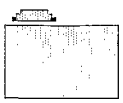
Statistique et ordinateur : un mariage de raison



Souvent considérée comme aride, l'analyse des données statistiques, associée à notre ordinateur, montre une très grande fécondité qui s'exprime par la grande diversité de l'offre du marché.

La statistique est, sans doute, l'une des premières branches d'application de l'informatique. Dès les années 1960-70, plusieurs logiciels de haut niveau font leur apparition sur les gros systèmes IBM. Citons, par exemple, BMDP, dans le domaine biomédical, et SPSS, pour les sciences sociales. Malgré une très grande puissance de calcul, l'utilisation de ces logiciels apparaît difficile et lourde, conduisant le statisticien, ou plus simplement le détenteur de données, à se transformer en «pseudo-informaticien». De plus, à cette époque de règne de la carte perforée, interactivité et convivialité ne font pas encore partie du vocabulaire de base de l'informatique.

A l'avènement des micro-ordinateurs, on n'assiste malheureusement pas à une diffusion de la statistique vers ces nouvelles machines encore trop modestes. En effet, les capacités de stockage et les vitesses de calcul restent trop réduites pour pénétrer dans un domaine professionnel où s'est établie une tradition d'informatique lourde. Par contre, dans le domaine de l'enseignement de la statistique, plusieurs «bricolages de génie» voient le jour. Programmés en FORTRAN, BASIC ou LSE, ils ne méritent pas encore le nom de logiciel, mais rendent de nombreux services dans les facultés encore mal équipées.

L'apparition, au début des années 1980, de micro-ordinateurs plus performants comme l'IBM PC et le Macintosh conforte le développement de ces nouvelles technologies : les champs d'applications se diversifient. Pourtant, la statistique continue à rester en marge de ces mutations jusqu'à l'émergence des mémoires de masses de grandes capacité (disques durs) et la diffusion de micro-processeurs plus rapides. A partir de 1985, des versions pour micro-ordinateur des logiciels pour gros systèmes comme BMDP, SPSS, MINITAB ou SAS deviennent disponibles aux côtés d'autres logiciels développés directement sur PC comme MICROSTAT, STATGRAPHICS, SYSTAT, etc.

Les nombreux logiciels disponibles sur le marché ouvrent un vaste champ d'étude pour qui veut extraire la «substantifique moelle» d'un tableau de nombres. En effet, on dispose d'une grande variété de méthodes qui permettent d'envisager sérieusement le dépouillement d'enquêtes statistiques de plusieurs centaines d'individus et de quelques dizaines de variables.

Pour justifier le terme «statistique», un logiciel d'analyse des données (statisticiel) doit assurer un nombre minimum de traitements usuels comme l'analyse de la variance, la régression multiple ou bien encore l'analyse factorielle. De plus, un grand nombre de méthodes statistiques considèrent que le tableau de données à analyser n'est en fait qu'un échantillon extrait d'une population plus large; les paramètres calculés doivent donc être assortis de tests de significativité qui donnent un seuil de confiance au delà duquel les valeurs calculées ne peuvent être dues à des fluctuations aléatoires d'échantillonnage.

Ainsi, les tableurs du genre *Excel* ou les grapheurs comme *CricketGraph* ou *MacSpin* ne peuvent pas être considérés comme des logiciels d'analyse statistique, bien qu'ils offrent la possibilité d'examiner des données numériques et même de les représenter sous forme d'histogrammes ou de diagrammes bivariés. Mais leur domaine d'application est limité à la simple description uni- ou bivariée, sans aucune possibilité d'inférer les résultats obtenus sur échantillon à l'ensemble de la population.

Si l'on écarte ces «croqueurs de nombres» (les «numbers crunchers» américains), les logiciels d'analyse statistique pour Macintosh demeurent assez nombreux. Il faut distinguer les logiciels généraux, ceux qui couvrent un large spectre d'applications, de ceux, plus spécialisés, qui répondent à des besoins particuliers, comme ceux des économètres, des laboratoires ou des études de marché. Les prix sont en général assez élevés, de 2 000 à 6 000 francs en moyenne, les différences s'expliquant (en partie seulement) par la richesse en techniques d'analyse, la vitesse et la capacité de calcul.

Les statisticiels diffèrent sur de nombreux points, et en premier lieu sur le plan de leur conception générale. D'une part, les logiciels gouvernés par des menus (*Statview II*, par exemple), qui sont très faciles d'accès pour des usagers occasionnels : une séquence

composée de choix dans un ou plusieurs menus permet de sélectionner méthodes statistiques et options complémentaires. Naturellement, cette approche guide l'utilisateur et, si le système est bien conçu, limite considérablement les possibilités de choix erronés; le prix à payer pour cette assurance s'élève à une limitation parfois contraignante des possibilités de traitement. A l'opposé, on trouve les logiciels gouvernés par un langage de commande (comme *Systat*) proche du langage de programmation BASIC. Le statisticien professionnel, ayant souvent l'habitude de la conception de programmes, appréciera la possibilité de jongler avec les instructions et d'adapter le logiciel à ses besoins particuliers. De plus, en conservant son programme dans un fichier, il pourra recommencer périodiquement la même séquence d'instructions, sur des données mises à jour. Le prix à payer pour cette liberté se monte au temps nécessaire à l'apprentissage du langage de commande du logiciel, souvent plusieurs heures pour un programmeur expérimenté.

Une seconde différence très importante réside dans l'adoption par certains logiciels (*Data Desk*, par exemple) de la méthode (au sens plein du terme) connue sous le nom américain «Exploratory Data Analysis» (EDA) ou analyse exploratoire des données. Proposée par le statisticien J. Tuckey, EDA cherche à prendre en compte les anomalies ou les cas extrêmes. Contrairement aux pratiques statistiques classiques, l'analyse exploratoire ne cherche pas l'adhésion quasi-rituelle au test d'hypothèse et à la prise de décision de type probabiliste. Elle est moins normative et peut très bien s'intégrer dans un processus de recherche mélangeant les deux approches; ainsi, par complémentarité, elle peut jouer un rôle exploratoire pour «radiographier les données» et isoler un problème qui, par la suite, se traitera par des méthodes moins intuitives. On est ainsi conduit à utiliser de manière extensive les représentations graphiques, souvent en combinant plusieurs modes de visualisation, et cela de manière interactive par un retour constant aux tableaux d'origine.

En regardant les données depuis des perspectives variées, on maîtrise mieux les relations entre les variables et l'on reconnaît plus facilement les groupes d'individus. C'est en fait un retour aux vieilles habitudes du «crayon-papier» que propose EDA. Ainsi, on comprend mieux la relation quasi-organique liant EDA à notre micro favori, qui n'élimine pas la statistique conventionnelle, celle des modèles et des tests d'hypothèse, mais incite à ne plus considérer l'ordinateur sous le seul angle du calculateur.

Enfin, dans le cadre de cette approche très générale, notons que la vitesse de calcul dépend très largement du microprocesseur, du bus de transmission de données, et du temps moyen d'accès au disque de l'ordinateur utilisé. Sur Mac II, le même logiciel verra ses performances multipliées par 2 ou 3 par rapport à un modeste Mac Plus, améliorant de manière notable le confort de l'utilisateur. Cependant, la vitesse du microprocesseur de base peut être grandement accrue par la présence d'un coprocesseur arithmétique Motorola 68881. *Statview II*, par exemple, ne fonctionne qu'en présence de cet «esclave calculateur» monté en standard sur Mac II, mais nécessitant un équipement complémentaire sur Mac SE. Avant tout achat il faut donc s'assurer de l'adéquation matériel-logiciel. En fait, l'analyse de tableaux de dimensions usuelles ne nécessite pas d'équipement particulier; au-delà de quelques centaines d'individus et de quelques dizaines de variables, l'ajout d'un coprocesseur arithmétique devient souhaitable.

Afin d'assurer une certaine homogénéité à ce dossier, le lecteur retrouvera, en général, le même exemple de données dans la majorité des articles. Il s'agit de la répartition de la population active des communes de la Martinique par secteurs d'activité. Lorsque cela s'est avéré impossible en raison des particularités d'un logiciel, c'est un des exemples fournis avec le logiciel qui a été retenu.

Micheline Cosinschi & Philippe Waniez

Glossaire

■ **Statistique descriptive** comprend l'ensemble des méthodes de description des distributions statistiques. Il s'agit soit du calcul des paramètres des variables (moyenne, écart-type, quantiles, etc...), soit de représentations graphiques comme les histogrammes ou les diagrammes à bâtons.

■ **Statistique inferentielle** diffère de la précédente par le recours à des distributions de probabilités théoriques que l'on compare aux distributions observées. Ces comparaisons sont faites à l'aide de **TESTS D'HYPOTHESES** qui permettent de conclure si une distribution observée peut avoir été engendrée par un processus dont on connaît les caractéristiques; l'acceptation ou le rejet de l'hypothèse se fait avec un risque d'erreur (un seuil) choisi par l'utilisateur en fonction de la marge de sécurité qu'il peut s'accorder.

■ **Anova** «Analysis Of Variance», ou Analyse de la variance. Il s'agit de la technique la plus courante pour comparer des résultats d'expériences faites sur des échantillons indépendants. On recherche en particulier si les différences observées sur la moyenne de chaque variable dans tous les échantillons peuvent s'expliquer ou non par des fluctuations aléatoires.

■ **Correlation** est sans doute l'un des principaux concepts de la statistique. On recherche si la variation d'une grandeur mesurée par une variable est liée à la variation d'autres variables, soit par des liens de cause à effet, soit par l'action de facteurs qui leur sont communs. On distingue le coefficient de corrélation linéaire (R de Pearson), le coefficient de corrélation des rangs (rho de Spearman) et divers coefficients d'association (Tau de Kendall par exemple).

■ **Régression** recouvre une famille de techniques d'ajustement d'une fonction mathématique dans le but de modéliser une relation. Dans le cas d'une variable à expliquer et d'une seule variable explicative, on parle de régression simple ; lorsque les variables explicatives sont plus nombreuses, on a affaire à une régression multiple.

■ **Analyse factorielle** se compose d'une grande variété de techniques de recherche de facteurs latents, comme par exemple le facteur d'intelligence mesuré par un ensemble de variables. Ce ou ces facteurs sont des combinaisons des variables d'origine dont elles donnent une expression synthétique. Selon la nature des données, on utilise soit l'analyse en composantes principales (ACP), soit l'analyse factorielle des correspondances (AFC), soit encore l'analyse factorielle discriminante.

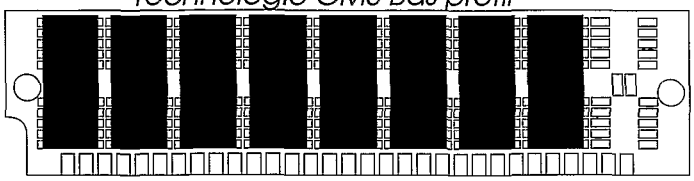
■ **Classification automatique** forme avec l'analyse factorielle l'essentiel des techniques dites multivariées, dans lesquelles entrent simultanément un grand nombre de variables. Ici, on ne cherche pas de nouvelles variables synthétiques, mais des groupes d'individus, ou classes, homogènes vis-à-vis de certains critères choisis par l'utilisateur. Le mode de formation des groupes permet de distinguer les techniques hiérarchiques (classification ascendante hiérarchique) des techniques non-hiérarchiques (nuées dynamiques par exemple).

PériMac

Nous préférons afficher des petits prix... Plutôt qu'une grande Pub !

La Barrette Mémoire 1 Mo : 695 Frs TTC

-Téchnologie CMS Bas profil-



Nos barrettes sont livrées avec schéma pour la pose.

2 Barrettes 1 390 Frs TTC (1 172,00 Fr HT)
4 Barrettes 2 750 Frs TTC (2 318,72 Fr HT)
8 Barrettes 5 500 Frs TTC (4 637,44 Fr HT)

Disques durs

Boîtier externe SCSI, accès rapide

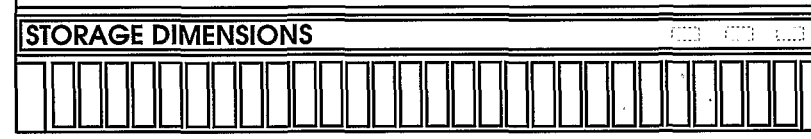
90 Mo 7 490 Frs TTC

6 315,35 Fr HT)

Autres Marques, Autres capacités... Nous consulter...

MacinStor™ 5 090 Frs TTC

(4 291,74 Fr HT) **le 45 Mo**



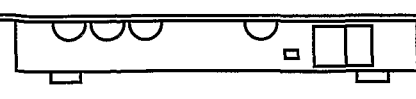
Le seul boîtier disque dur externe évolutif du marché, pouvant recevoir une seconde unité disque 3"1/2, doublant ainsi la capacité initiale, ou autorisant une sauvegarde "Miroir".
Livré formaté, prêt à l'emploi, manuel, câbles, terminateur...

29 990 Frs TTC

(25 287,00 Fr HT)

Configuration Graphique Couleur

Carte et Moniteur 19" Trinitron
Haute résolution 1024 x 826 (vrai A4, A3)
Mode 1, 2, 4 et 8 bits,
256 couleurs ou niveaux de gris
Pour toute famille Macintosh II,
Disponible pour Macintosh SE/30
Rafraîchissement de 70 hz



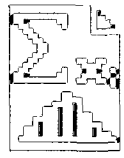
Matériel disponible sur stock !

Un Matériel vous plait ...
Votre choix est déjà fait ...
Avant tout Achat ...
Consultez PériMac !

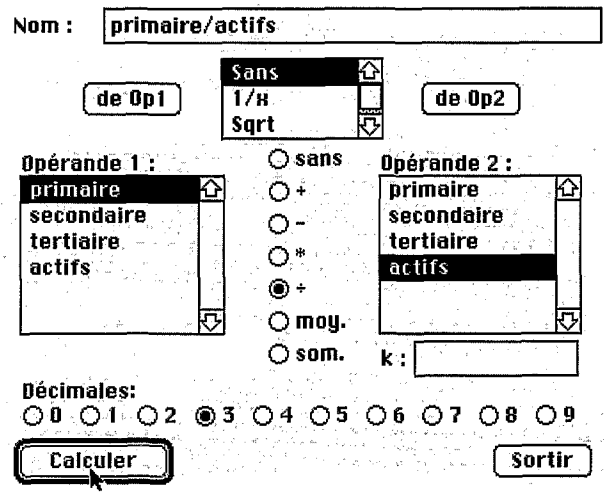
Pour nos amis de Province Nous expédions sous 24 H

PériMac 1, rue de Casablanca - 75015 PARIS - Tel : (1) 40 60 13 17

StatView : le préféré de nos lecteurs



Evitant l'inflation de techniques peu usuelles, simple d'emploi, ce statisticiel bénéficie d'une conception bien équilibrée, fruit d'une expérience déjà ancienne.



Calcul du rapport primaire/actif (qui, multiplié par 100 donnera le pourcentage des actifs dans le secteur primaire). Cette opération est réalisée avec l'article FORMULE du menu OUTILS.

Statview II et Statview SE+ Graphics ne sont en fait que deux versions du même logiciel. Pour utiliser Statview II, il faut obligatoirement doter son Macintosh II ou SE d'un coprocesseur mathématique 68881; Statview SE+ Graphics ne nécessite pas le même équipement, mais, bien évidemment, les performances s'en ressentent. De plus, ces deux logiciels peuvent utiliser la couleur, mais Statview II propose une palette plus étendue. Cela précisé, il faut savoir qu'il s'agit de la dernière évolution en date de Statview, initialement conçu par la société Brain Power, et toujours disponible sous le nom Statview 512+. Abacus concept l'a repris et doté de nombreuses améliorations. Dans le texte qui suit, Statview, sans autre précision, désigne à la fois Statview II et Statview SE+Graphics.

