

REVUE DE STATISTIQUE APPLIQUÉE

G. CARAUX

Exploration interactive d'une correspondance

Revue de statistique appliquée, tome 37, n° 1 (1989), p. 53-65

http://www.numdam.org/item?id=RSA_1989__37_1_53_0

© Société française de statistique, 1989, tous droits réservés.

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

NUMDAM

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

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

EXPLORATION INTERACTIVE D'UNE CORRESPONDANCE*

G. CARAUX**

RÉSUMÉ

D'un moyen de calcul puissant, l'ordinateur est devenu, pour beaucoup, un outil simulant un environnement de travail familier. La micro-informatique, en généralisant les écrans aux possibilités graphiques étendues et en introduisant des outils de pointage simples, a contribué fortement à accentuer ce phénomène. Les logiciels émanant de cette nouvelle génération d'outils informatiques ont tiré partie d'une nouvelle manière de concevoir l'interactivité et la convivialité des rapports homme-machine. Ces transformations posent en d'autres termes notre manière d'exploiter l'informatique dans l'analyse des faits statistiques. Partant d'une approche essentiellement calculatoire, dans laquelle un programme fournit la solution algorithmique d'un modèle prédéfini, ne faut-il pas imaginer un environnement de travail interactif propice à l'exploration des faits statistiques? L'utilisateur y explorera ses données par un certain nombre de jeux scéniques dans lesquels sa perception intuitive du scénario proposé lui permettra de rester maître de la logique de son application.

A titre d'exemple nous aborderons l'exploration graphique et interactive d'une correspondance statistique entre deux variables qualitatives.

Mots-clés : Graphique, Interactivité, Analyse exploratoire, Tableau de contingence.

1. Introduction

En janvier 1987 notre unité de recherches s'est dotée d'une station de travail graphique. Celle-ci, organisée autour d'un micro-ordinateur Goupil G40 compatible IBM PC-AT, dispose d'un écran graphique couleur haute résolution de 1024×768 pixels muni de sa carte d'interfaçage propre. Cette dernière dispose d'un processeur d'affichage graphique spécifique et d'une mémoire bitmap de 1 Moctet.

L'exposé proposé ici va essayer de présenter les idées qui incitent le travail développé sur ce matériel et les premiers résultats concrets qui ont déjà été obtenus. Nous examinerons dans un premier temps les perspectives qu'offre l'exploitation d'outils graphiques interactifs dans l'exploration de données statistiques.

Nous illustrerons ensuite notre propos par un exemple original de graphisme qui nous permettra d'explorer la dépendance statistique de deux variables qualitatives à travers un tableau de contingence.

* Ce texte reprend les termes d'un exposé présenté en juin 1987 à l'assemblée Générale du département de Biométrie de l'INRA.

** Unité de Biométrie, Ecole Nationale Supérieure Agronomique, Place P.Viala, 34060 Montpellier Cédex

2. Pour une approche interactive de l'exploration de données statistiques

Les outils de l'informatique disponibles dans les années 60 et 70 ont permis, dans notre discipline, de programmer et de résoudre numériquement les modèles mathématiques que nous écrivions. Bien adaptée à l'exploitation de données numériques, l'informatique d'alors a permis aux statisticiens de trouver les moyens de calcul performants dont ils avaient besoin. Ceci a fortement contribué au développement et à la vulgarisation de nombreuses méthodes, tant descriptives qu'inférentielles, devenus maintenant d'usage courant.

La contre-partie de cette vulgarisation est que les programmes, comme les logiciels, sont devenus des intermédiaires abstraits entre des données brutes et des résultats statistiques parfois très élaborés. Les programmes ont pris l'aspect de "boîtes noires" dont la logique interne n'est perceptible qu'aux spécialistes.

Or dans le même temps, l'accessibilité aux bibliothèques de logiciels s'est étendue avec la généralisation des moyens d'accès à l'informatique. La facilité d'usage de ces bibliothèques de programmes leur a même donné l'aspect rassurant de produits d'usage courant, sans contrainte d'utilisation.

Dans ce contexte les résultats d'un processus expérimental ne sont plus le résultat d'une rencontre entre la capacité d'observation de l'expérimentateur et sa propre perspicacité intellectuelle dans l'analyse de faits expérimentaux.

