

# STATISTIQUE ET ANALYSE DES DONNÉES

PHILIPPE BESSE

ANTOINE DE FALGUEROLLES

## **Le choix de logiciels statistiques**

*Statistique et analyse des données*, tome 12, n° 1-2 (1987), p. 154-170

[http://www.numdam.org/item?id=SAD\\_1987\\_\\_12\\_1-2\\_154\\_0](http://www.numdam.org/item?id=SAD_1987__12_1-2_154_0)

© Association pour la statistique et ses utilisations, 1987, tous droits réservés.

L'accès aux archives de la revue « Statistique et 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/>

LE CHOIX DE LOGICIELS STATISTIQUES

Philippe BESSE, Antoine de FALGUEROLLES  
Laboratoire de Statistique et Probabilités  
U. A. CNRS 745  
Université Paul Sabatier  
118, route de Narbonne  
31062 TOULOUSE CEDEX

Résumé :

*Les logiciels statistiques deviennent de plus en plus accessibles à des utilisateurs de formation informatique ou statistique réduite. L'objet de cet article est de donner quelques éléments susceptibles d'éclairer le choix d'un statisticien confronté à une offre abondante de logiciels.*

Abstract :

*Statistical computer programs are becoming increasingly accessible to people with limited computer or data training. The aim of this paper is to provide some elements to be considered in the decision-process of the statistician looking for packages.*

Mot-clé : logiciel statistique.

Indices de classification STMA : 15070 15090.

## INTRODUCTION

Contrairement aux autres domaines des mathématiques appliquées qui font surtout appel à des bibliothèques scientifiques générales (IMSL, NAG), le développement des statistiques et, plus récemment, de l'analyse exploratoire des données suscite l'élaboration de logiciels spécifiques car les manipulations de données et la présentation des résultats, dans ce domaine, prennent autant d'importance et nécessitent souvent beaucoup plus de code (d'instructions) que la simple exécution des algorithmes de calcul. Par ailleurs, renforcée par la banalisation des micros-ordinateurs, la tendance est d'ouvrir à des utilisateurs peu informaticiens voire peu statisticiens le champ des méthodes statistiques. Ceci entraîne la conception et la diffusion d'un nombre impressionnant de produits. Corrélativement apparaissent dans la littérature des articles proposant des catalogues de logiciel de plus en plus fournis (NEFFENDORF 1983, CHANDON 1985 et SIEGEL 1985 pour les micro-ordinateurs), des comparaisons ou évaluations de logiciels (BERNARD 1978, FRANCIS 1983, LAURO et SERIO 1982, Mc CALLUM 1983, WETHERHILL et CURRAM 1985, FRIDLUND A. J. 1986 et dans la revue "American Statistician"), des réflexions sur la formation de base des futurs utilisateurs (BESSE et FALGUEROLLES 1985), leurs comportements (MOLENAAR 1984) ou sur la place de l'informatique dans le métier de statisticien (VICTOR 1984, GOWER 1985). A signaler aussi la publication d'algorithmes pour les méthodes statistiques (NASH 1979, KENNEDY et GENTLE 1984 et la rubrique "statistical algorithms" de la revue Applied Stat.).

Le but de cet article n'est pas d'offrir un nouveau catalogue de logiciels ni de chercher à faire une étude comparative; les produits cités ne le sont qu'à titre d'exemple sans soucis d'exhaustivité. On se propose simplement de jeter quelques jalons afin de guider le futur utilisateur parmi une offre abondante mais disparate et de rappeler quelques contraintes liées aux choix tant conceptuels que matériels auxquelles tout utilisateur doit être sensibilisé. On adoptera donc le point de vue de l'utilisateur qui, devant préciser ses besoins, est plus intéressé par les fonctionnalités du logiciel qu'il va mettre en oeuvre, que par la façon dont elles sont élaborées. On s'efforcera cependant de signaler, le cas échéant, le poids de certaines contraintes matérielles ou choix conceptuels. Le lecteur intéressé par l'optique du concepteur peut se reporter, par exemple, à CHAMBERS 1980, NELDER 1984, BARRA et BECKER 1985 ou LEMAIRE 1985.