Statview est livré sur une seule disquette. La documentation se compose d'un seul volume de 279 pages en français. Il faut féliciter Alpha Systèmes, le distributeur français qui a fait l'effort de traduction du logiciel et de son manuel. Malgré son volume réduit, la documentation est bien faite et propose de nombreux exemples et figures commentés et agrémentés de rappels sur les méthodes statistiques. On y trouve même un appendice qui précise les formules des paramètres calculés.

Une interface conviviale

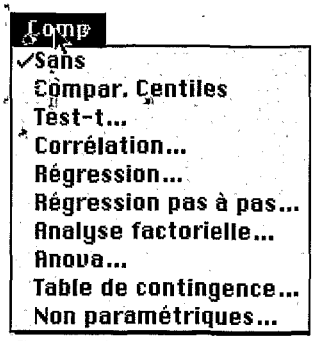
Toutes les opérations sont commandées directement à l'écran, par l'intermédiaire de la souris. C'est le règne des menus déroulants, des boutons et des boîtes de dialogue. Cela est extrêmement pratique pour des études isolées, mais gênant pour des travaux répétitifs, même si l'on dispose de Macromaker; en effet, on n'est jamais assuré d'une reproduction absolument identique d'un cas de figure donné.

Le fonctionnement de Statview repose sur un tableau rectangulaire de données où les lignes figurent les individus et les colonnes les variables. Chaque variable possède un nom (par défaut colonne 1, etc.) qui apparaît sur la première ligne du tableau et qui sera utilisé pour les calculs ultérieurs. Chaque individu est repéré par un numéro d'ordre dans ce tableau, mais il est bien entendu possible de créer une variable alphanumérique d'identification.

Notons que le tableau occupe la majeure partie de l'écran, ce qui autorise une vision synoptique avec les grands écrans. De plus, le redimensionnement de la fenêtre contenant le tableau reste toujours possible. Enfin, pour chaque variable, la largeur des cases peut aussi être changée.

Dans le menu *fichier* figure un article *importer* qui facilite la récupération de fichiers déjà enregistrés en format texte.

La création d'un tableau de données, après sélection de l'article *nouveau* du menu *fichier* est un modèle du genre. Pour chaque variable à entrer dans le nouveau tableau, une boîte de dialogue demande son nom, son type, et, s'il s'agit d'une variable numérique, le nombre de décimales. Le bouton *autre* conduit à définir une nouvelle variable, alors que *exécuter* provoque l'affichage du tableau prêt pour la saisie composé des variables préalablement définies et d'une seule ligne qui correspond au premier individu statistique.



Comparaison traite plusieurs variables à la fois.

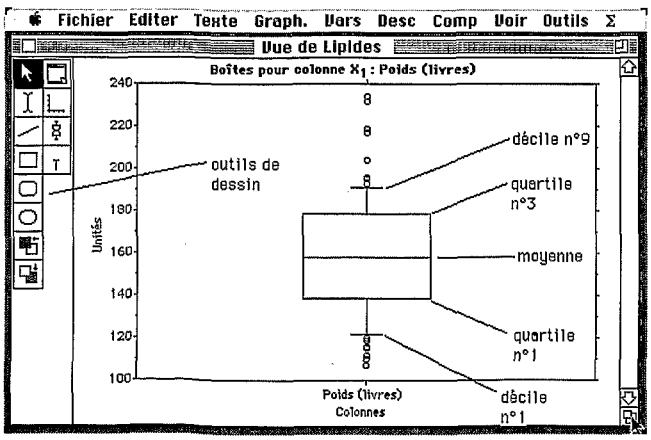
Préparation interactive du tableau de données

Statview offre un grand choix d'options de transformation des variables qui assurent la mise en conformité du tableau de données pour son analyse statistique. Le menu *outils* réalise toutes les opérations de transformation et de création des nouvelles variables nécessaires.

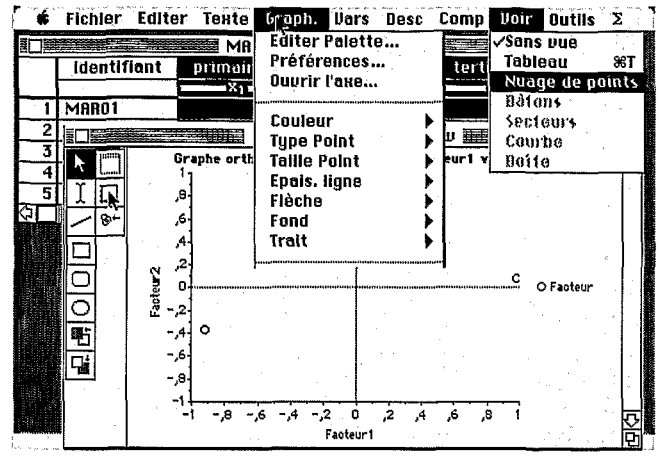
L'article *formule* réalise des calculs impliquant les valeurs de deux variables à la fois. C'est très pratique pour additionner, multiplier ou diviser deux colonnes, et cela d'autant plus qu'on peut directement appliquer aux valeurs de chaque variable l'une des 28 fonctions mathématiques disponibles. Avec *transformer*, une seule colonne à la fois est mise en jeu ; elle peut donc être transformée à l'aide d'une des fonctions mathématiques ou bien encore donner lieu à un cumul des valeurs ou à un lissage par moyennes mobiles, très utile pour dé-)saisonnaliser des séries chronologiques. *Recoder* transforme une variable continue en variable discrète par découpage en classes dont on peut spécifier l'amplitude dans une boîte de dialogue. Dans le menu *outils*, on trouve aussi un article de tri par ordre croissant ou décroissant, sur une seule colonne à la fois. Enfin, Statview peut éclater toute variable continue en autant de nouvelles variables qu'il y a de modalités présentes dans une variable discrète choisie pour l'éclatement (par exemple une variable continue donnant l'âge, et une variable discrète indiquant le sexe donnera lieu à la création de deux variables Ages/sexe). Cette dernière option facilite grandement le dépouillement des enquêtes par questionnaires.

Réaliser une analyse

La sélection des variables à analyser se fait par un clic dans chaque colonne retenue. Ces colonnes sont donc noircies et attendent une définition statistique. En effet, Statview réalise ses traitements sur un groupe composé d'au moins une variable. On distingue les variables X, sur lesquelles les mêmes calculs seront faits, des variables Y, en général une variable explicative pour les régressions. Cette affectation en X ou en Y se fait à l'aide des articles *Choisir les X* ou *Choisir les Y* du menu *Variables*.



Les graphiques sont tracés dans une fenêtre spéciale ressemblant à un MacPaint élémentaire. Ici, le diagramme en boîte et moustache (box plots) résume les principales caractéristiques de la distribution d'une variable.



Affichage du plan factoriel. A gauche, les outils. Le menu Graph quant à lui donne accès à une palette d'options graphiques étendue.

Ainsi, on se trouve alors en situation de sélection d'une méthode d'analyse statistique ou de représentation graphique. Statview ne propose qu'un nombre limité de méthodes, les plus courantes, en général bien suffisant pour couvrir un grand nombre de besoins. Le menu *Description* réalise des traitements univariés alors que *Comparaison* traite plusieurs variables à la fois.

Examinons une séquence d'opération type pour mener à bien une analyse factorielle. Lorsque les variables ont été sélectionnées, il faut alors choisir cette méthode dans le menu *Comparaison* et préciser les options de traitement (classiques mais assez complètes) retenues. La réalisation des calculs commence après avoir demandé l'affichage des *tableaux* de résultats dans le menu *voir*. La visualisation du plan factoriel croisant les composantes principales 1 & 2 s'obtient par activation de l'article *nuage de points* du menu *voir*. Notons la présence de quelques outils de modification du graphique standard ainsi qu'une assez grande variété d'options graphiques donnée par le menu *graph*. Enfin, si cette option a été cochée, on récupère les coordonnées factorielles dans le tableau.

La plupart des autres méthodes proposées par Statview fonctionnent de la même manière ou, en tous cas, adoptent la même conception. Dans les cas des techniques exigeant la présence d'une variable expliquée et d'une ou plusieurs variables explicatives comme la régression ou la corrélation, il faut sélectionner séparément les ensembles de variables.

Un excellent statisticiel d'enseignement

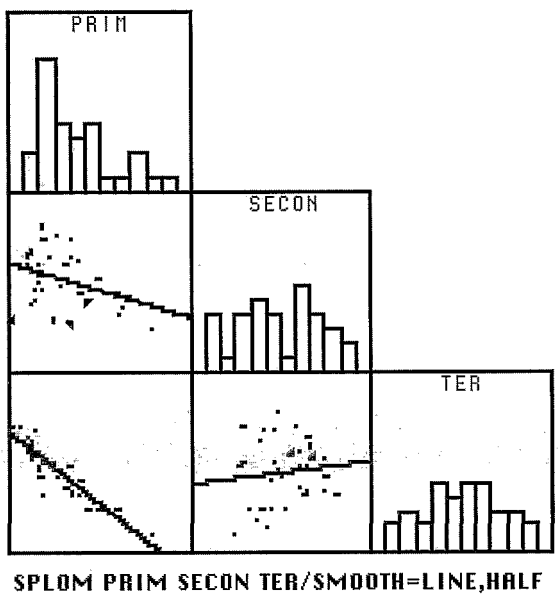
Statview est un statisticiel classique et bien conçu. L'interactivité est assez étendue mais pas totale, en particulier sur le plan des graphiques qui restent assez statiques (comparés à ceux de DataDesk). Statview peut être recommandé pour l'enseignement de la statistique, et pour les non-statisticiens (économistes, géographes, etc.) car il ne requiert pratiquement aucune connaissance préalable et ne demande pas l'apprentissage d'un langage de programmation (comme Systat). Dans le cadre professionnel, il trouvera sa place chez tous ceux qui n'ont à étudier leurs données que de manière occasionnelle. Enfin, Statview pâtit des inconvénients dûs à ses avantages, en particulier l'obligation de cliquer sans arrêt pour obtenir un résultat, ce qui est quelque peu agaçant dans le cadre d'une utilisation continue ou répétitive.

Rappelons également que Statview a été élu Icône d'or des logiciels de statistiques dans le cadre de notre référendum 89.

Philippe Waniez

Systat : pour statisticiens

Considéré depuis longtemps comme le statisticiel de référence sur PC, cette véritable encyclopédie de la statistique comblera les statisticiens les plus exigeants.



3. Chaque graphique bivarié comprend une droite de régression.

Edité par la société du même nom, Systat est l'archétype du logiciel de statistique gouverné par son propre langage de commande, proche du langage Basic. Systat s'adresse plus particulièrement aux statisticiens désirant convertir une partie de leurs applications, les plus légères, d'un ordinateur central vers un micro.

Systat 3.2 est livré sur cinq disquettes. La documentation, très épaisse, se compose d'un manuel de référence de plus de 400 pages, commun aux différents systèmes d'exploitation, intitulé « Systat, *the system for statistics* », d'un mode d'emploi du module graphique nommé (930 pages), et d'un abrégé des commandes formant la « Reference Card ». A tout cela s'ajoute le bulletin d'information trimestriel « Sysnet » dans lequel on trouve une présentation des nouvelles versions et des compléments d'information technique. A l'usage, cette documentation en anglais (il n'existe pas de traduction en français) s'est révélée très pratique et sans ambiguïté grâce aux nombreux exemples et figures. Les différents manuels facilitent le véritable apprentissage, nécessaire à la maîtrise des méthodes employées. On retrouve partout le même exemple relatif aux diverses formes de criminalité aux Etats-Unis, ce qui simplifie beaucoup la compréhension.

Au premier contact, Systat apparaît un peu comme une « usine à gaz », c'est-à-dire un enchevêtrement complexe de fonctions et d'options. Ses concepteurs l'ont voulu modulaire : chaque module correspond à une famille de méthodes statistiques. La conséquence directe de cette conception est une économie de mémoire centrale. L'organisation modulaire présente également un intérêt pédagogique : un groupe de fonctions est rassemblé dans un module donné et peut donc être étudié séparément des autres. Tous les modules proposent un menu « Transfert » permettant de quitter le module en cours pour un autre, et cela sans repasser par le bureau.

Les douze modules couvrent le large spectre des méthodes statistiques nécessaires.

DATA assure la constitution d'un fichier Systat, donc la saisie des données à l'aide d'un éditeur, leur lecture dans un fichier ASCII, leur transformation à l'aide d'opérateurs arithmétiques ou de fonctions mathématiques.

STATS calcule les paramètres des distributions statistiques (moyenne, écart-type, etc.) et réalise des tests de différences de ces paramètres calculés sur des groupes d'individus.

TABLES produit des tableaux croisés de profondeur multiple et ajuste un modèle Log-linéaire à des données discrètes.

NPAR: Tests non-paramétriques de Wilcoxon, Kruskal-Wallis, Kolmogorov-Smirnov....

CORR: Calcul des coefficients de corrélation linéaire de Pearson, de corrélation des rangs de Spearman ainsi que divers coefficients de similarité.

MGLH: signifie « multivariate general linear hypothesis ». C'est une adaptation du programme de calcul des moindres-carrés généralisés de Wilkinson nommé

REGM. Il ajuste en particulier l'ensemble des principaux modèles de régression simple, multiple, polynomiale, avec variables muettes, avec ou sans terme constant. Par ailleurs, il réalise des analyses de variance du genre ANOVA ou MANOVA. Ce module apparaît donc d'une très grande richesse.

FACTOR: Classique module d'analyse en composantes principales muni d'une grande variété d'options de rotation. L'utilisateur francophone sera déçu de ne pas y retrouver l'analyse des correspondances.

MDS: «Multidimensional scaling», c'est à dire représentation des similitudes dans un espace non-métrique. Plusieurs méthodes sont proposées, Kruskal, Shepard et Guttman.

CLUSTER: Propose une grande variété de méthodes de classification hiérarchique (avec plusieurs critères de calcul des distances) ou non (K-moyennes).

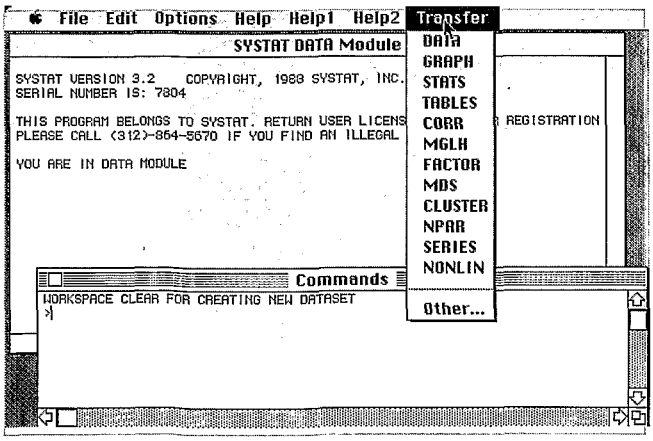
SERIES: Sous les commandes SMOOTH, ARIMA et FOURIER se cachent de nombreuses méthodes d'analyse des séries chronologiques y compris Box-Jenkins. Comme MGLH, SERIES forme un très puissant ensemble de techniques d'analyse.

NONLIN : calcule les paramètres d'une très grande variété de modèles non-linéaires comme la régression logistique avec estimation du maximum de vraisemblances.

GRAPH: trace une grande variété de graphiques, histogrammes, diagrammes à bâtons, diagrammes bi ou trivariés, mais aussi des courbes de niveaux, des surfaces tridimensionnelles, des cartes géographiques, des diagrammes triangulaires, etc. Les graphiques spécifiques à l'analyse statistique ne sont pas oubliés comme les diagrammes en boîte («Box plot») ou en tronc et feuilles («stem and leaf plot»).

