangle_{E^*}$.

$$\begin{aligned}
\| a'-a'_h \|_t^2 &= Tr((a'-a'_h)^*(a'-a'_h)) = Tr(q(a-a_h)d(a'-a'_h)) \\
&= Tr(b'rb(a-a_h)d(a'-a'_h)) \\
&= Tr(rb(a-a_h)d(a'-a'_h)b') \\
&= \sum_{j=1}^{m} \langle rb(a-a_h)d(a'-a'_h)b'(f^*_j) \,|\, f_j \rangle_{F^{**}} \\
&= \sum_j \langle (a'-a'_h)b'(f^*_j) \,|\, d(a'-a'_h)b'r(f_j) \rangle_{F^*} \\
&= \sum_j \| (a'-a'_h)b'(f^*_j) \|_d^2
\end{aligned}$$

$$= \sum_j \sum_i \langle e_i | (a' - a'_h) b' (f^*_j) \rangle^2_{E^*} \, (e^*_i | e^*_i)_d$$

$$= \sum_i \sum_j p_i \left(\langle e_i | a' b' (f^*_j) \rangle - \langle e_i | a'_h b' (f^*_j) \rangle \right)^2$$

$$= \sum_i \sum_j p_i \, (Y - Y_h)^2_{ij} \quad \square$$

6.3. DISTANCES EUCLIDIENNES USUELLES ENTRE INDIVIDUS

Pour illustrer le paragraphe 6.1, nous présentons quelques produits scalaires q, définis sur l'espace vectoriel des individus. Les deux premiers produits scalaires sont définis à partir des valeurs du tableau de données X, les deux autres ne dépendent que de sa dimension.

Produit scalaire associé à la réduction des variables. La réduction des variables du tableau X est la transformation linéaire inversible $Y_{ij} = \frac{X_{ij}}{\sigma_j}$, dans laquelle σ_j est l'écart-type non nul de la variable j.

Réduire les variables équivaut à munir l'espace vectoriel F des individus du produit scalaire q, défini par les valeurs suivantes sur la base $(f_1, \ldots, f_m)$:

$$(f_j | f_j)_q = \frac{1}{\sigma_j^2} \quad , \quad (f_i | f_j)_q = 0 \quad \text{si } i \neq j \ .$$

Dans la base $(\sigma_1 f_1, \ldots, \sigma_m f_m)$ q-orthonormée, $(x^i | x^j)_q$ se réduit à $\sum_{k=1}^{m} Y_{ik} \, Y_{jk}$. On en conclut que $(x^i | x^j)_q$ est indépendant des changements d'unité de mesure, car Y_{ik} est un nombre sans dimension.

Par contre, nous avons vu (p. 73) que la réduction des variables et le produit scalaire q ne conduisent pas à la même image euclidienne des variables. Dans le premier cas, l'image euclidienne reconstruit les corrélations, dans le second cas, elle reconstruit les covariances.

Distance de Mahalanobis. Ce produit scalaire est défini par sa matrice dans les bases $(f_1, \ldots, f_m)$ de F et $(f^*_1, \ldots, f^*_m)$ de F^*, qui est l'inverse V^{-1} de la matrice de variances.

La distance de Mahalanobis est indépendante des transformations linéaires inversibles, effectuées sur les données.

□ *Démonstration.* Notons $v_X = ada'$ l'application linéaire associée à la matrice de variances, calculée sur le tableau de données X. Une transformation linéaire des données se traduit par $y^i = b(x^i) = ba(e_i)$, et conduit à l'application linéaire $v_Y = bad(ba)' = b\, v_X\, b'$. La distance de Mahalanobis se déduit du produit scalaire :

$$(x^i|x^j)_{V_X^{-1}} = \langle x^i | v_X^{-1}(x^j) \rangle = \langle x^i | (b^{-1} v_Y (b')^{-1})^{-1} (x^j) \rangle$$

$$= \langle x^i | b'\, v_Y^{-1}\, b(x^j) \rangle = \langle b(x^i) | v_Y^{-1}\, b(x^j) \rangle = (y^i \mid y^j)_{V_Y^{-1}} \quad \square$$

Transformation linéaire des données par accumulation ou par différence. Si par exemple, le tableau de données X décrit l'acquisition successive de biens, on peut considérer la transformation linéaire suivante :

$$\forall\, j \in \{1,\ldots,m\} \qquad Y_{ij} = X_{i1} + X_{i2} + \ldots + X_{ij} \ ,$$

qui revient à considérer, à chaque étape, l'accumulation de biens.

On peut envisager la transformation linéaire inverse :

$$\forall\, j \in \{1,\ldots,m\} \qquad Y_{ij} = X_{ij} - X_{i,j-1} \ ,$$

utile pour prendre en compte un accroissement (ou une diminution) de population entre deux recensements, par exemple.

Autres choix de produits scalaires. Pour filtrer un effet de contiguité, ou éliminer une information peu intéressante, on peut utiliser un produit scalaire approprié sur l'espace vectoriel F des individus (BESSE-RAMSAY (1986), ESCOUFIER (1987)). Un souci analogue de filtrage (LE FOLL (1982), MOM (1988)) conduit à munir l'espace vectoriel E^* d'un produit scalaire différent de d. C'est encore la méthode utilisée pour éliminer les corrélations entre individus (ARAGON-CAUSSINUS (1980)), ou en analyse des covariances partielles (LEBART-MORINEAU-FENELON (1979)).

IV. Description théorique de la méthode STATIS

1. PRINCIPE DE LA METHODE

1.1. QUEL GENRE DE DONNEES PEUT-ON TRAITER ?

La méthode STATIS s'applique à des données quantitatives, collectées dans l'une des deux situations suivantes.

Situation 1 : p tableaux de mesures ont été recueillies en différentes occasions, sur les mêmes individus.

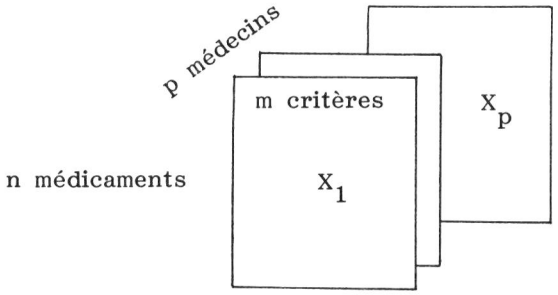
Situation 2 : les mêmes variables ont été mesurées sur p groupes d'individus différents.

Mais dans la plupart des cas, on est en présence des mêmes variables mesurées sur les mêmes individus en différentes occasions (le plus souvent à différentes dates). Alors, suivant qu'on s'intéresse aux individus ou aux variables, on se considère dans l'une ou l'autre des situations décrites. A chacune de ces situations correspond une stratégie appropriée, que nous présenterons au paragraphe 1.3. Auparavant, prenons un exemple concret.

Exemple d'un jeu de données. Un laboratoire pharmaceutique demande à p médecins de se prononcer sur n médicaments, selon m critères bien définis. Pour que les critères puissent être considérés comme des variables quantitatives, on va supposer que les médecins répondent par des notes comprises entre 0 et 100.

(Cet exemple ne peut qu'être fictif car dans ce genre d'enquête, les personnes interrogées cochent leur réponse sur une grille qui ne comporte que quelques appréciations qualitatives, éventuellement ordonnées, comme par exemple : "défavorable - peu favorable - sans opinion - favorable - très favorable").

Les données ainsi recueillies se présentent sous la forme de p tableaux. Chaque tableau fournit l'appréciation d'un médecin sur les n médicaments, selon les m critères. Les médicaments jouent le rôle d'individus, et les critères, le rôle de variables quantitatives.



1.2. LA NOTION D'ETUDE (X, q, d)

Décisions préliminaires à une analyse. Avant de mettre en oeuvre une méthode d'analyse sur un tableau de mesures, on doit évidemment s'interroger sur l'objectif poursuivi lors du recueil des données, car cette réflexion oriente le choix de la méthode à utiliser.

Mais on doit également s'interroger sur le poids qui doit être attribué à chacun des individus, et sur le calcul de la distance entre ces individus. Dans l'enquête sur les médicaments, on aurait pu écarter pour une première analyse, les médicaments peu connus car mis récemment sur le marché, en leur affectant un poids nul. Faire le choix, par exemple, de travailler avec des variables simplement centrées, ou centrées et réduites, détermine des distances différentes entre individus. Nous avons vu également qu'on pouvait filtrer certaines informations en utilisant une distance appropriée.

Rappel de la définition des espaces vectoriels euclidiens (F, q) et (E, d^*). X désigne un tableau de m mesures, *centrées* pour les poids $p_1, \ldots, p_n$ des individus.

Rappelons l'enchaînement des définitions du chapitre précédent (p. 67). (F, r) est un espace vectoriel euclidien, dont la dimension est égale au nombre m de variables. Soit $(f_1,\ldots,f_m)$ une base r-orthonormée de F. On associe à l'individu i, l'élément x^i de F, dont les coordonnées dans la base $(f_1,\ldots,f_m)$ sont les valeurs $X_{i1},\ldots,X_{im}$. On choisit, éventuellement, un produit scalaire q sur F, différent de r, pour calculer les distances entre individus.

E est un espace vectoriel, dont la dimension est égale au nombre n d'individus. Soit $(e_1,\ldots,e_n)$ une base de E. A partir du tableau W de produits scalaires centrés entre individus, on définit la forme bilinéaire w, symétrique, positive, sur E :

$$\forall\ i,\ \forall\ j \in \{1,\ldots,n\} \qquad (e_i|e_j)_w = W_{ij} = (x^i|x^j)_q \quad ,$$

et l'application linéaire w de $\mathcal{L}(E, E^*)$ qui lui est associée. On munit l'espace dual E^* du produit scalaire d, défini à partir des poids non nuls des individus, sur la base $(e_1^*,\ldots,e_n^*)$ duale de la base $(e_1,\ldots,e_n)$:

$$\forall\ i \in \{1,\ldots,n\} \qquad (e_i^*|e_i^*)_d = p_i$$

$$\forall\ i,\ \forall\ j \in \{1,\ldots,n\},\ i \neq j \qquad (e_i^*|e_j^*)_d = 0 \quad .$$

Définition d'une étude. Une étude est un triplet (X,q,d), dans lequel on a précisé les produits scalaires q et d, choisis pour analyser le tableau de données X.

Rappel des relations existant entre individus et variables. Soit a l'application linéaire de (E, d^*) dans (F, q), définie par :

$$\forall\ i \in \{1,\ldots,n\} \qquad a(e_i) = x^i \quad .$$

Le tableau de données X est la matrice de l'application transposée a', dans les bases $(f_1^*,\ldots,f_m^*)$ de F^* et $(e_1^*,\ldots,e_n^*)$ de E^*.

On associe à la variable j, l'élément $x_j = a'(f_j^*)$ de E^*. Les variables étant centrées, le produit scalaire $(x_i|x_j)_d$ est égal à la covariance entre les variables i et j. A partir de la matrice de variances V, on définit la forme bilinéaire v, symétrique, positive, sur F^* :

$$\forall\ i,\ \forall\ j \in \{1,\ldots,m\} \qquad (f_i^*|f_j^*)_v = (x_i|x_j)_d \quad ,$$

et l'application linéaire v de $\mathcal{L}(F^*, F)$ qui lui est associée.

Enfin, rappelons que $wd = a'(a')^*$ et $vq = aa^*$ ont les mêmes valeurs propres non nulles $\lambda_1 \geqslant \ldots \geqslant \lambda_\rho$. Ce qui entraîne :

$$Tr(wd) = Tr(vq) = \| a \|^2_{t\,\mathcal{L}(E,F)} = \lambda_1 + \ldots + \lambda_\rho \ .$$

Par contre, les valeurs propres de wd étant les valeurs singulières de w (p. 45) :

$$\| w \|^2_{t\,\mathcal{L}(E,E^*)} = Tr(w^*w) = \sigma_1^2(w) + \ldots + \sigma_\rho^2(w) = \lambda_1^2 + \ldots + \lambda_\rho^2$$

$$= Tr(dwdw) = Tr(wdwd) = Tr(a'(a')^*a'(a')^*)$$

$$= Tr((a')^*a'(a')^*a') = Tr(qvqv) = Tr(v^*v) = \| v \|^2_{t\,\mathcal{L}(F^*,F)} \ .$$

Rappel de la relation entre les produits scalaires q et r.
Supposons qu'on calcule les distances entre individus (considérés comme vecteurs de F) à partir d'un produit scalaire q, différent du produit scalaire initial r. Il existe un opérateur injectif b de (F, r), défini à la multiplication par un opérateur orthogonal près, tel que $(x^i | x^j)_q = (b(x^i) | b(x^j))_r$. Ceci équivaut à : $q = b'rb$.

Utiliser le produit scalaire q revient à travailler sur le tableau de données Y, dont l'élément général est égal à :

$$Y_{ij} = \langle y^i | f_j^* \rangle_{F^*} = \langle b(x^i) | f_j^* \rangle_{F^*} = \langle ba(e_i) | f_j^* \rangle_{F^*}$$

$$= \langle e_i | a'b'(f_j^*) \rangle_{E^*} = \langle e_i | y_j \rangle_{E^*} \ .$$

L'élément $y_j = a'b'(f_j^*)$ de E^* est associé à la variable j du tableau Y. Puisque les variables du tableau X sont centrées, les variables du tableau Y sont également centrées. En effet :

$$\sum_{i=1}^{n} p_i \, Y_{ij} = \langle b(\sum_i p_i \, x^i) | f_j^* \rangle = 0 \ .$$

Si on note v_X l'application linéaire de $\mathcal{L}(F^*, F)$, associée à la matrice de variances calculée sur le tableau de données X, $v_Y = bv_Xb'$.

1.3. STRATEGIE DE LA METHODE STATIS

L'idée essentielle de la méthode est la recherche d'une structure commune aux études, qu'on appelle *intrastructure*. Suivant les données, ou l'objectif de l'analyse, on s'intéresse aux individus ou aux variables.

Recherche d'une structure des individus, commune aux études. Cette stratégie est adaptée à la situation 1 : les tableaux de données X_k concernent les mêmes individus, les variables peuvent ne pas être les mêmes d'un tableau à l'autre. Cela implique que, si la définition de l'espace vectoriel euclidien (E, d^*) est valable pour l'ensemble des études, l'espace vectoriel euclidien des individus (F_k, q_k) dépend de l'étude considérée.

La question "existe-t-il une structure des individus, commune aux études ?", peut se poser autrement : "les relations inter-individuelles, c'est-à-dire les distances entre individus, sont-elles stables d'une étude à l'autre ?"

Pour répondre à cette question, on compare les formes bilinéaires w_k définies sur (E, d^*) ou, ce qui revient au même, les applications linéaires w_k de $\mathcal{L}(E, E^*)$ qui leur sont associées, au moyen de la distance déduite du produit scalaire de Hilbert-Schmidt sur $\mathcal{L}(E, E^*)$. Par opposition au terme "intrastructure" qui décrit la structure des individus à l'intérieur d'une étude, on appelle *interstructure* les relations entre études, décrites par les distances entre w_k.

A partir de l'interstructure, on construit une forme bilinéaire, symétrique, positive, *compromis* w. Lorsque les distances entre w_k sont faibles, on peut affirmer qu'il existe bien une structure des individus commune aux études. Cette structure est décrite par les distances entre individus, données par le compromis w.

Dans la dernière étape de la méthode, on trace les *trajectoires* des individus. L'interstructure a mis en évidence, sans les expliquer, les écarts des études, entre elles et avec le compromis. Les trajectoires permettent de décrire ces écarts, au niveau individuel.

Recherche d'une structure des variables, commune aux études. Cette stratégie est adaptée à la situation 2, dans laquelle les mêmes variables sont mesurées sur p groupes d'individus différents. On considère que les individus des différents groupes appartiennent au même espace vectoriel euclidien (F, q). En particulier cela implique que, si on calcule les distances entre individus à partir d'un produit scalaire q, différent du produit scalaire initial r, q ne doit pas dépendre des données. Par exemple, q peut correspondre à l'accumulation ou à la différence, mais pas à la réduction des données (p. 75).

On compare les applications linéaires v_k de $\mathcal{L}(F^*, F)$, associées aux matrices de variances de chaque étude, au moyen de la distance déduite du produit scalaire de Hilbert-Schmidt sur $\mathcal{L}(F^*, F)$. Si l'interstructure ne met pas en évidence de trop grandes différences entre les v_k, le compromis fournit une matrice de covariances entre les variables, qui décrit la structure des variables, commune aux études. Enfin, les trajectoires permettent de déceler quelles sont les variables responsables des écarts entre v_k.

Exemple. Reprenons l'enquête du laboratoire pharmaceutique. Il est logique de s'intéresser au classement des médicaments, fourni par chaque médecin. Ce qui conduit à adopter la première stratégie, et à déclarer que deux médecins sont proches s'ils ont donné la même répartition des médicaments. En particulier, ils ont rapproché ou opposé les mêmes médicaments, mais peut être pas pour les mêmes raisons.

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

Si on adopte la seconde stratégie, on ne s'intéresse plus à la répartition des médicaments, mais aux critères d'appréciation. Supposons qu'on ait centré et réduit les variables, pour chaque médecin.

On déclare alors que deux médecins sont proches, lorsque les matrices de corrélations sont voisines : ces médecins fournissent les mêmes relations linéaires entre critères.

Nous démontrerons que, lorsque deux médecins k_1 et k_2 sont proches pour les deux stratégies, les tableaux X_{k_1} et X_{k_2} sont voisins : les médecins ont classé les médicaments de la même façon, et pour les mêmes raisons.

2. COEFFICIENTS D'ASSOCIATION ENTRE ETUDES, PROPOSES PAR Y. ESCOUFIER

Pour comparer deux études (X_1, q, d) et (X_2, q, d) portant sur les mêmes variables, mesurées sur les mêmes individus dans des circonstances différentes, il semble naturel d'utiliser (JAFFRENOU (1978)) le produit scalaire de Hilbert-Schmidt sur $\mathcal{L}(F^*, E^*)$, entre les applications linéaires a'_1 et a'_2 associées aux tableaux de données :

$$(a'_1 | a'_2)_{t \atop \mathcal{L}(F^*,E^*)} = \mathrm{Tr}(q a_1 \, d \, a'_2) \ .$$

La distance déduite de ce produit scalaire, est une distance entre les tableaux de données X_1 et X_2, et entre les études (X_1, q, d) et (X_2, q, d).

Si on n'est pas dans cette situation (mêmes variables mesurées sur les mêmes individus), ou si on veut associer deux études autrement que par l'égalité des tableaux de données, on peut utiliser d'autres coefficients d'association entre tableaux (RAMSAY et al. (1984) - MARCOTORCHINO (1986) - COPPI (1986) - CLARET (1987)). La méthode STATIS est basée sur l'emploi des coefficients proposés par ESCOUFIER (1973, 1977 et 1986), qui mettent en évidence une structure commune aux études.

2.1. PRODUITS SCALAIRES ENTRE OBJETS w_k ou v_k

Produit scalaire entre objets w_k. Soient (X_1, q_1, d) et (X_2, q_2, d) deux études portant sur les mêmes individus, affectés des mêmes poids $p_1, \ldots, p_n$, mais pas forcément sur les mêmes variables. Les produits scalaires q_1 et q_2 entre individus peuvent être différents.

On appelle coefficient COVV (covariance vectorielle) entre les deux études, le produit scalaire de Hilbert-Schmidt sur $\mathcal{L}(E, E^*)$ entre les applications linéaires w_1 et w_2, associées aux produits scalaires entre individus : $(w_1 | w_2)_{t\ \mathcal{L}(E, E^*)}$.

Ce produit scalaire définit une distance entre objets w_k, et non entre études. Par contre la condition $w_1 = w_2$ définit une relation d'équivalence entre les études (X_1, q_1, d) et (X_2, q_2, d). Deux études sont équivalentes lorsque les images euclidiennes des individus, associées aux tableaux de produits scalaires W_1 et W_2, sont équivalentes.

Coefficient RV entre études. On appelle coefficient RV (corrélation vectorielle) entre les études (X_1, q_1, d) et (X_2, q_2, d), le produit scalaire de Hilbert-Schmidt sur $\mathcal{L}(E, E^*)$ entre les applications linéaires w_1 et w_2 normées :

$$RV(1,2) = \left(\frac{w_1}{\| w_1 \|_t} \,\middle|\, \frac{w_2}{\| w_2 \|_t} \right)_t .$$

La distance entre les applications linéaires w_1 et w_2 normées est égale à $(2(1 - RV(1,2)))^{1/2}$. Lorsque $RV(1,2) = 1$, cette distance est nulle et $\frac{w_1}{\| w_1 \|_t} = \frac{w_2}{\| w_2 \|_t}$. La condition $RV(1,2) = 1$ définit une relation d'équivalence entre les études (X_1, q_1, d) et (X_2, q_2, d). Cette équivalence se traduit par : l'image euclidienne des individus de la première étude (associée au tableau W_1 de produits scalaires) se déduit de l'image euclidienne des individus de la seconde étude (associée au tableau W_2), par une homothétie de rapport $\frac{\| w_1 \|_t}{\| w_2 \|_t}$.

Produit scalaire entre objets $\mathbf{v_k}$. Soient (X_1, q, d_1) et (X_2, q, d_2) deux études portant sur les mêmes variables, mesurées sur des groupes d'individus qui peuvent être différents. Remarquons que les distances entre individus, dans chacun des groupes, sont calculées à partir du même produit scalaire q sur F.

On utilise le produit scalaire de Hilbert-Schmidt sur $\mathcal{L}(F^*, F)$ entre les applications linéaires v_1 et v_2, associées aux produits scalaires entre variables : $(v_1 | v_2)_{t\ \mathcal{L}(F^*,F)}$. Les variables étant centrées, deux études sont équivalentes lorsque leurs matrices de variances sont égales.

On peut également utiliser le produit scalaire de Hilbert-Schmidt sur $\mathcal{L}(F^*, F)$ entre les applications linéaires v_1 et v_2 normées : $\left(\frac{v_1}{\| v_1 \|_t} \middle| \frac{v_2}{\| v_2 \|_t}\right)_t$. Alors les études sont équivalentes lorsque les matrices de variances sont homothétiques.

2.2. PROPRIETES DE CES PRODUITS SCALAIRES

2.2.1. PROPRIETES GENERALES

Egalité des études. Lorsque les études (X_1, q, d) et (X_2, q, d) sont équivalentes, pour le produit scalaire $(v_1 | v_2)_{t\ \mathcal{L}(F^*,F)}$ et l'un des deux produits scalaires COVV ou RV, elles sont égales.

□ *Démonstration.* Puisqu'on reconstruit un tableau de données à partir des valeurs propres et des vecteurs propres de wd et de vq, il y a égalité entre les tableaux de données X_1 et X_2, dès que $w_1 = w_2$ et $v_1 = v_2$.

On a démontré (p. 80) que $\| v \|_{t\ \mathcal{L}(F^*,F)} = \| w \|_{t\ \mathcal{L}(E,E^*)}$.

Alors $v_1 = v_2$ entraîne $\| w_1 \|_t = \| w_2 \|_t$. Si de plus $\frac{w_1}{\| w_1 \|_t} = \frac{w_2}{\| w_2 \|_t}$, on en déduit $w_1 = w_2$, et l'égalité des tableaux de données X_1 et X_2 □

Positivité. Les produits scalaires $(v_1|v_2)_{t}$ et $(w_1|w_2)_{t}$ sont positifs. Les produits scalaires entre applications normées sont compris entre 0 et 1.

□ *Démonstration.*

$$(w_1|w_2)_{t} = \mathrm{Tr}(dw_1dw_2) = \mathrm{Tr}(w_1dw_2d) = (w_1d|w_2d)_{t}.$$

Nous avons démontré (p. 44) que w_1d et w_2d sont des opérateurs autoadjoints et positifs de E^*, et que leur produit scalaire est positif (p. 39). Il en est de même pour $(v_1|v_2)_{t} = (v_1q|v_2q)_{t}$.

D'après l'inégalité de Cauchy-Schwartz, les produits scalaires normés sont inférieurs à 1 □

2.2.2. INTERPRETATION DU PRODUIT SCALAIRE ENTRE OBJETS w_k

Notation. On note $(x_k)^i$ le vecteur de F_k associé à l'individu i du tableau X_k, et $(x_k)_j$ l'élément de E^*, associé à la variable j de ce même tableau.

Expression de la distance entre w_1 et w_2, en fonction des produits scalaires entre individus. Soient deux études (X_1, q_1, d) et (X_2, q_2, d) portant sur les mêmes individus, munis des poids $p_1,\ldots,p_n$. Alors $\| w_1 - w_2 \|_t^2$ se décompose sur E^* en :

$$\| w_1 - w_2 \|_t^2 = \sum_{j=1}^{n} p_j \| w_1(e_j) - w_2(e_j) \|_d^2$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{n} p_i p_j (W_1 - W_2)_{ij}^2$$

$$= \sum_i \sum_j p_i p_j (((x_1)^i|(x_1)^j)_{q_1} - ((x_2)^i|(x_2)^j)_{q_2})^2 .$$

□ *Démonstration* par un développement analogue à celui qu'on a effectué p. 57, pour mesurer la qualité de l'approximation □

Expression de $(w_1|w_2)_t$ en fonction des données.

Lorsqu'on utilise la distance classique entre les individus, c'est-à-dire les produits scalaires r_1 sur F_1, et r_2 sur F_2, $(w_1|w_2)_t$ est égal à la somme des carrés des covariances des m_1 variables du tableau X_1, avec les m_2 variables du tableau X_2 :

$$(w_1|w_2)_t = \sum_{i=1}^{m_1} \sum_{j=1}^{m_2} ((x_1)_i|(x_2)_j)_d^2 \quad .$$

Mais, lorsqu'on utilise des produits scalaires q_1 et q_2 différents des produits scalaires initiaux r_1 et r_2, nous avons démontré que cela revient à travailler sur des tableaux de données Y_1 et Y_2, déduits des tableaux de données X_1 et X_2 par une transformation linéaire, définie à une transformation orthogonale près. Alors :

$$(w_1|w_2)_t = \sum_{i=1}^{m_1} \sum_{j=1}^{m_2} ((y_1)_i|(y_2)_j)_d^2 \quad .$$

□ *Démonstration.* Les calculs sont faits dans le cas de produits scalaires $q_1 = b_1' r_1 b_1$ et $q_2 = b_2' r_2 b_2$.

Remarquons que l'application linéaire $a_1 da_2'$ de $\mathcal{L}(F_2^*, F_1)$ a pour matrice dans les bases $(\phi_1^*, \ldots, \phi_{m_2}^*)$ r_2^*-orthonormée de F_2^* et et $(f_1, \ldots, f_{m_1})$ r_1-orthonormée de F_1, la matrice de variances entre les variables de X_1 et les variables de X_2 :

$$\langle a_1 da_2' (\phi_j^*)|f_i^* \rangle_{F_1^*} = \langle da_2' (\phi_j^*)|a_1'(f_i^*) \rangle_{E^{**}}$$

$$= (a_1'(f_i^*)|a_2'(\phi_j^*))_d = ((x_1)_i|(x_2)_j)_d \quad .$$

On la note $v_{X_1X_2}$, et on note $v_{Y_1Y_2} = b_1 v_{X_1X_2} b_2'$, l'application linéaire qui correspond à la matrice de variances entre les variables de Y_1 et les variables de Y_2. Alors :

$$(w_1|w_2)_t = \mathrm{Tr}(dw_1\, dw_2) = \mathrm{Tr}(da_1' q_1 a_1 da_2' q_2 a_2)$$

$$= \mathrm{Tr}(q_1 v_{X_1X_2} q_2 v_{X_1X_2}') = \mathrm{Tr}(b_1' r_1 b_1 v_{X_1X_2} b_2' r_2 b_2 v_{X_1X_2}')$$

$$= \mathrm{Tr}(r_1 v_{Y_1Y_2} r_2 v_{Y_1Y_2}') = \sum_{i=1}^{m_1} (f_i^* | r_1 v_{Y_1Y_2} r_2 v_{Y_1Y_2}' (f_i^*))_{r_1^*}$$

$$= \sum_i \langle f_i^* \mid v_{Y_1Y_2}\, r_2 v'_{Y_1Y_2}\, (f_i^*) \rangle = \sum_i \langle v'_{Y_1Y_2}\, (f_i^*) \mid r_2 v'_{Y_1Y_2} (f_i^*) \rangle$$

$$= \sum_i \| v'_{Y_1Y_2}(f_i^*) \|^2_{r_2} = \sum_i \sum_{j=1}^{m_2} \langle v'_{Y_1Y_2}(f_i^*) \mid \phi_j^* \rangle^2_{F_2^*}$$

$$= \sum_i \sum_j \langle v_{Y_1Y_2}\, (\phi_j^*) \mid f_i^* \rangle^2_{F_1^*} = \sum_i \sum_j ((y_1)_i \mid (y_2)_j)^2_d \quad \square$$

Orthogonalité entre w_1 et w_2. Si $(w_1 \mid w_2)_{t_{\mathcal{L}(E,E^*)}} = 0$, les variables de X_1 ont des covariances nulles avec les variables de X_2. De plus, $(a'_1 \mid a'_2)_{t_{\mathcal{L}(F^*,E^*)}} = 0$.

□ *Démonstration.* De la décomposition de $(w_1 \mid w_2)_t$ en $\sum_i \sum_j \langle v_{Y_1Y_2}\, (\phi_j^*) \mid f_i^* \rangle^2$, on déduit que :

$$(w_1 \mid w_2)_t = 0 \iff v_{Y_1Y_2} = b_1 v_{X_1X_2} b'_2 = 0 \ .$$

b_1 et b_2 étant des opérateurs injectifs, $b_1 v_{X_1X_2} b'_2 = 0$ entraîne que :

$$v_{X_1X_2} b'_2 = 0 \iff b_2 v'_{X_1X_2} = 0 \implies v'_{X_1X_2} = 0 \iff v_{X_1X_2} = 0$$

$$\iff \forall\, i \in \{1,\ldots,m_1\} \ , \ \forall\, j \in \{1,\ldots,m_2\} \quad ((x_1)_i \mid (x_2)_j)_d = 0 \ .$$

D'autre part :

$$v_{X_1X_2} = a_1 d a'_2 = 0 \implies (a'_1 \mid a'_2)_{t_{\mathcal{L}(F^*,E^*)}} = \mathrm{Tr}(q a_1 d a'_2) = 0 \quad \square$$

2.2.3. INTERPRETATION DU PRODUIT SCALAIRE ENTRE OBJETS v_k

Expression de $(v_1 \mid v_2)_t$. Lorsqu'on utilise la distance classique entre les individus, c'est-à-dire le produit scalaire r sur F, $(v_1 \mid v_2)_t$ est égal à la somme des produits de la covariance de deux variables du tableau X_1 avec les deux mêmes variables du tableau X_2 :

$$(v_1 \mid v_2)_t = \sum_{i=1}^{m} \sum_{j=1}^{m} ((x_1)_i \mid (x_1)_j)_{d_1} ((x_2)_i \mid (x_2)_j)_{d_2} \ .$$

□ *Démonstration.* Les calculs sont faits dans le cas général d'un produit scalaire q = b'rb. Dans la base $(f_1,\ldots,f_m)$ r-orthonormée de (F, q) :

$$(v_1|v_2)_t = \mathrm{Tr}(qv_{X_1}qv_{X_2}) = \mathrm{Tr}(b'rbv_{X_1}b'rbv_{X_2}) = \mathrm{Tr}(rv_{Y_1}rv_{Y_2})$$

$$= \sum_{j=1}^{m} (f_j^*|\, rv_{Y_1}rv_{Y_2}(f_j^*))_{r^*} = \sum_j (v_{Y_1}(f_j^*)|v_{Y_2}(f_j^*))_r$$

$$= \sum_j \sum_{i=1}^{m} \langle v_{Y_1}(f_j^*)|f_i^*\rangle \langle v_{Y_2}(f_j^*)|f_i^*\rangle$$

$$= \sum_i \sum_j ((y_1)_i|(y_1)_j)_{d_1} ((y_2)_i|(y_2)_j)_{d_2} .$$

Il en résulte que :

$$\| v_1 - v_2 \|_t^2 = \sum_j \| v_{Y_1}(f_j^*) - v_{Y_2}(f_j^*) \|_r^2 = \sum_i \sum_j (V_{Y_1} - V_{Y_2})_{ij}^2 \quad \square$$

Cas particulier des variables centrées et réduites. On utilise fréquemment la distance euclidienne classique, calculée à partir des variables centrées et réduites. Si les variables de X sont déjà réduites, cela revient à choisir $q = r$, et la matrice V de l'application linéaire v dans les bases $(f_1^*,\ldots,f_m^*)$ de F^* et $(f_1,..,f_m)$ de F, est une matrice de corrélations. Sinon, on utilise le produit scalaire q, correspondant à la réduction des variables. La matrice V est une matrice de variances. Par contre, les variables de Y sont centrées et réduites. Mais, dans les deux cas, $(w_1|w_2)_t$ et $(v_1|v_2)_t$ s'expriment en fonction des corrélations.

3. IMAGE EUCLIDIENNE DES ETUDES, ASSOCIEE AUX PRODUITS SCALAIRES ENTRE OBJETS

3.1. CHOIX DE L'OBJET REPRESENTANT L'ETUDE

Suivant les données, ou l'objectif de l'analyse, nous avons montré l'intérêt de choisir l'objet w_k, ou l'objet v_k, pour représenter l'étude n° k.

Le choix des objets normés $\frac{w_k}{\| w_k \|_t}$, ou $\frac{v_k}{\| v_k \|_t}$, sera discuté p. 96 , à propos du compromis.

Ce choix étant fait, on obtient un tableau S de produits scalaires entre objets, auquel on peut appliquer les résultats du chapitre III. Soient $\Pi_1, \ldots, \Pi_p$ les poids affectés aux études.

Exemple de poids Π_k. Dans la comparaison des études au moyen des objets v_k, on peut donner à l'étude un poids proportionnel au nombre d'individus pris en compte dans le calcul de la matrice de corrélation V_k. Ce poids représente la confiance qu'on peut accorder aux coefficients de corrélation.

3.2. CONSTRUCTION DE L'IMAGE EUCLIDIENNE DES ETUDES

Soit G un espace vectoriel, dont la dimension est égale au nombre p d'études. Soit $(g_1, \ldots, g_p)$ une base de G. A partir du tableau S de produits scalaires entre objets, on définit la forme bilinéaire s , symétrique, positive sur G :

$$\forall\, k,\ \forall\, \ell \in \{1,\ldots,p\} \qquad (g_k | g_\ell)_s = S_{k\ell} \ ,$$

et l'application linéaire s de $\mathcal{L}(G, G^*)$ qui lui est associée. On munit l'espace dual G^* du produit scalaire δ , défini à partir des poids $\Pi_1, \ldots, \Pi_p$, sur la base $(g_1^*, \ldots, g_p^*)$ duale de la base $(g_1, \ldots, g_p)$:

$$\forall\, k \in \{1,\ldots,p\} \qquad (g_k^* | g_k^*)_\delta = \Pi_k$$

$$\forall\, k,\ \forall\, \ell \in \{1,\ldots,p\} ,\ k \neq \ell \qquad (g_k^* | g_\ell^*)_\delta = 0 \ .$$

Image euclidienne des études. La décomposition en valeurs singulières de s permet de construire une image euclidienne $O, M_1, \ldots, M_p$ des études, associée aux produits scalaires entre objets, à partir des vecteurs propres γ_ℓ^* de $s\delta$, dans un espace dont la dimension est égale au rang de s (p. 53).

En se limitant aux vecteurs propres de $s\delta$, associés aux h plus grandes valeurs propres, on obtient une image euclidienne des études, associée à l'approximation d'ordre h des produits scalaires entre objets.

Etudes supplémentaires. Lorsqu'on on ne veut pas inclure une étude dans l'analyse, on lui affecte un poids nul. On peut toutefois représenter cette étude supplémentaire dans l'image euclidienne précédente, en utilisant la technique décrite p. 55 .

3.3. CONSEQUENCE DE LA POSITIVITE DES PRODUITS SCALAIRES ENTRE OBJETS

Dans l'image euclidienne O, $M_1,\ldots,M_p$ des études, associée aux produits scalaires entre objets, la longueur du vecteur OM_k est égale à la norme de l'objet. Puisque $\| w_k \|_{t_{\mathcal{L}(E,E^*)}} = \| v_k \|_{t_{\mathcal{L}(F^*,F)}}$ la longueur du vecteur OM_k est la même, quel que soit l'objet, w_k ou v_k , choisi. Dans le cas d'objets normés, la longueur du vecteur OM_k est égale à 1.

Le cosinus de l'angle entre deux vecteurs OM_{k_1} et OM_{k_2} est égal au produit scalaire entre les objets normés. Par exemple, qu'on choisisse l'objet w_k, ou l'objet normé $\frac{w_k}{\| w_k \|_t}$, ce cosinus est égal à $RV(k_1,k_2)$.

Les produits scalaires entre objets étant positifs, les cosinus sont positifs et les angles sont aigus. Les vecteurs $OM_1,\ldots,OM_p$ ont, par conséquent, la particularité de former les arêtes d'un cône convexe. La propriété suivante donne une caractérisation mathématique de cette particularité.

Propriété. Les coordonnées sur la base $(g_1^*,\ldots,g_p^*)$ du vecteur propre de $s\delta$, associé à la plus grande valeur propre, sont toutes de même signe.

□ *Démonstration.* Tout d'abord, rappelons que la base $(g_1^*,\ldots,g_p^*)$ étant δ-orthogonale mais non δ-normée, $\delta(g_k^*) = \mathbb{I}_k\, g_k$.

Considérons le quotient de Rayleigh $R_{s\delta}(u)$ pour un élément u δ-normé de G^*, et décomposons u sur la base $(g_1^*,\ldots,g_p^*)$:

$$R_{s\delta}(u) = (s\delta(u)|u)_\delta = \langle s\delta(u) \mid \delta(u) \rangle$$

$$= \sum_{k=1}^{p} \sum_{\ell=1}^{p} < g_k | u > < g_\ell | u > < s\,\delta\,(g_\ell^*) | \delta\,(g_k^*) >$$

$$= \sum_{k} \sum_{\ell} \Pi_k \Pi_\ell < g_k | u > < g_\ell | u > < s(g_\ell) | g_k >_{G^{**}} \quad .$$

Soit γ_1^* un vecteur propre de $s\,\delta$, associé à la plus grande valeur propre λ_1. Soit u l'élément $\sum_{k=1}^{p} | < g_k | \gamma_1^* > |\ g_k^*$, obtenu en prenant les valeurs absolues des coordonnées de γ_1^* sur la base $(g_1^*, \ldots, g_p^*)$. Les termes $S_{k\ell} = < s(g_\ell) | g_k >$ étant tous positifs,

$$R_{s\,\delta}(u) = \sum_{k} \sum_{\ell} \Pi_k \Pi_\ell \, | < g_k | \gamma_1^* > | \ | < g_\ell | \gamma_1^* > | < s(g_\ell) | g_k >$$

est obtenu à partir de $R_{s\delta}(\gamma_1^*)$, en remplaçant chaque terme par sa valeur absolue. Ce qui entraîne $R_{s\delta}(u) \geqslant R_{s\delta}(\gamma_1^*) = \lambda_1$. Or la caractérisation de la plus grande valeur propre, comme borne supérieure du quotient de Rayleigh, implique l'inégalité contraire, et donc l'égalité $R_{s\delta}(u) = \lambda_1$.

S'il existait deux coordonnées $< g_k | \gamma_1^* >$ et $< g_\ell | \gamma_1^* >$ de signe contraire, cela impliquerait $R_{s\,\delta}(\gamma_1^*) < R_{s\,\delta}(u)$, alors qu'il y a égalité. On en conclut que toutes les coordonnées de γ_1^* sont de même signe, et qu'on peut les choisir positives □

Image euclidienne centrée. En général, on se limite à l'examen d'un image euclidienne plane. Les points, groupés du côté positif de l'axe 1, ne se séparent que sur l'axe 2. Par conséquent, si on s'intéresse davantage aux distances entre objets qu'aux produits scalaires entre objets normés, l'image euclidienne des études, associée au tableau de produits scalaires centrés entre objets, fournit une information supplémentaire, même lorsqu'on se limite à l'examen d'une image plane.

Les associations et oppositions entre études sont mises en évidence sur cette image euclidienne centrée. Cela peut inciter à rechercher une partition de l'ensemble des études, par une méthode de classification basée sur la distance déduite du produit scalaire entre les objets. On peut, également, chercher à expliquer la position des études le long d'un axe, en calculant les corrélations de cet axe avec les variables quantitatives qu'on pourrait connaître sur les études (par exemple l'âge des médecins...).

4. LE COMPROMIS, RESUME MAJORITAIRE DES OBJETS

On cherche un compromis, de la même nature que les objets choisis pour représenter les études, qui soit un "bon" résumé de l'ensemble des objets.

4.1. DEFINITION ET PROPRIETES DU COMPROMIS

Pour ne pas introduire une notation nouvelle, la définition, ainsi que les propriétés du compromis sont écrites pour les objets w_k, mais sont bien entendu valables pour les autres objets.

Définition du compromis. Soient $< g_k | \gamma_1^* >$, k=1,...,p, les coordonnées positives du vecteur propre γ_1^* de s_δ associé à la plus grande valeur propre λ_1, dans la base $(g_1^*,\ldots,g_p^*)$. Le compromis w est l'application linéaire de $\mathcal{L}(E, E^*)$ qui est la moyenne pondérée des applications linéaires w_k :

$$w = \alpha \sum_{k=1}^{p} \Pi_k < g_k | \gamma_1^* > w_k \ .$$

α est un coefficient positif de normalisation, spécifique du type d'objets, qu'on précisera par la suite.

Propriété. La forme bilinéaire associée au compromis est symétrique, et positive.

□ *Démonstration.* A l'application linéaire w de $\mathcal{L}(E, E^*)$, on associe la forme bilinéaire w sur E, définie par :

$$\forall\, i,\ \forall\, j \in \{1,\ldots,n\} \qquad (e_i | e_j)_w = < e_i | w(e_j) > \ .$$

On obtient le résultat, en remarquant que w est une combinaison linéaire des formes bilinéaires w_k, symétriques, positives, dont les coefficients sont positifs □

Représentation du compromis. Dans l'image euclidienne des études, associée aux produits scalaires entre objets, le compromis est situé sur le premier axe, à la distance $\alpha\, \lambda_1^{1/2}$ de l'origine O.

□ *Démonstration.* La coordonnée du compromis sur l'axe ℓ est obtenue par combinaison linéaire des coordonnées $\lambda_\ell^{1/2} < g_k | \gamma_\ell^* >$ des points M_k. Π_k étant égal à $(g_k^* | g_k^*)_\delta$, cette coordonnée vaut :

$$\alpha \sum_{k=1}^{p} \Pi_k < g_k | \gamma_1^* > \lambda_\ell^{1/2} < g_k | \gamma_\ell^* > = \alpha \lambda_\ell^{1/2} (\gamma_1^* | \gamma_\ell^*)_\delta \quad .$$

Les vecteurs propres γ_ℓ^* de $s\delta$ étant δ-orthonormés, la première coordonnée du compromis vaut $\alpha \lambda_1^{1/2}$, et les autres sont nulles □

Détermination du coefficient de normalisation. On choisit le coefficient α , pour que le compromis soit un objet de même nature que les objets intervenant dans sa construction.

Dans le cas d'objets w_k, on choisit α de telle sorte que la norme de w (égale à $\alpha \lambda_1^{1/2}$) soit égale à la moyenne des normes des w_k. Ce qui donne :

$$w = (\lambda_1^{-1/2} \sum_{\ell=1}^{p} \Pi_\ell \| w_\ell \|_t) \sum_{k=1}^{p} \Pi_k < g_k | \gamma_1^* > w_k \quad .$$

Ainsi, la matrice de w dans les bases $(e_1,\ldots,e_n)$ de E, et $(e_1^*,\ldots,e_n^*)$ de E^* peut être considérée comme un tableau de produits scalaires entre individus, du même ordre de grandeur que les produits scalaires des différents tableaux W_k.

Dans le cas d'objets normés, en prenant $\alpha = \lambda_1^{-1/2}$, on obtient un compromis normé :

$$w = \lambda_1^{-1/2} \sum_k \Pi_k < g_k | \gamma_1^* > \frac{w_k}{\| w_k \|_t} \quad \Longrightarrow \quad \| w \|_t = 1 \quad .$$

Enfin, dans le cas d'objets v_k, si on écrit le compromis v sous la forme $\sum_k \alpha_k v_k$, le coefficient de normalisation est choisi de telle sorte que $\sum_k \alpha_k = 1$. Ce qui donne :

$$v = \frac{1}{\sum_{\ell=1}^{p} \Pi_\ell < g_\ell | \gamma_1^* >} \sum_{k=1}^{p} \Pi_k < g_k | \gamma_1^* > v_k \quad .$$

Alors, si les variables de chaque étude sont centrées et réduites, la matrice de v dans les bases $(f_1^*,\ldots,f_m^*)$ de F^*, et $(f_1,\ldots,f_m)$ de F peut être considérée comme une matrice de corrélations entre les variables communes aux études.

Propriété d'optimisation. Le compromis w est l'objet le plus "corrélé" (au sens du produit scalaire de Hilbert-Schmidt entre les objets) avec les objets w_k.

Ceci s'énonce de la façon suivante. Parmi les combinaisons linéaires $\sum_{\ell=1}^{p} \alpha_\ell \, w_\ell$, le compromis w réalise le maximum de :

$$\frac{\sum_{k=1}^{p} \Pi_k \, (\sum_{\ell=1}^{p} \alpha_\ell \, w_\ell \mid w_k)_t^2}{\| \sum_{\ell=1}^{p} \alpha_\ell \, g_\ell \|_{\delta^*}^2} \, ,$$

et ce maximum est égal à λ_1^2.

□ *Démonstration.* Rappelons que $(w_k | w_\ell)_t = \langle g_k | s(g_\ell) \rangle$, et que $\delta(g_k^*) = \Pi_k \, g_k$. Le numérateur est égal à :

$$\sum_k \Pi_k \, (\sum_{\ell_1} \alpha_{\ell_1} w_{\ell_1} \mid w_k)_t \, (\sum_{\ell_2} \alpha_{\ell_2} w_{\ell_2} \mid w_k)_t =$$

$$\sum_{\ell_1} \sum_{\ell_2} \alpha_{\ell_1} \alpha_{\ell_2} \sum_k \Pi_k \, (w_{\ell_1} | w_k)_t \, (w_k | w_{\ell_2})_t .$$

Or $\sum_k \Pi_k \langle g_{\ell_1} | s(g_k) \rangle \langle g_k | s(g_{\ell_2}) \rangle =$

$$\sum_k \langle g_{\ell_1} | s(\Pi_k \, g_k) \rangle \langle g_k | s(g_{\ell_2}) \rangle =$$

$$\sum_k \langle g_{\ell_1} | s \, \delta(g_k^*) \rangle \langle g_k | s(g_{\ell_2}) \rangle =$$

$$\langle g_{\ell_1} | s \, \delta (\sum_k \langle g_k | s(g_{\ell_2}) \rangle \, g_k^*) \rangle = \langle g_{\ell_1} | s \, \delta \, s(g_{\ell_2}) \rangle .$$

Si on note x le vecteur $\sum_{\ell=1}^{p} \alpha_\ell \, g_\ell$, le numérateur est égal à $\langle x | s \, \delta \, s(x) \rangle = (x | \delta \, s \, \delta \, s(x))_{\delta^*} = (s^* s(x) | x)_{\delta^*}$. Le dénominateur est égal à $(x|x)_{\delta^*}$, et le quotient est le quotient de Rayleigh $R_{s^*s}(x)$.

Ce quotient est maximum lorsque x est vecteur propre de s^*s, associé à la plus grande valeur propre $\lambda_1(s^*s) = \sigma_1^2(s)$. D'après les propriétés de s (p. 45), $\sigma_1(s) = \lambda_1(s \, \delta)$, x est vecteur propre du transposé $\delta \, s \, \delta$ de $s \, \delta$, et $\delta^{-1}(x)$ est vecteur propre de $s \, \delta$. On peut prendre $\delta^{-1}(x)$ égal à γ_1^*, ou égal à $\alpha \, \gamma_1^*$.

Après identification de G avec son bidual, δ est égal à son transposé, et α_ℓ est égal au coefficient de w_ℓ dans le compromis :

$$\alpha_\ell = \langle x | g_\ell^* \rangle = \alpha \langle \delta(\gamma_1^*) | g_\ell^* \rangle_{G^*} = \alpha \langle \gamma_1^* | \delta(g_\ell^*) \rangle_{G^{**}}$$

$$= \alpha \Pi_\ell \langle g_\ell | \gamma_1^* \rangle_{G^*} \quad \square$$

4.2. INTERPRETATION DU COMPROMIS

Nous allons expliquer ce que représente le compromis dans les cas types présentés p. 97 . Remarquons tout d'abord que les images euclidiennes approchées qu'on se contente de lire habituellement, ne sont qu'un reflet déformé de la réalité. Il faut compléter cette lecture par celle des normes des objets, et des produits scalaires entre objets normés.

Considérons le cas d'études (X_k, q_k, d), pour lesquelles les objets w_k ont des normes voisines, et les coefficients RV sont élevés. Ces deux conditions entraînent que les objets w_k sont proches les uns des autres. Sur la figure n° 1, cela se traduit par des vecteurs OM_k, de même longueur, formant un faisceau resserré. Le fait que les distances entre objets w_k soient faibles, implique qu'il existe une structure des individus, commune aux études, reflétée par le compromis. Les coordonnées des points M_k sur le premier axe sont voisines, et les contributions des différents objets w_k à la construction du compromis sont du même ordre.

Sur la figure n° 2, nous avons rajouté un vecteur OM_5 perpendiculaire aux autres. L'objet w_5, différent des autres objets, a une coordonnée faible sur le premier axe et n'intervient pratiquement pas dans la construction du compromis. Ceci justifie l'emploi du terme "résumé majoritaire", pour qualifier le compromis.

Dans le cas de la figure n° 3, les normes des objets w_k ne sont pas du même ordre de grandeur, et les coordonnées des points sur le premier axe reflètent ces différences de normes. Seuls les objets dont la norme est élevée contribuent à la construction du compromis. Il est préférable, dans ce cas, de caractériser l'étude k par l'objet normé $\dfrac{w_k}{\| w_k \|_t}$.