## 1. PORTABILITE, COMPATIBILITE

De toute évidence, l'évolution et le raffinement des logiciels d'application sont strictement dépendants de l'évolution de la capacité des matériels (taille des mots, de la mémoire vive, de la mémoire de masse, vitesse d'exécution, ...) ainsi que du développement des possibilités offertes par les systèmes d'exploitation (fichiers à accès multicritère, mémoire virtuelle, arborescence de catalogues, "pipes", interactivité, ...). Malheureusement, l'absence de normes internationales (COMPSTAT 1984), dans le domaine de l'informatique appliquée aux statistiques, impose des limites sévères aux concepteurs soucieux de réaliser des produits portables et donc, en principe, indépendants d'un constructeur. Cette portabilité est assurée en adoptant un des trois niveaux de compatibilité les plus fréquemment rencontrés:

i) Le premier niveau de compatibilité historiquement apparu est associé à un langage de programmation. La grande majorité des logiciels scientifiques sont écrits en fortran standard (ANSI66 puis ANSI77). Malgré son caractère archaïque et ses possibilités limitées, le choix de ce langage est motivé par la disponibilité de compilateurs pour tous les types de machines. Mais il est alors impossible d'utiliser ou d'intégrer toutes les fonctionnalités offertes par les systèmes d'exploitation sans les réécrire en fortran comme cela a été fait, par exemple, dans SYSTAT, pour un éditeur plein écran ou sans l'adjonction de procédures en code machine donc non portables.

ii) Le deuxième niveau de compatibilité se superpose au précédent et tient à l'utilisation d'un même système d'exploitation, par exemple celui de MICROSOFT ou encore UNIX système V dont un standard semble maintenant se dégager sur différents types de machines.

iii) Le troisième niveau repose sur une "identité" de matériel. Actuellement il concerne essentiellement la "norme" qui s'est commercialement imposée autour de la famille des IBM PC/XT ou AT et compatibles. Ceci permet alors une utilisation optimale des capacités de la machine et, en particulier, de ses extensions mémoires ou de ses possibilités graphiques mais nécessite, de la part du concepteur, une bonne connaissance de la programmation du processeur.

## 2. FONCTIONS "STATISTIQUES"

On ne considerera pas ici les tableurs et autres logiciels de "décideurs" dont l'objectif n'est pas explicitement de produire des statistiques. Ce faisant il convient néanmoins de reconnaître qu'ils sont sur ce plan de plus en plus élaborés (JAVELIN) et peuvent prétendre devenir des concurrents sérieux par leur qualités d'intégration (traitement de texte, base de donnée, communication, graphiques) et de commodité d'emploi.

En général les logiciels statistiques comportent quatre couches fonctionnelles plus ou moins développées d'un produit à l'autre que l'on explicite ci-après.

### a. la gestion des données

Ce terme générique regroupe tout un ensemble de commandes de manipulations des données qui s'avèrent indispensables pour une mise en oeuvre rapide et efficace des méthodes statistiques:

- saisie en conversationnel des données ou à partir d'un fichier ASCII standard, modification, sauvegarde.
- transformations (log, racine, ...) et codage en classes des variables quantitatives, recodage des variables qualitatives.
- sélection d'individus, de variables à l'aide d'opérateurs logiques et d'instructions gérées par un interpréteur plus ou moins élaboré.
- création de nouvelles variables par combinaisons de celles existantes.
- intégration ou interfaçage d'une base de données (ROOS 1985), lecture de fichiers issus de logiciels classiques (DBASE, LOTUS, SQL, ...).
- ...

### b. les statistiques élémentaires

On considère ici les indices ou les graphiques de la statistique descriptive classique ou exploratoire (TUKEY 1977). L'aspect graphique est ici souvent essentiel dans cette approche des données. Ce point de vue est un élément important dans la comparaison des logiciels.

