

REVUE DE STATISTIQUE APPLIQUÉE

A. BABILLIOT

Typologie des progiciels statistiques pour micro-ordinateur

Revue de statistique appliquée, tome 36, n° 3 (1988), p. 15-32

http://www.numdam.org/item?id=RSA_1988__36_3_15_0

© Société française de statistique, 1988, tous droits réservés.

L'accès aux archives de la revue « *Revue de statistique appliquée* » (<http://www.sfds.asso.fr/publicat/rsa.htm>) implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

*Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques*

<http://www.numdam.org/>

TYPOLOGIE DES PROGICIELS STATISTIQUES POUR MICRO-ORDINATEUR

A. BABILLIOT

*Université Paris IX Dauphine (Lea Reinsch) – Centre Scientifique IBM France
Place Mal De Laitre de Tassigny Paris 16^{ème} / 5 Place Vendôme 75001 Paris*

RÉSUMÉ

Cet article présente la typologie d'un échantillon de progiciels statistiques fonctionnant sur micro-ordinateurs. Cette typologie a été réalisée en utilisant deux méthodes statistiques d'analyse de données : l'analyse des correspondances et la méthode de classification par comparaison par paires. Notre étude a permis d'une part de recenser et de mettre en valeur des types particuliers de progiciels et d'autre part d'étudier les différents résultats générés par les deux méthodes statistiques.

Mots clés : Classification automatique, Typologie, Progiciels statistiques, Progiciels micro-informatique, Analyse des correspondances, Comparaison par paires.

ABSTRACT

This paper describes the typology of a sample of statistical software on micro-computers. This typology was made with two clustering methods: correspondence analysis and paired comparison analysis. Our work allow us to make an inventory and to lead to particular classes of software and also to study the discrepancies among results produced by the two statistical methods.

Keywords: Automatic clustering, Typology, Statistical software, Micro-computers software, Correspondence analysis, Paired comparison analysis.

1. Introduction

1.1. Contexte de l'étude

Il existe de nombreux logiciels de traitements statistiques fonctionnant sur micro-ordinateurs. Ces logiciels n'ont pas toujours les mêmes objectifs et n'utilisent pas tous les mêmes moyens. Les utilisateurs potentiels de ces logiciels qui ne sont pas dans tous les cas familiers de l'informatique et parfois des statistiques, ont des difficultés pour choisir un logiciel parmi le grand nombre de ceux disponibles. On aimerait avoir une vue structurée des logiciels qui nous permette de discerner des groupes cohérents correspondant à des besoins généraux, limitant ainsi l'éventail des choix possibles.

1.2. Objectifs de l'étude

1. Nous désirons recenser et décrire de façon détaillée un grand nombre de logiciels statistiques.
2. Il serait intéressant d'avoir une information synthétique, qui permette de voir si derrière une dénomination générique commune (logiciel statistique) il n'existe pas plusieurs groupes homogènes de logiciels.
3. Nous aimerions savoir si les caractéristiques de ces groupes de logiciels de statistiques sont propres à la nature des logiciels ou à leur destination.
4. Enfin, cette étude donne l'occasion de comparer les résultats obtenus par deux méthodes de classification aux approches différentes.

Nous voulons donc réaliser une typologie des logiciels statistiques sur micro-ordinateurs, c'est-à-dire déterminer, à l'intérieur d'un corpus le plus représentatif possible du marché, les traits caractéristiques et les groupes de logiciels qui s'y rattachent. La typologie finale devra prendre en compte à la fois les similarités et les différences entre les logiciels.

1.3. Fichiers de données

- Au cours d'une étude préalable [1] concernant les éléments d'évaluation des progiciels micro-informatiques, nous avons déterminé 37 variables informatiques et statistiques permettant de décrire les logiciels statistiques (tableau IV). Cette même étude [1] nous a amené à recenser environ 160 progiciels. Nous avons ensuite adressé un questionnaire, reprenant les éléments d'évaluation, à 145 concepteurs de progiciels statistiques commercialisés (hors prototypes) dont 69 nous ont répondu (tableau V).

Ce fichier donne une image partielle de "l'état de l'art" dans le domaine des techniques statistiques et informatiques implantées dans des progiciels statistiques micro-informatiques.

- Les données que nous allons analyser se présentent sous la forme d'une matrice croisant 69 individus (progiciels) et 37 variables qualitatives, dont chaque élément est la modalité prise par une variable pour un individu. Le fichier comporte une douzaine de données manquantes.

