

J.-P. BENZÉCRI

F. BENZÉCRI

**Liste des programmes du logiciel 'MacSAIF',
Système d'Analyse des InFormations, et
description de leurs fonctions**

Les cahiers de l'analyse des données, tome 15, n° 3 (1990),
p. 359-366

http://www.numdam.org/item?id=CAD_1990__15_3_359_0

© Les cahiers de l'analyse des données, Dunod, 1990, tous droits réservés.

L'accès aux archives de la revue « Les cahiers de l'analyse des données » 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/>

LISTE DES PROGRAMMES DU LOGICIEL 'MacSAIF', Système d'Analyse des InFormations, ET DESCRIPTION DE LEURS FONCTIONS

[MacSAIF]

J.-P. & F. BENZÉCRI

Nous donnerons successivement une liste des programmes, énumérés dans l'ordre alphabétique, et muni chacun d'une brève notice; puis les sommaires des articles de la revue *CAD* qui offrent une vue d'ensemble du logiciel 'MacSAIF'.

1 Liste alphabétique des programmes

NB: La présente liste ne comprend pas les programmes d'élaboration linguistique, brièvement décrits dans le sommaire de [LING. TRI]; ces programmes ne font pas partie de 'MacSAIF' mais peuvent être fournis aux linguistes.

'CAH': Classification Ascendante Hiérarchique; avec tracé d'arbres et tabulation d'une partition définie par les nœuds les plus hauts ou par des nœuds spécifiés; listages Facor et Vacor d'aide à l'interprétation en terme de facteurs, de variables, ou d'agrégats de variables; classification et aides à l'interprétation pour des ensembles supplémentaires.

'Carthage': programme de cartographie automatique, permettant de représenter sur une carte les résultats d'une classification automatique effectuée sur un ensemble d'unités territoriales tel que l'ensemble des départements français. Les trames affectées aux zones peuvent être choisies en dialogue à l'écran en s'aidant des résultats de l'analyse factorielle affichés à la demande au cours du déroulement du programme.

'corel': programme de calcul de coefficients de corrélations entre colonnes d'un tableau de nombres; permet d'effectuer des calculs sur des sous-ensembles d'individus, définis par une condition sur l'une des deux variables, et donc de réaliser une régression linéaire par morceaux; les résultats, demandés par dialogue, s'affichent à l'écran et sont conservés sur listage.

'creseri': permet de créer un tableau dont les lignes sont indicées par un ensemble de temps consécutifs et les colonnes sont des combinaisons de fonctions du temps, sinus, cosinus, exponentielle,... suffisant pour expérimenter sur des modèles complexes.

'cums': complète un tableau de correspondance en lui adjoignant des lignes ou colonnes définies par cumul; le programme peut être commandé par dialogue à l'écran ou par un listage.

'décale': à partir d'un tableau dont les lignes sont indicées par un ensemble de temps consécutifs, crée un tableau où à chaque colonne est associé un bloc de colonnes décalées en avance et en retard, selon les spécifications de l'utilisateur; l'analyse de ce tableau peut révéler inductivement des 'cycles'.

'discri': cf. analyse discriminante; grâce aux programmes 'qori', 'qorlsup', 'CAH', on dispose de fichiers de coordonnées sur les axes factoriels pour de nombreux ensembles (lignes ou colonnes, éléments principaux ou supplémentaires, centres de classes); étant donné un ensemble des *centres* et un ensemble des *individus*, 'discri' détermine, pour chaque *individu* le *centre* dont il est le plus proche.

'juxtab': pour juxtaposer horizontalement ou verticalement deux tableaux; permet notamment de calculer, par 'corel', des corrélations entre facteurs issus d'analyses différentes, de réunir en un seul tableau des données afférentes à des périodes de temps successives (analyse de tableaux ternaires)...

'litxcarte': à partir de deux fichiers texte donnant les coordonnées de points relevés sur une carte et énumérant les sommets sur le périmètre des unités territoriales d'un pays ou d'une région, on crée un fichier numérique de 'fond de carte' permettant de cartographier des données de base ou les résultats d'analyse de celles-ci. Le programme, par de nombreux affichages aide l'utilisateur à corriger ses erreurs de mesure.

