

ASU

BURO

Introduction
à
l'Analyse des Données

F. CAILLIEZ et J.P. PAGES

SMASH

F. CAILLIEZ

*Centre Technique
Forestier Tropical*

J.P. PAGES

*Commissariat à
l'Energie Atomique*

INTRODUCTION A L'ANALYSE DES DONNEES

Sous la direction de

G. MORLAT

*Professeur à l'Institut de Statistique
des Universités de Paris*

Avec des contributions de

J .C. AMIARD - J. ANDRES - M.F. BARA - J .M. BRAUN- J. BRENOT
P. CAZES- J. DEHEDIN- B. DIOP- Y. ESCOUFIER- C. GUEGUEN-
N. LACOURLY- J.P. MAILLES- B. MARCHADIER- M. PIET'RI- E. ROY-
G. SAPORITA- F. TESTU- R. THOMAS-

- 1976 -

SOCIETE de MATHEMATIQUES APPLIQUEES et de SCIENCES HUMAINES
9 rue Duban 75016 PARIS

Ce livre est patronné par

l'ASSOCIATION des STATISTICIENS UNIVERSITAIRES (A.S.U)

LE BUREAU UNIVERSITAIRE DE RECHERCHE OPERATIONNELLE (B.U.R.O.)

©SMASH 1976

Tous droits de traduction, de reproduction et d'adaptation adaptés pour tous pays.

Remerciements

Cette version électronique de l'ouvrage « Introduction à l'analyse de données » (J.P. PAGES, F. CAILLIEZ) dont la première édition date de 1976 a été rendue possible grâce :



A l'initiative du Centre International de Recherche sur l'Environnement et le Développement (CIRED) - (UMR 8568 du CNRS) qui a mené le projet et contribué à la relecture du document électronique



Au service de documentation de l'École Nationale des ponts et chaussée qui a réalisé la scannérisation de l'ouvrage



A des associés fondateurs de la SMASH qui ont assuré la mise en forme du document électronique et contribué à sa relecture

Avertissement

Cette première version électronique de l'ouvrage (2016) intègre les besoins de correction signalés mais non intégrés dans la réédition de 1983.

PREFACE

"Pendant longtemps, j'ai cru que j'étais un statisticien qui s'intéressait aux inférences allant du particulier au général. Mais, "attentif au développement de la statistique mathématique, j'ai trouvé des raisons d'étonnement et de doute".

*Ainsi s'exprimait, il y a une dizaine d'années, John W. TUKEY, dans les premières phrases d'un article prophétique publié dans les : **Annals of Mathematical Statistics**, sous un titre percutant: **The Future of Data Analysis**.*

Depuis cette époque, et tout particulièrement au cours des années les plus récentes, on a vu la gent statisticienne se scinder, grosso modo, en deux classes : la première catégorie est celle des statisticiens d'âge moyen qui ont appris et pratiqué la statistique mathématique classique, celle qui prétend formaliser l'induction, à la suite des statisticiens anglo-saxons, notamment, des années 1900 à 1950. La seconde classe est formée de gens en général plus jeunes, qui ont appris sous la même étiquette de "statistique" des techniques bien différentes, s'appuyant sur un outil mathématique purement algébrique, et visant à décrire, réduire, classer, des observations multidimensionnelles; ceux-là n'ont cure de l'induction et sont volontiers portés à proclamer que le statisticien doit mettre en œuvre ses techniques d'analyse sans faire aucune hypothèse sur les phénomènes observés. Ils pratiquent "l'analyse des données".

Au vrai, ce qui vient d'être dit ne décrit que fort incomplètement les grands courants de la pensée statisticienne. Il y a vingt ou trente ans, à l'époque où fut introduite en France, sous l'impulsion de Georges DARMOIS, la statistique mathématique, les statisticiens se divisaient déjà en deux classes, peut-être bien plus opposées encore dans leurs conceptions que celles d'aujourd'hui.

D'aucuns, que nous appellerons "les anciens", ne pouvaient accepter l'idée d'une inférence qui ne fût point basée sur des connaissances a priori exprimables sous forme de distributions de probabilité concernant les paramètres des lois des

phénomènes étudiés, et sur l'emploi de la formule de BAYES; les autres, que l'on considérait à l'époque comme modernes, mais que nous appellerons ici "les classiques", adoptaient les vues des statisticiens anglo-saxons (R.A. FISCHER, J. NEYMAN, A. WALD) et considéraient comme le premier devoir du statisticien d'établir des méthodes d'inférence qui ne laissent point de place aux idées a priori, sauf pour ce qui regarde la forme analytique des lois de probabilité utilisées (mais c'est peut-être là un a priori non négligeable puisque le statisticien classique se demande si cette forme analytique doit quelque chose aux observations qu'il utilise, témoin le problème des degrés de liberté dans le test d'ajustement du khi deux).

En fait, si les "anciens" réussissaient bien -au prix du choix de probabilités a priori quelque peu subjectif ou arbitraire, en tout cas prêtant à discussion- à formaliser le raisonnement inductif, les "classiques" traitaient un problème un peu différent, ainsi que J. NEYMAN a eu le mérite de le souligner: il s'agit pour eux de "comportement inductif", et l'on comprend bien comment leurs points de vue mènent à la théorie des décisions statistiques. Cela dit, à l'origine de leur débat, se situe une divergence concernant la signification concrète du concept de probabilité.

Pour les "anciens", la probabilité s'applique à tout ce qui est incertain, pour les "classiques" -et il est arrivé à quelques-uns d'entre eux de le proclamer de façon fort explicite- il s'agit d'une notion qui ne peut s'appliquer valablement qu'à des phénomènes susceptibles d'observations répétées, dans des conditions identiques (autant qu'il est possible) de sorte que la probabilité n'est autre chose que l'idéalisation de la notion de fréquence de tel évènement lié aux expériences répétées.

Si la race des "anciens" s'était éteinte ou était en voie de s'éteindre, notre tableau des querelles intestines de la gent statisticienne s'en trouverait simplifié. Mais certains développements théoriques sur l'induction et la décision, de même que de nombreuses possibilités d'application, s'en trouveraient appauvris. Car il s'est toujours trouvé, fort heureusement, même à l'époque de la plus grande gloire de la statistique mathématique classique, quelques statisticiens pour douter de la justesse des conceptions majoritaires. Parmi quelques autres -en petit nombre, il faut le reconnaître- Bruno DE FINETTI qui décrivait à peu près comme ceci l'évolution de la statistique entre 1930 et 1950 :

"On s'est aperçu à un certain moment que la

statistique mathématique était bâtie sur du sable (les probabilités a priori, si discutées), alors on a retiré le sable et la statistique a été bâtie sur rien !".