IMAGES EUCLIDIENNES TYPES

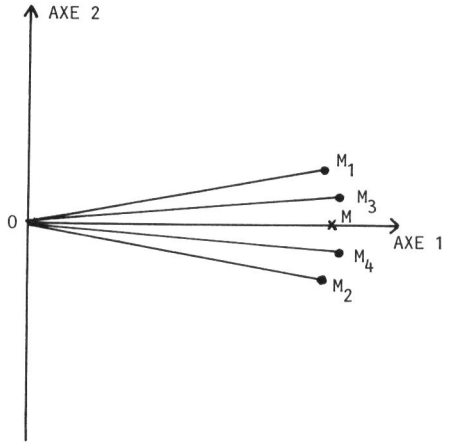


Figure n° 1

LE COMPROMIS w (REPRESENTE PAR M) REFLETE UNE STRUCTURE COMMUNE

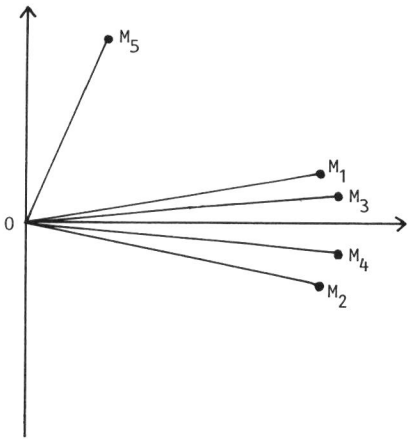


Figure n° 2

L'OBJET w_5 EST MAL REPRESENTE DANS LE COMPROMIS

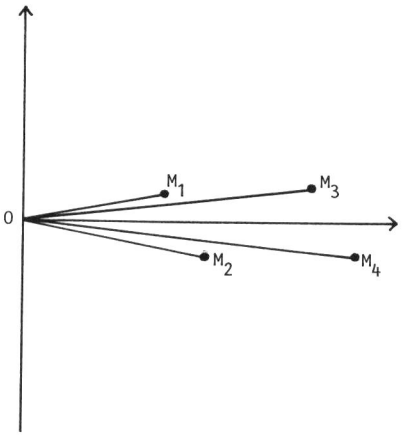


Figure n° 3

LE COMPROMIS N'EST PAS UN BON RESUME, CAR LES NORMES SONT TROP DIFFERENTES

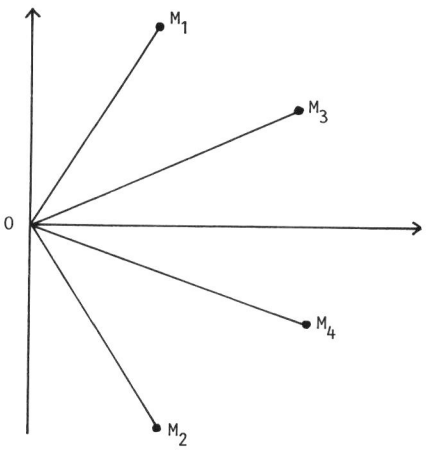


Figure n° 4

IL N'Y A PAS DE STRUCTURE COMMUNE

Ce qui revient à analyser les études "normées" $\left(\frac{X_k}{\| w_k \|^{1/2}}, q_k, d\right)$, et à comparer les images euclidiennes "normées" des individus. Lorsque les objets normés sont proches les uns des autres, on obtient une figure analogue à la figure n° 1.

Sur la figure n° 4, les cosinus sont faibles, les w_k sont éloignés les uns des autres, et il n'y a pas de structure des individus, commune aux études. Le compromis est simplement la moyenne pondérée des objets, qui est la plus "corrélée" (au sens du produit scalaire entre objets) à ces objets.

5. REPRESENTATION DES INDIVIDUS, DANS LE CAS D'ETUDES (X_k, q_k, d)

5.1. IMAGE EUCLIDIENNE COMPROMIS DES INDIVIDUS

Rappel sur le compromis. On est dans la situation 1 : on compare les études (X_k, q_k, d), portant sur les mêmes individus, au moyen des objets w_k ou $\frac{w_k}{\| w_k \|_t}$. Le compromis $w = \Sigma\, \alpha_k\, w_k$ est une application linéaire de (E, d^*) dans (E^*, d), associée à une forme bilinéaire, symétrique, positive sur E. La matrice W du compromis dans les bases $(e_1,\ldots,e_n)$ de E et $(e_1^*,\ldots,e_n^*)$ de E^*, peut être considérée comme un tableau de produits scalaires compromis entre les individus, centré pour les poids des individus. En effet, les tableaux W_k étant centrés :

$$\sum_{i=1}^{n} p_i \langle e_i | w(e_j) \rangle = \sum_i p_i \sum_k \alpha_k \langle e_i | w_k(e_j) \rangle = \sum_k \alpha_k \sum_i p_i (W_k)_{ij} = 0 \quad .$$

Image euclidienne compromis. La décomposition en valeurs singulières du compromis w s'exprime en fonction des valeurs propres non nulles $\lambda_1 \geqslant \ldots \geqslant \lambda_\rho$, et des vecteurs propres associés, d-orthonormés, $\varepsilon_1^*,\ldots,\varepsilon_\rho^*$ de l'opérateur wd de E^*.

A partir de cette décomposition, on sait construire une image euclidienne compromis des individus, dans un espace dont la dimension est égale au rang ρ de w. Cette image euclidienne est centrée. Soit $G, M_1, \ldots, M_n$ le nuage obtenu.

La distance $d_{M_iM_j}$ entre les points M_i et M_j de cette image euclidienne, représente la distance compromis entre les individus i et j. $d_{M_iM_j}$ se déduit des distances entre les individus i et j, dans chaque étude :

$$d^2_{M_iM_j} = W_{ii} + W_{jj} - 2\,W_{ij} = \sum_k \alpha_k \left((W_k)_{ii} + (W_k)_{jj} - 2(W_k)_{ij}\right)$$

$$= \sum_k \alpha_k \, \| (x_k)^i - (x_k)^j \|^2_{q_k} \quad .$$

Remarque. L'image euclidienne compromis est équivalente à celle qu'on aurait obtenue, en faisant l'analyse en composantes principales du tableau de données, construit en juxtaposant les tableaux $\alpha_k^{1/2} X_k$.

Corrélations des variables avec les axes du compromis. L'image euclidienne compromis étant centrée, on peut interpréter la position des individus le long d'un axe ℓ . On calcule la corrélation de la composante principale du compromis, correspondant à cet axe, avec les variables de chaque étude, y compris celles qui n'ont pas été prises en compte dans le calcul des produits scalaires entre individus. Si la variable x est centrée et réduite, cette corrélation est égale à $(\varepsilon^*_\ell | x)_d$.

Le graphique des corrélations des variables avec les composantes principales du compromis, résume ces informations. Une variable x est représentée par le point dont la coordonnée sur l'axe ℓ est égale à $(\varepsilon^*_\ell | x)_d$.

5.2. TRAJECTOIRES DES INDIVIDUS

L'interstructure - c'est-à-dire les produits scalaires entre objets w_k ou objets normés, et l'image euclidienne des études, associée à ces produits scalaires - a mis en évidence, sans les expliquer, les écarts des objets entre eux et avec le compromis. Les trajectoires permettent de décrire ces écarts au niveau individuel.

Remarque. On a pris l'habitude d'utiliser le terme "trajectoire" car le plus souvent, les données décrivent un phénomène évolutif. Pourtant dans l'exemple des médecins, ce terme ne convient pas, puisqu'une "trajectoire" représente les écarts entre les jugements des médecins sur un médicament donné.

5.2.1. CONSTRUCTION DES TRAJECTOIRES

Définition d'une trajectoire. Dans l'image euclidienne compromis, on place les différentes positions d'un individu tel qu'il est décrit par chaque étude, selon une technique rappelant la représentation des points supplémentaires. Les différentes positions d'un individu définissent sa trajectoire.

Construction d'une trajectoire. Soient $\varepsilon_1^*, \ldots, \varepsilon_\rho^*$, les vecteurs propres de wd, associés aux valeurs propres non nulles $\lambda_1 \geqslant \ldots \geqslant \lambda_\rho$. Rappelons (p. 53) que la $\ell^{\text{ième}}$ coordonnée du point M_i de l'image euclidienne compromis, dans le repère $(G, \mathcal{B})$, est égale à :

$$\lambda_\ell^{1/2} < e_i | \varepsilon_\ell^* > = \lambda_\ell^{-1/2} \, (w(e_i) | \varepsilon_\ell^*)_d \, .$$

On représente l'individu i (y compris un individu supplémentaire), décrit par l'étude k (y compris une étude supplémentaire), par le point M_i^k dont la $\ell^{\text{ième}}$ coordonnée dans le repère $(G, \mathcal{B})$ est égale à :

$$\lambda_\ell^{-1/2} \, (w_k(e_i) \mid \varepsilon_\ell^*)_d \quad ,$$

ou $\lambda_\ell^{-1/2} \left(\dfrac{w_k(e_i)}{\| w_k \|_t} \mid \varepsilon_\ell^*\right)_d$, lorsque les études sont représentées par les objets normés $\dfrac{w_k}{\| w_k \|_t}$.

Position compromis par rapport aux points de la trajectoire. Le point compromis M_i est le centre de gravité des différents points M_i^k, pondérés par les α_k, et cette propriété se conserve par projection.

Les individus supplémentaires ne sont pas intervenus dans l'interstructure, ni dans le compromis. Par conséquent, leur position compromis n'existe pas.

Individus absents de certaines études. Si un individu est absent de certaines études, on peut le situer quand même dans l'image euclidienne compromis. Il faut le traiter comme un individu supplémentaire, et calculer les coordonnées des points de sa trajectoire, correspondant aux études dans lesquelles il est présent.

Trajectoire de l'individu "moyen". L'individu moyen d'une étude est l'individu fictif qui aurait pour valeurs, les moyennes des variables de l'étude. Les variables étant centrées par étude, l'individu moyen de chaque étude est situé à l'origine G de l'image euclidienne compromis.

5.2.2. EXPLICATION DE LA DISTANCE ENTRE w_{k_1} ET w_{k_2} PAR LES ECARTS INDIVIDUELS

Décomposition de la distance entre les objets w_k. La distance entre deux applications linéaires w_{k_1} et w_{k_2} de $\mathcal{L}(E, E^*)$ se décompose (p. 86) en fonction des distances entre les vecteurs $w_k(e_i)$ de E^* :

$$\| w_{k_1} - w_{k_2} \|^2_{t\mathcal{L}(E,E^*)} = \sum_{i=1}^{n} p_i \| w_{k_1}(e_i) - w_{k_2}(e_i) \|^2_d \quad .$$

Relation entre $\| w_{k_1}(e_i) - w_{k_2}(e_i) \|^2_d$ et la distance entre les points $M_i^{k_1}$ et $M_i^{k_2}$. Exprimons le vecteur $w_k(e_i)$ dans la base $(\varepsilon^*_1, \ldots, \varepsilon^*_n)$ d-orthonormée, formée de vecteurs propres de wd. Alors :

$$\| w_{k_1}(e_i) - w_{k_2}(e_i) \|^2_d = \sum_{\ell=1}^{n} ((w_{k_1} - w_{k_2})(e_i) | \varepsilon^*_\ell)^2_d \quad .$$

Si on se limite aux ρ premières coordonnées, on remplace le vecteur $w_k(e_i)$ par le vecteur projeté $\sum_{\ell=1}^{\rho} (w_k(e_i) | \varepsilon^*_\ell)_d \, \varepsilon^*_\ell$, et le rapport de leurs normes donne une mesure de l'erreur commise quand on raisonne sur le vecteur projeté.

Or, les coordonnées des points M_i^k des trajectoires sont, au coefficient de dilatation $\lambda_\ell^{-1/2}$ près, les composantes du vecteur $w_k(e_i)$ sur $\varepsilon_1^*,\ldots,\varepsilon_\rho^*$. Les trajectoires permettent de détecter quels sont les individus i pour lesquels la distance entre les points $M_i^{k_1}$ et $M_i^{k_2}$ est grande. Puisque cette distance reflète la norme $\| w_{k_1}(e_i) - w_{k_2}(e_i) \|_d$, ce sont ces individus là qui sont responsables de l'écart entre w_{k_1} et w_{k_2} .

Remarque. Les coefficients de dilatation $\lambda_\ell^{-1/2}$ étant dans l'ordre croissant, les écarts entre les coordonnées des points $M_i^{k_1}$ et $M_i^{k_2}$ reflètent de plus en plus exagérément les projections de $(w_{k_1} - w_{k_2})(e_i)$ sur les axes. C'est la raison pour laquelle on se limitera à l'examen des trajectoires dans les deux premiers axes ou, s'il n'y a pas une grande différence entre λ_2 et λ_3, dans les trois premiers axes.

5.3. INTERPRETATION DU SENS DE PARCOURS D'UNE TRAJECTOIRE, EN FONCTION DES VARIABLES

On peut interpréter de façon plus détaillée les trajectoires des individus, lorsque les points se regroupent nettement par variable, sur le graphique des corrélations des variables avec les axes du compromis. Ce cas se rencontre fréquemment, car les variables sont souvent fortement auto-corrélées dans les différentes études. Il est alors facile, en général, d'expliquer les axes du compromis en fonction des variables.

Si, de plus, les applications linéaires w_k sont proches du compromis dans l'interstructure, alors on peut interpréter le sens de parcours d'une trajectoire le long d'un axe dont on connaît la signification.

A l'opposé, lorsque les corrélations à l'intérieur d'une étude sont fortes, on constate que les points du graphique des corrélations des variables avec les axes du compromis, se regroupent plutôt par étude que par variable. Il n'est alors plus possible de décrire les axes en fonction des variables, ni d'interpréter les trajectoires.

Nous allons montrer comment interpréter le déplacement d'une trajectoire, en projection sur un axe, en raisonnant pour simplifier sur le premier axe.

Interprétation des positions individuelles M_i^k de l'étude k, en projection sur un axe. Si wd et $w_k d$ ont en commun un vecteur propre associé à la plus grande valeur propre, alors, en projection sur le premier axe, la répartition des points M_i^k a la même interprétation que la répartition des points compromis.

□ *Démonstration.* $(\varepsilon_k^*)_\ell$ désigne un vecteur propre de $w_k d$, associé à la valeur propre $\lambda_\ell (w_k d)$. En utilisant la décomposition en valeurs singulières de w_k, $w_k(e_i)$ s'écrit :

$$w_k(e_i) = \sum_\ell \lambda_\ell (w_k d) \, \langle e_i | (\varepsilon_k^*)_\ell \rangle \, (\varepsilon_k^*)_\ell \quad .$$

Reportant cette expression dans la première coordonnée du point M_i^k, on obtient :

$$\lambda_1^{-1/2} \, (w_k(e_i) | \varepsilon_1^*)_d = \lambda_1^{-1/2} \sum_\ell \lambda_\ell (w_k d) \, \langle e_i | (\varepsilon_k^*)_\ell \rangle \, ((\varepsilon_k^*)_\ell | \varepsilon_1^*)_d$$

qui se simplifie, lorsque $(\varepsilon_k^*)_1 = \varepsilon_1^*$, en :

$$\lambda_1^{-1/2} \, \lambda_1 (w_k d) \, \langle e_i | \varepsilon_1^* \rangle \quad .$$

On peut considérer que les coordonnées des points M_i^k sur le premier axe, pour $i \in \{1,\ldots,n\}$, sont les valeurs d'une variable, à laquelle on associe l'élément y_k de E^*. Cette variable est centrée. Sous l'hypothèse $(\varepsilon_k^*)_1 = \varepsilon_1^*$, sa norme $\| y_k \|_d$ est égale à $\lambda_1^{-1/2} \, \lambda_1 (w_k d)$, et sa corrélation avec n'importe quelle variable x (centrée et réduite) est égale à :

$$\frac{1}{\| y_k \|_d} \, (y_k | x)_d = \sum_{i=1}^{n} p_i \, \langle e_i | \varepsilon_1^* \rangle \, \langle e_i | x \rangle = (\varepsilon_1^* | x)_d \quad .$$

Or $(\varepsilon_1^* | x)_d$ est la corrélation de la première composante principale de l'opérateur wd avec la variable x □

Remarque. En pratique, lorsque la distance entre les applications linéaires w et w_k est faible, on constate que les vecteurs propres de wd et de $w_k d$, associés à la plus grande valeur propre, sont en général voisins. Ceci est encore souvent vrai pour les vecteurs propres associés à la deuxième valeur propre, moins souvent pour les vecteurs propres associés à la troisième valeur propre (p. 166).

Si deux études k_1 et k_2 sont proches du compromis, et si les deux plus grandes valeurs propres des opérateurs $w_{k_1} d$ et $w_{k_2} d$ sont du même ordre de grandeur, alors les projections des points $M_i^{k_1}$ et $M_i^{k_2}$ dans le plan, représentent ce qu'on aurait obtenu en superposant les plans principaux des deux études.

Explication de l'écart entre $M_i^{k_1}$ et $M_i^{k_2}$, en projection sur un axe. Plaçons-nous dans le cas où l'explication du premier axe est stable d'une étude à l'autre. Cela suppose, en particulier, que les variables qui expliquent cet axe sont les mêmes pour les études k_1 et k_2 ou, du moins, qu'elles décrivent plusieurs aspects d'un critère général qui permettra de donner un nom à l'axe, qui soit le même pour les deux études.

Si $w_{k_1} d$, $w_{k_2} d$ et wd ont en commun un vecteur propre associé à la plus grande valeur propre, et si $\lambda_1(w_{k_1} d)$ est du même ordre de grandeur que $\lambda_1(w_{k_2} d)$, alors on peut expliquer le déplacement des trajectoires entre k_1 et k_2, le long du premier axe.

Supposons que les points $M_i^{k_1}$ et $M_i^{k_2}$ se projettent loin l'un de l'autre sur l'axe. Cela reflète une variation importante des valeurs des variables qui expliquent l'axe, pour l'individu i entre k_1 et k_2. Mais cette variation doit être interprétée par rapport à la variation de la moyenne de ces variables.

□ *Démonstration.* Supposons que le premier axe soit expliqué par le développement économique (comme dans l'exemple traité au chapitre VII). Appelons x_{k_1}, l'élément de E^* dont les coordonnées dans la base $(e_1^*,\ldots,e_n^*)$ sont les valeurs de la variable développement, centrée et réduite, sur les n individus dans l'étude k_1, et x_{k_2} l'élément correspondant à la variable développement dans l'étude k_2.

Soient y_{k_1} et y_{k_2}, les éléments de E^*, définis comme dans la démonstration précédente, par les coordonnées des points $M_i^{k_1}$ et $M_i^{k_2}$ sur le premier axe.

L'hypothèse de stabilité de l'explication du premier axe s'écrit : $(\varepsilon_1^*|x_{k_1})_d = (\varepsilon_1^*|x_{k_2})_d$. L'égalité des vecteurs propres de wd, $w_{k_1}d$ et $w_{k_2}d$, associés à la plus grande valeur propre, entraîne que la corrélation entre y_{k_1} et x_{k_1} (égale à $(\varepsilon_1^*|x_{k_1})_d$) est égale à la corrélation entre y_{k_2} et x_{k_2} (égale à $(\varepsilon_1^*|x_{k_2})_d$). Enfin, l'égalité des valeurs propres $\lambda_1(w_{k_1}d)$ et $\lambda_1(w_{k_2}d)$ implique que la dispersion des points $M_i^{k_1}$ sur l'axe, est comparable à celle des points $M_i^{k_2}$ □

Exemple. Illustrons ce dernier point avec l'exemple des médecins, en raisonnant sur un axe d'appréciation des médicaments selon un certain critère (les médicaments bien notés étant du côté positif de l'axe, par exemple). Le déplacement d'un médicament entre k_1 et k_2 vers le côté positif de l'axe, traduit un jugement plus favorable de la part du médecin k_2, compte tenu de l'habitude de notation (note moyenne) de chaque médecin. En effet, si le médecin k_2 est, dans l'ensemble, plus sévère que le médecin k_1, le déplacement de la trajectoire peut provenir de deux notes, en réalité, voisines.

Cas des objets normés. On désigne encore par $w = \Sigma\, \alpha_k \dfrac{w_k}{\|w_k\|_t}$, le compromis calculé à partir des objets normés. L'explication de l'écart entre $M_i^{k_1}$ et $M_i^{k_2}$, en projection sur le premier axe, sera possible si wd, $\dfrac{w_{k_1}d}{\|w_{k_1}\|_t}$ et $\dfrac{w_{k_2}d}{\|w_{k_2}\|_t}$ (ou, ce qui revient au même, si wd, $w_{k_1}d$ et $w_{k_2}d$) ont le même vecteur propre associé à la plus grande valeur propre, et si $\dfrac{\lambda_1(w_{k_1}d)}{\|w_{k_1}\|_t}$ est du même ordre de grandeur que $\dfrac{\lambda_1(w_{k_2}d)}{\|w_{k_2}\|_t}$.

Lorsque ces conditions sont réalisées pour les vecteurs propres associés aux deux plus grandes valeurs propres, les projections des points $M_i^{k_1}$ et $M_i^{k_2}$ sur le plan, représentent ce qu'on aurait obtenu en superposant les plans principaux des études $\left(\frac{X_{k_1}}{\| w_{k_1} \|_t}, q_{k_1}, d\right)$ et $\left(\frac{X_{k_2}}{\| w_{k_2} \|_t}, q_{k_2}, d\right)$.

Interprétation des points de la trajectoire, par rapport à la position compromis. Lorsque l'explication d'un axe est commune à toutes les études, on peut interpréter les positions des différents points d'une trajectoire, par rapport à la position compromis. Dans l'exemple des médecins, si les points de la "trajectoire" d'un médicament et la position compromis se projettent au même endroit sur l'axe, cela signifie que les jugements des médecins, selon les critères qui expliquent l'axe, concordent pour ce médicament. Dans le cas contraire, des écarts importants entre les points de la trajectoire montrent un désaccord des médecins sur ce médicament, toujours selon les critères qui expliquent l'axe.

6. REPRESENTATION DES VARIABLES, DANS LE CAS D'ETUDES (X_k, r, d_k)

6.1. IMAGE EUCLIDIENNE COMPROMIS DES VARIABLES

Rappel sur le compromis. On est dans la situation 2 : on compare les études (X_k, q, d_k), portant sur les mêmes variables, au moyen des objets v_k. Le compromis $v = \Sigma\, \alpha_k\, v_k$ est une application linéaire de (F^*, q^*) dans (F, q), associée à une forme bilinéaire, symétrique, positive sur F^*. La matrice V du compromis dans les bases $(f_1^*, \ldots, f_m^*)$ r^*-orthonormée de F^*, et $(f_1, \ldots, f_m)$ r-orthonormée de F, peut être assimilée à une matrice de variances entre les variables. En effet, puisque les variables sont centrées par étude, et que $\Sigma_k\, \alpha_k$ est égal à 1, l'élément :

$$V_{ij} = \langle v(f_j^*) \mid f_i^* \rangle = \sum_k \alpha_k \left((x_k)_i \mid (x_k)_j \right) d_k \quad ,$$

peut être considéré comme une "covariance compromis" entre les variables i et j. C'est la covariance entre les variables i et j, mesurées sur la totalité des individus, alors que les poids des individus de l'étude k sont multipliés par α_k .

Remarque. La matrice V représente la matrice de variances qu'on aurait obtenue à partir du tableau de données, construit en empilant les tableaux $\alpha_k^{1/2} X_k$.

Image euclidienne compromis. v étant une forme bilinéaire, symétrique, et positive, sur l'espace vectoriel euclidien (F^*, q^*), on sait construire (p. 52) une image euclidienne des variables, à partir de la décomposition en valeurs singulières de v. Soit O, $N_1,\ldots,N_m$ cette image euclidienne. Le produit scalaire entre les vecteurs ON_i et ON_j est égal à la covariance compromis entre les variables i et j. Le cosinus de l'angle des vecteurs ON_i et ON_j représente la corrélation compromis entre les variables i et j.

Image euclidienne approchée. En ne prenant que les vecteurs propres de vq, associés aux h plus grandes valeurs propres, on obtient une image euclidienne des variables, associée à l'approximation d'ordre h des covariances compromis.

Mais cette image, ainsi que la mesure de l'approximation, dépendent du produit scalaire q. L'image euclidienne approchée ne peut pas être interprétée lorsque les termes $(f_i|f_j)_q$, pour $i \neq j$, ne sont pas nuls. Et même lorsque ces termes sont nuls, l'interprétation reste délicate car la mesure de l'approximation fait intervenir les quantités $(f_j|f_j)_q$ qui ne représentent pas des poids. On suppose donc, à partir de maintenant, que le produit scalaire q est égal à r .

Alors la mesure de la qualité de l'approximation s'exprime en fonction des écarts entre la matrice de variances compromis V, et son approximation V_h d'ordre h :

$$\| v-v_h \|^2_{\mathcal{L}(F^*,F)} = \sum_{i=1}^{m}\sum_{j=1}^{m} (V-V_h)^2_{ij} = \sum_{\ell=h+1}^{\rho} \lambda^2_\ell(vr) ,$$

où ρ désigne le rang de v.

Enfin, si les variances compromis ne sont pas du même ordre de grandeur, l'image euclidienne approchée restitue essentiellement les variables dont la variance est importante. Il est préférable, dans ce cas, de représenter les corrélations compromis. Ce qui revient à calculer le compromis $v = \Sigma\, \alpha_k\, v_k$, à partir des matrices de corrélation v_k, c'est-à-dire à centrer et réduire les variables de chaque étude (X_k, r, d_k).

6.2. TRAJECTOIRES DES VARIABLES

Construction d'une trajectoire. $\phi_1, \ldots, \phi_\rho$ sont des vecteurs propres r-orthonormés de vr, associés aux valeurs propres non nulles $\lambda_1 \geqslant \ldots \geqslant \lambda_\rho$. La $\ell^{\text{ième}}$ coordonnée du point N_j de l'image euclidienne compromis, dans le repère $(O, \mathcal{B})$ est égale à :

$$\lambda_\ell^{1/2} < \phi_\ell \mid f_j^* > = \lambda_\ell^{-1/2} (\phi_\ell \mid v(f_j^*))_r \quad .$$

Dans l'image euclidienne compromis, on représente la variable j, mesurée dans l'étude k (y compris une étude supplémentaire) par le point N_j^k, dont la $\ell^{\text{ième}}$ coordonnée dans le repère $(O, \mathcal{B})$ est égale à :

$$\lambda_\ell^{-1/2} (\phi_\ell \mid v_k(f_j^*))_r \quad .$$

Position compromis par rapport aux points de la trajectoire. Le point compromis N_j est le centre de gravité des différents points N_j^k, pondérés par les α_k.

Décomposition de la distance entre les objets v_k. La distance entre les applications linéaires v_{k_1} et v_{k_2} de $\mathcal{L}(F^*, F)$ se décompose (p. 89) en fonction des distances entre les vecteurs $v_k(f_j^*)$ de F :

$$\| v_{k_1} - v_{k_2} \|^2_{\mathcal{L}(F^*,F)} = \sum_{j=1}^{m} \| v_{k_1}(f_j^*) - v_{k_2}(f_j^*) \|^2_r \quad .$$

Relation entre $\| v_{k_1}(f_j^*) - v_{k_2}(f_j^*) \|_r^2$ et la distance entre les points $N_j^{k_1}$ et $N_j^{k_2}$. Exprimons le vecteur $v_k(f_j^*)$ dans la base $(\phi_1,\ldots,\phi_m)$ r-orthonormée, formée de vecteurs propres de vr. Alors :

$$\| v_{k_1}(f_j^*) - v_{k_2}(f_j^*) \|_r^2 = \sum_{\ell=1}^{m} ((v_{k_1} - v_{k_2})(f_j^*) \mid \phi_\ell)_r^2 \quad .$$

Les coordonnées des points N_j^k des trajectoires sont, au coefficient de dilatation $\lambda_\ell^{-1/2}$ près, les composantes du vecteur $v_k(f_j^*)$ sur $\phi_1,\ldots,\phi_\rho$. Les trajectoires permettent de détecter quelles sont les variables j pour lesquelles la distance entre les points $N_j^{k_1}$ et $N_j^{k_2}$ est grande, et qui sont responsables de l'écart entre v_{k_1} et v_{k_2}.

Comparaison des matrices de corrélations. Supposons que les variables de chaque étude soient centrées et réduites. Alors si deux études k_1 et k_2 sont proches du compromis, et si les deux plus grandes valeurs propres des opérateurs $v_{k_1} r$ (ou $w_{k_1} d_{k_1}$) et $v_{k_2} r$ sont du même ordre de grandeur, on peut considérer que les projections des points $N_j^{k_1}$ et $N_j^{k_2}$ sur le plan, représentent ce qu'on aurait obtenu en superposant les cercles de corrélations des deux études.

V. Programmation de la méthode STATIS

La méthode STATIS est simple à mettre en oeuvre. Les seuls calculs nécessaires sont la multiplication de matrices, le calcul de la trace d'une matrice carrée, et la diagonalisation d'une matrice symétrique. Si on dispose d'un langage matriciel (comme SAS IML, GAUSS ou APL), il suffit d'écrire une vingtaine d'instructions pour obtenir rapidement les résultats essentiels sur un petit jeu de données.
Par contre, si les données sont volumineuses, il devient nécessaire d'organiser la sortie des graphiques pour éviter la superposition des trajectoires, et pouvoir les lire.

Nous avons détaillé les calculs matriciels nécessaires à la programmation de la méthode dans ce chapitre, et donné les programmes correspondants en annexe (p. 203 et suivantes). Ces programmes, écrits en Turbo Pascal, poursuivent essentiellement un but pédagogique : chaque programme effectue une tâche élémentaire, et conserve les résultats dans un fichier "texte", lisible à l'écran. Les graphiques sont rudimentaires car volontairement programmés en mode graphique simple. Ces deux derniers points demandent à être améliorés si on veut articuler ces programmes en un logiciel performant.

Par ailleurs, nous proposons au lecteur un logiciel disponible sur micro-ordinateur de type IBM PC mais également sur gros système, dont nous décrivons les fonctions p. 120.

Notation matricielle. Dans ce chapitre, les expressions algébriques du chapitre précédent sont écrites sous forme matricielle. La matrice de l'application linéaire a est désignée par la lettre majuscule A. Un vecteur x est noté matriciellement $\mathbf{x}$ en tant que vecteur colonne, et $\mathbf{x}'$ en tant que vecteur ligne. La matrice identité de dimension k est notée I_k.

1. DONNEES RECUEILLIES SUR LES MEMES INDIVIDUS

1.1. ELEMENTS DE L'ETUDE (X_k, Q_k, D)

Tableaux de données. p désigne le nombre de tableaux, n le nombre d'individus, m_k le nombre de variables du $k^{ième}$ tableau. $p_1,\ldots,p_n$ sont les poids non nuls des individus, rappelons que $p_1 + \ldots + p_n = 1$. Le tableau de données X_k comporte n lignes et m_k colonnes, et contient les valeurs des variables **centrées** pour les poids des individus :

$$\forall\, j \in \{1,\ldots,m_k\} \qquad \sum_{i=1}^{n} p_i \, (X_k)_{ij} = 0 \quad .$$

Expression matricielle du produit scalaire de deux vecteurs. Soit (F, q) un espace vectoriel euclidien, rapporté à la base $(f_1,\ldots,f_m)$. Notons Q la matrice du produit scalaire q, dans les bases $(f_1,\ldots,f_m)$ de F et $(f_1^*,\ldots,f_m^*)$ de F^*. Le produit scalaire des vecteurs x et y s'écrit matriciellement $(x|y)_q = \mathbf{x}'\, Q\, \mathbf{y}$, et le carré de la distance entre x et y est égal à $\| x - y \|_q^2 = (\mathbf{x} - \mathbf{y})'\, Q\, (\mathbf{x} - \mathbf{y})$.

Si la base $(f_1,\ldots,f_m)$ est q -orthonormée, Q est la matrice identité, et l'expression matricielle du produit scalaire se réduit à $(x|y)_q = \mathbf{x}'\, \mathbf{y}$.

Objet W_k : tableau des produits scalaires entre individus. Si on prend pour distance entre individus la distance euclidienne classique, cela revient à utiliser le produit scalaire r_k sur l'espace vectoriel euclidien F_k des individus (p. 67). Puisque la base de référence est r_k-orthonormée, r_k a pour matrice la matrice identité I_{m_k} de dimension m_k. Les produits scalaires entre individus sont les éléments du tableau $W_k = X_k\, X_k'$, symétrique de dimension n.

Lorsqu'on utilise sur F_k un produit scalaire q_k, différent de r_k, il faut connaître sa matrice Q_k dans la base de référence et sa base duale. Alors le tableau des produits scalaires entre individus est le tableau $W_k = X_k\, Q_k\, X_k'$.

Matrice diagonale des poids des individus. Les poids $p_1,\ldots,p_n$ des individus définissent un produit scalaire d sur E^* (p. 69), dont la matrice D est diagonale :

$$\forall\ i \in \{1,\ldots,n\} \qquad D_{ii} = p_i \ .$$

Si les poids des individus sont égaux, $D = \frac{1}{n} I_n$.

1.2. CALCUL DES PRODUITS SCALAIRES ENTRE W_k

Tableau S des produits scalaires entre objets W_k. Le produit scalaire entre les objets w_{k_1} et w_{k_2} (p. 84) est égal à la somme des éléments diagonaux de la matrice $DW_{k_1}DW_{k_2}$:

$$(w_{k_1} \mid w_{k_2})_t = \mathrm{Tr}(DW_{k_1}DW_{k_2}) \ .$$

On range ces produits scalaires dans le tableau symétrique S de dimension p. Ainsi $S_{k_1k_2} = (w_{k_1} \mid w_{k_2})_t$.

Norme de W_k. $\| W_k \| = \| w_k \|_t = S_{kk}^{1/2}$.

Tableau des coefficients RV. A partir du tableau S, on calcule les coefficients RV :

$$\forall\ k_1,\ \forall\ k_2 \in \{1,\ldots,p\} \qquad RV(k_1, k_2) = \frac{S_{k_1k_2}}{S_{k_1k_1}^{1/2} S_{k_2k_2}^{1/2}} \ .$$

Produits scalaires entre objets normés. Si on décide de caractériser une étude par l'objet normé $\dfrac{w_k}{\| w_k \|_t}$, le tableau S des produits scalaires entre objets normés (ou produits scalaires normés) est le tableau des coefficients RV.

2. IMAGE EUCLIDIENNE DES ETUDES, ASSOCIEE AUX PRODUITS SCALAIRES ENTRE W_k

2.1. COORDONNEES DES POINTS

Matrice diagonale des poids des études. Les poids $\Pi_1,\ldots,\Pi_p$ des études définissent un produit scalaire δ sur G^* (p. 90), dont la matrice Δ est diagonale :

$$\forall\ k \in \{1,\ldots,p\} \qquad \Delta_{kk} = \Pi_k \ .$$

On calcule les coordonnées des points M_k de l'image euclidienne des études, à partir des valeurs propres et des vecteurs propres, δ -orthonormés, de $S\Delta$.

Diagonalisation de $S\Delta$. Si les poids des études sont différents, $S\Delta$ n'est pas symétrique. Les vecteurs propres γ_ℓ, δ -orthonormés, de $S\Delta$ se déduisent des vecteurs propres $\mathbf{z}_\ell$, orthonormés, de la matrice symétrique $\Delta^{1/2}\, S\, \Delta^{1/2}$: $\gamma_\ell = \Delta^{-1/2}\, \mathbf{z}_\ell$.

□ *Démonstration.* La matrice $\Delta^{1/2}\, S\, \Delta^{1/2}$ a pour terme général $S_{k_1k_2}\, \Pi_{k_1}^{1/2}\, \Pi_{k_2}^{1/2}$. Soit $\mathbf{z}$ un vecteur propre normé de $\Delta^{1/2}\, S\, \Delta^{1/2}$, associé à la valeur propre λ . $\mathbf{z}$ vérifie $\Delta^{1/2}\, S\, \Delta^{1/2}\, \mathbf{z} = \lambda\, \mathbf{z}$, et $\mathbf{z}'\mathbf{z} = 1$. En multipliant les deux membres par $\Delta^{-1/2}$ (matrice diagonale de terme général $\Pi_k^{-1/2}$), on obtient $S\Delta\,(\Delta^{-1/2}\, \mathbf{z}) = \lambda\,(\Delta^{-1/2}\, \mathbf{z})$. Par conséquent, $\gamma = \Delta^{-1/2}\, \mathbf{z}$ est vecteur propre de $S\Delta$, associé à la valeur propre λ . On vérifie que la δ -norme de γ est égale à 1, et que deux vecteurs propres γ_1 et γ_2, associés à deux valeurs propres distinctes, sont δ -orthogonaux. En effet $\gamma_1'\, \Delta\, \gamma_2 = (\Delta^{-1/2}\, \mathbf{z}_1)'\, \Delta\, (\Delta^{-1/2}\, \mathbf{z}_2) = \mathbf{z}_1'\, \mathbf{z}_2$ □

Remarque. Si les valeurs propres sont très voisines, les vecteurs propres ne sont pas stables (GOLUB - VAN LOAN (1983)). Il faut se méfier des vecteurs propres fantaisistes que les algorithmes peuvent fournir dans ce cas-là.

Construction de l'image euclidienne (p. 53). Les coordonnées des points M_k sur l'axe ℓ , sont les composantes du vecteur $\lambda_\ell^{1/2} \gamma_\ell = \lambda_\ell^{-1/2} S \Delta \gamma_\ell$.

Notation matricielle des éléments propres de S Δ . Soit h le nombre d'axes qu'on veut représenter. Les h plus grandes valeurs propres de S Δ, rangées dans l'ordre décroissant, définissent la matrice diagonale Λ . Les vecteurs propres γ_ℓ correspondants sont les colonnes de la matrice Γ . Alors les coordonnées des points M_k sont les éléments de la colonne ℓ de la matrice $\Gamma \Lambda^{1/2} = S \Delta \Gamma \Lambda^{-1/2}$.

Poids des études identiques. Si les poids Π_k sont égaux, il suffit de diagonaliser la matrice S pour obtenir les coordonnées des points M_k.

□ *Démonstration.* Soit **z** un vecteur propre normé de S, associé à λ (S). Alors $\Delta^{1/2} S \Delta^{1/2} = S \Delta = \frac{1}{p} S$ admet **z** comme vecteur propre normé, associé à la valeur propre $\lambda(S \Delta) = \frac{\lambda(S)}{p}$, et $\gamma = p^{1/2}$ **z** comme vecteur propre δ -normé.

Les coordonnées des points M_k sur l'axe ℓ sont les éléments du vecteur

$$\lambda_\ell^{1/2}(S \Delta) \gamma_\ell = \lambda_\ell^{1/2}(S \Delta) p^{1/2} \mathbf{z}_\ell = \lambda_\ell^{1/2}(S) \mathbf{z}_\ell \quad \square$$

2.2. IMAGE EUCLIDIENNE CENTREE

Centrage du tableau S. On note I la matrice identité, et **1** le vecteur colonne dont tous les éléments sont égaux à 1. A partir du tableau S des produits scalaires entre objets W_k, on calcule le tableau C, centré en lignes et en colonnes par rapport aux poids $\Pi_1, \ldots, \Pi_p$:

$$C = (I - \mathbf{1}\,\mathbf{1}' \Delta)\, S\, (I - \Delta\, \mathbf{1}\,\mathbf{1}') ,$$

I et **1** sont de dimension p.

Image euclidienne centrée. Comme précédemment, on calcule les coordonnées des points M_k de l'image euclidienne centrée, à partir des valeurs propres $\lambda_\ell(C \Delta)$ et des vecteurs propres χ_ℓ ,

δ-orthonormés de C Δ , déduits des valeurs propres et des vecteurs propres orthonormés de $\Delta^{1/2}$ C $\Delta^{1/2}$. Les coordonnées des points M_k sur l'axe ℓ sont les éléments du vecteur

$$\lambda_\ell^{1/2} (C\Delta)\ \chi_\ell = \lambda_\ell^{-1/2} (C\Delta)\ C\Delta\ \chi_\ell .$$

2.3. ETUDE SUPPLEMENTAIRE

Rappelons qu'une étude supplémentaire (X_{p+1}, Q_{p+1}, D) est une étude à laquelle on a affecté un poids nul.

Coordonnées d'une étude supplémentaire. On calcule les produits scalaires de W_{p+1} avec chacun des W_k, correspondant aux p études actives. Ces produits scalaires sont les $(g_k|g_{p+1})_{s_1} = \langle g_k|s_1(g_{p+1})\rangle$ de la p. 55 (avec les notations de la p. 90). On les range dans le vecteur colonne $\mathbf{s_1(g_{p+1})}$ de dimension p. L'étude supplémentaire p+1 est représentée dans l'image euclidienne des études actives, par le point M_{p+1} dont la coordonnée sur l'axe ℓ est égale à $\lambda_\ell^{-1/2}$ (SΔ) $\mathbf{s_1'\ (g_{p+1})}\ \Delta\ \gamma_\ell$.

Produits scalaires centrés. Soit S_1 le tableau des produits scalaires entre objets W_k pour toutes les études, actives et supplémentaires. A partir de S_1, on calcule le tableau C_1 centré pour les poids des études actives, par la même formule qu'au paragraphe précédent :

$$C_1 = (I - \mathbf{1}\ \mathbf{1}'\ \Delta_1)\ S_1\ (I - \Delta_1\ \mathbf{1}\ \mathbf{1}') ,$$

I et **1** ont pour dimension le nombre total d'études, et Δ_1 est la matrice diagonale des poids de l'ensemble des études. Soit $\mathbf{c_1(g_{p+1})}$ le vecteur colonne dont les éléments sont les produits scalaires centrés de W_{p+1} avec chacun des W_k actifs.

Coordonnée d'une étude supplémentaire dans l'image euclidienne centrée. Dans l'image euclidienne centrée des études actives, on représente l'étude supplémentaire p+1 par le point M_{p+1}, dont les coordonnées sont calculées de la même manière que précédemment. Les vecteurs propres, δ-orthonormés, de CΔ étant notés χ_ℓ , la coordonnée de M_{p+1} sur l'axe ℓ est égale à $\lambda_\ell^{-1/2}$(CΔ) $\mathbf{c_1'\ (g_{p+1})}\ \Delta\ \chi_\ell$.

3. POSITIONS COMPROMIS ET TRAJECTOIRES DES INDIVIDUS

3.1. POSITIONS COMPROMIS DES INDIVIDUS

Calcul de l'objet compromis. Les éléments du vecteur propre γ_1 de $S\Delta$, associé à la plus grande valeur propre λ_1 sont tous de même signe. On choisit pour vecteur propre γ_1 celui dont tous les éléments sont positifs. Les coefficients α_k du compromis $W = \sum_{k=1}^{p} \alpha_k W_k$ sont les composantes du vecteur $\alpha = \lambda_1^{-1/2} \left(\sum_{\ell=1}^{p} \Pi_\ell S_{\ell\ell}^{1/2} \right) \Delta \gamma_1$.

Cas d'objets normés.

$$\alpha = \lambda_1^{-1/2} \Delta \gamma_1 \quad , \quad \text{et} \quad W = \sum_{k=1}^{p} \alpha_k \frac{W_k}{\| W_k \|} .$$

Image euclidienne compromis. Comme pour l'image euclidienne des études, on déduit les valeurs propres λ_ℓ (WD) et les vecteurs propres ε_ℓ, d-orthonormés, de l'opérateur compromis WD, de la diagonalisation de la matrice symétrique $D^{1/2}\ W\ D^{1/2}$. Les coordonnées des positions compromis des individus sur l'axe ℓ sont les éléments du vecteur $\lambda_\ell^{1/2}(WD)\ \varepsilon_\ell = \lambda_\ell^{-1/2}(WD)\ WD\ \varepsilon_\ell$.

3.2. EXPLICATION DES AXES DU COMPROMIS PAR LES VARIABLES

Corrélations des variables avec les axes du compromis. Les variables du tableau X_k sont centrées. Pour calculer les corrélations de ces variables avec les axes du compromis, il faut les réduire. Soit Y_k le tableau des données centrées et réduites. Les corrélations de ces variables avec l'axe ℓ du compromis sont les éléments du vecteur $Y_k'\ D\ \varepsilon_\ell$.

On peut également calculer la corrélation des axes du compromis, avec n'importe quelle variable exogène x dont on connaît les valeurs sur les n individus. **x** désignant le vecteur colonne dont les éléments sont les valeurs centrées et réduites de la variable x, la corrélation de x avec l'axe ℓ est égale à $\mathbf{x}'\ D\ \varepsilon_\ell$.

Cas où les coefficients α_k du compromis sont égaux. Dans le cas d'études (X_k, I_{m_k}, D), X_k désignant le tableau des données centrées et réduites, supposons que les coefficients α_k du compromis soient égaux. Alors si on note X le tableau de données, obtenu par juxtaposition des tableaux $X_1,\ldots,X_p$, W est proportionnel à XX'. Les vecteurs propres d-orthonormés de WD sont vecteurs propres de XX'D, et le graphique des corrélations des variables avec les axes du compromis est le cercle des corrélations des variables de X (p. 72).

3.3. TRAJECTOIRES DES INDIVIDUS

Construction des trajectoires. Dans l'image euclidienne compromis, on représente les individus décrits par l'étude k, par les points M_i^k dont les coordonnées sur l'axe ℓ sont les éléments du vecteur $\lambda_\ell^{-1/2}(WD)\ W_k\ D\ \varepsilon_\ell$.

Trajectoire d'un individu supplémentaire. Pour chaque étude, on calcule les produits scalaires de l'individu n+1 avec chacun des n individus actifs. On range ces produits scalaires dans le vecteur colonne $\mathbf{w_k(e_{n+1})}$ de dimension n. Les points M_{n+1}^k, dont les coordonnées sur l'axe ℓ sont égales à $\lambda_\ell^{-1/2}\ (WD)\ \mathbf{w'_k(e_{n+1})}\ D\ \varepsilon_\ell$, pour $k \in \{1,\ldots,p\}$, forment la trajectoire de l'individu supplémentaire n+1.

Cas d'objets normés. Les coordonnées des points M_i^k sur l'axe ℓ sont les éléments du vecteur $\lambda_\ell^{-1/2}(WD)\dfrac{W_k}{\|W_k\|}\ D\ \varepsilon_\ell$

(ou $\lambda_\ell^{-1/2}(WD)\dfrac{\mathbf{w'_k(e_{n+1})}}{\|W_k\|}\ D\ \varepsilon_\ell$ pour un individu supplémentaire).

4. RESUME DES CALCULS DE BASE

On se place dans le cas le plus courant : distance euclidienne classique entre individus ($\forall\ k \in \{1,\ldots,p\}\ \ Q_k = I_{m_k}$), poids des individus égaux à $\frac{1}{n}$ ($D = \frac{1}{n} I_n$),

poids des études égaux à $\frac{1}{p}$ ($\Delta = \frac{1}{p} I_p$), pas d'étude ni d'individu supplémentaire.

Image euclidienne des études. A partir des tableaux de données centrées X_k, on calcule les élements de la matrice S des produits scalaires entre objets W_k :

$$\forall\, k \in \{1,\ldots,p\} \qquad W_k = X_k X_k' \quad ,$$

$$\forall\, k_1, \forall\, k_2 \in \{1,\ldots,p\} \qquad S_{k_1 k_2} = \frac{1}{n^2} \mathrm{Tr}(W_{k_1} W_{k_2}) \quad ,$$

$$RV(k_1, k_2) = \frac{S_{k_1 k_2}}{S_{k_1 k_1}^{1/2} S_{k_2 k_2}^{1/2}} \quad .$$

On calcule ensuite les h (en pratique 2 ou 3) plus grandes valeurs propres $\lambda_\ell(S)$ de S, et les vecteurs propres orthonormés $\mathbf{z}_\ell$ correspondants. Ces éléments propres fournissent les coordonnées des points de l'image euclidienne (qui sont les éléments du vecteur $\lambda_\ell^{1/2}(S)\ \mathbf{z}_\ell$ sur l'axe ℓ), et les coefficients α_k du compromis. En effet, puisque $\Delta = \frac{1}{p} I_p$, $\lambda_1(S\Delta) = \frac{1}{p} \lambda_1(S)$ et $\gamma_1 = p^{1/2}\ \mathbf{z}_1$, le vecteur $\alpha = \lambda_1^{-1/2}(S\Delta)\ \frac{1}{p} (\sum_\ell S_{\ell\ell}^{1/2})\ \frac{1}{p} \gamma_1$ qui donne les coefficients α_k du compromis, se simplifie en $\lambda_1^{-1/2}(S)\ \frac{1}{p} (\sum_\ell S_{\ell\ell}^{1/2})\ \mathbf{z}_1$.

Remarque. On peut prendre pour S, la matrice dont les éléments sont $S_{k_1 k_2} = \mathrm{Tr}(W_{k_1} W_{k_2})$. Cela ne modifie ni les coefficients RV, ni les coefficients α_k du compromis. Par contre, les valeurs propres étant multipliées par n^2, les coordonnées de l'image euclidienne sont multipliées par n.

Image euclidienne compromis et trajectoires. Les deux ou trois plus grandes valeurs propres de $W = \sum_k \alpha_k W_k$, et les vecteurs propres orthonormés $\mathbf{x}_\ell$ correspondants, fournissent les coordonnées des positions compromis des individus sur l'axe ℓ : ce sont les composantes du vecteur $\lambda_\ell^{1/2}(W)\ \mathbf{x}_\ell$ ($= \lambda_\ell^{-1/2}(W)\ W\ \mathbf{x}_\ell$).

Les corrélations de l'axe ℓ du compromis avec les variables centrées et réduites du tableau de données Y_k, sont les éléments du vecteur $n^{-1/2} Y'_k \mathbf{x}_\ell$ (puisque $\varepsilon_\ell = n^{1/2} \mathbf{x}_\ell$).

Dans l'image euclidienne compromis, on représente les individus décrits par l'étude k, par les points M_i^k dont les coordonnées sur l'axe ℓ sont les composantes du vecteur $\lambda^{-1/2}(W)\, W_k \mathbf{x}_\ell$.

5. DONNEES CONCERNANT LES MEMES VARIABLES

Tableaux de données. Les mêmes m variables ont été mesurées sur p groupes d'individus qui peuvent être différents. Pour les raisons exposées p. 107, on se place dans le cas d'études (X_k, I_m, D_k), où X_k désigne le tableau des données, centrées pour les poids des n_k individus de l'étude k.

Objet V_k : matrice de variances de l'étude k. L'étude k est caractérisée par l'objet $V_k = X'_k D_k X_k$ qui est la matrice de variances des variables. Lorsque les données sont centrées et réduites par étude, V_k est la matrice des corrélations.

Image euclidienne des études. La matrice S des produits scalaires entre objets V_k a pour terme général $S_{k_1 k_2} = \mathrm{Tr}(V_{k_1} V_{k_2})$. A partir de cette matrice S, on détermine de la même manière que pour les objets W_k, l'image euclidienne (éventuellement centrée) des études, ainsi que les coefficients du compromis.

Objet compromis. Soit γ_1 le vecteur propre de $S\Delta$ dont toutes les composantes sont positives, associé à la plus grande valeur propre. L'objet compromis est la matrice de variances compromis

$$V = \frac{\Sigma\, \alpha_k V_k}{\Sigma\, \alpha_k}$$

, dans laquelle les coefficients α_k sont les éléments du vecteur $\alpha = \Delta \gamma_1$.

Image euclidienne des variables, associée aux variances compromis. On déduit les coordonnées des points N_j de l'image euclidienne des variables, des valeurs propres λ_ℓ (V) et des vecteurs propres ϕ_ℓ , orthonormés (puisque Q = I), de la matrice compromis V. Les coordonnées des points N_j sur l'axe ℓ sont les éléments du vecteur $\lambda_\ell^{1/2}(V)\ \phi_\ell = \lambda_\ell^{-1/2}(V)\ V\ \phi_\ell$.

Dans cette image euclidienne, on représente les variables, d'après leurs mesures dans l'étude k, par les points N_j^k , dont les coordonnées sur l'axe ℓ sont les composantes du vecteur $\lambda_\ell^{-1/2}(V)\ V_k\phi_\ell$.