2. Méthodes statistiques

Nous voulons réaliser une typologie de façon automatique grâce à des moyens informatiques. La statistique offre plusieurs méthodes, dont les algorithmes ont été implantés sur ordinateurs, permettant d'effectuer une typologie. Il faut choisir parmi ces techniques de classification, la plus adaptée à nos objectifs et aux contraintes que nous nous sommes fixées.

Nous recherchons une partition formée des classes homogènes de progiciels se ressemblant. Cette partition doit être réalisée au regard de l'ensemble des variables décrivant notre fichier de données. Notre absence de connaissance du sujet

a priori, rend difficile la fixation à l'avance, du nombre de classes de progiciels que nous désirons obtenir, ainsi que du "niveau" de la partition dans le cas de classes emboîtées. Notre fichier de base comportant plusieurs données manquantes (non réponses), il faut que la méthode réalisant la typologie prenne ce type de données en compte et propose des éléments permettant de caractériser les classes obtenues.

Quatre méthodes de classification semblent mieux répondre à nos contraintes : la méthode de classification hiérarchique, la méthode des nuées dynamiques (Diday), l'analyse des correspondances et une méthode de classification par comparaison par paires (Marcotorchino, Michaud).

Les deux dernières étant implantées au Centre Scientifique IBM où nous effectuons nos recherches, nous allons principalement utiliser la méthode de classification par comparaison par paires [3,4] pour réaliser deux classifications avec des critères d'agrégation différents. Nous appliquerons ensuite un algorithme d'analyse factorielle des correspondances multiples (AFC) [2] pour vérifier la stabilité des résultats obtenus précédemment (l'AFC traitera un tableau disjonctif complet).

2.1. Méthode de classification par comparaison par paires

Cette méthode qui est fondée sur une approche de comparaison par paires, développée par Marcotorchino et Michaud [3, 4, 5], permet de réaliser la classification de données par agrégation des similarités.

2.1.1. Représentation des données

On distingue trois étapes dans la représentation des données à classifier.

- Le tableau de données brutes, sous la forme d'une matrice croisant individus et variables.
- Les tableaux de comparaison par paires, appelés tableaux C^k qui croisent les individus entre eux pour chaque variable en indiquant pour chaque paire d'individus si ils ont la même modalité pour une variable k .
- Le tableau C ou tableau collectif de base qui somme l'ensemble des tableaux C^k .

2.1.2. Forme des tableaux C^k de données

Le tableau initial de données (n individus et m variables), donnant la modalité de chaque variable pour chaque individu, est représentée par m tableaux carrés C^k de comparaison par paires, de taille $n \times n$, dont le terme général est de la forme C_{ij}^k et dont les valeurs sont :

$$C_{ij}^k = 1 \text{ si les individus } i \text{ et } j \text{ ont la même modalité (pour la variable } k)$$

$$C_{ij}^k = 0 \text{ autrement.}$$

2.1.3. Propriétés des tableaux

- Tableaux C^k

Chaque tableau C^k représente une relation d'équivalence, vérifiant les conditions suivantes : Réflexivité, Symétrie et Transitivité.

- Tableaux \bar{C}^k

On peut associer à chaque tableau C^k son tableau complémentaire à 1, appelé \bar{C}^k .

Ce tableau \bar{C}^k vérifie les conditions suivantes : Non réflexivité, Symétrie et Inégalité triangulaire. Par construction on a :

$$C = \sum_{k=1}^m C^k \quad \text{avec} \quad C_{ij} = \sum_{k=1}^m C_{ij}^k$$

On définit C_{ij} comme le nombre de variables pour lesquelles i et j ont même modalité.

- De même, il est possible de construire un tableau $\bar{C} = \sum_{k=1}^m \bar{C}^k$ pour lequel \bar{C}_{ij} est égal au nombre de variables où i et j ont des modalités différentes.

2.1.4. Le tableau collectif Y

On appelle Y le tableau représentant la partition collective que l'on recherche. Y est un tableau binaire de comparaison par paires ($n \times n$) possédant les propriétés d'une relation d'équivalence avec :

$$y_{ij} = 1 \text{ si } i \text{ appartient à la même classe que } j$$

$$y_{ij} = 0 \text{ sinon}$$

Le tableau Y doit approximer au mieux les m relations C^k .

2.1.5. Définition des critères d'agrégation

Nous recherchons une "partition s'ajustant au mieux avec un ensemble de relations de similarités". Il faut donc définir une mesure de similarité, c'est-à-dire un critère d'agrégation permettant de réaliser la partition [3, 4, 5].

Plusieurs critères peuvent être utilisés dans le cadre des comparaisons par paires, nous voulons étudier les résultats générés par deux critères différents.