Quatre autres modules supplémentaires correspondant à des besoins particuliers peuvent être acquis séparément. *DESIGN* réalise les opérations nécessaires aux plans d'expérience aléatoires. *LOGIT* offre la régression logistique sur des variables binaires avec estimation par le maximum de vraisemblances. *PROBIT* propose une méthode de régression appropriée pour l'estimation des paramètres d'un modèle de régression multiple et l'analyse de covariance sur des variables dépendantes catégorielles ne pouvant prendre qu'une seule modalité parmi deux. Enfin, *TESTAT* calcule des tests statistiques, des coefficients d'association, etc., sur des questionnaires à réponses multiples.

Chaque module de Systat définit l'environnement nécessaire au dialogue. Celui-ci se fait au moyen de deux fenêtres (écran 1). La fenêtre inférieure permet d'entrer les instructions en langage Systat afin de réaliser un traitement donné.



1. L'écran standard d'un module de SYSTAT. Dans la partie supérieure, la fenêtre de sortie des résultats; en bas, la fenêtre de commande. Le menu déroulé, TRANSFER, donne accès aux autres modules.

get "MAR"	Sélection du fichier ASCII nommé MAR.
input code$ prim secon ter popac	Lecture des 5 variables composant le fichier MAR.
save SECT	Sauvegarde de la lecture *dans un fichier Systat* SECT.
run	Exécution de cette étape.
use SECT	Sélection du fichier Systat SECT créé ci-dessus.
*let prim=(prim/popac)*100*	Calcul des pourcentages.
*let secon=(secon/popac)*100*	
*let ter=(ter/popac)*100*	
drop popac	Suppression de la variable popac devenue inutile.
save PCTSECT	Sauvegarde des calculs *dans le fichier Systat*
PCTSECT.	
run.	Exécution de cette étape.
use PCTSECT	Sélection du fichier Systat PCTSECT créé ci-dessus.
list	Liste du fichier.
run.	Exécution de cette étape.

Voici un exemple de création et d'utilisation d'un fichier Systat. Des données relatives à la population active des communes de la Martinique ont été saisies à l'aide d'Excel, puis sauvegardées en mode texte, dans un fichier nommé MAR dans l'ordre suivant : le code alphanumérique de la commune, la population active du secteur primaire, du secteur secondaire et du secteur tertiaire et, enfin, la population active totale. On cherche à calculer, dans le module DATA, la part (en %) de chaque secteur dans la population totale. Cette courte démonstration montre la simplicité d'utilisation des commandes de Systat, mais aussi le nécessaire apprentissage préalable à toute analyse. Sa ressemblance avec le langage Basic le rend très facile d'apprentissage à tous les utilisateurs ayant une connaissance rudimentaire de ce langage.

Chaque module comprend un jeu d'instructions qui lui est propre ; ces commandes doivent être entrées au clavier dans la fenêtre inférieure, après le signe «>», soit in-extenso, soit réduites à leurs deux premières lettres, ce qui permet de gagner du temps.

La présence de menus d'aide, dont les articles sont les mots du langage, facilite la rédaction des programmes. L'écran 2 présente un exemple d'utilisation de ces menus, à propos de la commande INPUT : son rôle est d'abord succinctement décrit, puis quelques exemples en présentent les principales formes. Ceci est très pratique, à condition de savoir lire l'anglais. Enfin, notons qu'en cas d'erreur de frappe, Systat affiche un message qui en indique la cause (mais de manière parfois difficile à comprendre).

Pour réaliser des traitements statistiques simples, il suffit d'entrer quelques commandes dans la fenêtre prévue à cet effet; on obtient alors une multitude de paramètres statistiques. Examinons le déroulement du processus nécessaire à l'application de trois méthodes courantes au fichier PCTSECT précédemment créé : le calcul des paramètres des distributions, la régression linéaire simple, l'analyse en composantes principales.

Deux instructions seulement sont nécessaires au calcul sur les variables numériques présentes dans le fichier. En premier lieu, il faut sélectionner le fichier à l'aide de la commande «USE». Puis, la commande STATISTICS (ou ST) déclenche le calcul proprement dit. De manière standard, le module STATS calcule le nombre d'individus, le minimum, le maximum, la moyenne arithmétique et l'écart-type.

L'ajustement d'un modèle de régression se fait dans le module MGLH.

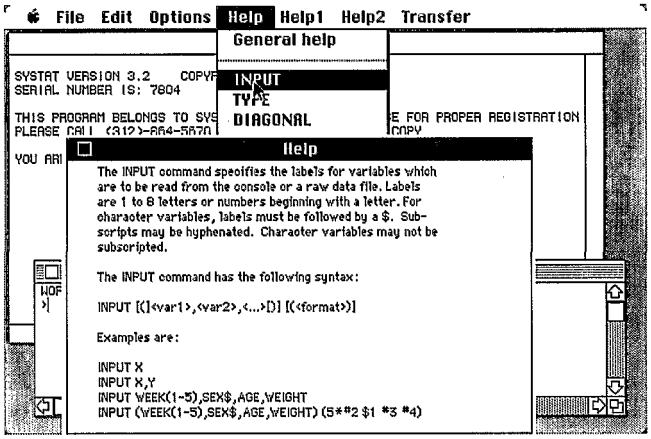
Le module FACTOR est très complet, avec notamment tout une batterie de rotations orthogonales ou obliques. Cependant, les sorties ne sont pas très heureuses, en particulier pour les graphiques. On a donc tout intérêt à ne faire que les calculs avec FACTOR et à procéder aux représentation graphiques à l'aide du module GRAPH. Le stockage dans un fichier des coordonnées des individus sur les composan-

tes principales se fait par l'instruction :
SAVE nom du fichier/SCORES

Les sorties numériques du module FACTOR sont classiques.

Ces quelques exemples peuvent donner au lecteur une impression de sécheresse du logiciel Systat. Celui-ci se veut un système sérieux, dépouillé de tout gadget inutile. A celui qui ne cherchera pas à connaître l'ensemble de ses options, Systat ne donnera que le strict minimum, bien suffisant dans la majorité des applications il est vrai.

Autant les modules statistiques de Systat apparaissent très complets, mais classiques sur le plan des méthodes et des sorties, autant le module graphique offre une grande variété de diagrammes. Ceux-ci simplifient l'étude des distributions et des relations entre variables. Bien sûr, le module GRAPH propose les représentations courantes : diagrammes à bâtons et autres histogrammes sont bien présents et s'affichent relativement vite. Mais le principal intérêt de ce module réside ailleurs, dans les graphiques proprement statistiques, qu'on ne trouve pas dans les tableurs, par exemple.



2. Le menu HELP. Ici, la syntaxe de l'instruction INPUT.

Les graphiques réalisés par Systat se répartissent en 4 groupes, selon le nombre simultané de variables qu'ils permettent d'étudier. En premier lieu, on trouve ceux ne traitant que des caractéristiques de chacune des distributions statistiques. Le BOX PLOT et le STEM-LEAF PLOT sont de ceux-ci. Leur construction est assez différente des histogrammes classiques puisqu'ils facilitent l'appréciation de la forme des distributions en fonction de leurs paramètres, médiane et intervalle interquartile.

Systat propose toute une batterie de graphiques bivariés, autorisant l'examen précis des relations entre deux variables. Parmi toutes les options proposées, celle du tracé d'une droite de régression est très intéressante. Sont tracés, non seulement les points représentant les individus par leurs valeurs sur les 2 variables du graphique, mais aussi la droite de régression figurant la forme de la relation et l'intervalle de confiance d'après un seuil choisi par l'utilisateur (CONFI=.95 pour un seuil à 5%). Le repérage des individus particuliers ne se conformant pas à la relation générale présentée par la régression est ainsi simplifié.

Systat réalise des graphiques en perspective cavalière pour localiser des individus dans un espace à trois dimensions, figurant trois variables différentes. L'option *line* trace une ligne verticale entre le plan de base et chaque point donnant ainsi l'impression de relief recherché. Le résultat n'est vraiment satisfaisant que si les points présentent une tendance décroissante orientée du fond vers l'extérieur du graphique. Dans tous les autres cas, la lecture est difficile.

Beaucoup plus satisfaisante est l'étude des relations entre un groupe de variables prises 2 à 2. Les graphiques de type *splom* (écran 3) se présentent sous la forme d'une matrice carrée, comme les matrices de coefficient de corrélation. Dans la diagonale, on trouve les histogrammes des variables choisies; ailleurs, des graphiques bivariés, avec ou sans droite de régression, donnent une idée des relations entre les variables. Ce mode de représentation est d'une exceptionnelle efficacité puisqu'il communique d'un seul coup, et de manière complémentaire à la matrice de corrélation, l'ensemble des relations entretenues par les variables d'un fichier.

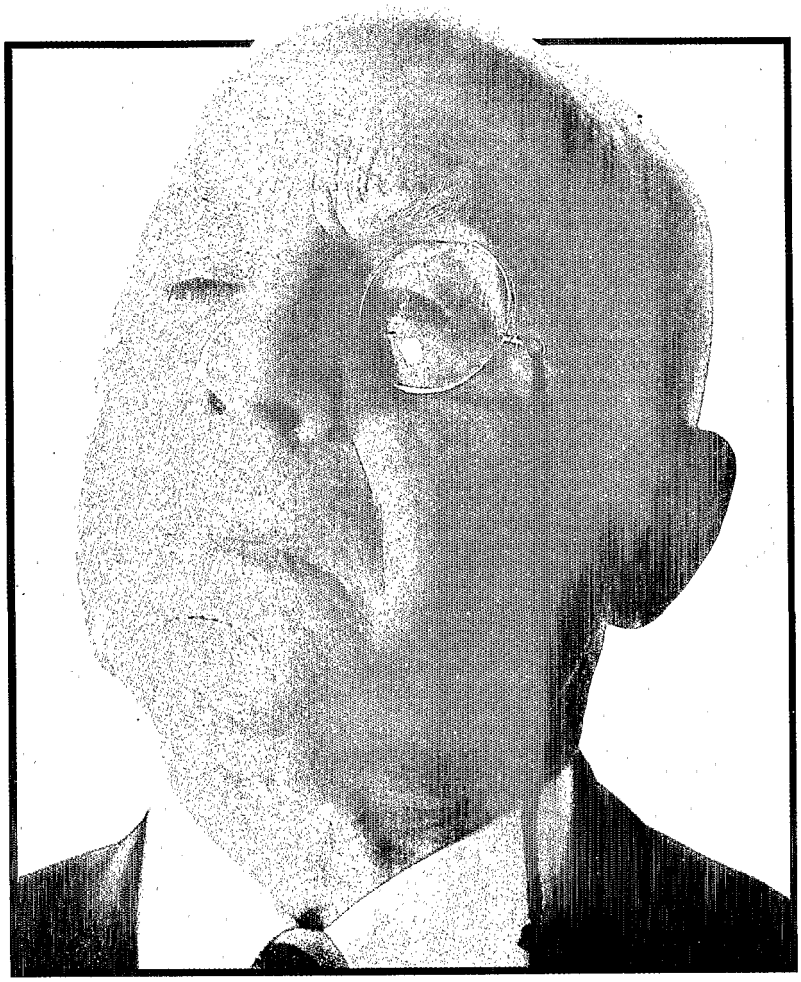
Les graphiques de Systat sont du type PICT et peuvent donc être récupérés dans *MacDraw, SuperPaint, Canvas*, etc. Une méthode pratique consiste à utiliser l'article *Copy Graph* du menu *Edit*, et à coller le dessin dans l'album. On conserve ainsi une séquence de graphiques qu'il sera facile d'utiliser ou de modifier plus tard. Remarquable logiciel d'analyse statistique, Systat offre au statisticien un très grand nombre de méthodes d'analyse. Sa mise en œuvre n'est pas immédiate car elle nécessite l'apprentissage du langage de commande propre au logiciel.

A ceux qui ne possèdent aucune notion de programmation, Systat semblera difficile d'accès ; il leur est donc conseillé de s'orienter vers d'autres logiciels plus conviviaux. Par contre, tout programmeur en langage Basic élémentaire trouvera en Systat un outil performant, complet, et surtout adaptable à ses problèmes particuliers, précisément grâce au langage de programmation interprété par le logiciel. Les instructions, préalablement enregistrées dans un fichier «texte», peuvent être soumises en traitement par lot (du genre «batch», comme avec les gros systèmes), ce qui se révèle très utile pour des utilisations répétitives.

Langage de programmation, variété des options de traitement et richesse graphique sont, en définitive, les atouts maîtres de Systat.

Philippe Waniez

On peut réussir sans la Presse Professionnelle. Mais tellement moins vite.



La Presse Professionnelle s'engage en permanence sur la qualité de sa rédaction. C'est sa raison d'être.
Tout ce qui est nouveau, utile, performant est d'abord dans cette presse-là.
Avec elle, on progresse plus vite.

La Presse Professionnelle sait mettre en valeur tous les acteurs d'une profession. C'est sa vocation.
Tout ce qui bouge, se fait, se dit, c'est d'abord dans cette presse-là.
Avec elle on réussit plus vite.

Moteur de tous les progrès et de tous les succès, la Presse Professionnelle est
le miroir fidèle de chaque profession.

La Presse Professionnelle, le média de tous les succès.

SYNDICAT DE LA PRESSE
SPEP
DES ENTREPRISES
ET DES PROFESSIONNELS

Parameter Manager : pour les ingénieurs

Contrôle de qualité, surveillance des processus, marketing : l'analyse des séries chronologiques ne passe pas nécessairement par l'emploi des méthodes mathématiques les plus sophistiquées.

File Edit Windows Schedule Data Analyze Preferences Report

Weather-Parameters ID:San Jose

Parameter	001	002	003	004	005	006
Name	High Temp	Low Temp	Avg. Temp		Pressure	Humidity
Units	°F	°F	°F		Bars	Percent
Ref Time						
			- Condition Limits -			
High Alarm	100.0	60.0	70.0		32.0	100 %
Low Alarm	60.0	18.0	46.6		28.0	0 %
High Alert	70.0		59.0			
Low Alert	46.6	32.0	46.6			
High Normal	58.3	58.2	58.2			
Low Normal		46.6	46.6			
Ref Value					25.0	
% Change						
Max Value	91.0	61.0	74.0		30.2	84 %
Min Value	72.0	49.0	62.0		29.8	35 %

1. La fiche des paramètres permettant de définir le contenu de la base de données.

Avec Parameter Manager Plus (pmPlus), les ingénieurs disposent maintenant d'un remarquable outil d'analyse de la variation dans le temps de paramètres (en fait, de variables, au sens statistique du terme) mesurés au cours du déroulement d'un processus technique ou d'une expérience scientifique, à partir de divers capteurs. pmPlus est livré avec Parameter Manager Talk (pmTalk) qui offre une grande variété de modes d'acquisition de données à partir d'informations stockées dans une base de données, à laquelle on accède par modem et ligne téléphonique; pmTalk peut aussi enregistrer des données en provenance d'unités de mesure assurant une conversion analogique/numérique selon la norme IEEE 488.

Réalisés par la société Rebus Development Corporation, pmPlus et pmTalk sont livrés sur deux disquettes : la première contient l'application, la seconde offre un grand nombre d'exemples relatifs à des domaines variés comme des tests de laboratoire, une étude de contrôle de qualité, des données médicales ou encore des séries climatiques où nous irons chercher les exemples présentés ici. La très volumineuse documentation en anglais comprend 3 volumes. D'une part, les 80 pages du très accessible manuel d'introduction complètent le diaporama de démonstration figurant dans la disquette d'exemples. Ceci facilite l'apprentissage de ce système assez complexe. Avec ses 470 pages, le manuel de référence présente méthodiquement les phases d'élaboration de la base de données, d'analyse descriptive et prévisionnelle, et de présentation des résultats. Enfin, le manuel qui se rapporte à pmTalk, intitulé « communications interface », décrit les diverses procédures d'acquisition de données.