6. DESCRIPTION DU LOGICIEL STATIS

6.1. SUR QUEL MATERIEL PEUT-ON L'INSTALLER ?

Nous proposons un logiciel, exécutable sur micro-ordinateur IBM PC, ou compatible (*). Lorsque les données ont des dimensions raisonnables, il suffit d'appeler le nom du programme pour l'exécuter. Mais si le volume des données est important, il se peut que les dimensions standard soient insuffisantes. L'exécution du programme nécessite alors la compilation du programme principal, qui a pour unique fonction de gérer l'allocation dynamique de la mémoire. Ce programme appelle des programmes objets, écrits en Fortran 77 et compilés sous Microsoft version 3.2. L'édition de liens utilise les bibliothèques Fortran Microsoft : Math.Lib. et Fortran.Lib.

Il existe également une version pour gros système, utilisable sur IBM TSO et VM, MULTICS, CDC, VAX, UNIVAC, HP 9000, etc. Cette version fait partie de la bibliothèque modulaire de programmes d'analyse de données MODULAD (**) qui est implantée sur un grand nombre de centres de calcul.

(*) S'adresser à l'auteur, Unité de Biométrie, ENSA-INRA-USTL, Place Viala, 34060 MONTPELLIER (FRANCE), 67.61.24.22 - 67.61.24.28.

(**) commercialisée par l'INRIA, Domaine de Voluceau, Rocquencourt, BP 105, 78150 LE CHESNAY (FRANCE).

Le logiciel STATIS est adapté à la situation la plus courante, celle où les données sont recueillies sur les mêmes individus dans des circonstances différentes. Le logiciel STATIS DUALE s'applique lorsque ce sont les mêmes variables qui sont mesurées sur des groupes d'individus, éventuellement différents. Dans ce chapitre, nous décrivons les fonctions du logiciel STATIS.

6.2. CHOIX PROPOSES POUR L'UTILISATION DE LA METHODE

Nous avons fait le choix de ne proposer que les options les plus utilisées, pour ne pas alourdir le programme.

Individus supplémentaires. Les tableaux ont tous le même nombre de lignes. Chaque ligne correspond à un enregistrement contenant l'identificateur (jusqu'à 10 caractères), et les valeurs des variables d'un individu. On peut déclarer que certains individus sont supplémentaires en donnant un filtre, ou en demandant un tri sur la valeur d'un caractère ou d'une variable. Le programme rejette ces individus en dernières lignes.

La notion d'individus supplémentaires est utile pour comprendre l'influence d'une variable qualitative. Pour chaque modalité de la variable, on crée un individu moyen, à partir des individus correspondant à cette modalité. On compare ensuite les trajectoires de ces individus moyens.

On peut également, par ce biais, tracer les trajectoires d'individus dont on n'aurait pas les mesures pour les derniers tableaux. Il faut, auparavant, avoir déclaré ces individus supplémentaires et remplacé les lignes manquantes des derniers tableaux par des valeurs quelconques, par exemple zéro.

Variables exogènes. Dans chaque tableau, on peut déclarer que certaines variables n'interviendront pas dans le calcul des produits scalaires entre individus. Ces variables exogènes sont triées par les valeurs d'un filtre, et rejetées en dernières colonnes.

Tableaux supplémentaires. Certains tableaux peuvent être déclarés supplémentaires, toujours à l'aide d'un filtre.

Ces tableaux ne sont pas rejetés en fin de fichier car dans le cas de données évolutives, la place d'un tableau est importante dans la lecture des trajectoires.

Poids des individus et des tableaux. Les poids des individus actifs, peuvent être quelconques, mais les poids des tableaux actifs sont égaux.

Produits scalaires entre individus. A l'intérieur de chaque tableau, les variables sont automatiquement centrées pour les poids des individus. On peut les réduire sur demande. Ces transformations effectuées, la distance entre individus est la distance euclidienne classique.

En résumé, on travaille sur des études (X_k, I_{m_k}, D). La matrice des poids est $\Delta = \frac{1}{p} I_p$. Les variables de X_k sont éventuellement réduites.

Objet caractéristique d'une étude. Par défaut, on caractérise l'étude k par l'objet W_k et on utilise le produit scalaire $Tr(DW_{k_1} DW_{k_2})$. Si les normes $\| W_k \|$ sont trop différentes on peut, en option, caractériser l'étude k par l'objet normé $\frac{W_k}{\| W_k \|}$, et prendre pour produit scalaire le coefficient $RV(k_1, k_2)$.

6.3. MODALITES D'UTILISATION DU LOGICIEL

Séparation des traitements. Pour réduire la place mémoire nécessaire et simplifier son utilisation, le logiciel STATIS est formé de trois parties qu'on lance successivement. La première assure la gestion du fichier de données, effectue les calculs de l'interstructure et fournit l'image euclidienne des études. La deuxième décrit l'image euclidienne compromis, et la troisième trace les trajectoires.

Si par exemple, le nombre de tableaux est élevé, on est limité par la place mémoire dont on dispose, mais aussi par la lisibilité des résultats graphiques. On peut dans ce cas, ne pas être intéressé par le tracé des trajectoires, mais seulement par les similitudes ou oppositions entre tableaux, décrites par l'interstructure,

ou par l'image euclidienne compromis. Suivant les besoins, on arrêtera alors le traitement après la première ou la deuxième partie.

Fichier des paramètres. Nous avons préféré la création d'un fichier de paramètres à la solution des questions posées à l'écran, car le nombre d'informations à fournir est important. Mais nous avons simplifié au maximum la saisie de ces informations. Les paramètres sont définis à l'aide de mots-clés sans ordre, les options standard sont prises par défaut, et nous fournissons un fichier de paramètres modèle pour un jeu d'essai. Ces informations sont ensuite transmises d'une étape à l'autre, par des fichiers de travail.

Diagnostic des erreurs. Certaines erreurs provoquent l'arrêt du traitement, et un message en expliquant la cause. Le plus souvent, ces erreurs proviennent d'enregistrements non conformes dans le fichier des paramètres.

6.4. RESULTATS FOURNIS PAR LE LOGICIEL

Interstructure. La première partie restitue les données (sur demande), et les statistiques élémentaires de chaque tableau : moyenne (et éventuellement écart-type et corrélations) des variables. Ce qui permet de définir l'individu "moyen", dont l'évolution sert de référence pour l'interprétation des trajectoires. Sont ensuite listés, le tableau S des produits scalaires entre W_k, le tableau des coefficients RV, la combinaison linéaire définissant le compromis, et l'image euclidienne (éventuellement centrée) des études, associée aux produits scalaires entre W_k.

De plus, les huit premières coordonnées des points de l'image euclidienne sont stockées sur un fichier. Car si le nombre de tableaux est important, il peut être intéressant d'effectuer une classification des tableaux à partir de ce fichier.

Intrastructure. La deuxième partie fournit l'image euclidienne compromis des individus actifs, ainsi que le graphique des corrélations des variables (endogènes et exogènes) de chaque tableau, avec les axes du compromis.

Afin d'obtenir un graphique précis, on a la possibilité de tracer le premier axe (de plus grande dispersion) verticalement, et d'étirer le graphique en largeur sur plusieurs pages. Mais le graphique est alors encombrant... C'est pourquoi les coordonnées des points sont stockées sur un fichier, pour tracer le graphique sur un matériel plus performant dont on disposerait.

Tracé des trajectoires. La représentation des trajectoires des individus autour de leur point compromis, ne peut pas se faire par un graphique classique pour des raisons évidentes de non lisibilité. Réaliser le graphique en couleurs sur une table traçante ne le rend pas plus lisible pour autant car si les trajectoires sont nombreuses et enchevêtrées, l'oeil ne sait pas les "photographier". Il faut, par conséquent, trier les trajectoires.

CARLIER (1986) propose une classification automatique des trajectoires selon leur forme (la pente de la trajectoire entre deux points successifs) et leur amplitude (les coordonnées des points).

Nous avons choisi une autre optique, en offrant à l'utilisateur la possibilité de choisir le sous-ensemble d'individus dont il veut tracer les trajectoires, soit par tri sur une partie de l'identificateur, soit en donnant une liste de noms. On peut, de plus, imposer un nombre maximum de trajectoires par graphique. En combinant ces deux possibilités, on évite le croisement des trajectoires, et les graphiques dessinés sur table traçante peuvent être directement insérés dans une publication. Les figures des chapitres VI et VIII en sont une illustration.

Le cadre du graphique, calculé sur l'ensemble des trajectoires permet d'apprécier la situation du sous-ensemble demandé sur le plan factoriel, et de comparer ainsi les différents graphiques. La partie occupée par les trajectoires est alors agrandie automatiquement si nécessaire. Si les tableaux sont indicés par le temps, on peut développer les trajectoires, axe par axe, en fonction du temps. Ces graphiques sont plus faciles à interpréter que les graphiques des trajectoires dans le plan. Enfin, on peut tronquer la trajectoire en ne dessinant pas les points correspondant aux derniers tableaux. Ceci permet de tracer les trajectoires des individus absents des dernières études.

VI. Explication détaillée des résultats fournis par la méthode STATIS, appliquée à des données de croissance

1. DESCRIPTION DES DONNEES

Pour illustrer les résultats fournis par la méthode STATIS, nous allons utiliser des données séquentielles, extraites d'un fichier tout à fait exceptionnel, par la durée de l'enquête qu'il a nécessitée et par son contenu.

1.1. PRESENTATION RAPIDE DU FICHIER AUXOLOGIQUE FRANÇAIS

Le "fichier auxologique français" a été patiemment établi en 22 ans, de 1953 à 1975, par le suivi des *mêmes* enfants, de leur naissance à la fin de leur adolescence. L'acquisition des données a fait l'objet d'une enquête engagée à Paris, parallèlement à d'autres enquêtes semblables par leur méthodologie d'investigation, réalisées à Bruxelles, Londres, Stockholm, Zurich, Louisville et Kampala. L'ensemble de ces enquêtes fut coordonné par F. FALKNER, en collaboration avec le Centre International de l'Enfance. L'enquête française a été effectuée dans le cadre du Centre d'Etude de la Croissance et du Développement de l'Enfant - section française des études internationales coordonnées par le Centre International de l'Enfance - principalement par M.P. ROY-PERNOT jusqu'en 1959, puis par M. SEMPÉ jusqu'en 1975 (sous les auspices de l'Institut National de la Santé et de la Recherche Médicale à partir de 1965).

(*) Ce chapitre a été écrit en collaboration avec Marie-Odile PERNIN. Assistante, Laboratoire d'Analyse de Données et Biométrie, Université Claude Bernard. LYON.

Les critères de recrutement des enfants ont été les suivants : parents d'origine française et métropolitaine, habitant Paris ou sa proche banlieue à la naissance de l'enfant ; poids de naissance compris entre 2,5 kg et 4,7 kg ; absence d'anomalie dans le comportement des premiers jours, et de malformation décelable à la maternité. Ce recrutement s'est étalé sur sept ans. Mais sur un effectif de 588 enfants, 75 % ont été recrutés dans les deux premières années, durée assez courte pour que l'effet de la tendance séculaire puisse être considéré comme négligeable. Le rythme de surveillance des enfants fut rigoureusement respecté, les mesures ont été faites à intervalles de temps fixes, identiques pour tous les individus. Les techniques de mesures ont, elles aussi, été strictement définies.

L'historique de l'enquête, la description du fichier et son exploitation ont fait l'objet de nombreuses publications, dont on trouvera les références dans SEMPÉ (1979). En particulier, les courbes de poids, taille et périmètre crânien, figurant dans les carnets de santé, ont été établies à partir de ce fichier. Actuellement, l'exploitation de ces données se poursuit au sein du laboratoire d'analyse de données et biométrie, de l'Université Claude Bernard à Lyon (*).

Le fichier, directement issu de l'enquête est un fichier brut constitué de 588 dossiers individuels, chaque dossier comportant 131 enregistrements de 80 caractères chacun. L'utilisation de ce fichier nécessite son organisation en base de données, ce qui n'a été effectué que partiellement jusqu'à présent. Les données que nous traitons dans ce chapitre, sont extraites d'une base de données partielle, réalisée par Marie-Odile PERNIN, et exploitées dans sa thèse (1986). Signalons enfin que plusieurs méthodes ont été confrontées, sur ces mêmes données, lors du "fourth international symposium on data analysis" (1985), publié par Plenum en 1987.

SEMPÉ M., PEDRON G., ROY-PERNOT M.P. - Auxologie, Méthode et séquences, Laboratoire THERAPLIX, Paris, 1979.

(*) dirigé par Monsieur le Professeur Jacques PONTIER, 43, boulevard du 11 Novembre 1918, 69622 VILLEURBANNE.

1.2. EXTRAIT DU FICHIER EXPLOITE PAR LA METHODE STATIS ET OBJECTIF POURSUIVI

Nous reprenons une partie du travail présenté par M.O. PERNIN dans sa thèse, à laquelle nous renvoyons le lecteur intéressé par l'interprétation biologique des résultats.

Données. Les données décrivent la morphologie de l'enfant par le poids, la taille (taille de l'enfant couché), le segment supérieur ou buste (hauteur de l'enfant assis), les périmètres crânien, thoracique, du bras gauche et du mollet gauche, et le diamètre bi-iliaque (largeur du bassin). Nous disposons des mesures annuelles de ces huit paramètres, sur un échantillon de 30 filles, entre 4 et 15 ans.

Ce fichier de données, publié p. 150 (*), comporte 12 tableaux. A chaque âge k, correspond un tableau X_k, possédant 30 lignes et 8 colonnes. Chaque ligne contient les mesures des huit caractères somatiques retenus, pour un individu. Il n'y a pas de données manquantes.

Objectif. La croissance est un phénomène complexe dont nous n'observons ici que l'aspect morphologique. Est-il possible de caractériser les différents âges étudiés, en dehors du bouleversement bien connu de la puberté ? Peut-on mettre en évidence des types morphologiques caractéristiques ?

1.3. CHOIX EFFECTUES POUR UTILISER LA METHODE STATIS

Choix des produits scalaires sur les individus et sur les variables. Les individus sont affectés des mêmes poids. Les variables sont centrées et réduites à l'intérieur de chaque tableau.

(*) avec l'autorisation du Docteur Michel SEMPÉ, pédiatre, Maître de Recherche INSERM, Laboratoire d'Analyse de Données et Biométrie de l'Université Claude Bernard.

On a ainsi défini des études (X_k, r, d), (ou (X_k, I_8, $\frac{1}{30}$ I_{30}) avec les notations du chapitre V), dans lesquelles le tableau X_k est un tableau de mesures centrées et réduites. Les études ont toutes le même poids $\frac{1}{12}$.

Choix des coefficients d'association entre études (p. 83). Cette étude a deux objectifs : décrire l'évolution de chaque individu pour mettre en évidence des types morphologiques, mais aussi caractériser les différents âges. On s'intéresse donc a priori aux trajectoires des individus. Dans ce but, on choisit l'objet w_k pour représenter un âge, et on étudie ainsi l'évolution des relations inter-individuelles au cours des années.

Puisque les variables sont les mêmes pour tous les tableaux, on peut également envisager la comparaison des matrices de corrélations entre variables à chaque âge. Un âge, représenté par l'objet v_k, sera alors caractérisé par les relations linéaires existant entre les variables.

Enfin, les variables étant centrées et réduites par tableau, le carré de la norme $\| w_k \|$ (égale à $\| v_k \|$) est la somme des corrélations au carré des variables du tableau X_k (p. 87). Cette remarque nous amène à étudier d'abord les corrélations entre variables, puis les relations inter-individuelles.

2. ETUDE DES CORRELATIONS ENTRE VARIABLES

L'interstructure, réalisée sur les objets v_k, fournit des produits scalaires (et par conséquent des distances) entre matrices de corrélations, et un objet compromis qui peut être considéré comme une matrice de corrélations moyennes entre les variables. On représente ensuite la trajectoire de chaque variable dans le cercle des corrélations compromis.

2.1. INTERSTRUCTURE SUR LES OBJETS v_k

Les matrices de corrélations, reproduites p. 132 et 133, sont relativement stables au cours du temps.

Pour les applications linéaires v_k, cela se traduit par des normes $\| v_k \|$ du même ordre de grandeur (p. 130), et des produits scalaires normés $\frac{(v_{k_1} \mid v_{k_2})}{\| v_{k_1} \| \ \| v_{k_2} \|}$ voisins de 1. Rappelons qu'un produit scalaire normé égal à 1 indique que les matrices de corrélations sont homothétiques. Si de plus, leurs normes sont égales, alors les matrices de corrélations sont égales.

La distance entre les points M_{k_1} et M_{k_2} de l'image euclidienne des âges (centrée ou non), construite à partir des produits scalaires $(v_{k_1} | v_{k_2})$ est égale à :

$$d_{v_{k_1} v_{k_2}} = (\| v_{k_1} \|^2 + \| v_{k_2} \|^2 - 2 (v_{k_1} | v_{k_2}))^{1/2} \quad .$$

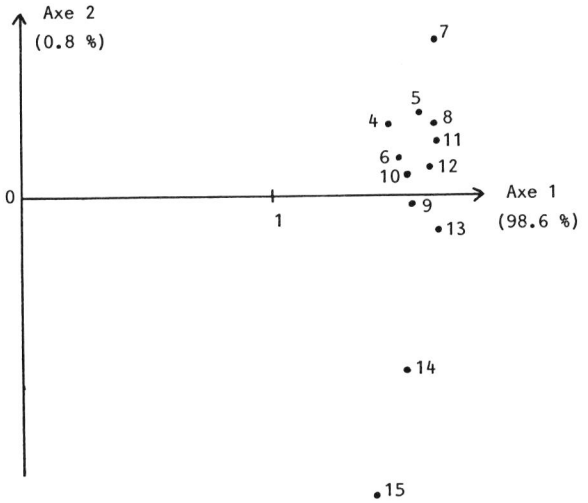
Mais, dans la pratique, on n'examine que des images euclidiennes approchées. Nous allons montrer en détail comment l'image euclidienne approchée (figure n° 1) restitue les produits scalaires normés, et comment l'image euclidienne centrée (figure n° 2) restitue les distances entre v_k.

Image euclidienne des âges, associée à l'approximation d'ordre 2 des produits scalaires $(v_{k_1} | v_{k_2})$ (figure n° 1). Les produits scalaires normés sont les cosinus des angles entre les vecteurs OM_k de l'image euclidienne des âges, associée aux produits scalaires $(v_{k_1} | v_{k_2})$. Les produits scalaires normés étant pratiquement tous voisins de 1, cela entraîne que le faisceau, formé par les vecteurs joignant l'origine O aux points 4 - 5 - 6 - 8 - 9 - 10 - 11 - 12 et 13 de la figure n° 1, est très resserré. C'est également la raison pour laquelle le pourcentage d'inertie (98.6 %) expliqué par le premier axe est si élevé.

Pour expliquer la position isolée de 7 ans, opposée à 14 et 15 ans sur le deuxième axe, on se reporte aux valeurs des produits scalaires normés. On constate que le produit scalaire normé entre un âge et l'âge suivant est en général très élevé. Plus deux âges sont éloignés, plus ce produit scalaire normé est faible. Cet écart, particulièrement marqué pour 14 et 15 ans, explique la position isolée de ces deux points sur la figure n° 1.

TABLEAU DES PRODUITS SCALAIRES NORMES : $\dfrac{(v_{k_1} \mid v_{k_2})}{\|v_{k_1}\| \, \|v_{k_2}\|}$

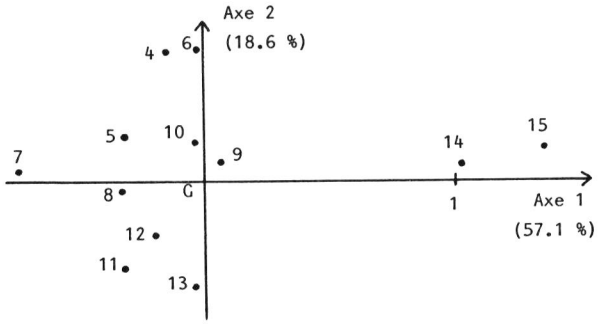
	4 ans	5	6	7	8	9	10	11	12	13	14
4 ans											
5	.996										
6	.991	.992									
7	.990	.994	.986								
8	.993	.997	.991	.996							
9	.987	.989	.990	.984	.992						
10	.988	.990	.990	.988	.994	.993					
11	.988	.993	.984	.991	.996	.990	.992				
12	.989	.994	.989	.990	.996	.992	.994	.996			
13	.985	.991	.986	.986	.994	.991	.990	.996	.996		
14	.974	.977	.977	.966	.981	.987	.981	.981	.983	.991	
15	.949	.950	.956	.932	.952	.961	.957	.953	.959	.969	.987



$\|v_k\| = \|w_k\|$

4 ans	5.12
5	5.40
6	5.15
7	5.63
8	5.52
9	5.29
10	5.25
11	5.62
12	5.56
13	5.62
14	5.30
15	4.93

Figure n° 1 : IMAGE EUCLIDIENNE DES AGES, ASSOCIEE A L'APPROXIMATION D'ORDRE 2 DES PRODUITS SCALAIRES $(v_{k_1} \mid v_{k_2})$



Coordonnées des points sur l'axe 3 (8.1 %)

4 ans	-0.23
5	-0.21
6	0.13
7	-0.16
8	-0.03
9	0.39
10	0.33
11	-0.02
12	0.06
13	-0.09
14	0.01
15	-0.17

Figure n° 2 : IMAGE EUCLIDIENNE CENTREE

La position isolée de 7 ans provient de ce que, contrairement à la tendance générale, le produit scalaire normé entre 6 et 7 ans est moins élevé qu'entre 6-8 ans, 6-9 ans et 6-10 ans. De même, 4-7 ans est moins fort que 4-8 ans. 7-11 ans et 7-12 ans sont plus élevés que 7-9 ans et 7-10 ans. Enfin, on remarque que 7-15 ans est la valeur la plus faible du tableau des produits scalaires normés.

De l'étude des produits scalaires normés, on conclut que les matrices de corrélations des âges 4 - 5 - 6 - 8 - 9 - 10 - 11 - 12 et 13 ans sont, dans l'ensemble, homothétiques. Par contre, les matrices de corrélations à 7 ans, et plus encore à 14-15 ans, sont différentes des autres.

Objet compromis v. Les coefficients de la combinaison linéaire donnant la matrice de corrélations compromis sont proportionnels aux abscisses des points de la figure n° 1.

4 ans	.080	8 ans	.086	12 ans	.087
5 ans	.084	9 ans	.082	13 ans	.088
6 ans	.080	10 ans	.082	14 ans	.082
7 ans	.087	11 ans	.088	15 ans	.075

L'objet compromis est situé sur le premier axe de l'image euclidienne des âges. On comprend bien, pourquoi il reflète essentiellement les âges qui sont groupés autour du premier axe.

Image euclidienne centrée (figure n° 2, qu'il faut imaginer dans l'espace). Pour avoir une idée plus précise des distances entre les matrices de corrélations, on centre (p. 62) le tableau des produits scalaires $(v_{k_1} | v_{k_2})$ par rapport aux poids des études. Les proximités et oppositions entre les points de l'image euclidienne centrée, reflètent les distances $d_{v_{k_1} v_{k_2}}$ et s'expliquent, puisque les produits scalaires normés sont relativement forts, essentiellement par les différences de normes.

Par exemple, la proximité des points 5 et 8 ans d'une part, 11-12 et 13 ans d'autre part, indique que les normes sont voisines et les matrices de corrélations semblables.

MATRICES DE CORRELATIONS AUX DIFFERENTS AGES

	POID	TAIL	SSUP	TETE	THOR	BRAS	MOLL
POID							
TAIL	.65					4 ans	
SSUP	.69	.77					
TETE	.67	.35	.52				
THOR	.83	.39	.44	.54			
BRAS	.77	.24	.36	.47	.77		
MOLL	.73	.53	.53	.41	.53	.45	
ILIA	.63	.69	.56	.47	.40	.44	.48
POID							
TAIL	.71					5 ans	
SSUP	.75	.86					
TETE	.68	.39	.51				
THOR	.79	.41	.51	.62			
BRAS	.75	.29	.41	.44	.76		
MOLL	.79	.51	.56	.48	.57	.63	
ILIA	.72	.66	.72	.57	.52	.48	.50
POID							
TAIL	.75					6 ans	
SSUP	.76	.91					
TETE	.67	.41	.56				
THOR	.69	.37	.30	.46			
BRAS	.61	.19	.15	.35	.70		
MOLL	.75	.48	.45	.48	.55	.49	
ILIA	.76	.65	.66	.48	.49	.48	.56
POID							
TAIL	.75					7 ans	
SSUP	.82	.93					
TETE	.69	.53	.66				
THOR	.78	.62	.63	.51			
BRAS	.76	.38	.51	.50	.73		
MOLL	.79	.56	.63	.59	.49	.60	
ILIA	.72	.68	.71	.59	.55	.44	.59
POID							
TAIL	.73					8 ans	
SSUP	.77	.90					
TETE	.68	.52	.62				
THOR	.80	.48	.53	.51			
BRAS	.79	.30	.39	.42	.81		
MOLL	.83	.52	.54	.51	.65	.70	
ILIA	.67	.65	.71	.59	.51	.43	.54
POID							
TAIL	.61					9 ans	
SSUP	.74	.89					
TETE	.64	.49	.64				
THOR	.72	.28	.37	.25			
BRAS	.79	.16	.33	.32	.74		
MOLL	.83	.46	.59	.40	.68	.68	
ILIA	.74	.61	.74	.58	.41	.52	.55
POID							
TAIL	.67					10 ans	
SSUP	.74	.90					
TETE	.56	.54	.59				
THOR	.81	.44	.52	.35			
BRAS	.74	.14	.28	.16	.76		
MOLL	.83	.46	.54	.45	.65	.60	
ILIA	.67	.64	.66	.53	.48	.34	.53

	POID	TAIL	SSUP	TETE	THOR	BRAS	MOLL
POID							
TAIL	.67					11 ans	
SSUP	.76	.91					
TETE	.55	.43	.59				
THOR	.89	.55	.62	.43			
BRAS	.83	.29	.44	.43	.82		
MOLL	.85	.43	.54	.46	.78	.78	
ILIA	.73	.65	.76	.55	.61	.52	.59
POID							
TAIL	.71					12 ans	
SSUP	.77	.94					
TETE	.61	.49	.59				
THOR	.89	.51	.57	.44			
BRAS	.78	.22	.30	.39	.81		
MOLL	.82	.44	.49	.39	.72	.73	
ILIA	.77	.65	.68	.73	.56	.52	.55
POID							
TAIL	.63					13 ans	
SSUP	.73	.92					
TETE	.61	.53	.62				
THOR	.88	.39	.54	.52			
BRAS	.82	.23	.38	.42	.86		
MOLL	.85	.36	.47	.48	.78	.82	
ILIA	.76	.57	.67	.63	.60	.62	.68
POID							
TAIL	.46					14 ans	
SSUP	.66	.87					
TETE	.64	.54	.71				
THOR	.84	.12	.34	.43			
BRAS	.79	.05	.29	.40	.86		
MOLL	.82	.22	.41	.51	.81	.81	
ILIA	.70	.45	.56	.58	.47	.50	.61
POID							
TAIL	.34					15 ans	
SSUP	.57	.85					
TETE	.64	.40	.65				
THOR	.80	.06	.34	.50			
BRAS	.76	-.22	.06	.35	.78		
MOLL	.76	.00	.21	.41	.69	.78	
ILIA	.67	.29	.39	.46	.47	.43	.58

CORRELATIONS COMPROMIS

	POID	TAIL	SSUP	TETE	THOR	BRAS	MOLL
POID							
TAIL	.64						
SSUP	.73	.89					
TETE	.64	.47	.60				
THOR	.81	.39	.48	.47			
BRAS	.77	.19	.33	.39	.78		
MOLL	.80	.42	.50	.46	.66	.67	
ILIA	.71	.60	.66	.57	.51	.48	.56

Les fluctuations entre 4 et 7 ans, ainsi que le rapprochement entre 7 et 8 ans, sont expliqués par la variation des valeurs absolues de certaines corrélations d'une année à l'autre. Par contre, on retrouve l'opposition sur le premier axe entre 7 ans et 15 ans, due au produit scalaire normé "faible" et à la différence de normes (5.63 et 4.93). Enfin, le rôle central du point 8 ans est expliqué par la valeur élevée des produits scalaires normés de cet âge avec les autres.

2.2. EVOLUTION DES CORRELATIONS ENTRE VARIABLES

Les trajectoires des variables sont dessinées sur la figure n° 3, et reflètent les différentes matrices de corrélations. Le point compromis ne fait pas partie de la trajectoire et n'est pas relié aux autres points. Pour ne pas alourdir le graphique, nous ne l'avons pas souligné. Si sa position est claire pour le périmètre du mollet par exemple, elle est nettement plus difficile à deviner pour les autres variables.

Nous avons remarqué que la matrice de corrélations compromis est un bon reflet des différentes matrices de corrélations (sauf à 15 ans). On peut donc considérer que la figure n° 3 est, à peu de choses près, ce qu'on obtiendrait si on superposait les cercles de corrélations des différents âges, sur le cercle des corrélations compromis.

D'une manière générale, les variables sont situées du côté négatif de l'axe 1, et se dispersent largement sur l'axe 2. On distingue cinq catégories de variables : le poids, les variables mesurant la longueur du squelette (taille et longueur du segment supérieur), sa largeur (diamètre bi-iliaque), le périmètre crânien et enfin les variables mesurant la corpulence (périmètres du bras, du mollet et du thorax).

Puisque nous utilisons ces données comme un exemple d'application de la méthode, nous allons montrer en détail comment les trajectoires reflètent les particularités des matrices de corrélations, avant de proposer une rapide explication biologique en conclusion.

Poids. Le poids joue un rôle particulier. La position centrale de sa trajectoire sur l'axe 2,montre que les corrélations de cette variable avec toutes les autres sont fortes et,stables au cours des années.

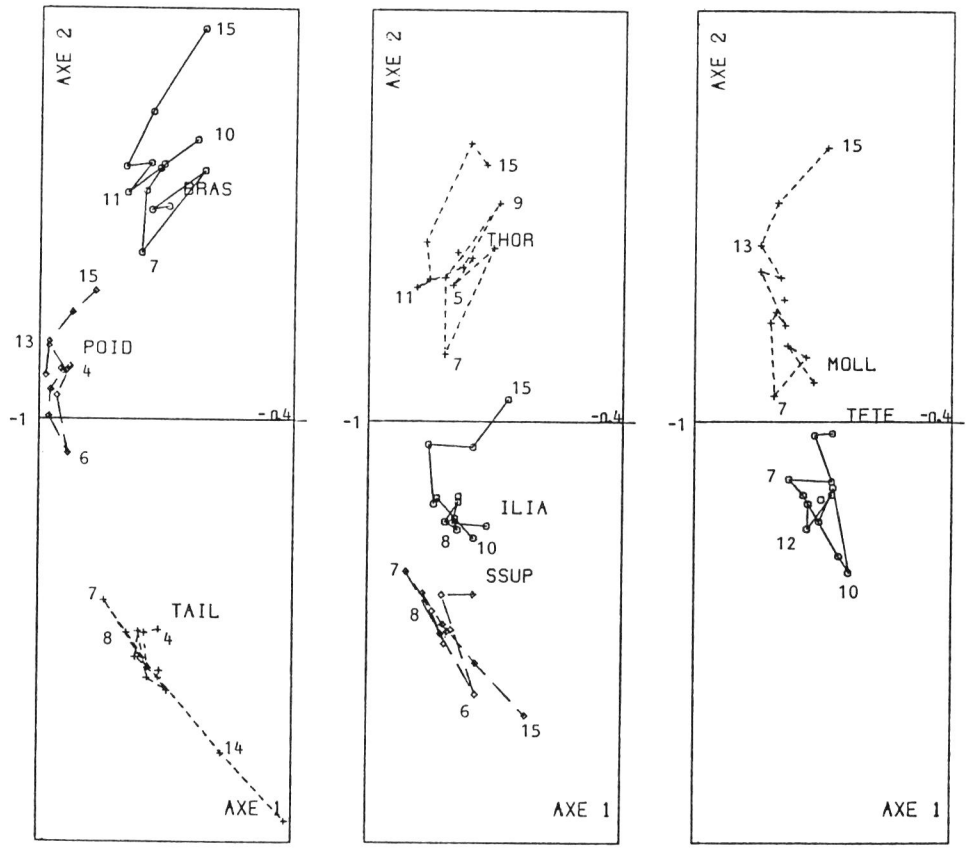


Figure n° 3 : TRAJECTOIRES DES VARIABLES

(AFIN D'EVITER LA SUPERPOSITION DES TRAJECTOIRES,
CE GRAPHIQUE EST ECLATE EN TROIS)

Toutefois, à partir de 13 ans, la trajectoire du poids s'oppose aux trajectoires des variables mesurant la longueur du squelette : les corrélations ont en effet tendance à décroître fortement.

Diamètre bi-iliaque et périmètre crânien. Les corrélations du diamètre bi-iliaque avec toutes les autres variables sont très stables. Toutefois la corrélation avec la taille chute brutalement à 15 ans, ce qui explique la remontée de la trajectoire.

Le périmètre crânien a des corrélations relativement stables avec les variables mesurant la longueur et la largeur du squelette, sauf à 12 ans où la corrélation avec le diamètre bi-iliaque est particulièrement élevée. Si cette particularité n'apparaît pas très nettement sur la trajectoire, on remarque par contre une plongée de la trajectoire du périmètre crânien à 10 ans, opposée à un retrait vers le haut de la trajectoire du périmètre du bras. En effet, la corrélation entre ces deux variables chute brutalement à 10 ans. Cet âge correspond également au minimum de la corrélation entre périmètre du bras et diamètre bi-iliaque.

Variables mesurant la corpulence. D'une manière générale, ces variables sont fortement corrélées entre elles. Ceci est très net pour les périmètres du bras et du thorax, dont les trajectoires se suivent.

Après quelques fluctuations jusqu'à 11 ans, la corrélation entre les périmètres du bras et du mollet a tendance à croître nettement.

La corrélation entre les périmètres du thorax et du mollet atteint un minimum à 7 ans, pour augmenter régulièrement ensuite. Le retour de la trajectoire du périmètre du thorax à 15 ans reflète la baisse de la corrélation entre périmètres du thorax et du bras d'une part, et périmètres du thorax et du mollet d'autre part.

Variables mesurant la longueur du squelette. Taille et longueur du segment supérieur sont toujours très corrélées et s'opposent aux variables mesurant la corpulence. Pour ces deux groupes de variables, les trajectoires changent brutalement de direction à 7 ans. Cet âge correspond en effet à un maximum dans l'évolution des corrélations.

Les corrélations de la taille et de la longueur du segment supérieur, avec les périmètres du thorax et du bras évoluent de façon semblable : elles fluctuent irrégulièrement jusqu'à 11 ans, avec un maximum à 7 et 11 ans, pour décroître ensuite, particulièrement à 14 et 15 ans. Toutefois la situation éloignée de la trajectoire du périmètre du bras indique que les corrélations des variables mesurant la longueur du squelette avec cette variable, sont toujours plus faibles qu'avec le périmètre du thorax.

Par contre, la trajectoire du périmètre du mollet évolue peu jusqu'à 7 ans et se situe au centre de l'axe 2. Cela reflète des corrélations stables et assez élevées avec les variables mesurant la longueur du squelette.

Remarquons enfin que la longueur du segment supérieur a une trajectoire qui se situe, dans l'ensemble, plus près de celles des variables de corpulence et du poids, que la taille. En effet les variables mesurant la corpulence sont toujours plus corrélées avec la longueur du segment supérieur qu'avec la taille, sauf à 6 ans, ce qui se traduit par une pointe de la trajectoire vers le bas.

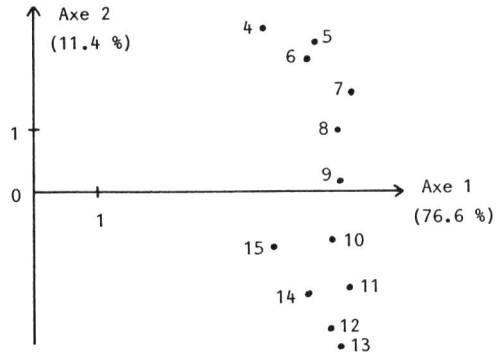
Dispersion de la trajectoire autour du point compromis. Lorsque les corrélations d'une variable avec les autres sont stables au cours des années, la trajectoire de cette variable est ramassée autour du point compromis. C'est le cas des variables mesurant la longueur et la largeur du squelette, ainsi que du poids, jusqu'à 13 ans. Ce n'est pas le cas, par contre, des variables mesurant la corpulence car leurs corrélations avec les autres variables fluctuent de façon irrégulière.

2.3. CONCLUSIONS DE L'ETUDE DES VARIABLES

L'étude des corrélations entre les huit paramètres retenus pour caractériser la morphologie des filles entre 4 et 15 ans, met en évidence une relative stabilité jusqu'à 13 ans. 14 et 15 ans se distinguent des âges précédents par une opposition très nette entre les paramètres mesurant la corpulence et ceux mesurant la taille du squelette. 8 ans est un âge "moyen", pour lequel la matrice de corrélations est un bon reflet de la période stable 4-13 ans.

TABLEAU DES PRODUITS SCALAIRES NORMES : $RV(k_1, k_2) = \dfrac{(w_{k_1} \mid w_{k_2})}{\| w_{k_1} \| \; \| w_{k_2} \|}$

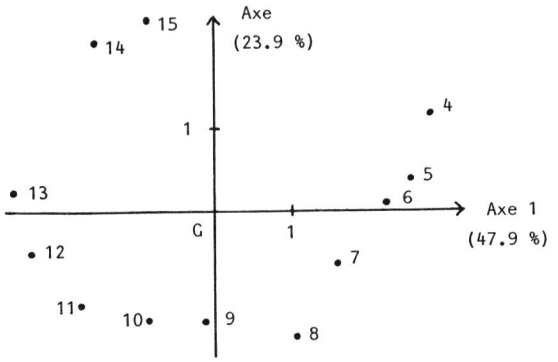
	4 ans	5	6	7	8	9	10	11	12	13	14
4 ans											
5	.893										
6	.864	.935									
7	.810	.923	.911								
8	.702	.815	.856	.909							
9	.638	.751	.809	.865	.928						
10	.558	.694	.719	.818	.845	.914					
11	.520	.634	.662	.753	.808	.851	.925				
12	.453	.568	.595	.696	.714	.794	.858	.910			
13	.478	.557	.570	.673	.675	.750	.819	.880	.946		
14	.553	.613	.616	.653	.623	.694	.724	.771	.808	.875	
15	.562	.631	.639	.661	.611	.676	.672	.694	.700	.755	.926



Coordonnées des points sur l'axe 3 (5.4 %)

4 ans	1.22
5	0.60
6	0.22
7	-0.36
8	-1.39
9	-1.26
10	-1.31
11	-1.04
12	-0.47
13	0.38
14	2.14
15	2.30

Figure n° 4 : IMAGE EUCLIDIENNE DES AGES, ASSOCIEE A L'APPROXIMATION D'ORDRE 2 DES PRODUITS SCALAIRES $(w_{k_1} \mid w_{k_2})$



Coordonnées des points sur l'axe 3 (8.7 %)

4 ans	1.11
5	0.44
6	-0.05
7	0.02
8	-0.78
9	-0.99
10	-0.30
11	0.29
12	0.97
13	1.03
14	-0.39
15	-1.35

Figure n° 5 : IMAGE EUCLIDIENNE CENTREE

7 ans joue toutefois un rôle à part, car les corrélations entre variables sont systématiquement plus fortes qu'aux autres âges. Après les fluctuations des corrélations entre 4 et 6 ans, 7 ans apparaît donc comme un âge d'harmonie morphologique.

Pour la période 4-13 ans, les variations de la norme de la matrice de corrélations permettent de distinguer plusieurs étapes. Une première période de fluctuations s'étend de 4 à 8 ans, avec toutefois une grande similitude entre les matrices de corrélations à 5 et 8 ans. Cette particularité est mise en évidence sur les graphiques de l'interstructure par la proximité des points 5 et 8, mais aussi sur la figure n° 3, où les variables à 5 et 8 ans ont des positions très voisines, sauf pour le périmètre crânien.

Une deuxième période correspond à 9 et 10 ans. Les corrélations sont moins fortes, particulièrement celles des variables mesurant la corpulence avec les autres variables. Cette période est moins harmonieuse pour le rapport corpulence-squelette.

Enfin, une nouvelle période d'harmonie morphologique par rapport à la période 9-10 ans d'une part, et le bouleversement pubertaire d'autre part, s'étale sur 11-12 et 13 ans. Les matrices de corrélations sont semblables, les corrélations entre variables mesurant la corpulence, et variables mesurant la longueur du squelette, étant à nouveau élevées.

3. ETUDE DES EVOLUTIONS INDIVIDUELLES

3.1. INTERSTRUCTURE SUR LES OBJETS w_k

Le tableau des coefficients RV entre les âges (p. 138), ainsi que les figures n° 4 et n° 5 (qu'il faut imaginer dans l'espace) font apparaître une évolution chronologique régulière de 4 à 13 ans. Comme dans l'étude des variables, 14 et 15 ans se distinguent de la période précédente.

Les normes des w_k étant du même ordre de grandeur, un coefficient RV (k_1, k_2) proche de 1, indique que les images euclidiennes des produits scalaires entre individus, aux âges k_1 et k_2, sont équivalentes.

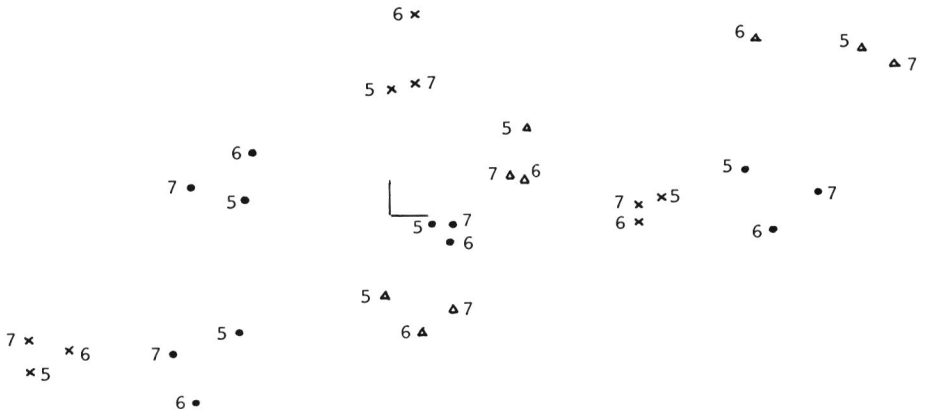


Figure n° 6 : PLAN 1.2. DU COMPROMIS. SELECTION DES POINTS DES TRAJECTOIRES INDIVIDUELLES CORRESPONDANT A 5 ANS, 6 ANS ET 7 ANS. LES POSITIONS MUTUELLES DES INDIVIDUS VARIENT PEU D'UNE ANNEE A L'AUTRE.

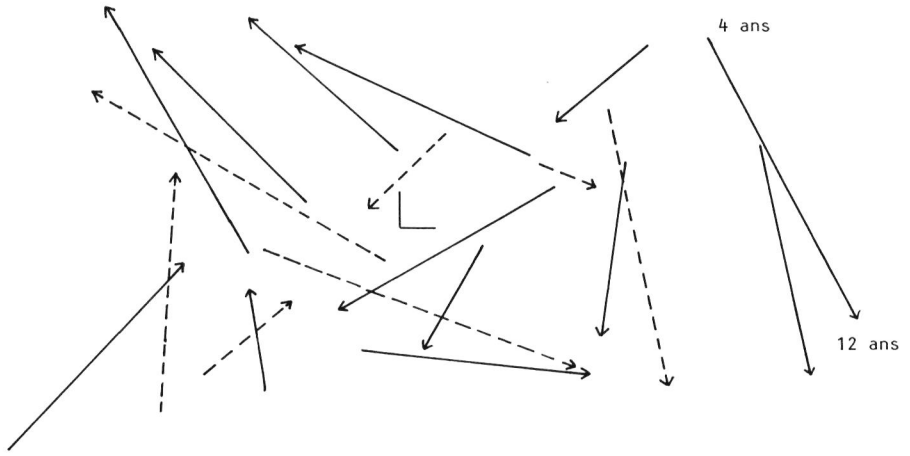


Figure n° 7 : PLAN 1.2. DU COMPROMIS. SELECTION DES POINTS DES TRAJECTOIRES INDIVIDUELLES CORRESPONDANT A 4 ANS ET 12 ANS.

C'est le cas par exemple de 5 et 6 ans dont le RV, égal à 0.935, est la valeur la plus élevée du tableau : la répartition des individus est stable d'une année à l'autre (figure n° 6), par opposition aux âges 4 et 12 ans (figure n° 7) dont le RV, égal à 0.453, est la valeur la plus faible du tableau. Pourtant on a vu que les matrices de corrélations à 5 et 6 ans n'étaient pas égales. Cela signifie que les changements constatés dans les relations entre variables, entre 5 et 6 ans, sont valables pour tous les individus.

Examinons le cas inverse des âges 5 et 8 ans, proches pour les objets v_k, mais éloignés pour les objets w_k. Les répartitions des individus ne sont plus comparables, ce qui traduit une diversité des évolutions individuelles entre 5 et 8 ans.

Jusqu'à 13 ans, les âges sont répartis régulièrement sur les graphiques de l'interstructure. On en conclut que les distances entre les individus changent d'une année à l'autre, mais pas de façon brutale.

3.2. POSITIONS COMPROMIS DES INDIVIDUS

Objet compromis w. Contrairement à l'objet compromis v, les coefficients de la combinaison linéaire donnant l'objet compromis w sont assez différents les uns des autres :

4 ans	.078	8 ans	.102	12 ans	.098
5 ans	.093	9 ans	.100	13 ans	.098
6 ans	.091	10 ans	.098	14 ans	.090
7 ans	.105	11 ans	.103	15 ans	.080

Cela reflète les valeurs faibles des coefficients RV, et l'écartement du faisceau formé par les vecteurs OM_k de l'image euclidienne des âges.

Interprétation des axes du compromis. Pour interpréter les positions compromis des individus dans le plan 1.2 (figure n° 9), on calcule les corrélations des variables avec les axes du compromis. Sur la figure n° 8 , on retrouve évidemment les associations et oppositions entre variables, mises en évidence au paragraphe 2.

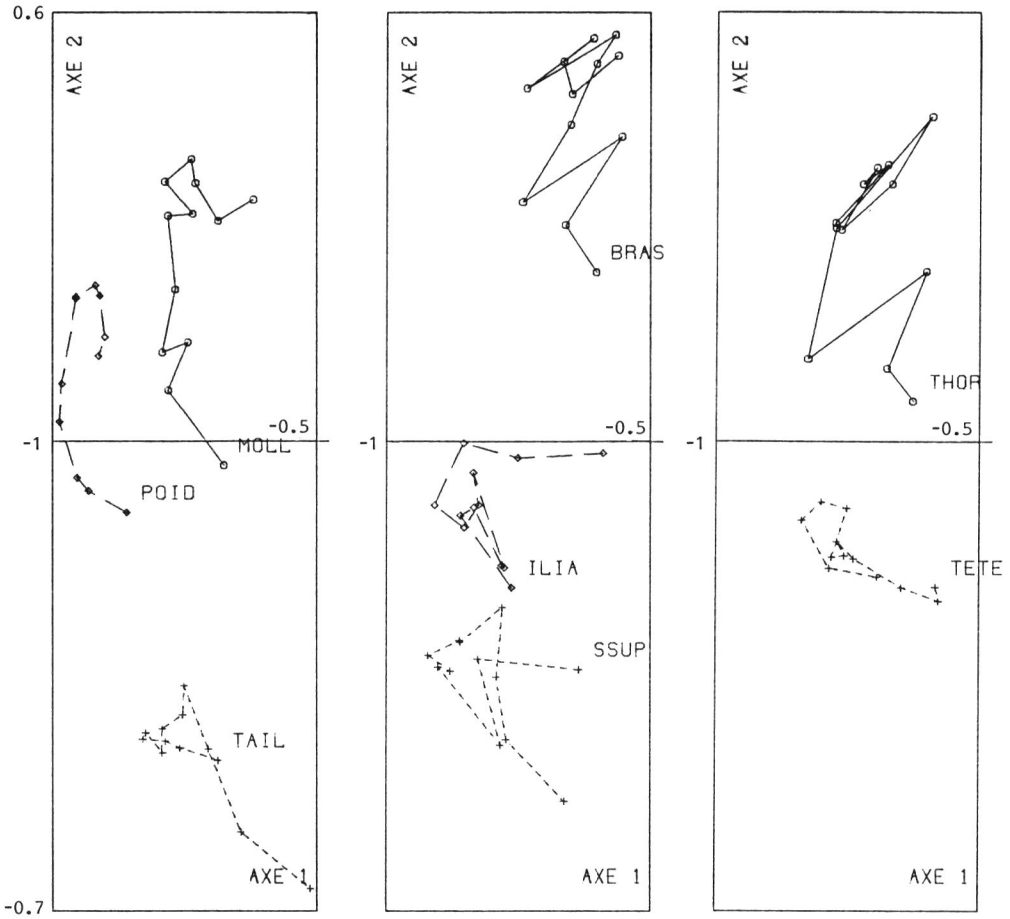


Figure n° 8 : CORRELATIONS DES VARIABLES AVEC LES AXES DU COMPROMIS : CHAQUE VARIABLE A POUR COORDONNEES SES CORRELATIONS AVEC LES AXES 1 ET 2 DU COMPROMIS.

(LE NOM EST SITUE A COTE DE LA POSITION DE LA VARIABLE A 4 ANS).

Toutes les variables, mais plus particulièrement le poids, sont corrélées avec la partie négative de l'axe 1 : cet axe distingue les enfants de forte constitution des enfants menus. L'axe 2 oppose les variables mesurant la corpulence aux variables mesurant la longueur du squelette : il sépare donc les enfants grands et maigres, des enfants petits et gros. Le diamètre bi-iliaque n'est pratiquement pas corrélé, ni avec l'axe 2, ni avec l'axe 3, et n'explique donc pas la dispersion des individus. Le périmètre crânien est corrélé avec la partie négative de l'axe 3. Mais comme nous nous limitons ici à l'explication des deux premiers axes, nous n'exploiterons pas cette information.

3.3. TRAJECTOIRES INDIVIDUELLES

On a représenté quelques trajectoires individuelles sur la figure n° 10. Ces trajectoires ont des formes différentes, des positions globales dans le plan différentes, et des sens de parcours différents.

Rappelons que les trajectoires s'interprètent par rapport à l'évolution moyenne. C'est-à-dire l'évolution d'un individu fictif moyen qui aurait pour valeurs les moyennes des variables par année. Les variables étant centrées par année, la trajectoire de cet individu moyen est réduite à un point, qui est l'origine du plan.

Forme d'une trajectoire. On distingue deux types de trajectoires. Une trajectoire qui est peu étendue et qui tourne sur elle-même, comme celle de l'individu n° 8, correspond à un individu dont l'évolution suit l'évolution moyenne. Cela signifie que, pour chaque variable, l'écart entre la valeur de la variable pour cet individu et la moyenne, est régulier d'une année à l'autre.

A l'opposé, une trajectoire de grande amplitude reflète une modification morphologique de l'individu au cours des années, différente de l'évolution moyenne.

