

PRÉSENTATION D'OUVRAGES

JD_STAT, un ensemble de programmes pour étudiants, médecins, épidémiologistes et biologistes (logiciel bilingue).

Jean Dutertre, MCSH (CR)

Logiciel présenté au congrès SPE de l'île Maurice, novembre 1996.

JD_STAT, a set of Programs for students, physicians, epidemiologists or biologists (a bilingual program).

* * The data entry program is a worksheet. Its " look and feel " allows you to use it without any documentation or training. Just rely on instinct. Your data will automatically be set in such a way that any program of this set should find all what it needs to perform the computation, without annoying you any further with bothersome queries : as long as Artificial Intelligence is concerned, you supply Intelligence, the machine supplies Artificial. The Worksheet is three dimensional :

- 1 - The Columns relate to sets of data,
- 2 - The Rows, to successive data in each set,
- 3 - The Pages, to replications, if any.

For the dose-response analysis there is one dimension more, on the same page :

- s - the Standard preparation,
- t - the Test preparation.

You will then be asked if your data are weighted or not. Usually, you get Columns A, B, C, D, ... and Rows 1, 2, 3, 4, 5, ... but if you said " weighted ", the Columns become Xi, Ni(A), Ni(B), Ni(C), Ni(D), ... and for a correlation template (with weighted data), you will get a first Row for the Xi, a first Column for the Yj, the Cells waiting for the size of the corresponding samples. For a multiple step wise correlation, the first Column, Y, is devoted to the independent variable, the next ones, X1, X2, X3, ... and so on, to the predictive variates. You will find out that other templates are available.

You may choose " labels ". To write data in a cell, get the cursor in that cell with the arrows and write the digits with the keyboard. To validate, just arrow out of that cell. To rectify a data, come back in the cell and write over. It's all you will have to do and it's easy. To rub out, come back in the cell and hit the Del key. You may get at the head or bottom of the template with the Home and End keys.

* * The programs provide classical analysis : multivariate data analysis, factorial data analysis, analysis of variance, comparison of two means, correlation & regression, factorial assay, chi square, dose-response (quantal) and dose-response (quantitative), discriminant analysis, multiple regression step by step, Survival tables. No statistical tables are used, the programs compute the significance levels from scratch using, to do so, the genuine distribution functions.

As already said above, the different programs of data analysis ask no questions, and compute as much statistics as the quality and number of the data allows them to do. All depends of the unique and initial choice of the user, as performed while the completion of the worksheet is done.

The language used in the documentation, programs and forms is french or english, at will.

jean.dutertre@wanadoo.fr

Matériel nécessaire

Ce logiciel fonctionne :

sur PC ;

sous DOS ;

avec 640 k de mémoire vive ;

les programmes et fichiers tiennent en 3 méga sur disquette.

L'interface semi-graphique (en français ou en anglais, au choix) se pilote à la souris (ou au clavier, au choix). On voit qu'un PC " élémentaire " tel qu'on en trouve au rebut suffit donc largement, toutefois - qui peut le plus peut le moins - un matériel plus moderne n'ajoute rien, mais ne nuit pas. Pour la disponibilité du programme, voir à la fin du texte.

Installation

(à sauter en première lecture)

Le logiciel se présente sous la forme de deux disquettes de trois pouces et demi étiquetées respectivement " JD_Stat Programmes " et " JD_Stat Exemples ". Préparez un sous-répertoire JD_STAT (par exemple, vous avez le choix) pour les programmes, placez-vous dans ce sous-répertoire et recopiez-y la totalité de la disquette " programmes " (copy a:*. * /v) et un sous-répertoire JD_DATA (par exemple, mais vous faites comme vous voulez), placez-vous dans ce sous-répertoire et recopiez-y la tota-

lité de la disquette " exemples " (copy a:*. * /v). Pour vos données personnelles, vous pouvez préparer un troisième sous répertoire JD_MIEN (par exemple...) Il est bon d'ajouter dans votre fichier CONFIG.SYS (si elles n'y sont pas déjà) les instructions suivantes :

FILES = 30

BUFFERS = 30

L'affichage du tableur ira beaucoup plus vite.