Une structure de données rigoureuse

pmPlus dispose d'un rigoureux système de gestion de base de données intégré. Il se compose de cinq éléments. En premier lieu, chaque base de données relative à un thème, comme par exemple, le temps dans la ville californienne de San José, comprend une fiche descriptive qui indique le thème d'étude, l'intervalle de temps retenu pour toutes les mesures, le nombre de paramètres enregistrés, le nombre de mesures effectivement opérées, et la date du dernier enregistrement.

En second lieu, la fiche des paramètres (écran 1) donne une description précise des variables. Lors de la phase de définition de la base, l'utilisateur doit donner, pour chaque variable, son nom, son unité de mesure et son type (numérique, alphanumérique, temporel, pourcentage, etc.). A cela s'ajoute une possibilité fort intéressante pour l'analyse des processus physiques : on peut indiquer des valeurs particulières des variables, nommées « conditions limites », qui faciliteront ultérieurement l'observation des variations dans le temps; il s'agit, en quelque sorte, de signaux d'urgence qui rappellent les valeurs normales, d'alerte et d'alarme.

Le tableau de mesures proprement dit (écran 2) contient les valeurs des variables. C'est un tableur d'un genre très particulier. On y retrouve le numéro, le nom et l'unité de mesure de chaque paramètre tels qu'ils ont été définis lors de la création de la base. Chaque ligne représente un enregistrement composé, bien entendu, de la date et de l'heure de la mesure, suivies des valeurs des paramètres.

File Edit Windows Schedule Data Analyze Preferences Report

Weather-Measurements ID:San Jose

Parameter	001	002	003	004	005	006
Name	High Temp	Low Temp	Avg. Temp		Pressure	Humidity
Units	°F	°F	°F		Bars	Percent
Date / Time	- Measurement Values -					
'85 Jul 15,12:00	87.7	53.4	70.0		29.9	39 %
'85 Jul 16,12:00	85.0	55.0	70.0		29.9	45 %
'85 Jul 17,12:00	84.0	54.0	69.0		29.9	43 %
'85 Jul 18,12:00	84.0	54.0	69.0		29.9	35 %
'85 Jul 28,12:00	79.0	59.0	69.0		29.9	48 %
'85 Jul 29,12:00	79.0	59.0	69.0		30.0	47 %
'85 Jul 30,12:00	76.0	61.0	69.0		30.0	59 %
Sample No.						

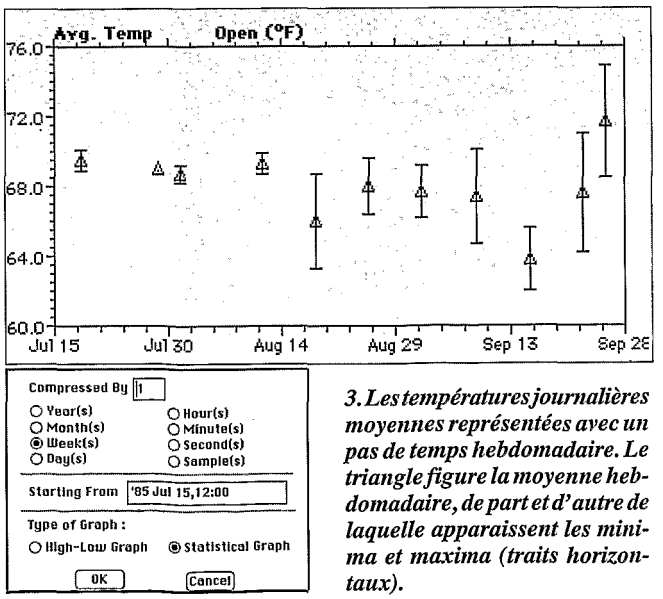
2. Le tableau de mesures.

Enfin, la structuration des données dans la base est complétée par deux ensembles d'informations. D'une part, le calepin donne la possibilité d'écrire et de conserver des remarques relatives à l'enregistrement des données, comme, par exemple, les éventuels incidents survenus au cours d'une expérimentation. D'autre part, on peut conserver une figure, un schéma indiquant la localisation des capteurs, le dessin de la pièce mécanique étudiée, etc. Chaque élément de cette structure assez complexe est accessible via les articles du menu *WINDOWS*. Ce dernier est complété par le menu *DATA* qui contient tous les outils nécessaires à l'importation de fichiers externes, au calcul de nouvelles variables, au tri de la base à partir des données qu'elle renferme (dans l'ordre des températures, par exemple), à la sélection d'une partie seulement des observations, selon un critère choisi par l'utilisateur, ou bien encore l'inclusion (*JOIN*) de données nouvelles à une base existant déjà.

Des possibilités de traitement limitées

A l'usage, pmPlus semble être davantage un système d'analyse graphique de données statistiques d'un genre particulier, les séries chronologiques, qu'un véritable logiciel d'analyse statistique de ces données, comme en demandent les économètres, par exemple.

On choisit les méthodes de représentation graphique dans le menu *ANALYZE* qui propose 10 articles différents. On n'y trouve que des modes de représentation très classiques, mais aussi très utiles. Ces méthodes se répartissent en deux familles. D'une part, celles qui prennent directement en compte le temps, comme l'habituel graphique de tendance *TREND PLOT*. Il représente la variation d'un paramètre figuré en ordonnée, comme par exemple la température, en fonction du temps, qui apparaît en abscisse; l'article *STRIP CHART* superpose plusieurs paramètres sur le même graphique. Comme les points représentant chaque observation sont reliés par des lignes, on peut ainsi détecter visuellement l'existence d'une tendance. Beaucoup plus originale est la fonction de compression des données: les valeurs relevées sur un pas de temps donné, chaque jour, par exemple, peuvent être compressées sur un pas de temps plus long, comme la semaine (écran 3). Enfin, un e fonction de prévision estime, à partir d'une famille de fonctions de lissage (exponentiel, etc.), les valeurs futures d'un paramètre donné, et quand il risque de dépasser l'une des valeurs critiques qui lui ont été assignées lors de la création de la base.



3. Les températures journalières moyennes représentées avec un pas de temps hebdomadaire. Le triangle figure la moyenne hebdomadaire, de part et d'autre de laquelle apparaissent les minima et maxima (traits horizontaux).

L'autre famille de méthodes d'analyse apparaît dans la plupart des logiciels d'analyse statistiques : statistiques descriptives, histogrammes, graphiques bivariés, et corrélation.

Un puissant éditeur de rapports

Parameter Manager Plus propose un puissant éditeur de rapport, sous forme numérique et graphique, qui permet la constitution de véritables dossiers de mesures et de bulletins d'information comme en diffusent les stations météorologiques du monde entier. Ainsi, malgré des limites vite atteintes sur le plan de l'analyse statistique numérique, pmPlus semble être un logiciel bien ciblé, et qui rendra de nombreux services aux techniciens et ingénieurs dont l'activité principale est l'acquisition et la diffusion de mesures relevées dans le temps.

Philippe Waniez

LADDAD : l'analyse des données à la française

Analyse des correspondances, nuées dynamiques, aides à l'interprétation, la place laissée vacante par les logiciels anglo-saxons dans le domaine de l'analyse multivariée apparaît désormais solidement occupée.

LADDAD est le logiciel diffusé par l'Association pour le Développement et la diffusion de l'Analyse des Données. Il rassemble l'essentiel de la méthodologie acquise depuis les années 60 par une trentaine d'enseignants du supérieur, de chercheurs et d'ingénieurs. L'Analyse des Données est une branche particulière de la statistique regroupant un ensemble de méthodes dites multidimensionnelles, par opposition aux méthodes de la statistique descriptive qui ne traitent, en général, qu'une seule variable à la fois ; le terme analyse multivariée a également cours. On recourt à l'Analyse des Données pour obtenir une information qui résume des ensembles de données trop grands et trop complexes pour être appréhendés directement. Derrière le foisonnement des statistiques, les résultats mettent en évidence les tendances les plus marquantes et les hiérarchies, tout en éliminant tout ce qui perturbe une perception globale. Sans chauvinisme mal placé, on doit reconnaître l'importance et l'originalité des apports de l'Ecole française d'analyse des données, ce qui justifie parfaitement le titre de cet article.

LADDAD est un logiciel portable: il fonctionne donc sur une vaste gamme d'ordinateurs, des plus grands systèmes IBM aux micro-ordinateurs PC/PS et Macintosh. L'un des avantages de cette portabilité réside dans la possibilité pour les utilisateurs de partager des savoir-faire et de changer de machine sans avoir à apprendre le fonctionnement d'un nouveau logiciel. Cette médaille a son revers: le logiciel n'utilise pas l'ensemble des possibilités de chaque machine. Ainsi, la version pour Macintosh ne fait pas appel aux possibilités graphiques de cette machine, ni à l'interactivité permise par l'interface utilisateur qui nous est chère. Cependant, LADDAD propose une telle richesse de techniques d'analyse introuvables ailleurs que tout «analyste de données» se doit d'en connaître les possibilités étendues, ce qui le conduira, sans doute, vers une utilisation intensive, parallèlement avec d'autres statisticiels plus classiques.

Le logiciel est livré sur cinq disquettes, ce qui représente plus

```
LES POIDS DES LIGNES ET DES COLONNES SONT MULTIPLIES PAR  10 ** -2
 NOMJ(J)!  PRIM  SECO  TERT
 PJ(J)  !    98   147   623   869

 LES VALEURS  PROPRES        VAL(1)=   1.00000
 INUM ! VAL PROPRE ! POURC.! CUMUL !VARIAT.!*! HISTOGRAMME DES VALEURS PROPRES
 !  2 !      .17163 ! 93.910! 93.910!*******!*!***************!***************!
 !  3 !      .01113 !  6.090!100.000! 87.820!*!**

     ! I1 ! QLT POID INR!  1#F  COR CTR!  2#F  COR CTR!
    1!MA01! 1000    5  50!  1356  976  52!  -213   24  20!
    2!MA02! 1000   16  32!   590  964  33!   114   36  19!
   33!MA33! 1000    5  77!  1614  947  78!  -381   53  67!
   34!MA34! 1000    3  22!  1168  946  23!  -280   54  20!
     !    !          1000!             1000!             1000!
     ! J1 ! QLT POID INR!  1#F  COR CTR!  2#F  COR CTR!
    1!PRIM! 1000  113 790!  1127  996 838!   -69    4  49!
    2!SECO! 1000  169  55!    68   79   5!   233  921 826!
    3!TERT! 1000  717 155!  -194  951 157!   -44   49 125!
     !    !          1000!             1000!             1000!

AXE HORIZONTAL( 1)--AXE VERTICAL( 2)--TITRE:SECTEURS D'ACTIVITE EN MARTINIQUE

NOMBRE DE POINTS :   37
==ECHELLE :  4 CARACTERE(S) = .110    1 LIGNE = .046
+-----------+-------MA26-------------------------------------------------+ 0 01
!           SECO                                                         ! 0 01
!           !     MA20                                                   ! 0 01
!           MA18    MA25MA03    MA02                                     ! 0 01
!       MA12MA21MA09MA24        MA29                                     ! 1 01
!   MA16    !   MA19            MA17MA28                                 ! 0 01
MA11--------+---MA15-------------------------------------------------+ 0 01
!   TERT    !   MA14    MA23                        PRIM                ! 0 01
!       MA08MA27                                MA31                    ! 1 01
MA10        !                       MA05    MA30                        ! 0 01
!           !           MA06                        MA32    MA01        ! 0 01
!           !               MA07                        MA04            ! 0 01
!           !                                       MA34                ! 0 01
!           !                                                           ! 0 01
!           !                                                       MA33! 0 01
+-----------+-------------------------------------------------------+ 0 01

NOMBRE DE POINTS SUPERPOSES :   2

 MA13(MA12)  MA22(MA08)
```

Extraits des sorties du programme ANCORR.
La première partie (non représentée ici) rappelle les caractéristiques de l'analyse. Suivent l'histogramme des valeurs propres, les coordonnées (1#F et 2#F) et les contributions absolues (COR) et relatives (CTR) des individus, puis des variables et, enfin, le premier plan factoriel sur lequel on peut apprécier, au travers des communes (MAR01 à MAR34), le caractère plus ou moins dominant de chaque secteur (PRIM, SECO, TERT).

de 3 Mo. sur disque. La documentation, forte de 250 pages rédigées en français, est très complète. Les méthodes d'analyse ne sont pas décrites, ce que justifie l'abondante bibliographie qui s'adresse à des lecteurs de tous niveaux en mathématique. Il s'agit, pour l'essentiel, du mode d'emploi de chaque programme qui comprend une présentation du type de tableau de données en entrée, des sorties attendues, des paramètres à fixer et des options à choisir.

LADDAD se compose de trois sous-ensembles de programmes indépendants dans leur fonctionnement mais qui peuvent échanger des données et des résultats sous forme de fichiers.

Recodage et description : une importante étape préalable

Le premier sous-ensemble comprend toute une série de procédures de préparation des tableaux préalable à l'analyse des données proprement dites. Cette opération est très importante car elle permet d'adapter les données aux conditions exigées par chaque méthode.

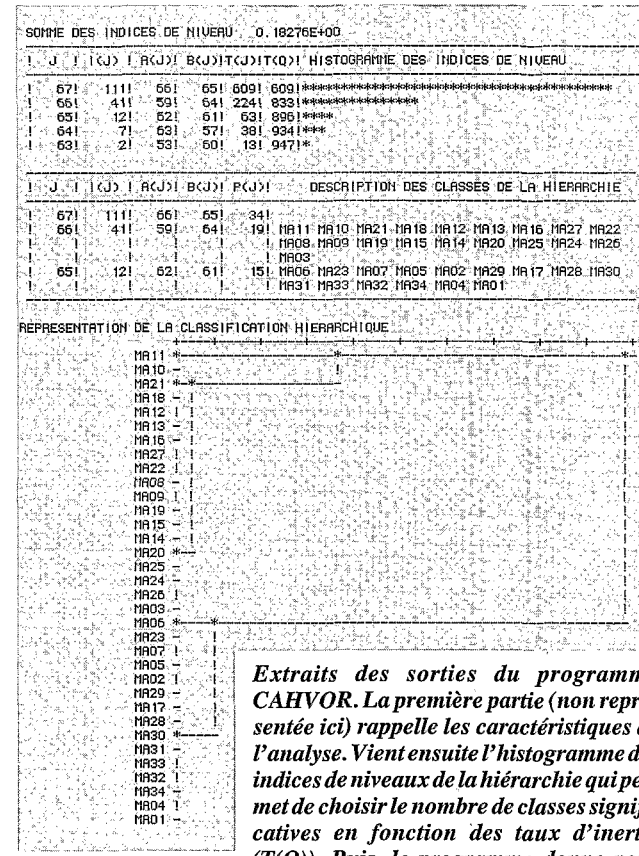
Le programme *DEDOUB* assure le dédoublement d'un tableau de notes, par exemple des notes obtenues à divers tests psychologiques, afin de contrôler le poids de chaque test dans les analyses ultérieures. *DISJON* met sous forme disjonctive complète un tableau de variables logiques codées 0 ou 1, par exemple les réponses OUI et NON à un questionnaire ; on confère ainsi une importance égale dans l'analyse à chacune de ces deux modalités. *RECOD1* étend ce type de recodage à des questions ayant plus de deux modalités de réponse. Ces trois programmes sont extrêmement utiles, en particulier, à tous ceux qui doivent traiter des données provenant de questionnaires; ils rendent compatibles les réponses avec les diverses techniques d'analyse des données qui requièrent une certaine homogénéité des tableaux d'entrée. Enfin, *RECOD2* permet de transformer une variable continue en variable discrète découpée en classe d'effectifs égaux (quartiles, déciles, etc.), ou bien encore de centrer et réduire chaque variable (la variable recodée a une moyenne arithmétique nulle et un écart-type égal à l'unité).