'litxicqki': permet de cartographier une classification quelconque, non obtenue par la chaîne d'analyse factorielle et de CAH.

'manq': à partir d'un tableau de données comportant des données manquantes repérables par un code uniforme, par exemple '0', '9' ou '10000', on crée un listage décrivant en détail la distribution des manques, ce qui aide l'utilisateur à choisir un sous-tableau propre à être analysé.

'pers': cf. 'équation personnelle'; prépare le codage, par 'zrang', des réponses fournies, à un questionnaire comportant des batteries d'échelles, par des sujets faisant des notes un usage non uniforme.

'planF'; puisqu'on dispose de fichiers de coordonnées sur les axes factoriels pour de nombreux ensembles (lignes ou colonnes, éléments principaux

ou supplémentaires, centres de classes), le programme conversationnel 'planF' affiche à l'écran un nombre quelconque de ces ensembles sur des plans croisant deux axes spécifiés par l'utilisateur. Les individus de chaque ensemble sont représentés soit par leurs sigles, soit (pour plus de clarté si les individus sont nombreux) par un caractère unique. L'utilisateur saisit en copie d'écran les graphiques de son choix.

'**planX**': offre à l'utilisateur la même liberté de choix que 'planF', mais crée des graphiques qui sont des fichiers de texte.

'**printx**': conçu pour imprimer, sur une imprimante 'ImageWriter', sous forme de listages, des fichiers textes, dont les lignes ont, au plus, 136 caractères.

'**qori**': Analyse des Correspondances; mise en supplémentaire d'éléments désignés un par un ou par blocs, en dialogue à l'écran; avec création de listage des facteurs et aides à l'interprétation; les graphiques plans sont créés par 'planF' et 'planX'.

'**qorlsup**': adjoint à l'analyse d'un tableau (dont éventuellement certaines lignes ou colonnes ont été mises en supplémentaire) des tableaux extérieurs de lignes ou colonnes; sert notamment à adjoindre un tableau binaire (en 0,1) à l'analyse d'un tableau de Burt.

'**soustab**': extrait d'un tableau des sous-tableaux dont les ensembles de lignes et colonnes peuvent être spécifiés soit en dialogue à l'écran, soit par un fichier texte. Sert à créer un tableau sans données manquantes (cf. 'manq'); à séparer des sous-ensembles d'individus qui seront représentés à l'écran (cf. 'planF') par un caractère unique symbolisant une particularité qu'ils ont en commun...

'**zBurt**': à partir d'un tableau de numéros de modalités, créé par 'zrang', le programme 'zBurt' crée rapidement le tableau de Burt, ou tableau de cooccurrence des modalités.

'**zrang**': offre de nombreux choix pour examiner un tableau de données, en créant des histogrammes de tout format, des tableaux de tris croisés, des cartes (s'il s'agit de données relatives à un ensemble d'unités territoriales pour lequel existe un fichier de fond de carte créé par 'litxcarter'); et ensuite, ou simultanément, coder ces données, par découpage en classes (codage en 0,1) ou par codage flou (i.e. linéaire par morceaux: cf. [CODAGE LIN.]).

2 Bibliographie de synthèse sur le logiciel 'MacSAIF'

Trois articles du n°1 du Vol XIV (1989), de la revue *Cahiers de l'Analyse des Données* ont pour objet de présenter en détail le logiciel 'MacSAIF', bibliothèque de programmes d'analyse des données conçus pour être exécutés par les ordinateurs Macintosh+, SE, SE30, ou Macintosh II & IIX. Les

programmes nouveaux et les mises à jour font l'objet d'articles ultérieurs. Dans le présent sommaire, nous reprenons, en les complétant, des résumés de ces articles publiés dans *CAD*. Ces mêmes articles, éventuellement modifiés, servent de notice au logiciel 'MacSAIF'.

Tous ces programmes ont servi à des études publiées dans les Cahiers de l'Analyse des Données, en sorte que les exemples ne manquent pas à l'utilisateur qui désire acquérir ces programmes et se livrer à des essais. Un dossier 'Data', contenant des fichiers de données, est distribué avec le logiciel 'MacSAIF', ainsi qu'un dossier 'Atlas', contenant les fonds de carte disponibles (cf. [NOT. PROG. CART.] et [STAT. CART.]): France, Paris, Grèce, Italie, Portugal...