A partir des années cinquante, cette voix a commencé à trouver quelque écho même chez les statisticiens américains, et ce fut le développement de la statistique "néobayésienne" - qu'il faut bien regarder aussi comme une des branches de la statistique moderne-.

Revenons à l'analyse des données.

Contrairement à la statistique néobayésienne, qui vise à consolider la formalisation de l'induction, en consolidant "le sable" sur lequel s'appuyait la théorie statistique, on peut dire qu'on a retiré, en analyse des données, non seulement le sable, mais tout ce qui reposait dessus, c'est-à-dire le modèle probabiliste.

Dès lors, que reste-t-il de la statistique? Pas grand- chose -aux yeux des classiques ou des néobayésiens- sinon des procédés descriptifs. Et c'est bien de cela qu'il s'agit: l'analyse des données n'est pas autre chose que la forme moderne de la statistique descriptive. Mais, parce qu'on dispose de moyens de calcul un peu plus puissants qu'autrefois, il ne s'agit plus, comme autrefois, d'un chapitre préliminaire, mineur et ennuyeux de la statistique. On peut se livrer à une étude purement descriptive, mais puissante cependant, de tableaux de données multidimensionnelles, dans lesquels on ne pouvait étudier naguère que les marges ou les liaisons par couples. Et la description s'avère si efficace que d'aucuns n'hésitent guère à considérer que désormais l'essentiel de la statistique, voire la statistique toute entière, doit se réduire à l'analyse des données, la statistique mathématique classique n'étant qu'un ensemble de jeux de l'esprit, quelque peu arbitraires, destinés à meubler les loisirs des statisticiens qui à l'époque ne disposaient pas d'ordinateurs pour résoudre des problèmes plus réalistes.

C'est peut-être aller un peu loin. Dans l'article déjà cité, TUKEY écrivait:

"Dans la mesure où certaines parties de la statistique mathématique ne réussissent pas à contribuer, ou ne sont pas destinées à contribuer, même par un chemin long et

tortueux, à la pratique de l'analyse des données, elles doivent être regardées comme des éléments de mathématiques pures et critiquées selon les normes correspondantes. Les chapitres de la statistique mathématique doivent chercher leur justification soit dans l'analyse des données, soit dans les mathématiques pures. Un travail qui n'obéit à aucun de ces maîtres -et certains statisticiens prétendent qu'il en est ainsi pour leur propre travail- ne peut manquer d'être provisoire et voué à l'oubli. Et nous devons être attentifs à ce qu'il n'entraîne pas dans sa disparition des contributions d'intérêt plus durable".

Pouvait-on plus clairement accuser les statisticiens de byzantinisme? Personne ne songerait, certes, que ces accusations puissent s'appliquer aux techniques modernes d'analyse des données. Mais faut-il croire pour autant que la statistique mathématique classique est vouée à l'oubli ?

Il n'y a pas très longtemps que les statisticiens peuvent utiliser des ordinateurs. Il y a encore moins longtemps qu'on a vu se répandre en France les diverses méthodes d'analyse des données multidimensionnelles, principalement grâce à l'impulsion donnée par les travaux du laboratoire de statistique mathématique de J.P. BENZECRI et les services rendus à des utilisateurs les plus variés (comment s'en étonner? L'analyse des données doit rendre service partout où l'on se soucie d'accumuler des observations, et on doit accumuler des observations partout où l'on veut faire œuvre scientifique). Ces services rendus montrent bien que l'analyse des données constitue aujourd'hui, et de loin, la partie la plus immédiatement rentable de la statistique.

Tout se passe comme si l'absence de moyens de calcul avait longtemps barré une des voies d'exploration les plus fécondes du réel. Ce verrou vient de sauter d'une façon rapide et on est partout encouragé non seulement à exploiter globalement des données nombreuses, mais aussi à en recueillir de nouvelles, ce qui donne naissance à un torrent d'applications efficaces des techniques telles que l'analyse en composantes principales, l'analyse des correspondances, la classification automatique et d'autres encore. Cela permet selon les cas de découvrir dans les phénomènes étudiés des structures directement visibles sur les résultats de l'analyse, alors qu'elles ne l'étaient pas sur les données originelles, ou de retrouver en les précisant des structures que l'on soupçonnait déjà pour telle ou telle raison.

Mais après? Que va-t-il se passer quand on aura

mouliné à l'ordinateur quasiment toutes les données pertinentes à telle discipline?

L'application des techniques d'analyse des données met en évidence des traits parfois tellement saillants qu'on est tenté d'en rester là; et, en effet, le client venu avec des masses de chiffres repart content avec quelques graphiques et des idées pour un bout de temps. Mais peut-être un jour voudra-t-il prévoir les phénomènes qu'il étudie (et il lui faudra adopter un modèle probabiliste, suggéré sans doute par les résultats de l'étape descriptive), ou éclairer des décisions (et il se posera des problèmes du type de ceux qu'abordait la statistique mathématique classique).

On peut alors penser que les résultats classiques s'avèreront utiles à l'analyse des données -ou plutôt à des prolongements naturels- même si c'est "par un chemin long et tortueux", selon l'expression de TUKEY. Et l'on verra peut-être se développer davantage de recherches visant à proposer des modèles probabilistes précis dans des espaces multidimensionnels -un peu plus sophistiqués que la classique loi de LAPLACE GAUSS- et à permettre l'application des théories de la décision statistique, sur des exemples moins simples qu'un échantillon de variables normales réduites indépendantes.

En attendant, et pour quelques années encore, priorité à l'analyse des données.

G. MORLAT.

-1971 -

PRÉFACE A L'ÉDITION DE 1983

Cette Introduction à l'Analyse des Données date déjà de 1976 ; qu'en dire en 1982 ?

Nous ne remettons rien en cause au niveau du langage adopté, tant nous sommes persuadés que le parti pris pour la géométrie et l'exploitation systématique de la dualité était le bon choix. En s'efforçant par exemple, avec le schéma de dualité, de présenter de façon symétrique les individus et les variables, on insistait sur un fait fondamental que les factorialistes du début de ce siècle ne perdaient jamais de vue; en analyse multidimensionnelle, l'univers des variables joue un rôle au moins aussi important dans la stabilité d'un résultat, que l'univers des individus. Si on remet en cause le modèle probabiliste en analyse de données, on tombe donc sur le paradoxe suivant: d'un côté, l'individu n'est plus une simple unité statistique permettant de définir un continuum : la loi de probabilité; de l'autre, les variables, et ceux qui analysent les résultats d'enquête le savent bien, doivent être considérées comme extraites d'un univers dont il est difficile de distinguer, en général, les unités. Mais nous n'avons pas abordé, sauf en régression, ce grand problème de la stabilité et le langage choisi a permis surtout de bien montrer l'importance du choix des métriques et des critères d'optimalité en analyse factorielle.