### c. les méthodes statistiques

Ce point fait apparaître une grande diversité car chaque logiciel a été développé dans le cadre d'un domaine d'application particulier (psychométrie, biométrie, sciences sociales, économétrie, enseignement, ...) ou autour de méthodes statistiques particulières (modèle linéaire généralisé, analyse exploratoire, analyses factorielles, analyse de variance, classification, séries chronologiques, ...). C'est pourquoi, il est indispensable qu'un logiciel prévoit une interface, par l'intermédiaire de fichiers ASCII standards, afin de pouvoir facilement utiliser une technique statistique plus adaptée seulement disponible sur un autre produit.

### d. les sorties

Deux types de stratégies se rencontrent fréquemment, la première, héritière des habitudes du traitement en différé, propose l'édition exhaustive des résultats fournis par une méthode, alors que d'autres logiciels, plus interactifs, ne fournissent que les résultats explicitement demandés. Cette dernière pratique constitue au moins un garde-fou contre des impressions systématiques souvent inutilement volumineuses.

Mais, dans les deux cas, un point important est la qualité graphique des sorties. Celui-ci est souvent négligé ou laissé de côté à cause des contraintes matérielles. Si la plupart des produits proposent des histogrammes et des "biplots" sur imprimante, il est rare de pouvoir utiliser facilement un traceur ou une imprimante de qualité. Le coût du matériel est un problème mais c'est aussi l'absence de normes qui bloque les investissements de la part des concepteurs ou des utilisateurs.

## 3. FONCTIONS UTILITAIRES

Des aspects autres que statistiques apparaissent lors de l'utilisation d'un logiciel et notamment leur facilité d'accès. Les types de mise en oeuvre peuvent schématiquement s'ordonner par degré croissant de sophistication ce qui ne signifie pas difficulté croissante d'utilisation:

- Les programmes "batch" (traitement différé)
- les programmes dirigés par menus
- les interprètes de commandes
- les "langages statistiques"

a. "conversationnalité"

Les premiers logiciels, élaborés du temps maintenant révolu des cartes perforées et du traitement exclusivement différé, évoluent mais se trouvent concurrencés par des produits offrant toujours plus d'interactivité. Dans le premier cas, l'aspect conversationnel est souvent très limité et n'autorise que la constitution d'un fichier de paramètres définissant la liste des opérations à effectuer sans aucun contrôle possible de la part de l'utilisateur au cours de l'exécution.

A l'opposée, en phase expérimentale ou de mise au point d'un traitement statistique, une interactivité élaborée permet d'éviter beaucoup de calculs et d'éditons inutiles par un contrôle en temps réel des résultats intermédiaires et par la possibilité, à tout moment, de choisir la suite des opérations, revenir en arrière, modifier des paramètres ...

D'autres produits (LEAS, STAT-ITCF ou en option dans SICLA) proposent des menus arborescents qui interrogent l'utilisateur sur tous les choix et paramètres nécessaires. Cette optique est sans doute pertinente en utilisation pédagogique pour des néophytes mais devient très vite lourde et fastidieuse une fois le premier apprentissage acquis. Dans certains logiciels, cette contrainte est assouplie par la présence d'options par défaut: STATGRAPHICS et SHADOCK V.S. qui permet la réédition du fichier des commandes créé auparavant par des menus déroulants.

b. chaînage des analyses

Il est fréquent d'avoir à enchaîner différentes méthodes: représentation graphique de résidus à un modèle, classification sur des facteurs, ... Ce n'est possible que lorsque le logiciel gère données et résultats sous forme de structures standardisées (programmation orientée objet). Il est cependant regrettable qu'aucune norme ne se soit dégagée pour faciliter les chaînages entre logiciels (COMPSTAT 1984). Aussi l'utilisateur a souvent besoin de connaître quelques rudiments de programmation (fortran ou autre) pour réaliser ces interfaces.