Programmes d'analyse des correspondances et de classification ascendante hiérarchique: notice d'utilisation, [NOT. CORR. CAH], par J.-P. et F. Benzécri, *CAD*, Vol XIV, n°1, pp. 7-34.

Dans la présente notice, les programmes d'analyse des correspondances et de CAH ne seront pas décrits d'abord avec toutes leurs possibilités; mais seulement présentés sur un exemple simple, afin de permettre à un utilisateur de s'accoutumer à l'enchaînement des étapes du traitement, certains compléments étant donnés en addenda.

Les programmes 'qori' et 'CAH' d'analyse et de classification proprement dits comprennent des procédures créant des listages d'aide à l'interprétation et de tracé d'arbre. Les graphiques plans issus de l'analyse des correspondances sont créés par des programmes séparés 'planF' et 'planX'. Le programme 'planF' permet de créer en dialogue et d'afficher à l'écran une grande variété de graphiques, qui peuvent être saisis par copie d'écran pour être ensuite retouchés par un logiciel tel que 'MacPaint' ou 'FullPaint'.

Programmes de création de tableaux: notice d'utilisation, [NOT. CRÉ. TAB.], par J.-P. et F. Benzécri, *CAD*, Vol XIV, n°1, pp. 35-54.

Quelques programmes de création de tableaux suffisent à l'utilisateur non programmeur pour dépouiller par lui-même une enquête par questionnaire, et analyser un ensemble de variables hétérogènes recueillies au laboratoire, ou dans une expérimentation clinique. La présente notice est consacrée à un ensemble de tels programmes.

Le programme 'zrang' sert à ranger, observer et découper les variables d'un tableau: ranger, c'est-à-dire ordonner par valeurs croissantes; observer, en affichant des tableaux de valeurs, des histogrammes, des croisements, voire des cartes; découper, en fixant un ensemble de modalités et créant un tableau de modalités, un tableau binaire, ou un tableau de codage flou (cf. [CODAGE LIN.]).

Le programme 'zBurt' crée le tableau de Burt à partir du tableau des numéros de modalités; 'qorlsup' adjoint à une analyse factorielle un tableau externe de lignes ou de colonnes supplémentaires; 'soustab' crée un sous-tableau d'un tableau donné, en retenant des sous-ensembles spécifiés de lignes et de colonnes; 'discri' à partir de deux tableaux donnant les coordonnées sur les axes

factoriels de deux ensembles d'éléments appelés respectivement ensemble des individus et ensemble des centres, attache chaque individu au centre dont il est le plus proche; 'décale', crée à partir d'un ensemble de séries temporelles un ensemble de séries décalées...

Cette série de programmes, qui est sans cesse complétée, a déjà fait l'objet de nombreuses mises au point: cf. [CODAGE LIN.], [ÉQU. PERS.], [CORREL. JUXT.], [CUM. LI. COL.],...

Programme de représentation cartographique des résultats d'une analyse multidimensionnelle: notice d'utilisation, [NOT. PROG. CART.], par J.-P. et F. Benzécri, *CAD*, Vol XIV, n°1, pp. 55-72.

Dans [NOTE CORR. CAH], on part d'un exemple de géographie économique, le tableau de données croisant l'ensemble des départements de la France avec un ensemble de variables. La présente notice reprend ce même exemple pour montrer comment le programme 'carthage' permet de représenter sur une carte les résultats de la Classification Ascendante Hiérarchique, en tenant compte au mieux des proximités entre départements et variables, révélées par l'analyse des correspondances.

La notice explique également comment créer, par le programme 'litxcarte', un fichier numérique décrivant un fond de carte; et comment modifier l'ensemble des trames disponibles pour couvrir les régions sur la carte.

L'article [STAT. CART.] offre une mise à jour de [NOT. PROG. CART.].

L'analyse des séries chronologiques décalées: principes d'interprétation sur des cas modèles, [MODELE CHRON. DÉCAL.], par A. Quaazeh, *CAD*, Vol XII, n°4, pp.407-418.