Si nous défendons notre langage, nous sommes les premiers à souligner les insuffisances de l'ouvrage au niveau des méthodes décrites. Il s'agissait, non pas de développer, mais d'introduire de façon rigoureuse et synthétique les principales techniques de régression, d'analyse factorielle et de classification automatique. Malgré les remarques pratiques qui parsèment les différents chapitres, ces introductions se veulent avant tout mathématiques.

Nous pensons qu'en statistique la révolution informatique est loin encore d'être digérée; le grand ouvrage de l'analyse des données, c'est-à-dire de la statistique moderne, est encore à écrire. Si nous avons su échapper (pour un temps ?) au discours réducteur, et singulièrement peu adapté à l'approche multidimensionnelle, du statisticien mathématicien « classique », nous sommes encore loin d'une construction logique, structurée... et esthétique. Parmi les travaux qui nous semblent contribuer de façon significative à cette construction, nous plaçons d'abord ceux où l'on s'efforce d'exploiter la diversité à l'intérieur de ce que nous appelons toujours des « échantillons ». Cette diversité contient tous les ingrédients qui permettent d'envisager une statistique, qui ne ferait référence au modèle probabiliste a priori, que dans la mesure où il apporterait véritablement quelque chose, compte-tenu de ce que l'on a effectivement à faire.

F. CAILLIEZ & J.P. PAGES
Novembre 1982

Edition 1982

*AVIS AUX LECTEURS
Aucun changement n'a été
apporté au texte d'origine
(édition de 1976)*

AVANT PROPOS

Ce livre, qui fait suite à un cours édité en 1971 par le Centre d'Etudes Economiques d'Entreprises (C 3E), ne doit être considéré que comme une introduction à l'analyse des données multidimensionnelles.

En effet, si les bases et les principaux résultats utiles de l'analyse des données sont fournis, les propositions qui sont faites au niveau de la pratique ne prétendent pas à l'universalité. Cette pratique, en pleine évolution, ne saurait être réduite à l'application de quelques techniques standard sur des données toutes préparées; à l'heure actuelle, les tendances sont variées et dépendent du domaine d'application (économie, biologie, psychologie, marketing, ...); les tendances dominantes ne sont pas les mêmes en France et dans les pays anglo-saxons.

Les modifications apportées au cours édité au C 3E n'en ont pas changé l'esprit: nous cherchons à fournir des connaissances solides sur les méthodes de l'analyse des données, aussi bien à des non mathématiciens de formation auxquels les rappels algébriques (chapitres I à V) sont en premier destinés, qu'à des mathématiciens qui trouveront l'occasion de revoir de façon géométrique des connaissances algébriques fondamentales qui apparaîtront ici comme opérationnelles. Aussi ce livre pourra-t-il être utile aux praticiens de l'analyse des données recherchant des justifications théoriques et aux étudiants de 2ème ou de 3ème cycle des Universités: son contenu a fait l'objet d'enseignements dans le cadre de la formation permanente notamment au Bureau Universitaire de Recherche Opérationnelle, et dans le cadre universitaire, en particulier à l'Université PARIS VI (cycle supérieur de l'Institut de Statistique des Universités de Paris et 3ème cycle de Statistique Mathématique). Il s'inspire, au niveau de la forme et du langage, à la fois de T.W. ANDERSON et de J.P. BENZECRI.

Un choix a été fait dans la façon de présenter les techniques d'analyse des données : nous utilisons constamment les notions de projecteur, d'application M -symétrique, de codage, et surtout le "schéma de dualité" qui est un instrument de langage efficace permettant de présenter d'un seul jet et sous toutes leurs facettes les techniques relevant de l'algèbre linéaire. P. CAZES et J.P. MAILLES ont largement contribué à ce choix.

Le chapitre 6 est consacré à un rappel de statistique descriptive; on y souligne l'importance de la notion de distance qui assure la cohérence dans le choix de caractéristiques de valeur centrale et de dispersion. Toutes les notions clé permettant d'introduire et de justifier les techniques d'analyse linéaire sont développées au chapitre 7 : notions d'individu et de caractère, codage, schéma de dualité. Au chapitre 8 sont présentées, simultanément grâce à la dualité, l'analyse en composantes principales et l'analyse factorielle sur tableau de distances; on trouve en fin de chapitre quelques théorèmes qui jouent un rôle important en analyse factorielle classique (Spearman, Thurstone, Harman,...) et en régression (théorème de Gauss-Markov). Les exemples développés au chapitre 9 ont pour objectif d'illustrer dans le détail la pratique de l'analyse en composantes principales:

les documents de base que doit fournir tout bon programme sont présentés et commentés. Les autres techniques d'analyse factorielle, qui peuvent d'ailleurs être présentées comme des analyses en composantes principales particulières, conduisent à des documents semblables. Le chapitre 10 qui traite de la régression est particulier dans la mesure où la discussion sur le choix de la métrique, obligeant à s'interroger sur les objectifs d'un "modèle", conduit à osciller entre l'optique "analyse des données" et l'optique "interférentielle"; on trouvera dans ce chapitre une présentation entièrement géométrique du théorème de Gauss-Markov qui joue un rôle fondamental dans l'étude du modèle général linéaire. L'analyse canonique fait l'objet du chapitre 11; elle est assimilée à l'étude des positions relatives de deux sous espaces vectoriels; trois techniques usuelles relèvent de l'analyse canonique: l'analyse canonique classique que l'on utilise parfois pour dresser un bilan des corrélations entre deux paquets de caractères quantitatifs (chapitre 11), l'analyse factorielle discriminante (chapitre 12) qui permet d'étudier les liaisons entre une variable qualitative et un paquet de variables quantitatives et l'analyse factorielle des correspondances, introduite au chapitre 13, pour décrire la liaison entre deux variables qualitatives. A la fin des chapitres 11,12 et 13 on donnera sous forme d'exercices une manière d'étendre l'analyse canonique à l'étude des positions relatives de plus de deux sous espaces vectoriels. Au début du chapitre 14 on dresse un bilan des techniques introduites dans les chapitres précédents; on souligne en particulier que la stratégie à adopter pour l'étude de variables quantitatives est non seulement fonction des objectifs mais aussi du nombre d'observations. Puis on introduit une méthode, basée sur les opérateurs développés par Y. ESCOUFIER, permettant de décrire les proximités entre tableaux individus x caractères ou entre paquets de variables. Le dernier chapitre est consacré à la classification automatique: une grande place est accordée à la discussion sur les indices (proximités entre éléments, entre parties, entre relations binaires); les principales techniques permettant d'obtenir un arbre de classification ou une partition sont passées en revue; à la fin du chapitre on discute des méthodes de segmentation.