L'expérimentateur s'est reposé, parfois d'une façon aveugle, sur la crédibilité donnée aux méthodes statistiques et à l'utilisation de l'ordinateur. Il a pu s'ensuire parfois une démobilisation intellectuelle importante du biologiste à une étape pourtant essentielle du processus expérimental.

Or n'oublions pas que nos modèles sont construits dans le cadre d'hypothèse simplificatrices qui sont le reflet d'une situation supposée idéale. L'optimalité des solutions que nous proposons est le fruit d'un processus simplificateur.

Trop souvent le biologiste prend le fruit de nos algorithmes comme des oracles et non pas comme le reflet d'une réalité complexe, aléatoire et contradictoire qui doit alimenter et éclairer ses réflexions personnelles.

Nos logiciels ont créé une génération de consommateurs ou plutôt de spectateurs. Il faut maintenant essayer de leur rendre un rôle d'acteur.

Les moyens, tant matériels que conceptuels, que met à notre disposition l'informatique ont changé radicalement. Ceci nous permet de poser en d'autres termes les problèmes issus de l'exploration ou de l'exploitation de données statistiques.

Nombre de nos modèles ont été conçus et développés pour des machines dont l'interactivité était nulle. Le dialogue était unilatéral et s'opérait sous la forme de commande dactylographiée sur un clavier alphanumérique. Les résultats fournis étaient publiés sous une forme séquentielle sur un support le plus souvent en papier.

De même les langages employés alors, s'ils sont encore bien adaptés pour la description d'algorithmes numériques, ne permettent que la manipulation d'objets mathématiques rudimentaires (scalaires, matrices indicées,...). On ne peut les utiliser

pour décrire des objets conceptuels évoluées ou pour développer l'environnement d'accueil de l'utilisateur final.

En micro-informatique les logiciels de base ont réconcilié les non-informaticiens avec l'usage de l'informatique.

Les outils du dialogue homme-machine ont été révolutionnés par la généralisation d'écrans aux possibilités graphiques étendues et par des systèmes interactifs de pointage comme la "souris" ou le "photostyle".

Plus qu'une machine à calculer ou à ordonner, l'ordinateur est apparu comme un environnement de travail simulant des tâches et des procédures familières à l'utilisateur (KAY [15]).

Les logiciels apparaissent alors comme des outils de facilitation et non d'exécution. L'utilisateur de par sa perception intuitive de l'environnement où il évolue, reste maître de son application et de la logique du traitement qu'il applique.

Dans ce contexte une partie de notre travail de statisticien est de concevoir l'univers théâtral dans lequel l'utilisateur pourra se déplacer pour explorer ses données par différents jeux scéniques.

Nous ne lui fournirons pas, dans cette optique, une solution mécaniste à son problème, mais les moyens de regarder dynamiquement ses résultats en monopolisant sa perception des phénomènes.

Un travail de conceptualisation d'objets statistiques nouveaux est à entreprendre et une partie de la statistique descriptive est à réinvestir.

Les outils d'affichage graphique dont nous disposons maintenant permettent d'utiliser des graphismes simples au pouvoir évocateur puissant :

- L'intensité d'un phénomène est représentée d'une façon naturelle par une certaine graduation d'intensité de lumière.
- Un phénomène chronologique peut être suggéré par le déplacement d'un spot dans un graphique approprié,
- etc....

Ces artifices, qui s'apparentent à un codage, sont autant de métaphores qui, étant prélevées dans l'univers habituel de l'utilisateur, lui permettent de conserver son esprit d'initiative, de libérer son imagination et de conserver bon sens et perspicacité.

L'interactivité peut s'appliquer à des méthodes connues et banalisées. On connaît par exemple l'intérêt que présente la faculté d'observer visuellement et instantanément les répercussions de la modification de l'ordonnée d'un point sur la pente de la droite de régression. Cette faculté enrichit beaucoup, voire supplante chez l'utilisateur de la statistique, le concept de variance attachée à l'estimateur de la pente d'une droite de régression.

Dans ce domaine des travaux sont menés aux Etats-Unis, depuis le début des années 70, dans le prolongement des idées de TUKEY [20]. L'importante bibliographie publiée récemment par CLEVELAND [8] est là pour en témoigner.