Le démarrage du logiciel se fait en lançant le programme de " menu " appelé JD_MENU.EXE. On se place donc dans le sous-répertoire JD_STAT, on frappe JD_Menu puis <Entrée> et c'est parti. Les astucieux sauront préparer (dans le répertoire racine) un fichier batch nommé J.BAT dans lequel figure la seule instruction :

C:\JD_STAT\JD_MENU.EXE

au démarrage de l'ordinateur, il suffira donc de taper <J> et <Entrée> au clavier, le reste suivra tout seul.

Pour conserver une copie papier des résultats affichés, il suffit de frapper la touche < impr. écran > ou < print screen >. Trois pilotes d'impression en assembleur, pour trois styles d'imprimantes, sont fournis. Il suffit d'installer celui qui convient dans le répertoire racine et de l'appeler dans le fichier AUTOEXEC.BAT. La procédure est légè-

ment plus compliquée sous Windows que sous Dos, comme toujours. Pour ceux qui préfèrent travailler avec l'interface Windows, donc, cette même touche envoie la copie dans le presse papier, que l'on peut secondairement imprimer.

Description du logiciel

Il y a deux parties essentielles dans le logiciel " JD_Stat " :

- La première est constituée par l'interface avec l'utilisateur. Elle est destinée à lui faciliter la tâche, sans l'embarrasser jamais de questions auxquelles il est souvent difficile de répondre, faute de vue d'ensemble préalable des techniques mises en oeuvre : lorsque l'on commence, en principe, on ne sait rien, ou peu de choses, des données elles-mêmes, ni de ce que l'on peut en attendre.

- La seconde concerne l'ensemble des programmes de calcul ou de présentation des statistiques classiques.

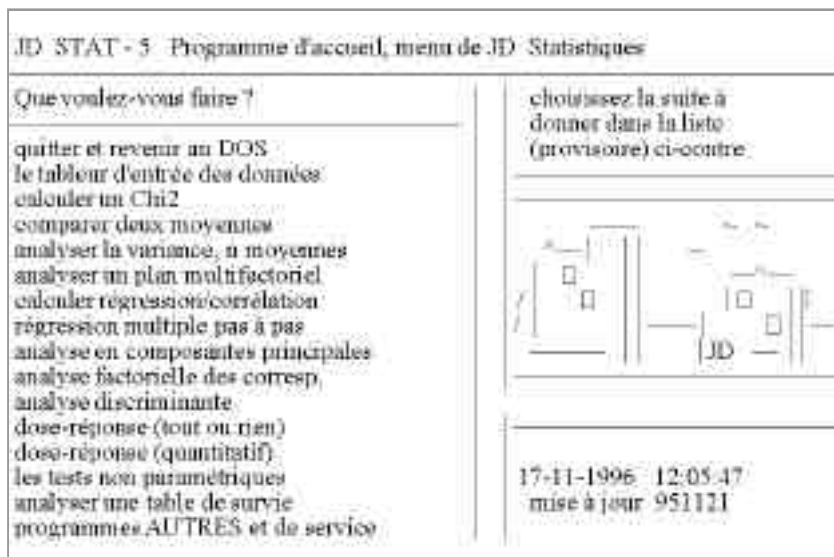
Le logiciel lui-même comprend la totalité de la documentation disponible sur écran, y compris tout ce qui concerne le matériel et l'installation mais... pour en prendre connaissance, encore faut-il l'avoir déjà installé... d'où les paragraphes qui précèdent. La suite concerne une présentation rapide du contenu, ainsi que quelques tableaux représentatifs de l'interface écran mise à la disposition de l'utilisateur.

L'interface avec l'utilisateur

L'interface avec l'utilisateur est semi-graphique, elle comprend un système de menu assez banal, qui permet d'exercer des choix, et enfin un tableau plus original. Le menu, le tableau, et l'ensemble des programmes, fonctionnent sous DOS. Il s'agit là d'une option délibérée, car un programme scientifique doit pouvoir faire appel à des routines en langage machine, et s'abstraire de toutes les complications et lourdeurs inévitablement induites par une interface graphique. En revanche, les avantages de l'interface graphique ont été conservés, et tout le système est commandé à l'aide de fenêtres dans lesquelles l'utilisateur exerce ses choix à l'aide des flèches du clavier, ou, plus généralement et plus aisément encore, à l'aide de la souris. Il en résulte que le logiciel est plus compact, plus rapide, se contente d'une mémoire vive minimale et d'un processeur éventuellement lent. JD_Stat par conséquent se trouve parfaitement fonctionnel et tout à fait à l'aise sur un ordinateur PC d'un modèle ancien, retrouvé sur une étagère, dans la cave, ou d'occasion. Pour n'effaroucher personne, notons qu'il est parfaitement possible de piloter ce même programme dans une fenêtre windows, une icône est à la disposition de ceux qui voudraient en faire usage dans leur gestionnaire de programmes : tant il est vrai que si l'on peut faire simple, rien n'interdit de faire compliqué.