Si nous avons exploité en partie les idées développées récemment par Y.ESCOUFIER de l'Université de Montpellier pour l'analyse des tableaux à plusieurs dimensions, nous n'avons pas présenté certains travaux français récents effectués sur l'analyse des données et qui peuvent être considérés comme des aménagements ou des généralisations des techniques décrites; citons parmi ces travaux, outre ceux effectués au sein du Laboratoire de Statistique Mathématique de J.P. BENZECRI, ceux développés à l'Université Paul SABATIER de Toulouse par J.DAUXOIS et A.POUSSE, ceux de G.SAPORTA et de M.MASSON à l'Université de Paris VI, ceux de E.DIDAY et de son équipe à l'IRIA, ceux de J .M. BOUROCHE et de ses collaborateurs, etc ...

F. CAILLIEZ J.P. PAGES

Novembre 1975

TABLE DES MATIERES

CHAPITRE I ENSEMBLES - APPLICATIONS - RELATIONS BINAIRES

0	Introduction	1
1	Ensembles - Applications	3
11	<i>Quelques généralités sur les ensembles</i>	3
121	<i>Inclusion - Ensemble des parties</i>	5
122	<i>Définition - Exemples</i>	5
13	<i>Ensemble des parties d'un ensemble</i>	6
131	<i>Opérations élémentaires $n \cup \cdot$</i>	6
132	<i>Intersection</i>	6
133	<i>Réunion</i>	8
134	<i>Complémentaire</i>	9
14	<i>Produit cartésien</i>	11
15	<i>Applications</i>	12
151	<i>Définition</i>	12
152	<i>Exemples</i>	12
153	<i>Fonction caractéristique</i>	14
154	<i>Applications surjectives, injectives et bijective</i>	15
155	<i>Application inverse</i>	16
156	<i>Composition d'application</i>	17
	<i>Bilan des connaissances- Notations</i>	18
2	Relations binaires	20
21	<i>Exemples</i>	20
22	<i>Définition</i>	21
23	<i>Etude des relations binaires sur un ensemble</i>	22
231	<i>Représentation des relations binaires</i>	22
232	<i>Quelques types de relations binaires</i>	22
233	<i>Quelques compléments sur les relations d'ordre</i>	24
233.1	<i>Vocabulaire</i>	24
233.2	<i>Treillis</i>	25
233.21	<i>Exemples</i>	25
233.22	<i>Définition</i>	27
233.23	<i>Ensembles flous</i>	27
234	<i>Fermeture transitive d'une relation binaire</i>	29
234.1	<i>Exemple</i>	29
234.2	<i>Définitions</i>	30
235	<i>Etude des relations d'équivalence</i>	32
235.1	<i>Relation d'équivalence et partition</i>	32
235.2	<i>Treillis des partitions d'un ensemble</i>	32
236	<i>Etude des préordres</i>	33
236.1	<i>Un préordre est un ordre sur une partition</i>	33
236.2	<i>Treillis des préordres d'un ensemble</i>	34
237	<i>Image réciproque d'une relation binaire</i>	34
	<i>Bilan</i>	36

CHAPITRE II ESPACES VECTORIELS

0	Avertissement.....	37
1	Trois exemples.....	38
11	<i>Ensemble des points du plan</i>	38
12	<i>Un ensemble d'applications</i>	40
13	<i>Les caractères quantitatifs</i>	41
2	Définition générale d'un vectoriel (espace vectoriel).....	42
3	Combinaison linéaire - Base - Indépendance linéaire.....	43
31	<i>Un exemple dans le plan P</i>	43
32	<i>Définition et vocabulaire : notations</i>	44
4	Exemples.....	46
41	<i>Le plan P ou R</i>	46
42	<i>Représentation d'un n-échantillon p dimensionnel</i>	47
43	<i>Les engrais ternaires</i>	48
44	<i>Le portefeuille de valeurs mobilières</i>	49
45	<i>Quelques réflexions sur le panier de la ménagère</i>	50
5	Extension de la notion de base : décomposition en somme directe.....	52
51	<i>Sous-espace vectoriel</i>	52
52	<i>Exemples de sous-espaces vectoriels</i>	54
53	<i>Décomposition en somme directe</i>	55
54	<i>Exemple</i>	55
55	<i>Quelques propriétés</i>	56
Bilan	57

CHAPITRE III APPLICATIONS LINEAIRES - MATRICES

1	Applications linéaires d'un vectoriel dans un autre.....	59
11	<i>Exemple : symétrie dans le plan par rapport à une droite passant par l'origine</i>	59
12	<i>Exemple : projection orthogonale dans le plan sur une droite passant par l'origine</i>	60
13	<i>Définition d'une application linéaire</i>	61
14	<i>L'ensemble des applications linéaires d'un vectoriel dans un autre</i>	61
141	<i>C'est un vectoriel</i>	61
142	<i>Cas particulier : forme linéaire - dual</i>	61
142.1	<i>Base duale</i>	61
142.2	<i>Bidual</i>	63
142.3	<i>Notation</i>	63
15	<i>Transposée d'une application linéaire</i>	64
16	<i>Rang d'une application linéaire</i>	65
2	Introduction au langage matriciel.....	66
21	<i>Reprenons l'exemple 1</i>	66
22	<i>Reprenons l'exemple 2</i>	68
23	<i>Autre exemple : matrices unités</i>	70

24	Définition générale d'une matrice.....	71
25	Exemple (récapitulatif) : les sacs d'engrais	72
3	Opérations sur les matrices (calcul matriciel).....	75
31	Addition de deux matrices.....	75
32	Produit d'une matrice par un nombre.....	75
33	Produit de deux matrices.....	76
34	Inverse d'une matrice - noyau d'une application linéaire.....	78
341	Reprenons l'exemple 1.....	78
342	Reprenons l'exemple 2	79
343	Reprenons les matrices unités	81
344	Applications linéaires inverses; matrices inverses	81
344.1	Noyau	82
344.2	Applications linéaires injectives	82
344.3	Applications linéaires bijectives : applications et matrices inverses.....	83
345	Résolution d'un système linéaire	84
346	Inverses généralisées.....	85
35	Tableau récapitulatif des opérations sur les matrices.....	87
36	Matrice transposée	88
4	Quelques matrices particulières	89
5	Quelques propriétés des opérations sur les matrices - Bilan.....	94
	Un survol.....	94
6	Changement de base dans un vectoriel.....	95
7	Influence d'un changement de base (à la fois dans l'espace de départ et dans l'espace d'arrivée) sur l'expression de la matrice associée à une application linéaire	97
8	Valeurs et vecteurs propres	100
81	Exemples.....	100
82	Définition générale des valeurs et vecteurs propres	107
83	Quelques propriétés	108
831	Nombre de valeurs propres associées à une matrice carrée	108
832	Indépendance linéaire des vecteurs propres	110
84	Diagonalisation d'une matrice carrée.....	112
	Bilan	113