A titre d'exemple, un travail important sur la recherche d'un langage de commandes interactives propres à l'exploration des faits statistiques est réalisé depuis plusieurs

années au Département de Statistique de l'Université de Harvard (DONOHO, HUBER *et alt.* [9]).

Parallèlement, un prototype de station de travail dédié à l'exploration de données statistiques a été élaboré sous le nom d'Orion I, autour d'un ordinateur Prime, au Département de Statistique de l'Université de Stanford (MCDONALD [17]).

On peut noter que l'interactivité dans les procédures d'analyse de données exploratoires est privilégiée. Les outils informatiques de l'interface homme-machine, comme l'organisation multifenêtre d'un écran graphique, sont de plus en plus employés (Cf : *Plot Windows*, STUETZLE [18] et [19]).

Plus récemment, BECKER et CLEVELAND [4] ont montré l'intérêt de l'utilisation d'une "souris" dans l'exploitation interactive d'un "scatterplot". La pertinence des résultats trouvés est probante.

Enfin notons que certains logiciels proposent maintenant des outils d'exploration interactive de données statistiques. Citons par exemple :

- MacSpin un logiciel dédié à l'exploration de données multidimensionnelles ([2], [5], [10], [14]) et

Profession :	Sources d'information								Total
	Télé	Journal	Radio	Livre	Ami	Mairie	Association	Plus d'une source	
Agriculteur	26	18	9	8	5	6	4	38	114
Industriel	4	5	0	3	1	2	0	4	19
Artisan	36	28	7	5	5	2	3	13	99
Cadre Supérieur	15	44	4	13	5	1	5	21	108
Profession libérale	1	7	1	1	3	0	1	4	18
Cadre moyen	44	87	4	39	13	3	14	36	240
Employé	83	87	13	24	19	1	5	43	275
Ouvrier	181	107	16	31	41	7	7	70	460
Etudiant	14	43	1	17	15	0	5	21	116
Femme au foyer	158	87	12	33	20	5	17	72	404
Retraité	167	95	29	15	15	7	7	89	424
Sans profession	27	9	4	2	0	2	2	5	51
Total	756	617	100	191	142	36	70	416	2328

FIGURE 1
Sources d'information des catégories socio-professionnelles sur les problèmes liés à l'environnement et au cadre de vie

- S environnement capable d'accueillir des programmes interactifs (BECKER et CHAMBERS [3]).
- XploRe un outil d'exploration d'un modèle de régression et d'estimation de densités (AERTS et HOLTSBERG [1]).
- ...

Comme nous le voyons le champ d'investigation est ici très ouvert et prometteur.

Nous allons maintenant illustrer ces idées générales par l'exploration interactive d'un tableau de contingence.

3. Exploration d'une correspondance statistique

3.1. Notations

Soit $N = (n_{ij})_{IJ}$ un tableau de contingence, d'élément générique n_{ij} , croisant les modalités de deux variables qualitatives.

I et J seront respectivement l'ensemble des modalités de ces deux variables : $I = \{1, \dots, n\}$, $J = \{1, \dots, p\}$.

Communément on associe à N la loi de fréquence f_{IJ} sur l'ensemble des paires (i, j) de $I \times J$:

$$f_{ij} = \frac{n_{ij}}{n_{..}} \quad \forall (i, j) \in I \times J$$

On note alors $f_{i.}$ et $f_{.j}$ les deux lois de fréquences marginales définies respectivement sur I et J :

$$f_{i.} = \sum_{j \in J} f_{ij} \quad \text{avec } i \in I$$

$$f_{.j} = \sum_{i \in I} f_{ij} \quad \text{avec } j \in J$$

Enfin nous noterons plus loin par p_{IJ} la loi de fréquences conjointes respectant la propriété d'indépendance des lois marginales :

$$p_{IJ} = f_{i.} \cdot f_{.j}$$

$$p_{ij} = f_{i.} \cdot f_{.j} \quad \forall (i, j) \in I \times J$$

3.2. La représentation graphique proposée

Nous avons déjà (CARAUX [7]), exploré les problèmes liés à la représentation visuelle d'un tableau de données numériques. Ces travaux s'inspiraient alors des graphismes proposés par BERTIN [6] dans ses recherches en sémiologie graphique. Certains auteurs ont proposé des constructions graphiques appropriées à l'exploration d'un tableau de contingence. Par exemple GENEST et GREEN [11] ont repris récemment

l'idée de construction d'une "mosaïque" proposée par HARTIGAN et KLEINER [12] [13].