Le menu (que l'on peut lancer à partir d'un fichier batch voir plus haut), n'appelle pas de commentaires. Il prend l'aspect de la copie d'écran (fig. 1) suivante :

Figure 1.



On déplace la souris dans la partie gauche du tableau (elle n'en peut sortir), surlignant ainsi la ligne voulue, on

clique, c'est fait. Le choix de la dernière ligne conduit au tableau complémentaire (fig. 2) :

Figure 2.

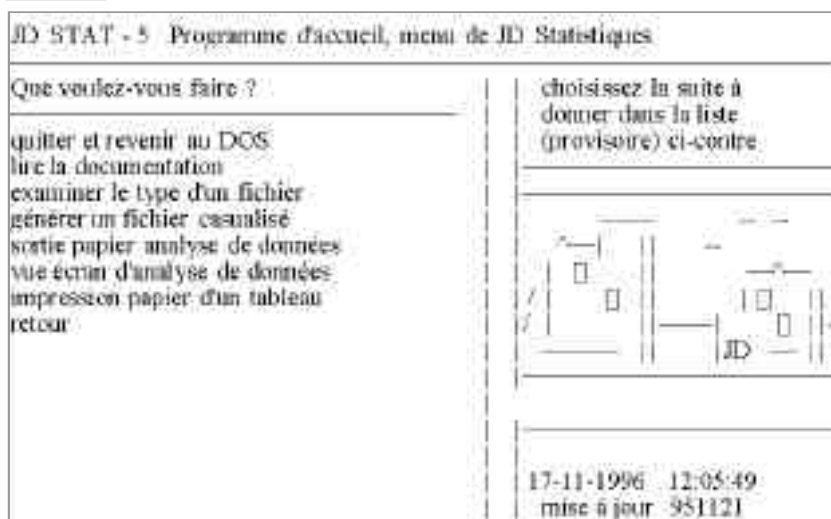


Figure 3.

JD_STAT - 1. Tableau d'entrée des données — Mouvements : (6 flèches)					
Label/Donn.	(Tab)	Valider (flèches)	Effacer (Suppr/Del)	Finir (Echap/Esc)	
	A	B	C	D	E
	de trait	de maud	depart	reclute	succes
1 atoxyl_1	16	56	235	495	2094
2 cesuin1	1	0	14	36	230
3 cesury1	2	1	10	23	139
4 lomidm1	0	1	20	16	95
5 arsoba1	1	0	7	0	54
6 cesuin2	14	14	75	107	413
7 cesury2	59	271	302	815	1326
8 lomity2	5	12	49	21	89
9 motury2	12	47	57	83	162
10 arsoba2	5	6	31	4	214
11
12

demande complète A 1 : 16

Bilan Trypano Group. Lobi

Le tableau est plus particulier. C'est, comme on le voit ci-dessous, un tableau d'aspect ordinaire, avec des chiffres en lignes et des lettres en colonnes, comme toujours, ainsi que des libellés de pieds de lignes et des libellés de têtes de colonnes, à ceci près que... lorsque l'on se déplace dans le (parfois vaste) corps du tableau, certaines cases disparaissent momentanément de l'écran, mais jamais les pieds de ligne, ni les têtes de colonne, ni les libellés, qui sont toujours visibles, ce qui est essentiel pour éviter les erreurs de saisie. Le corps du tableau reçoit simplement les données, il sert donc, essentiellement, à entrer les données.

La copie d'écran se présente (fig. 3) ainsi : Le tableau est à trois dimensions :

- 1 - Les Colonnes pour les " séries " de données à entrer,
- 2 - Les Lignes pour les " données " successives,
- 3 - Les Pages pour les " répétitions " s'il y en a.