Analyse factorielle, classification automatique et discrimination : variété et puissance de l'Analyse des Données

L'Analyse des Données comprend principalement trois familles de méthodes. D'une part, les méthodes factorielles utilisent des calculs d'ajustement faisant appel à l'algèbre linéaire pour localiser les objets à décrire (variables, individus, ou les deux) par rapport à tous les autres objets, sur un axe ou dans un plan. D'autre part, les méthodes de classification mettent en jeu des procédures algorithmiques pour rassembler et ranger les objets à décrire, en fonction de leur degré de ressemblance, dans des classes plus ou moins homogènes. Enfin, les méthodes de discrimination ont pour principale finalité le classement d'individus testés, dont on cherche à connaître les caractéristiques vis-à-vis d'une population de base connue ; l'aspect décisionnel et parfois même prévisionnel est dans ce cas le plus important. LADDAD couvre l'ensemble de ces méthodes très variées.

Les méthodes factorielles sont représentées par les programmes *ANCORR* pour l'analyse des correspondances, *ESCOF2* pour l'analyse des correspondances par sous tableaux et *ANCOMP* pour l'analyse en composantes principales. La figure n°1 présente un exemple de sortie du programme ANCORR. Notons que LADDAD permet d'étudier le comportement de variables et d'individus supplémentaires, c'est-à-dire leur localisation dans l'espace factoriel, sans qu'ils contribuent à la définition des facteurs proprement dite; cette option, inexistante dans les autres logiciels, accroît considérablement les possibilités d'analyse.

Les techniques de classification proposées comprennent la classification ascendante hiérarchique sous diverses formes : maximisation du moment centré d'ordre deux d'une partition (voisins réductibles et voisins réciproques) et d'après le critère de l'information mutuelle (il s'agit, respectivement, des programmes *CAHVOR*, *CAH2CO* et *CAH-2IN*). Les méthodes non-hiérarchiques sont représentées par les nuées dynamiques (*NUEDYN*) et les boules optimisées (*BOULOP*). Les sorties de ces programmes de classification sont plus complètes que ce que proposent d'autres statisticiels (figure n°2). Enfin, la discrimination comprend les programmes *MAHAL2* pour le cas de deux groupes et *MAHAL3* qui est une généralisation du précédent au cas de trois ou plusieurs groupes.



Extraits des sorties du programme CAHVOR. La première partie (non représentée ici) rappelle les caractéristiques de l'analyse. Vient ensuite l'histogramme des indices de niveaux de la hiérarchie qui permet de choisir le nombre de classes significatives en fonction des taux d'inertie (T(Q)). Puis, le programme donne pour chaque nouvelle classe, la liste des individus qui la composent. Enfin on représente la hiérarchie par un arbre sur lequel on observe les principales ruptures dans le processus de hiérarchisation.

Un peu lourd

Comme nous l'avons déjà signalé, le mode de fonctionnement de LADDA est le même sur tous les types d'ordinateur. Il consiste principalement par la rédaction d'un fichier de commande qui est ensuite soumis pour exécution à l'application ADDAD qui en vérifie la syntaxe et déclenche les traitements.

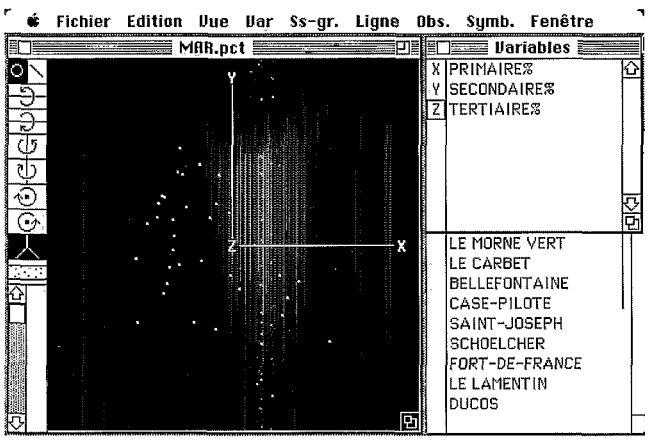
Ce mode de fonctionnement un peu lourd doit être corrigé dans un avenir proche.

Reste qu'en l'état, l'utilisateur de ce logiciel dispose d'une extraordinaire bibliothèque de programmes, unique à notre connaissance, et qui rendra de nombreux services à tous ceux qui désirent pénétrer dans l'univers passionnant de l'Analyse des Données.

Philippe Waniez

MacSpin: l'analyse graphique des données

En aiguisant son sens de l'observation ainsi que son intuition, on aboutit à une analyse très subtile de son information qui ne se limite pas aux structures les plus évidentes.



1. L'écran de travail de MacSpin. Au centre, le graphique tri-dimensionnel, à droite les fenêtres se rapportant aux divers objets manipulés par le logiciel. A gauche, les outils de sélection et de rotation.

MacSpin n'est pas à proprement parler un statisticiel car il ne répond que très partiellement aux différents critères énoncés dans l'introduction du présent dossier. Cependant, il s'agit bien d'un système original d'étude des données statistiques qui rendra bien des services à tous ceux qui ne veulent pas (ou ne peuvent pas) suivre les lois contraignantes de la statistique classique. Une telle conception fait de MacSpin un logiciel d'analyse exploratoire (EDA) proche de DataDesk, même s'il ne dispose pas, loin s'en faut, de son immense variété de méthodes. Comme l'indique intelligemment la documentation, on aura intérêt à utiliser MacSpin conjointement avec un véritable statisticiel afin de préciser les structures découvertes par des paramètres statistiques plus précis.

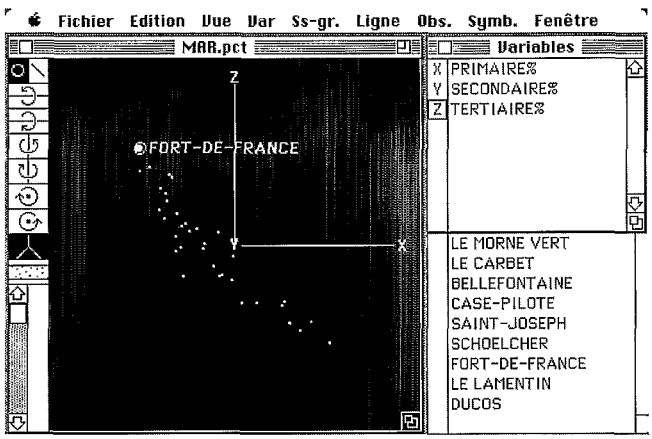
MacSpin est livré sur une seule disquette. La documentation se compose d'un seul volume de 228 pages en français. Malgré son volume réduit, la documentation apparaît très claire et repose sur des exemples facilement compréhensibles. L'effort consenti pour présenter des figures claires permet de bien saisir l'originalité de cette approche.

La fonction essentielle de MacSpin consiste en l'affichage d'un nuage de points en trois dimensions.

Cela revient à considérer chaque variable comme un axe d'un repère orthonormé, où chaque individu est un point dont les coordonnées sur les axes sont les valeurs qu'il prend sur ces variables. L'écran 1 présente le bureau de MacSpin.

La plus grande partie de l'écran est occupée par le nuage de points blancs sur fond noir, ce qui renforce l'impression de galaxie. Le système d'axes permet de savoir sous quel angle le nuage de points est observé.

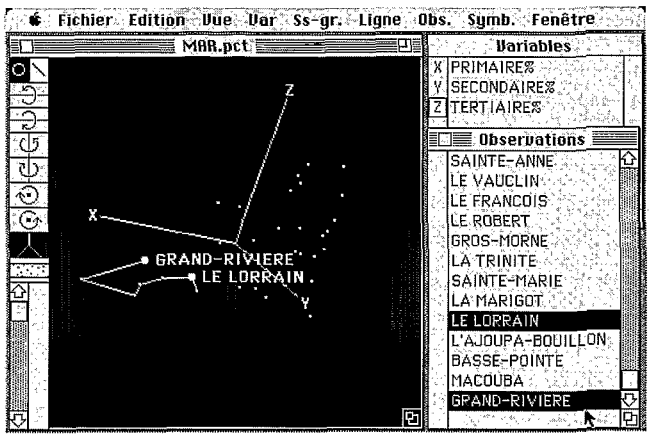
Sur la gauche, on trouve une boîte à outils qui assure les fonctions d'identification et de sélection des points ainsi que la rotation du système d'axes. Lorsqu'on clique sur un point avec l'outil d'identification (le petit cercle en haut et à gauche de la boîte à outils), son nom apparaît



2. Le graphique Primaire/Secondaire/Tertiaire après rotation autour de l'axe X. On observe très bien la corrélation négative entre le primaire et le tertiaire.

en regard. Les outils de rotation permettent de faire tourner le nuage de points autour des trois axes et, ainsi, facilitent la détection de structures intéressantes. Par exemple, en faisant pivoter la galaxie autour de l'axe X, on détecte une corrélation linéaire négative du secteur tertiaire avec le secteur primaire (écran 2).

La partie droite de l'écran est réservée à un ensemble de fenêtres dans lesquelles apparaissent les noms des variables et des individus, ainsi que des groupes d'individus sélectionnés directement sur le graphique et qui deviennent des «objets» statistiques à part entière.



3. Interaction avec le graphique : pour trouver la position d'un individu dans l'espace tri-dimensionnel, il suffit de choisir son nom dans la fenêtre des observations à l'aide de la souris.

Une fois enregistrés, ces groupes peuvent être examinés séparément et vis-à-vis d'autre variables que celles qui ont servi à les définir. La méthode d'analyse ne se limite donc pas aux seuls tableaux ternaires, mais s'étend, de proche en proche, à des tableaux plus conséquents. Cette véritable observation interactive et graphique du tableau de données est sans doute la plus grande richesse de MacSpin.

Les autres menus assurent soit le contrôle de l'environnement de travail, soit la définition et l'observation des objets statistiques.

VUE offre le choix entre plusieurs positions d'origine pour les axes et permet l'affichage des points en noir sur fond blanc ou en blanc sur fond noir. SYMB donne une palette de symboles simplifiant l'identification d'individus ou de groupes d'individus particuliers.

VAR comprend tous les articles nécessaires au recodage des variables existant déjà dans le fichier en cours d'analyse, ou à la création de nouvelles variables comme, par exemple, des pourcentages ou des rapports. SS-GR permet de réunir plusieurs sous-groupes et d'en extraire les individus qui composent leur intersection. LIGNE est un menu à utiliser conjointement avec l'outil ligne de la boîte à outils (en haut à droite).

En joignant un ensemble de points avec cet outil, on obtient une ligne brisée reliant des points aux caractéristiques proches. On peut tracer et enregistrer plusieurs lignes sur le même graphique. OBS isole ou exclut un groupe de points du graphique et permet de rechercher un point donné parmi tous les points du graphique.

Tous ces menus sont couplés avec les fenêtres de la droite de l'écran. Lorsqu'on désigne à l'aide de la flèche de la souris, un élément particulier d'une de ces fenêtres (une variable, un sous-groupe, un domaine ou un individu), le point, le groupe de points ou les lignes correspondants s'affichent en double brillance sur le graphique (écran 3). Réciproquement, lorsqu'on désigne sur le graphique un point ou un ensemble de points, il sont écrits dans la fenêtre correspondante en fond inversé (blanc sur fond noir).

Là encore, l'interaction de l'utilisateur avec ses données a été particulièrement soignée.

Au premier contact, MacSpin déconcerte un peu celui qui a une certaine habitude de l'analyse des données statistiques.

Passé le stade d'apprentissage, ce logiciel se révèle très agréable à utiliser et extrêmement convivial. Sans doute la méthode EDA y est-elle pour quelque chose, mais sa traduction informatique par MacSpin est une incontestable réussite.

Nous encourageons donc vivement tous les utilisateurs de statistiques à explorer les galaxies de leurs informations avec cet outil si plaisant.

Philippe Waniez

Les auteurs du dossier

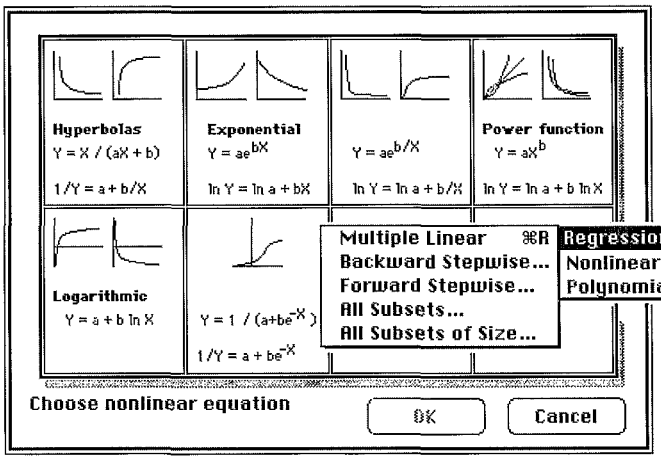
■ Micheline Cosinschi est maître-assistant, agrégée de Faculté à l'Université de Lausanne. Dans le cadre de l'Institut de Géographie (IGUL), elle est chargée de méthodes quantitatives en géographie et en cartographie automatique.

Chercheur de l'ORSTOM, Philippe Waniez mène, à la Maison de la Géographie de Montpellier (GIP RECLUS), des recherches sur la différenciation spatiale au Brésil tout en coordonnant diverses productions cartographiques sur les départements et territoires d'Outre-Mer. Il est l'auteur de «Cartographie sur Macintosh» aux Editions Eyrolles.

Micheline Cosinschi et Philippe Waniez ont publié «Pratique de l'analyse statistique, SAS sur PC/PS, minis et Gros systèmes», au GIP RECLUS.

Exstatix : la statistique extensive

Exstatix est l'un des derniers nés de la famille des statisticiels. Il occupe une position intermédiaire, le plaçant après les statisticiels classiques, tels _Systat_ ou _Statview_, et le rapprochant de ceux de l'analyse exploratoire, tel _Data Desk_.



Exstatix *offre une large gamme de modèles de régression non-linéaire.*

Cet «Expandable Statistical Analysis System», expression à l'origine du terme *Exstatix*, permet à un utilisateur averti de créer ses propres extensions au statisticiel, à la condition de savoir programmer en Pascal ou en C par exemple. On peut ainsi ajouter de nouveaux articles aux menus pour exécuter des fonctions ou commandes spécialisées qui vous sont propres.

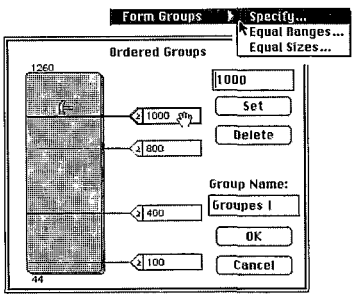
Il s'agit d'un des rares statisticiels qui permette à un utilisateur d'intégrer ses procédures favorites (analyse factorielle ou intégration à un tableur, par exemple). Bien qu'intéressante, cette nouvelle facilité n'est cependant pas à la portée de tout le monde et pour la grande majorité d'entre nous, il faudra attendre que des groupes d'utilisateurs mettent à disposition ce genre de développement pour enrichir la liste des outils disponibles sous *Exstatix* puisque Select Micro Systems Inc. n'a pas l'intention de développer lui-même un tel marché.