Position d'une trajectoire. Comme l'évolution morphologique d'un enfant est relativement régulière au cours des années, les points de sa trajectoire sont toujours plus ou moins groupés dans le plan. La position globale de la trajectoire situe "en gros" l'individu, ce qui permet de le caractériser avec les mêmes règles d'interprétation que pour le compromis.

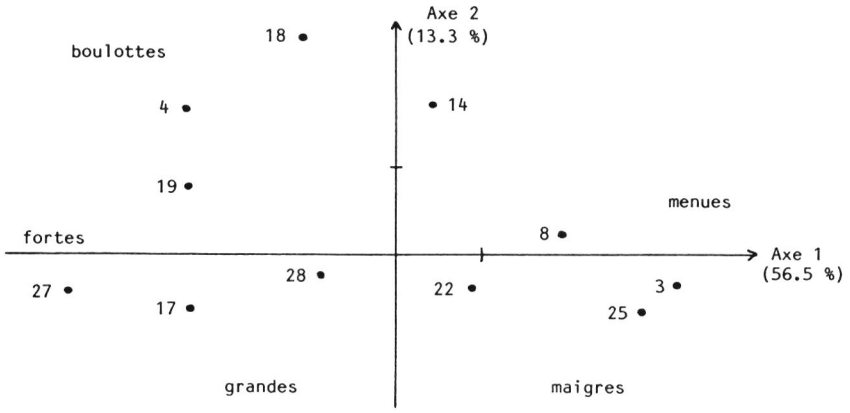


Figure n° 9 : POSITIONS COMPROMIS DES INDIVIDUS

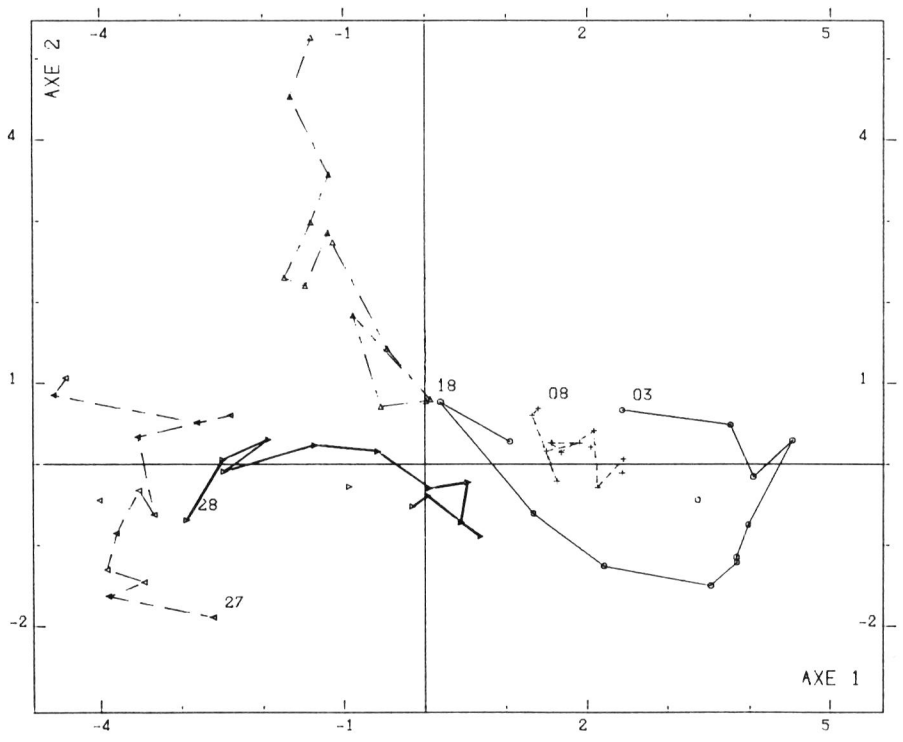


Figure n° 10 : TRAJECTOIRES INDIVIDUELLES DANS LE PLAN 1.2 DU COMPROMIS

(LE NUMERO DE L'INDIVIDU EST SITUE A COTE DE SA POSITION A 4 ANS. LES FIGURES N° 9 ET N° 10 SONT REALISEES AVEC LA MEME ECHELLE).

Par exemple, l'enfant n° 27, de constitution plutôt forte, s'oppose globalement à l'enfant n° 3, qui reste dans l'ensemble menu.

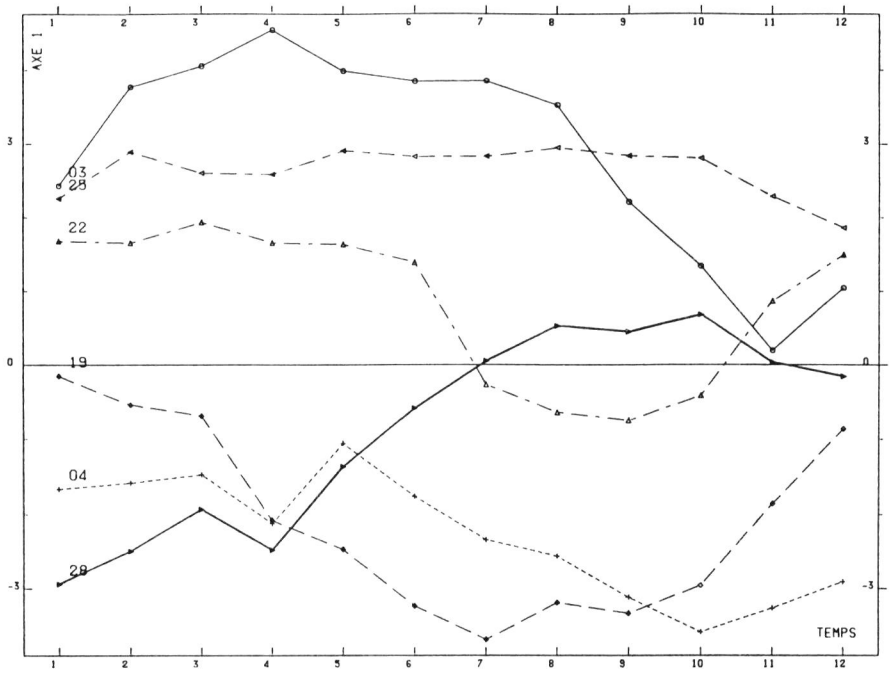
Sens de parcours. La stabilité relative des corrélations des variables avec les axes du compromis au cours des années, permet de donner une signification biologique aux axes, valable pour l'ensemble des années : le premier axe est expliqué par le poids de l'individu, le second axe oppose corpulence à taille. On peut alors interpréter le sens de parcours d'une trajectoire par rapport à un axe. En développant les trajectoires axe par axe en fonction du temps, on obtient les figures n° 11 et n° 12, plus faciles à interpréter que la figure n° 10 des trajectoires dans le plan.

Nous allons détailler les trajectoires des individus 27, 18, 29 qui sont, plus ou moins, parallèles à un axe sur la figure n° 10, et dont le sens de parcours s'interprète en fonction des variables qui expliquent cet axe.

Enfant n° 27 (figures n° 10 et n° 12). Sur la figure n° 10, la trajectoire est située globalement sur la partie gauche du graphique, car les valeurs des variables sont toujours au-dessus de la moyenne. La partie verticale de la trajectoire entre 5 et 11 ans s'explique par l'évolution des variables mesurant la corpulence, particulièrement le périmètre du bras, qui s'écartent de plus en plus des valeurs moyennes.

Par la suite, toutes les variables accusent un léger repli à 12 ans puis une nette augmentation. L'axe 1 étant corrélé avec l'ensemble des variables, ces variations expliquent la partie horizontale de la trajectoire entre 11 et 15 ans.

Enfant n° 18 (figures n° 10 et n° 12). La situation globale de la trajectoire dans la partie gauche supérieure de la figure n° 10 s'explique par les valeurs des variables mesurant la longueur du squelette qui sont toujours au-dessous de la moyenne, et les valeurs des variables mesurant la corpulence, y compris le poids, qui sont toujours au-dessus de la moyenne. Mais, à partir de 8 ans, l'augmentation des paramètres de corpulence est très nette, renforcée par une taille nettement plus petite que la moyenne à partir de 12 ans.



TRAJECTOIRES INDIVIDUELLES PAR RAPPORT A L'AXE 1 DU COMPROMIS

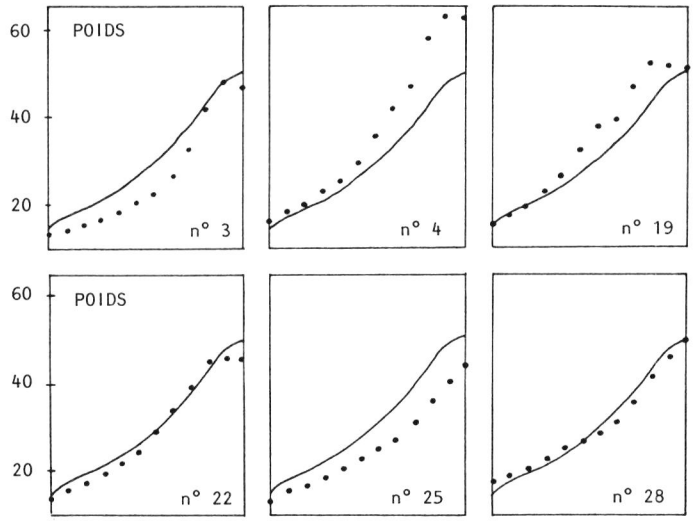
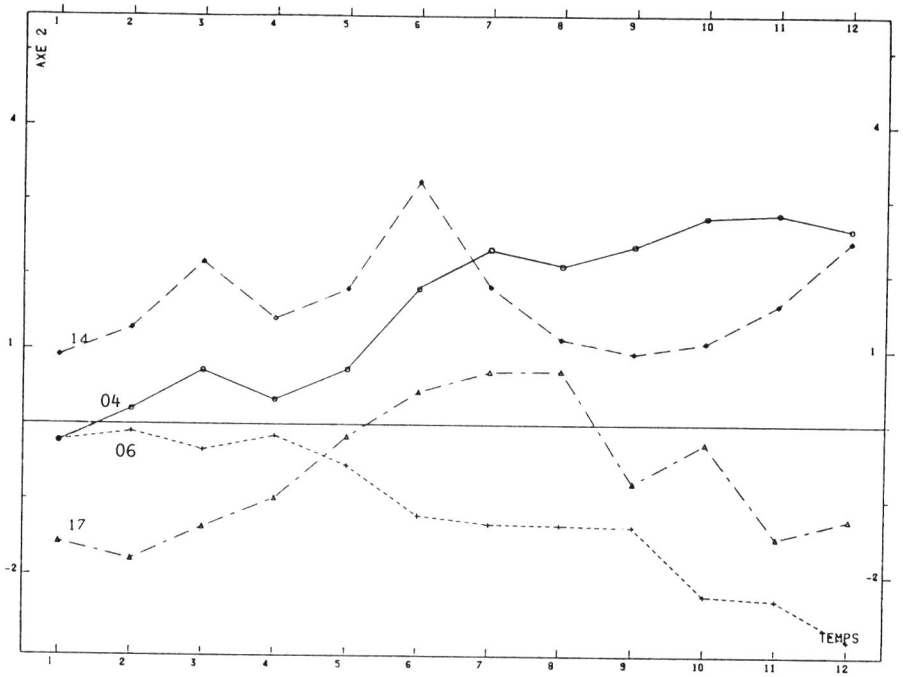


Figure n° 11 : EXPLICATION DES VARIATIONS DES TRAJECTOIRES LE LONG DE L'AXE 1 PAR LA VARIABLE **POIDS** QUI A UNE FORTE CORRELATION AVEC CET AXE.

(LA COURBE CONTINUE CORRESPOND A L'EVOLUTION DU POIDS MOYEN, LES POINTS REPRESENTENT LE POIDS DE L'INDIVIDU AUX DIFFERENTS AGES).



TRAJECTOIRES INDIVIDUELLES PAR RAPPORT A L'AXE 2 DU COMPROMIS

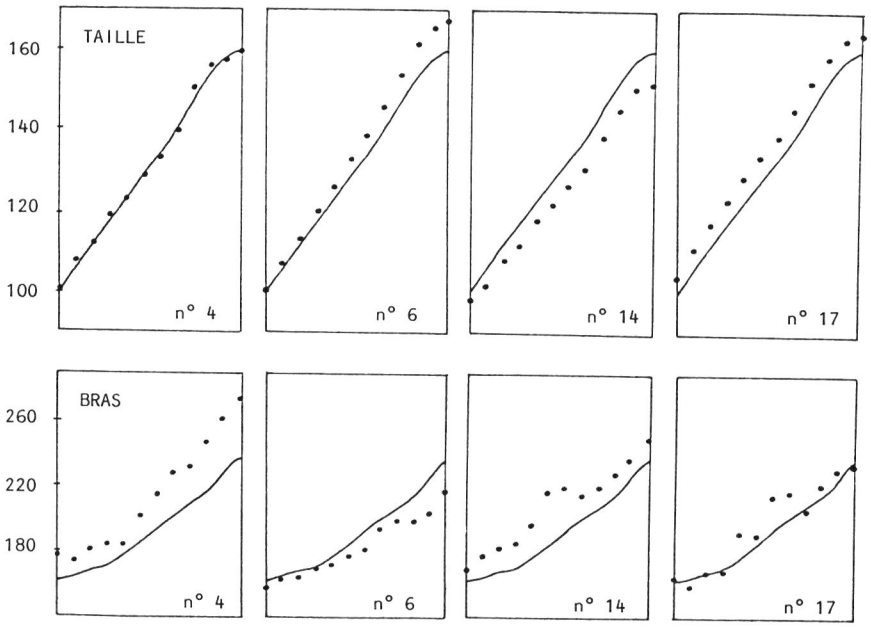
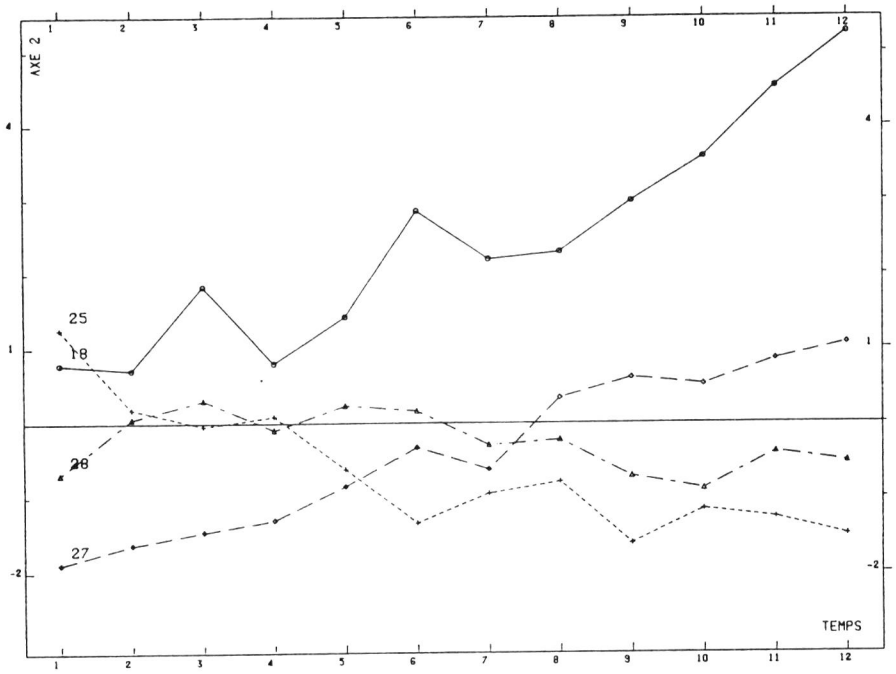


Figure n° 12 : EXPLICATION DES VARIATIONS DES TRAJECTOIRES LE LONG DE L'AXE 2 PAR L'OPPOSITION ENTRE LES VARIABLES **TAILLE** ET **PERIMETRE DU BRAS.**



TRAJECTOIRES INDIVIDUELLES PAR RAPPORT A L'AXE 2 DU COMPROMIS

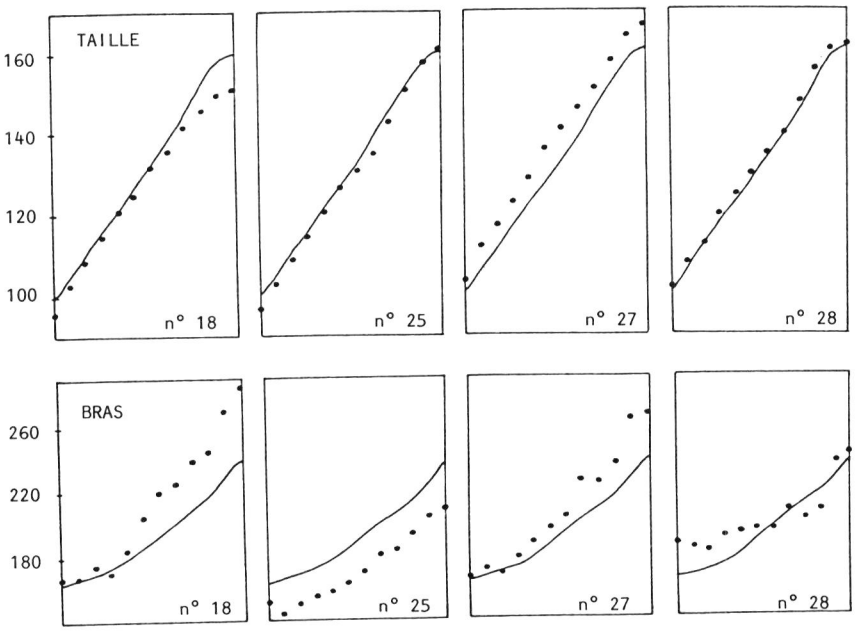


Figure n° 12 (suite) : EXPLICATION DES VARIATIONS DES TRAJECTOIRES LE LONG DE L'AXE 2 PAR L'OPPOSITION ENTRE LES VARIABLES **TAILLE** ET **PERIMETRE DU BRAS.**

Le retour de la trajectoire entre 9 et 10 ans semble s'expliquer par la légère augmentation de la longueur du segment supérieur pour rattraper la moyenne, et l'augmentation moins forte des périmètres du thorax et du mollet qui, tout en restant élevés, suivent l'évolution de la moyenne.

Enfant n° 28 (figures n° 10, n° 11 et n° 12). Pour cet enfant, poids, taille et périmètres du thorax et du mollet suivent en gros l'évolution moyenne. Mais bien que cet enfant démarre à 4 ans avec une constitution légèrement plus forte que la moyenne, ses paramètres de corpulence (particulièrement le bras) et son poids décroissent régulièrement pour devenir moins élevés que la moyenne à partir de 9 ans, et décroître encore jusqu'à 13 ans pour remonter ensuite légèrement.

3.4. CONCLUSIONS DE L'ETUDE DES INDIVIDUS

L'interstructure sur les objets w_k confirme la stabilité de la période 4-13 ans, et la rupture entre 13 ans d'une part, et 14-15 ans d'autre part. Les distances entre w_k mettent en évidence une évolution régulière des relations inter-individuelles. Les fluctuations des matrices de corrélations ne correspondent pas à des changements importants des distances entre individus. Cela signifie que les changements observés, à un âge donné, dans les relations entre variables sont caractéristiques de cet âge et affectent tous les individus. C'est le cas par exemple de la période 14-15 ans.

Par contre la rupture entre 13 et 14 ans est due à une diversité dans les évolutions individuelles, et se traduit par de grands écarts sur les trajectoires des individus.

Les points compromis donnent les différents types morphologiques, et les trajectoires permettent de décrire l'évolution de chaque individu par rapport à l'évolution moyenne.

ETUDE DE LA CROISSANCE

PARAMETRES MORPHOLOGIQUES

12 TABLEAUX CORRESPONDANT AUX AGES 4 A 15 ANS. 30 INDIVIDUS. 8 PARAMETRES. UNE LIGNE CORRESPOND A UN INDIVIDU POUR LEQUEL ON DONNE, DANS L'ORDRE :

NUMERO D'ORDRE,
POIDS (dcg),
TAILLE (mm),
SEGMENT SUPERIEUR (mm),
PERIMETRE CRANIEN (mm),
PERIMETRE DU THORAX (mm),
PERIMETRE DU BRAS GAUCHE (mm),
PERIMETRE DU MOLLET GAUCHE (mm),
DIAMETRE BI-ILIAQUE (mm).

4 ans

1.1456.1025. 602. 486. 520. 157. 205. 170.
2.1426. 998. 572. 501. 520. 150. 215. 169.
3.1335. 961. 560. 494. 495. 145. 214. 158.
4.1607.1006. 595. 497. 560. 178. 218. 172.
5.1684.1012. 584. 490. 553. 165. 220. 158.
6.1374.1012. 580. 492. 525. 158. 202. 167.
7.1570.1040. 586. 511. 540. 153. 220. 180.
8.1450. 990. 561. 488. 520. 159. 210. 158.
9.1214. 968. 571. 481. 476. 145. 198. 150.
10.1456. 983. 563. 485. 532. 158. 219. 154.
11.1290. 952. 553. 465. 510. 158. 203. 155.
12.1625. 990. 565. 500. 540. 175. 216. 177.
13.1431. 995. 547. 475. 510. 162. 204. 177.
14.1530. 968. 564. 505. 565. 170. 215. 161.
15.1628.1015. 590. 506. 540. 170. 220. 169.
16.1573. 995. 578. 505. 538. 165. 214. 167.
17.1706.1040. 600. 506. 552. 165. 216. 177.
18.1530. 962. 573. 500. 539. 168. 223. 161.
19.1515.1031. 567. 484. 518. 160. 226. 171.
20.1339. 932. 553. 484. 515. 158. 211. 159.
21.1842.1065. 624. 504. 570. 170. 242. 176.
22.1414.1005. 565. 486. 537. 152. 201. 156.
23.1617.1011. 571. 507. 552. 174. 220. 174.
24.1515.1025. 602. 497. 521. 157. 210. 173.
25.1360. 970. 546. 480. 527. 153. 207. 162.
26.1685.1013. 578. 500. 582. 176. 213. 168.
27.1660.1037. 605. 506. 538. 166. 224. 188.
28.1740.1020. 600. 502. 561. 185. 225. 177.
29.1509. 954. 568. 494. 556. 164. 208. 167.
30.1565. 963. 571. 520. 532. 170. 200. 161.

5 ans

1.1635.1082. 614. 488. 537. 160. 206. 180.
2.1503.1052. 589. 508. 530. 144. 211. 179.
3.1394.1012. 570. 498. 495. 145. 218. 163.
4.1820.1067. 610. 503. 580. 175. 228. 180.
5.1860.1070. 610. 500. 552. 170. 230. 176.
6.1615.1071. 597. 500. 555. 163. 215. 177.
7.1765.1102. 612. 515. 550. 162. 230. 190.
8.1636.1051. 579. 494. 532. 164. 222. 169.
9.1375.1034. 583. 486. 522. 153. 206. 157.
10.1667.1046. 570. 495. 550. 163. 227. 170.
11.1330.1000. 548. 470. 515. 149. 205. 161.
12.1857.1062. 595. 503. 563. 173. 228. 189.
13.1605.1056. 585. 487. 531. 167. 215. 186.
14.1735.1020. 574. 514. 557. 179. 226. 172.
15.1880.1089. 613. 515. 558. 167. 239. 175.
16.1830.1062. 610. 510. 552. 170. 227. 186.
17.1870.1106. 626. 522. 555. 160. 222. 192.
18.1731.1027. 593. 505. 565. 168. 234. 181.
19.1735.1095. 600. 484. 540. 167. 233. 184.
20.1560.1002. 586. 495. 553. 164. 230. 171.
21.2070.1141. 635. 512. 564. 168. 250. 189.
22.1555.1067. 596. 496. 546. 152. 209. 167.
23.1810.1078. 595. 512. 589. 163. 225. 180.
24.1660.1082. 618. 502. 542. 160. 213. 181.
25.1545.1032. 575. 490. 518. 145. 209. 173.
26.1875.1070. 609. 510. 575. 175. 226. 180.
27.1880.1115. 649. 510. 562. 171. 241. 205.
28.1900.1078. 615. 510. 575. 183. 233. 182.
29.1610.1012. 586. 496. 548. 162. 217. 172.
30.1705.1020. 576. 522. 562. 170. 212. 188.

6 ans

1.1800.1132. 631. 491. 563. 168. 213. 188.
2.1720.1095. 625. 515. 560. 154. 230. 183.
3.1530.1063. 590. 502. 508. 146. 220. 169.
4.2000.1122. 623. 508. 583. 182. 237. 199.
5.2060.1132. 632. 498. 573. 170. 239. 192.
6.1740.1134. 620. 505. 565. 165. 222. 187.
7.2030.1168. 667. 520. 592. 160. 230. 196.
8.1750.1112. 620. 500. 546. 160. 227. 172.
9.1610.1087. 606. 490. 524. 154. 216. 167.
10.1840.1103. 593. 500. 585. 170. 234. 187.
11.1517.1052. 573. 480. 540. 165. 217. 177.
12.2125.1126. 631. 507. 596. 188. 236. 203.
13.1780.1111. 626. 504. 556. 171. 226. 193.
14.1835.1075. 592. 513. 598. 182. 234. 179.
15.2065.1159. 647. 521. 552. 170. 244. 190.
16.1960.1130. 637. 519. 556. 169. 235. 195.
17.2135.1170. 648. 523. 582. 168. 235. 196.
18.1985.1085. 603. 512. 613. 175. 250. 182.
19.1940.1160. 623. 493. 575. 170. 240. 190.
20.1725.1065. 595. 494. 560. 165. 233. 179.
21.2250.1200. 673. 516. 577. 167. 250. 200.
22.1725.1126. 615. 497. 566. 158. 216. 170.
23.1960.1145. 620. 511. 610. 170. 233. 192.
24.1810.1143. 636. 506. 553. 160. 220. 190.
25.1680.1089. 600. 493. 550. 150. 218. 182.
26.2093.1131. 636. 518. 595. 178. 223. 181.
27.2160.1174. 656. 518. 582. 168. 248. 213.
28.2060.1134. 635. 521. 590. 181. 242. 185.
29.1810.1074. 606. 509. 565. 160. 221. 182.
30.1865.1070. 600. 524. 565. 175. 220. 191.

7 ans

1.2105.1183. 650. 499. 588. 173. 226. 198.
2.1860.1173. 630. 517. 562. 151. 230. 192.
3.1682.1113. 608. 504. 513. 152. 225. 180.
4.2300.1190. 660. 510. 630. 185. 245. 198.
5.2257.1187. 644. 515. 573. 172. 253. 193.
6.1925.1195. 642. 503. 590. 170. 233. 192.
7.2260.1235. 671. 524. 584. 168. 252. 211.
8.1970.1157. 624. 505. 562. 162. 241. 192.
9.1705.1126. 610. 491. 538. 151. 218. 172.
10.2045.1162. 623. 506. 596. 168. 243. 191.
11.1800.1090. 580. 475. 545. 154. 220. 177.
12.2350.1186. 640. 517. 610. 184. 239. 205.
13.1985.1157. 628. 507. 577. 173. 236. 198.
14.2105.1123. 618. 518. 576. 186. 247. 190.
15.2343.1218. 663. 526. 591. 180. 259. 194.
16.2300.1209. 663. 528. 594. 173. 253. 208.
17.2385.1227. 663. 527. 611. 169. 254. 203.
18.2165.1146. 632. 515. 585. 170. 260. 182.
19.2285.1218. 654. 503. 608. 185. 260. 197.
20.1990.1127. 617. 502. 570. 175. 257. 192.
21.2490.1256. 694. 520. 588. 173. 266. 207.
22.1920.1184. 627. 505. 598. 159. 227. 178.
23.2130.1202. 643. 515. 582. 170. 240. 199.
24.1960.1199. 657. 508. 575. 158. 230. 194.
25.1855.1149. 624. 500. 565. 155. 230. 185.
26.2255.1186. 650. 518. 600. 183. 231. 183.
27.2410.1232. 680. 525. 604. 177. 259. 223.
28.2282.1202. 673. 528. 602. 189. 250. 191.
29.1985.1145. 634. 514. 570. 170. 235. 183.
30.2070.1111. 618. 530. 595. 172. 235. 198.

8 ans

1.2321.1242. 673. 510. 596. 178. 245. 207.
2.2090.1210. 655. 520. 600. 156. 240. 200.
3.1855.1168. 637. 510. 542. 154. 234. 191.
4.2540.1234. 675. 515. 642. 185. 255. 199.
5.2465.1238. 645. 518. 600. 175. 263. 198.
6.2200.1258. 676. 514. 617. 173. 247. 202.
7.2550.1294. 691. 528. 619. 180. 264. 219.
8.2175.1213. 641. 508. 574. 170. 257. 188.
9.1845.1195. 655. 493. 560. 155. 225. 178.
10.2230.1218. 654. 514. 613. 175. 255. 193.
11.1900.1145. 618. 480. 570. 172. 232. 182.
12.2864.1254. 676. 529. 664. 208. 268. 214.
13.2220.1214. 648. 510. 622. 182. 243. 207.
14.2470.1183. 652. 523. 620. 198. 259. 196.
15.2600.1278. 683. 537. 592. 178. 267. 199.
16.2500.1266. 689. 535. 605. 167. 253. 215.
17.2770.1294. 699. 539. 645. 193. 264. 210.
18.2430.1208. 655. 520. 630. 185. 270. 192.
19.2655.1283. 690. 509. 630. 197. 269. 210.
20.2345.1174. 643. 504. 610. 188. 272. 197.
21.2875.1330. 717. 517. 619. 182. 276. 213.
22.2200.1248. 655. 508. 602. 167. 233. 187.
23.2260.1249. 666. 515. 580. 171. 243. 208.
24.2170.1253. 677. 511. 587. 162. 240. 199.
25.2040.1212. 643. 503. 575. 158. 233. 192.
26.2480.1244. 671. 518. 612. 185. 238. 190.
27.2720.1291. 704. 529. 620. 188. 272. 233.
28.2500.1250. 683. 531. 613. 191. 258. 194.
29.2288.1194. 662. 518. 606. 174. 250. 192.
30.2225.1168. 652. 530. 590. 179. 237. 206.

9 ans

1.2730.1291. 690. 513. 635. 192. 255. 214.
2.2340.1285. 683. 525. 590. 163. 245. 206.
3.2045.1220. 649. 510. 580. 155. 245. 190.
4.2965.1288. 698. 520. 662. 205. 277. 213.
5.2810.1303. 684. 519. 610. 184. 266. 207.
6.2490.1330. 680. 515. 595. 177. 257. 206.
7.2792.1348. 718. 530. 632. 180. 266. 225.
8.2375.1270. 664. 511. 628. 177. 259. 194.
9.2050.1247. 667. 500. 588. 164. 237. 185.
10.2580.1282. 670. 516. 620. 184. 263. 200.

	11.	2075.	1195.	637.	483.	570.	186.	241.	191.
	12.	3160.	1310.	699.	528.	655.	211.	271.	224.
	13.	2580.	1261.	678.	514.	660.	192.	263.	214.
	14.	3105.	1219.	661.	531.	656.	218.	272.	209.
	15.	2940.	1341.	712.	547.	594.	183.	274.	207.
	16.	2980.	1328.	716.	539.	605.	189.	265.	225.
	17.	3123.	1340.	723.	536.	671.	192.	274.	213.
	18.	2910.	1253.	678.	520.	660.	206.	297.	202.
	19.	3231.	1351.	710.	512.	668.	210.	293.	221.
	20.	2540.	1219.	658.	510.	622.	173.	272.	199.
	21.	3150.	1372.	740.	530.	623.	185.	291.	219.
	22.	2467.	1312.	665.	513.	638.	170.	247.	193.
	23.	2460.	1290.	682.	519.	602.	172.	251.	208.
	24.	2373.	1300.	696.	514.	590.	167.	245.	203.
	25.	2270.	1266.	661.	506.	588.	163.	237.	199.
	26.	2688.	1296.	690.	519.	620.	188.	242.	200.
	27.	3060.	1355.	725.	531.	625.	195.	283.	242.
	28.	2700.	1297.	694.	534.	625.	193.	266.	198.
	29.	2641.	1262.	681.	519.	610.	177.	255.	204.
	30.	2453.	1222.	663.	540.	593.	185.	240.	214.
10 ans	1.	3140.	1340.	710.	517.	674.	206.	265.	220.
	2.	2535.	1330.	696.	525.	641.	162.	255.	207.
	3.	2260.	1271.	671.	511.	567.	161.	252.	201.
	4.	3562.	1332.	718.	521.	711.	215	282.	218.
	5.	3220.	1353.	713.	518.	650.	203.	285.	217.
	6.	2880.	1394.	724.	525.	628.	182.	275.	216.
	7.	3142.	1403.	708.	530.	652.	188.	272.	232.
	8.	2700.	1314.	684.	514.	630.	189.	274.	203.
	9.	2280.	1280.	685.	505.	590.	168.	245.	191.
	10.	2817.	1336.	701.	516.	650.	188.	275.	205.
	11.	2250.	1245.	649.	489.	585.	187.	246.	195.
	12.	3510.	1362.	710.	528.	677.	217.	277.	216.
	13.	2970.	1311.	694.	521.	675.	207.	278.	219.
	14.	3088.	1271.	676.	525.	655.	220.	274.	211.
	15.	3325.	1403.	719.	550.	631.	194.	291.	210.
	16.	3135.	1383.	732.	542.	655.	183.	270.	230.
	17.	3720.	1385.	738.	536.	687.	214.	294.	218.
	18.	3360.	1315.	695.	530.	680.	220.	310.	215.
	19.	3792.	1413.	735.	520.	741.	226.	311.	222.
	20.	2815.	1272.	679.	518.	631.	196.	293.	204.
	21.	3565.	1426.	754.	534.	634.	191.	297.	230.
	22.	2930.	1392.	723.	522.	682.	183.	262.	206.
	23.	2670.	1345.	690.	525.	633.	180.	255.	217.
	24.	2555.	1350.	714.	517.	592.	171.	253.	206.
	25.	2500.	1308.	677.	512.	600.	170.	251.	203.
	26.	2945.	1346.	712.	530.	640.	198.	254.	204.
	27.	3390.	1409.	745.	533.	649.	205.	294.	250.
	28.	2920.	1345.	706.	537.	637.	194.	274.	202.
	29.	2915.	1324.	713.	523.	646.	192.	273.	205.
	30.	2720.	1270.	688.	543.	641.	186.	266.	217.

11 ans	1.	3549.	1390.	728.	521.	712.	220.	275.	228.
	2.	2790.	1386.	721.	525.	650.	174.	265.	219.
	3.	2658.	1344.	704.	514.	612.	164.	262.	211.
	4.	4200.	1403.	736.	528.	756.	228.	313.	234.
	5.	3470.	1400.	726.	520.	665.	204.	288.	223.
	6.	3180.	1456.	748.	530.	650.	194.	275.	226.
	7.	3616.	1469.	748.	536.	698.	196.	285.	241.
	8.	3000.	1378.	700.	517.	680.	190.	284.	210.
	9.	2490.	1336.	703.	505.	606.	170.	255.	197.
	10.	3165.	1399.	716.	519.	663.	194.	279.	208.
	11.	2540.	1301.	670.	499.	625.	188.	265.	202.
	12.	4065.	1431.	759.	532.	745.	231.	302.	231.
	13.	3365.	1373.	723.	524.	695.	218.	293.	225.
	14.	3185.	1312.	690.	532.	645.	216.	279.	207.
	15.	3845.	1478.	774.	551.	675.	198.	310.	223.
	16.	3660.	1470.	786.	544.	680.	199.	285.	246.
	17.	3968.	1463.	764.	539.	740.	217.	307.	227.
	18.	3750.	1362.	728.	535.	735.	225.	323.	223.
	19.	3920.	1498.	773.	527.	760.	226.	313.	231.
	20.	3090.	1320.	696.	520.	645.	203.	303.	220.
	21.	3880.	1493.	795.	542.	745.	209.	313.	250.
	22.	3420.	1473.	769.	524.	705.	202.	273.	219.
	23.	2840.	1377.	712.	524.	620.	181.	262.	223.
	24.	2790.	1409.	739.	518.	609.	184.	264.	213.
	25.	2700.	1354.	684.	526.	630.	180.	264.	209.
	26.	3400.	1410.	744.	539.	696.	205.	271.	215.
	27.	3960.	1458.	780.	537.	699.	225.	310.	258.
	28.	3135.	1397.	726.	543.	645.	205.	282.	209.
	29.	3600.	1382.	735.	529.	720.	214.	297.	217.
	30.	3020.	1310.	718.	556.	675.	207.	276.	226.
12 ans	1.	4040.	1450.	756.	528.	747.	226.	299.	244.
	2.	3135.	1436.	748.	531.	640.	179.	284.	231.
	3.	3305.	1433.	760.	522.	650.	186.	292.	222.
	4.	4685.	1502.	803.	535.	830.	232.	331.	253.
	5.	3975.	1485.	770.	530.	760.	206.	309.	222.
	6.	3500.	1541.	797.	531.	712.	198.	288.	232.
	7.	4130.	1549.	798.	544.	735.	206.	300.	255.
	8.	3350.	1428.	720.	517.	692.	208.	298.	219.
	9.	2785.	1397.	725.	515.	605.	186.	273.	200.
	10.	3695.	1471.	765.	521.	720.	205.	304.	223.
	11.	2770.	1365.	716.	500.	630.	192.	269.	202.
	12.	4672.	1512.	786.	540.	790.	237.	318.	251.
	13.	4090.	1464.	766.	537.	785.	232.	313.	236.
	14.	3650.	1392.	739.	537.	701.	221.	296.	223.
	15.	4475.	1581.	833.	560.	730.	203.	322.	249.
	16.	4220.	1543.	822.	554.	720.	220.	295.	270.
	17.	3930.	1531.	809.	546.	693.	204.	310.	242.
	18.	4265.	1420.	755.	540.	763.	239.	337.	232.
	19.	4700.	1584.	824.	533.	785.	236.	328.	248.
	20.	3345.	1373.	708.	527.	675.	214.	314.	232.

	21.	4350.	1576.	823.	540.	720.	208.	319.	254.
	22.	3910.	1548.	808.	535.	740.	206.	291.	234.
	23.	3270.	1443.	742.	531.	670.	187.	278.	238.
	24.	3170.	1473.	762.	524.	643.	195.	281.	223.
	25.	3100.	1429.	734.	525.	660.	183.	274.	223.
	26.	4090.	1486.	780.	546.	760.	219.	283.	239.
	27.	4240.	1506.	779.	550.	715.	223.	319.	272.
	28.	3550.	1479.	763.	549.	700.	200.	293.	225.
	29.	3865.	1470.	789.	537.	750.	205.	300.	233.
	30.	3395.	1372.	737.	556.	695.	214.	282.	237.
13 ans	1.	4530.	1520.	790.	535.	783.	233.	323.	261.
	2.	3485.	1497.	782.	540.	680.	186.	297.	242.
	3.	4140.	1523.	800.	527.	720.	198.	313.	241.
	4.	5790.	1555.	832.	540.	870.	248.	350.	275.
	5.	4450.	1556.	798.	537.	745.	207.	312.	238.
	6.	4020.	1624.	832.	543.	710.	198.	299.	247.
	7.	4610.	1598.	825.	550.	770.	217.	312.	267.
	8.	3780.	1505.	764.	530.	725.	203.	302.	234.
	9.	3015.	1449.	749.	515.	630.	190.	276.	208.
	10.	4170.	1540.	793.	529.	710.	214.	313.	239.
	11.	3235.	1439.	745.	501.	665.	197.	280.	228.
	12.	5060.	1576.	820.	545.	785.	232.	320.	263.
	13.	4580.	1560.	825.	542.	775.	239.	327.	267.
	14.	4145.	1458.	773.	539.	737.	229.	307.	239.
	15.	5135.	1648.	865.	563.	750.	222.	341.	262.
	16.	4790.	1585.	860.	558.	790.	217.	315.	270.
	17.	4625.	1592.	828.	546.	755.	222.	320.	270.
	18.	5030.	1458.	757.	547.	830.	245.	351.	245.
	19.	5245.	1610.	853.	537.	835.	245.	343.	260.
	20.	3780.	1438.	750.	530.	740.	219.	330.	247.
	21.	5090.	1626.	883.	549.	796.	216.	330.	269.
	22.	4505.	1578.	828.	537.	795.	223.	300.	243.
	23.	3900.	1530.	791.	536.	695.	195.	298.	251.
	24.	3970.	1576.	823.	530.	710.	201.	301.	238.
	25.	3600.	1506.	763.	528.	699.	192.	288.	237.
	26.	4600.	1555.	831.	547.	808.	224.	302.	242.
	27.	4960.	1584.	838.	554.	758.	235.	335.	286.
	28.	4145.	1556.	803.	549.	748.	205.	302.	229.
	29.	4540.	1520.	811.	545.	825.	229.	326.	245.
	30.	3740.	1455.	772.	558.	724.	210.	291.	251.
14 ans	1.	4880.	1568.	816.	535.	811.	237.	328.	262.
	2.	4020.	1564.	823.	545.	745.	197.	314.	258.
	3.	4735.	1559.	829.	537.	810.	230.	337.	257.
	4.	6285.	1576.	843.	547.	935.	262.	357.	280.
	5.	5030.	1594.	831.	539.	790.	216.	333.	253.
	6.	4400.	1658.	856.	546.	780.	206.	318.	255.
	7.	4970.	1618.	842.	548.	795.	225.	322.	273.
	8.	4160.	1549.	792.	532.	780.	207.	312.	239.
	9.	3650.	1513.	804.	523.	704.	206.	296.	221.
	10.	4365.	1572.	836.	535.	760.	218.	316.	243.
	11.	3765.	1501.	781.	509.	712.	210.	302.	240.
	12.	5325.	1597.	837.	552.	816.	238.	325.	270.
	13.	4840.	1580.	839.	542.	795.	246.	341.	274.
	14.	4500.	1509.	806.	543.	825.	237.	319.	248.
	15.	5445.	1654.	886.	580.	780.	224.	347.	269.
	16.	5235.	1601.	871.	559.	810.	228.	340.	276.
	17.	4840.	1640.	863.	553.	765.	232.	323.	271.
	18.	5430.	1504.	800.	548.	915.	270.	374.	256.
	19.	5170.	1627.	863.	545.	840.	253.	343.	265.
	20.	3910.	1492.	764.	532.	760.	218.	324.	257.
	21.	5520.	1657.	890.	554.	860.	232.	337.	276.
	22.	4540.	1591.	837.	537.	798.	227.	306.	246.
	23.	4400.	1593.	830.	544.	750.	212.	298.	269.
	24.	4420.	1635.	843.	543.	752.	213.	316.	243.
	25.	4050.	1579.	804.	538.	728.	203.	304.	251.
	26.	5330.	1581.	850.	555.	850.	239.	326.	245.
	27.	5870.	1645.	876.	566.	880.	262.	369.	300.
	28.	4620.	1605.	835.	558.	800.	234.	323.	236.
	29.	4685.	1547.	829.	551.	865.	236.	328.	250.
	30.	4630.	1522.	826.	566.	780.	230.	315.	270.
15 ans	1.	5050.	1585.	828.	543.	826.	246.	336.	267.
	2.	4620.	1606.	849.	550.	800.	211.	323.	269.
	3.	4660.	1582.	835.	537.	770.	235.	344.	269.
	4.	6275.	1600.	857.	549.	905.	275.	365.	280.
	5.	5000.	1615.	852.	541.	790.	222.	339.	259.
	6.	4725.	1682.	878.	547.	820.	217.	318.	259.
	7.	5340.	1632.	850.	552.	815.	240.	328.	279.
	8.	4365.	1573.	812.	535.	790.	215.	324.	248.
	9.	4120.	1560.	840.	527.	750.	217.	313.	233.
	10.	4880.	1590.	837.	537.	782.	238.	327.	255.
	11.	4260.	1520.	800.	516.	771.	224.	314.	249.
	12.	5380.	1611.	850.	552.	830.	248.	327.	274.
	13.	4965.	1598.	851.	544.	810.	249.	348.	279.
	14.	4960.	1519.	817.	552.	845.	251.	342.	262.
	15.	5700.	1668.	905.	585.	820.	243.	360.	270.
	16.	5435.	1607.	880.	577.	855.	241.	346.	277.
	17.	5167.	1650.	873.	559.	835.	235.	338.	272.
	18.	5485.	1509.	810.	548.	915.	285.	375.	262.
	19.	5100.	1635.	866.	545.	850.	246.	349.	267.
	20.	4365.	1505.	776.	540.	800.	236.	347.	265.
	21.	5450.	1670.	898.	554.	873.	230.	345.	275.
	22.	4610.	1601.	847.	538.	808.	230.	313.	248.
	23.	4720.	1633.	860.	553.	780.	222.	318.	279.
	24.	4780.	1656.	864.	545.	780.	227.	329.	254.
	25.	4400.	1613.	827.	543.	784.	209.	316.	263.
	26.	5100.	1593.	861.	563.	880.	247.	319.	251.
	27.	6185.	1655.	892.	560.	925.	265.	380.	305.
	28.	5050.	1623.	856.	564.	825.	240.	338.	241.
	29.	4880.	1568.	852.	551.	865.	239.	332.	255.
	30.	5300.	1555.	857.	571.	840.	252.	340.	285.

VII. Évolution économique des provinces espagnoles entre 1960 et 1979

1. DESCRIPTION DES DONNEES

1.1. EVOLUTION GLOBALE DE L'ECONOMIE ESPAGNOLE ENTRE 1960 ET 1979

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

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

(*) Ce chapitre a été écrit en collaboration avec Carlos PEREZ-HUGALDE, Professeur, Département de Statistique, Escuela Tecnica Superior de Ingenieros Agronomos, MADRID, ESPAGNE.

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

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

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

1.2. DESCRIPTION DES DONNEES ET DE L'OBJECTIF POURSUIVI

Dans l'exemple sur la croissance morphologique des enfants, présenté au chapitre VI, on cherchait essentiellement à caractériser les âges comme étapes particulières dans le développement de l'enfant. Dans cet exemple, nos préoccupations sont différentes. On s'intéresse aux évolutions régionales, afin de comprendre certains aspects du développement économique de l'Espagne, pendant la période étudiée. Les années n'ont plus l'importance qu'elles avaient dans l'exemple précédent. C'est la raison pour laquelle on ne considère que 10 années, réparties dans la période étudiée : 1960 - 62 - 64 - 67 - 69 - 71 - 73 - 75 - 77 et 79.

L'unité géographique considérée est la province, qui est le plus petit niveau pour lequel nous disposions de données homogènes.

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

Chacune des 50 provinces espagnoles (individus) est ainsi caractérisée par 15 variables quantitatives, pour 10 années (tableaux). Les provinces ont toutes le même poids $\frac{1}{50}$, les années ont toutes le même poids $\frac{1}{10}$. Les variables ont été centrées et réduites, pour chaque année.

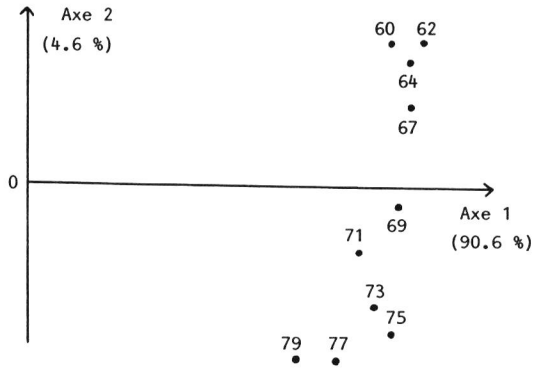
La méthode STATIS appliquée à ces données, met en évidence les déséquilibres régionaux, et permet de les expliquer. Quelques questions nous aideront à préciser notre objectif : les provinces qui ont les meilleurs résultats économiques dans le secteur agricole sont-elles aussi, de manière générale, les plus développées ? Le processus de croissance a-t-il réduit les déséquilibres existant entre les résultats économiques des agricultures régionales ? Les résultats économiques du secteur agricole ont-ils évolué plus rapidement, dans les provinces à fort développement global ?

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

JUDEZ L., GARCIA VELAZQUEZ A. - Analyse des relations entre développement agricole et développement global au niveau régional en Espagne, Economie Rurale, n° 134, p. 45-53, 1979.

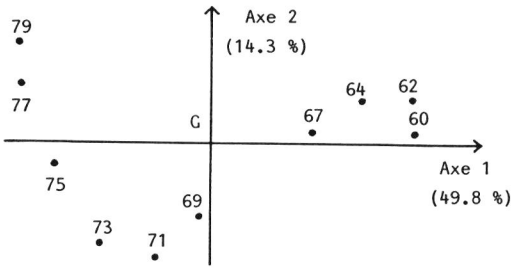
TABLEAU DES PRODUITS SCALAIRES NORMES : $RV(k_1,k_2) = \dfrac{(w_{k_1} \mid w_{k_2})}{\| w_{k_1} \| \; \| w_{k_2} \|}$

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



	$\| w_k \|$
1960	9.03
62	8.95
64	9.13
67	8.97
69	8.75
71	8.54
73	8.20
75	8.21
77	7.93
79	7.66

Figure n° 1 : IMAGE EUCLIDIENNE DES ANNEES, ASSOCIEE A L'APPROXIMATION D'ORDRE 2 DES PRODUITS SCALAIRES $(w_{k_1} | w_{k_2})$



Coordonnées des points sur l'axe 3 (10.4 %)	
1960	-1.60
62	-0.50
64	0.83
67	0.96
69	1.44
71	0.21
73	-0.90
75	-0.35
77	-0.12
79	0.03

Figure n° 2 : IMAGE EUCLIDIENNE CENTREE

2. INTERSTRUCTURE ET DETERMINATION DU COMPROMIS

Interstructure sur les w_k. Les coefficients RV entre les années (p. 157) sont particulièrement élevés. Ce qui signifie que les positions mutuelles des provinces évoluent peu d'une année à l'autre, principalement entre 1960 et 1967.

Par contre, les normes des w_k évoluent. La diminution régulière de la norme, à partir du maximum atteint en 1967, indique que les corrélations entre les variables d'une même année sont de moins en moins fortes. Nous verrons que cela se traduit par une spécialisation, de plus en plus marquée, des provinces.

Définition du compromis. Les coefficients de la combinaison linéaire donnant le compromis, privilégient les années 60 à 71 :

1960	62	64	67	69	71	73	75	77	79
.109	.109	.113	.113	.110	.107	.100	.101	.095	.091

Malgré tout, le compromis est un bon résumé car les w_k sont, dans l'ensemble, proches les uns des autres. En effet, les coefficients RV sont élevés, et les normes des w_k sont du même ordre de grandeur.

3. SIGNIFICATION ECONOMIQUE DES AXES DU COMPROMIS

Le graphique des corrélations des variables de chaque tableau avec les axes du compromis, permet de comparer l'évolution des différents secteurs, de donner une signification économique aux axes, et d'interpréter la situation compromis et les trajectoires des provinces (Figures n° 3 et n° 4, complétées par les matrices de corrélations entre variables en 1960 - 67 - 79, p.162).

Degré de développement. Toutes les variables, sauf celles qui représentent le poids du secteur agricole dans l'ensemble économique (production et emploi dans l'agriculture), sont corrélées avec le premier axe du compromis. Cet axe, représentant un fort pourcentage de l'inertie, discrimine les provinces selon leur degré de développement.