CHAPITRE IV DISTANCES ET DISTANCES EUCLIDIENNES

0	Introduction	115
1	Rappel sur les distances (ou métriques).....	116
11	Un exemple : les distances euclidiennes dans \mathbb{R}^2	116
12	Définition d'une distance sur un ensemble.....	118
13	Trois distances utiles.....	119
131	Deux distances non euclidiennes dans \mathbb{R}^n	119
132	Les distances ultramétriques	120
2	Espaces euclidiens	122

21	<i>Le produit scalaire classique</i>	122
211	<i>Définition</i>	123
212	<i>Quelques exemples</i>	123
213	<i>Propriétés du produit scalaire classique</i>	124
214	<i>Norme et distance euclidiennes classiques dans \mathbb{R}^n</i>	124
215	<i>Utilité du produit scalaire classique</i>	125
22	<i>Les espaces euclidiens</i>	127
221	<i>Définition générale d'un espace euclidien</i>	127
222	<i>Aperçu matriciel</i>	128
223	<i>Isomorphisme associé à une métrique euclidienne</i>	131
224	<i>Bilan et notations</i>	133
23	<i>Notion de M-angle</i>	134
231	<i>M-projection sur un axe</i>	134
232	<i>Inégalité de Schwarz</i>	134
233	<i>M-angle : définition</i>	135
24	<i>Isométries</i>	136
25	<i>Exercice</i>	138
	<i>Bilan</i>	140

CHAPITRE V PROJECTEURS ASSOCIES A UNE DECOMPOSITION EN SOMME DIRECTE

- . Applications linéaires idempotentes
- . Applications linéaires M-symétriques

1	Projecteurs associés à une décomposition en somme directe - applications idempotentes	141
11	<i>Exemple dans \mathbb{R}^3</i>	141
12	<i>Cas général= projecteurs associés à une décomposition en somme directe</i>	142
121	<i>Définition</i>	143
122	<i>Propriétés</i>	143
123	<i>Exercice</i>	144
13	<i>Applications idempotentes</i>	145
2	Décomposition en somme directe en sous-espaces M-orthogonaux	147
21	<i>Exemple dans $E=\mathbb{R}^3$ muni de la distance euclidienne classique ($M=I_3$)</i>	147
22	<i>Sous-espaces vectoriels M-orthogonaux</i>	147
23	<i>Sous-espaces vectoriels supplémentaires M-orthogonaux</i>	148
24	<i>Décomposition en somme directe en sous-espaces M-orthogonaux</i>	149
3	Projecteurs associés à une décomposition de E en deux sous-espaces supplémentaires M-orthogonaux = applications linéaires idempotentes M-symétriques	150
4	Quelques propriétés des applications linéaires M-symétriques	152
	<i>Du vocabulaire</i>	154

CHAPITRE VI INTRODUCTION A LA DESCRIPTION STATISTIQUE

1	Echantillon d'une variable	155
---	---	-----

11	Deux représentations de E	156
12	Trois caractéristiques de "valeur centrale".....	156
121	Moyenne.....	156
122	Médiane.....	157
123	Moyenne des valeurs extrêmes.....	157
13	Trois caractéristiques de dispersion.....	158
131	Variance et écart-type.....	158
132	Ecart-moyen.....	158
133	Etendue.....	158
14	Caractéristiques de valeur centrale et de dispersion associées au choix d'une distance dans R^n	159
141	R^n est muni de la distance d_1	159
142	R^n est muni de la distance d_2	160
143	R^n est muni de la distance d_3	160
144	Généralisation.....	161
144.1	Choix d'autres distances dans R^n	161
144.2	Interprétation géométrique de la moyenne et de la variance dans R^n	161
15	Diagrammes - Histogrammes.....	162
151	La variable x est entière.....	162
151.1	Diagramme en bâtons.....	162
151.2	Mode.....	163
152	La variable x est quelconque= histogramme.....	163
153	Quelques types d'histogramme.....	165
16	Comment décrire au mieux un échantillon?.....	167
2	Echantillon d'un couple de variables.....	168
21	Représentation de l'échantillon dans R^2	168
22	Covariance et coefficient de corrélation linéaire.....	169
221	Définitions.....	169
222	Représentation de l'échantillon dans R^n - Signification géométrique du coefficient de corrélation.....	171
223	Interprétation des coefficients de corrélation.....	173
3	Echantillon de p variables.....	180
31	Définition et représentation.....	180
32	Matrices de variance et de corrélation - Tableau de description centré.....	181
33	Comment décrire un tableau individus \times caractères.....	183
Bilan	184

CHAPITRE VII INTRODUCTION A L'ANALYSE MULTIDIMENSIONNELLE LINEAIRE

0	Introduction.....	185
1	Individus - Caractères - Codage.....	185
11	Exemple.....	185
12	Individus - Caractères.....	187
121	Caractères.....	187
121.1	Qualitatif - Quantitatif.....	188
121.2	Codage.....	189
122	Individus.....	193
2	Notations et dualité en analyse linéaire.....	194

21	Notations	195
211	Représentation des caractères dans E et E^* et des individus dans F	197
212	Les applications X et X'	198
22	Proximités	199
221	Rappel= espaces euclidiens	200
222	Métrique euclidienne dans l'espace des individus E	202
223	Métrique euclidienne dans l'espace des caractères F	205
23	Bilan général : le schéma de « dualité »	206
24	Métrique des poids dans l'espace F des caractères ; centre de gravité ; forme quadratique d'inertie	207
241	Inertie en un point - Centre de gravité	207
242	Métrique des poids - Forme quadratique d'inertie	208
3	Panorama des techniques d'analyse multidimensionnelle linéaire	210
31	Relation de finesse entre caractères quantitatifs = le rapport de corrélacion - Equivalence entre caractères quantitatifs	211
311	Définition	211
312	Comment juger si un caractère quantitatif est plus fin qu'un autre caractère quantitatif : le rapport de corrélacion	213
313	Comment juger si deux caractères quantitatifs sont équivalents ?	214
314	Equivalence entre caractères quantitatifs en analyse linéaire	215
32	Equivalence entre individus	216
33	Comment reconnaître des caractères ?	217
34	Panorama des techniques d'analyse multidimensionnelle linéaire	218
Bilan	220