Un premier critère basé sur la notion de "votes supportant une opinion collective" développée par Condorcet [4]. Un second critère dérivé du critère défini en statistique des contingences par Jordan [5].

Critère de Condorcet

Ce critère correspond à la règle de la majorité par paires.

Cette règle consiste à retenir pour chaque valeur de y_{ij} celle qui est supportée par une majorité de variables.

L'opinion collective Y que nous recherchons représente la partition qui maximise le critère $V(Y)$ avec $V(Y) = \sum \sum C_{ij} Y_{ij} + \bar{C}_{ij} (1 - Y_{ij})$. Y étant le tableau de comparaison par paires d'une relation d'équivalence.

Critère de Jordan

Le critère de Jordan dont on trouvera le détail des propriétés dans [5], centre les variables par rapport à la moyenne en ligne et en colonne.

Ce critère s'exprime sous la forme suivante :

$$\text{Jordan}(Y, C) = \sum_i \sum_j \left(c_{ij} - \left(\frac{c_i + c_j}{2N} \right) y_{ij} \right)$$

c_i (resp. c_j) désignant la somme de la ligne i (resp. j) du tableau C .

On notera [3, 4] que l'ensemble du problème est linéarisé pour les deux critères.

3. Analyses statistiques

3.1. Résultats

TABLEAU I

Classes obtenues par le critère de Condorcet

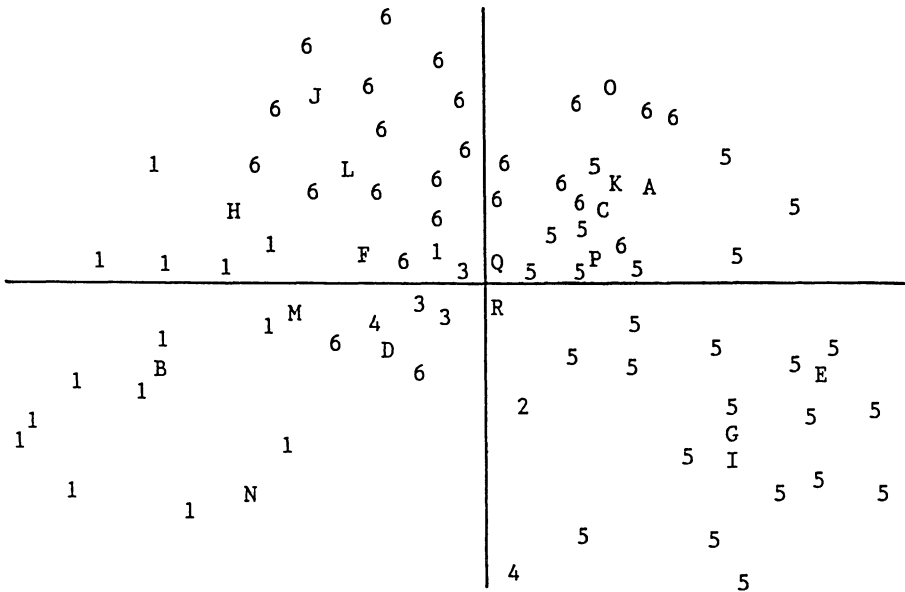
CLASSE 1	CLASSE 3	CLASSE 5
1 AB STAT	28 MACROSTU	10 DATAVIEW
14 EAL	20 INTROSTA	29 MODLER D
17 EXEC U C	35 PC ANOVA	16 ESP
31 NUMBER C	6 BIOMPC	CLASSE 6
33 NWA STAT	37 PC SCA	12 DCS MULT
36 PC ISP	39 PC STATI	CLASSE 7
40 PRODAS	7 BMDPC	13 DELTA PL
42 SAS/STAT	41 SAM-86	CLASSE 8
45 SMART FO	43 SIGSTAT	66 TRIAGE
46 SORITEC	44 SL-MICRO	19 GERIATRI
49 SPSS/PC+	9 CRUNCH	CLASSE 9
52 STATA	48 SPS	22 LEAS
55 STATGRAP	56 STAT IA	CLASSE 10
59 STATPACG	57 STATISTI	30 MULT DPR
62 STATPRO	58 S MACE	CLASSE 11
69 YSTAT	60 STAT PAC	32 NTSYS-PC
CLASSE 2	15 ELF	CLASSE 12
3 ANAFAC	63 STATS-2	34 OTIS
25 MICROMAR	18 FLURP	51 STAN
27 MICROSTF	65 SYSTAT	CLASSE 13
4 ANCOPR	CLASSE 4	50 SST
5 ANACONDA	23 LIDE	CLASSE 14
47 SPMC	38 PCSS(SM)	54 STATAPL
11 DATAVISI	24 MARCOS	67 123 FORE
2 AFC	21 ITEM	53 STATAIDS
64 STRADES	8 CHADOC	61 STATPLAN
	26 MEDLOG	
	68 UNIX STA	