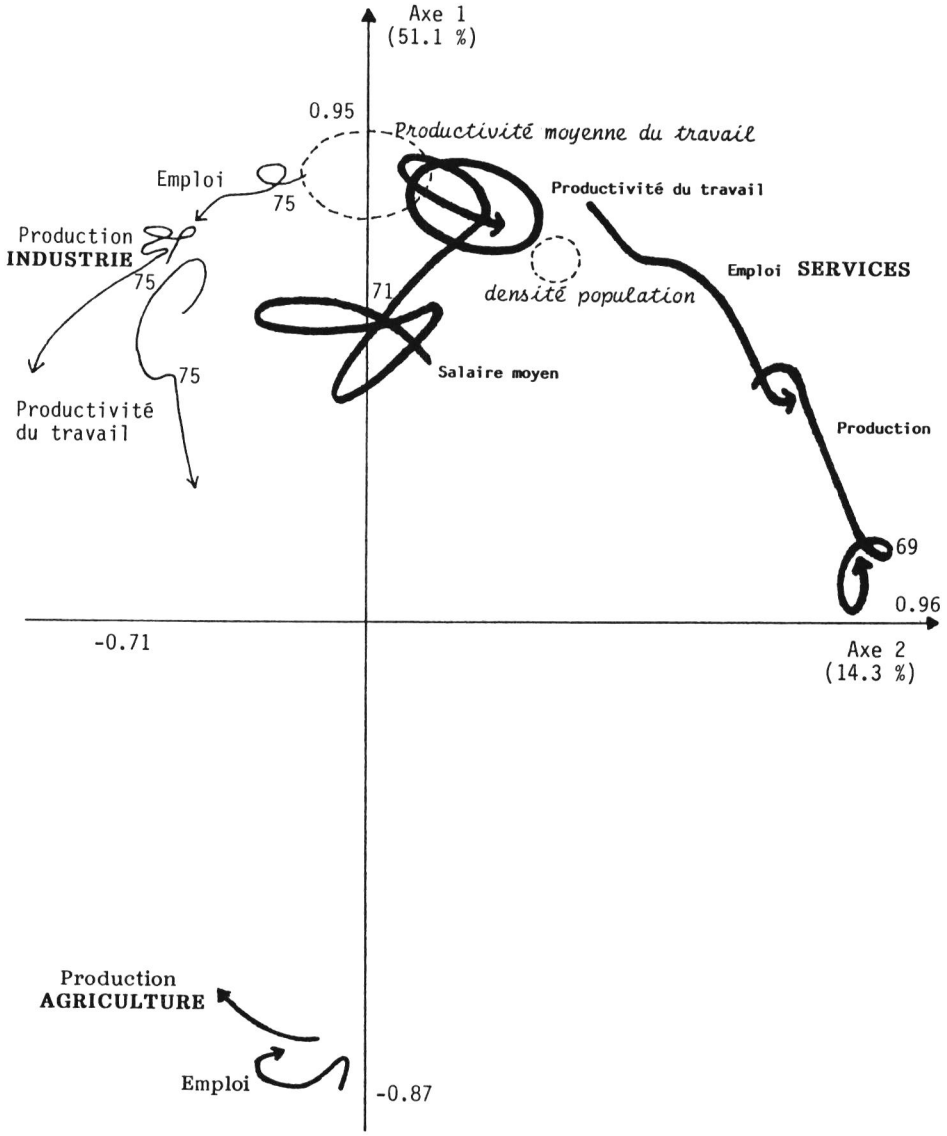


Figure n° 3 : SIGNIFICATION ECONOMIQUE DES AXES DU COMPROMIS. L'AXE 1 INDIQUE LE DEGRE DE DEVELOPPEMENT, L'AXE 2 OPPOSE LES SECTEURS INDUSTRIEL ET TERTIAIRE.

Il oppose les provinces industrialisées, à forte densité de population, qui ont une productivité moyenne du travail élevée, et de bons résultats économiques dans les secteurs non agricoles, à celles qui ont les caractéristiques contraires parmi lesquelles on remarque, en particulier, les provinces qui ont un secteur agricole prépondérant.

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

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

Agriculture. La productivité du travail agricole, et le revenu par actif non salarié dans ce secteur, sont très corrélés (la corrélation oscille entre .64 et .77 durant la période étudiée), et définissent le troisième axe. Parallèlement, ces variables sont de moins en moins corrélées avec le premier axe. Cela signifie que les agricultures riches ne se trouvent pas nécessairement dans les régions à fort degré de développement.

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

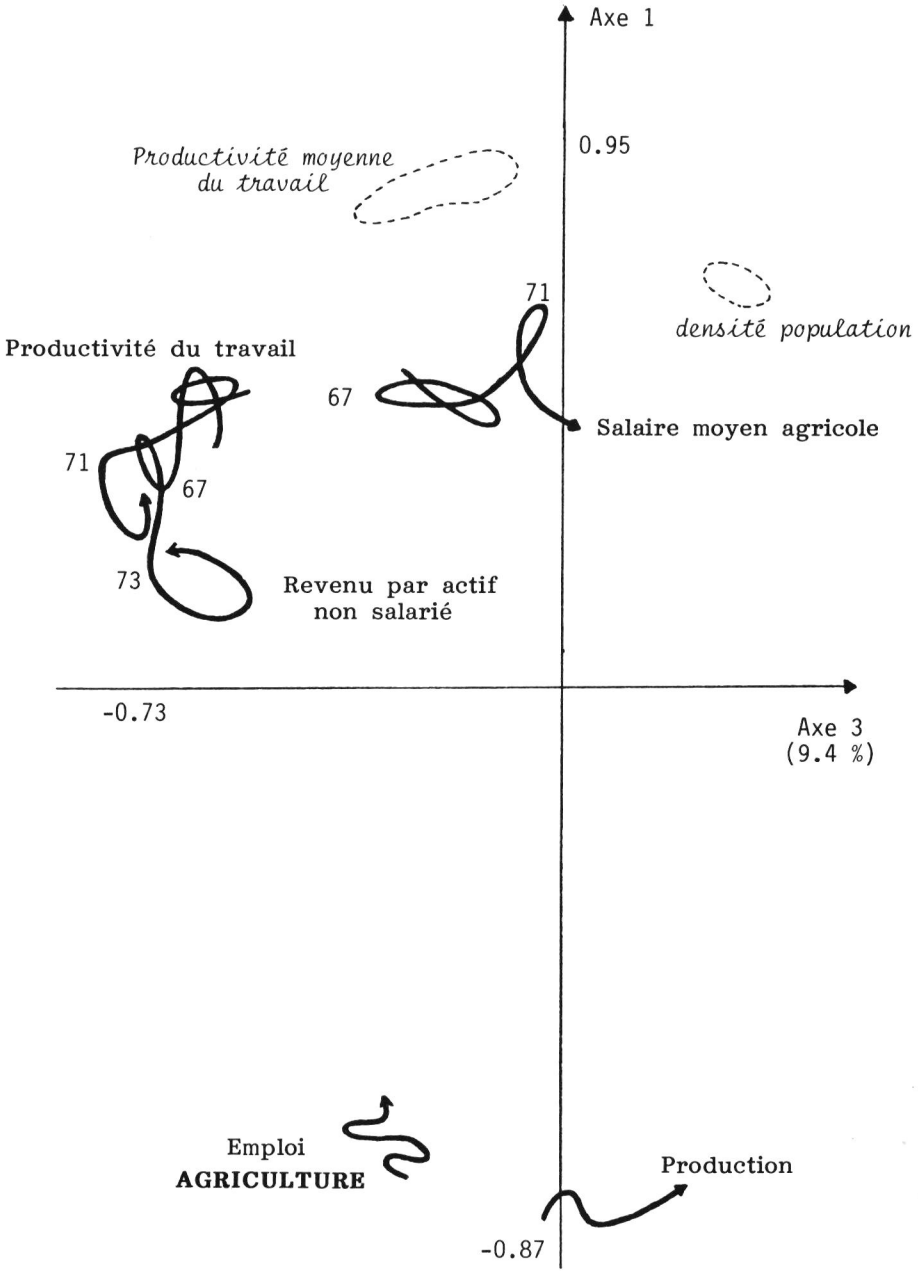


Figure n° 4 : L'AXE 3 EST LIE AU SECTEUR AGRICOLE : PRODUCTIVITE DU TRAVAIL ET REVENU PAR ACTIF NON SALARIE.

CORRELATIONS ENTRE LES VARIABLES

		POPUL.	PROD. MOYENNE	EMPLOI	PIB	PROD. INDUSTRIE	SALAIRE	EMPLOI	PIB	PROD. SERVICES	SALAIRE
	PROD.	.65									
I	EMPLOI	.69	.89								
I	PIB	.46	.71	.87							
I	PROD.	.19	.73	.63	.82						
I	SALAIRE	.58	.81	.83	.81	.73					
S	EMPLOI	.67	.77	.70	.41	.33	.50				
S	PIB	.56	.32	.30	-.02	-.13	.12	.80			
S	PROD.	.73	.83	.73	.46	.45	.64	.77	.63		
S	SALAIRE	.20	.81	.40	.28	.27	.36	.41	.34	.59	
A	EMPLOI	-.75	-.89	-.95	-.74	-.54	-.76	-.88	-.54	-.79	-.41
A	PIB	-.71	-.74	-.86	-.78	-.56	-.73	-.80	-.58	-.72	-.38
A	PROD.	.19	.68	.44	.23	.39	.36	.40	-.01	.50	.38
A	SALAIRE	.26	.59	.62	.51	.39	.44	.40	-.02	.38	.17
	REVENU/ NON SAL.	.25	.53	.34	.08	.19	.18	.47	.23	.50	.19

1960

		EMPLOI	PIB	PROD. AGRICULTURE	SALAIRE
A	PIB	.92			
A	PROD.	-.42	-.11		
A	SALAIRE	-.55	-.38	.58	
	REVENU/NON SAL.	-.40	-.16	.75	.30

		POPUL.	PROD. MOYENNE	EMPLOI	PIB	PROD. INDUSTRIE	SALAIRE	EMPLOI	PIB	PROD. SERVICES	SALAIRE
	PROD.	.66									
I	EMPLOI	.62	.89								
I	PIB	.46	.72	.89							
I	PROD.	.40	.72	.64	.81						
I	SALAIRE	.58	.80	.79	.83	.82					
S	EMPLOI	.59	.70	.45	.19	.32	.40				
S	PIB	.43	.21	.03	-.22	-.16	-.05	.78			
S	PROD.	.73	.89	.85	.65	.55	.68	.63	.37		
S	SALAIRE	.54	.65	.48	.40	.56	.56	.51	.23	.63	
A	EMPLOI	-.72	-.94	-.91	-.72	-.60	-.75	-.77	-.38	-.88	-.55
A	PIB	-.71	-.76	-.80	-.74	-.61	-.70	-.68	-.47	-.79	-.47
A	PROD.	.12	.65	.51	.29	.33	.35	.31	-.18	.44	.37
A	SALAIRE	.25	.60	.56	.40	.34	.46	.25	-.11	.51	.28
	REVENU/ NON SAL.	.07	.47	.39	.18	.15	.09	.37	.02	.28	.14

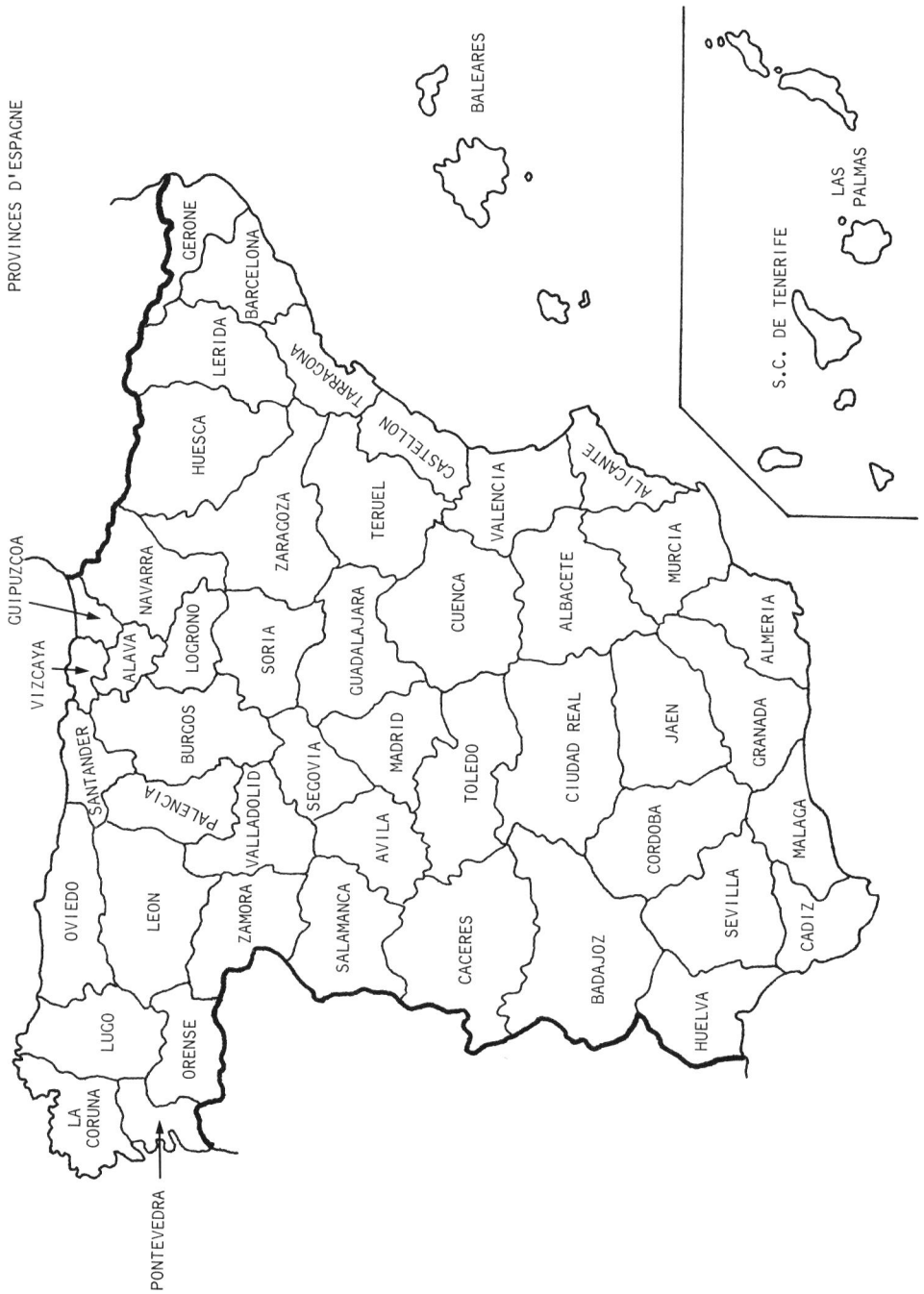
1967

		EMPLOI	PIB	PROD. AGRICULTURE	SALAIRE
A	PIB	.90			
A	PROD.	-.46	-.08		
A	SALAIRE	-.48	-.24	.68	
	REVENU/NON SAL.	-.43	-.17	.68	.32

		POPUL.	PROD. MOYENNE	EMPLOI	PIB	PROD. INDUSTRIE	SALAIRE	EMPLOI	PIB	PROD. SERVICES	SALAIRE
	PROD.	.54									
I	EMPLOI	.44	.74								
I	PIB	.19	.41	.81							
I	PROD.	.02	.30	.15	.51						
I	SALAIRE	.53	.50	.48	.47	.33					
S	EMPLOI	.47	.62	.07	-.06	-.06	.24				
S	PIB	.36	.12	-.30	-.64	-.39	.01	.78			
S	PROD.	.58	.77	.68	.33	.00	.43	.42	.30		
S	SALAIRE	.72	.69	.65	.33	-.01	.46	.41	.30	.87	
A	EMPLOI	-.63	-.93	-.78	-.40	-.10	-.51	-.66	-.25	-.76	-.73
A	PIB	-.64	-.61	-.55	-.39	-.19	-.54	-.53	-.43	-.70	-.68
A	PROD.	.04	.59	.52	.27	.09	.10	.14	-.34	.25	.21
A	SALAIRE	.12	.38	.57	.60	.28	.45	-.12	-.34	.35	.28
	REVENU/ NON SAL.	.09	.48	.30	.05	.02	.00	.31	-.11	.10	.16

1979

		EMPLOI	PIB	PROD. AGRICULTURE	SALAIRE
A	PIB	.77			
A	PROD.	-.43	.16		
A	SALAIRE	-.33	-.22	.31	
	REVENU/NON SAL.	-.39	.08	.76	-.11


PROVINCES D'ESPAGNE
BALEARES
S.C. DE TENERIFE
LAS PALMAS
GERONE
BARCELONA
LERIDA
TARRACONA
CASTELLON
ALICANTE
HUESCA
ZARAGOZA
TERUEL
VALENCIA
GUIPUZCOA
VIZCAYA
NAVARRA
ALAVA
LOGRONO
SORIA
CUADALAJARA
CUENCA
ALBACETE
MURCIA
ALMERIA
SANTANDER
BURGOS
PALENCIA
VALLADOLID
SEGOVIA
MADRID
TOLEDO
CIUDAD REAL
JAEN
GRANADA
OVIEDO
LEON
ZAMORA
SALAMANCA
AVILA
CACERES
BADAJOZ
CORDOBA
SEVILLA
MALAGA
CADIZ
LUGO
ORENSE
HUELVA
LA CORUNA
PONTEVEDRA

4. POSITIONS COMPROMIS DES PROVINCES

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

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

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

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

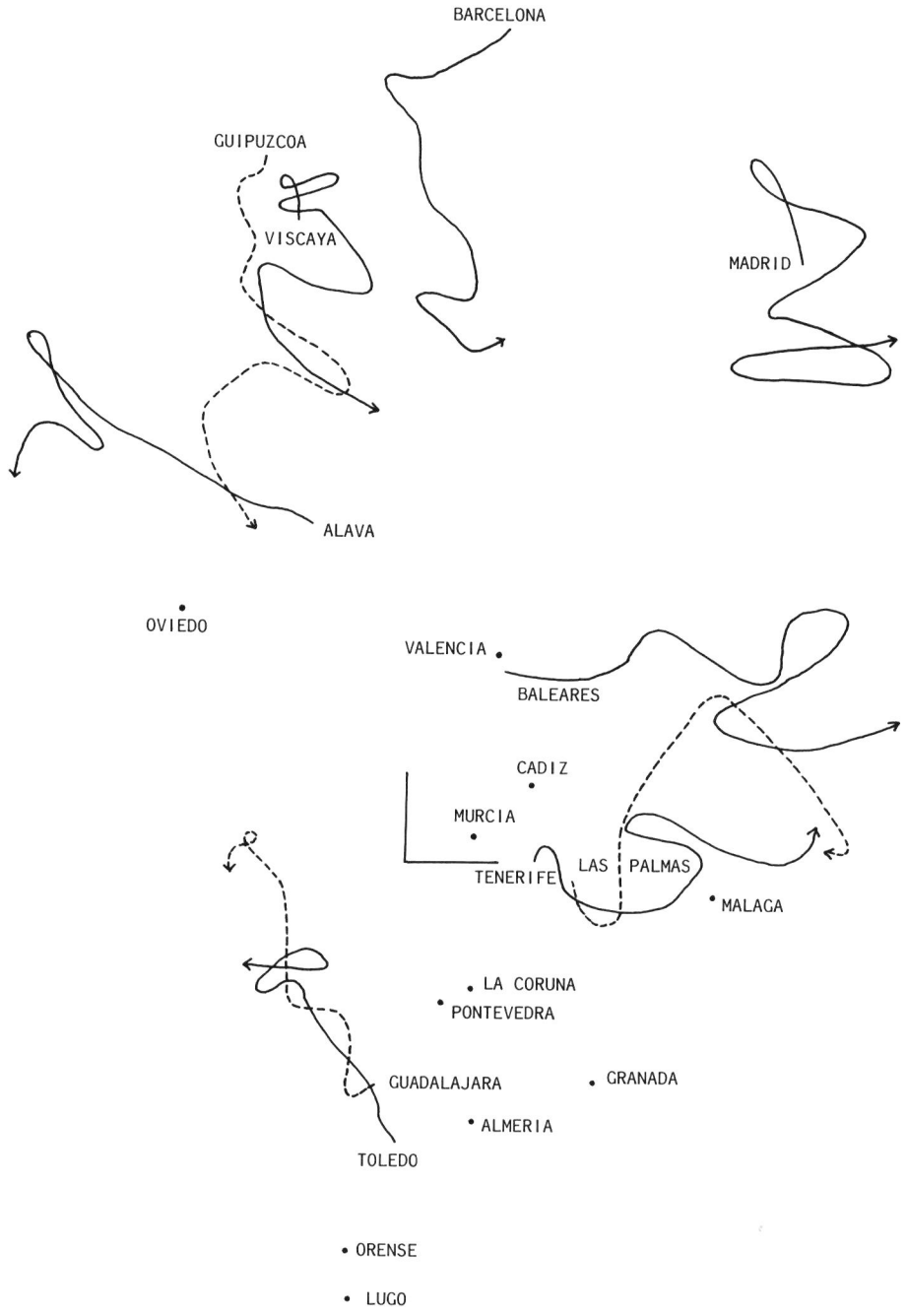


Figure n° 5 : POSITIONS COMPROMIS ET TRAJECTOIRES DES PROVINCES DANS LE PLAN 1.2.

5. TRAJECTOIRES DES PROVINCES

Relation entre les vecteurs propres de wd et de $w_k d$. Ces données ont été traitées, en utilisant une ancienne version du programme qui calculait les éléments propres de chaque opérateur $w_k d$, ainsi que les corrélations de ces vecteurs propres avec les vecteurs propres de l'opérateur compromis wd.

La version actuelle du programme n'effectue plus ces diagonalisations, car l'information est redondante avec les résultats de l'interstructure, comme nous allons le constater sur cet exemple. En effet, lorsque les applications linéaires w_k sont proches du compromis, les vecteurs propres des opérateurs $w_k d$ et wd, associés aux plus grandes valeurs propres sont en général voisins.

VALEURS PROPRES DE wd ET DES $w_k d$

	wd	1960	62	64	67	69	71	73	75	77	79
λ_1	8.03	8.54	8.49	8.67	8.44	8.16	7.88	7.42	7.41	6.96	6.53
λ_2	2.24	2.02	1.98	2.12	2.23	2.45	2.38	2.39	2.52	2.84	3.13
λ_3	1.48	1.74	1.67	1.54	1.72	1.59	1.92	2.13	1.98	2.10	2.00

CORRELATIONS DES VECTEURS PROPRES $\varepsilon^*_{k_\ell}$ DE $w_k d$ AVEC LES VECTEURS PROPRES ε^*_ℓ DE wd

	1960	62	64	67	69	71	73	75	77	79
$(\varepsilon^*_{k_1} \mid \varepsilon^*_1)$	.98	.98	.99	.99	.99	.99	.98	.98	.98	.97
$(\varepsilon^*_{k_2} \mid \varepsilon^*_2)$	.86	.90	.93	.95	.94	.92	.88	.94	.94	.95
$(\varepsilon^*_{k_3} \mid \varepsilon^*_3)$	.79	.88	.84	.84	.87	.86	.79	.85	.79	.84

Sur ces tableaux, on constate que les corrélations entre les vecteurs propres des opérateurs $w_k d$ et wd sont élevées. Puisque ces vecteurs sont d-normés, cela signifie qu'ils sont voisins. Les valeurs propres étant, grosso modo, du même ordre de grandeur, on peut considérer que les projections des trajectoires des provinces, sur les deux premiers axes, représentent ce qu'on aurait obtenu, en superposant les plans principaux de chaque étude.

Explication des trajectoires. Nous allons expliquer l'évolution des provinces dont les déplacements sont importants, sur les figures n° 5 et n° 6. Les tableaux de la p.170 permettent un retour aux données, pour les années extrêmes 1960 et 1979.

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

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

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

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

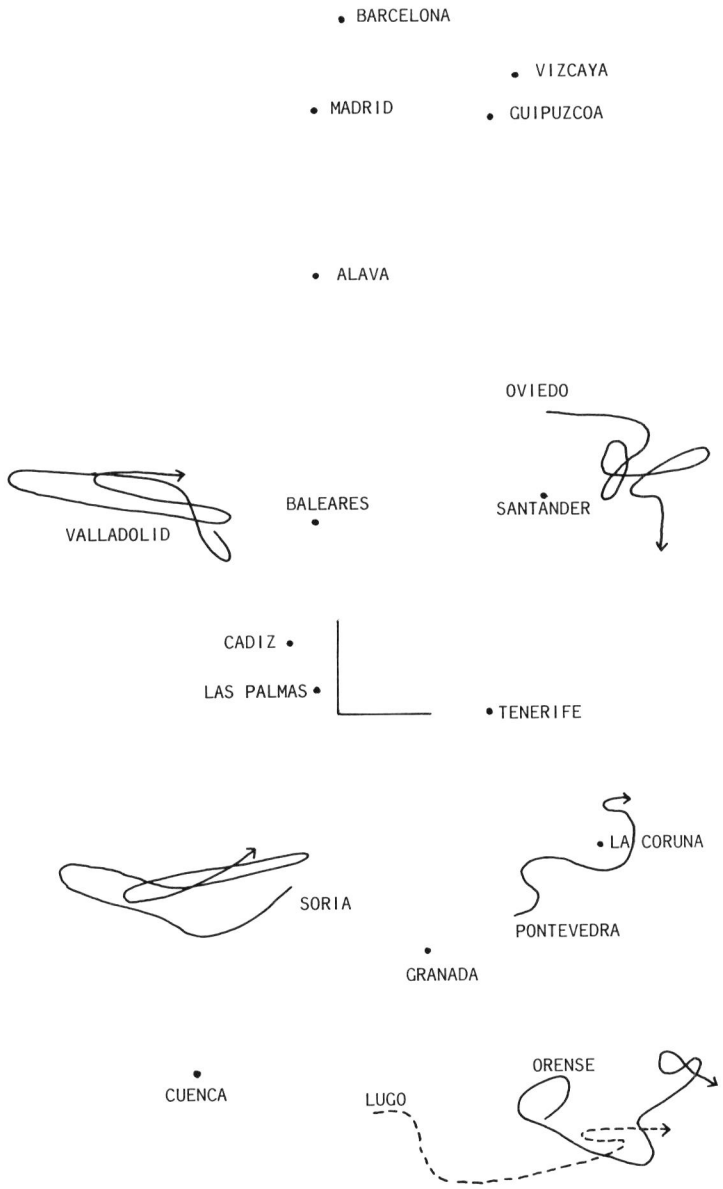


Figure n° 6 : POSITIONS COMPROMIS ET TRAJECTOIRES DES PROVINCES DANS LE PLAN 1.3.

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

6. CONCLUSIONS

Dans cet exemple, on s'est intéressé essentiellement aux évolutions individuelles. L'interstructure a montré une évolution régulière des distances entre provinces, au cours de la période étudiée. Les corrélations des variables avec les axes du compromis, ont été calculées dans le but d'interpréter les positions compromis des provinces, et les mouvements de leurs trajectoires. De plus, ces corrélations ont permis de mettre en évidence des tendances générales dans l'évolution des relations entre les différents secteurs, que nous résumons brièvement.

Les résultats des secteurs industriel et tertiaire, contrairement à ceux de l'agriculture, sont associés au niveau de développement, mais avec des rythmes et des intensités différents. Les résultats du secteur agricole peuvent être caractérisés par la productivité du travail, et le revenu des non salariés, sans tenir compte du niveau des salaires qui dépend des caractéristiques économiques générales.

La signification économique des axes est simple, et permet une interprétation claire des mouvements des trajectoires des provinces. Les trajectoires montrent la diversité des situations rencontrées. Le phénomène le plus remarquable est la rapide expansion du secteur tertiaire, dans quelques provinces. La crise économique des années 70 est une crise essentiellement industrielle, et sa manifestation la plus aiguë a eu lieu au Pays Basque. Quant aux résultats économiques du secteur agricole, les écarts entre les régions qui affichent des niveaux extrêmes semblent s'être creusés, entre le début et la fin de la période étudiée.

Tableau n° 1 : DONNEES CONCERNANT LES PROVINCES QUI ONT CONNU UNE PROGRESSION, OU UNE REGRESSION IMPORTANTE DE LEUR NIVEAU DE DEVELOPPEMENT. LA MOYENNE DE CHAQUE VARIABLE POUR UNE ANNEE A ETE RAMENEE A 100.

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

Tableau n° 2 : PROVINCES EN FORTE PROGRESSION DANS LE SECTEUR SERVICES

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

Tableau n° 3 : RESULTATS DU SECTEUR AGRICOLE POUR LES PROVINCES A ORIENTATION BOVINE.

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

VIII. Typologie des cantons ruraux de l'Hérault, décrits par l'évolution de leur population active entre 1954 et 1982

1. DESCRIPTION DES DONNEES

1.1. LE CHANGEMENT SOCIAL DANS LES CAMPAGNES LANGUEDOCIENNES

Depuis la fin de la seconde guerre mondiale, la structure socio-économique des campagnes de la région Languedoc-Roussillon a été profondément bouleversée. L'évolution de la répartition des actifs selon leur catégorie socioprofessionnelle, au cours des différents recensements, donne une idée de ce bouleversement.

Il y a une trentaine d'années, l'économie régionale reposait sur la viticulture, pratiquement monoculture, qui utilisait une abondante main d'oeuvre salariée. Cela donnait une certaine uniformité à la plaine languedocienne et aux coteaux de l'arrière-pays. Seuls de petits centres industriels, héritiers d'une tradition textile, ou implantés sur les bassins miniers des Cévennes, utilisaient la main d'oeuvre disponible localement puisque l'agriculture, dans ces régions de moyenne montagne, avait pratiquement disparu. Les villes vivaient du commerce du vin comme Béziers et Narbonne, ou de fonctions administratives comme Montpellier.

AURIAC F., BERNARD M.C., LOCHARD E. - Le changement social dans les campagnes languedociennes, Espace Géographique, n° 4, p. 239-250, 1975.

(*) Ce chapitre a été écrit en collaboration avec Marie-Claire BERNARD, Maître de Conférences, Laboratoire de Géographie Rurale, Université Paul Valéry, MONTPELLIER.

Les changements observés depuis cette époque peuvent s'expliquer par les difficultés rencontrées par les viticulteurs, le déclin des petites industries, conjugués au développement urbain et à une croissance démographique, sensible dès 1962 avec le retour des rapatriés d'Afrique du Nord, et amplifiée depuis 1975 au point d'atteindre les taux records de France. Dans ce contexte, les campagnes ont évolué de façon différente, suivant qu'elles étaient proches ou non des centres urbains.

D'une manière générale, les cantons situés dans l'arrière pays déclinent régulièrement, au rythme de la disparition des exploitations agricoles. Toutefois, certains hauts cantons de la bordure du Massif Central ont bénéficié de la tendance écologiste qui a entraîné des retours au pays, et l'implantation de familles attirées par un mode de vie différent. Mais nous verrons que les changements les plus notables sont dus à l'installation des "nouveaux languedociens", en majorité des cadres et des employés, dans les bourgs et villages proches des agglomérations.

1.2. LES CANTONS RURAUX DE L'HERAULT

Choix du département de l'Hérault. Entre 1975 et 1982, la définition des catégories socioprofessionnelles a changé. Notre but est de montrer que la méthode STATIS peut encore s'appliquer. Pour cela, nous avons retenu les cantons ruraux du département de l'Hérault car ce fichier, malgré sa petite taille, résume les différentes évolutions des cantons des cinq départements de la région (BERNARD - LAVIT (1985)). La dominante est agricole pour la moitié d'entre eux, industrielle pour le quart environ, et tertiaire pour le reste. On retrouve les différents profils, des stations balnéaires du littoral méditerranéen au rebord du Massif Central, sur le plateau du Larzac, en passant par les petits centres industriels de Ganges, Lodève et Bédarieux ; et des cantons périurbains de l'est qui profitent des retombées tertiaires de la métropole régionale montpelliéraine, aux cantons de l'ouest du département, restés pendant longtemps essentiellement viticoles, aux portes de Béziers.

Définition d'un canton rural. Les données ont été collectées dans le cadre des cantons "ruraux", c'est-à-dire qu'ont été supprimées les communes urbaines de plus de 5 000 habitants en 1954. Ce seuil, différent de celui de l'INSEE (2 000 habitants) a paru plus adapté au Languedoc où, à l'époque, de nombreux bourgs viticoles ne pouvaient pas être déclarés urbains pour autant.

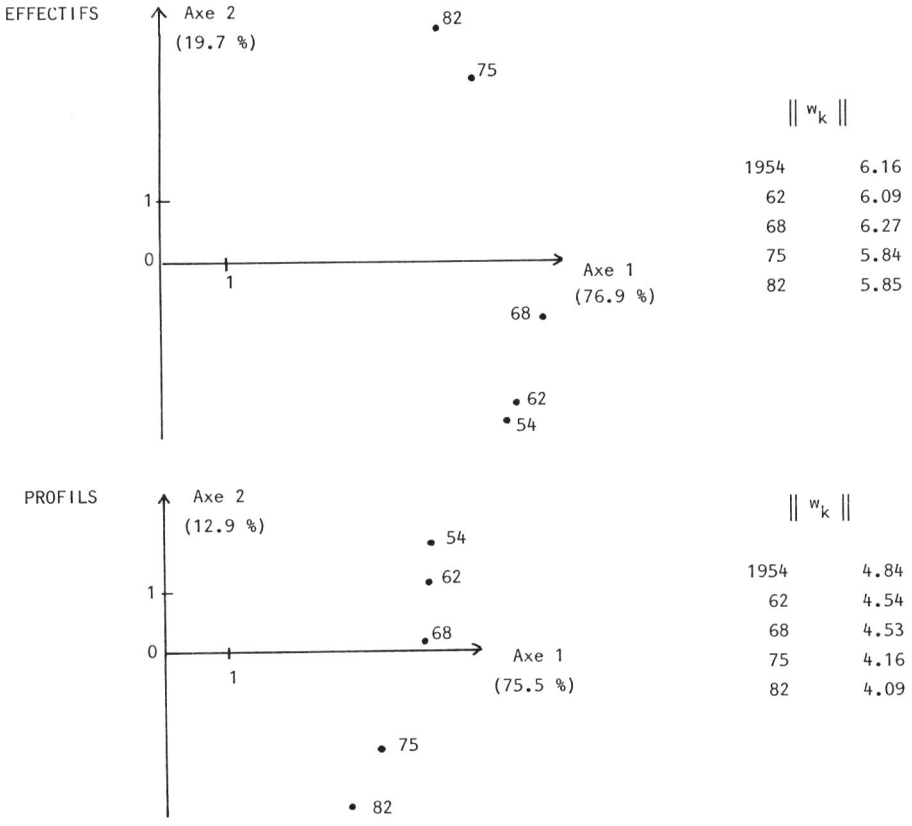
Changement dans la définition des catégories socioprofessionnelles. La nomenclature de l'INSEE a été modifiée en 1982. Pour les recensements de 1954, 1962, 1968 et 1975, la population des actifs est répartie en neuf catégories socioprofessionnelles : exploitants agricoles - salariés agricoles - artisans, et patrons de l'industrie et du commerce - professions libérales et cadres supérieurs - cadres moyens - employés - ouvriers - personnels de service - clergé, armée, police. En 1982, il ne reste que huit catégories socioprofessionnelles. La catégorie disparate "clergé, armée, police" a été supprimée : le clergé est désormais rattaché à la nouvelle catégorie des professions intermédiaires (anciennement cadres moyens), l'essentiel de l'armée et de la police est classé avec les employés. La catégorie personnels de service est restreinte aux "personnels des services aux particuliers", ce qui la réduit sensiblement. Précisons enfin qu'en dehors des cantons à dominante industrielle comme Ganges, Bédarieux ..., la catégorie ouvriers se compose essentiellement de personnes travaillant dans le bâtiment, ou dans de petites entreprises.

Validité des données. Il faut signaler également que les valeurs recueillies en 1975 et en 1982, ne sont fournies par l'INSEE qu'au travers de sondages (respectivement au 1:5 et au 1:4), alors que les dépouillements des recensements de 1954, 1962 et 1968 sont exhaustifs. L'erreur aléatoire n'est donc pas négligeable sur des effectifs aussi faibles. C'est la raison pour laquelle nous tenons à préciser que ces données ont pour seul but l'illustration des possibilités de la méthode STATIS. Néanmoins, les trajectoires des cantons suivent des chemins vraisemblables au regard des évolutions constatées sur le terrain.

TABLEAUX DES PRODUITS SCALAIRES NORMES : $RV(k_1, k_2) = \frac{(w_{k_1} \mid w_{k_2})}{\| w_{k_1} \| \; \| w_{k_2} \|}$

		1954	62	68	75
EFFECTIFS	1954				
	62	.958			
	68	.873	.936		
	75	.525	.576	.734	
	82	.413	.442	.603	.934

		1954	62	68	75
PROFILS	1954				
	62	.871			
	68	.723	.868		
	75	.600	.676	.735	
	82	.475	.530	.602	.705



	$\| w_k \|$
1954	6.16
62	6.09
68	6.27
75	5.84
82	5.85

	$\| w_k \|$
1954	4.84
62	4.54
68	4.53
75	4.16
82	4.09

Figures n° 1 : IMAGE EUCLIDIENNE DES RECENSEMENTS, ASSOCIEE A L'APPROMIMATION D'ORDRE 2 DES PRODUITS SCALAIRES $(w_{k_1} | w_{k_2})$

1.3. CHOIX EFFECTUES POUR L'UTILISATION DE LA METHODE

Le fichier des effectifs, donné p.188 avec le fichier des profils en pourcentage, comporte 5 tableaux de contingence, croisant les 32 cantons et les catégories socioprofessionnelles. Chaque tableau correspond à un recensement. Sète ne figure pas dans ce fichier car c'est un canton urbain. Les chiffres de Montpellier et de Béziers, représentant la couronne rurale de ces villes, sont disproportionnés par rapport aux autres.

Nous avons analysé simultanément les effectifs bruts et les profils en pourcentage, car les informations sont complémentaires. Nous avons supprimé Montpellier et Béziers de l'analyse des effectifs, puisque les cantons se répartissent essentiellement d'après leur taille, et nous les avons traités en éléments supplémentaires dans l'analyse des profils.

Les catégories socioprofessionnelles, jouant le rôle des variables, ne sont pas les mêmes d'un tableau à l'autre. On est, par conséquent, dans la situation n° 1 (p. 81). La comparaison des recensements au moyen des objets w_k nous fournira une typologie des cantons. Montpellier et Béziers ont un poids nul, les autres cantons ont un poids égal à $\frac{1}{30}$. Les variables ont été centrées et réduites par tableau dans les deux analyses. Nous avons accordé aux cinq recensements, le même poids. Mais vu le changement de définition des catégories socioprofessionnelles, on aurait pu traiter le dernier recensement en élément supplémentaire, sur la base formée par les quatre autres.

2. ANALYSE DE L'INTERSTRUCTURE

On a noté sur la même page, les coefficients RV entre études, pour les deux analyses (sur les effectifs et sur les profils). On remarque que les coefficients RV, calculés sur les profils, décroissent régulièrement au cours des recensements alors que pour les effectifs, les coefficients RV entre deux tableaux successifs sont très élevés, sauf entre 1968 et 1975.

Disparité entre 1968 et 1975. Cette cassure entre 1968 et 1975 est d'ailleurs importante dans les deux analyses, pour les coefficients RV et pour les normes des w_k.

Si on étudie l'évolution du canton moyen, on s'aperçoit que cette période est marquée par de profonds bouleversements, caractérisés par une baisse importante des exploitants et ouvriers agricoles, et une augmentation des catégories tertiaires.

EVOLUTION DU CANTON "MOYEN"

EFFECTIFS

	EXP.AGR.	OUV.AGR.	IND.COM.	PROF.LIB.	CAD.MOY.	EMPLOYES	OUVRIERS	SERVICE	CLERGE...	TOTAL
1954	796	660	264	21	77	90	397	49	27	2381
62	649	534	243	26	97	111	472	51	40	2223
68	541	392	238	37	123	152	611	66	33	2193
75	396	247	244	90	204	224	660	83	40	2188

	EXP.AGR.	OUV.AGR.	IND.COM.	PROF.LIB.	PROF.INTER.	EMPLOYES	OUVRIERS	SERVICE	TOTAL
82	382	155	313	164	423	570	715	81	2803

PROFILS EN POURCENTAGE

	EXP.AGR.	OUV.AGR.	IND.COM.	PROF.LIB.	CAD.MOY.	EMPLOYES	OUVRIERS	SERVICE	CLERGE...
1954	35.42	26.46	10.37	0.85	3.27	3.69	16.68	2.02	1.25
62	30.86	23.10	10.31	1.15	4.41	4.86	21.17	2.18	1.97
68	27.32	16.82	10.20	1.69	5.58	6.67	27.14	2.89	1.70
75	20.97	10.52	11.15	3.74	8.91	9.49	29.56	3.54	2.12

	EXP.AGR.	OUV.AGR.	IND.COM.	PROF.LIB.	PROF.INTER.	EMPLOYES	OUVRIERS	SERVICE
82	17.01	5.76	11.38	4.83	13.95	18.73	25.68	2.66

Compte tenu des ressemblances entre les répartitions des cantons selon leurs effectifs, en 54 et 62, 62 et 68, 75 et 82, la rupture entre 68 et 75 est plus marquée que pour les profils. Bien que la taille du canton moyen soit stable, les trajectoires mettront en évidence des variations d'effectifs très importantes, pour le tiers des cantons, essentiellement pendant cette période. Les cantons à dominante industrielle comme Ganges, St Pons, et les cantons agricoles comme St Chinian, Gignac et Capestang perdent une grande partie de leur population. Alors que pendant la même période, certains cantons proches de l'agglomération montpelliéraine, comme Les Matelles, Mauguio, Castries ou Frontignan, connaissent une très forte augmentation des catégories tertiaires.

Ressemblance entre 1975 et 1982. Dans l'analyse des effectifs, w_{75} et w_{82} ont pratiquement des normes égales, et sont donc très voisins : on en conclut qu'entre ces deux dates, les positions mutuelles des cantons n'évoluent pas. Ceci paraît surprenant, d'autant plus que tous les effectifs augmentent dans de très fortes proportions entre ces deux recensements. Nous verrons que cette stabilité provient de ce que les cantons se répartissent principalement d'après leur taille (la dispersion selon la taille représente 57 % de la dispersion totale). Par conséquent, les changements de profil affectent peu cette répartition. Les petits cantons de La Salvetat et de Claret en seront un bon exemple.

3. TYPOLOGIE DES CANTONS, D'APRES LEURS POSITIONS COMPROMIS

3.1. ANALYSE DES EFFECTIFS

Explication des axes du compromis. Les corrélations des catégories socioprofessionnelles avec les axes du compromis (figure n° 3) montrent que le premier axe est un axe de taille. En effet, toutes les catégories, et plus particulièrement les catégories tertiaires, sont corrélées négativement avec cet axe. L'explication du deuxième axe n'est pas stable au cours du temps. Il oppose les catégories agricoles aux ouvriers, en début de période, et aux catégories tertiaires, en fin de période.

Positions compromis des cantons. Sur la figure n° 2, les cantons se répartissent d'abord par leur taille, sur le premier axe qui représente 57 % de l'inertie totale. Le deuxième axe oppose les cantons agricoles aux cantons à dominante ouvrière, ou tertiaire (en fin de période). Sur le troisième axe (non représenté) les cantons ouvriers (St Pons, St Gervais, Ganges, Bédarieux et Mèze) s'opposent, très nettement, aux cantons tertiaires (Les Matelles, Mauguio et Castries).

La séparation très marquée entre les cantons agricoles et les cantons à dominante ouvrière ou tertiaire, s'explique par le fort pourcentage de cantons purement agricoles, peu touchés par la tertiarisation, même en fin de période. On peut citer Gignac, Capestang, Servian, Olonzac, Roujan, Olargues (mais aussi Montagnac, St Chinian et Murviel qui se superposent à Servian sur le figure n° 2).

CANTONS DE
L'HERAULT
LUNEL
MAUGUIO
CASTRIES
CLARET
LES
MATELLES
MONTPELLIER
FRONTIGNAN
SETE
GANGES
SAINT MARTIN
ANIANE
GIGNAC
MEZE
LE CAYLAR
LODEVE
CLERMONT
MONTAGNAC
FLOREN-
SAC
AGDE
PEZENAS
ROUJAN
SERVIAN
BEZIERS
LUNAS
BEDARIEUX
MURVIEL
SAINT
GERVAIS
OLARGUES
SAINT CHINIAN
CAPESTANG
LA SALVETAT
SAINT PONS
OLONZAC

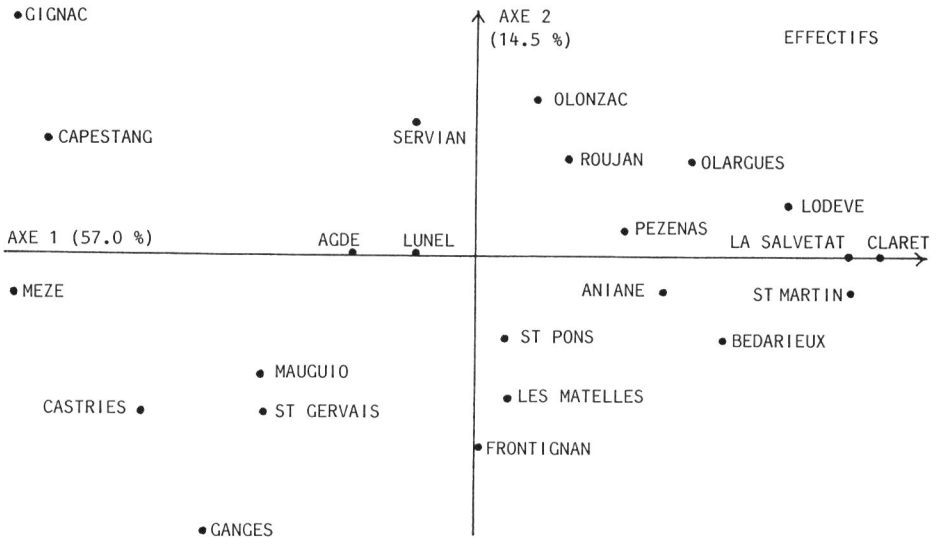


Figure n° 2 : POSITIONS COMPROMIS DES CANTONS
DANS LE PLAN 1.2

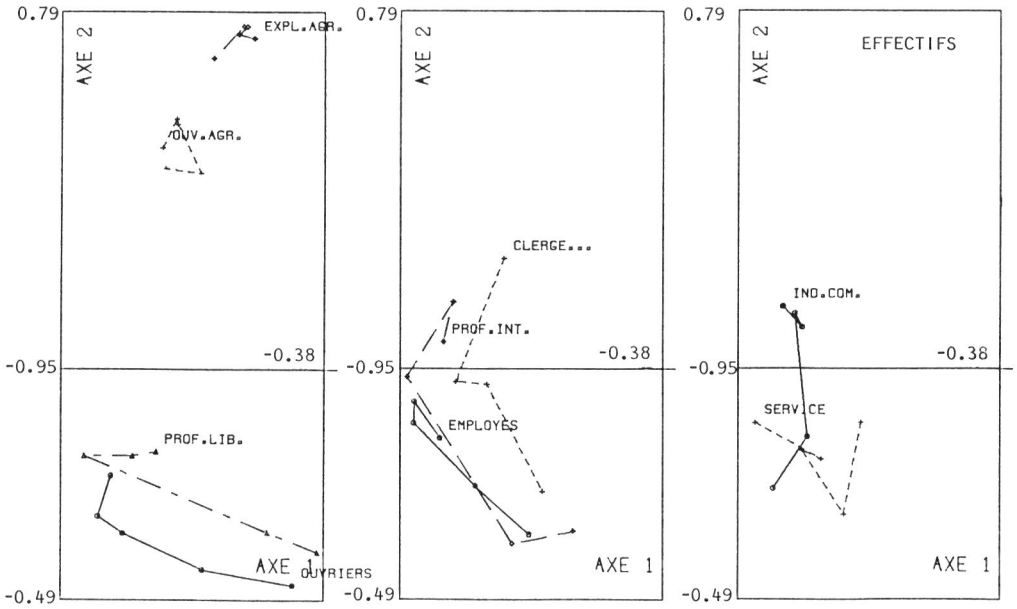


Figure n° 3 : CORRELATIONS DES CATEGORIES SOCIOPROFESSIONNELLES
AVEC LES AXES DU COMPROMIS

(AFIN D'EVITER LA SUPERPOSITION DES TRAJECTOIRES, CE GRAPHIQUE
EST ECLATE EN TROIS, LE NOM APPARAIT AU DEBUT DE LA TRAJECTOIRE)

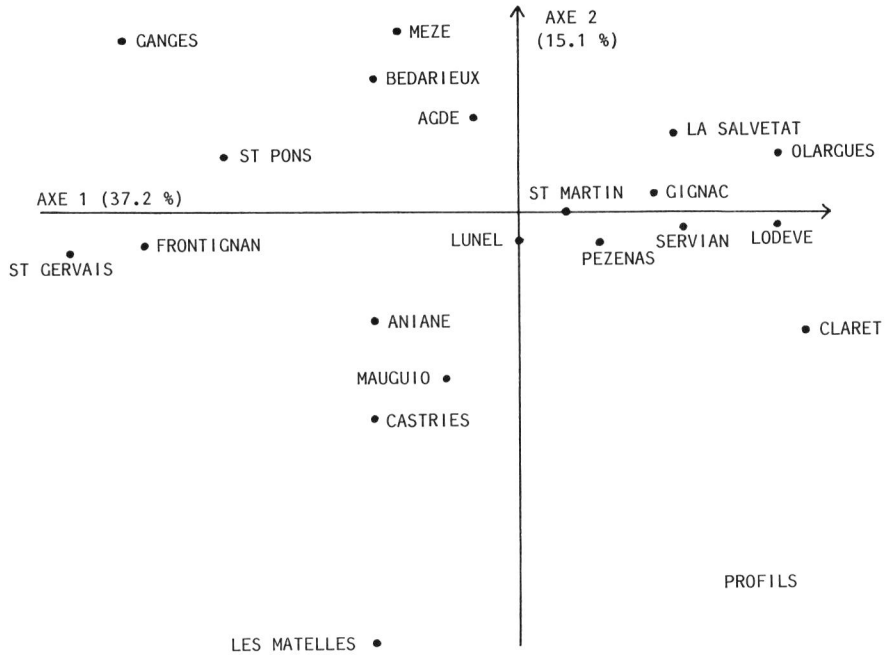


Figure n° 4 : POSITIONS COMPROMIS DES CANTONS DANS LE PLAN 1.2

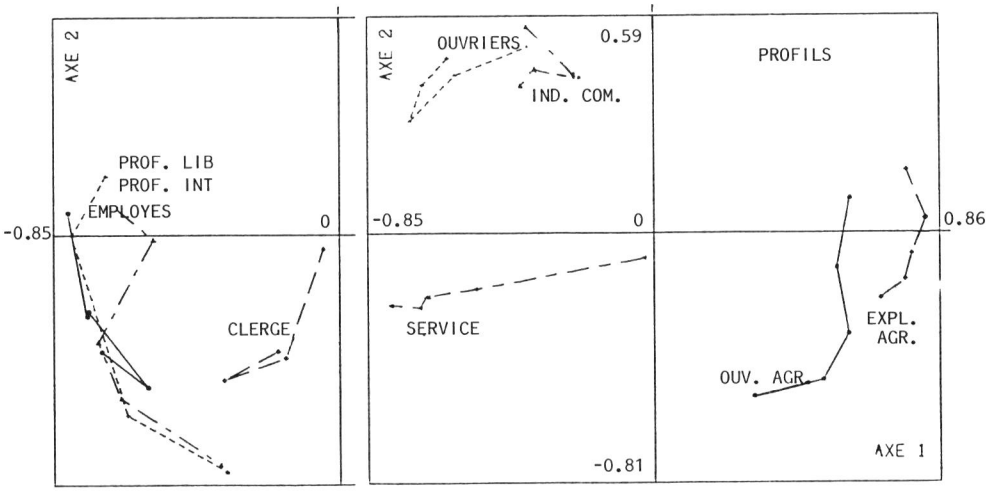


Figure n° 5 : CORRELATIONS DES CATEGORIES SOCIOPROFESSIONNELLES AVEC LES AXES DU COMPROMIS

(AFIN D'EVITER LA SUPERPOSITION DES TRAJECTOIRES, LA PARTIE NEGATIVE DE L'AXE 1 EST ECLATEE EN DEUX)

3.2. ANALYSE DES PROFILS

Explication des axes du compromis. La taille d'un canton n'intervenant plus, on retrouve évidemment sur le premier axe de la figure n° 5, l'opposition entre catégories agricoles et autres catégories (principalement les catégories tertiaires). L'explication du deuxième axe n'est pas stable au cours du temps, puisqu'il oppose les catégories secondaires (ouvriers, patrons de l'industrie et du commerce) aux ouvriers agricoles en 1954, et aux catégories tertiaires une trentaine d'années plus tard. Remarquons enfin que la restriction en 1982 de la catégorie "personnels de service" aux "services aux particuliers" lui fait perdre son rôle explicatif de l'axe 1, aux côtés des catégories tertiaires.