Visualiser les données à sa convenance

Exstatix offre trois possibilités de gestion et de visualisation des données à convenance, la sélection se faisant par le menu *View*. La plus classique, *by TABLE*, permet une représentation sous forme de tableau chiffré, *by ICON*, identifie chaque variable par une icône rectangulaire. On déplace et réorganise l'agencement des variables pour faciliter les sélections ou mettre de l'ordre dans la liste. La troisième, *by LIST*, permet de visualiser chaque variable sur une ligne où apparaît le texte descriptif la définissant. Chacun de ces trois modes de gestion donne la possibilité, à l'aide des petits boîtiers, de définir les variables dépendantes (les Y, à ex-pliquer) et indépendantes (les X, explicatives) ; une variable alphanumérique est notée d'un A tandis que les variables inactives sont en grisé et les données manquantes annotées d'un • visible en mode *TABLE*.

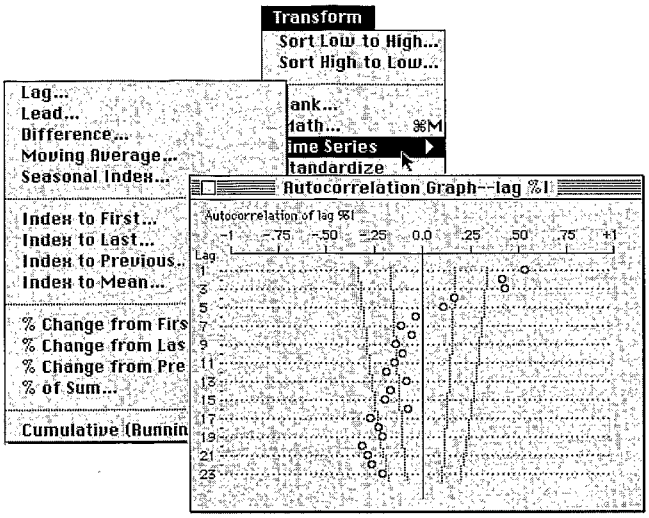
Des traitements dignes des gros systèmes

Tout à fait comparable à ses concurrents sérieux tels *Statview II* ou *Data Desk Professionnal*, *Exstatix* offre une gamme très large de procédures statistiques dont les résultats sont riches et complets, allant même jusqu'à diagnostiquer «en bon anglais» la signification des tests et des comparaisons. On trouve sous le menu *STATISTICS* un ensemble d'analyses de régression, à la fois linéaire et non-linéaire, simple, multiple ou pas-à-pas dont la gamme des résultats est digne des statisticiels sur gros ordinateurs. Les corrélations et autocorrélations, statistiques descriptives, tableaux croisés complètent le menu. On n'y trouvera pas cependant d'analyses multivariées de la famille des analyses factoriel-



Une manière originale et commode de partitionner une variable : entre les valeurs minimum et maximum, glissez une petite main pour définir une limite de classe, donnez un coup de marteau pour ajouter une nouvelle partition et* Exstatix *crée automatiquement une nouvelle variable catégorielle.

les ou typologiques, et c'est regrettable. Par contre, le menu *TEST* fournit une belle batterie de tests inférentiels ou non-paramétriques dont on peut définir librement le niveau de signification, du moins pour les tests F, chi et t. *Exstatix* permet de définir le niveau de détail des calculs et donc d'affichage des résultats. Vous pouvez également spécifier si vous désirez revenir à ces fenêtres de dialogue chaque fois que vous réalisez une analyse. Il s'agit d'une option utile lors des calculs répétitifs.



***L'option* Time Series *du menu* Transform *permet l'analyse des séries chronologiques qui peuvent par la suite être visualisées par l'option* Autocorrelation *du menu* Statistics.**

Il est possible de manipuler les données, sous le menu *TRANSFORM*, pour leur faire subir des transformations algébriques ou fonctionnelles. Plusieurs facilités y sont offertes pour mettre en rang, trier, standardiser, générer des nombres aléatoires (six lois sont disponibles), effectuer des transformations mathématiques ou encore définir des groupes. A ce titre, on découvre une méthode originale pour grouper des données en catégories en utilisant une approche visuelle (figure n°5): l'option *GROUP* permet ainsi la création de N partitions des données sur une variable selon votre propre critère, selon des effectifs égaux ou selon des amplitudes égales. Une main permet de varier les limites de classes tandis qu'un marteau ajoute une nouvelle partition. Il suffit de cliquer-glisser pour établir des catégories. Sous l'item *TIME SERIES* on découvre un ensemble de quatorze fonctions pour analyser des données temporelles dont l'autocorrélation peut être visualisée sous forme graphique.

Des graphiques classiques (et sans surprise)

Gérant la couleur, *Exstatix* met à disposition une variété de représentations graphiques dans le menu *GRAPH*. Il offre, pour une variable, la possibilité de faire des graphiques séquentiels, des histogrammes cumulés ou non, et des diagrammes en secteurs. Toutes ces représentations utilisent les données brutes ou les données groupées lorsque vous désirez visualiser vos propres limites de classes. Pour traiter deux variables, on peut utiliser les graphiques de nuages de points, les bâtonnets 3D (plus simples que dans *MacSpin* ou *Data Desk* cependant) ou les boîtes et moustaches. Les graphiques bi-dimensionnels offrent à gauche de la fenêtre d'affichage, une série d'outils d'interaction : changer l'échelle des X ou/et des Y en logarithme, dessiner la droite de régression et calculer la corrélation, standardiser les axes, lier les points et enfin, intervertir les axes. On retrouve d'autres outils appropriés également sur les représentations 3D. Lorsque vous travaillez sur trois variables ou plus, *Exstatix* peut les représenter en nuage de points tri-dimensionnel. Ici encore, la fenêtre met à disposition, à gauche, des outils pour dessiner un cube autour du nuage, tracer les axes X, Y et Z, leur affecter un libellé, ajouter une perspective, ou encore retravailler la rotation. Les graphiques produits par Ex*statix* sont des objets et peuvent être sauvés en PICT ou copiés dans le presse-papier pour une édition subséquente dans une autre application.

Des outils originaux de mise en page

Un des grands avantages de *Exstatix* par rapport à ses concurrents est d'offrir des résultats éditables en format TEXT. Faites une analyse, une fenêtre déroulante de résultats s'affiche et vous pourrez à loisir les modifier ou les traduire en français, soit directement, soit en les copiant dans un traitement de texte. Ces fenêtres peuvent être automatiquement datées et inclure un titre standard par le menu *HEADER*; vous pouvez choisir les polices de caractères, leur taille, leur style et la couleur pour soigner la présentation. Outre l'édition des résultats numériques, le statisticiel met à disposition des fenêtres graphiques spéciales, appelées «*Layout Window*» qui vous permettent de combiner textes et graphiques. On peut y copier des graphiques tels un histogramme ou un nuage de points et les coller dans une ou plusieurs fenêtres de présentation. Le texte de n'importe quelle autre fenêtre de résultats peut également être sélectionné, copié et collé avec les graphiques.

Dans une telle fenêtre de présentation, vous pouvez arranger ces items à votre convenance et utiliser un des outils de dessin disponibles.

Un système intéressant, surtout dans le domaine des tests statistiques

Exstatix est un programme qui, malgré son orientation au goût très «business», devrait plaire à un large public. On y trouve en particulier un excellent répertoire de tests et des résultats statistiques bien développés, permettant une riche évaluation numérique des traitements statistiques.

Des possibilités ou fonctions d'édition uniques par rapport à d'autres statisticiels le rendent encore plus attrayant. Les spécialistes pourront programmer des procédures externes afin de répondre à leurs besoins spécifiques ; c'est ce qu'il faudra faire si vous désirez obtenir des analyses multivariées ou une interface particulière.

Exstatix a néanmoins quelques limites. Signalons l'impossibilité d'importer automatiquement des fichiers de données au format SYLK ou DIFF propres aux tableurs (il importe ses propres fichiers ou des fichiers TEXT en code ASCII). Les chercheurs et le milieu académique le trouveront incomplet, par rapport à *Systat* par exemple, même s'il s'avère très riche pour des analyses sur une ou deux variables ainsi qu'au niveau des tests statistiques, tout en combinant d'assez bonnes représentations graphiques.

La documentation est correcte, mais est loin d'être aussi complète que celle de *Data Desk*, la référence en la matière.

Micheline Cosinschi

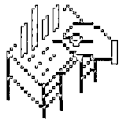
Quelques livres

Statlab. HODGES J.R., D. KRECH, R.S. CRUTCHFIELD (1979) Paris, Economica, 373 p.

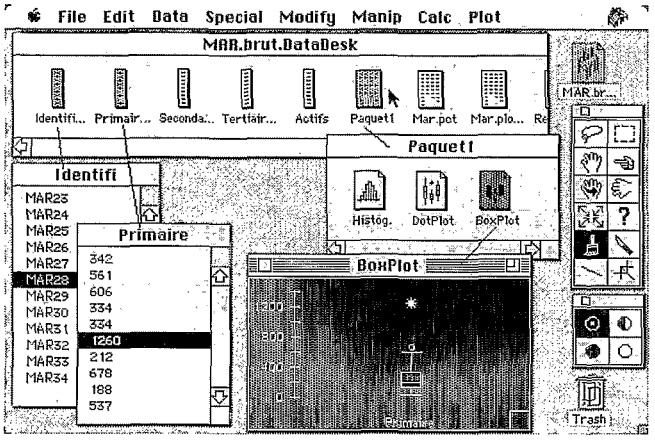
Initiation pratique à la statistique. LIORZOU A. (1973) Paris, Eyrolles, 314 p.

Qu'est-ce que l'analyse des données. FENELON J.P. (1981) Paris, Lefonen, 311 p.

Data Desk : la statistique relationnelle



Développé à l'Université de Cornell pour des besoins liés à l'enseignement de la statistique, *Data Desk Professional* a été repris par Odesta Corporation qui l'a commercialisé pour un plus large public.



1. Le bureau de* Data Desk.** ***Une représentation iconique des variables et des paquets de travail (les «bundles») permet une manipulation très flexible d'un environnement qui rompt avec les tableaux habituels.

Data Desk est l'un des premiers statisticiels cherchant explicitement à mettre en pratique la démarche de l'Analyse exploratoire des données (EDA) et les idées de J. Tukey (de l'Université de Princeton et des Laboratoires AT&T Bell) connaissant un succès croissant sur le vieux continent. Moins normative que la statistique classique, l'EDA reconnaît que nous n'avons très souvent que peu ou prou d'hypothèse forte à tester au départ ; nous cherchons d'abord à voir ce qui se passe dans nos chiffres, sans à priori. J. Tukey propose de les examiner comme un détective examinerait la scène d'un crime : gardant l'esprit ouvert, cherchant, un indice après l'autre, les vérités enfouies sous la masse des données. A ce titre, *Data Desk Professional* fournit tous les outils pour manipuler et inspecter visuellement vos données d'une manière nouvelle et intuitive tout en offrant les procédures d'analyse statistique classique.

Data Desk Professional V2.0 est livré sur deux disquettes, l'une contenant le programme et un utilitaire d'accès au SGDB *Double Helix*, l'autre le fichier d'aide, des exemples et un utilitaire de gestion des fichiers. La documentation très complète, et même exemplaire, se compose de trois volumes : un petit manuel d'introduction rapide, « Handbook », un excellent manuel décrivant comment analyser les données dans l'esprit EDA, définissant les termes et concepts statistiques, et enfin « Statistics Guide » et « Reference Guide », le premier volume décrivant les méthodes de la statistique confirmatoire, le second fournissant un très bon guide des menus.

Data Desk peut prendre en compte le coprocesseur 68881, et utilise la couleur dans sa version 3.0, s'interface avec *Double Helix II* et peut importer des données de tableurs ou d'autres bases de données de même que des fichiers ASCII en provenance de gros ordinateurs. Il peut analyser des données provenant d'un VAX, transférées dans votre micro-ordinateur à l'aide de *Helix VMX*.

Data Desk fournit les outils graphiques essentiels à la visualisation des structures et des relations entre les nombres dans un environnement où il est possible de traiter dynamiquement l'information et de relier les différents traitements entre eux. Ce statisticiel intègre aux outils graphiques des procédures de transformation des données et de statistiques descriptives, les tableaux de contingence et chi-deux, les tests de comparaison de moyennes ou de variances, et le calcul des intervalles de confiance; les modèles linéaires, les corrélations paramétriques et non-paramétriques, les régressions simples et multiples, linéaires et polynominales, avec ajouts et suppressions dynamiques des variables dans l'équation ; l'analyse de la variance jusqu'à trois facteurs sur des plans équilibrés ou non-équilibrés ; l'analyse des résidus et le calcul des valeurs prévues. Si d'autres analyses multivariées de typologie et d'analyse en composantes principales complètent les menus statistiques, elles sont cependant plus sommaires.

La gestion des variables se fait au moyen d'icônes. Celles-ci représentent des colonnes de données et s'ouvrent en fenêtres permettant de visualiser et d'éditer des chiffres et du texte. Des variables allant ensemble peuvent être regroupées en paquets, analogues aux dossiers du Finder.

Ces paquets (appelés «bundles») sont aussi utilisés pour organiser une collection de résultats graphiques ou tabulaires et sont à la base de la gestion du travail sous *Data Desk* (écran 1). Si vous êtes un habitué des tableurs, il vous faudra cependant un peu de pratique pour maîtriser cette interface qui s'avère d'une grande flexibilité à l'usage, mais qui demande un peu de réflexion et de rigueur pour classer (et retrouver !) les nombreux résultats qui s'empilent très vite.

L'interactivité prend tout son sens dans le traitement de l'information. Les tableaux et graphiques sont inter-reliés, une sélection d'une partie d'un graphique met en évidence les données correspondantes sur les autres représentations; on mesure la véritable puissance du statisticiel à travers ces réponses dynamiques aux actions que l'on fait sur les résultats (écran 2). *Data Desk* offre à ce titre, dans le menu *Modify*, deux palettes pour travailler sur les graphiques : la première, *Plot Tools*, fournit douze outils pour manipuler, déplacer, lier, isoler, identifier ou sélectionner des sous-ensembles de données ; la seconde, *Selection Modes*, l'accompagne pour gérer quatre modes de sélection des données. Même si ce statisticiel n'est pas dédié à des présentations graphiques dignes d'être immédiatement publiées, il fournit l'ensemble des modules courants pour visualiser vos données : diagrammes en bâtons et graphiques en secteurs, boîtes et moustaches (les «box plot»), histogrammes, nuages de points, courbes et même les graphiques rotatifs en trois dimensions de nuages de points et de plans, avec affichage et mise à jour des équations de projection au fur et à mesure de la rotation, cette dernière possibilité le mettant presque sur le même pied que *MacSpin* (écran 3). Evidemment tous ces graphiques peuvent être exportés vers d'autres logiciels pour la touche finale.

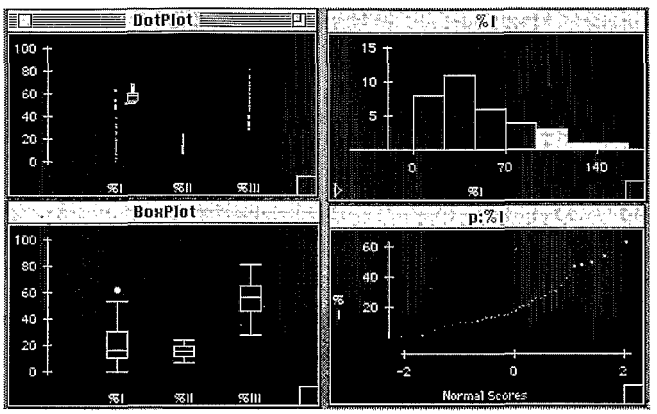
Certaines parties de graphiques et tableaux proposent des sous-menus qui suggèrent des graphiques et analyses apparentés (on retrouve là l'une des idées de base d'HyperCard). Cliquez un bouton d'hypervue (un petit triangle) dans une fenêtre de résultats ou cliquez la petite main au niveau d'un résultat, on vous offre de continuer plus avant l'analyse dans un sous-menu adapté.