TABLEAU II
Classes obtenues par le critère de Jordan

CLASSE 1	CLASSE 5	CLASSE 6
2 AFC	1 AB STAT	28 MICROSTU
3 ANAFAC	26 MEDLOG	39 PC STATI
4 ANCOPR	12 DCS MULT	24 MARCOS
5 ANACONDA	29 MODLER D	41 SAM-86
11 DATAVISI	30 MULT DPR	35 PC ANOVA
18 FLURP	31 NUMBER C	54 STATAPL
19 GERIATRI	33 NWA STAT	43 SUGSTAT
21 ITEM	8 CHADOC	56 STAT IA
22 LEAS	36 PC ISP	57 STATISTI
23 LIDE	38 PCSS(SM)	58 S MAGE
25 MICROMAR	40 PRODAS	44 SL-MICRO
27 MICROSTF	42 SAS/STAT	60 STAT PAC
47 SPMC	45 SMART FO	7 BMDPC
53 STATAIDS	46 SORITEC	6 BIOMPC
64 STRADES	49 SPSS/PC+	63 STATS-2
66 TRIAGE	50 SST	15 ELF
CLASSE 2	10 DATAVIEW	65 SYSTAT
16 ESP	52 STATA	37 PC SCA
CLASSE 3	14 EAL	20 INTROSTA
32 NTSYS-PC	55 STATGRAP	9 CRUNCH
61 STATPLAN	59 STATPACG	48 SPS
67 123 FORE	62 STATPRO	
CLASSE 4	17 EXEC U S	
34 OTIS	68 UNIX STA	
13 DELTA PL	69 YSTAT	
51 STAN		

Tableau III
Graphique obtenu par l'AFC

Les individus ont été numérotés en fonction de leur appartenance aux principales classes obtenues par le critère de Jordan.



Valeurs propres : 0.25/0.15 Inertie 1^{er} axe 7,8 % Inertie 2^{ème} axe 4.8 %

Modalités prises par les variables les plus significatives

A : anglo-saxon B : français C : pas d'analyse des données D : analyse des données E : présence didacticiel F : pas de didacticiel G : couleur H : noir et blanc I : haute résolution J : mode texte K : ms/dos L : intégrateur non M : pas de fonction d'édition N : pas d'anova O : anova ancova P : régression Q : fortran R : mode commande

3.2. Discussion

3.2.1. Statistiques descriptives simples

Nous avons réalisé des histogrammes donnant la répartition des modalités de chaque variable de notre fichier, dont voici les traits marquants.

Caractéristiques informatiques

Dans leur majorité les progiciels mettent en œuvre des moyens informatiques classiques : système d'exploitation MS/DOS (80 %), conduite du programme au clavier (95 %). On discerne une orientation normale vers le calcul mathématique : langage FORTRAN et BASIC compilé, présence d'un coprocesseur arithmétique (64 %). Les fonctions progiciels sont essentiellement centrées autour du calcul statistique, les fonctions de gestion des données ne sont pas systématiques. Le mode de fonctionnement est partagé entre mode menu et mode commande. Des fonctions "avancées" orientées vers un meilleur confort d'utilisation du progiciel ou vers son ouverture à l'extérieur (aide en ligne, souris, didacticiel, intégrateur, interfaces, accès source, utilitaires graphiques) ne sont que très rarement proposées.

Caractéristiques statistiques

Il apparaît que les progiciels sont très complets en ce qui concerne les techniques statistiques de base, généralement considérées comme indispensables : moyenne, variance, écart type, etc (95 %), tests usuels, régression linéaire (75 %). En revanche, des méthodes plus complexes comme l'analyse de variance, l'analyse de covariance et certaines formes de régression sont moins présentes (55 %). L'analyse de données n'est pas souvent traitée. Il faut peut être voir là, d'une part l'influence des progiciels d'origine anglo-saxonne, qui privilégient les méthodes descriptives ainsi que la régression et ne développent pas l'analyse de données et d'autre part la faible diffusion de ces méthodes dans le grand public.

3.2.2. Classification des progiciels : Comparaison par paires

Utilisation du critère de Condorcet

Les résultats détaillés de cette classification figurent dans le tableau I.