CHAPITRE VIII ANALYSE EN COMPOSANTES PRINCIPALES ET ANALYSE FACTORIELLE SUR TABLEAU DE DISTANCES

1	Introduction - Notations	221
11	Schéma de dualité - Notations	223
12	La tactique en analyse en composantes principales	224
2	Inertie et moments d'inertie	226
21	Inertie en un point (rappel)	226
22	Moment d'inertie par rapport à un sous-espace	228
221	Moment d'inertie par rapport à un sous-espace vectoriel	228
222	Moment d'inertie par rapport à un sous-espace affine: théorème de Huygens ..	229
23	Conséquence du théorème de Pythagore	231
24	Moment d'inertie par rapport à une droite $\Delta_{\underline{u}}$ et par rapport à l'hyperplan M -orthogonal $\Delta_{\underline{u}^\perp}$	231
25	Conséquences de l'inégalité de Schwarz	237
3	Etude des vecteurs et valeurs propres de VoM	238
4	Axes et plans principaux	239
41	Premier axe principal	239
42	Plan principal	241
43	Sous-espace principal de dimension k	243
44	Bilan et dualité	244
5	Description du nuage des individus	245

6	Description des caractères	249
7	Analyse factorielle sur tableau de distances	251
71	<i>Deux théorèmes</i>	251
711	<i>Unicité</i>	251
712	<i>Existence</i>	252
72	Analyse factorielle sur tableau de distances	254
721	<i>W est semi-définie positive</i>	254
722	<i>Cas général: W n'est pas semi-définie positive</i>	255
723	<i>Si d est ultramétrique il existe une image euclidienne de (l,d)</i>	260
8	Deux métriques particulières	261
81	<i>Exemple</i>	261
82	<i>Représentation des caractères si $M=D 1/\sigma^2$</i>	264
9	Quelques cas d'interprétation	266
91	<i>"Taille" et "Forme"</i>	266
92	<i>Un caractère indépendant</i>	267
93	<i>Un exemple simplet</i>	268
10	Quelques théorèmes annexes	270
101	<i>Vers l'analyse factorielle au sens des psychologues</i>	270
102	<i>Théorème réciproque</i>	272
103	<i>Axes de M-symétrie</i>	274
Bilan	275

CHAPITRE IX DEUX EXEMPLES D'ANALYSE EN COMPOSANTES PRINCIPALES

1	Les poissons d'Amiard	277
11	<i>Objectifs</i>	277
12	<i>Expérience</i>	277
13	<i>Résultats et commentaires</i>	279
131	<i>Caractéristiques des variables</i>	279
132	<i>Histogrammes</i>	279
133	<i>Matrice de corrélation R</i>	285
134	<i>Analyse en composantes principales</i>	286
134.1	<i>Un coup d'œil sur les corrélations</i>	288
134.2	<i>Etude de la matrice de corrélation: vecteurs et valeurs propres</i>	288
134.3	<i>Corrélations entre les 16 caractères et les 4 premières composantes principales</i>	289
134.4	<i>Cosinus carrés des angles entre les observations et leurs projections dans les différents sous-espaces</i>	290
134.5	<i>Représentation des individus dans le plan principal</i>	290
134.6	<i>Représentation des caractères par leurs corrélations avec les deux premières composantes principales</i>	292
134.7	<i>Conclusion</i>	293
2	La sociomatrice de Thomas	294
21	<i>Examen rapide des données</i>	296
211	<i>Histogramme des 24 x (24-1) notes attribuées</i>	296

212	Statuts sociométriques	297
22	Description de la sociomatrice par l'analyse en composantes principales.....	298
221	Première analyse.....	298
222	Deuxième analyse.....	301
23	Conclusion.....	302

CHAPITRE X REGRESSION LINEAIRE


1	Quelques exemples	303
11	Exemple : enquête alimentaire	303
12	Exemple : analyse de la variance - Plan factoriel à deux facteurs sans interaction et sans répétitions	305
13	Exemple 1 : Enquête alimentaire (suite) - Analyse de covariance.....	306
14	Exemple 2 : Ajustement polynomial.....	308
15	Exemple 3 : Test d'intelligence et réussite scolaire.....	309
16	Bilan	310
2	Introduction mathématique ; notations.....	311
3	Solution.....	314
31	Le point \underline{y}^{\wedge}	314
32	Une façon de mesurer la qualité de la représentation de y par \underline{y}^{\wedge}	315
33	Recherche du projecteur A : solution \underline{b}	315
34	Changement d'échelle.....	317
35	Autre expression de la solution \underline{b} : optique de régression entre variables aléatoires (exemple 5).....	318
4	Choix de la métrique N	320
41	Inégalités entre formes quadratiques semi-définies positives	323
42	Théorème de Gauss-Markov	325
421	Conséquences si X' est injective	327
422	Conséquence dans le cas où X' est quelconque	327
43	Conclusion : choix de la métrique dans le cadre du modèle général linéaire	328
5	Discussion des résultats.....	330
51	Proximité globale de \underline{y} et \underline{y}^{\wedge}	330
511	Cas général	330
512	Cas de la métrique $N = D_p$; coefficient de corrélation multiple	331
52	Examen détaillé de la proximité entre y et \underline{y} : résidus	332
53	Discussion des hypothèses dans le cadre de modèle général linéaire: le rapport F	335
531	Introduction à la pratique du test en régression.....	336
532	Application : simplification du modèle	339
54	Régression entre variables aléatoires et modèle conditionnel.....	341
55	Introduction à la régression pas à pas	343
551	Elimination descendante des variables.....	344
552	Introduction ascendante des variables.....	345
553	Exemple.....	345
6	Cas particulier : droite de régression	346
	Bilan	349