D'autre part, si l'interactivité présente de nombreux avantages sur une machine rapide ou avec de petits jeux de

données, elle devient rédhibitoire lorsque le temps de réponse dépasse quelques minutes; aussi, le chaînage de différentes étapes doit pouvoir être automatisé. Ceci se fait par la constitution d'un fichier de paramètres ou encore par l'écriture d'un "programme" dans un langage interprété propre au logiciel.

#### c. commandes interprétées

Les langages interprétés (par opposition à compilés) intégrés à des logiciels "batchs" ou interactifs permettent d'enchaîner des commandes propres aux outils statistiques manipulés. La syntaxe peut en être plus ou moins modulaire et faire appel à des structures plus ou moins complexes (test, itérations). Cette caractéristique est utile pour exécuter une même séquence de traitements sur différents jeux de données. Elle est d'autant plus pratique que la liste de ces commandes peut être directement intégrée au logiciel sous la forme d'une macro-commande (GLIM).

#### d. langage statistique

Certains produits offrent de plus la possibilité d'opérer directement sur des matrices ou tableaux de données avec une syntaxe comparable à celle d'un langage évolué pour des scalaires (SAS/IML, MATLAB, GAUSS, SYSTEME S, PC-ISP, ...). Il faut cependant distinguer les langages qui ne font qu'interpréter (SAS/IML, ISP, MATLAB) et dont l'exécution peut être longue et coûteuse de ceux qui permettent de sauvegarder des procédures compilées avec un code efficace (GAUSS).

L'intérêt de tels produits est double. Ils constituent d'abord, pour le statisticien testant une nouvelle méthode, des outils très efficaces car proches de sa formulation mathématique. Ils facilitent ensuite la diffusion de ces nouvelles méthodes par l'environnement statistique et fonctionnel qu'ils garantissent.

#### e. documentation

Une documentation complète et détaillée est évidemment indispensable au bon apprentissage puis usage d'un logiciel; malheureusement elle est souvent volumineuse, difficilement accessible (déposé à la bibliothèque du centre de calcul ou encore empruntée par un collègue ou périmée ou ... donc



indisponible juste au moment où un problème surgit inopinément). La seule parade est de pouvoir disposer d'une documentation "en ligne" accessible à tous les niveaux d'utilisation. Un logiciel auto-documenté affiche alors, à la demande, la liste des commandes et mots clefs disponibles ainsi qu'une description de leur syntaxe ou encore offre la possibilité d'exécuter une succession de menus arborescents.

#### 4. FONCTIONS SYSTEMES

##### a. extensions

Le paragraphe 3.c précédent pose le problème de l'"ouverture" d'un logiciel, c'est-à-dire de la possibilité pour l'utilisateur de modifier ou d'ajouter certaines procédures. Les modifications sont dépendantes de la disponibilité du code source (fortran ou autre), ce qui est rarement le cas pour des produits commerciaux. Elles sont de toutes façons délicates. La solution consistant à ajouter des procédures spécifiques (méthode personnelle, interface vers un autre logiciel, sortie graphique...) est préférable si la possibilité en est explicitement prévue par le concepteur et clairement décrite dans la documentation (procédures en APL dans STATGRAPHICS).

D'autre part un logiciel n'est jamais fini, il comporte toujours quelques bogues et des compléments ou améliorations sont indispensables pour préserver sa compétitivité. Aussi est-il important que le contrat passé avec le diffuseur prévoit une maintenance, c'est-à-dire l'accès aux corrections et versions successives.

##### b. communication avec le système

Il est très utile de pouvoir communiquer avec le système d'exploitation ou, plus simplement, lancer des commandes de l'intérieur du logiciel afin, par exemple, de lister des fichiers, faire des copies, des assignations ou encore entrer dans un éditeur. C'est rarement possible car ceci suppose que le logiciel s'exécute au même niveau que le système d'exploitation ou que la communication ait été prévue entre les deux niveaux. La conséquence en est une plus grande dépendance vis à vis du support et donc une moindre portabilité sauf lorsque celle-ci repose sur une compatibilité suffisante (comme UNIX ou IBM PC).

c. volume des données

On s'intéresse ici à la taille maximum du tableau de données susceptible d'être traité. Différentes stratégies ont été élaborées : ou bien les données restent sur disque, ou bien tout est chargé en mémoire centrale. Dans le premier cas le nombre d'individus-lignes est théoriquement illimité. Dans le second cas, le nombre de lignes et le nombre de colonnes ou le produit des deux sont limités par les dimensionnements initiaux (en fortran) des programmes et impérativement par la place offerte par le système d'exploitation.