Le critère de Condorcet détermine 14 classes qui se partagent en quatre classes à effectif important (1, 2, 3, 4), deux classes (5 et 14) à effectif moyen (3 ou 4 progiciels), et huit classes composées d'un ou deux progiciels.

Nous allons nous attacher à caractériser les quatre principales classes de progiciels.

- Classe 1 (16 progiciels)
 - Progiciels en langue anglaise
 - Mode graphique haute résolution couleur
 - Système d'exploitation MS/DOS
 - Option sortie sans risque du progiciel
 - Possibilité de paramétrage du progiciel

- Sauvegarde sophistiquée des résultats
- Gestion évoluée des données
- Présence de toutes les statistiques descriptives et de quelques techniques de régression. Plusieurs tests sont proposés.

Il s'agit d'une classe de progiciels en langue anglaise, d'un assez haut degré de complexité informatique, proposant essentiellement des méthodes statistiques simples.

- Classe 2 (9 progiciels)

- Progiciels en langue française
- Mode graphique texte noir et blanc
- Absence de didacticiel
- Absence de fonctions d'édition évoluées
- Pas de possibilité de fonctionnement sous un intégrateur
- Pas de mode de fonctionnement différé
- Propose essentiellement des techniques statistiques d'analyse de données

Il s'agit de progiciels en langue française, à faible degré de complexité informatique, spécialisés dans l'analyse de données.

- Classe 3 (20 progiciels)

- Progiciels en langue anglaise
- Mode graphique texte noir et blanc
- Absence de didacticiel
- Absence de fonctions graphiques évoluées
- Utilise entre 60 et 270 Ko de mémoire vive
- Pas de possibilité de fonctionnement sous un intégrateur
- Absence de fonctions de sécurité des données
- Absence de protection (copie) des progiciels
- Les techniques statistiques descriptives, l'analyse de variance et la régression sont particulièrement développées.

Il s'agit de progiciels en langue anglaise dont les possibilités informatiques sont moyennes et dont les capacités statistiques sont bien développées, notamment la régression et l'ANOVA.

- Classe 4 (7 progiciels)

- Progiciels en langue française
- Mode graphique texte couleur
- Système d'exploitation MS/DOS
- Absence de fonctions graphiques évoluées

- Utilise entre 60 et 270 Ko de mémoire vive
- Absence de fonctions d'édition évoluées
- Pas de possibilité de fonctionnement sous un intégrateur
- Les techniques statistiques descriptives sont moyennement développées, l'analyse de variance et la régression sont plus présentes.

Cette classe est composée de progiciels en langue française dont les possibilités informatiques sont moyennes et dont les capacités statistiques sont moyennement développées, notamment la régression et l'ANOVA.

Utilisation du critère de Jordan

Les résultats de cette classification figurent dans le tableau II.

Le critère de Jordan détermine 6 classes qui se partagent en trois classes à effectif important (1, 5, 6), deux classes (3 et 4) à effectif moyen (3 progiciels) et une classe (la 2) composée d'un seul progiciel. Nous allons nous attacher à caractériser les trois principales classes de progiciels.

• Classe 1 (16 progiciels)

- Progiciels en langue française
- Pas de possibilité de fonctionnement sous un intégrateur
- Absence de fonctions d'édition évoluées
- Fonctionnement en mode menus
- Documentation papier
- Absence de didacticiel
- Mode graphique texte noir et blanc
- Absence de fonctions de sécurité des données
- Absence de fonctions graphiques évoluées
- Propose essentiellement des techniques statistiques d'analyse de données, à un moindre degré quelques techniques descriptives simples, mais ni l'analyse de variance ni la régression ne sont particulièrement développées.

Il s'agit de progiciels en langue française, à faible degré de complexité informatique, spécialisés dans l'analyse de données.

• Classe 5 (25 progiciels)

- Progiciels en langue anglaise
- Mode graphique haute résolution couleur
- Présence de didacticiel
- Mode de gestion des données évolué
- Système d'exploitation MS/DOS
- Absence de fonctions graphiques évoluées
- Structure sous la forme d'un bloc de programme gérant ses fonctions

- Mode d'impression évolué
- Possibilité de paramétrage du progiciel
- Les techniques statistiques descriptives, l'analyse de variance et la régression sont particulièrement développées.

Il s'agit d'une classe de progiciels en langue anglaise, d'un assez haut degré de complexité informatique, proposant à la fois des méthodes statistiques simples et plusieurs techniques d'analyse de variance et de régression.