CHAPITRE XI ANALYSE CANONIQUE

1	Introduction - Notations	351
11	<i>Equivalence entre deux paquets de caractères quantitatifs</i>	351
12	<i>Schéma de dualité ; notations</i>	352
2	Exemples	355
21	<i>La pollution des rivières</i>	355
22	<i>Tests psychotechniques et réussite scolaire</i>	358
3	Buts et étapes mathématiques de l'analyse canonique	358
31	<i>But</i>	358
32	<i>Un exemple</i>	360
33	<i>Étapes mathématiques</i>	361
4	Recherche des vecteurs η^i et ξ^i	364
41	<i>Étape 1 : recherche de η^1 et ξ^1</i>	366
411	<i>Conséquences de l'inégalité de Schwarz</i>	367
412	<i>Étude des valeurs et vecteurs propres de AoA</i>	369
413	<i>Recherche du vecteur ξ^1</i>	371
414	<i>Recherche du vecteur $\xi^j \in W_2$</i>	373
42	<i>Étape 2 : recherche de ξ^2 et η^2</i>	375
43	<i>Étape k : recherche de ξ^k et η^k</i>	377
5	Caractères, facteurs et axes canoniques	378
6	Sorties graphiques	381
61	<i>Description de l'ensemble des caractères</i>	381
611	<i>Dans W_1</i>	381
612	<i>Dans W_2</i>	383
62	<i>Description de l'ensemble des individus</i>	384
7	Cas particulier de l'analyse canonique classique	386
71	<i>Caractères et facteurs canoniques</i>	388
72	<i>Coefficients de corrélation canoniques</i>	388
73	<i>Remarques sur les sorties graphiques</i>	389
8	Un exercice : vers l'analyse canonique généralisée	390
	<i>Les points importants</i>	392

CHAPITRE XII ANALYSE FACTORIELLE DISCRIMINANTE

1	Exemples	393
11	<i>Exemple 1 : origine sociale et réussite scolaire</i>	393
12	<i>Exemple 2 : amélioration d'un diagnostic</i>	394
13	<i>Exemple 3 : analyse de la variance p dimensionnelle sur plan facto</i>	394
2	Notations	396
21	<i>Ensemble des caractères quantitatifs</i>	396

22	<i>La variable qualitative y</i>	397
23	<i>Partition de ...</i> 	397
231	<i>Centres de gravité "intra-classe"</i>	398
232	<i>Forme quadratique d'inertie "intra-classe"</i>	398
233	<i>Forme quadratique d'inertie "inter-classe"</i>	399
234	<i>Relation entre les formes quadratiques intra et inter-classe</i>	399
3	Potentiels de prévision	400
31	<i>Potentiel de prévision associé à l'ensemble des caractères quantitatifs x_1, x_2, \dots, x_p</i>	400
32	<i>Potentiel de prévision associé à la variable qualitative y</i>	400
4	Analyse factorielle discriminante vue comme un cas particulier de l'analyse canonique	404
41	<i>Pouvoir discriminant des caractères éléments de $W_1 n W_2$: exemple</i>	405
42	<i>Recherche des caractères canoniques $\xi^i \in W_1$ et des facteurs canoniques $a_i \in E_1^*$: caractères et facteurs discriminants</i>	408
421	<i>Expressions des projecteurs</i>	408
422	<i>Caractères et facteurs discriminants</i>	409
5	En quoi l'analyse discriminante est-elle une analyse en composantes principales particulière	412
6	Sorties graphiques	415
61	<i>Description dans $E_1 = \mathbb{R}^p$</i>	415
62	<i>Description dans $F = \mathbb{R}^n$</i>	417
7	Un exercice : de nouveau vers l'analyse canonique généralisée	418
	<i>Bilan - Notations</i>	419

CHAPITRE XIII ANALYSE FACTORIELLE DES CORRESPONDANCES

1	Exemples	423
11	<i>Elections présidentielles</i>	423
12	<i>Budgets x pays</i>	424
2	Mesure et mesure de probabilité sur un ensemble fini I	425
21	<i>Exemple</i>	425
22	<i>Mesure sur $(I, P(I))$</i>	426
23	<i>Mesure de probabilité sur $(I, P(I))$</i>	426
24	<i>Application: mesures, lois de probabilité et variables aléatoires en analyse factorielle des correspondances</i>	427
25	<i>Bilan des notations</i>	430
3	Application des résultats de l'analyse canonique	431
31	<i>Schéma de dualité et potentiels de prévision des variables qualitatives x et y</i>	432
32	<i>Analyse canonique sur variables indicatrices</i>	434
33	<i>Précisions sur la dimension de $W_1 n W_2$</i>	436
331	<i>Caractères et facteurs centrés</i>	436
332	<i>Dimension de $W_1 n W_2$</i>	438

332.1	Partition du tableau P en "rectangles maximaux"	438
332.2	Application : dimension de W_1nW_2	440
	Bilan	441
34	Vers une interprétation « probabiliste » à l'aide de l'analyse en composantes principales.....	442
4	Simplexe des lois de probabilité et distance du x^2	444
41	Représentation vectorielle des mesures et mesures de probabilité sur un ensemble fini	444
42	Distance du x^2 entre lois de probabilité	446
43	Application : « x^2 » de contingence	446
431	\emptyset^2 et x^2	449
432	Décomposition du x^2 de contingence	451
5	Interprétation à l'aide de l'analyse en composantes principales	452
51	Analyses en composantes principales sur les nuages des lois conditionnelles ..	453
52	Composantes principales et facteurs principaux	455
6	Sorties graphiques : représentation simultanée des ensembles I et J	456
61	Optique "analyse canonique"	456
62	Optique "analyse en composantes principales"	457
63	Bilan	458
7	Deux exercices importants	460
71	Un exercice sur les indicatrices : réduction du tableau de contingence (Equivalence distributionnelle)	460
72	Tableau de Burt et analyse canonique généralisée ; cas des données dichotomiques	463
	Les points importants	465

CHAPITRE XIV VERS L'ANALYSE DES TABLEAUX A PLUS DE DEUX DIMENSIONS

1	Equivalence entre tableaux de données et opérateurs en analyse linéaire	470
11	Equivalence entre tableaux de distances	470
12	Equivalence entre tableaux "individus x caractères" dans une optique de description	472
121	Equivalences et opérateurs	472
122	Images obtenues par les opérateurs	473
13	Equivalence entre tableaux "individus x caractères" dans une optique de prévision	474
14	Equivalence entre variables qualitatives	475
15	Equivalence entre une variable qualitative et un paquet de variables quantitatives	477
16	Bilan	477
2	Produit scalaire et distance entre opérateurs D_p -symétriques	478
21	Optique de description : proximité entre triplets.....	479
211	Choix de la métrique euclidienne classique	481
212	Choix de la métrique $D_{11\sigma^2}$	482