Pour l'analyse à droites parallèles de la relation dose-réponse il y a même une dimension de plus, sur la même Page : r - la préparation " de référence ", t - la préparation " à titrer " .

Les données peuvent être pondérées, ou non. Si c'est non pondérées, le tableur se présente normalement ainsi (fig. 4) :

	A	B	C	D	E	F	G	...
1	47	7	90	83	10	
2	38	11	67	80	14	
3	2617	389	2117	1656	406	
...	

mais si les données sont pondérées, les colonnes deviennent automatiquement Xi, Ni(A), Ni(B), Ni(C) ... (fig. 5) :

	Xi	Ni(A)	Ni(B)	Ni(C)	...
1	2100	2	1
2	2300	4	1
3	2500	6	3
...

En matière de table de corrélation on aura une première ligne pour les Xi, une première colonne pour les Yj, les cases recevant alors les effectifs correspondants... (fig. 6) :

	yj	x1	x2	x3
xi		16.5	18.5	20.5
y4	2700	1	5	2
y5	2900	-	-	3
y6	3100	-	1	1
...

et dans le cas de la corrélation multiple pas à pas, la première colonne Y concerne la variable à expliquer, les suivantes X1, X2, X3 etc. concernent les variables explicatives (fig. 7) :

	y	x1	x2	x3	...
1	78.5	7	26	6	
2	74.5	1	29	15	
3	104.3	11	56	8	
...	

mais jamais les programmes de traitement des données ne posent de questions indiscrettes sur ces détails. En effet, dès que les données sont présentées ainsi (on a passé rapidement sur la question, car il existe encore d'autres possibilités pour l'analyse dose-réponse, les tables de survie etc.) les paramètres de lecture des tables et de calcul sont fixés et l'utilisateur n'a plus à se soucier de rien, que de l'interprétation du résultat. Redisons-le, ce n'est pas lors du traitement des données, qui se fait toujours automatiquement sans aucune intervention de l'utilisateur, mais au moment de leur introduction dans le tableur que les caractéristiques des données, et le choix du traitement qui devra leur être appliqué, sont fixés par l'utilisateur lui-même. En effet, au bout de quelques

semaines ou mois il peut être difficile de se souvenir de l'origine ou du caractère particulier de certains fichiers. Au moment même où l'on le constitue, on le sait parfaitement bien. On a donc tenu compte ici de cette expérience.

Mode d'emploi du tableur : Pour écrire un chiffre dans une case, amener le curseur dans cette case avec les flèches et écrire le chiffre avec le clavier. Pour valider, un mouvement de flèche vers la case suivante y pourroit simplement. Bien sûr, on peut tout corriger, ajouter ou supprimer lignes et colonnes. Tout est instinctif, il n'existe, pour utiliser ce logiciel, aucune période d'apprentissage.

Ce tableur calcule les statistiques des variables qu'il contient. Vous pouvez choisir ainsi de voir la moyenne, avec la plus petite, la plus grande valeur et l'étendue, les sommes, sommes de carrés, moyenne, variance, écart type, symétrie et courbure avec leur test t de signification contre zéro. Ci-dessous, une copie d'écran (fig. 8) :

Figure 8.