La plupart des logiciels de statistique traditionnels ont fini quand ils ont imprimé un graphique ou un tableau. Pas *Data Desk*. Ici, les tableaux et graphiques ne font que démarrer votre analyse. Faites glisser une nouvelle variable dans la fenêtre de régression ; substituez une variable à une autre dans une analyse ou un graphique... et tout est recalculé (écran 4).

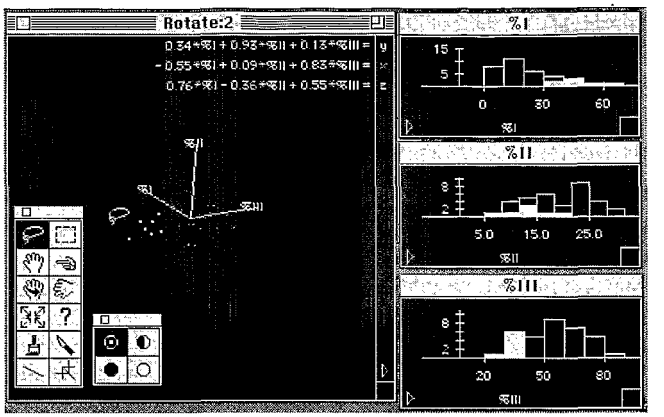
Grâce aux outils disponibles sous *Data Desk*, vous saurez faire «parler» vos données ! *Data Desk* s'écarte des logiciels classiques en mettant l'accent sur l'exploration des données (tant graphique que statistique) plutôt que sur l'interprétation numérique de l'information. On peut regretter que *Data Desk* ne soit pas complet au niveau des tests non-paramétriques ou des analyses multivariées qui restent sommaires et qui n'offrent évidemment pas de modèles d'analyses des données «à la française». *Data Desk* s'adresse surtout à un utilisateur semi-professionnel (chercheur, gestionnaire, enseignant, etc.), mais néophytes comme professionnels y découvriront une large palette d'outils interactifs pour traiter l'information, tant au niveau de l'analyse exploratoire que de la statistique confirmatoire plus classique. Il est d'un grand confort d'utilisation, même si son environnement de travail très iconique peut paraître déroutant au départ.

La documentation qui l'accompagne est de premier niveau, dépassant le simple lexique d'un mode d'emploi de programme pour toucher au vif du sujet de la statistique. Data Desk propose également une aide en ligne très claire. Appartenant à une toute nouvelle génération de statisticiels *Data Desk* est un gagneur ; ses concurrents les plus sérieux pourraient être *Exstatix* ou encore *JMP*.

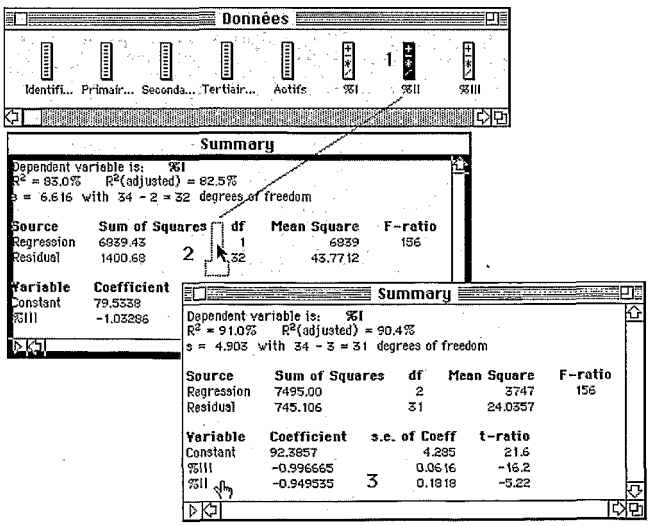
Micheline Cosinschi



2. Les graphiques de* Data Desk. *Une gamme très complète de visualisation de l'information. Ce qui est sélectionné sur un graphique se retrouve représenté instantanément sur les autres fenêtres qui y sont reliées, même au niveau des données.

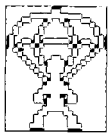


3. Les graphiques rotatifs permettent de visualiser l'information en trois dimensions. Les symboles font ressortir les sous-groupes et les points contrastés se retrouvent dans les autres fenêtres d'édition. On peut afficher et mettre à jour les équations de projections au fur et à mesure de la rotation.

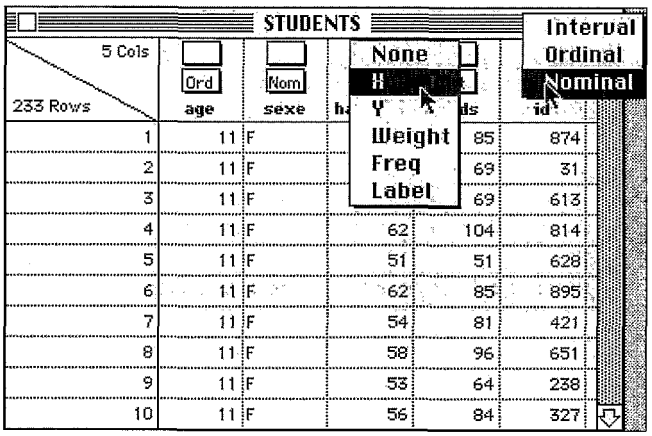


4. Une intégration instantanée des manipulations: sélectionnez une variable (1), faites-la glisser (2) et observez le nouveau résultat (3)!

JMP : un saut dans l'analyse exploratoire



Logiciel de visualisation des données statistiques, JMP (prononcer "jump") va bien au-delà de la seule réalisation de graphiques sophistiqués. Il offre une vaste panoplie d'outils orientés vers l'analyse exploratoire.



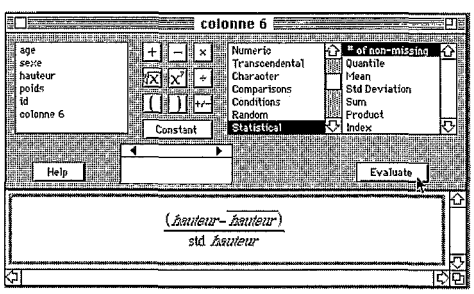
1. Le tableau de JMP. Les menus de définition des échelles de mesure et de choix du rôle des variables dans l'analyse sont déroulés.

La réputation mondiale (n'ayons pas peur des mots) de SAS Institute n'est plus à faire. Depuis de nombreuses années, cette grosse société américaine disposant d'une assise planétaire (SAS Institute S.A. est son représentant en France) développe et diffuse le principal logiciel d'analyse statistique du marché.

Le Système d'Analyse Statistique (Statistical Analysis System, SAS) apparaît aujourd'hui comme le statisticiel le mieux diffusé et le plus complet du marché. Son langage de programmation est devenu la langue commune de nombreux statisticiens. On comprend donc qu'avec JMP, l'entrée (le saut!) de SAS Institute dans le monde du Macintosh constitue un événement. Disons tout de suite que JMP n'est pas SAS, ni même l'un des nombreux modules de ce système. Il s'agit en fait nous dit l'éditeur d'un *«prototype de ce que sera prochainement le système SAS en matière de statistiques et de représentations interactives des données»* sous les systèmes d'exploitation IBM, DEC, UNIX, DOS et OS/2. Pour les actuels utilisateurs de SAS, indiquons que JMP assure l'essentiel des traitements offerts par les procédures PRINT, FREQ, UNIVARIATE, GLM (dans tous les domaines des méthodes des moindres carrés généralisés), TTEST, LOGIST et SORT. On y trouve aussi une grande partie des opérations assurées par l'étape DATA.

JMP est livré sur deux disquettes. La première renferme l'application proprement dite et un dossier d'exemples ne contenant pas moins de 24 fichiers dans des domaines d'utilisation très divers. La seconde disquette contient un fichier d'aide pouvant être consulté à partir de l'application. La documentation se compose d'un volume de 464 pages en anglais (SAS Institute France nous offrira-t-il un jour une version en français?). Cette documentation est très bien conçue. *Getting started* introduit aux principales opérations nécessaires au fonctionnement du système. Suit le *Guide de référence* où chaque menu déroulant fait l'objet d'un examen détaillé. Enfin, le troisième et dernier chapitre expose le mode de «navigation» sur les plates-formes d'analyse. On reconnaît dans la conception de cette documentation, et dans le soin apporté à sa réalisation, un professionnalisme qui a fait ses preuves.

Les tableaux de données utilisés par JMP sont semblables à ceux de SAS (une option importe directement des fichiers SAS en format de transport) : les données sont organisées sous forme de tableaux rectangulaires où les lignes figurent les observations, et les colonnes les variables. Les valeurs peuvent être numériques ou alphanumériques et expriment des mesures réalisées sur des échelles d'intervalles, ordinales,



2. Utilisation du calculateur pour centrer et réduire une variable et enregistrer le résultat dans une nouvelle variable.

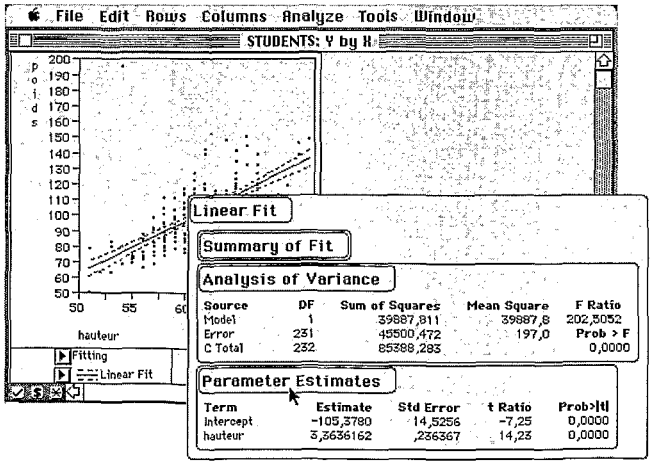
ou nominales. Ces échelles peuvent être modifiées (pour peu que cela ait un sens) et sont prises en compte pour le choix d'une méthode d'analyse (écran 1).

Par exemple, dans le cas d'une étude de causalité, si la variable endogène (Y) relève d'une échelle d'intervalle, JMP procédera à une régression multiple, alors que si l'échelle est ordinale ou nominale une régression logistique sera directement calculée. Il en est de même pour les variables explicatives (X) : la sélection de l'échelle nominale ou ordinale sera traitée comme une variable de classification avec un nombre de degrés de liberté égal au nombre de modalités. De ce point de vue, JMP possède une «intelligence» qui le distingue de ses concurrents.

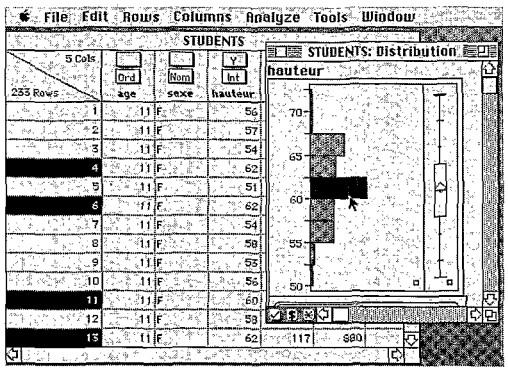
La conception du tableau de JMP a été particulièrement soignée. Cette qualité s'exprime sur de nombreux points. En premier lieu, on trouve le concept de statut d'une ligne. Le menu *Rows* permet d'affecter des caractéristiques aux observations sélectionnées, caractéristiques qui seront activées lors des traitements ; les observations peuvent être exclues ou incluses dans l'analyse, cachées ou visibles, identifiées par un libellé, colorées ou marquées. Des icônes indiquent pour chaque ligne quel statut la caractérise. En second lieu, la préparation du tableau de données en vue de son analyse est dotée d'outils originaux comme, par exemple, le calculateur. Celui-ci permet de créer toute nouvelle variable à partir des variables d'origine. Par exemple, si une nouvelle variable nommée «colonne 6» doit contenir les valeurs centrées réduites de la variable «hauteur», on écrira la formule avec le calculateur (écran 2) : celle-ci s'affichera en clair dans la partie inférieure de la fenêtre. Cette capacité à comprendre les formules fait de JMP un outil très sympathique.

Ajoutons qu'avec la commande *JOIN*, JMP sait joindre différents tableaux de données pour constituer un nouveau tableau, et cela de plusieurs manières : selon les numéros de lignes, selon les valeurs d'une variable commune, et selon la «méthode cartésienne» nécessaire à la définition des tableaux de contingence.

Le menu *Analyse* donne accès à l'une des six plates-formes d'analyse statistique proposées par JMP. Une plate-forme est une fenêtre interactive permettant d'analyser les données, d'explorer les graphiques et d'enregistrer les résultats obtenus. Pour réaliser une analyse, il faut procéder en deux étapes : choisir l'échelle de mesure et la fonction de chaque variable (endogène ou exogène, X ou Y) dans le tableau, puis sélectionner l'une des six plates-formes suivantes :



4. Régression linéaire simple. Des boutons font successivement apparaître les diverses parties de l'analyse, comme celle de la variance ou l'estimation des paramètres de la régression.



3. Etude de la distribution statistique de la variable HAUTEUR. Des ascenseurs donnent accès à l'ensemble du traitement, graphiques (histogrammes, box plots) et paramètres statistiques. Notons que lorsqu'on clique sur un des bâtons de l'histogramme, JMP souligne dans les tableaux les observations comprises dans cet intervalle de valeurs.

- distribution des Y : décrit la distribution de chaque variable Y à l'aide d'histogrammes ainsi que d'autres graphiques et paramètres statistiques (écran 3).
- distribution des Y et des X : décrit chaque paire de variables (X,Y).
- ajustement des Y par les X : ajuste une variable Y par toutes les variables X, et cela conformément aux échelles de mesure adoptées. Selon le cas, il s'agit de régression, d'analyse de la variance, d'analyse de la covariance ou bien encore de modèles d'ajustement des données catégorielles (écran 4).
- spécification d'un modèle : permet de définir les termes d'un modèle complexe tout en indiquant la nature des effets et des termes d'erreur.
- SPIN : produit un graphique en trois dimensions pouvant être examiné sous divers angles afin de détecter des regroupements ou des corrélations.
- Y par Y : calcule les corrélations entre les variables Y.

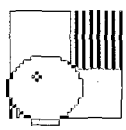
Tous les articles du menu *Analyse* se caractérisent par une interactivité poussée à l'extrême. D'une part, l'utilisateur peut cliquer sur les éléments des graphiques pour mieux les appréhender : il obtient comme réponse l'identification des observations par leur numéro d'ordre et leur signalement (en blanc sur fond noir) dans le tableau. Pour les sorties numériques, de multiples boutons permettent de visualiser les éléments nécessaires à l'analyse et, le cas échéant, de les imprimer. De nombreuses options donnent accès à une très grande variété de traitements.

La récente sortie de JMP confirme le très grand dynamisme de l'école d'analyse exploratoire EDA et la très bonne adaptation de l'interface du micro aux icônes à ces méthodes d'analyse statistique. En tous points remarquable, JMP surprend par l'intelligence qu'y ont mis ses auteurs et l'extraordinaire imagination dont ils ont fait preuve dans la conception de ce statisticiel.