- Classe 6 (21 progiciels)
 - Progiciels en langue anglaise
 - Absence de didacticiel
 - Mode graphique texte noir et blanc
 - Absence de fonctions graphiques évoluées
 - Utilise entre 60 et 270 Ko de mémoire vive
 - Pas de possibilité de fonctionnement sous un intégrateur
 - Absence de protection (copie) des progiciels
 - Absence de fonctions de sécurité des données
 - Les techniques statistiques descriptives, l'analyse de variance et la régression sont particulièrement développées.

Il s'agit de progiciels en langue anglaise, d'un faible niveau de complexité informatique, proposant à la fois des méthodes statistiques simples et plusieurs techniques d'analyse de variance et de régression.

Comparaison des critères

La comparaison des deux classifications résultant de l'utilisation du critère de Jordan et du critère de Condorcet fait apparaître des différences dans la structure des classes.

- Le critère de Jordan est plus synthétique que le critère de Condorcet, et a tendance à réunir dans de grandes classes des éléments isolés par le critère de Condorcet. En effet, la classe 1 obtenue par le critère Jordan regroupe les classes 2, 4, 8, 9 obtenues par le critère de Condorcet. La classe 5 obtenue par le critère de Jordan regroupe les classes 1, 4, 5, 6, 10, 13 obtenues par le critère de Condorcet. La classe 6 obtenue par le critère de Jordan est identique à la classe 3 obtenue par le critère de Condorcet.

Ainsi la classification issue du critère de Condorcet est plus détaillée mais moins facile à interpréter du fait de l'abondance d'éléments isolés.

- L'analyse des caractéristiques de ces classes comportant des effectifs importants est à peu près identique pour les deux classifications, les critères agissant plus sur la densité des classes que sur leur signification.

Le critère de Jordan réunit donc les éléments isolés par le critère de Condorcet sans dénaturer la signification des classes importantes.

La classification fait apparaître trois pôles de description des progiciels.

1. Le potentiel informatique

Caractérisé par la qualité et le nombre des moyens mis à la disposition des utilisateurs.

2. Le potentiel statistique

Caractérisé par la qualité et le nombre des techniques statistiques proposées aux utilisateurs.

3. L'origine (française, anglo-saxonne)

Ces trois pôles se regroupent en quatre classes de progiciels.

1. Les progiciels d'origine française, à potentiel informatique faible et à potentiel statistique moyen.
2. Les progiciels d'origine française, à potentiel informatique faible et à potentiel statistique orienté vers l'analyse de données.
3. Les progiciels d'origine anglo-saxonne, à potentiel informatique faible et à potentiel statistique fort.
4. Les progiciels d'origine anglo-saxonne, à potentiel informatique fort et à potentiel statistique moyen.

3.2.3. Classification des progiciels : Analyse des correspondances

La représentation graphique des individus et des variables projetés sur les axes 1 et 2 forme un nuage continu de points. L'absence de groupe distinct de points ne permet pas de mettre en valeur des classes de progiciels.

Lorsque l'on numérote les individus en fonction de leur appartenance aux classes obtenues par le critère de Jordan on voit apparaître des groupes homogènes de points ce qui confirme la stabilité des données par rapport aux méthodes statistiques. La projection des variables montre que la signification des classes est respectée.

4. Conclusion

- Les méthodes statistiques de classification que nous avons utilisées nous ont permis de réaliser une typologie des progiciels statistiques sur micro-ordinateur. Ces méthodes de structuration des données sont des outils à la fois rapides et puissants permettant de réaliser et d'interpréter facilement une classification. Cette typologie fait apparaître un nombre limité de classes de progiciels dont l'interprétation est cohérente et stable pour les deux critères d'agrégation utilisés.

- On remarque, dans ce cas précis, que ce n'est que rétrospectivement que l'AFC confirme l'autre classification. En effet l'étude de la représentation graphique des principaux axes de l'AFC ne permet pas de discerner des groupes de points

pouvant former des classes. La méthode d'agrégation des similarités produit donc pour cet exemple des résultats plus facilement interprétables.

- Les principales caractéristiques des classes obtenues mettent en valeur à la fois la composante informatique (technique et ergonomique), la composante statistique et l'origine des progiciels. Ces résultats sont originaux, ils donnent une vision structurée du marché de ce type de progiciels et permettent à un utilisateur potentiel de situer rapidement le groupe de progiciels qui correspond à ses besoins.

- Il n'apparaît pas de groupe homogène de progiciels que leurs caractéristiques rendent particulièrement adaptés à un domaine d'application particulier. Les deux classifications font plutôt apparaître des classes de progiciels caractérisées par leur degré de complexité informatique ou statistique et leur nationalité.