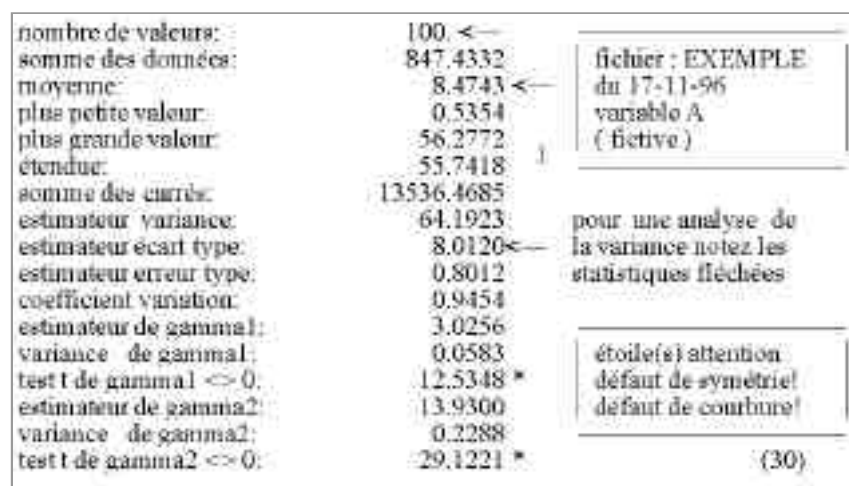
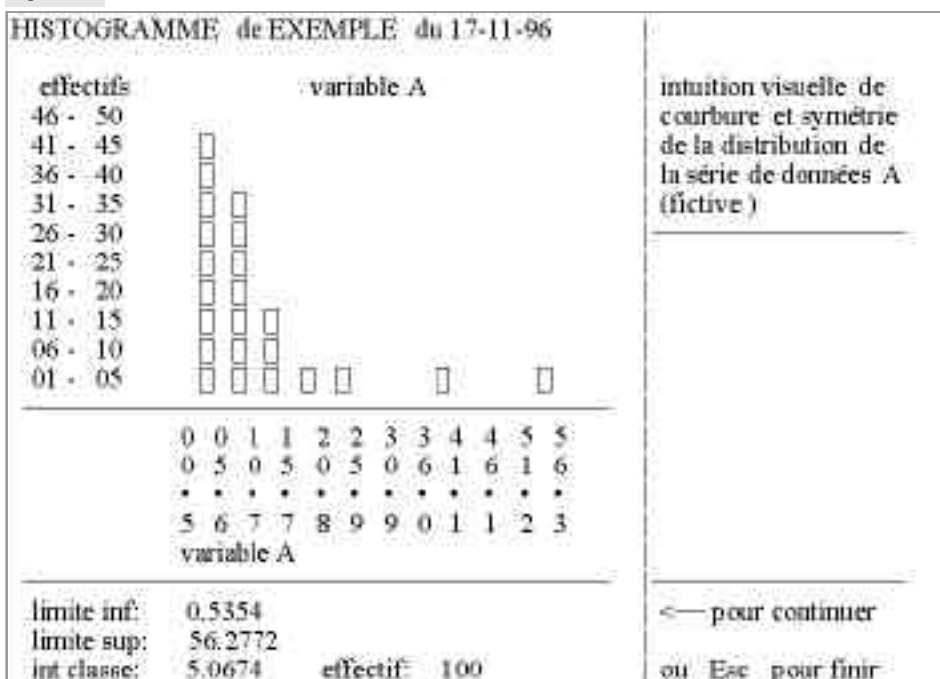


Figure 9.



Puis le tableur présente des histogrammes, sans demander ni maximum, ni minimum, ni nombre de classes, ni intervalle de classe, ni quoi que ce soit. Si la distribution est asymétrique (il est important de le savoir), essayez une transformation log (automatiquement), cela peut sauver bien des situations difficiles. Ci-contre, (fig. 9) une copie d'écran.

Les programmes de calcul

Les méthodes de calcul sont puisées dans quelques livres classiques. On citera en tête les « *Méthodes statistiques à l'usage des médecins et des biologistes* » (Flammarion, Paris, première édition 1963, quatrième édition 1993) de Daniel SCHWARTZ, l'ouvrage de statistique médicale le plus pédagogique jamais écrit en aucune langue. De nombreux exemples sont choisis dans ce livre, ou dans « *L'Essai thérapeutique chez l'homme* » du même auteur. Certaines méthodes et quelques exemples sont pris aussi dans « *Statistical Methods* » de SNEDECOR et COCHRAN, dans « *Statistical Methods in Medical Research* » d'ARMITAGE, d'un abord certes plus austère. On a eu largement recours aussi à « *Théorie et méthodes statistiques* » et « *Analyse statistique à plusieurs variables* » de Pierre DAGNÉLIE (Duculot S.A., Gembloux), ainsi qu'à « *Méthodes en épidémiologie* » de C. RUMEAU-ROUQUETTE, G. BRÉART et R. PADIEU (Flammarion, Paris). Pour tout ce qui concerne l'analyse des données, on s'est référé d'abord à l'« *Initiation à l'analyse des données* » (Dunod, Paris) de Jean de LAGARDE, excellente introduction dans le domaine, où à mon « *Analyse des données sur HP-41* » (éd. du Cagire, 1984 Toulouse), puis à la « *Pratique de l'analyse des données* » de Jean Pierre BENZÉCRI, et à « *L'Analyse des données en médecine, pharmacologie et physiologie clinique* » de J-P & F. BENZÉCRI. Pour la relation dose-effet, l'ouvrage de référence est la « *Statistical Method in Biological Assay* », de D.J. FINNEY (Griffin, London),