Positions compromis des cantons. Sur la figure n° 4, les trois types de cantons : agricoles, industriels et tertiaires se séparent nettement. Afin de rendre ce graphique plus lisible, nous n'avons pas noté Murviel, Montagnac, Florensac, Olonzac, St Chinian, Roujan, Le Caylar qui se trouveraient au centre du groupe formé par les cantons agricoles, ni Capestang et Clermont qui se situeraient au voisinage de Lunel. Le groupe dense des cantons agricoles se sépare sur le troisième axe (non représenté) suivant l'importance du pourcentage d'ouvriers agricoles, ou d'exploitants agricoles. Notons enfin la position centrale de Lunel, et la position isolée de Claret par rapport aux groupes agricoles et tertiaires.

4. TYPOLOGIE DES CANTONS, D'APRES LEURS TRAJECTOIRES

Nous avons tracé les trajectoires de quelques cantons, représentatifs des groupes observés sur les positions compromis. On interprète les déplacements d'une trajectoire, par les écarts de l'évolution du canton à l'évolution moyenne (p. 176). Entre 1954 et 1982, le canton moyen perd son caractère agricole (chute de la moitié des exploitants agricoles, et des trois quarts des ouvriers agricoles) pour gagner des actifs dans les secteurs secondaire (la population ouvrière passe du simple au double), et tertiaire (le nombre d'actifs quintuple).

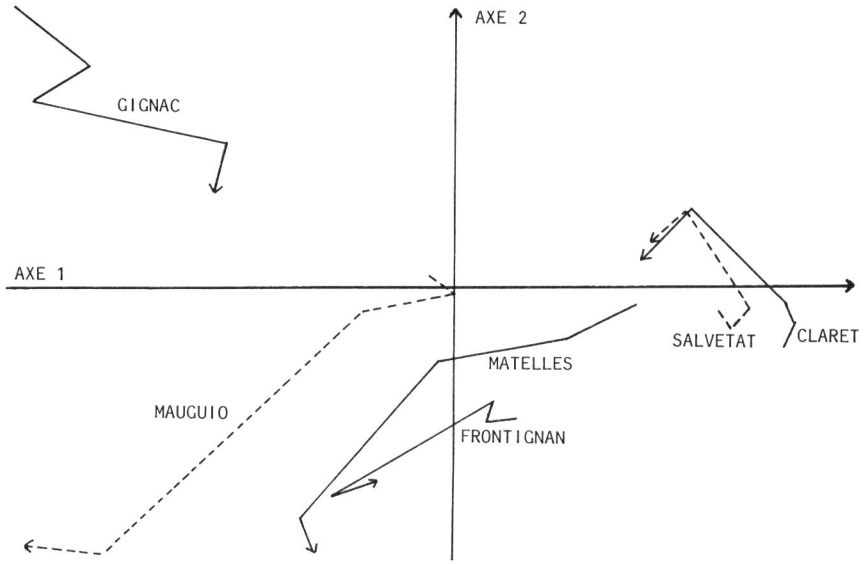
4.1. ANALYSE DES EFFECTIFS

Trajectoires amples. Le premier axe est un axe de taille. Par conséquent, les trajectoires (des figures n° 6 et n° 8) qui présentent des déplacements importants le long de cet axe, sont celles de cantons en expansion (déplacement vers la gauche, comme Mauguio et Les Matelles), ou en récession (déplacement vers la droite comme Gignac, Ganges et St Pons). Remarquons que ces déplacements, très importants entre 1968 et 1975 expliquent l'écart que nous avions constaté entre w_{68} et w_{75}, lors de l'étude de l'interstructure.

Le deuxième axe oppose les catégories agricoles, aux ouvriers en début de période, puis aux catégories tertiaires en fin de période. Une tertiarisation très forte se traduit par une "plongée" des trajectoires le long de l'axe 2. Ce phénomène touche les cantons proches de Montpellier, et s'accompagne d'un développement important de la population. Il en résulte que ces trajectoires sont, en réalité, obliques comme celles de Mauguio et des Matelles ou, plus modestement, comme celle de Frontignan.

A l'opposé, les cantons industriels comme Ganges, St Pons et Bédarieux ont des trajectoires qui évoluent exactement dans l'autre sens. Ces cantons, ouvriers en début de période, perdent leur population ouvrière, sans pour autant gagner de façon sensible des actifs dans le secteur tertiaire. Les trajectoires se déplacent par conséquent selon un mouvement ascendant comme Bédarieux (et Lunas, non représenté car semblable à Bédarieux), ou oblique lorsque cette récession s'accompagne d'un important dépeuplement, comme Ganges et St Pons (ou encore St Gervais entre 1968 et 1975, que nous n'avons pas représenté).

Trajectoires restreintes. Certaines trajectoires tournent sur elles-mêmes comme celles de Servian, Agde, Aniane, Olargues et Lunel. En effet, ces cantons suivent l'évolution moyenne, tout en gardant leur caractéristique : Servian est un canton plus important (plus à gauche sur l'axe 1) et plus agricole (plus haut sur l'axe 2) qu'Aniane. Agde est le canton moyen par excellence.



TRAJECTOIRES DES CANTONS DANS LE PLAN 1.2

Figure n° 6 : EFFECTIFS

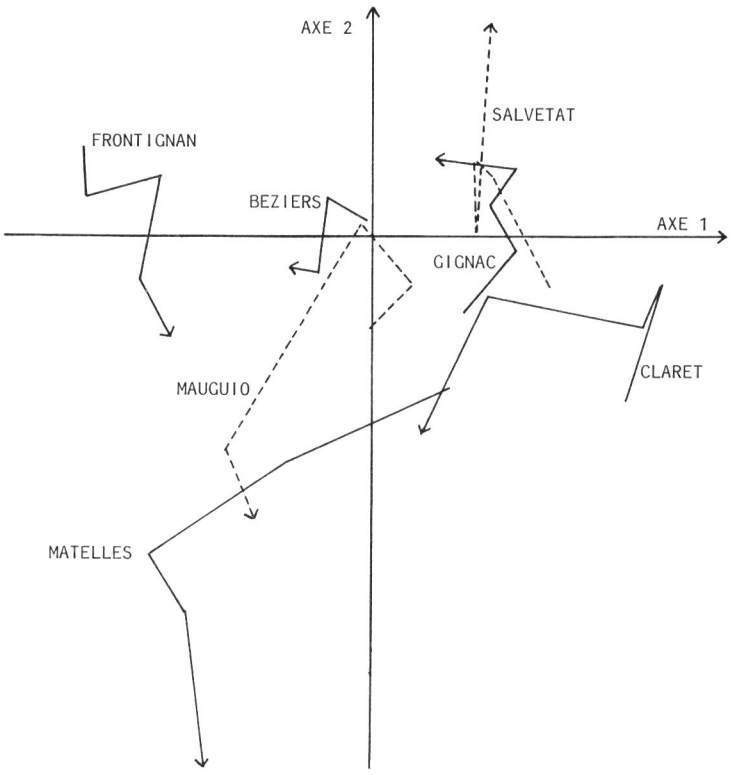
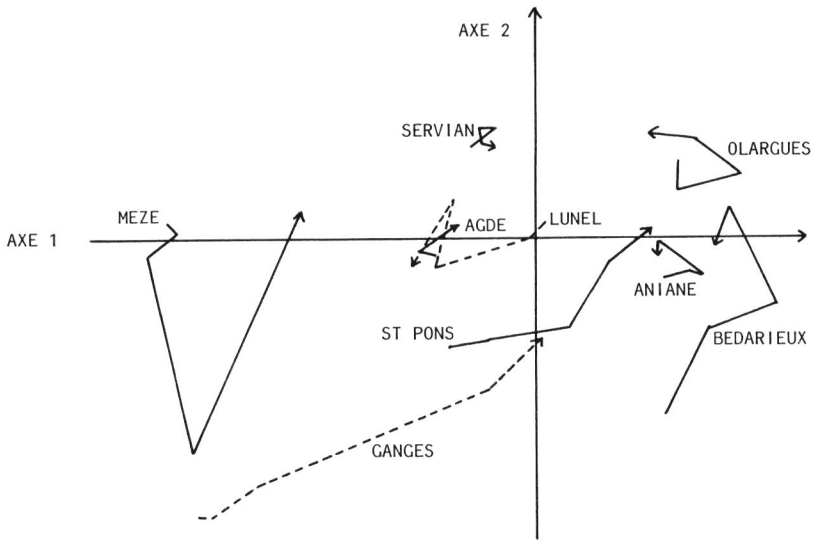


Figure n° 7 : PROFILS



TRAJECTOIRES DES CANTONS DANS LE PLAN 1.2

Figure n° 8 : EFFECTIFS

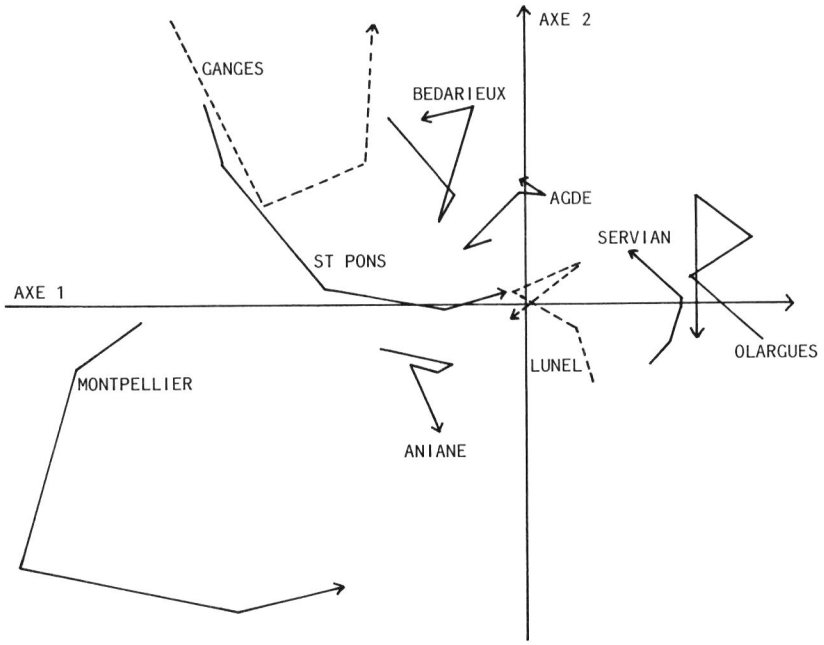


Figure n° 9 : PROFILS

Trajectoires irrégulières. D'autres trajectoires sont moins régulières que celles que nous venons de décrire. Ainsi la chute brutale de la trajectoire de Mèze le long de l'axe 2 s'explique par la forte baisse des viticulteurs et des ouvriers agricoles, accompagnée d'un gain en ouvriers, plus forts que l'évolution moyenne. La remontée de la trajectoire après 1975 est une anomalie, due au changement dans la définition des catégories socioprofessionnelles. En effet, les communes du canton, riveraines de l'étang de Thau, sont spécialisées dans la conchyliculture. Or les éleveurs d'huîtres et de moules ont été décomptés comme artisans jusqu'en 1975, puis assimilés aux agriculteurs en 1982. Ce changement de catégorie a eu pour effet de doubler localement l'effectif des exploitants agricoles, contrairement à la tendance générale.

Remarquons les trajectoires semblables de La Salvetat, de Claret (et du Caylar, de Lodève et de St Martin non représentés). Le point de départ, assez bas sur l'axe 2, montre que ces petits cantons ne sont pas exclusivement agricoles (contrairement à Gignac, Servian et Olonzac dont la trajectoire se situe entièrement du côté positif de l'axe 2). L'évolution suit l'évolution du canton moyen jusqu'en 1968, puis s'en éloigne. Mais la tertiarisation, pas assez marquée, provoque la remontée de la trajectoire le long de l'axe 2. La reprise démographique, essentiellement dans le secteur tertiaire, infléchit ensuite la trajectoire à partir de 1975.

4.2. ANALYSE DES PROFILS

Rappelons que, dans cette analyse, l'axe 1 oppose les catégories agricoles aux autres catégories. L'axe 2 oppose les ouvriers et patrons de l'industrie et du commerce, aux catégories tertiaires en fin de période.

Cantons agricoles. Sur les figures n° 7 et n° 9, examinons tout d'abord les trajectoires des cantons agricoles. Servian et Gignac (mais aussi Montagnac, Roujan, Olonzac, St Chinian, Murviel, Capestang et Le Caylar dont les trajectoires se superposeraient à celle de Gignac) sont des cantons purement agricoles. Leur trajectoire suit le déplacement des catégories agricoles, parallèlement à l'axe 2.

Par contre, La Salvetat a une trajectoire légèrement différente : le changement de direction entre 1968 et 1975 s'explique par la perte d'ouvriers, et le gain en professions libérales. Les trajectoires d'Olargues et de Pézenas (non dessinée) reviennent sur elles-mêmes après 1975, et reflètent de ce fait une tertiarisation assez forte (accompagnée dans le cas d'Olargues, d'une perte en ouvriers et artisans).

Cantons tertiaires en fin de période. Une très forte poussée des catégories tertiaires se traduit par une "plongée" de la trajectoire vers la partie négative de l'axe 2. C'est le cas de Montpellier entre 1954 et 1968, et des cantons proches comme Les Matelles (ou Castries non représenté), Mauguio à partir de 1968, et Claret dont la trajectoire est, cette fois-ci, franchement différente de celle de La Salvetat.

Cantons industriels en début de période. Ganges, St Pons, Bédarieux (ainsi que St Gervais et Lunas, non représentés) sont des cantons à dominante ouvrière : leur trajectoire est située dans la partie gauche supérieure du graphique. Mais si le profil de St Pons se rapproche du profil moyen en 1982, il n'en est pas de même pour Ganges et Bédarieux. La remontée de leurs trajectoires montre que ces cantons gardent encore un profil ouvrier en fin de période, car leur régénération par le tertiaire n'a pas été suffisante.

Cantons moyens. Enfin, on retrouve au centre du graphique Agde, Aniane et Lunel (particulièrement à partir de 1968) ainsi que Béziers (et Clermont l'Hérault non représenté). Ces cantons suivent pratiquement l'évolution du canton moyen.

5. CONCLUSIONS

La modification dans la définition des catégories socioprofessionnelles n'a pas été gênante, dans la mesure où les axes s'interprètent en fonction des pôles agricole, secondaire et tertiaire. La catégorie clergé, armée, police, supprimée en 1982, et la catégorie personnels de service, profondément modifiée, n'interviennent pratiquement pas dans l'explication des positions compromis des cantons.

Dans les deux analyses, les grandes variations des corrélations des catégories tertiaires avec les axes du compromis, donnent une première explication des transformations intervenues pendant cette période. En 1954, dans un monde rural relativement homogène, dominé par la viticulture, ce sont les cantons industriels qui s'individualisent. A partir de 1968, la croissance accélérée et la diffusion des catégories tertiaires opposent les cantons des zones périurbaines, aux campagnes restées essentiellement viticoles. Parallèlement, le développement du secondaire (surtout le bâtiment) gomme les spécialisations locales, héritées d'une période d'industrialisation antérieure.

L'examen des trajectoires permet de donner une explication plus détaillée. L'importance des cantons, et la variation du nombre des actifs, sont prises en compte par les trajectoires sur les effectifs. Alors que la forme, et l'ampleur des évolutions, sont reflétées par les trajectoires sur les profils. Les deux approches, complémentaires, mettent en relief les disparités croissantes des évolutions locales.

En effet, au seul examen de l'évolution du canton moyen, on serait tenté d'interpréter le phénomène de tertiarisation générale comme un processus d'homogénéisation des sociétés locales. Or il n'en est rien. Les cantons à dominante agricole, dans leur majorité, changent peu : leurs effectifs sont stables ou diminuent, leur évolution tertiaire reste très en deçà de celle du canton moyen. En revanche, la mutation de petits cantons comme Claret ou Les Matelles qui sont profondément modifiés dans leur comportement socio-économique, s'explique par l'arrivée d'éléments extérieurs. Les nouveaux venus contribuent à la diversification des activités préexistantes (comme à Mauguio ou à Frontignan), ou en substituent de nouvelles à celles qui ont décliné (comme à St Pons).

La typologie des transformations suggère par conséquent, une évolution à deux vitesses d'un monde rural qui a perdu sa relative unité, en subissant l'évolution de la société globale et en recevant, de façon inégale, d'importants apports extérieurs.

POPULATION ACTIVE AU LIEU DE RESIDENCE, REGROUPEE SELON LES CATEGORIES SOCIOPROFESSIONNELLES

(DONNEES ISSUES DES RECENSEMENTS GENERAUX DE POPULATION. INSEE : DIRECTION REGIONALE DE MONTPELLIER)

Effectifs		EAGR	OAGR	INDU	LIBE	CADR	EMPL	OUVR	SERV	CLER	
	54										
AGDE		678	1072	420	26	88	141	506	61	17	3009
BEDARIEUX		312	123	111	10	68	84	1024	31	16	1779
BEZIERS		2237	2596	926	51	241	334	1135	194	50	7764
CAPESTANG		1234	1919	516	34	147	181	476	86	35	4628
FLORENSAC		686	802	212	16	55	53	201	29	20	2074
MONTAGNAC		1105	1145	291	21	91	91	242	52	34	3072
MURVIEL		1437	1129	320	27	86	105	223	46	25	3398
OLARGUES		1284	279	165	8	57	56	156	15	23	2043
OLONZAC		1362	668	315	15	77	82	235	69	28	2851
PEZENAS		480	675	191	11	58	58	183	25	15	1696
ROUJAN		932	718	274	20	78	99	197	35	19	2372
ST CHINIAN		1356	826	385	23	97	101	343	51	33	3215
ST GERVAIS		737	247	363	44	134	142	663	90	29	2449
ST PONS		567	175	297	47	115	91	736	48	38	2114
SALVETAT		612	84	101	4	23	28	86	14	19	971
SERVIAN		951	1306	336	23	75	87	208	49	29	3064
LE CAYLAR		232	99	48	2	18	11	78	9	13	510
CLERMONT		906	678	264	32	96	109	423	48	24	2580
GIGNAC		1989	1484	529	34	145	114	487	108	88	4978
LODEVE		836	285	70	2	37	23	128	15	11	1407
LUNAS		641	198	207	20	84	96	852	37	30	2165
ANIANE		532	266	123	18	58	58	133	40	26	1254
CASTRIES		935	1040	273	23	96	122	366	77	40	2972
CLARET		306	234	35	2	18	12	22	8	9	646
FRONTIGNAN		220	405	162	12	71	132	547	51	19	1619
GANGES		339	310	529	62	127	180	1787	115	31	3480
LUNEL		729	1006	212	21	54	101	243	55	24	2445
MATELLES		483	455	96	7	36	50	118	30	28	1303
MAUGUIO		646	1012	223	21	71	87	252	77	28	2417
MEZE		1011	942	798	37	124	179	877	93	47	4108
MONTPELLIER		1329	1705	741	102	288	450	1441	354	142	6552
ST MARTIN		335	213	61	9	22	17	133	14	17	821
	62										
AGDE		546	834	397	42	116	141	578	90	27	2771
BEDARIEUX		331	115	105	13	66	66	628	14	29	1367
BEZIERS		1969	2035	895	76	265	457	1440	215	99	7451
CAPESTANG		1095	1582	477	52	173	239	699	89	79	4485
FLORENSAC		593	625	196	20	82	78	286	32	27	1939
MONTAGNAC		887	956	241	26	120	117	373	30	53	2803
MURVIEL		1091	931	271	24	120	182	457	68	43	3187
OLARGUES		915	211	163	15	77	62	226	20	27	1716
OLONZAC		1143	632	266	22	101	89	267	35	34	2586
PEZENAS		450	560	162	21	76	69	276	26	47	1687
ROUJAN		820	611	251	25	102	109	198	44	37	2197
ST CHINIAN		1263	694	377	23	123	135	370	55	51	3091
ST GERVAIS		603	164	353	57	187	132	696	195	57	2444
ST PONS		394	179	268	47	151	106	713	40	41	1939
SALVETAT		367	139	99	8	26	32	183	14	15	883
SERVIAN		801	1086	274	30	99	117	364	44	34	2849
LE CAYLAR		144	84	41	2	11	10	55	6	16	369
CLERMONT		925	550	250	41	136	146	409	43	29	2529
GIGNAC		1717	1302	456	39	191	151	686	94	80	4716
LODEVE		568	237	55	4	50	43	130	17	15	1119
LUNAS		429	153	160	18	88	94	641	41	34	1658
ANIANE		427	203	128	16	68	68	160	30	38	1138
CASTRIES		693	740	220	26	127	217	579	93	57	2752
CLARET		262	181	39	4	26	12	43	2	7	576
FRONTIGNAN		165	301	162	28	82	142	713	64	31	1688
GANGES		216	176	481	67	141	209	1968	78	76	3412
LUNEL		610	739	194	22	91	128	585	64	32	2465
MATELLES		339	338	81	18	47	76	193	41	67	1200
MAUGUIO		518	831	204	25	69	140	499	68	30	2384
MEZE		918	722	878	45	135	202	1002	80	75	4057
MONTPELLIER		1158	1267	910	199	522	806	2397	416	279	7954
ST MARTIN		229	143	51	6	24	24	182	5	14	678

Effectifs	EAGR	OAGR	INDU	LIBE	CADR	EMPL	OUVR	SERV	CLER	
68										
AGDE	469	630	432	40	154	185	867	103	18	2898
BEDARIEUX	220	57	73	11	62	78	492	22	16	1031
BEZIERS	1530	1527	979	120	451	640	2261	292	86	7886
CAPESTANG	886	1173	452	77	274	338	1020	99	45	4364
FLORENSAC	505	464	199	19	106	82	393	17	24	1809
MONTAGNAC	833	700	259	42	131	172	532	54	52	2775
MURVIEL	944	708	258	32	156	232	642	82	36	3090
OLARGUES	751	134	135	14	65	62	205	17	22	1405
OLONZAC	992	404	218	28	90	127	305	51	30	2245
PEZENAS	377	446	165	28	82	106	379	46	42	1671
ROUJAN	723	474	228	23	108	110	356	57	16	2095
ST CHINIAN	981	496	341	48	138	141	452	75	45	2717
ST GERVAIS	482	100	291	71	232	174	708	233	59	2350
ST PONS	248	170	231	41	150	139	684	52	30	1745
SALVETAT	272	75	95	7	34	39	148	13	17	700
SERVIAN	716	861	275	42	117	137	548	89	32	2817
LE CAYLAR	123	46	28	3	12	12	80	4	11	319
CLERMONT	773	387	278	50	159	202	573	65	27	2514
GIGNAC	1481	980	497	63	229	247	1047	121	78	4743
LODEVE	456	155	57	14	44	54	142	16	11	949
LUNAS	334	125	132	26	79	100	532	34	30	1392
ANIANE	358	113	116	25	96	71	285	53	25	1142
CASTRIES	611	541	265	75	244	345	1114	110	48	3353
CLARET	248	139	28	10	24	19	55	9	11	543
FRONTIGNAN	164	195	199	31	109	198	895	72	25	1888
GANGES	147	113	353	72	176	296	2024	106	64	3351
LUNEL	514	574	202	39	136	246	1039	86	37	2873
MATELLES	296	256	106	67	128	129	375	60	65	1482
MAUGUIO	465	635	241	50	117	223	923	97	27	2778
MEZE	658	547	919	63	220	266	1306	132	50	4161
MONTPELLIER	841	943	1206	609	1230	1503	3957	663	230	11182
ST MARTIN	210	77	54	9	24	43	208	18	10	653
75										
AGDE	335	555	420	80	185	260	995	125	15	2970
BEDARIEUX	125	15	105	15	65	60	390	10	25	810
BEZIERS	1300	1045	1015	305	770	1250	2565	375	120	8745
CAPESTANG	755	660	440	125	275	495	1050	150	40	3990
FLORENSAC	425	240	215	50	160	150	695	85	0	2020
MONTAGNAC	780	480	200	50	160	230	625	85	40	2650
MURVIEL	650	470	210	105	215	305	705	60	15	2735
OLARGUES	420	75	140	20	80	55	200	15	35	1040
OLONZAC	740	220	180	35	130	175	275	35	40	1830
PEZENAS	300	285	155	30	110	115	430	30	40	1495
ROUJAN	480	275	150	45	100	180	380	50	25	1685
ST CHINIAN	705	180	235	70	200	175	470	30	35	2100
ST GERVAIS	250	40	305	110	320	210	815	265	35	2350
ST PONS	205	135	165	65	195	120	580	80	35	1580
SALVETAT	160	55	60	25	30	30	85	30	10	485
SERVIAN	605	590	205	85	140	195	585	95	45	2545
LE CAYLAR	70	25	40	5	25	15	40	5	15	240
CLERMONT	480	240	205	110	215	210	635	60	20	2175
GIGNAC	1310	535	390	85	245	295	1030	125	90	4105
LODEVE	340	60	80	25	70	45	210	0	10	840
LUNAS	185	20	155	30	90	90	545	10	45	1170
ANIANE	230	25	130	35	160	165	380	50	35	1210
CASTRIES	430	300	390	355	745	735	1535	190	90	4770
CLARET	160	65	40	35	30	45	120	15	5	515
FRONTIGNAN	95	130	355	105	350	410	1100	190	50	2785
GANGES	100	60	335	65	215	315	1360	60	60	2570
LUNEL	435	615	215	50	235	265	1115	110	30	3070
MATELLES	215	205	270	405	410	350	545	95	115	2610
MAUGUIO	380	470	435	360	570	590	1105	230	85	4225
MEZE	385	345	1020	120	365	375	1580	160	105	4455
MONTPELLIER	615	680	1905	1830	2810	3485	5210	1160	295	17990
ST MARTIN	130	35	75	5	40	75	225	40	25	650

Effectifs 82	EAGR	OAGR	INDU	LIBE	INTE	EMPL	OUVR	SERV	
AGDE	416	260	436	116	416	608	1036	124	3412
BEDARIEUX	104	8	132	28	132	176	384	8	972
BEZIERS	1148	462	1300	528	1744	2520	3066	324	11092
CAPESTANG	656	376	532	152	600	860	1200	100	4476
FLORENSAC	308	116	236	80	300	392	604	40	2076
MONTAGNAC	628	244	244	60	332	368	640	72	2588
MURVIEL	584	328	340	72	320	520	800	120	3084
OLARGUES	532	76	140	56	188	212	196	36	1436
OLONZAC	616	172	232	64	196	244	316	48	1888
PEZENAS	280	128	152	68	256	320	448	32	1684
ROUJAN	476	196	176	68	196	356	300	48	1816
ST CHINIAN	692	168	304	64	208	348	400	76	2260
ST GERVAIS	192	48	380	92	368	740	620	104	2544
ST PONS	164	120	164	64	236	316	488	32	1584
SALVETAT	176	44	84	32	32	92	140	4	604
SERVIAN	560	348	312	92	304	564	648	108	2936
LE CAYLAR	88	32	40	4	52	48	72	12	348
CLERMONT	488	108	320	76	304	420	780	64	2560
GIGNAC	944	352	524	172	592	796	1432	184	4996
LODEVE	320	28	112	24	160	188	252	52	1136
LUNAS	240	32	148	36	144	176	424	28	1228
ANIANE	220	24	176	132	268	356	408	40	1624
CASTRIES	376	200	800	800	1776	2104	1780	152	7988
CLARET	168	48	68	68	168	124	172	4	820
FRONTIGNAN	88	48	484	228	660	1112	1132	172	3924
GANGES	112	72	500	96	332	556	1132	60	2860
LUNEL	336	328	376	184	620	792	1360	120	4116
MATELLES	228	108	464	912	1328	1172	804	120	5136
MAUGUIO	492	324	712	764	1420	2004	1680	328	7724
MEZE	832	300	684	224	636	968	1480	120	5244
MONTPELLIER	1192	374	3140	3264	5848	7700	6594	672	28784
ST MARTIN	152	24	120	84	148	176	312	16	1032

Profils en Pourcentage 54	EAGR	OAGR	INDU	LIBE	CADR	EMPL	OUVR	SERV	CLER
AGDE	22.53	35.63	13.96	0.86	2.92	4.69	16.82	2.03	0.56
BEDARIEUX	17.54	6.91	6.24	0.56	3.82	4.72	57.56	1.74	0.90
BEZIERS	28.81	33.44	11.93	0.66	3.10	4.30	14.62	2.50	0.64
CAPESTANG	26.66	41.46	11.15	0.73	3.18	3.91	10.29	1.86	0.76
FLORENSAC	33.08	38.67	10.22	0.77	2.65	2.56	9.69	1.40	0.96
MONTAGNAC	35.97	37.27	9.47	0.68	2.96	2.96	7.88	1.69	1.11
MURVIEL	42.29	33.23	9.42	0.79	2.53	3.09	6.56	1.35	0.74
OLARGUES	62.85	13.66	8.08	0.39	2.79	2.74	7.64	0.73	1.13
OLONZAC	47.77	23.43	11.05	0.53	2.70	2.88	8.24	2.42	0.98
PEZENAS	28.30	39.80	11.26	0.65	3.42	3.42	10.79	1.47	0.88
ROUJAN	39.29	30.27	11.55	0.84	3.29	4.17	8.31	1.48	0.80
ST CHINIAN	42.18	25.69	11.98	0.72	3.02	3.14	10.67	1.59	1.03
ST GERVAIS	30.09	10.09	14.82	1.80	5.47	5.80	27.07	3.67	1.18
ST PONS	26.82	8.28	14.05	2.22	5.44	4.30	34.82	2.27	1.80
SALVETAT	63.03	8.65	10.40	0.41	2.37	2.88	8.86	1.44	1.96
SERVIAN	31.04	42.62	10.97	0.75	2.45	2.84	6.79	1.60	0.95
LE CAYLAR	45.49	19.41	9.41	0.39	3.53	2.16	15.29	1.76	2.55
CLERMONT	35.12	26.28	10.23	1.24	3.72	4.22	16.40	1.86	0.93
GIGNAC	39.96	29.81	10.63	0.68	2.91	2.29	9.78	2.17	1.77
LODEVE	59.42	20.26	4.98	0.14	2.63	1.63	9.10	1.07	0.78
LUNAS	29.61	9.15	9.56	0.92	3.88	4.43	39.35	1.71	1.39
ANIANE	42.42	21.21	9.81	1.44	4.63	4.63	10.61	3.19	2.07
CASTRIES	31.46	34.99	9.19	0.77	3.23	4.10	12.31	2.59	1.35
CLARET	47.37	36.22	5.42	0.31	2.79	1.86	3.41	1.24	1.39
FRONTIGNAN	13.59	25.02	10.01	0.74	4.39	8.15	33.79	3.15	1.17
GANGES	9.74	8.91	15.20	1.78	3.65	5.17	51.35	3.30	0.89
LUNEL	29.82	41.15	8.67	0.86	2.21	4.13	9.94	2.25	0.98
MATELLES	37.07	34.92	7.37	0.54	2.76	3.84	9.06	2.30	2.15
MAUGUIO	26.73	41.87	9.23	0.87	2.94	3.60	10.43	3.19	1.16
MEZE	24.61	22.93	19.43	0.90	3.02	4.36	21.35	2.26	1.14
MONTPELLIER	20.28	26.02	11.31	1.56	4.40	6.87	21.99	5.40	2.17
ST MARTIN	40.80	25.94	7.43	1.10	2.68	2.07	16.20	1.71	2.07

Profils en Pourcentage	EAGR	OAGR	INDU	LIBE	CADR	EMPL	OUVR	SERV	CLER
62									
AGDE	19.70	30.10	14.33	1.52	4.19	5.09	20.86	3.25	0.97
BEDARIEUX	24.21	8.41	7.68	0.95	4.83	4.83	45.94	1.02	2.12
BEZIERS	26.43	27.31	12.01	1.02	3.56	6.13	19.33	2.89	1.33
CAPESTANG	24.41	35.27	10.64	1.16	3.86	5.33	15.59	1.98	1.76
FLORENSAC	30.58	32.23	10.11	1.03	4.23	4.02	14.75	1.65	1.39
MONTAGNAC	31.64	34.11	8.60	0.93	4.28	4.17	13.31	1.07	1.89
MURVIEL	34.23	29.21	8.50	0.75	3.77	5.71	14.34	2.13	1.35
OLARGUES	53.32	12.30	9.50	0.87	4.49	3.61	13.17	1.17	1.57
OLONZAC	44.15	24.41	10.27	0.85	3.90	3.44	10.31	1.35	1.31
PEZENAS	26.67	33.20	9.60	1.24	4.51	4.09	16.36	1.54	2.79
ROUJAN	37.32	27.81	11.42	1.14	4.64	4.96	9.01	2.00	1.68
ST CHINIAN	40.86	22.45	12.20	0.74	3.98	4.37	11.97	1.78	1.65
ST GERVAIS	24.67	6.71	14.44	2.33	7.65	5.40	28.48	7.98	2.33
ST PONS	20.32	9.23	13.82	2.42	7.79	5.47	36.77	2.06	2.11
SALVETAT	41.56	15.74	11.21	0.91	2.94	3.62	20.72	1.59	1.70
SERVIAN	28.12	38.12	9.62	1.05	3.47	4.11	12.78	1.54	1.19
LE CAYLAR	39.02	22.76	11.11	0.54	2.98	2.71	14.91	1.63	4.34
CLERMONT	36.58	21.75	9.89	1.62	5.38	5.77	16.17	1.70	1.15
GIGNAC	36.41	27.61	9.67	0.83	4.05	3.20	14.55	1.99	1.70
LODEVE	50.76	21.18	4.92	0.36	4.47	3.84	11.62	1.52	1.34
LUNAS	25.87	9.23	9.65	1.09	5.31	5.67	38.66	2.47	2.05
ANIANE	37.52	17.84	11.25	1.41	5.98	5.98	14.06	2.64	3.34
CASTRIES	25.18	26.89	7.99	0.94	4.61	7.89	21.04	3.38	2.07
CLARET	45.49	31.42	6.77	0.69	4.51	2.08	7.47	0.35	1.22
FRONTIGNAN	9.77	17.83	9.60	1.66	4.86	8.41	42.24	3.79	1.84
GANGES	6.33	5.16	14.10	1.96	4.13	6.13	57.68	2.29	2.23
LUNEL	24.75	29.98	7.87	0.89	3.69	5.19	23.73	2.60	1.30
MATELLES	28.25	28.17	6.75	1.50	3.92	6.33	16.08	3.42	5.58
MAUGUIO	21.73	34.86	8.56	1.05	2.89	5.87	20.93	2.85	1.26
MEZE	22.63	17.80	21.64	1.11	3.33	4.98	24.70	1.97	1.85
MONTPELLIER	14.56	15.93	11.44	2.50	6.56	10.13	30.14	5.23	3.51
ST MARTIN	33.78	21.09	7.52	0.88	3.54	3.54	26.84	0.74	2.06
68									
AGDE	16.18	21.74	14.91	1.38	5.31	6.38	29.92	3.55	0.62
BEDARIEUX	21.34	5.53	7.08	1.07	6.01	7.57	47.72	2.13	1.55
BEZIERS	19.40	19.36	12.41	1.52	5.72	8.12	28.67	3.70	1.09
CAPESTANG	20.30	26.88	10.36	1.76	6.28	7.75	23.37	2.27	1.03
FLORENSAC	27.92	25.65	11.00	1.05	5.86	4.53	21.72	0.94	1.33
MONTAGNAC	30.02	25.23	9.33	1.51	4.72	6.20	19.17	1.95	1.87
MURVIEL	30.55	22.91	8.35	1.04	5.05	7.51	20.78	2.65	1.17
OLARGUES	53.45	9.54	9.61	1.00	4.63	4.41	14.59	1.21	1.57
OLONZAC	44.19	18.00	9.71	1.25	4.01	5.66	13.59	2.27	1.34
PEZENAS	22.56	26.69	9.87	1.68	4.91	6.34	22.68	2.75	2.51
ROUJAN	34.51	22.63	10.88	1.10	5.16	5.25	16.99	2.72	0.76
ST CHINIAN	36.11	18.26	12.55	1.77	5.08	5.19	16.64	2.76	1.66
ST GERVAIS	20.51	4.26	12.38	3.02	9.87	7.40	30.13	9.91	2.51
ST PONS	14.21	9.74	13.24	2.35	8.60	7.97	39.20	2.98	1.72
SALVETAT	38.86	10.71	13.57	1.00	4.86	5.57	21.14	1.86	2.43
SERVIAN	25.42	30.56	9.76	1.49	4.15	4.86	19.45	3.16	1.14
LE CAYLAR	38.56	14.42	8.78	0.94	3.76	3.76	25.08	1.25	3.45
CLERMONT	30.75	15.39	11.06	1.99	6.32	8.03	22.79	2.59	1.07
GIGNAC	31.22	20.66	10.48	1.33	4.83	5.21	22.07	2.55	1.64
LODEVE	48.05	16.33	6.01	1.48	4.64	5.69	14.96	1.69	1.16
LUNAS	23.99	8.98	9.48	1.87	5.68	7.18	38.22	2.44	2.16
ANIANE	31.35	9.89	10.16	2.19	8.41	6.22	24.96	4.64	2.19
CASTRIES	18.22	16.13	7.90	2.24	7.28	10.29	33.22	3.28	1.43
CLARET	45.67	25.60	5.16	1.84	4.42	3.50	10.13	1.66	2.03
FRONTIGNAN	8.69	10.33	10.54	1.64	5.77	10.49	47.40	3.81	1.32
GANGES	4.39	3.37	10.53	2.15	5.25	8.83	60.40	3.16	1.91
LUNEL	17.89	19.98	7.03	1.36	4.73	8.56	36.16	2.99	1.29
MATELLES	19.97	17.27	7.15	4.52	8.64	8.70	25.30	4.05	4.39
MAUGUIO	16.74	22.86	8.68	1.80	4.21	8.03	33.23	3.49	0.97
MEZE	15.81	13.15	22.09	1.51	5.29	6.39	31.39	3.17	1.20
MONTPELLIER	7.52	8.43	10.79	5.45	11.00	13.44	35.39	5.93	2.06
ST MARTIN	32.16	11.79	8.27	1.38	3.68	6.58	31.85	2.76	1.53

Profils en Pourcentage 75	EAGR	OAGR	INDU	LIBE	CADR	EMPL	OUVR	SERV	CLER
AGDE	11.28	18.69	14.14	2.69	6.23	8.75	33.50	4.21	0.51
BEDARIEUX	15.43	1.85	12.96	1.85	8.02	7.41	48.15	1.23	3.09
BEZIERS	14.87	11.95	11.61	3.49	8.81	14.29	29.33	4.29	1.37
CAPESTANG	18.92	16.54	11.03	3.13	6.89	12.41	26.32	3.76	1.00
FLORENSAC	21.04	11.88	10.64	2.48	7.92	7.43	34.41	4.21	0.00
MONTAGNAC	29.43	18.11	7.55	1.89	6.04	8.68	23.58	3.21	1.51
MURVIEL	23.77	17.18	7.68	3.84	7.86	11.15	25.78	2.19	0.55
OLARGUES	40.38	7.21	13.46	1.92	7.69	5.29	19.23	1.44	3.37
OLONZAC	40.44	12.02	9.84	1.91	7.10	9.56	15.03	1.91	2.19
PEZENAS	20.07	19.06	10.37	2.01	7.36	7.69	28.76	2.01	2.68
ROUJAN	28.49	16.32	8.90	2.67	5.93	10.68	22.55	2.97	1.48
ST CHINIAN	33.57	8.57	11.19	3.33	9.52	8.33	22.38	1.43	1.67
ST GERVAIS	10.64	1.70	12.98	4.68	13.62	8.94	34.68	11.28	1.49
ST PONS	12.97	8.54	10.44	4.11	12.34	7.59	36.71	5.06	2.22
SALVETAT	32.99	11.34	12.37	5.15	6.19	6.19	17.53	6.19	2.06
SERVIAN	23.77	23.18	8.06	3.34	5.50	7.66	22.99	3.73	1.77
LE CAYLAR	29.17	10.42	16.67	2.08	10.42	6.25	16.67	2.08	6.25
CLERMONT	22.07	11.03	9.43	5.06	9.89	9.66	29.20	2.76	0.92
GIGNAC	31.91	13.03	9.50	2.07	5.97	7.19	25.09	3.05	2.19
LODEVE	40.48	7.14	9.52	2.98	8.33	5.36	25.00	0.00	1.19
LUNAS	15.81	1.71	13.25	2.56	7.69	7.69	46.58	0.85	3.85
ANIANE	19.01	2.07	10.74	2.89	13.22	13.64	31.40	4.13	2.89
CASTRIES	9.01	6.29	8.18	7.44	15.62	15.41	32.18	3.98	1.89
CLARET	31.07	12.62	7.77	6.80	5.83	8.74	23.30	2.91	0.97
FRONTIGNAN	3.41	4.67	12.75	3.77	12.57	14.72	39.50	6.82	1.80
GANGES	3.89	2.33	13.04	2.53	8.37	12.26	52.92	2.33	2.33
LUNEL	14.17	20.03	7.00	1.63	7.65	8.63	36.32	3.58	0.98
MATELLES	8.24	7.85	10.34	15.52	15.71	13.41	20.88	3.64	4.41
MAUGUIO	8.99	11.12	10.30	8.52	13.49	13.96	26.15	5.44	2.01
MEZE	8.64	7.74	22.90	2.69	8.19	8.42	35.47	3.59	2.36
MONTPELLIER	3.42	3.78	10.59	10.17	15.62	19.37	28.96	6.45	1.64
ST MARTIN	20.00	5.38	11.54	0.77	6.15	11.54	34.62	6.15	3.85

Profils en Pourcentage 82	EAGR	OAGR	INDU	LIBE	INTE	EMPL	OUVR	SERV
AGDE	12.19	7.62	12.78	3.40	12.19	17.82	30.36	3.63
BEDARIEUX	10.70	0.82	13.58	2.88	13.58	18.11	39.51	0.82
BEZIERS	10.35	4.17	11.72	4.76	15.72	22.72	27.64	2.92
CAPESTANG	14.66	8.40	11.89	3.40	13.40	19.21	26.81	2.23
FLORENSAC	14.84	5.59	11.37	3.85	14.45	18.88	29.09	1.93
MONTAGNAC	24.27	9.43	9.43	2.32	12.83	14.22	24.73	2.78
MURVIEL	18.94	10.64	11.02	2.33	10.38	16.86	25.94	3.89
OLARGUES	37.05	5.29	9.75	3.90	13.09	14.76	13.65	2.51
OLONZAC	32.63	9.11	12.29	3.39	10.38	12.92	16.74	2.54
PEZENAS	16.63	7.60	9.03	4.04	15.20	19.00	26.60	1.90
ROUJAN	26.21	10.79	9.69	3.74	10.79	19.60	16.52	2.64
ST CHINIAN	30.62	7.43	13.45	2.83	9.20	15.40	17.70	3.36
ST GERVAIS	7.55	1.89	14.94	3.62	14.47	29.09	24.37	4.09
ST PONS	10.35	7.58	10.35	4.04	14.90	19.95	30.81	2.02
SALVETAT	29.14	7.28	13.91	5.30	5.30	15.23	23.18	0.66
SERVIAN	19.07	11.85	10.63	3.13	10.35	19.21	22.07	3.68
LE CAYLAR	25.29	9.20	11.49	1.15	14.94	13.79	20.69	3.45
CLERMONT	19.06	4.22	12.50	2.97	11.87	16.41	30.47	2.50
GIGNAC	18.90	7.05	10.49	3.44	11.85	15.93	28.66	3.68
LODEVE	28.17	2.46	9.86	2.11	14.08	16.55	22.18	4.58
LUNAS	19.54	2.61	12.05	2.93	11.73	14.33	34.53	2.28
ANIANE	13.55	1.48	10.84	8.13	16.50	21.92	25.12	2.46
CASTRIES	4.71	2.50	10.02	10.02	22.23	26.34	22.28	1.90
CLARET	20.49	5.85	8.29	8.29	20.49	15.12	20.98	0.49
FRONTIGNAN	2.24	1.22	12.33	5.81	16.82	28.34	28.85	4.38
GANGES	3.92	2.52	17.48	3.36	11.61	19.44	39.58	2.10
LUNEL	8.16	7.97	9.14	4.47	15.06	19.24	33.04	2.92
MATELLES	4.44	2.10	9.03	17.76	25.86	22.82	15.65	2.34
MAUGUIO	6.37	4.19	9.22	9.89	18.38	25.95	21.75	4.25
MEZE	15.87	5.72	13.04	4.27	12.13	18.46	28.22	2.29
MONTPELLIER	4.14	1.30	10.91	11.34	20.32	26.75	22.91	2.33
ST MARTIN	14.73	2.33	11.63	8.14	14.34	17.05	30.23	1.55

Références bibliographiques

Algèbre linéaire. Analyse numérique.

CHARLES B., ALLOUCH D. - Algèbre générale, P.U.F., 1984.

CIARLET P.G. - Introduction à l'analyse numérique et à l'optimisation, Masson, 1982.

FRENKEL J. - Géométrie pour l'élève professeur, Hermann, 1973.

GANTMACHER F.R. - Théorie des matrices, tomes 1 et 2, Dunod, 1966.

GODEMENT R. - Cours d'algèbre, Hermann, 1973.

GOLUB G.H., VAN LOAN C.F. - Matrix computations, North Oxford Academic, 1983.

PARLETT B.N. - The symmetric eigenvalue problem, Prentice Hall, 1980.

RAO C.R. - Linear statistical inference and its applications, Wiley, 1973.

RAO C.R. - Matrix approximations and reduction of dimensionality in multivariate statistical analysis, Multivariate analysis, KRISHNAIAH ed., North Holland, p. 3-22, 1980.

SABATIER R. - Méthodes factorielles en analyse des données : approximations et prise en compte de variables concomitantes, Thèse de doctorat d'état, Montpellier II, 1987.

SABATIER R., JAN Y., ESCOUFIER Y. - Approximations d'applications linéaires et analyse en composantes principales, Data analysis and informatics, DIDAY ed., North Holland, p. 569-580, 1984.

Analyse d'un tableau de données.

ARAGON Y., CAUSSINUS H. - Une analyse en composantes principales pour des unités statistiques corrélées, Data analysis and informatics, DIDAY ed., North-Holland, p. 121-131, 1980.

BENZECRI J.P. - Sur l'analyse factorielle des proximités. Publications de l'ISUP, n° XIII, p. 235-281, 1964.

BENZECRI J.P. - L'analyse des données. Tome 2 : l'analyse des correspondances, Dunod, 1973.

BESSE Ph. - Optimal metric in principal components analysis of longitudinal data, 5ièmes journées analyse des données et informatique, INRIA, tome 2, p. 117-128, 1987.

BESSE Ph., RAMSAY J.O. - Principal components analysis of sampled functions, Psychometrika, n° 51, p. 285-311, 1986.

BOUROCHE J.M., SAPORTA G. - L'analyse des données, P.U.F., coll. Que sais-je ?, 1987.

CAILLIEZ F. - Analyse des données, P.U.M., Eyrolles, 1980.

CAILLIEZ F., PAGES J.P. - Introduction à l'analyse des données, SMASH, Paris, 1976.

CRITCHLEY F. - Optimal norm characterizations of multidimensional scaling methods and some related data analysis problems, Data analysis and informatics, DIDAY ed., North-Holland, p. 209-229, 1980.

CRITCHLEY F. - Dimensionality theorems in multidimensional scaling and hierarchical cluster analysis, Data analysis and informatics, DIDAY ed., North-Holland, p. 45-70, 1986.

DIDAY E., LEMAIRE J., POUGET J., TESTU F. - Eléments d'analyse de données, Dunod, 1982.

ESCOUFIER Y. - The duality diagram. A means for better practical applications. Developments in numerical ecology, LEGENDRE ed., Springer-Verlag, p. 139-156, 1987.

FICHET B. - Analyse factorielle sur tableaux de dissimilarité, Thèse de doctorat d'état, Aix-Marseille II, 1983.

FICHET B. - Distances and euclidean distances for presence-absence characters and their application to factor analysis, Multidimensional data analysis, de LEEUW and al. ed., DSWO Press, p. 23-46, 1986.

FICHET B., LE CALVE G. - Structure géométrique des principaux indices de dissimilarité sur signes de présence-absence, Statistique et Analyse de Données, n° 9 (3), p. 11-44, 1984.

GOWER J.C. - Properties of euclidean and non-euclidean distance matrices, Linear Algebra and its Applications, n° 67, p. 81-97, 1985.

KRUSKAL J.B., WISH M. - Multidimensional scaling, Sage publications, 1978.

LEBART L. - Analyse statistique de la contiguïté, Thèse de 3e cycle, Paris, 1966.

LEBART L., MORINEAU A., FENELON J.P. - Traitement des données statistiques, Dunod, 1979.

LE CALVE G. - Quelques remarques sur certains aspects de l'analyse factorielle. Université Rennes II, cahier n° 2, 1976.

LE FOLL Y. - Pondération des distances en analyse factorielle, Statistique et Analyse de Données, n° 7 (1), p. 13-31, 1982.

MOM A. - Quantification et analyse des contiguïtés : application à des problèmes de trafic, Thèse de doctorat, Montpellier II, 1988.

PAGES J.P., CAILLIEZ F., ESCOUFIER Y. - Analyse factorielle : un peu d'histoire et de géométrie, Revue de Statistique Appliquée, XXVII (1), p. 5-28, 1979.

SAPORTA G. - Théories et méthodes de la statistique, Technip, 1980.

SCHEKTMAN Y. - Métriques et analyse de données. Contributions et synthèse, 2e colloque Franco-Soviétique, 1984.