Le traitement de gros fichiers (plusieurs milliers de lignes) impose d'adopter la première approche; cela pénalise évidemment l'aspect conversationnel d'un logiciel car les temps d'exécution sont sévèrement majorés par les accès disques. Evidemment la deuxième approche entraîne une limitations de la taille du tableau mais offre plus de rapidité, de souplesse et souvent de possibilités. Une façon de résoudre le problème consiste à offrir les deux possibilités de traitement en option. L'utilisateur choisit alors, en fonction de ses données et des capacités de sa machine.

Dans les deux cas il est souhaitable de pouvoir disposer d'un dimensionnement dynamique (impossible en fortran) afin d'éviter des compilations intempestives ou la réservation d'une place démesurée pour la gestion de l'espace de travail. La limitation serait alors celle imposée par le système d'exploitation indépendamment du logiciel.

## 5. CONTRAINTES MATERIELLES

Les possibilités des micro-ordinateurs (mémoire, puissance de calcul) évoluent très rapidement. Ceux-ci permettent d'entreprendre des analyses statistiques même volumineuses à un coût bien moindre et avec plus de souplesse que sur un gros système dont l'utilisation n'est justifiée que pour des très grands tableaux de données ou des logiciels spécifiques.

a. mémoire vive

En attendant la généralisation des microprocesseurs 32 bits et d'un système d'exploitation gérant une mémoire virtuelle, la limitation de la taille de la mémoire

- est très pénalisante pour les micros 8 bits, dont le basic ne peut adresser que 64 ko,
- demeure réelle pour les micros 16 bits dont les versions actuelles de MS-DOS ne peuvent gérer que 640 ko,
- ne sera plus très contraignante avec le successeur (OS/2) de MS-DOS qui permettra de gérer plusieurs mega-octets.

b. mémoire de masse

Le code exécutable des logiciels statistiques prend une place importante. C'est notamment le cas pour des adaptations sur micro de produits initialement conçus pour gros systèmes. Aussi la présence d'un disque dur, parfois impérative, est souhaitable afin d'éviter des manipulations acrobatiques de disquettes et cela améliore nettement les temps d'exécution lors du traitement de gros fichiers car les accès sur le disque sont beaucoup plus rapides.

c. coprocesseur

La présence d'un coprocesseur arithmétique gérant les calculs en virgule flottante améliore très sensiblement les temps d'exécution. Imposée pour certains logiciels, elle est recommandée dans la plupart des cas.

d. graphique

La bonne exécution d'une application graphique nécessite la compatibilité de trois éléments: le logiciel, la carte extension graphique et le moniteur, qui doivent accepter un standard commun. Trois "normes" se rencontrent couramment: C.G.A. monochrome ou couleur mais dont la définition (640x200) est insuffisante pour les graphiques rencontrés en statistique, Hercules monochrome (720x348) et E.G.A. couleur (640x350). Avec la série PS/2, IBM propose un nouveau standard (V.G.A.) qui pourrait se généraliser.

6. FORMATION

Tout logiciel, aussi convivial ou "amical" soit-il nécessite une période plus ou moins longue d'apprentissage. La formation d'un utilisateur est un investissement important, souvent plus important que le simple achat du produit; elle

concerne des domaines variés afin de permettre:

- une utilisation rationnelle de l'outil informatique,
- une bonne compréhension des méthodes statistiques et de leurs limites,
- une certaine maîtrise des méthodes numériques utilisées,
- une mise en oeuvre efficace du logiciel.