ainsi que l'annexe statistique de la « *Pharmacopée Européenne* ». Les techniques de calcul des fonctions de distribution sont dues au « *Handbook of mathematical Function* » d'ABRAMOWITZ et STEGUN, National Bureau of Standards, 1968, le calcul matriciel des vecteurs propres et des valeurs propres fait appel à l'algorithme n° 13 des « *Numerical methods for small computers* » de J.C. NASH (Adam Hilger Ltd, Bristol).

Vient ensuite normalement l'interprétation des résultats, en termes de signification. On a renoncé à l'usage des tables. La plupart du temps, celles-ci sont l'objet d'un copyright qui interdit leur reproduction sans autorisation. En fait, avec un ordinateur, rien n'empêche

de recalculer les fonctions de distribution, et c'est le parti qui a été choisi. De ce fait, les résultats sont présentés toujours ainsi (par exemple) :

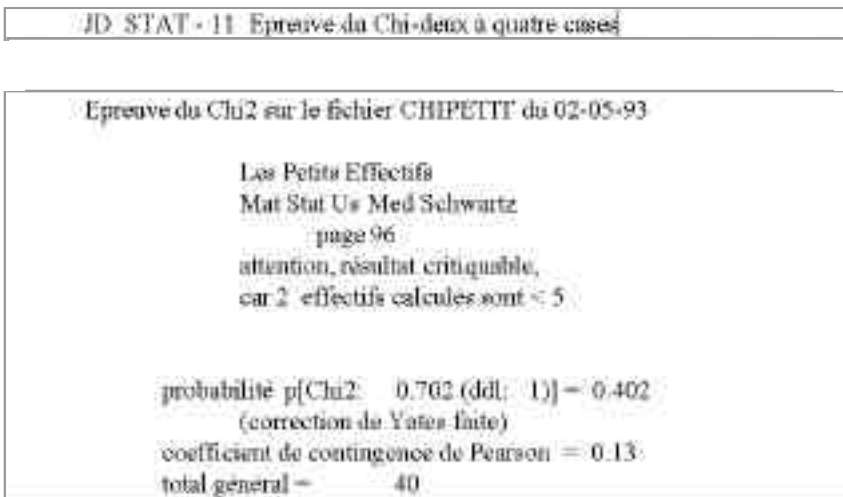
$$p\{t : 4,583 \text{ (ddl : 6)}\} = 0,004$$

La probabilité de l'occurrence (du fait du hasard) des observations est donc calculée exactement, et ne fait pas référence à un intervalle de probabilité, comme lors de l'usage d'une table. L'utilisateur reportera simplement la ligne ci-dessus à l'appui de ses dires dans sa communication, ou même la résumera plus encore ainsi: ($p = .004$).

En l'état actuel, la liste des programmes est la suivante (fig. 10) :

Liste des programmes de calcul : (* pour les projets)		
JD_STAT-1	JD_Calc	Tableur d'entrée
JD_STAT-2	JD_CoPri	Analyse de données multivariates
JD_STAT-3	JD_FaCor	Analyse de données à double entrée
JD_STAT-4	JD_AnVar	Analyse de la variance
JD_STAT-5	JD_Menu	Programme d'accueil
JD_STAT-6	JD_Moyen	Comparaison de moyennes
JD_STAT-7	JD_Corrl	Corrélation et Régression
JD_STAT-8	JD_Exame	Coup d'oeil sur un fichier
JD_STAT-9	JD_Gener	Génère un fichier exemple
JD_STAT-10	JD_PlanF	Plan factoriel
JD_STAT-11	JD_TChi2	Chi2 ou table de contingence
JD_STAT-12	JD_Doslo	Dose-réponse (succès sur total)
JD_STAT-13	JD_Dosav	Dose-réponse (quantitative)
JD_STAT-14	JD_NoPar *	Tests non paramétriques
JD_STAT-15	JD_Grafx	Graphique imprimé pour 2 & 3
JD_STAT-16	JD_Tablo	Impression des données tabulées
JD_STAT-17	JD_Dscri *	Analyse discriminante
JD_STAT-18	JD_Docum	Lecture de la documentation
JD_STAT-19	JD_Ecran	Graphique sur écran pour 2, 3 et 20
JD_STAT-20	JD_RegMu	Régression multiple pas à pas
JD_STAT-21	JD_Survi	Tables de survie

Figure 11.