- Dans le cadre d'une étude menée par un utilisateur potentiel, il serait intéressant de disposer de deux moyens de choix : un logiciel statistique donnant une classification optimisée comme celui que nous avons utilisé et un logiciel de segmentation dynamique et hiérarchique permettant d'éliminer et de conserver des progiciels pas à pas en fonction de critères (valeur des variables) importants pour l'utilisateur.

Références

- [1] A. BABILLIOT (29 Juin 1988). — Typologie critique des méthodologies informatiques pour l'analyse des données en épidémiologie, *Nouvelle Thèse en informatique*, Paris IX Dauphine.
- [2] J.M. BOUROCHE, G. SAPORTA (1980). — l'Analyse des Données, Presse Universitaires de France, Que Sais-je, Paris.
- [3] J.F. MARCOTORCHINO (1985). — Utilisation des Comparaisons par Paires en Statistiques des Contingences (Partie I et II), Centre Scientifique IBM France, Étude N° F-069.
- [4] J.F. MARCOTORCHINO, P. MICHAUD (1981). — Agrégation des Similarités en Classification Automatique, Centre Scientifique IBM France, Étude N° F-012.
- [5] P. MICHAUD (1985). — Agrégation à la Majorité II : Analyse du Résultat d'un Vote, Centre Scientifique IBM France, Étude N° F-052.

TABLEAU IV

Définition des variables utilisées

VARIABLE 1 : coprocesseur-arithmétique

Non : 1, Optionnel : 2, Obligatoire : 3

VARIABLE 2 : progiciel protégé

Non : 0, Oui : 1

VARIABLE 3 : affichage couleur

Noir/blanc : 1, Couleur : 2

VARIABLE 4 : langue utilisée pour les messages

Français : 1, Anglais : 2

VARIABLE 5 : structure du progiciel

plusieurs sous programmes : 1, plusieurs sous programmes liés par un menu : 2, un seul programme intégré gérant ses fonctions : 3

VARIABLE 6 : accès au système d'exploitation (SE)

pas d'accès au SE : 0, accès aux fonctions : renommer et effacer : 1, accès à quelques fonctions du SE : 2, accès à toutes les fonctions du SE : 3

VARIABLE 7 : possibilité de paramétrage du progiciel

Non : 0, Oui : 1

VARIABLE 8 : présence d'un didacticiel

Non : 0, Oui : 1

VARIABLE 9 : possibilité de sortie sans risque du progiciel

Non : 0, Oui : 1

VARIABLE 10 : nombre d'interfaces de communications

pas d'interface : 0, ASCII : 1, 2 → 9

VARIABLE 11 : fonctions d'édition

Non : 0, Rapports : 1, Traitement texte : 2, 1 et 2 : 3

VARIABLE 12 : création de formules ou de fonctions mathématiques

pas de possibilité de description de formules : 0, création de formules utilisant des opérateurs simples : 1, création de formules utilisant des opérateurs complexes : 2, création des fichiers de formules : 3, 2 et 3 : 4

VARIABLE 13 : mémoire vive utilisée

$0 \text{Ko} < V \leq 60 \text{Ko}$: 1, $60 < V \leq 270$: 2, $270 < V \leq 400$: 3, $400 < V \leq 700$: 4, $700 < V \leq 2000$: 5

VARIABLE 14 : mode de stockage des données

Disquette : 1, Disque Dur : 2, Disquette RAM : 3

VARIABLE 15 : mode de stockage des programmes

Disquette : 1, Disque Dur et master : 2, Disque Dur : 3, Disquette RAM : 4

TABLEAU IV (suite)
Définition des variables utilisées

VARIABLE 16 : mode d'impression

Copie d'écran : 1, Fichier : 2, 1 et 2 : 3

VARIABLE 17 : langage de programmation

Fortran : 1, Cobol : 2, Basic compilé : 3, Pascal : 4, C : 5, Basic interprété : 6, Pascal UCSD : 7, Assembleur 8, APL : 9

VARIABLE 18 : système d'exploitation

MS/DOS : 1, Apple : 2, Machintosh : 3, 1 + 2 : 4, CP/M + 1 + 2 + 3 : 5 UNIX et XENIX : 6, HP/DOS : 7, ATARI + 1 : 8, 6 + 1 + 3 : 9, MS/DOS XENIX : 10, B/DOS + 1 + 6 : 11, MS/DOS CP/M : 12, 12 + 2 + 3 + 6 : 13