De toute évidence certains logiciels sont plus adaptés aux débutants que d'autres mais encore faut-il cerner le champ des lacunes (statistique ou informatique ou les deux) et le type d'utilisation (occasionnelle ou à "temps complet"). Il est bon de se renseigner sur les possibilités de stage, l'existence de manuels pédagogiques ou de clubs d'utilisateurs. Enfin, et c'est là un point sensible, il est préférable de pouvoir tester en vraie grandeur un logiciel avant de l'acquérir. La tentation de copiage illicite, préalablement à une éventuelle acquisition, est réelle. Pour éviter les abus d'une telle pratique certains concepteurs proposent des versions "bridées" de démonstration de leur produit. Une autre approche, qui semble se dessiner, consisterait à proposer des stages de formations incluant des essais comparatifs.

## 7. LE FUTUR

Même s'il est en rapide progression grâce à l'expansion de la micro-informatique, le marché du logiciel statistique est relativement restreint car très spécifique. La statistique bénéficie des retombées technologiques de progrès fondamentaux (logiciels ou matériels) réalisés dans d'autres domaines plus qu'elle ne les suscite. En suivant l'actualité informatique, on peut tenter de prévoir, sans trop de risques, quelles sont les avancées actuelles qui auront des conséquences importantes sur la conception et l'usage des logiciels statistiques:

- Des processeurs de plus en plus rapides associés au développement du calcul parallèle ou vectoriel améliorent très sensiblement les vitesses d'exécution et donc ira dans le sens de plus d'interactivité. Ceci nécessite au préalable de reconsidérer les algorithmes des méthodes numériques pour les adapter au parallélisme (LAFAYE de MICHEAUX 1984).

- Des outils graphiques se développent rapidement avec l'expansion de la C. A. O. et du traitement d'images mais, comme ils nécessitent des moyens de calcul importants, on les rencontre surtout sur des stations de travail plutôt que sur des micro-ordinateurs. Des normes s'imposent autour de GRAPHIGS ou GKS.

- Les inévitables systèmes experts (NELDER 1984, DEMONCHAUX et al. 1985, PREGIBON et GALE 1984, HAHN 1985, DAMBROISE et MASSOTTE 1986, LEMAIRE 1986) peuvent trouver un champ d'application important en statistique en proposant des systèmes d'aide adapté à l'utilisateur, à ses besoins et au contexte dans lequel il évolue. Cette aide peut intervenir sur le plan informatique, ou sur celui des méthodes statistiques en tenant compte de ce qui a déjà été exécuté par l'utilisateur et éventuellement en l'interrogeant sur ses objectifs. Un tel outil remplacera très avantageusement des documentations volumineuses ou des menus arborescents fastidieux.

- En ce qui concerne les systèmes d'exploitation, la plupart des stations de travail adoptent UNIX ou l'un de ses clones qui est également disponible sur la famille des PC/AT (XENIX de MicroSoft) et annoncé par IBM pour le haut de gamme de la série PS/2 à base de processeurs 80386. Ce système d'exploitation ne supplantera pas MS.DOS ou son successeur OS/2 mais se répand très rapidement pour les applications scientifiques. Il semble que l'on tende vers une cohabitation, gérée par un superviseur, de ces deux standards sur une même machine. Ceci permettra une meilleure intégration d'outils développés hors cadre statistique (gestion de bases de données par exemple) dans un fonctionnement multi-fenêtres et multi-tâches.

- Si le fortran reste utilisé à travers l'emploi de bibliothèques scientifiques (NAG, IMSL, ...), le développement de nouveaux produits fera appel à des langages alliant efficacité humaine (en phase de mise au point ou de mise en oeuvre) et efficacité machine (en phase d'exploitation). C'est par exemple TURBO PASCAL (LEMAIRE 1985) pour les micros actuels, le langage C associé à UNIX ou, dans un proche avenir, ADA armé pour la parallélisation des tâches.