Le programme demande ensuite s'il convient, oui ou non, de calculer l'épreuve "exacte" de FISHER. En cas de réponse affirmative, il vient (fig. 12) :

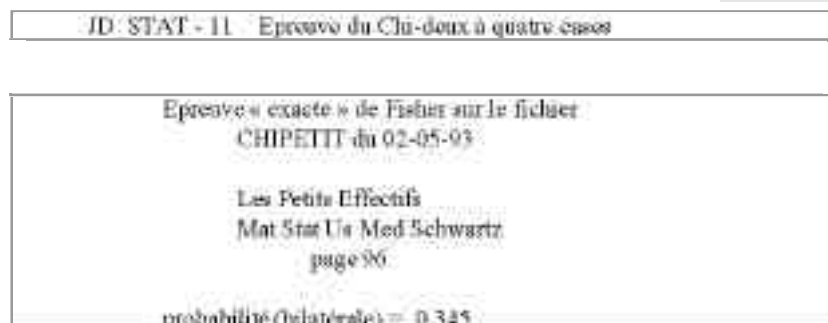


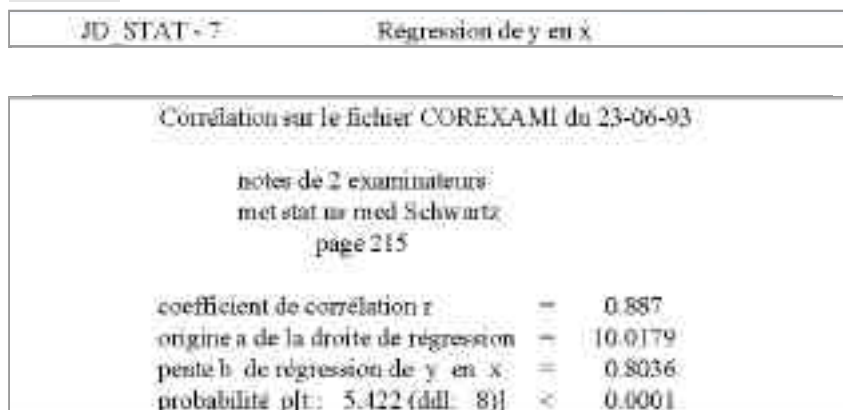
Figure 12.

Quelques exemples

Ce logiciel permet de pratiquer des essais assez élaborés des données, des analyses de données multivariates à la Benzécri, des régressions multiples pas à pas, des comparaisons d'activité de substances biologique ou de vaccins à droites parallèles... Nous nous contenterons, pour aller vite, d'exemples simples et familiers à tous. Prenons donc un "classique" chi-deux, pour commencer : On voit que la présentation contient l'essentiel, y compris un avertissement concernant les effectifs calculés. La correction de Yates est appliquée dès lors que les conditions sont réunies. Le titre, sous titre, référence, date des données sont rappelées systématiquement. Voir ci-contre (fig. 11) les copies d'écran :

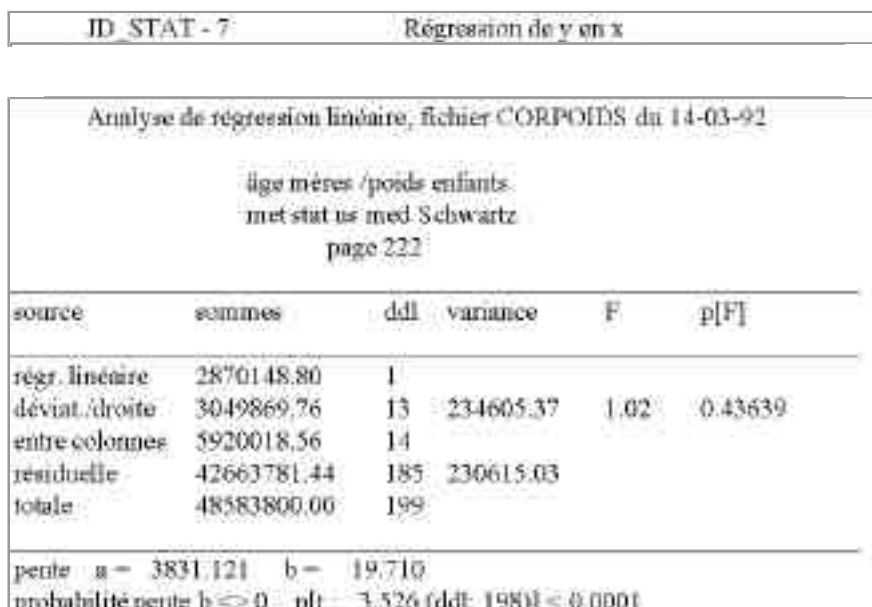
Une régression ou corrélation, maintenant, d'abord dans un cas simple (fig. 13) :