Par analyse des séries décalées, nous entendons traitement simultané, en un seul tableau de correspondance, d'une ou plusieurs séries de base $k(t,j)$ et de séries décalées en avance ou en retard : $k(t,j+h)=k(t+h,j)$, $k(t,j-h)=k(t-h,j)$. La méthode est ici appliquée au modèle de la croissance exponentielle, au modèle périodique sinusoïdal, au modèle polynomial ainsi qu'à des combinaisons de modèles ; ces modèles étant étudiés à la fois par l'analyse mathématique et par des simulations numériques sur ordinateur.

Le présent article constitue une notice détaillée d'utilisation pour les programmes 'creseri' et 'decale'.

Le codage linéaire par morceaux: réalisation et applications, [CODAGE LIN.], par J.-P. & F. Benzécri, *CAD*, Vol XIV, n°2, pp. 203-210.

On présente, avec une application à l'analyse des performances de matières plastiques, un type de codage flou désormais réalisé par le programme 'zrang'.

Codage linéaire par morceaux et équation personnelle, [ÉQU. PERS.], par J.-P. & F. Benzécri, *CAD*, Vol XIV, n°3, pp. 331-336.

Le codage de chaque note j suivant trois modalités $\{j+, j=, j-\}$, avec une formule, dite équation personnelle, propre à chaque sujet a déjà servi dans plusieurs études. On montre ici, sur l'exemple du questionnaire objet de l'article

[QUEST. MEM. RECH.] publié dans le même cahier, comment plusieurs programmes nouveaux permettent de choisir des données sans lacune et mettre en œuvre ce codage.

Calculs de corrélation entre variables et juxtaposition de tableaux, [CORREL. JUXT.], par J.-P. & F. Benzécri, *CAD*, Vol XIV, n°3, pp. 347-354.

On montre ici, sur deux exemples, l'un simulé et l'autre extrait de l'article [ÉCHELLES] publié dans le même cahier, l'usage de deux nouveaux programmes du logiciel MacSAIF.

État des recherches sur la cartographie automatique des données statistiques, [STAT. CART.], par J.-P. & F. Benzécri, *CAD*, Vol XIV, n°3, pp. 449-466.

On considère successivement les thèmes traités jusqu'ici en cartographie automatique avec le logiciel MacSAIF (économie, sociologie, géographie électorale); les perfectionnements apportés aux programmes 'carthage' et 'litxcarte' depuis la publication de la notice [NOT. PROG. CART.]; le déroulement d'une étude complexe (agriculture de la Sierra équatorienne) où ces perfectionnements jouent un rôle essentiel.

Programmes de statistique linguistique fondés sur le tri par fusion de fichiers de texte, [LING. TRI.], par J.-P. & F. Benzécri, *CAD*, Vol XV, n°1, pp. 59-82.

Sur l'exemple du texte grec de l'Évangile selon Saint Jean, on montre comment, à partir de la saisie de textes, construire indices, concordances et, finalement, tableaux de correspondance. Il s'agit d'un logiciel nouveau, qui à partir de textes saisis sur un éditeur, donne accès à l'analyse des données telle qu'elle est réalisée par le logiciel 'MacSAIF'.

Extension d'un tableau par adjonction de lignes et colonnes créées par cumul, [CUM. LI. COL.], par J.-P. & F. Benzécri, *CAD*, Vol XV, n°1, pp. 109-116.

On présente un programme de cumul qui, procédant, soit entièrement par dialogue à l'écran, soit avec l'aide d'un fichier texte, sert, notamment, à préparer une analyse discriminante (comme on l'a fait dans l'article [NOUV. TEST. GREC.], publié dans le même cahier).

La nouvelle version de ce programme pose la question suivante:

'faut-il éliminer des cumuls une fraction des éléments O ou N '

si la réponse est 'O' (oui), le programme précise la question:

'éliminer un demi(2), un tiers(3), ou un quart(4) '

si, par exemple, la réponse de l'utilisateur est '3', le programme 'cums' ne réalise pas exactement les cumuls qui lui sont demandés; mais élimine les éléments dont le rang est divisible par 3. Cette variante permet de réserver un échantillon d'épreuve en analyse discriminante: en effet, en bref, les centres de gravité des classes à discriminer sont calculés en n'utilisant que les (2/3) des individus; et l'affectation du tiers restant (i.e. des individus dont le rang est divisible par 3) se fait comme s'il s'agissait de données nouvelles non prises en compte dans l'étude.