Par ailleurs, toujours pour des raisons incontournables "d'efficacité humaine", il est probable que le développement des logiciels statistiques se hiérarchise en deux niveaux:

1) réalisation de langages statistiques efficaces (de "quatrième génération") par des "informaticiens"; LOM (SCHEKTMAN 1982) est resté un produit trop confidentiel sous Siris8; GAUSS, PC-ISP, SAS/IML sur micro et SYSTEM S sous UNIX sont très bien positionnés en ce sens;

ii) application de ces outils au développement de méthodes statistiques ouvertes et flexibles ou au moins de

prototypes par des "statisticiens" ainsi dégagés des bogues et des contraintes informatiques.

Enfin l'évolution de l'informatique impulse le développement de nouvelles méthodes statistiques largement calculatoires et essentiellement graphiques (HUBER 1985, YOUNG 1985, DONOMO 1986) fondées sur l'exploration dynamique de nuages de point ou sur la recherche de "Projections Révélatrices" (Projection Poursuit).

## CONCLUSION

Le choix d'un logiciel statistique apparaît donc être un problème délicat. Il doit tenir compte du matériel utilisé, de la finalité de l'exploitation et du niveau de formation ou de motivation de l'utilisateur. Pour ces derniers points TUKEY, 1985 distingue :

- 1. the very innocent with zero background,*
- 2. the innocent, trained to use a similar system about five years ago,*
- 3. the casual user, who uses it about four times in a year*
- 4. a regular user without specific training in statistics, and*
- 5. a regular user, who is a journeyman data analyst with a good statistical background.*

Cet aspect est déterminant car un utilisateur a ses habitudes, son passé, un type de problématique. Aussi adoptera-t-il, souvent a priori, le logiciel qui a été développé dans un domaine proche de son champ d'investigation (biométrie, sociométrie, psychométrie, économétrie, gestion de la qualité ...).

Ce peut être là une démarche raisonnable à condition d'éviter l'écueil important consistant à se laisser enfermer dans UN logiciel par crainte d'avoir à subir une nouvelle formation et l'investissement qu'elle suppose. L'objectif global, qui reste l'efficacité humaine, ne doit plus passer par une pratique virtuose d'un outil, comme cela a été vrai pour le fortran ou pour le basic, mais de l'emploi standard du bon outil parmi la vaste panoplie qui s'offre maintenant au statisticien.

Références:

J. R. Barra et M. Becker (1985) : "*conception d'un logiciel d'assistance - intelligence en analyse factorielle et canonique*"

Edition provisoire des actes des quatrièmes journées internationales "Analyse de Données et Informatique". INRIA, tome 2, pp.739-750.

G. Bernard (1978) : "*quelques réflexions sur des logiciels pour l'analyse statistique - comparaison de GENSTAT, BMDP et SPSS*"  
Statistique et Analyse des Données, vol. 3, N°1, pp.53-62.

Ph. Besse et A. de Falguerolles (1985) : "*exploration interactive d'un nuage de points dans le plan*".

Bulletin de l'APMEP, N° 347, février, pp.21-29.

P. M. Caporal and G. J. Hahn (1984) : "*general software for statistical graphics - a survey*".

Statistical Software Newsletter, vol.10, N°1, pp.3-13.

J. M. Chambers (1980) : "*statistical computing : history and trends*".

The American Statistician, vol. 34, N°4, pp.238-243.

J. L. Chandon (1985) : "*panorama des logiciels français d'analyse des données sur micro-ordinateurs*".

Edition provisoire des actes des quatrièmes journées internationales "Analyse de Données et Informatique". INRIA, tome 2, pp.709-720.

COMPSTAT 84 : "*standard packages versus tailor made software - panel discussion at COMPSTAT 84 in Prague*".

Statistical Software Newsletter, vol.10, N°2, pp.56-67.

E. Dambroise et P. Massotte (1986) : "*MUSE : an expert system in statistic*".

in COMPSTAT 86, Physica-Verlag, Vienne, pp.271-276.