Figure 13.



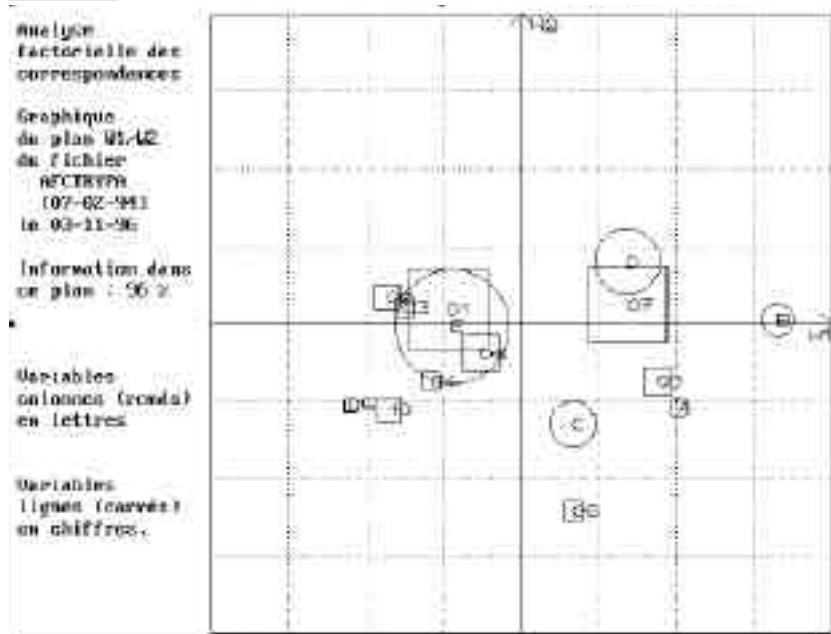
Le programme a calculé le plus de statistiques possibles, en fonction des données disponibles. Ici, il était encore faisable de calculer les éléments de la droite de régression, cela a donc été fait. L'exemple suivant permet d'en voir plus (fig. 14) :

Figure 14.



Cette fois-ci, il y avait assez de données pour calculer une analyse de la variance. Le programme l'a constaté, c'est donc ce qui a été fait sans que l'utilisateur ait eu à intervenir. Le programme fait systématiquement " au mieux ".

Figure 15.



Graphiques

Avant de terminer, notons que les analyses de données (en composantes principales ou factorielles des correspondances), la régression multiple aussi, aboutissent à des présentations graphiques. C'est ainsi que les données de la fig. 3, se traduisent par la copie d'écran suivante (fig. 15) :

Bien entendu, il est possible de sortir une copie papier de ce graphique, toujours plus fine qu'une copie d'écran, en vue d'une publication.

On pourrait allonger encore cette présentation, sans en dire beaucoup plus, que ceci : tout les calculs se font automatiquement, en fonction des données disponibles et du choix unique, initial, de l'utilisateur. Rappelons que le présent texte représente un simple survol et que toute la documentation détaillée est disponible en ligne, en français et en anglais ; de même tout le logiciel, formulaires et tableaux, se présente au choix dans ces deux langues.

Disponibilité

JD_STAT est gratuit et se trouve auprès de l'auteur, adresse e-mail : jean.dutertre@wanadoo.fr

Il est actuellement proposé en téléchargement à l'adresse suivante : <http://perso.wanadoo.fr/jdtr/>