VARIABLE 19 : fonctionnement sous intégrateurs

Non : 0, 1 intégrateur, 2 intégrateurs, 3 intégrateurs

VARIABLE 20 : éléments sauvegardés sur fichier

Non : 0, Résultats seuls : 1, Graphiques + 1 : 2, Résultats intermédiaires : 3, 1 + 3 : 4, 1 + 2 + 3 : 5

VARIABLE 21 : mode graphique

Mode texte : 1, Basse résolution : 2, Haute résolution : 3

VARIABLE 22 : accès au fichier source

Non : 0, Description des E/S : 1, Oui : 2

VARIABLE 23 : sécurité des données

Non : 0, Codage des données : 1, Mot de passe : 2, 1 + 2 : 3

VARIABLE 24 : mode différé

Non : 0, Simple : 1, Fichier de commandes : 2, Fichier d'instructions : 3

VARIABLE 25 : mode commande

Non : 0, Commandes : 1, Touches fonctions : 2, Macro Commandes : 3, Macro-commandes programmées : 4, 1 + 2 + 3 : 6, 1 + 2 + 3 + 4 : 7

VARIABLE 26 : mode menu

Non : 0, Menus simples : 1, Menus dynamiques : 2, Menus programmables : 3, 2 + 3 + 4 : 5

VARIABLE 27 : type de documentation

Pas de documentation : 0, Documentation papier : 1, Documentation à l'écran : 2

VARIABLE 28 : gestion des données

Non : 0, Éditeur : 1, Tableur : 2, Fonctions SGBD : 3, 1 + 2 + 3 : 4

VARIABLE 29 : fonctions graphiques avancées

Non : 0, Oui : 1

TABLEAU IV (suite)
Définition des variables utilisées

VARIABLE 30 : aide

Non : 0, Aide en contexte : 1, Solutions proposées : 2, 1 + 2 : 3

VARIABLE 31 : type de périphériques utilisé

Clavier : 0, Souris : 1, 0 + 1 : 2

VARIABLE 32 : statistiques descriptives (histogrammes courbes)

Ni l'un ni l'autre : 0, L'un ou l'autre : 1, Les deux : 2

VARIABLE 33 : statistiques descriptives (Variance Écart Type/Fréquence Quantiles Minimum Maximum)

Ni l'un ni l'autre : 0, L'un ou l'autre : 1, Les deux : 2

VARIABLE 34 : tests

Pas de tests : 0, Entre 1 et 4 tests : 1, Plus de 4 tests : 2

VARIABLE 35 : régression

Non traite : 0, régression simple : 1, 1 + reg complexe : 2

VARIABLE 36 : analyse de variance (ANOVA) et de covariance

Pas d'ANOVA : 0, ANOVA seule : 1, ANOVA et ANCOVA : 2

VARIABLE 37 : analyse de données (AD)

AD non traite : 0, AD traitée : 1

TABLEAU V

Liste des progiciels statistiques sur micro-ordinateurs étudiés

1 : AB STAT	2 : AFC	3 : ANAFAC
4 : ANCOPR	5 : ANACONDA	6 : BIOMPC
7 : BMDPC	8 : CHADOC	9 : CRUNCH
10 : DATAVIEW	11 : DATAVISION	12 : DCS MULTISTAT
13 : DELTAPLAN	14 : EAL	15 : ELF
16 : ESP	17 : EXEC U STAT	18 : FLURP
19 : GERIATRI	20 : INTROSTA	21 : ITEM
22 : LEAS	23 : LIDE	24 : MARCOS
25 : MICRO-MARKETING	26 : MEDLOG	27 : MICRO-TSP
28 : MICROSTAT	29 : MODLER DSS	30 : MULTIDIM DPR
31 : NUMBER CRUNCHER	32 : NTSYS-PC	33 : NWA STAT
34 : OTIS	35 : PC ANOVA	36 : PC ISP
37 : PC SCA	38 : PCSS(SM)	39 : PC STATISTICIAN
40 : PRODAS	41 : SAM-86	42 : SAS/STAT
43 : SIGSTAT	44 : SL-MICRO	45 : SMART FORECAST
46 : SORITEC	47 : SPMC	48 : SPS
49 : SPSS/PC+	50 : SST	51 : STAN
52 : STATA	53 : STATAIDS	54 : STATAPL
55 : STATGRAPHICS	56 : STAT IA	57 : STATISTICS
58 : STATISTICIAN MACE	59 : STATPAC GOLD	60 : STAT PACK
61 : STATPLAN	62 : STATPRO	63 : STATS-2
64 : STRADES	65 : SYSTAT	66 : TRIAGE
67 : 123 FORCAST	68 : UNIX STATISTI	69 : Y STAT