E. Demonchaux, J. Quinqueton, H. Ralambondrainy (1985) :

"CLAVECIN : un système expert en analyse de données".  
Rapport de Recherche N°431, INRIA, Rocquencourt.

D. Donoho (1986) : "two talks on data analysis".  
Communication aux Journées de Statistique de Lille 1986

I. Francis (1983) : "a survey of statistical softwares".  
Computational Statistics and Data Analysis, vol.1, N°1,  
pp.17-27.

J. C. Gower (1985) : "the developement of statistical computing at  
Rothamsted".  
In Rothamsted Report for 1985, Part 2, pp.221-235.

G. H. Hahn (1985) : "more intelligent statistical software and  
statistical expert systems : future directions".  
The American Statistician, vol.39, N°1, pp.1-8.

P. J. Huber (1985) : "Projection Pursuit", with discussion.  
The Annals of Statistics, Vol.13, N°2, pp.435-525.

J. Jida et J. Lemaire (1986) : "expert systems and data analysis  
package management".  
in COMPSTAT 86, Physica-Verlag, Vienne, pp.251-258.

W. J. Kennedy et J. E. Gentle (1984) : "statistical computing"  
(Marcel Dekker inc., New-York, 1980).

D. Lafaye de Micheaux (1984) : "parallelization of algorithms in  
the practice of statistical data analysis".  
In COMPSTAT 84, Physica-Verlag, Vienne, pp.293-300.

N. Lauro and G. Serio (1982) : "criteria for evaluating and  
comparing statistical software : a multidimensional data  
analysis approach".  
Statistical Software Newsletter, vol.8, N°3, pp.102-119.

J. Lemaire (1985) : "choix informatiques du concepteur de  
logiciels d'analyse de données sur micro-ordinateurs".  
Edition provisoire des actes des quatrièmes journées  
internationales "Analyse de Données et Informatique". INRIA,  
tome 2, pp.699-708.



I. W. Molenaar (1984) : "*behavioral studies of the software user*".

Computational statistics and Data Analysis, 2, pp.1-12.

M. E. Muller (1980) : "*aspects of statistical computing : what packages for the 1980's ought to do*".

The American Statistician, vol.34, N°3, pp.159-168.

J. C. Nash : "*compact numerical methods for computers : linear algebra and function minimisation*".

(Adam Hilger Ltd, Bristol, 1979).

H. Neffendorf (1983) : "*statistical packages for microcomputers: a listing*".

The American Statistician, vol.37, N°1, pp.83-86.

J. A. Nelder (1984) : "*present position and potential developments : some personal views on statistical computing*".

J. R. Statist. Soc. A, vol.147, part 2, pp.151-160.

D. Pregibon et W. A. Gale (1984) : "*REX : an expert system for regression analysis*".

In COMPSTAT 84, Physica-Verlag, Vienne, pp.242-248.

J. L. Roos (1985) : "*recherche et traitement des données dans une base de données statistique orientée objet*".

In Selected Papers of the 45th-ISI-session, édité par l'IASC, pp. 301-316.

Y. Schektman et ali (1982) : "*computer science as a tool improving data analysis researches and uses*".

In COMPSTAT 82, Physica-Verlag, Vienne, pp.103-112.

J. B. Siegel : "*statistical software for microcomputers*".

(North-Holland, Amsterdam, 1985).

J. W. Tukey : "*exploratory data analysis*"

(Addison Wesley, Reading, Mass, 1977).

J. W. Tukey (1985) : "*comments [on the article of G. H. Hahn (1985)]*".

The American Statistician, vol.39, N°1, pp.12-14.

N. Victor (1984) : "*statistique informatique : science ou outil?*"  
Statistique et Analyse des Données, vol. 9, N°1, pp. 75-101.

G. B. Wetherill et J. B. Curram (1985) : "*the design and evaluation  
of statistical software for microcomputers*".  
The statistician, vol. 34, N° , pp. 391-427.

F. W. Young (1985) : "*new developments in describing and  
visualizing multivariate data*".  
To appear in 1985 SUGI Proceedings, SAS Institute, Cary, NC.