TENENHAUS M., YOUNG F.W. - An analysis and synthesis of multiple correspondence analysis, optimal scaling, dual scaling, homogeneity analysis and other methods for quantifying categorical multivariate data, Psychometrika, n° 50, p. 91-119, 1985.

TORGERSON W.S. - Theory and methods of scaling, Wiley, 1958.

Analyse simultanée de tableaux de données.

BENZECRI J.P. - Analyse de l'inertie intra-classe par l'analyse d'un tableau de correspondance, Les Cahiers de l'Analyse des Données, n° VIII, p. 351-358, 1983.

BOLASCO S. - Per una teoria sulla costruzione e l'analisi delle matrici a tre modi, Proceedings of the 33rd meeting of the Italian Statistical Society, n° 1, p. 183-195, 1986.

BOUROCHE J.M. - Analyse des données ternaires : la double analyse en composantes principales, Thèse de 3e cycle, Paris VI, 1975.

CARLIER A. - Analyse des évolutions sur tables de contingence : quelques aspects opérationnels, 4ièmes journées analyse des données et informatique, INRIA, tome 2, p. 421-428, 1985.

CARLIER A. - Factor analysis of evolutions and cluster methods on trajectories, Physica-Verlag, p. 139-145, Compstat 1986.

CARROLL J.D., CHANG J.J. - Analysis of individual differences in multidimensional scaling via an N-way generalization of "Eckart-Young" decomposition, Psychometrika, n° 35, p. 283-319, 1970.

CLARET A. - Contribution au problème de l'approximation factorielle d'un tableau de données, Thèse de doctorat, Montpellier II, 1987.

COPPI R. - Analysis of three-way data matrices based on pairwise relation measures, Physica-Verlag, p. 129-139, Compstat 1986.

D'ALESSIO G. - L'analisi di successioni di matrici di dati qualitativi, Thèse de doctorat, Rome, 1986.

D'AMBRA L., MARCHETTI G.M. - Un metodo per l'analisi interstrutturale di più matrici basato su misure di relazione tra le unità statistiche, Proceedings of the 33rd meeting of the Italian Statistical Society, n° 1, p. 171-182, 1986.

ESCOFIER B. - Analyse factorielle en référence à un modèle. Application à l'analyse de tableaux d'échanges, Revue de Statistique Appliquée, XXXII (4), p. 25-36, 1984.

ESCOFIER B. - Analyse des correspondances multiples conditionnelle, 5ièmes journées analyse des données et informatique, INRIA, tome 2, p. 13-22, 1987.

ESCOFIER B., PAGES J. - L'analyse factorielle multiple, Cahier du B.U.R.O., n° 2, 1984.

ESCOUFIER Y. - Echantillonnage dans une population de variables aléatoires réelles, Publications de l'ISUP, n° XIX (4), p. 1-47, 1970.

ESCOUFIER Y. - Le traitement des variables vectorielles, Biometrics, n° 29, p. 751-760, 1973.

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

ESCOUFIER Y. - L'analyse conjointe de plusieurs matrices de données, Biométrie et temps, JOLIVET et al. ed., p. 59-76, 1980.

ESCOUFIER Y. - A propos du choix des variables en analyse des données, Metron, n° XLIV, p. 31-47, 1986.

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

FLURY B. - A generalization of principal component analysis of k groups, Technical report 83-14, Dept. of Stat., Purdue University, 1983.

FOUCART T. - Suites de tableaux et de sous-tableaux, Revue de Statistique Appliquée, XXIX (2), p. 31-42, 1981.

FOUCART T. - Analyse factorielle des tableaux multiples, Masson, 1984.

GLAÇON F. - Analyse conjointe de plusieurs matrices de données. Comparaison de différentes méthodes, Thèse de 3e cycle, Grenoble, 1981.

GOWER J.C. - Generalized procuste analysis, Psychometrika, n° 40, p. 33-51, 1975.

HARSHMAN R.A., LUNDY M.E. - The PARAFAC model for three-way factor analysis and multidimensional scaling, Research methods for multimode data analysis, LAW and al. ed., New-York, Praeger, p. 122-215, 1984.

Van der HEIJDEN P.G.M. - Correspondence analysis of longitudinal categorical data, DSWO Press, 1987.

Van der HEIJDEN P.G.M., de LEEUW J. - Correspondence analysis used complementary to loglinear analysis, Psychometrika, n° 50, p. 429-447, 1985.

JAFFRENOU P.A. - Sur l'analyse des familles finies de variables vectorielles, Thèse de 3e cycle, Lyon I, 1978.

KETTENRING J.R. - Canonical analysis of several sets of variables, Biometrika, n° 58, p. 433-451, 1971.

KROONENBERG P.M. - Three-mode principal component analysis, DSWO Press, 1983.

KRZANOWSKI W.J. - Principal component analysis in the presence of group structure, Applied Statistics, n° 33, p. 164-168, 1984.

LASTOVICKA J.L. - The extension of component analysis to four-mode matrices, Psychometrika, n° 46, p. 47-57, 1981.

LE FOLL Y. - L'analyse factorielle des évolutions, Annales de l'Université des sciences sociales de Grenoble, n° 2, p. 23-27, 1972.

LERMAN I.C., TALLUR B. - Classification des éléments consécutifs d'une juxtaposition de tableaux de contingence, Revue de Statistique Appliquée, XXVIII (3), p. 5-28, 1980.

LEVIN J. - Simultaneous factor analysis of several Gramian matrices, Psychometrika, n° 31, p. 413-419, 1966.

L'HERMIER DES PLANTES H. - Structuration des tableaux à trois indices de la statistique, Thèse de 3e cycle, Montpellier II, 1976.

MARCOTORCHINO F. - Maximal association as a tool for classification, Classification as a tool of research, GAUL and SCHADER ed., North-Holland, p. 275-288, 1986.

PAGES J.P., ESCOUFIER Y., CAZES P. - Opérateurs et analyse des tableaux à plus de deux dimensions, Cahier du B.U.R.O., n° 25, p. 61-89, 1976.

PERNIN M.O. - Contribution à la méthodologie d'analyse de données longitudinales. Exemple de croissance chez l'être humain, Thèse de doctorat, Lyon I, 1986.

PICARD J. - Classification des profils évolutifs incomplets et asynchrones, Revue de Statistique Appliquée, XXXV (2), p. 27-38, 1987.

POLIT E. - Une n-ACP d'un hypercube de données, Thèse de 3e cycle, Grenoble, 1986.

PONTIER J., PERNIN M.O. - Méthode LONGI, Data analysis : Ins and outs of solving real problems, JANSSEN and al. ed., Plenum, p. 49-66, 1987.

RALAMBONDRAINY H. - Problèmes d'optimisation relatifs aux tableaux multiples. Point de vue de la classification automatique, Rapport de recherche INRIA, n° 576, 1986.

RAMSAY J.O., ten BERGE J., STYAN G.P.H. - Matrix correlation, Psychometrika, n° 49, p. 403-423, 1984.

Research methods for multimode data analysis, LAW, SNYDER, HATTIE and Mc DONALD ed., New-York, Praeger, 1984.

SCHEKTMAN Y. - A general euclidean method for measuring and describing associations between several sets of variables, Recent developments in clustering and data analysis, Institute of statistical mathematics, Tokyo, p. 31-42, 1987.

SEMPÉ M. ; KROONENBERG P.M. ; LAVIT Ch. and PERNIN M.O. ; LEWI P.J., CALOMME G. and VAN HOOF J. ; PONTIER J. ; MINEO A. - Multivariate and longitudinal data on growing children, Data analysis : the ins and outs of solving real problems, JANSSEN, MARCOTORCHINO and PROTH ed., Plenum, p. 7-112, 1987.

TENENHAUS M. - L'analyse canonique généralisée de variables numériques, nominales ou ordinales par des méthodes de codage optimal, Data analysis and informatics, DIDAY ed., North-Holland, p. 71-84, 1984.

TUCKER L.R. - Some mathematical notes on three-mode factor analysis, Psychometrika, n° 31, p. 279-311, 1966.

Exemples d'application de la méthode STATIS.

AMANIEU M., GUELORGET O., NOUGUIER J. - Analyse de la diversité de la macrofaune benthique d'une lagune littorale méditerranéenne, Vie Milieu, n° 31, p. 303-312, 1981.

BEGEL M. - Les facteurs de réussite et d'échec en première année d'études. Le numerus clausus en pharmacie et ses conséquences, Thèse de doctorat d'état, Montpellier I, 1987.

BERNARD M.C., LAVIT Ch. - L'évolution des sociétés rurales du Languedoc-Roussillon (1954-1975) : application de la méthode STATIS, 110e congrès national des sociétés savantes, géographie, p. 183-196, 1985.

BOLASCO S. - L'evoluzione della struttura dei mercati del lavoro regionali del periodo 1977-1982 in italia, Publication du département de statistique, Série A, Université La Sapienza, Rome, n° 15, 1984.

BOUCHER DE LA RUPELLE M.L. - Techniques graphiques en analyse de données multidimensionnelles, Thèse de 3e cycle, Montpellier II, 1980.

ESCOUFIER Y., L'HERMIER H. - A propos de la comparaison graphique des matrices de variance, Biometrischc Zeitschrift, n° 20, p. 477-483, 1978.

GOULARD M. - Champs spatiaux et statistique multidimensionnelle, Thèse de doctorat, Montpellier II, 1988.

JUNCA S. - Outils informatiques pour l'évaluation de la pertinence d'un résultat en analyse des données, Thèse de 3e cycle, Montpellier II, 1985.

L'HERMIER DES PLANTES H., THIEBAUT B. - Etude de la pluviosité au moyen de la méthode STATIS, Revue de Statistique Appliquée, XXV (2), p. 57-58, 1977.

NORMAND B. - Analyse de données appliquée au suivi de sujets hémodialysés : méthode STATIS, Mémoire de DERBH, Paris-Sud, 1987.

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

PEREZ HUGALDE C., GARCIA VELAZQUEZ A., JUDEZ L. - Analisis de los resultados economicos de los sectores agrario y no agrario en el proceso de desarrollo, Revista de Estudios Agro-Sociales, n° 137 extra, p. 317-351, 1986.

TEYSSIER J. - Evolution de la distribution du diamètre des adipocytes chez le lapin en croissance, Thèse de doctorat, INRA Montpellier, 1988.

Annexe

Programmes en Turbo Pascal

1. DONNEES RECUEILLIES SUR LES MEMES INDIVIDUS _ PRELIMINAIRES

```
Program Lecture_dimensions;
(************************)

  Const MaxIND  = 50; MaxVAR  = 25; MaxTAB  = 20;
  Var   NT,NI,NV,k :integer;
        Rep        :char;
        Fdim       :text;

  Begin
      assign (Fdim,'A:dim'); rewrite (Fdim);
    write ('nombre de tableaux = '); readln (NT);
      if NT>MaxTAB then
        begin
          writeln ('trop grand nombre de tableaux : modifier');
          writeln ('la constante MaxTAB dans les programmes');
          exit;
        end;
    write ('nombre d''individus par tableau = '); readln (NI);
      if NI>MaxIND then
        begin
          writeln ('trop grand nombre d''individus : modifier');
          writeln ('la constante MaxIND dans les programmes');
          exit;
        end;
    write (Fdim,NT,' ',NI,' ');

    write ('les variables sont-elles les memes ');
    write ('pour tous les tableaux? (O/N) ');
            readln (Rep);
    if Rep='O' then
        begin
          write ('nombre de variables = '); readln (NV);
          if NV>MaxVAR then
            begin
              writeln ('trop grand nombre de variables : modifier');
              writeln ('la constante MaxVAR dans les programmes');
              exit;
            end;
```

```
          for k:=1 to NT do write (Fdim, NV,' ');
        end
      else
        begin
          for k:=1 to NT do
            begin
              write ('nombre de variables du tableau n°',k,' = ');
                     readln (NV);
              if NV>MaxVAR then
                begin
                  writeln ('le nombre de variables est trop grand');
                  writeln ('modifier la constante MaxVAR ');
                  exit;
                end;
              write  (Fdim,NV,' ');
            end;
        end;

    close (Fdim);
    writeln ('               fin de l''execution');
    writeln ('les dimensions des tableaux sont dans A:dim');
  End.



Program Lecture_poids_individus;
(*****************************)

  Const MaxIND  = 50;
  Type  colonne = array [1..MaxIND] of real;
  Var   NT,NI,NIS,NIA,i :integer;
        POID            :colonne;
        Fdim,Fpoids     :text;

Procedure Poids (N:integer; Var POID:colonne);
(*************)
     (* calcul des poids des individus actifs *
      ****************************************)

  Var SOM :real;
      Rep :char;

  Begin
    write ('les individus actifs ont-ils le meme poids ? (O/N)  ');
             readln (Rep);
    if Rep='O' then
        for i:=1 to N do POID[i]:= 1
      else
        begin
          writeln ('rentrer les poids des ',N,' individus actifs :');
                   writeln ('(return apres chaque valeur)');
          for i:=1 to N do readln (POID[i]);
        end;
    SOM:= 0;
    for i:=1 to N do SOM:=SOM+POID[i];
    for i:=1 to N do POID[i]:= POID[i]/SOM;
  End;
```

```
  Begin
      assign (Fdim,'A:dim');     reset (Fdim);
      assign (Fpoids,'A:matD'); rewrite (Fpoids);

    read (Fdim,NT,NI);
    write ('nombre d''individus supplementaires parmi les ',NI,' = ');
          readln (NIS);
    if NIS > 0 then writeln
      ('les individus supplementaires sont supposes etre a la fin');
    writeln;
    NIA:= NI-NIS;

     (* calcul et stockage des poids des individus actifs *
      ****************************************************)
    Poids (NIA,POID);
    writeln (Fpoids,NIA,' ');
    for i:=1 to NIA do writeln (Fpoids,POID[i],' ');

    close (Fdim); close (Fpoids);
    writeln ('               fin de l''execution');
    write ('le nombre d''individus actifs et leurs poids sont dans ');
    writeln ('A:MatD');
  End.
```

```
Program Centrage_donnees;
(***********************)
  (*********************************************************************
   * centrage (et eventuellement reduction) des tableaux de donnees *
   * apres transformation , les tableaux de donnees sont dans A:Xk  *
   *********************************************************************)

  Const MaxIND  = 50; MaxVAR  = 25;
  Type  ligne   = array [1..MaxVAR] of real;
        colonne = array [1..MaxIND] of real;
        tableau = array [1..MaxIND,1..MaxVAR] of real;
  Var   NT,NI,NIA,NV,i,j,k        :integer;
        Rep                       :char;
        Fichier                   :string[20];
        POID                      :colonne;
        MOY,SIGMA                 :ligne;
        X                         :tableau;
        Fdon,Fdim,Fpoids,Fmoy,FXk :text;

Procedure Moyecart (NI,NIA,NV:integer; X:tableau; POID:colonne;
(****************)                    Var SIGMA,MOY:ligne);

  (* calcul des moyennes et ecart-types des variables du tableau X *
   ****************************************************************)

  Var A,C,D,SOM :real;
      B         :ligne;
```

```
Begin
  SOM:=POID[1];
  for j:=1 to NV do
    begin
      SIGMA[j]:=0;
      MOY[j]  :=X[1,j];
    end;
  for i:=2 to NIA do
    begin
      A:=SOM+POID[i];
      if abs(A)<1E-5 then A:=1E-5;
      for j:=1 to NV do
        begin
          B[j]    :=X[i,j]-MOY[j];
          C       :=B[j]*POID[i]/A;
          D       :=C*SOM;
          MOY[j]  :=MOY[j]+C;
          SIGMA[j]:=SIGMA[j]+D*B[j];
        end;
      SOM:=A;
    end;
    for j:=1 to NV do SIGMA[j]:=sqrt(SIGMA[j]/SOM);
End;


Begin
  write ('nom du fichier des donnees : ');
          readln (Fichier);
    assign (Fdon,Fichier);     reset (Fdon);
    assign (Fdim,'A:Dim');     reset (Fdim);
    assign (Fpoids,'A:MatD'); reset (Fpoids);
    assign (Fmoy,'A:Moy');     rewrite (Fmoy);
    assign (FXk,'A:Xk');       rewrite (FXk);

  write ('voulez-vous reduire les variables ? (O/N) ');
          Readln (Rep);
  read (Fdim,NT,NI); read (Fpoids,NIA);
  for i:=1 to NIA do read (Fpoids,POID[i]);

  for k:=1 to NT do
    begin
      read (Fdim,NV);
      for i:=1 to NI do
        for j:=1 to NV do read (Fdon,X[i,j]);

    (* moyennes et ecart-types des variables du tableau n°k *
     ******************************************************)
      Moyecart (NI,NIA,NV,X,POID,SIGMA,MOY);

      for j:=1 to NV do write (Fmoy,MOY[j]:9:2,' ');
      writeln (Fmoy);
      if Rep='O' then
          begin
            for j:=1 to NV do
              begin
                if SIGMA[j]<1E-5 then SIGMA[j]:=1;
                for i:=1 to NI do
```

```
                    X[i,j]:=(X[i,j]-MOY[j])/SIGMA[j];
                end;
            end
          else
            begin
              for j:=1 to NV do
                for i:=1 to NI do X[i,j]:=X[i,j]-MOY[j];
            end;

        for i:=1 to NI do
          begin
            for j:=1 to NV do write (FXk,X[i,j],' ');
            writeln (FXk);
          end;
        writeln (FXk);
      end;

  close (Fdon); close (Fdim); close (Fpoids); close (Fmoy);
  close (FXk);
  writeln ('fin de l''execution');
  writeln ('les moyennes des variables se trouvent dans A:Moy');
End.
```

2. PRODUITS SCALAIRES ENTRE Wk

```
Program Objet_Wk;
(**************)
     (*************************************************************
      * calcul des matrices de produits scalaires entre individus *
      *                      Wk=Xk*(Xk)'      ( p.111)            *
      *************************************************************)

  Const MaxIND  = 50; MaxVAR  = 25;
  Var   NT,NI,NIA,NV,i,j,k,l :integer;
        LigneW               :array [1..MaxIND] of real;
        X                    :array [1..MaxIND,1..MaxVAR] of real;
        Fdim,Fpoids,FXk,FWk  :text;

  Begin
      assign (Fdim,'A:Dim');     reset (Fdim);
      assign (Fpoids,'A:matD'); reset (Fpoids);
      assign (FXk,'A:Xk');       reset (FXk);
      assign (FWk,'A:Wk');       rewrite (FWk);

    read (Fdim,NT,NI); read (Fpoids,NIA);

    for k:=1 to NT do
      begin
        read (Fdim,NV);
        for i:=1 to NI do
          for j:=1 to NV do read (FXk,X[i,j]);
        for i:=1 to NIA do
          begin
            for j:=1 to NIA do
              begin
                LigneW[j]:=0;
                for l:=1 to NV do LigneW[j]:=LigneW[j]+X[i,l]*X[j,l];
              end;
            for j:=1 to NIA do write (FWk,LigneW[j],' ');
            writeln (FWk);
          end;
        writeln (FWk);
      end;

    close (Fdim); close (Fpoids); close (FXk); close (FWk);
    writeln ('fin de l''execution');
    writeln ('les matrices Wk se trouvent dans A:Wk');
  End.
```

```
Program Produits_scalaires_covv;
(*****************************)
  (*********************************************************************
   * calcul de la matrice S de produits scalaires entre objets Wk      *
   *       S[k1,k2]=covv(k1,k2)=Tr(DWk1DWk2)       (p.84 et 112)       *
   *********************************************************************)

  Const MaxIND  = 50; MaxTAB  = 20;
```

```
Var   NT,NI,NIA,i,j,k1,k2,l :integer;
      BIDON                 :real;
      LigneS                :array [1..MaxTAB] of real;
      POID                  :array [1..MaxIND] of real;
      Wk1,Wk2               :array [1..MaxIND,1..MaxIND] of real;
      Fdim,Fpoids,FWk,Fps   :text;

Begin
    assign (Fdim,'A:Dim');    reset (Fdim);
    assign (Fpoids,'A:MatD'); reset (Fpoids);
    assign (FWk,'A:Wk');      reset (FWk);
    assign (Fps,'A:PS');      rewrite (Fps);

  read (Fdim,NT,NI); read (Fpoids,NIA);
  for i:=1 to NIA do read (Fpoids,POID[i]);

  for k1:=1 to NT do
    begin
      reset (FWk);
      if k1>1 then
       (* position du pointeur sur Wk1 dans le fichier FWk *
        ****************************************************)
        begin
          for l:=1 to k1-1 do
            begin
              for i:=1 to NIA do
                for j:=1 to NIA do read (FWk,BIDON);
            end;
        end;
                  (* lecture de Wk1 et calcul de DWk1 *
                   ************************************)
      for i:=1 to NIA do
        begin
          for j:=1 to NIA do
            begin
              read (FWk,Wk1[i,j]);
              Wk1[i,j]:=Wk1[i,j]*POID[i];
            end;
        end;
      reset (FWk);
      for k2:=1 to k1 do
        begin
                  (* lecture de Wk2 et calcul de DWk2 *
                   ************************************)
          for i:=1 to NIA do
            begin
              for j:=1 to NIA do
                begin
                  read (FWk,Wk2[i,j]);
                  Wk2[i,j]:=Wk2[i,j]*POID[i];
                end;
            end;

 (* calcul du produit scalaire covv(k1,k2) , egal a Tr(DWk1DWk2) *
  ***************************************************************)
          LigneS[k2]:=0;
          for i:=1 to NIA do
```

```
            for j:=1 to NIA do
              LigneS[k2]:=LigneS[k2]+Wk1[i,j]*Wk2[j,i];
          end;
        for k2:=1 to k1 do write (Fps,LigneS[k2],' ');
        writeln (Fps);
      end;

    close (Fdim); close (Fpoids); close (FWk); close (Fps);
    writeln ('fin de l''execution');
    writeln ('les produits scalaires entre Wk se trouvent dans A:ps');
  End.
```

```
Program Choix_produit_scalaire;
(*****************************)
     (*****************************************************************
      * choix du produit scalaire entre etudes : Tr(DWk1DWk2) ou RV *
      *                 p.84 et discussion p.96                     *
      *****************************************************************)

  Const MaxTAB  = 20;
  Var   NT,k,k1,k2     :integer;
        Rep            :char;
        NORME          :array [1..MaxTAB] of real;
        S,RV           :array [1..MaxTAB,1..MaxTAB] of real;
        Fdim,Fps,Fnorm :text;

  Begin
      assign (Fdim,'A:dim');   reset (Fdim);
      assign (Fps,'A:ps');     reset (Fps);
      assign (Fnorm,'A:norm'); rewrite (Fnorm);

    read (Fdim,NT);
    for k1:=1 to NT do
      for k2:=1 to k1 do read (Fps,S[k1,k2]);

     (* calcul des normes des Wk *
      ***************************)
    writeln ('norme du tableau');
    for k:=1 to NT do
      begin
        NORME[k]:=sqrt(S[k,k]);
        writeln ('                n° ',k,' = ',NORME[k]:8:2);
      end;

     (* normalisation ? *
      ********************)
    write ('voulez-vous travailler avec les coefficients RV ? ');
    write ('(O/N) '); readln (Rep); writeln;

     (* calcul de la matrice RV *
      ***************************)
    writeln ('                                matrice RV');
    for k1:=1 to NT do
      begin
```

```
        for k2:=1 to k1 do RV[k1,k2]:=S[k1,k2]/(NORME[k1]*NORME[k2]);
        for k2:=1 to k1 do write (RV[k1,k2]:6:3,' '); writeln;
      end;

    write (Fnorm,Rep,' ');
      for k:=1 to NT do write (Fnorm,NORME[k],' ');
    if Rep='O' then
      begin
        close (Fps); rewrite (Fps);
        for k1:=1 to NT do
          begin
            for k2:=1 to k1 do write (Fps,RV[k1,k2],' ');
            writeln (Fps);
          end;
      end;

    close (Fdim); close(Fps); close(Fnorm);
    writeln ('fin de l''execution');
    writeln ('les normes des tableaux sont dans A:norm ');
    write ('les coefficients RV dans A:ps ');
    writeln ('si vous avez repondu Oui');
  End.



Program Lecture_poids_tableaux;
(*****************************)

  Const MaxTAB  = 20;
  Type colonne  = array [1..MaxTAB] of real;
  Var  NT,NTA,NTS,k :integer;
       POID         :colonne;
       Fdim,Fpoids  :text;

Procedure Poids (N:integer; Var POID:colonne);
(*************)
     (* calcul des poids des tableaux actifs *
      ***************************************)

  Var SOM :real;
      Rep :char;

  Begin
    write ('les tableaux actifs ont-ils le meme poids ? (O/N) ');
            readln (Rep);
    if Rep='O' then
        for k:=1 to N do POID[k]:=1
      else
        begin
          writeln ('rentrer les poids des  ',N,' tableaux actifs : ');
                   writeln ('(return apres chaque valeur)');
          for k:=1 to N do readln (POID[k]);
        end;
    SOM:=0;
    for k:=1 to N do SOM:=SOM+POID[k];
    for k:=1 to N do POID[k]:=POID[k]/SOM;
  End;
```

```
Begin
    assign (Fdim,'A:dim');         reset (Fdim);
    assign (Fpoids,'A:matDelta'); rewrite (Fpoids);

  read (Fdim,NT);

  write ('nombre de tableaux supplementaires parmi les ',NT);
  write (' tableaux = '); readln (NTS);
    if NTS>0 then writeln
      (' les tableaux supplementaires sont supposes etre a la fin');
    writeln;
    NTA:=NT-NTS;

   (* calcul et stockage des poids des tableaux actifs *
    ****************************************************)
  Poids (NTA,POID);
  writeln (Fpoids,NTA,' ');
  for k:=1 to NTA do writeln (Fpoids,POID[k],' ');

  close (Fdim); close (Fpoids);
  writeln ('               fin de l''execution');
  write ('le nombre de tableaux actifs et leurs poids sont dans ');
  writeln ('A:matDelta');
End.
```

```
Program Sauvegarde_produits_scalaires_actifs;
(*****************************************)

  Const MaxTAB  = 20;
  Var  NT,NTA,k,k1,k2       :integer;
       S                    :array [1..MaxTAB,1..MaxTAB] of real;
       Fdim,Fpoids,Fps,Fpsa :text;

  Begin
      assign (Fdim,'A:dim');         reset (Fdim);
      assign (Fpoids,'A:matDelta'); reset (Fpoids);
      assign (Fps,'A:ps');          reset (Fps);
      assign (Fpsa,'A:psa');        rewrite (Fpsa);

    read (Fdim,NT); read (Fpoids,NTA);
    for k1:=1 to NT do
      for k2:=1 to k1 do read (Fps,S[k1,k2]);

      (* copie de la partie active de S dans le fichier Fpsa *
       ******************************************************)
    writeln (Fpsa,NTA);
    for k1:=1 to NTA do
      begin
        for k2:=1 to k1 do write (Fpsa,S[k1,k2],' ');
        writeln (Fpsa);
      end;
```

```
  close (Fdim); close (Fpoids); close (Fps); close(Fpsa);
  writeln ('fin de l''execution');
  writeln ('la partie active du tableau des produits scalaires');
  writeln ('entre Wk se trouve dans A:psa');
End.
```

3. DECOMPOSITION EN VALEURS SINGULIERES DU TABLEAU DES PRODUITS SCALAIRES

```
Program Elements_propres_de_SD;
(**************************)
     (*****************************************
      * decomposition en valeurs singulieres de S *
      *    =  diagonalisation de SD   (p.53)      *
      *****************************************)

  Const NbMax   = 50; NbAxes  = 3;
  Type colonne  = array [1..NbMax] of real;
       tableau  = array [1..NbMax,1..NbMax] of real;
       vecteurs = array [1..NbMax,1..NbAxes] of real;
  Var  N,i,j,l                 :integer;
       Fichier                 :string[20];
       ValProp                 :array [1..NbAxes] of real;
       Diag,SsDiag,POID        :colonne;
       S                       :tableau;
       VectProp                :vecteurs;
       Fpoids,Fpsa,Fvecp,Fvalp :text;

Procedure Tridiagonalisation (N:integer; Var S:tableau;
(*************************)             Var Diag,SsDiag:colonne);

   (*****************************************************************
    * transformation orthogonale de la matrice S pour la rendre     *
    * tridiagonale - methode de Householder -                       *
    * ( EISPACK . Lecture Notes in Computer Science , n°6 , 1976 )  *
    *                                                               *
    * a la fin de la procedure :                                    *
    *   Diag contient les elements de la matrice transformee ,      *
    *   les N-1 premiers elements de SsDiag donnent la sous ( sur) *
    *   diagonale ,                                                 *
    *   S contient la matrice de la transformation orthogonale ,    *
    *   qui est le produit des N-2 transformations ayant annule     *
    *   chacune une ligne .                                         *
    *****************************************************************)

  Var i,j,k,l :integer; A,B,SOM,SOMCARRE,PROD :real;

  Begin
    if N>1 then for i:=N downto 2 do
      begin
        k:=i-1; SOM:=0; SOMCARRE:=0;
        if k=1 then SsDiag[2]:=S[2,1]
          else
            begin
              for j:=1 to k do SOM:=SOM+abs(S[i,j]);
              if SOM<>0 then
                  begin
                    for j:=1 to k do
                      begin
                       (* normalisation pour eviter les overflows *
                        ****************************************)
                        S[i,j]:=S[i,j]/SOM;
                        SOMCARRE:=SOMCARRE+sqr(S[i,j]);
```

```
                    end;
                  if S[i,k]<=0 then A:=sqrt(SOMCARRE)
                    else A:=-sqrt(SOMCARRE);
                  SsDiag[i]:=A*SOM;
                  SOMCARRE:=SOMCARRE-A*S[i,k];
                  S[i,k]:=S[i,k]-A;

                  A:=0;
                  for j:=1 to k do
                    begin
                      S[j,i]:=S[i,j]/SOMCARRE;
                      PROD:=0;
                      for l:=1 to j do PROD:=PROD+S[j,l]*S[i,l];
                      if k>=j+1 then for l:=j+1 to k do
                        PROD:=PROD+S[l,j]*S[i,l];
                      SsDiag[j]:=PROD/SOMCARRE;
                      A:=A+SsDiag[j]*S[i,j];
                    end;
                  B:=A/(2*SOMCARRE);
                  for j:=1 to k do
                    begin
                      A:=SsDiag[j]-B*S[i,j];
                      SsDiag[j]:=A;
                      for l:=1 to j do
                        S[j,l]:=S[j,l]-S[i,j]*SsDiag[l]-A*S[i,l];
                    end;
                end
              else SsDiag[i]:=S[i,k];
          end;
          Diag[i]:=SOMCARRE;
    end;

  for i:=2 to N do SsDiag[i-1]:=SsDiag[i];
                   SsDiag[N]:=0;
  Diag[1]:=0;
  for i:=1 to N do
    begin
      if Diag[i]<>0 then for j:=1 to i-1 do
        begin
          PROD:=0;
          for k:=1 to i-1 do PROD:=PROD+S[i,k]*S[k,j];
          for k:=1 to i-1 do S[k,j]:=S[k,j]-PROD*S[k,i];
        end;
      Diag[i]:=S[i,i];
      S[i,i]:=1;
      if i>1 then for j:=1 to i-1 do
        begin
          S[i,j]:=0; S[j,i]:=0;
        end;
    end;
End;
```

```
Procedure Diagonalisation (N:integer; Var S:tableau;
(**********************)              Var Diag,SsDiag:colonne);
   (*****************************************************************
    * diagonalisation d'une matrice tridiagonale symetrique par la *
    * methode QL implicite (EISPACK guide)                          *
    *****************************************************************)

  Const seuil=1E-5;
  Var i,j,k,l,NbIter :integer;
      A,B,C,D,E,F,G  :real;

Procedure Test (N,i:integer; Diag,SsDiag:colonne; Var j:integer);
(************)
  Var A:real;
  Begin
    j:=i-1;
    repeat
      j:=j+1;
      if j=N then exit;
      A:=seuil*(abs(Diag[j])+abs(Diag[j+1]));
    until abs(SsDiag[j])<=A;
  End;

  Begin
    if N>1 then
      begin
        for i:=1 to N do
          begin
            NbIter:=0;
            Test(N,i,Diag,SsDiag,j);
            while j<>i do
              begin
                if NbIter=70 then
                  begin
                    writeln;
                    write ('a l''etape ',i,' le calcul de la valeur');
                    writeln (' propre demande trop d''iterations ;');
                    write ('augmenter le seuil dans la procedure ');
                    writeln ('test');
                    exit;
                  end;
                NbIter:=NbIter+1;
                A:=(Diag[i+1]-Diag[i])/(2*SsDiag[i]);
                B:=sqrt(sqr(A)+1);
                if A<=0 then A:=Diag[j]-Diag[i]+SsDiag[i]/(A-B)
                        else A:=Diag[j]-Diag[i]+SsDiag[i]/(A+B);
                C:=1;  D:=1;  E:=0;
                for k:=j-1 downto i do
                  begin
                    F:=C*SsDiag[k];  G:=D*SsDiag[k];
                    if abs(F)<abs(A) then
                        begin
                          C:=F/A;  B:=sqrt(sqr(C)+1);
                          SsDiag[k+1]:=A*B;
                          D:=1/B;  C:=C*D;
                        end
```

```
                    else
                      begin
                        D:=A/F;  B:=sqrt(sqr(D)+1);
                        SsDiag[k+1]:=B*F;
                        C:=1/B;  D:=C*D;
                      end;
                  A:=Diag[k+1]-E;  B:=(Diag[k]-A)*C+2*D*G;
                  E:=B*C;  Diag[k+1]:=A+E;  A:=B*D-G;
                  for l:=1 to N do
                    begin
                      F:=S[l,k+1];
                      S[l,k+1]:=C*S[l,k]+D*F;
                      S[l,k]  :=D*S[l,k]-C*F;
                    end;
                end;
                Diag[i]:=Diag[i]-E;
                SsDiag[i]:=A;  SsDiag[j]:=0;
                Test(N,i,Diag,SsDiag,j);
            end;
        end;
    end;
End;


Procedure Ordre (N:integer; Var Diag:colonne; Var S:tableau);
(*************)
    (*******************************************************
     * rangement des valeurs propres par ordre decroissant *
     *******************************************************)

  Var i,k,max :integer; A :real; echange :boolean;

  Begin
    max:=N;
    echange:=true;
    while echange do
      begin
        echange:=false;
        for k:=1 to max-1 do
          if Diag[k]<Diag[k+1] then
            begin
              A:=Diag[k];
              Diag[k]:=Diag[k+1];
              Diag[k+1]:=A;
              for i:=1 to N do
                begin
                  A:=S[i,k];
                  S[i,k]:=S[i,k+1];
                  S[i,k+1]:=A;
                end;
              max:=k;
              echange:=true;
            end;
      end;
  End;
```

```
Begin
  write ('nom du fichier dans lequel se trouvent les poids ');
  write ('des objets : '); readln (Fichier);
    assign (Fpoids,Fichier); reset (Fpoids);
    assign (Fpsa,'A:psa');   reset (Fpsa);
    assign (Fvecp,'A:vecp'); rewrite (Fvecp);
    assign (Fvalp,'A:valp'); rewrite (Fvalp);

  read (Fpoids,N); read (Fpsa,N);
  for i:=1 to N do read (Fpoids,POID[i]);
  for i:=1 to N do
    for j:=1 to i do read (Fpsa,S[i,j]);

   (* calcul de (D)1/2*S*(D)1/2 *
    ***************************)
  for i:=1 to N do
    for j:=1 to i do S[i,j]:=S[i,j]*sqrt(POID[i]*POID[j]);

   (* diagonalisation de la matrice symetrique (D)1/2*S*(D)1/2 *
    ***********************************************************)
  Tridiagonalisation (N,S,Diag,SsDiag);
  Diagonalisation (N,S,Diag,SsDiag);
  Ordre (N,Diag,S);

  for l:=1 to NbAxes do
    begin
      ValProp[l]:=Diag[l];
      for i:=1 to N do VectProp[i,l]:=S[i,l];
    end;

   (* calcul des vecteurs propres D-orthonormes de SD *
    **************************************************)
  for l:=1 to NbAxes do write (Fvalp,ValProp[l],' ');
  for i:=1 to N do
    begin
      for l:=1 to NbAxes do
        VectProp[i,l]:=VectProp[i,l]/sqrt(POID[i]);
      for l:=1 to NbAxes do write (Fvecp,VectProp[i,l],' ');
      writeln (Fvecp);
    end;

  close (Fpoids); close (Fpsa); close (Fvecp); close (Fvalp);
  writeln ('fin de l''execution');
  writeln ('les valeurs propres sont dans A:valp');
  writeln ('les vecteurs propres sont dans A:vecp');
End.
```

4. IMAGE EUCLIDIENNE DES ETUDES

```
Program Coefficients_compromis;
(***************************)
   (*************************************************************
    * calcul des coefficients du compromis a partir des elements *
    * propres du tableau non centre des produits scalaires entre *
    * objets Wk    (p.116)                                       *
    *************************************************************)

  Const MaxTAB  = 20; NbAxes  =  3;
  Var  NTA,NIA,k,l      :integer;
       SOM              :real;
       Rep              :char;
       POID,NORME,COEFF :array [1..MaxTAB] of real;
       ValProp          :array [1..NbAxes] of real;
       VectProp         :array [1..MaxTAB,1..NbAxes] of real;
       Fpoids,Fnorm,Fvecp,Fvalp,FmatD,Fcoef :text;

  Begin
      assign (Fpoids,'A:matDelta'); reset (Fpoids);
      assign (Fnorm,'A:norm');      reset (Fnorm);
      assign (Fvecp,'A:vecp');      reset (Fvecp);
      assign (Fvalp,'A:valp');      reset (Fvalp);
      assign (FmatD,'A:matD');      reset (FmatD);
      assign (Fcoef,'A:coef');      rewrite (Fcoef);

    read (Fpoids,NTA); read (Fvalp,ValProp[1]);
    for k:=1 to NTA do read (Fpoids,POID[k]);
    for k:=1 to NTA do
      for l:=1 to NbAxes do read (Fvecp,VectProp[k,l]);
    read (Fnorm,Rep);
    for k:=1 to NTA do read (Fnorm,NORME[k]);

    writeln ('les coefficients du compromis sont : ');

    if Rep='O' then
               (* cas d'objets normes Wk/|Wk| *
                ******************************)
       for k:=1 to NTA do
         begin
           COEFF[k]:=POID[k]*abs(VectProp[k,1])/sqrt(ValProp[1]);
           writeln ('                              ',COEFF[k]:6:3);
           COEFF[k]:=COEFF[k]/NORME[k];
         end
      else
               (* cas d'objets Wk *
                ******************)
        begin
          SOM:=0; for k:=1 to NTA do SOM:=SOM+POID[k]*NORME[k];
          for k:=1 to NTA do
            begin
              COEFF[k]:=
                  SOM*POID[k]*abs(VectProp[k,1])/sqrt(ValProp[1]);
              writeln ('                              ',COEFF[k]:6:3);
            end;
        end;
```

```
    read (FmatD,NIA); writeln (Fcoef,NTA,' ',NIA);
    for k:=1 to NTA do write (Fcoef,COEFF[k],' ');

    close (Fpoids); close (Fnorm); close (Fvecp); close (Fvalp);
    close (FmatD); close (Fcoef);
    writeln ('fin de l''execution');
    writeln ('les coefficients du compromis sont dans A:coef');
  End.



Program Coordonnees_tableaux;
(*************************)
     (*************************************************
      * coordonnees de l'image euclidienne des tableaux *
      *************************************************)

  Const MaxTAB  = 20; NbAxes  = 3;
  Var   NT,NTA,k1,k2,k,l :integer;
        Rep              :char;
        Fichier          :string[20];
        ValProp          :array [1..NbAxes] of real;
        POID             :array [1..MaxTAB] of real;
        S                :array [1..MaxTAB,1..MaxTAB] of real;
        COORD,VectProp   :array [1..MaxTAB,1..NbAxes] of real;
        Fdim,Fpoids,Fps,Fvecp,Fvalp,Fcotab :text;

  Begin
      assign (Fdim,'A:dim');         reset (Fdim);
      assign (Fpoids,'A:matDelta'); reset (Fpoids);
      assign (Fps,'A:ps');           reset (Fps);
      assign (Fvecp,'A:vecp');       reset (Fvecp);
      assign (Fvalp,'A:valp');       reset (Fvalp);
      assign (Fcotab,'A:cotab');     rewrite (Fcotab);

    read (Fdim,NT); read (Fpoids,NTA);
    for l:=1 to NbAxes do read (Fvalp,ValProp[l]);
    for k:=1 to NTA do
      for l:=1 to NbAxes do read (Fvecp,VectProp[k,l]);

     (* calcul des coordonnees des tableaux actifs *
      *********************************************)
    for k:=1 to NTA do
      for l:=1 to NbAxes do
        COORD[k,l]:=VectProp[k,l]*sqrt(ValProp[l]);

    if NTA<NT then
      (* calcul des coordonnees des tableaux supplementaires *
       *****************************************************)
      begin
        for k:=1 to NTA do read (Fpoids,POID[k]);
        for k1:=1 to NT do
          for k2:=1 to k1 do read (Fps,S[k1,k2]);
        for k1:=NTA+1 to NT do
          for l:=1 to NbAxes do
```

```
            begin
              COORD[k1,l]:=0;
              for k2:=1 to NTA do COORD[k1,l]:=
                COORD[k1,l]+S[k1,k2]*VectProp[k2,l]*POID[k2];
              COORD[k1,l]:=COORD[k1,l]/sqrt(ValProp[l]);
            end;
      end;

    for k:=1 to NT do
      begin
        for l:=1 to NbAxes do write (Fcotab,COORD[k,l],' ');
        writeln (Fcotab);
      end;

    close (Fdim); close (Fpoids); close (Fps); close(Fvecp);
    close(Fvalp); close (Fcotab);
    writeln ('fin de l''execution');
    writeln ('les coordonnees des tableaux se trouvent dans A:cotab');
  End.
```

```
Program Nuage;
(***********)
  (*****************************************************************
   *            trace d'un nuage de points dans un plan            *
   * en mode graphique simple l'ecran constitue une grille de 24   *
   * lignes et 80 colonnes , l'origine etant situee en haut et     *
   * a gauche . le nuage est represente dans un repere orthonorme *
   *****************************************************************)

  Const NbMax  = 51; NbAxes =  3; NbLig  = 21; NbCol  = 74;
  Type  ligne          = array [1..NbAxes] of real;
        vecteurs       = array [1..NbMax,1..NbAxes] of real;
        identificateur = array [1..NbMax] of string[2];
  Var   N,AxeVert,i,l :integer;
        UnHor,UnVert  :real;
        Rep           :char;
        Fichier       :string[20];
        Min,Max       :ligne;
        COORD         :vecteurs;
        NOM           :identificateur;
        Fcoord        :text;

Procedure MinMax (N:integer; COORD:vecteurs; Var Min,Max:ligne);
(**************)
     (* recherche des valeurs minimum et maximum sur chaque axe *
      ***********************************************************)

  Begin
    for l:=1 to NbAxes do
      begin
        COORD[N+1,l]:=0;          (* on rajoute l'origine *
                                   ************************)
        Min[l]:=COORD[1,l];  Max[l]:=COORD[1,l];
        for i:=2 to N+1 do
```

```
          begin
            if Min[l]>COORD[i,l] then Min[l]:=COORD[i,l];
            if Max[l]<COORD[i,l] then Max[l]:=COORD[i,l];
          end;
      end;
  End;

Procedure Unite (Min,Max:ligne; Var UnHor,UnVert:real);
(*************)
   (* calcul du nombre de lignes (colonnes) correspondant a l'unite *
    * pour que le repere soit orthonorme                            *
    ****************************************************************)

  Const Largeur = 23; Hauteur = 16;
  Var   A :real;

  Begin
     (* nombre de lignes par unite sur l'axe vertical *
      ************************************************)
    if (Max[2]-Min[2])>=(Max[3]-Min[3]) then
        UnVert:=NbLig/(Max[2]-Min[2])
     else UnVert:=NbLig/(Max[3]-Min[3]);

     (* nombre de colonnes par unite sur l'axe horizontal *
      *****************************************************)
    UnHor:=UnVert*Hauteur/NbLig*NbCol/Largeur;

    if UnHor*(Max[1]-Min[1])>NbCol then
      begin
        A:=NbCol/(UnHor*(Max[1]-Min[1]));
        UnHor:=UnHor*A;
        UnVert:=UnVert*A;
      end;
  End;

Procedure Axes (AxeVert:integer; UnHor,UnVert:real; Min,Max:ligne);
(************)
     (* trace de l'axe horizontal puis de l' axe vertical *
      *****************************************************)

  Var Ligne0,Colonne0 :integer;

  Begin
    Ligne0:=round(Max[AxeVert]*UnVert)+1;
    for i:=1 to NbCol+1 do
      begin
        gotoxy (i,Ligne0); write ('-');
      end;
    gotoxy (NbCol-6,Ligne0+1); writeln ('Axe 1');

    Colonne0:=round(-Min[1]*UnHor)+1;
    gotoxy (Colonne0+1,1); writeln ('Axe ',AxeVert);
    for i:=1 to NbLig+1 do
      begin
        gotoxy (Colonne0,i); write('i');
      end;
  End;
```

```
Procedure Trace (N,AxeVert:integer; UnHor,UnVert:real; Min,Max:ligne;
(*************)                    COORD:vecteurs; NOM:identificateur);

  Var Col,Lig :integer;

  Begin
    clrscr;
    Axes (AxeVert,UnHor,UnVert,Min,Max);
    for i:=1 to N do
      begin
        Lig:=round((Max[AxeVert]-COORD[i,AxeVert])*UnVert)+1;
        Col:=round((COORD[i,1]-Min[1])*UnHor)+1;
        gotoxy (Col,Lig);
        write (NOM[i]);
      end;
  End;


  Begin
    write ('dans quel fichier se trouvent les coordonnees ? ');
           readln (Fichier);
           assign (Fcoord,Fichier); reset (Fcoord);
    write ('nombre de points ?'); readln (N);
    writeln ('identificateurs des points sur 2 caracteres : ');
             writeln ('(return apres chaque nom)');
             for i:=1 to N do readln (NOM[i]);
    for i:=1 to N do
      for l:=1 to NbAxes do read (Fcoord,COORD[i,l]);

    MinMax (N,COORD,Min,Max);
    Unite (Min,Max,UnHor,UnVert);

    for l:=2 to NbAxes do
      begin
        Trace (N,l,UnHor,UnVert,Min,Max,COORD,NOM);
        gotoxy (1,NbLig+3);
        write ('appuyer sur une touche , puis sur return ');
        write ('pour continuer '); readln (Rep);
      end;
    close (Fcoord);
  End.



Program Centrage_produits_scalaires;
(*******************************)
     (*****************************************************************
      * le tableau des produits scalaires S est centre par rapport *
      * aux poids des tableaux actifs  (p.62)                       *
      *****************************************************************)

  Const MaxTab  = 20;
  Var   NT,NTA,i,j      :integer;
        total           :real;
```

```
      POID,SOM        :array [1..MaxTab] of real;
      S               :array [1..MaxTab,1..MaxTab] of real;
      Fdim,Fpoids,Fps :text;

Begin
    assign (Fdim,'A:dim');         reset (Fdim);
    assign (Fpoids,'A:matDelta'); reset (Fpoids);
    assign (Fps,'A:ps');           reset (Fps);

  read (Fdim,NT); read (Fpoids,NTA);
  for i:=1 to NTA do read (Fpoids,POID[i]);
  for i:=1 to NT do
    for j:=1 to i do read (Fps,S[i,j]);

                  (* on complete le tableau S *
                   ***************************)
  for i:=1 to NT do
    for j:=1 to i do S[j,i]:=S[i,j];

                  (* calcul des sommes des lignes *
                   ********************************)
  total:=0;
  for i:=1 to NTA do
    begin
      SOM[i]:=0;
      for j:=1 to NTA do SOM[i]:=SOM[i]+POID[j]*S[i,j];
      total:=total+POID[i]*SOM[i];
    end;

        (* on remplace le tableau S par le tableau centre *
         ************************************************)
  for i:=1 to NT do
    for j:=1 to NT do S[i,j]:=S[i,j]-SOM[i]-SOM[j]+total;
  close (Fps); rewrite (Fps);
  for i:=1 to NT do
    begin
      for j:=1 to i do write (Fps,S[i,j],' '); writeln (Fps);
    end;

  close (Fdim); close (Fpoids); close (Fps);
  writeln ('fin de l''execution');
  write ('le tableau centre des produits scalaires ');
  writeln ('se trouve dans A:ps');
End.

(*************************************************************
 * pour obtenir l'image euclidienne centree , utiliser les  *
 * programmes : 'sauvegarde produits scalaires actifs' ,    *
 * 'elements propres de SD' , 'coordonnees tableaux'        *
 * (le fichier A:cotab contient alors les coordonnees des   *
 * tableaux dans l'image euclidienne centree ) , et 'nuage' *
 *************************************************************)
```