Entre DATA-DESK et JMP, le choix semble très difficile, mais JMP a pour lui d'exister en version limitée à 500 cellules (JMP-IN), ce qui facilitera sa diffusion dans le monde de l'enseignement, de la formation permanente et auprès de tous ceux qui souhaitent se faire une bonne idée des méthodes très attrayantes (voire amusantes) de l'analyse exploratoire. La remarquable aide en ligne devrait les y aider. Même les utilisateurs de SAS sur gros et mini-systèmes auront intérêt à acquérir JMP pour explorer des échantillons issus de leurs bases de données, préalablement à tout traitement exhaustif et plus systématique nécessitant des ressources informatiques plus importantes que ce que peut offrir un micro-ordinateur. L'existence de JMP ne comble pas le vide occasionné par l'absence de SAS sur notre machine préférée, mais il permet de patienter.

P. Waniez

Autres logiciels de statistiques en bref



La diversité des logiciels de traitement des données est telle qu'il apparaît difficile de tracer une limite précise entre les statisticiels et les autres. Voici quelques produits complémentaires.

L'analyse statistique touchant à tous les domaines des sciences, de la physique à la biologie, en passant par l'économie et la linguistique, il ne faut pas s'étonner du bouillonnement que connaît la production de logiciels dans ce domaine. Les différents articles qui constituent ce dossier ne peuvent pas en rendre compte totalement. Voici d'autres produits dont les domaines d'application sont en général moins larges que ceux analysés précédemment.

☐ FASTAT coûte environ deux fois moins cher que son frère aîné, SYSTAT. Il apparaît idéal à tous ceux dont la statistique n'est pas le métier et qui, de plus, ne souhaitent pas s'engager dans l'apprentissage d'un langage de programmation. Doté d'une feuille de calcul aux fonctions très limitées, il dispose d'une panoplie raisonnable de méthodes d'analyse, comme les tests non-paramétriques de Wilcoxon et de Kruskal-Wallis, les techniques de régression linéaire et d'analyse de variance (ANOVA), les procédés d'analyse des séries chronologiques (lissage, désaisonnalisation, autocorrélation), et bien entendu, les paramètres statistiques habituels (moyenne, etc.). Sur le plan graphique, FASTAT propose un assortiment d'outils d'habillage des graphiques statistiques, y compris la couleur.

☐ SHERLOCK est un programme de création et de traitement d'enquêtes et de sondages réalisé par la société française KYNOS et réalisé avec le logiciel de base de données relationnelles Quatrième Dimension. Pour un coût d'environ 5000 francs, l'utilisateur dispose d'une panoplie d'outils nécessaires à la gestion et l'interrogation d'une enquête. De manière classique avec 4D, il faut, préalablement à toute opération, décrire la structure de l'enquête, c'est-à-dire définir les types de questions, (fermées, à modalités simples, multiples et numériques et même ouvertes), ainsi que les écrans de saisie. Le concepteur de l'étude a ainsi le loisir de concevoir son enquête tout en imaginant son informatisation, ce qui constitue sans aucun doute un progrès. La saisie, conviviale, peut être faite par toute personne qui connaît le maniement de base du Macintosh, argument de poids lorsqu'on sait quels goulots d'étranglement cette phase de traitement occasionne en général.

SHERLOCK propose un module de recodage soit en cours de saisie, lorsqu'apparaissent des incohérences dans le codage des réponses, soit durant la préparation des traitements statistiques. L'analyse et l'édition des résultats statistiques sont réduites à leur plus simple expression : tris à plat, tris croisés avec tests du Khi-deux, paramètres élémentaires de distributions.

Heureusement, SHERLOCK dispose d'une fonction d'exportation vers d'autres logiciels, EXCEL en particulier. Nul doute que SHERLOCK rendra de nombreux services à tous ceux qui s'intéressent à ce que les gens ont dans la tête. A noter qu'une nouvelle version, utilisant 4D4, est en cours d'élaboration. Lire article détaillé dans Icônes n° 13.

☐ STATCALC, de la société Clear Lake Research, utilise les fonctions d'HyperCard pour calculer un nombre réduit de statistiques, paramètres des distributions et test T et F. Clear Lake Research propose aussi CLR ANOVA, un programme d'analyse de variance très complet pouvant prendre en compte jusqu'à 10 facteurs.

☐ RATS, de VAR Econometrics, est un logiciel sophistiqué d'analyse économétrique proposant les diverses méthodes couramment utilisées dans ce domaine. A partir de diverses techniques de régression, il permet de procéder à des estimations dont la validité peut être testée avec toute une panoplie de tests statistiques. Il comprend également les méthodes de traitement avancées des séries chronologiques comme Box-Jenkins, ARIMA, les modèles autorégressifs, ou la régression non-linéaire. On y trouve enfin un module d'analyse spectrale avec transformées de Fourier.

☐ STATISTICS FOR EXCEL, de Heizer Software, comprend un ensemble de macros utilisables avec le tableur de Microsoft. Elles permettent de réaliser des régressions, de calculer des coefficients de corrélation et des tests non-paramétriques, de construire des tableaux croisés, et de mener des analyses de variance (ANOVA).

☐ TRUESTAT se compose d'un ensemble de modules écrits en langage BASIC pouvant être appelés par l'excellent langage

TRUE BASIC. Il s'adresse plus particulièrement aux étudiants qui veulent se familiariser avec la programmation dans ce langage, tout en l'appliquant au traitement des données. La même société propose aussi CHIPENDALE qui est un programme de construction de tableaux croisés s'adressant particulièrement aux sciences sociales.

MONTE- CARLO SIMULATION d'Actuarial Micro Software est un système professionnel qui s'adresse plus particulièrement aux bureaux d'études des assurances. Grâce à diverses procédures de simulation, il permet de mesurer le risque encouru par une activité donnée, en faisant varier les contraintes extérieures (la météorologie, par exemple).

STATISTICS MODULES de Lionheart Press comprend cinq modules différents qui couvrent l'analyse de la variance, les séries chronologiques, l'analyse économétrique, les paramètres courants.

MAC-SAIF est un logiciel français d'analyse des données que son distributeur STATMATIC n'a pas voulu nous communiquer. Indiquons simplement qu'il comprend l'analyse des correspondances et la classification automatique, ainsi qu'un module de cartographie, le tout pour environ 20 000 francs.

CRYSTAL BALL occupe une place particulière parmi les nombreux logiciels d'analyse statistique. Il se présente comme un «programme de prévision et de gestion des risques». Il permet de répondre à des questions comme «*Quelle chance avons nous de terminer tel projet dans les délais prévus*», ou bien encore «*si nous ajoutons telle possibilité à notre réalisation, pourrons-nous respecter notre budget*», etc. Ce système offre de nombreuses possibilités de simulation par la méthode de Monte Carlo avec une grande variété de distributions de probabilités (binomiale, de Poisson, uniforme, etc.). Les données de départ sont saisies dans une feuille de calcul et l'utilisateur peut enregistrer les résultats de ses simulations successives afin d'apprécier les conséquences probables de ses choix multiples

Dernière heure. A l'image des autres champs d'application de l'informatique, le monde des statisticiels change très vite.

Nos informations en provenance directe des Etats-Unis nous incitent à vous signaler la sortie prochaine d'un logiciel qui va sans doute occuper l'une des premières places parmi les systèmes d'analyse statistique. En effet, SPSS Inc. annonce la sortie d'une version pour Mac SE et Mac II dès 1990. Elle devrait comprendre toutes les fonctions de l'actuel SPSS-PC+; cependant des modules supplémentaires seront disponibles ultérieurement afin de conduire à un produit semblable à SPSS-X, le logiciel phare de cette société, celui qui fonctionne sur gros et mini systèmes. Cette version avancée incluera entre autres les analyses multidimensionnelles. Des interfaces devraient permettre d'utiliser toutes les possiblités graphiques de *Cricket Graph*.

Enfin, notons que l'Université Carnegie-Mellon (USA) travaille à une conversion du célèbre logiciel MINITAB. Bien qu'aucune date de sortie ne soit connue, on nous promet une implantation de l'ensemble de la version 6.1 offrant toutes les possibilités courantes de MINITAB.

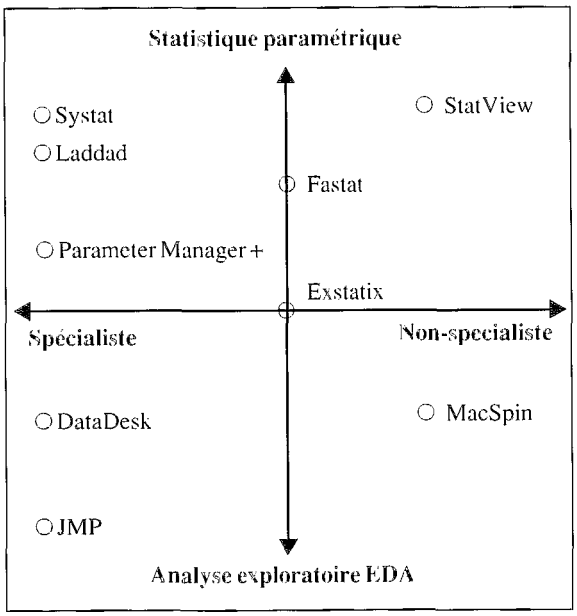
Nous ne manquerons pas de rendre compte dans nos colonnes de ces importantes nouveautés, dès que nous aurons pu en tester l'intérêt et la qualité.

P.W.

Service lecteur P 26 page 88

Statistique et souris : un mariage prolifique

Comment choisir le statisticiel répondant le mieux à vos besoins.



Le positionnement des logiciels

Par specialiste on entend un praticien de la statistique, qui fait du traitement des données l'essentiel de son travail.

La partie supérieure du graphique définit le domaine de la statistique paramétrique. StatView est sans doute le plus simple d'emploi, alors que Systat requiert une connaissance de de la programmation. Fastat et Parameter Manager Plus ne proposent qu'un nombre limité de techniques d'analyse. Laddad occupe une place à part dans l'analyse des données.

La partie inférieure du graphique représente le domaine de l'analyse exploratoire. JMP est sans doute le plus complet en ce domaine. DataDesk le rejoint sur ce plan, tout en proposant une tres grande variété de méthodes parametriques. Exstatix se situe à mi-chemin et se presente comme une bonne synthèse des diverses tendances. MacSpin n'offre qu'un éventail limité de techniques statistiques.

Au cours des quelques mois de travail qu'a nécessité cette étude, de nouveaux produits sont sans cesse apparus donnant l'impression d'une compétition très rude entre les fabricants de logiciels. Contrairement à ce qu'on a pu observer dans le passé, à propos des PC, les concurrents appartiennent moins à la classe des développeurs «poids lourds» qu'à celle des petites sociétés créatives qui ont rapidement compris qu'elles pourraient occuper ce «créneau», trop longtemps laissé vacant, de la statistique. De ce processus découlent, sans doute, les principales caractéristiques des statisticiels disponibles aujourd'hui sur le marché.

Force est de constater qu'il y en a pour tous les goûts, entendez par là qu'il existe un très large spectre d'application de ces logiciels, qui vont du monde des affaires au contrôle de processus, en passant par tous les domaines qui nécessitent l'analyse d'informations numériques.

Mais ces logiciels sont encore perfectibles, en particulier sur le plan des techniques d'analyse qu'ils proposent. De fait, aucun d'eux ne couvre entièrement l'ensemble d'un groupe de méthodes, comme par exemple, l'analyse multivariée, bien déficiente dans la majorité des cas. De plus, l'analyse des données ne permet que rarement d'assurer des traitements répétitifs qui sont le lot quotidien de nombreux analystes. Ce clivage, très important, devrait entrer pour une part non négligeable dans le choix d'un logiciel. De même, la rareté des connexions bases de données/statisticiels doit inciter à la plus grande circonspection, non seulement dans le choix d'un logiciel, mais également à propos du type d'informatique à mettre en œuvre.

Pour vous éviter le casse-tête que représente un choix difficile à faire, le tableau ci-contre résume les fonctions assurées par les principaux produits disponibles. Le prix reste un critère important car il varie de 1 à 3 environ. Ensuite, mis à part LADDAD, Parameter Manager et MacSpin (qui répondent à des besoins particuliers), on peut remarquer qu'il n'y a pas de différences notables entre les méthodes disponibles : seuls des détails, parfois importants, mais n'apparaissant pas dans ce tableau synoptique, peuvent justifier tel ou tel choix. L'essentiel reste pourtant le parti pris EDA ou non-EDA, et la convivialité plus ou moins bien mise en valeur.

M. C. & P. W.

	Exstatix	DataDesk	Statview	Systat	Laddad	Parameter	MacSpin	JMP
Version	1.01	prof. 2.0	SE+Gra.	3.2		Manager +	2.0	1.0
PRIX (environ)	220 $	3 000 F	4 000 F	7 000 F		4 000 F	3 000 F	8 500 F
STATISTIQUE DESCRIPTIVE								
moyenne, écart-type, etc...	■	■	■	■		■		■
distributions de fréquences	■	■	■	■				■
TESTS PARAMETRIQUES								
Khi-deux	■	■	■	■				■
t de Student	■	■	■	■				■
TESTS NON-PARAMETRIQUES								
Kolmogorov-Smirnov	■	■	■	■				
Wilcoxon	■	■	■	■				
Kruskal-Wallis	■	■		■				
ANALYSE DE LA VARIANCE								
ANOVA	■	■	■	■				■
ANCOVA				■				■
CORRELATION & REGRESSION								
Pearson R	■	■	■	■	■	■		■
Spearman	■	■	■	■				
Kendall	■	■	■	■				
régression	■	■	■	■	■	■		■
rég. polynomiale	■	■	■	■				■
pas à pas		■	■	■	■			■
rég. non-linéaire	■	■	■	■				
ANALYSE DES DONNEES								
analyse en composantes principales		■	■	■	■			■
analyse des correspondances					■			
analyse discriminante		■	■	■	■			
classification ascendante hiérarchique		■		■	■			
nuées dynamiques					■			
analyse canonique				■				
SERIES-CHRONOLOGIQUES								
lissages	■			■		■		
autocorrélation	■			■				
ARIMA				■				
transformées de Fourier				■				
GRAPHIQUES								
histogramme	■	■	■	■	■	■		■
diagramme en boîte (box plot)	■	■	■	■				■
diagramme en tronc et feuilles		■		■				
graphiques bivariés (x,y)	■	■	■	■	■	■	■	■
droite de régression	■	■	■	■		■		■
courbes de niveaux				■				
diagramme triangulaire				■				
graphiques trivariés (x,y,z)	■	■		■			■	■
histogramme en 3 dimensions	■		■	■				
surface en 3 dimensions		■		■				
cartes géographiques				■				
GESTION DES DONNEES								
valeurs manquantes	■	■	■	■			■	■
nombre maximum de variables	mémoire	mémoire	mémoire	200	mémoire	mémoire	mémoire	?
nombre maximum d'individus	mémoire	mémoire	mémoire	disque	disque	32000	mémoire	?
PREPARATION DES TABLEAUX								
sélection d'individus	■	■	■	■		■	■	■
calcul de nouvelles variables	■	■	■	■	■	■	■	■
pondération des observations				■	■			■
recodages	■	■	■	■	■	■	■	■
IMPRESSIONS GENERALES								
Convivialité	★★★	★★★★	★★★	★	★	★★★	★★★	★★★★
Qualité de la documentation	★★★	★★★★	★★★	★★	★★	★★★★	★★★★	★★★★
Logiciel en français	non	non	oui	non	oui	non	oui	non

N° 21
Février / Mars 90
5ème année - 30 FF

ICÔNES

Des souris et des hommes

Supplément
PAO

Dossier

Les statisticiels

La création de fontes

Pratique :
Illustrator
Excel
4D4

Design Studio
La micro-vidéo

L 1228 - 21 - 30,00 F
3791228030005 00210.
Belgique 180 FB - Suisse 8 FS - Canada 5,75 $
ROCCO