3 Indice et références

Afin de donner au lecteur une autre vue d'ensemble de la diversité des programmes et de leurs applications, nous reproduisons ci-dessous l'article 'programme' de l'Indice Systématique du volume XIV de la revue *CAD*.

programme: description et notices d'utilisation: **pr.** d'A. des Correspondances: p. 13 sqq.; **pr.** pour adjoindre des tableaux externes de lignes et colonnes supplémentaires: pp. 48-50; **pr.** d'affichage à l'écran de graphiques croisant deux facteurs: pp. 19-23; **pr.** créant un listage de graphiques: pp. 23-26; **pr.** de C.A.H.: pp.26-34; **pr.** conversationnel de *codage* pour ranger, observer, découper en classes, les variables d'un tableau: pp. 38-47; **pr.** de *codage* linéaire par morceaux: pp. 203-210; **pr.** pour créer un tableau de Burt: pp. 47-48; **pr.** pour créer des sous-tableaux d'un tableau de données ou d'un t. de facteurs: pp. 50-54; **pr.** pour affecter un ensemble d'individus à un ens. de centres (les éléments des deux ens. étant décrits par leurs coordonnées sur les axes issus d'une a. fact.): pp. 52-54; **pr.** de cartographie pour représenter les résultats d'une analyse multidimensionnelle: pp. 55-72, pp. 452-454; **pr.** de création du fond de carte: pp. 59-60; vérification et correction des erreurs: pp.454-459; **pr.** pour juxtaposer ou superposer deux tableaux: pp. 352, 354; **pr.** de calcul de corrélation et son utilisation pour la régression: pp. 348 sqq; **pr.** pour le recadrage, ou le double recadrage, des notes suivant l'*équation* personnelle: pp. 331-336; **pr.** pour recenser les données manquantes dans un tableau: pp. 333 sqq.

Version 2 (1990) du logiciel MacSAIF

Dans les versions antérieures du logiciel, le nombre des facteurs gardés sur un fichier en vue d'effectuer une classification automatique était limité à 29; dans la nouvelle version, ce nombre peut aller jusqu'à 93, ce qui nous paraît ne plus imposer aucune restriction à la pratique de la CAH. Pour réaliser un tel progrès, il a fallu modifier l'entête des fichiers binaires des tableaux; et, afin qu'aucune confusion ne soit possible pour un utilisateur disposant simultanément des versions ancienne et nouvelle des programmes, le nom même des fichiers numériques a été changé: au lieu des suffixes 'ww' et 'zz', on a désormais les suffixes '.w' et '.z'. Par exemple, le fichier de facteurs pour l'ensemble 'i',

résultant de l'analyse du tableau de données 'D:phos', aura pour nom 'D:phosiFac.w'; le tableau de Burt, obtenu après découpage en classes des variables du tableau de données 'D:quest', s'appellera 'D:questB.z'; etc...

À ceci près les modifications se bornent à des améliorations sans influence sur la structure des fichiers; en sorte que les exemples proposés dans les articles de *CAD*, peuvent être repris avec la nouvelle version, en suivant pas à pas les étapes indiquées.

Données et exemples

Nous rappelons que sont fournis, avec le logiciel Mac-SAIF, un dossier 'atlas', qui contient des fonds de carte (France, Paris, Grèce, Italie, Portugal); et plusieurs dossiers d'études ayant fait l'objet d'articles de *CAD*; nous conseillons à l'utilisateur de copier ces dossiers sur disque, de les consulter pour se pénétrer des résultats, puis de reprendre les étapes successives des calculs afin de maîtriser l'enchaînement des programmes.

N.B. Les lecteurs désireux d'acquérir les programmes du Logiciel **Mac-SAIF**, ('Système d'Analyse des InFormations'), décrits dans la présente note s'adresseront à la

Société **STATMATIC**: 4, rue de Fécamp 75012 Paris;
Téléphone: (16.1) 43.42.48.19 / (16.1) 47.98.77.39