5. POSITIONS COMPROMIS DES INDIVIDUS

```
Program Compromis;
(***************)
     (*******************************
      * calcul de l'objet compromis *
      *******************************)

  Const MaxIND  = 50; MaxTAB  = 20;
  Var   NTA,N,i,j,k     :integer;
        COEFF           :array [1..MaxTAB] of real;
        Wk,W            :array [1..MaxIND,1..MaxIND] of real;
        Fcoef,FWk,Fpsa :text;

  Begin
      assign (Fcoef,'A:coef'); reset (Fcoef);
      assign (FWk,'A:Wk');     reset (FWk);
      assign (Fpsa,'A:psa');   rewrite (Fpsa);

    read (Fcoef,NTA,N);
    for k:=1 to NTA do read (Fcoef,COEFF[k]);

    for i:=1 to N do
      for j:=1 to N do W[i,j]:=0;

    for k:=1 to NTA do
      begin
        for i:=1 to N do
          for j:=1 to N do
            begin
              read (FWk,Wk[i,j]);
              W[i,j]:=W[i,j]+COEFF[k]*Wk[i,j];
            end;
      end;

    writeln (Fpsa,N);
    for i:=1 to N do
      begin
        for j:=1 to i do write (Fpsa,W[i,j],' ');
        writeln (Fpsa);
      end;

    close (Fcoef); close (FWk); close (Fpsa);
    writeln ('fin de l''execution');
    writeln ('la matrice compromis W se trouve dans A:psa');
  End.

  (*****************************************************************
   * pour obtenir l'image euclidienne compromis , il faut calculer *
   * les valeurs singulieres de W en utilisant le programme :      *
   *        'elements propres de SD'                               *
   *****************************************************************)
```

```
Program Coordonnees_compromis;
(*************************)
     (*********************************************
      * coordonnees de l'image euclidienne compromis *
      *********************************************)

  Const MaxIND  = 50; NbAxes  =  3;
  Var  N,i,l                  :integer;
       ValProp                :array [1..NbAxes] of real;
       COORD,VectProp         :array [1..MaxIND,1..NbAxes] of real;
       Fpsa,Fvecp,Fvalp,Fcomp :text;

  Begin
      assign (Fpsa,'A:psa');    reset (Fpsa);
      assign (Fvecp,'A:vecp');  reset (Fvecp);
      assign (Fvalp,'A:valp');  reset (Fvalp);
      assign (Fcomp,'A:comp');  rewrite (Fcomp);

    read (Fpsa,N);
    for l:=1 to NbAxes do read (Fvalp,ValProp[l]);
    for i:=1 to N do
      for l:=1 to NbAxes do read (Fvecp,VectProp[i,l]);

    for i:=1 to N do
      for l:=1 to NbAxes do
        COORD[i,l]:=VectProp[i,l]*sqrt(ValProp[l]);

    for i:=1 to N do
      begin
        for l:=1 to NbAxes do write (Fcomp,COORD[i,l],' ');
        writeln (Fcomp);
      end;

    close (Fpsa); close (Fvecp); close (Fvalp); close (Fcomp);
    writeln ('fin de l''execution');
    write ('les coordonnees de l''image euclidienne compromis sont ');
    writeln ('dans A:comp');
  End.



Program Correlations_variables_avec_axes;
(************************************)
  Const MaxIND = 50; MaxVAR = 25; MaxTAB = 20; NbAxes = 3;
  Type  ligne       = array [1..MaxVAR] of real;
        colonne     = array [1..MaxIND] of real;
        tableau     = array [1..MaxIND,1..MaxVAR] of real;
        vecteurs    = array [1..MaxIND,1..NbAxes] of real;
  Var   NT,NI,NIA,NV,i,j,k,l        :integer;
        Rep                         :char;
        POID                        :colonne;
        MOY,SIGMA                   :ligne;
        X                           :tableau;
        VectProp,COORD              :vecteurs;
        Fdim,FmatD,FXk,Fpoids,Fvecp :text;
```

```
Procedure Moyecart (NI,NIA,NV:integer; X:tableau; POID:colonne;
(****************)                    Var SIGMA,MOY:ligne);

   (* calcul des moyennes et ecart-types des variables du tableau X *
    ***************************************************************)

  Var   A,C,D,SOM  :real;
        B          :ligne;

  Begin
    SOM:=POID[1];
    for j:=1 to NV do
      begin
        SIGMA[j]:=0;
        MOY[j]  :=X[1,j];
      end;
    for i:=2 to NIA do
      begin
        A:=SOM+POID[i];
        if abs(A)<1E-5 then A:=1E-5;
        for j:=1 to NV do
          begin
            B[j]    :=X[i,j]-MOY[j];
            C       :=B[j]*POID[i]/A;
            D       :=C*SOM;
            MOY[j]  :=MOY[j]+C;
            SIGMA[j]:=SIGMA[j]+D*B[j];
          end;
        SOM:=A;
      end;
      for j:=1 to NV do SIGMA[j]:=sqrt(SIGMA[j]/SOM);
  End;


  Begin
      assign (Fdim,'A:Dim');         reset (Fdim);
      assign (FmatD,'A:MatD');       reset (FmatD);
      assign (FXk,'A:Xk');           reset (FXk);
      assign (Fpoids,'A:matDelta'); reset (Fpoids);
      assign (Fvecp,'A:vecp');       reset (Fvecp);

    read (Fdim,NT,NI);  read (FmatD,NIA);
    for i:=1 to NIA do read (FmatD,POID[i]);
    for i:=1 to NIA do
      for l:=1 to NbAxes do read (Fvecp,VectProp[i,l]);

    writeln ('correlations des variables avec les ',NbAxes,' axes:');
    for k:=1 to NT do
     (* lecture du tableau n°k centre (eventuellement reduit) *
      ******************************************************)
      begin
        read (Fdim,NV);
        for i:=1 to NI do
          for j:=1 to NV do read (FXk,X[i,j]);

     (* moyennes et ecart-types des variables du tableau n°k *
      ******************************************************)
```

```
      Moyecart (NI,NIA,NV,X,POID,SIGMA,MOY);

    (* reduction des variables du tableau n°k *
     ****************************************)
      for j:=1 to NV do
        begin
          if SIGMA[j]<1E-5 then SIGMA[j]:=1;
          for i:=1 to NI do
            X[i,j]:=(X[i,j]-MOY[j])/SIGMA[j];
        end;

    (* correlations des variables du tableau n°k avec les axes *
     ********************************************************)
      for l:=1 to NbAxes do
        for j:=1 to NV do
          begin
            COORD[j,l]:=0;
            for i:=1 to NIA do
              COORD[j,l]:=COORD[j,l]+X[i,j]*VectProp[i,l]*POID[i];
          end;

      writeln; writeln ('tableau n°',k);
               writeln ('***********');
      for j:=1 to NV do
        begin
          for l:=1 to NbAxes do write (COORD[j,l]:6:3,' '); writeln;
        end;
      write ('appuyer sur une touche  , puis sur return  ');
      write ('pour continuer : '); readln (Rep);
    end;

  close (Fdim); close (FmatD); close (FXk); close (Fpoids);
  close (Fvecp);
  writeln ('fin de l''execution');
End.
```

6. TRAJECTOIRES DES INDIVIDUS

```
Program Coordonnees_trajectoires;
(*****************************)

  Const MaxIND = 50; MaxVAR = 25; MaxTAB = 20; NbAxes = 3;
  Var   NT,NI,NIA,NV,i,i1,j,k,l:integer;
        Rep          :char;
        NORME        :array [1..MaxTAB] of real;
        LigneW,POID  :array [1..MaxIND] of real;
        ValProp      :array [1..NbAxes] of real;
        X            :array [1..MaxIND,1..MaxVAR] of real;
        VectProp     :array [1..MaxIND,1..NbAxes] of real;
        COORD        :array [1..MaxIND,1..MaxTAB,1..NbAxes] of real;
        Fdim,Fpoids,Fnorm,FWk,Fvecp,Fvalp,FXk,Fcotraj :text;

  Begin
      assign (Fdim,'A:dim');        reset (Fdim);
      assign (Fpoids,'A:matD');     reset (Fpoids);
      assign (Fnorm,'A:norm');      reset (Fnorm);
      assign (FWk,'A:Wk');          reset (FWk);
      assign (Fvecp,'A:vecp');      reset (Fvecp);
      assign (Fvalp,'A:valp');      reset (Fvalp);
      assign (FXk,'A:Xk');          reset (FXk);
      assign (Fcotraj,'A:cotraj'); rewrite (Fcotraj);

    read (Fdim,NT,NI);  read (Fpoids,NIA);
    for i:=1 to NIA do read (Fpoids,POID[i]);
    for l:=1 to NbAxes do read (Fvalp,ValProp[l]);
    for i:=1 to NIA do
      for l:=1 to NbAxes do read (Fvecp,VectProp[i,l]);

    read (Fnorm,Rep);
    if Rep='O' then
      for k:=1 to NT do read (Fnorm,NORME[k]);

   (* calcul des coordonnees des trajectoires des individus actifs *
    ***************************************************************)
    for k:=1 to NT do
      begin
        for i:=1 to NIA do
          begin
            for j:=1 to NIA do read (FWk,LigneW[j]);
            for l:=1 to NbAxes do
              begin
                COORD[i,k,l]:=0;
                for j:=1 to NIA do COORD[i,k,l]:=
                  COORD[i,k,l]+LigneW[j]*VectProp[j,l]*POID[j];
                COORD[i,k,l]:=COORD[i,k,l]/sqrt(ValProp[l]);
                if Rep='O' then COORD[i,k,l]:=COORD[i,k,l]/NORME[k];
              end;
          end;
      end;

    if NIA<NI then
    (* coordonnees des trajectoires des individus supplementaires *
     ************************************************************)
```

```
      begin
        for k:=1 to NT do
          begin
           (* produits scalaires de l'individu supplementaire i1 *
            * avec les individus actifs                          *
            ****************************************************)
            read (Fdim,NV);
            for i:=1 to NI do
              for j:=1 to NV do read (FXk,X[i,j]);
            for i1:=NIA+1 to NI do
              begin
                for i:=1 to NIA do
                  begin
                    LigneW[i]:=0;
                    for j:=1 to NV do
                      LigneW[i]:=LigneW[i]+X[i1,j]*X[i,j];
                  end;

                for l:=1 to NbAxes do
                  begin
                    COORD[i1,k,l]:=0;
                    for j:=1 to NIA do
                      COORD[i1,k,l]:=COORD[i1,k,l]+
                                    LigneW[j]*VectProp[j,l]*POID[j] ;
                    COORD[i1,k,l]:=COORD[i1,k,l]/sqrt(ValProp[l]);
                    if Rep='O' then
                      COORD[i1,k,l]:=COORD[i1,k,l]/NORME[k];
                  end;
              end;
          end;
      end;

    for i:=1 to NI do
      begin
        for k:=1 to NT do
          begin
            for l:=1 to NbAxes do write (Fcotraj,COORD[i,k,l],' ');
            writeln (Fcotraj);
          end;
        writeln (Fcotraj);
      end;

    close (Fdim); close (Fpoids); close (Fnorm); close (FWk);
    close (Fvecp); close (Fvalp); close (FXk); close (Fcotraj);
    writeln ('fin de l''execution');
    write ('les coordonnees des trajectoires se trouvent dans ');
    writeln ('A:cotraj');
  End.



Program Trace_trajectoires;
(************************)

  Const MaxIND  = 50; MaxTAB  = 20; NbAxes  =  3;
        NbLig   = 20; NbCol   = 78;
```

```
  Type  ligne   = array [1..NbAxes] of real;
        colonne = array [1..MaxTAB] of real;
        cube    = array [1..MaxIND,1..MaxTAB,1..NbAxes] of real;
  Var   NT,N,i,k,l,m :integer;
        UnHor,UnVert :real;
        condition    :boolean;
        Rep          :char;
        Min,Max      :ligne;
        COORD        :cube;
        Fdim,Fcotraj :text;

Procedure MinMax (COORD:cube; Var Min,Max:ligne);
(**************)
     (* recherche des valeurs minimum et maximum sur chaque axe *
      * pour l'ensemble des trajectoires                         *
      ************************************************************)

  Begin
    for l:=1 to NbAxes do
      begin
        Min[l]:=COORD[1,1,l];  Max[l]:=COORD[1,1,l];
        for i:=1 to N do
          for k:=1 to NT do
            begin
              if Min[l]>COORD[i,k,l] then Min[l]:=COORD[i,k,l];
              if Max[l]<COORD[i,k,l] then Max[l]:=COORD[i,k,l];
            end;
      end;
  End;


Function Unite (AxeVert:integer; Min,Max:ligne):real;
(*************)
     (* nombre de lignes par unite sur l'axe vertical *
      *************************************************)
  Begin
    Unite:=NbLig/(Max[AxeVert]-Min[AxeVert]);
  End;


Procedure Axes (AxeVert:integer; UnHor,UnVert:real; Min,Max:ligne);
(*************)
     (* trace de l'axe horizontal puis de l' axe vertical *
      *****************************************************)

  Var j,Ligne0,Colonne0 :integer;

  Begin
    Ligne0:=round(Max[AxeVert]*UnVert)+1;
    for j:=1 to NbCol do
      begin
        gotoxy (j,Ligne0); write ('-');
      end;
    gotoxy (NbCol-5,Ligne0+1); writeln ('Index');

    Colonne0:=1;
    gotoxy (Colonne0+1,1); writeln ('Axe ',AxeVert);
```

```
    for j:=1 to NbLig+1 do
      begin
        gotoxy (Colonne0,j); write('i');
      end;
  End;


Procedure Trajectoire (i,AxeVert:integer; UnHor,UnVert:real;
(*******************)                    Min,Max:ligne; COORD:cube);
  Var x,y     :integer;
      a,b     :real;
      Lig,Col :array [1..MaxTAB] of integer;

  Begin
    for k:=1 to NT do
      begin
        Col[k]:=round((k-1)*UnHor)+2;
        Lig[k]:=round((Max[AxeVert]-COORD[i,k,AxeVert])*UnVert)+1;
      end;
    for k:=1 to NT-1 do
      begin
        a:=(Lig[k+1]-Lig[k])/(Col[k+1]-Col[k]);
        b:=Lig[k]-a*Col[k];
        for x:=Col[k] to Col[k+1] do
          begin
            y:=round(a*x+b);
            gotoxy (x,y);
            write ('.');
          end;
      end;
    gotoxy (Col[1],Lig[1]);
    write (i);
  End;


  Begin
      assign (Fdim,'A:dim');       reset (Fdim);
      assign (Fcotraj,'A:cotraj'); reset (Fcotraj);

    read (Fdim,NT,N);
    for i:=1 to N do
      for k:=1 to NT do
        for l:=1 to NbAxes do read (Fcotraj,COORD[i,k,l]);

    UnHor:=NbCol/NT;
    MinMax (COORD,Min,Max);

    for l:=1 to 2 do
      begin
        i:=1;
        while i<=N do
          begin
            clrscr;
            gotoxy (1,NbLig+2);
            write ('developpement des trajectoires le long de ');
            writeln ('l''axe ',l,' (vertical) ,en fonction');
            writeln ('de l''index du tableau (horizontal)');
```

```
            UnVert:=unite(l,Min,Max);
            Axes (l,UnHor,UnVert,Min,Max);

            m:=l;
            repeat
              Trajectoire (i,l,UnHor,UnVert,Min,Max,COORD);
              m:=m+l;  i:=i+l;
              condition:=(m>4) or (i>N);
            until condition;
            gotoxy (l,NbLig+4);
            write ('appuyer sur une touche , puis sur return ');
            write ('pour continuer : '); readln (Rep);
          end;
      end;
    close (Fdim); close (Fcotraj);
  End.
```

```
7. DONNEES CONCERNANT LES MEMES VARIABLES
  7.1. PRELIMINAIRES

Program Lecture_dimensions;
(************************)

  Const MaxIND  = 50; MaxVAR  = 25; MaxTAB  = 20;
  Var   NT,NI,NV,k :integer;
        Rep        :char;
        Fdim       :text;

  Begin
      assign (Fdim,'A:dim'); rewrite (Fdim);
    write ('nombre de tableaux = '); readln (NT);
      if NT>MaxTAB then
        begin
          writeln ('le nombre de tableaux est trop grand');
          writeln ('modifier la constante MaxTAB');
          exit;
        end;
    write ('nombre de variables par tableau = '); readln (NV);
      if NV>MaxVAR then
        begin
          writeln ('le nombre de variables est trop grand');
          writeln ('modifier la constante MaxVAR');
          exit;
        end;
    write (Fdim,NT,' ',NV,' ');

    write ('les individus sont-ils les memes ');
    write ('pour tous les tableaux? (O/N) '); readln (Rep);
    if Rep='O' then
        begin
          write ('nombre d''individus = '); readln (NI);
          if NI>MaxIND then
            begin
              writeln ('le nombre d''individus est trop grand');
              writeln ('modifier la constante MaxIND');
              exit;
            end;
          for k:=1 to NT do write (Fdim, NI,' ');
        end
      else
        begin
          for k:=1 to NT do
            begin
              write ('nombre d''individus du tableau n°',k,' = ');
                      readln (NI);
              if NI>MaxIND then
                begin
                  writeln ('le nombre d''individus est trop grand');
                  writeln ('modifier la constante MaxIND');
                  exit;
                end;
              write (Fdim,NI,' ');
            end;
        end;
```

```
    close (Fdim);
    writeln ('               fin de l''execution');
    writeln ('les dimensions des tableaux sont dans A:dim');
  End.



Program Lecture_poids_individus;
(*****************************)

  Const MaxIND  = 50;
  Type  colonne = array [1..MaxIND] of real;
  Var   NT,NV,NI,NIS,NIA,i,k :integer;
        Rep                  :char;
        POID                 :colonne;
        Fdim,Fpoids          :text;

Procedure Poids (N:integer; Var POID:colonne);
(*************)
     (* calcul des poids des individus actifs *
      ****************************************)

  Var SOM :real;

  Begin
    write ('les individus actifs ont-ils le meme poids ? (O/N)  ');
           readln (Rep);
    if Rep='O' then
        for i:=1 to N do POID[i]:= 1
      else
        begin
          writeln ('rentrer les poids des ',N,' individus actifs : ');
                   writeln ('(return apres chaque valeur)');
          for i:=1 to N do readln (POID[i]);
        end;
    SOM:= 0;
    for i:=1 to N do SOM:=SOM+POID[i];
    for i:=1 to N do POID[i]:= POID[i]/SOM;
  End;


  Begin
      assign (Fdim,'A:dim');     reset (Fdim);
      assign (Fpoids,'A:matD'); rewrite (Fpoids);

    read (Fdim,NT,NV);
    write ('les individus sont-ils les memes ');
    write ('pour tous les tableaux ? (O/N) '); readln (Rep);
    if Rep ='O' then
        begin
          read (Fdim,NI);
          write ('nombre d''individus supplementaires = ');
                  readln (NIS);
          if NIS > 0 then write ('les individus supplementaires ');
            writeln ('sont supposes etre a la fin'); writeln;
```

```
          NIA:= NI-NIS;
          Poids (NIA,POID);
          for k:=1 to NT do
            begin
              writeln (Fpoids,NIA,' ');
              for i:=1 to NIA do writeln (Fpoids,POID[i],' ');
            end;
        end
      else
        for k:=1 to NT do
          begin
            read (Fdim,NI);
            write ('nombre d''individus supplementaires parmi les ');
            write (NI,' individus du tableau n° ',k,' = ');
                   readln (NIS);
            if NIS > 0 then write ('les individus supplementaires ');
              writeln ('sont supposes etre a la fin'); writeln;
            NIA:= NI-NIS;
            Poids (NIA,POID);
            writeln (Fpoids,NIA,' ');
            for i:=1 to NIA do writeln (Fpoids,POID[i],' ');
          end;

    close (Fdim); close (Fpoids);
    writeln ('               fin de l''execution');
    write ('le nombre d''individus actifs et leurs poids sont ');
    writeln ('dans A:MatD');
  End.

     (*********************************************************
      * utiliser le programme 'centrage_donnees' pour saisir *
      * centrer et eventuellement reduire les donnees        *
      *********************************************************)
```

7.2. IMAGE EUCLIDIENNE DES ETUDES

```
Program Objet_Vk;
(**************)
     (*************************************************************
      * calcul des matrices de produits scalaires entre variables *
      *                    Vk=(Xk)'* D * Xk                       *
      *************************************************************)

  Const MaxIND  = 50; MaxVAR  = 25;
  Var   NT,NI,NIA,NV,i,j,k,l :integer;
        POID                 :array [1..MaxIND] of real;
        LigneV               :array [1..MaxVAR] of real;
        X                    :array [1..MaxIND,1..MaxVAR] of real;
        Fdim,Fpoids,FXk,FWk  :text;

  Begin
      assign (Fdim,'A:Dim');    reset (Fdim);
      assign (Fpoids,'A:matD'); reset (Fpoids);
      assign (FXk,'A:Xk');      reset (FXk);
      assign (FWk,'A:Wk');      rewrite (FWk);

    read (Fdim,NT,NV);

    for k:=1 to NT do
      begin
        read (Fdim,NI); read (Fpoids,NIA);
        for i:=1 to NIA do read (Fpoids,POID[i]);
        for i:=1 to NI do
          for j:=1 to NV do read (FXk,X[i,j]);

        for i:=1 to NV do
          begin
            for j:=1 to NV do
              begin
                LigneV[j]:=0;
                for l:=1 to NIA do
                  LigneV[j]:=LigneV[j]+X[l,i]*X[l,j]*POID[l];
              end;
            for j:=1 to NV do write (FWk,LigneV[j],' ');
            writeln (FWk);
          end;
        writeln (FWk);
      end;

    close (Fdim); close (Fpoids); close (FXk); close (FWk);
    writeln ('fin de l''execution');
    writeln ('les matrices Vk se trouvent dans A:Wk');
  End.


Program Produits_scalaires_entre_Vk;
(*********************************)
   (*****************************************************************
    * calcul de la matrice S de produits scalaires entre objets Vk *
    *                    S[k1,k2]=Tr(Vk1Vk2)                        *
    *****************************************************************)
```

```
Const MaxVAR  = 25; MaxTAB  = 20;
Type  ligne   = array [1..MaxTAB] of real;
      tableau = array [1..MaxVAR,1..MaxVAR] of real;
Var   NT,NI,NV,i,j,k1,k2,l :integer;
      BIDON                :real;
      LigneS               :ligne;
      Vk1,Vk2              :tableau;
      Fdim,FWk,Fps         :text;

Begin
    assign (Fdim,'A:Dim'); reset (Fdim);
    assign (FWk,'A:Wk');   reset (FWk);
    assign (Fps,'A:PS');   rewrite (Fps);

  read (Fdim,NT,NV);

  for k1:=1 to NT do
    begin
      reset (FWk);
      if k1>1 then
       (* position du pointeur sur Vk1 dans le fichier FWk *
        ************************************************)
       begin
         for l:=1 to k1-1 do
           begin
             for i:=1 to NV do
               for j:=1 to NV do read (FWk,BIDON);
           end;
       end;
                   (* lecture de Vk1 *
                    ******************)
      for i:=1 to NV do
        for j:=1 to NV do read (FWk,Vk1[i,j]);
      reset (FWk);
      for k2:=1 to k1 do
        begin
                   (* lecture de Vk2 *
                    ******************)
          for i:=1 to NV do
            for j:=1 to NV do read (FWk,Vk2[i,j]);

          LigneS[k2]:=0;
          for i:=1 to NV do
            for j:=1 to NV do
              LigneS[k2]:=LigneS[k2]+Vk1[i,j]*Vk2[j,i];
        end;
      for k2:=1 to k1 do write (Fps,LigneS[k2],' ');
      writeln (Fps);
    end;

  close (Fdim); close (FWk); close (Fps);
  writeln ('fin de l''execution');
  writeln ('les produits scalaires entre Vk se trouvent dans A:ps');
End.
```

```
(************************************************************
 * le programme 'choix_produit_scalaire' permet d'obtenir les *
 * normes des Vk et le tableau des produits scalaires normes  *
 * entre Vk . Mais dans le cas d'objets Vk , il n'est pas     *
 * justifie de travailler avec les produits scalaires normes  *
 *                                                            *
 * On obtient l'image euclidienne des tableaux en utilisant   *
 * les programmes : 'lecture_poids_tableaux' ,                *
 *                  'sauvegarde_produits_scalaires_actifs' ,  *
 *                  'elements_propres_de_SD' ,                *
 *                  'coordonnees_tableaux' , et 'nuage'       *
 * L'image euclidienne centree s'obtient en utilisant le      *
 * programme 'centrage_produits_scalaires' .                  *
 ************************************************************)
```

7.3. COVARIANCES COMPROMIS ET TRAJECTOIRES

```
Program Coefficients_compromis;
(***************************)
   (************************************************************
    * calcul des coefficients du compromis a partir des elements *
    * propres du tableau non centre des produits scalaires entre *
    * objets Vk                                                  *
    ************************************************************)

  Const MaxTAB  = 20; NbAxes  =  3;
  Var  NTA,N,k,l         :integer;
       SOM               :real;
       Rep               :char;
       POID,NORME,COEFF  :array [1..MaxTAB] of real;
       ValProp           :array [1..NbAxes] of real;
       VectProp          :array [1..MaxTAB,1..NbAxes] of real;
       Fpoids,Fvecp,Fvalp,Fdim,Fcoef :text;

  Begin
      assign (Fpoids,'A:matDelta'); reset (Fpoids);
      assign (Fvecp,'A:vecp');      reset (Fvecp);
      assign (Fvalp,'A:valp');      reset (Fvalp);
      assign (Fdim,'A:dim');        reset (Fdim);
      assign (Fcoef,'A:coef');      rewrite (Fcoef);

    read (Fpoids,NTA); read (Fvalp,ValProp[1]);
    for k:=1 to NTA do read (Fpoids,POID[k]);
    for k:=1 to NTA do
      for l:=1 to NbAxes do read (Fvecp,VectProp[k,l]);

    SOM:=0;
    for k:=1 to NTA do
      begin
        COEFF[k]:=POID[k]*abs(VectProp[k,1]);
        SOM:=SOM+COEFF[k];
```

```
      end;

    writeln ('les coefficients du compromis sont : ');
    for k:=1 to NTA do
      begin
        COEFF[k]:=COEFF[k]/SOM;
        writeln ('                                ',COEFF[k]:6:3);
      end;

    read (Fdim,N,N); writeln (Fcoef,NTA,' ',N);
    for k:=1 to NTA do write (Fcoef,COEFF[k],' ');

    close (Fpoids); close (Fvecp); close (Fvalp); close (Fdim);
    close (Fcoef);
    writeln ('fin de l''execution');
    writeln ('les coefficients du compromis sont dans A:coef');
  End.

       (*****************************************
        * utiliser ensuite le programme 'compromis' *
        *****************************************)



Program Elements_propres_de_V;
(**************************)

  Const NbMax   = 50; NbAxes  = 3;
  Type ligne    = array [1..NbAxes] of real;
       colonne  = array [1..NbMax] of real;
       tableau  = array [1..NbMax,1..NbMax] of real;
       vecteurs = array [1..NbMax,1..NbAxes] of real;
  Var  N,i,j,l          :integer;
       Fichier          :string[20];
       ValProp          :ligne;
       Diag,SsDiag,POID :colonne;
       S                :tableau;
       VectProp         :vecteurs;
       Fpsa,Fvecp,Fvalp :text;

Procedure Tridiagonalisation (N:integer; Var S:tableau;
                              Var Diag,SsDiag:colonne);

Procedure Diagonalisation (N:integer; Var S:tableau;
                                      Var Diag,SsDiag:colonne);

Procedure Ordre (N:integer; Var Diag:colonne; Var S:tableau);

  Begin
      assign (Fpsa,'A:psa');   reset (Fpsa);
      assign (Fvecp,'A:vecp'); rewrite (Fvecp);
      assign (Fvalp,'A:valp'); rewrite (Fvalp);

    read (Fpsa,N);
    for i:=1 to N do
      for j:=1 to i do read (Fpsa,S[i,j]);
```

```
    Tridiagonalisation (N,S,Diag,SsDiag);
    Diagonalisation (N,S,Diag,SsDiag);
    Ordre (N,Diag,S);

    for l:=1 to NbAxes do write (Fvalp,Diag[l],' ');
    for i:=1 to N do
      begin
        for l:=1 to NbAxes do write (Fvecp,S[i,l],' ');
        writeln (Fvecp);
      end;

    close (Fpsa); close (Fvecp); close (Fvalp);
    writeln ('fin de l''execution');
    writeln ('les valeurs propres sont dans A:valp');
    writeln ('les vecteurs propres sont dans A:vecp');
  End.

   (*************************************************************
    * utiliser les programmes 'coordonnees_compromis' et 'nuage'  *
    * pour obtenir l'image euclidienne des correlations compromis *
    *************************************************************)



Program Coordonnees_trajectoires;
(*****************************)

  Const MaxVAR = 25; MaxTAB = 20; NbAxes = 3;
  Var   NT,NV,i,j,k,l:integer;
        LigneV       :array [1..MaxVAR] of real;
        ValProp      :array [1..NbAxes] of real;
        VectProp     :array [1..MaxVAR,1..NbAxes] of real;
        COORD        :array [1..MaxVAR,1..MaxTAB,1..NbAxes] of real;
        Fdim,FWk,Fvecp,Fvalp,Fcotraj :text;

  Begin
      assign (Fdim,'A:dim');       reset (Fdim);
      assign (FWk,'A:Wk');         reset (FWk);
      assign (Fvecp,'A:vecp');     reset (Fvecp);
      assign (Fvalp,'A:valp');     reset (Fvalp);
      assign (Fcotraj,'A:cotraj'); rewrite (Fcotraj);

    read (Fdim,NT,NV);
    for l:=1 to NbAxes do read (Fvalp,ValProp[l]);
    for i:=1 to NV do
      for l:=1 to NbAxes do read (Fvecp,VectProp[i,l]);

    for k:=1 to NT do
      begin
        for i:=1 to NV do
          begin
            for j:=1 to NV do read (FWk,LigneV[j]);
            for l:=1 to NbAxes do
              begin
                COORD[i,k,l]:=0;
```

```
            for j:=1 to NV do
              COORD[i,k,l]:=COORD[i,k,l]+LigneV[j]*VectProp[j,l];
            COORD[i,k,l]:=COORD[i,k,l]/sqrt(ValProp[l]);
          end;
      end;
  end;

for i:=1 to NV do
  begin
    for k:=1 to NT do
      begin
        for l:=1 to NbAxes do write (Fcotraj,COORD[i,k,l],' ');
        writeln (Fcotraj);
      end;
    writeln (Fcotraj);
  end;

close (Fdim); close (FWk); close (Fvecp); close (Fvalp);
close (Fcotraj);
writeln ('fin de l''execution');
write ('les coordonnees des trajectoires se trouvent dans ');
writeln ('A:cotraj');
End.

 (*******************************************************
  * pour tracer les trajectoires , utiliser le programme *
  *      'trace_trajectoires'                            *
  *******************************************************)
```

```
8. JEU TEST

   Donnees extraites des cantons ruraux :
5 cantons (Gignac,Ganges,Matelles,Lunel,Aniane en supplementaire) ,
3 variables (secteurs agricole,ouvrier,tertiaire) ,
3 recensements (68,75,82) .

51.88 32.55 15.57
 7.76 70.93 21.31
37.24 32.45 30.31
37.87 43.19 18.94
41.24 35.12 23.64

44.94 34.59 20.47
 6.22 65.96 27.82
16.09 31.22 52.69
34.20 43.32 22.48
21.08 42.14 36.78

25.95 39.15 34.90
 6.44 57.06 36.50
 6.54 24.68 68.78
16.13 42.18 41.69
15.03 35.96 49.01


Si on choisit l'objet Wk pour representer le tableau n° k :
**********************************************************

nombre de tableaux ? 3
nombre d'individus par tableau ? 5
les variables sont-elles les memes pour tous les tableaux ? (O/N) O
nombre de variables ? 3
les dimensions des tableaux sont dans A:dim : 3 5 3 3 3

nombre d'individus supplementaires parmi les 5 ? 1
les individus actifs ont-ils le meme poids ? (O/N) O
le nombre d'individus actifs et leurs poids sont dans A:MatD

nom du fichier des donnees ? A:jeu
voulez-vous reduire les variables ? (O/N) O
les moyennes des variables se trouvent dans A:moy

les matrices Wk se trouvent dans A:Wk

les produits scalaires entre Wk se trouvent dans A:ps :

 4.9048675911E+00
 4.5458133619E+00  4.5486596329E+00
 4.1202224090E+00  4.2506842159E+00  4.9074159542E+00

norme du tableau n° 1 =     2.21
                 n° 2 =     2.13
                 n° 3 =     2.22

voulez-vous travailler avec les coefficients RV ? (O/N) N
```

```
                matrice RV
 1.000
 0.962  1.000
 0.840  0.900  1.000

nombre de tableaux supplementaires parmi les 3 ? 0
les tableaux actifs ont-ils le meme poids ? (O/N) O
le nombre de tableaux actifs et leurs poids sont dans A:MatDelta

la partie active du tableau des produits scalaires entre Wk se
trouve dans A:psa

Decomposition en valeurs singulieres du tableau des produits
scalaires entre Wk :
fichier dans lequel se trouvent les poids des objets ? A:MatDelta

les valeurs propres sont dans A:valp :

 4.4664519898E+00  2.7660521093E-01  4.3923858532E-02

les vecteurs propres sont dans A:vecp :

-1.0134208543E+00  9.9651594373E-01  9.8991623177E-01
-9.9611857569E-01  3.5074442271E-01 -1.3728532817E+00
-9.9031608852E-01 -1.3725648705E+00  3.6788547265E-01


le nombre de tableaux , la dimension des Wk et les coefficients
du compromis sont dans A:coef :
3 4
 3.4966320927E-01  3.4369335949E-01  3.4169131239E-01

les coordonnees des tableaux se trouvent dans A:cotab :

-2.1417618184E+00  5.2410032656E-01  2.0746683782E-01
-2.1051952138E+00  1.8446796325E-01 -2.8772285978E-01
-2.0929322478E+00 -7.2187675609E-01  7.7101509443E-02


Centrage du tableau des produits scalaires entre Wk :
le tableau centre des produits scalaires se trouve dans A:ps :

 3.2364125552E-01
 3.9835743417E-02  1.1793073152E-01
-3.6347699895E-01 -1.5776647495E-01  5.2124347389E-01

la partie active de ce tableau se trouve dans A:psa

Decomposition en valeurs singulieres du tableau centre des produits
scalaires entre Wk :
fichier dans lequel se trouvent les poids des objets ? A:MatDelta

les valeurs propres sont dans A:valp :

 2.7690344630E-01  4.4035040672E-02 -4.4802425658E-12
```

```
les vecteurs propres sont dans A:vecp :

-1.0057864684E+00  9.9417985291E-01 -1.0000000000E+00
-3.5809177431E-01 -1.3681265589E+00 -1.0000000000E+00
 1.3638782428E+00  3.7394670602E-01 -9.9999999999E-01

Comme il n'y a que 2 valeurs propres non nulles , il faut limiter le
calcul a deux coordonnees dans le programme 'coordonnees_tableaux',
tout en conservant NbAxes = 3 pour lire correctement les vecteurs
propres .

les coordonnees des tableaux se trouvent dans A:cotab :

-5.2926109247E-01  2.0862394755E-01
-1.8843367815E-01 -2.8709489800E-01
 7.1769477067E-01  7.8470950458E-02

Calcul du compromis : la matrice compromis W se trouve dans A:psa

Decomposition en valeurs singulieres du compromis :
nom du fichier dans lequel se trouvent les poids des objets ? A:MatD

les valeurs propres sont dans A:valp :

 1.6078258125E+00  1.4832998056E+00  1.4018025342E-02

les vecteurs propres sont dans A:vecp :

-7.6041741057E-01  1.1846271705E+00 -1.0091699374E+00
 1.6619071676E+00 -1.6769353021E-01 -4.5819592656E-01
-8.0675614034E-01 -1.5212501959E+00 -1.8692885346E-01
-9.4733616643E-02  5.0431655562E-01  1.6542944521E+00

les coordonnees de l'image euclidienne compromis sont dans A:comp :

-9.6420982008E-01  1.4427668638E+00 -1.1948344202E-01
 2.1072994763E+00 -2.0423528573E-01 -5.4249363159E-02
-1.0229673625E+00 -1.8527427268E+00 -2.2131954189E-02
-1.2012229361E-01  6.1421114881E-01  1.9586472795E-01

correlations des variables avec les 3 axes:

tableau n°1
***********
-0.936  0.352 -0.003
 1.000 -0.015  0.001
-0.122 -0.992  0.009
tableau n°2
***********
-0.663  0.744  0.089
 0.997  0.079  0.013
-0.271 -0.955 -0.118
tableau n°3
***********
-0.491  0.864 -0.114
 0.895  0.447 -0.011
-0.462 -0.883  0.076
```

```
les coordonnees des trajectoires se trouvent dans A:cotraj :

-1.3440727281E+00  1.2255341692E+00 -1.1700444907E-01
-1.0370834749E+00  1.3800416865E+00  1.6994874700E+00
-4.0328310359E-01  1.5801733030E+00 -1.9393932976E+00

 2.5069917022E+00 -4.5245776880E-01  5.2363469724E-02
 1.9995704657E+00 -4.8212134363E-01 -5.3647107694E-01
 1.5904929435E+00  3.5024143047E-01  3.2726208875E-01

-9.3677722790E-01 -1.2361420700E+00  1.0756665646E-01
-7.6984088959E-01 -1.7633278813E+00 -2.2447486340E+00
-1.2608504034E+00 -2.3836287628E+00  2.0830531097E+00

-2.2614174610E-01  4.6306566954E-01 -4.2925676979E-02
-1.9264610116E-01  8.6540753842E-01  1.0817322414E+00
 7.3640563461E-02  4.5321402939E-01 -4.7092190120E-01

-8.6881711894E-01 -1.7135556392E-01  1.1542390894E-02
-4.4808511600E-02 -5.4094226155E-01 -6.8193823897E-01
-4.5011597355E-01 -2.2969538749E-01  5.2623305173E-02


Si on choisit l'objet Vk pour representer le tableau n° k :
**********************************************************

nombre de tableaux ? 3
nombre de variables par tableau ? 3
les individus sont-ils les memes pour tous les tableaux ? (O/N) O
nombre d'individus ? 5
les dimensions des tableaux sont dans A:dim : 3 3 5 5 5

les individus sont-ils les memes pour tous les tableaux ? (O/N) O
nombre d'individus supplementaires ? 1
les individus actifs ont-ils le meme poids ? (O/N) O
le nombre d'individus actifs et leurs poids sont dans A:MatD

 les matrices Vk se trouvent dans A:Wk :

 1.0000000000E+00 -9.4118182658E-01 -2.3452267313E-01
-9.4118182658E-01  1.0000000000E+00 -1.0774822772E-01
-2.3452267313E-01 -1.0774822772E-01  1.0000000001E+00

 1.0000000000E+00 -6.0071859199E-01 -5.4137711273E-01
-6.0071859199E-01  1.0000000000E+00 -3.4695505683E-01
-5.4137711273E-01 -3.4695505683E-01  1.0000000000E+00

 1.0000000000E+00 -5.2537880962E-02 -5.4464145251E-01
-5.2537880962E-02  9.9999999997E-01 -8.0889643117E-01
-5.4464145251E-01 -8.0889643117E-01  1.0000000000E+00

les produits scalaires entre Vk se trouvent dans A:ps :

 4.9048675911E+00
 4.4594688437E+00  4.5486596328E+00
 3.5286712500E+00  4.2141352123E+00  4.9074159541E+00
```

```
norme du tableau
                n° 1 =      2.21
                n° 2 =      2.13
                n° 3 =      2.22

               produits scalaires normes
 1.000
 0.944  1.000
 0.719  0.892  1.000

la partie active du tableau des produits scalaires entre Vk
se trouve dans A:psa

Decomposition en valeurs singulieres du tableau des produits
scalaires entre Vk :
fichier dans lequel se trouvent les poids des objets ? A:MatDelta

les valeurs propres sont dans A:valp :

 4.3087436501E+00  4.6381024728E-01  1.4427161902E-02

les vecteurs propres sont dans A:vecp :

-9.9884028924E-01  1.1333495949E+00  8.4725248439E-01
-1.0233103127E+00  1.3772948903E-01 -1.3906353195E+00
-9.7731984557E-01 -1.3025165195E+00  5.9016661707E-01


le nombre de tableaux , la dimension des Vk et les coefficients
du compromis sont dans A:coef :
3 3
 3.3300554438E-01  3.4116365892E-01  3.2583079669E-013

Calcul du compromis . La matrice compromis V se trouve dans A:psa :

 9.9999999999E-01
-5.3548057895E-01  9.9999999998E-01
-4.4025650545E-01 -4.1781258250E-01  1.0000000000E+00

Diagonalisation de V .
les valeurs propres sont dans A:valp :

1.5367825328E+00  1.3942028623E+00  6.9014604826E-02

les vecteurs propres sont dans A:vecp :

 7.3666619982E-01  3.1853706345E-01 -5.9653755058E-01
-6.7130622510E-01  4.5099009113E-01 -5.8818015084E-01
-8.1675346270E-02 -8.3375180774E-01 -5.4606507021E-01

les coordonnees de l'image euclidienne compromis sont dans A:comp :

 9.1322323273E-01  3.7611699448E-01 -1.5671417813E-01
-8.3219841115E-01  5.3251271855E-01 -1.5451863650E-01
-1.0125050365E-01 -9.8446385070E-01 -1.4345473911E-01
```

```
les coordonnees des trajectoires se trouvent dans A:cotraj :

 1.1193638368E+00  7.5889096260E-02  3.2397987857E-01
 9.5521322277E-01  4.2260198784E-01  1.9953967884E-01
 6.5857749783E-01  6.3428339603E-01 -1.0210114092E+00

-1.0937123003E+00  2.0412558426E-01  1.2221896873E-01
-8.7563411395E-01  4.6487995112E-01 -1.5366557454E-01
-5.1944633448E-01  9.3894630366E-01 -4.3824316945E-01

-1.4690052017E-01 -8.1053420989E-01 -1.3048339861E+00
-1.9971154479E-01 -9.8467949025E-01 -7.2480999219E-02
 4.8499115275E-02 -1.1619976105E+00  9.6918430932E-01
```

Notations

NOTATIONS INTRODUITES AU CHAPITRE II

R : corps des nombres réels.

E, F, G : espaces vectoriels réels de dimension finie. Les dimensions de ces espaces sont : dim(E) = n, dim(F) = m, dim(G) = p. Les lettres x, y, z désignent des vecteurs.

$\mathcal{L}(E, F)$: espace vectoriel des applications linéaires de E dans F. Les applications linéaires sont notées a, b, c.

$E^* = \mathcal{L}(E, R)$: espace vectoriel des formes linéaires sur E.

Pour $u \in E^*$: u(x) est noté $\langle x|u \rangle_{E^*}$ (ou $\langle x|u \rangle$).

$E^{**} = \mathcal{L}(E^*, R)$: espace vectoriel bidual de E.

H : espace affine, dont les points sont désignés par des lettres majuscules O, M, N ...

$(e_1,\ldots,e_n)$: base de E,

$(e_1^*,\ldots,e_n^*)$: base de E^*, duale de $(e_1,\ldots,e_n)$.

$\langle x|e_i^* \rangle$: coordonnée de x sur e_i,

$\langle e_i|u \rangle_{E^*}$: coordonnée de u sur e_i^*.

Dans (E, q) espace vectoriel euclidien :

$$\left.\begin{array}{l} q : E \times E \longrightarrow R \\ q : E \longrightarrow E^* \\ (x|y)_q = \langle x|q(y) \rangle_{E^*} \end{array}\right\} \text{produit scalaire sur E.}$$

$\|x\|_q = (x|x)_q^{1/2}$ q-norme de x.

$d_{xy} = \|x - y\|_q$ distance euclidienne déduite du produit scalaire q.

$(x|y)_q = 0 \iff$ x et y q-orthogonaux.

$F^\perp$: sous-espace de E, q-orthogonal au sous-espace F de E.

Dans (E^*, q^*) :

$$(u|v)_{q^*} = \langle q^{-1}(u)|v\rangle_{E^*} = \langle u|q^{-1}(v)\rangle_{E^{**}} = (q^{-1}(u)|q^{-1}(v))_q \ .$$

Si $(e_1,\ldots,e_n)$ est q-orthonormée, $(e_1^*,\ldots,e_n^*)$ est q^*-orthonormée. Alors $e_i^* = q(e_i)$ et $\langle x|e_i^*\rangle = (x|e_i)_q$.

Pour $a \in \mathcal{L}(E, F)$:

$$\text{Ker}(a) = \text{noyau de } a = \{x \in E \ / \ a(x) = 0\} ,$$
$$\text{Im}(a) = \text{image de } a = \{y \in F \ / \ \exists\, x \in E \ \ y = a(x)\} ,$$
$$\text{rg}(a) = \text{rang de } a = \dim(\text{Im}(a)), \text{ noté souvent } \rho .$$

$a' \in \mathcal{L}(F^*, E^*)$: transposé de a,
$\langle x|a'(v)\rangle_{E^*} = \langle a(x)|v\rangle_{F^*}$.

La lettre majuscule A désigne la matrice de a par rapport à une base $(e_1,\ldots,e_n)$ de E et une base $(f_1,\ldots,f_m)$ de F. Si i est l'indice ligne et j l'indice colonne, $A_{ij} = \langle a(e_j)|f_i^*\rangle$ est la coordonnée de $a(e_j)$ sur f_i. On note A' la matrice transposée de A.

x désigne le vecteur colonne, dont les éléments sont les coordonnées du vecteur x de E, dans la base $(e_1,\ldots,e_n)$.

$\mathcal{L}(E) = \mathcal{L}(E, E)$: ensemble des opérateurs de E.

Pour $a \in \mathcal{L}(E)$:

$$\text{Tr}(a) = \sum_i \langle a(e_i)|e_i^*\rangle ,$$

$$x \neq 0 , \quad R_a(x) = \frac{(a(x)|x)_q}{(x|x)_q}$$

Pour a diagonalisable, à valeurs propres réelles :

$\lambda_1(a) \geqslant \ldots \geqslant \lambda_n(a)$: valeurs propres,

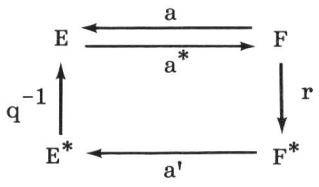
$E_k(a)$: sous-espace de E engendré par les vecteurs propres de a associés aux k plus grandes valeurs propres.

(E, q), (F, r) espaces vectoriels euclidiens.

Pour a et b $\in \mathcal{L}(E, F)$:

$a^* \in \mathcal{L}(E, F)$: adjoint de a,

$(x|a^*(y))_q = (a(x)|y)_r$.



$\sigma_1(a) \geqslant \ldots \geqslant \sigma_n(a)$: valeurs singulières de a,
$\sigma_i(a) = (\lambda_i(a^*a))^{1/2}$.

$(a|b)_t = Tr(a^*b)$: produit scalaire de Hilbert-Schmidt sur $\mathcal{L}(E, F)$.

NOTATIONS INTRODUITES AU CHAPITRE IV

$p_1, \ldots, p_n$: poids des individus. Produit scalaire associé d .

$\Pi_1, \ldots, \Pi_p$: poids des études. Produit scalaire associé δ .

X_k : tableau de données croisant individus (lignes) et variables (colonnes) centrées pour les poids des individus.

$(x_k)^i$: vecteur associé à l'individu i,

$(x_k)_j$: vecteur associé à la variable j.

W_k : tableau de produits scalaires entre individus,

V_k : matrice de covariances.

$$RV(k_1, k_2) = \left(\frac{w_{k_1}}{\| w_{k_1} \|_t} \,\middle|\, \frac{w_{k_2}}{\| w_{k_2} \|_t}\right)_t .$$

S : tableau de produits scalaires entre objets caractéristiques des études (w_k ou v_k, normés ou non).

Index alphabétique

actif (objet) 54

adjointe d'une application linéaire 30

auto-adjoint (opérateur) 31

base q-orthonormée 25

bidual 27

centré (tableau de produits scalaires) 62

cercle des corrélations 73

coefficient RV 84

composante principale 18, 66

compromis 18, 93

corrélations des variables avec les axes du compromis 18, 99

covariances compromis 107

décomposition en valeurs singulières 39

décomposition spectrale 33

distance euclidienne classique 67

dual 25

éléments propres 31

espace affine euclidien 28

espace vectoriel euclidien 23

espace vectoriel euclidien des individus 67

étude (X, q, d) 78

image euclidienne 49

image euclidienne centrée 63

individu moyen 101

inertie 58

interstructure 18, 81

intrastructure 17, 81

objet normé 90

objet v 71, 119

objet w 69, 111

opérateur 31

orthogonal (opérateur) 31

plan principal 69

produit scalaire 23

produit scalaire de Hilbert-Schmidt 38

produit scalaire normé, produit scalaire entre objets normés 84, 85

q-norme 24

q-orthogonalité 24

q-projecteur 31

quotient de Rayleigh 33

repère affine 28

semi-produit scalaire 43

supplémentaire (objet) 54

trace 37

trajectoires des individus 19, 99

trajectoires des variables 108

transposée 29

valeur propre 31

valeur singulière 39

vecteur propre 31

MASSON, Éditeur
120, bd Saint-Germain
75280 Paris Cedex 06
Dépôt légal : sept. 1988

Corlet, Imprimeur, S.A.
14110 Condé-sur-Noireau
N° d'Imprimeur : 4781
Dépôt légal : juillet 1988

Suite de la page 2 de couverture.

MÉTHODE+PROGRAMMES

Collection coordonnée par J.-F. Phélizon

STATISTIQUE ET ANALYSE DES DONNÉES

MÉTHODES PRATIQUES DE LA STATISTIQUE, par T. Foucart, A. Bensaber et R. Garnier.

RÉGRESSION LINÉAIRE SUR MICRO-ORDINATEURS, par T. Foucart et J.-Y. Lafaye.

ANALYSE FACTORIELLE. Programmation sur micro-ordinateur, par T. Foucart.

ANALYSE FACTORIELLE DE TABLEAUX MULTIPLES, par T. Foucart.

ALGORITHMES DE CLASSIFICATION, par M. Roux.

LES ARBRES ET LES REPRÉSENTATIONS DES PROXIMITÉS, par J.-P. Barthélémy et A. Guénoche.

ANALYSE CONJOINTE DES TABLEAUX QUANTITATIFS, par Ch. Lavit.

ANALYSE DE DONNÉES QUALITATIVES, par J. Herman, tome 1.-Traitement d'enquêtes, échantillon, répartitions, associations.

PRÉPARATION DES FICHIERS POUR LES ANALYSES MULTIVARIABLES. Sur micro-ordinateurs, par Ph. Laval.

GESTION

COMPTABILITÉ GÉNÉRALE AVEC dBASE III PLUS, par B.A. Oudet.

GESTION BUDGÉTAIRE ET MICRO-INFORMATIQUE, par J.-C. Courbon et G. Modoux.

GESTION STRATÉGIQUE DES COUTS ET DES PRIX DANS L'ENTREPRISE, par J.-F.Phélizon.

GESTION DU PERSONNEL SUR MICRO-ORDINATEUR, par B. Merck.

GESTION DE LA MASSE SALARIALE SUR MICRO-ORDINATEURS, par B. Merck.

GRAPHIQUES DE GESTION. Programmation en Basic sur micro-ordinateur et traceur, par D. Charpentier.

ANALYSE CONJOINTE DE TABLEAUX QUANTITATIFS

Ch. LAVIT

Cet ouvrage met à la disposition des utilisateurs de méthodes d'analyse de données une technique nouvelle — la méthode STATIS — permettant l'exploration simultanée de plusieurs tableaux. Les fondements théoriques sont exposés dans les chapitres II à IV ; on ne suppose connues que les notions classiques d'algèbre linéaire et les notions de moyenne, écart-type et coefficient de corrélation linéaire.

Les opérations matricielles nécessaires à la programmation sont décrites dans le chapitre V et complétées en annexe par les programmes correspondants écrits en Turbo Pascal.

Les trois derniers chapitres sont consacrés au traitement de jeux de données réels afin que l'utilisateur de méthodes statistiques puisse juger de l'intérêt de l'application de la méthode sur ses propres données sans qu'il lui soit nécessaire d'aborder la partie théorique.

Un logiciel est disponible sur micro-ordinateur, ainsi que sur gros système.

METHODE + PROGRAMMES

Derniers volumes parus :

THÉORIE DES POSSIBILITÉS. Application à la représentation des connaissances en informatique, par D. DUBOIS et H. PRADE.

LES ARBRES ET LES REPRÉSENTATIONS DES PROXIMITÉS, par J.-P. BARTHÉLEMY et A. GUÉNOCHE.

PROGRAMMATION LINÉAIRE SUR MICRO-ORDINATEUR. Programmes en Pascal, par B. LEMAIRE et C. LEMAIRE-MISONNE.

(Voir liste complète page 2)

Cette collection a pour objectif de présenter des applications de l'informatique, sous la forme de monographies assorties de programmes de calcul opérationnels. Elle est destinée à ceux qui désirent utiliser l'informatique pour résoudre un problème concret, en particulier dans un cadre professionnel.

Les sujets sélectionnés correspondent aux perspectives d'évolution de l'informatique au sens large. Ils répondent à une demande de plus en plus grande des milieux professionnels pour des ouvrages susceptibles de faciliter et d'accélérer l'analyse et la programmation des problèmes *informatisables*.

ISBN : 2-225-**81478**-3

© CORLET, IMPRIMEUR, S.A.