22	<i>Optique de prévision</i>	482
221	<i>Proximités entre paquets de caractères quantitatifs</i>	482
222	<i>Proximités entre caractères qualitatifs</i>	483
223	<i>Proximité entre un caractère qualitatif et un paquet de caractères quantitatifs</i>	486
3	Pratique des opérateurs	488
31	<i>Description des opérateurs à l'aide de l'analyse en composantes principales</i> ...	488
311	<i>Opérateurs "individus"</i>	488
31.2	<i>Opérateurs-caractères</i>	489
32	<i>Pratique des opérateurs dans une optique de description</i>	492
33	<i>Pratique des opérateurs dans une optique de prévision</i>	494
	Bilan	496

CHAPITRE XV CLASSIFICATION AUTOMATIQUE

1	Les mesures de proximité sur un ensemble	499
11	<i>Exemple : mesures de proximité utilisées en phytosociologie</i>	499
111	<i>Les données</i>	499
112	<i>Mesures de proximité dans le cas de caractères "présence-absence "</i>	499
113	<i>Mesures de proximité tenant compte de l'abondance des espèces</i>	501
114	<i>Conclusion</i>	502
12	Indice de similarité - Indice de dissimilarité	502
121	<i>Définitions</i>	502
122	<i>Différents types d'indices de dissimilarité</i>	503
123	<i>Compléments de vocabulaire</i>	505
123.1	<i>Espace métrique</i>	505
123.2	<i>Intermédiaire</i>	505
123.3	<i>Une relation d'ordre sur l'ensemble des indices de dissimilarité</i>	505
123.4	<i>Image réciproque d'un indice de dissimilarité</i>	505
124	Equivalence entre indices de dissimilarités - Transformation d'indices	507
124.1	<i>Préordonnance et équivalence</i>	507
124.2	<i>Transformations d'indices</i>	508
13	Indices de dissimilarité sur un produit cartésien	511
14	Indices de dissimilarité sur un vectoriel	513
141	<i>Semi-norme et norme</i>	513
142	<i>Ecart et distance associés</i>	513
143	<i>Discussion</i>	514
144	<i>Exemple : normes de Minkowski</i>	516
15	Mesures de dissimilarité sur l'ensemble des parties d'un ensemble	517
151	Dissimilarités sur $P(I)$ définies sans tenir compte de dissimilarités sur I ... 517	
151.1	<i>Distance du poids de la différence symétrique</i>	517
151.2	<i>Distance d'Ochiai et distance du x^2</i>	520
151.3	<i>Distances associées à des supervaluations</i>	522
151.31	<i>Définitions et propriétés</i>	523
151.32	<i>Quelques supervaluations sur le treillis $P(I)$ des parties d'un ensemble I</i> .. 525	
152	Dissimilarités sur $P(I)$ tenant compte de dissimilarités sur I	526
152.1	<i>Distance de Hausdorff</i>	526
152.2	<i>Distance du maximum</i>	528
152.3	<i>Distance des centres de gravité - Distance de l'inertie</i>	529
152.4	<i>Conclusion</i>	532

16	<i>Proximité entre relations binaires</i>	533
161	<i>Proximité entre relations binaires considérées comme parties de $I \times I$</i>	534
162	<i>Proximité entre relations binaires considérées comme applications de I dans $P(I)$ dans $P(I)$</i>	537
163	<i>Distances entre partitions associées à des supervaluations</i>	539
164	<i>Deux coefficients de corrélation pour comparer des préordres totaux</i>	541
2	Hiérarchie de parties et arbre de classification	544
21	<i>Hiérarchie et hiérarchie stratifiée</i>	544
211	<i>Hiérarchie totale : exemple et définitions</i>	544
212	<i>Hiérarchie stratifiée et préordonnance induite: exemples et définitions</i>	548
213	<i>Compatibilité entre partitions et hiérarchie</i>	552
214	<i>Exemple récapitulatif</i>	553
22	<i>Hiérarchie indicée</i>	556
221	<i>Définition</i>	556
222	<i>Distance ultramétrique associée à une hiérarchie totale indicée</i>	557
	<i>Les points importants</i>	559
3	Quelques techniques de classification automatique	560
31	<i>Panorama des techniques présentées</i>	560
311	<i>Vocabulaire</i>	560
311.1	<i>Classes monothétiques - Classes polythétiques</i>	560
311.2	<i>Algorithme agglomératif - Algorithme divisif</i>	560
312	<i>Présentation des techniques</i>	561
32	<i>Quelques techniques de classification hiérarchique fournissant des classes polythétiques</i>	561
321	<i>Méthode de l'ultramétrie sous-dominante (single linkage)</i>	561
321.1	<i>Ultramétrie sous-dominante d'un indice de dissimilarité</i>	562
321.11	<i>Définitions</i>	562
321.12	<i>Ultramétrie sous-dominante et chemin le plus court</i>	563
321.13	<i>Algorithme de Prim fournissant l'ultramétrie sous-dominante</i>	564
321.2	<i>Autres algorithmes permettant d'obtenir l'ultramétrie sous dominante</i>	567
321.21	<i>Algorithme de Johnson</i>	567
321.22	<i>Algorithme de Roux</i>	568
321.23	<i>Algorithme de Lerman</i>	569
321.24	<i>Algorithme Slink de Sibson</i>	569
321.3	<i>Critique de la méthode</i>	569
322	<i>Algorithmes procédant par fusions successives des deux classes les plus proches</i>	570
322.1	<i>Famille des algorithmes de Lance et Williams</i>	571
322.2	<i>Autres algorithmes</i>	573
323	<i>Algorithme de Sokal et Michener</i>	574
324	<i>Algorithmes basés sur l'optimisation d'un critère</i>	575
325	<i>Conclusion sur les méthodes de classification hiérarchique polythétique</i>	578
33	<i>Quelques techniques de classification non hiérarchique fournissant des classes polythétiques</i>	581
331	<i>Algorithme de Van Den Driessche</i>	581
332	<i>Algorithmes procédant par transferts</i>	583
333	<i>Algorithmes du type "nuées dynamiques"</i>	585
333.1	<i>L'algorithme de Fargy</i>	585
333.2	<i>Algorithmes contemporains de celui de Fargy</i>	589
333.3	<i>A propos de la méthode introduite par E. Diday</i>	591
34	<i>Procédés divisifs conduisant à des classes monothétiques</i>	593

341	<i>A propos des indices utilisés en segmentation</i>	594
341.1	<i>Proximité entre caractères qualitatifs</i>	594
341.2	<i>Proximité entre un caractère qualitatif et un caractère quantitatif</i>	595
342	<i>Algorithme de Williams et Lambert</i>	596
343	<i>Méthode de Belson et méthodes dérivées</i>	601
343.1	<i>Cas où y est quantitatif</i>	602
343.2	<i>Cas où y est qualitatif</i>	603
<i>Bilan</i>		604
BIBLIOGRAPHIE		605
INDEX		613