Le graphisme que nous proposons ici est original. Comme nous allons le voir les bases théoriques de sa construction sont élémentaires et tiennent bien compte de la nature du tableau de contingence N .

Dessignons un carré de surface unité. Subdivisons deux de ses côtés orthogonaux en respectivement n et p segments de longueur f_i ($i \in I$) et f_j ($j \in J$). Ces subdivisions induisent alors un maillage croisé de la surface du carré (figure 2).

Par construction, chacune des mailles obtenues a pour surface $p_{ij} = f_i f_j$, avec $(i,j) \in I \times J$.

Ainsi nous obtenons un découpage du carré de surface unité en éléments qui dessinent et respectent la loi des fréquences croisées sous l'hypothèse d'indépendance.

Dans le même temps cette figure nous offre une représentation de la loi des fréquences marginales. En effet la surface de chaque ligne (respectivement de chaque colonne) est égale à la fréquence de la marge qui lui est associée.

Dans notre construction nous retrouvons la propriété fondamentale :

$$\sum_{i \in I} f_i = \sum_{j \in J} f_j = \sum_{i \in I} \sum_{j \in J} f_i f_j = 1$$

Exemple

Pour illustrer notre propos, nous avons repris un tableau de contingence issu d'une enquête conduite par le CREDOC sur la qualité de la vie des Français (figure 1, page 56).

Ce tableau, déjà traité par LEBART et MORINEAU [16], croise deux variables qualitatives : La première code en douze modalités la catégorie socio-professionnelle. La seconde distingue suivant huit modalités les sources d'information des particuliers sur les problèmes touchant à l'environnement et au cadre de vie.

On note très facilement, à la figure 2 (page 59), que la télévision et les journaux constituent à eux seuls plus de la moitié des sources d'information des personnes interrogées.

On remarque également sur cette figure la marginalité des industriels et des professions libérales dans l'échantillon étudié (cf lignes 2 et 5 de la figure 2).

FIGURE 2
Représentation graphique de p_{ij}

3.3. Etude de l'écart à l'hypothèse d'indépendance

Nous venons de voir que la représentation maillée que nous proposons est représentative de la distribution des fréquences sous l'hypothèse d'indépendance des deux variables qualitatives mises en ligne et en colonne.

Le plus souvent, l'objet de l'analyse d'un tel tableau est de révéler la nature et l'intensité de l'écart de la loi des fréquences observées f_{ij} , à l'hypothèse d'indépendance :

$$f_{ij} \stackrel{?}{=} f_{i.} f_{.j}$$

Les écarts des fréquences f_{ij} à l'hypothèse d'indépendance peuvent s'exprimer par :

$$\begin{aligned} e_{ij} &= f_{ij} - p_{ij} && \text{avec } p_{ij} = f_{i.} f_{.j} \\ &= p_{ij} \left(\frac{f_{ij}}{p_{ij}} - 1 \right) \end{aligned}$$

En posant que $c_{ij} = \left(\frac{f_{ij}}{p_{ij}} - 1 \right)$ nous trouvons que :

$$e_{ij} = p_{ij} c_{ij}$$

Dans cette dernière équation, p_{ij} est la surface que nous avons affectée à la modalité (i,j) de $I \times J$ dans notre représentation graphique. On peut donc considérer ici que le coefficient c_{ij} est l'intensité de e_{ij} par unité de surface.

Ainsi, si sur l'écran graphique d'un ordinateur nous affectons une luminescence proportionnelle à c_{ij} à tous les "pixels" élémentaires de la maille associée à (i,j) , nous aurons un bon rendu visuel de e_{ij} . En effet la quantité de lumière émise par la maille de surface p_{ij} sera proportionnelle à la valeur de e_{ij} .

Notons ici que si nous devons avoir $f_{IJ} = p_{IJ} = f_I f_J$, alors l'ensemble des cases du tableau seraient noires. Dans ce cas l'intensité de lumière serait uniformément nulle car dans ce cas :

$$c_{ij} = 0 \quad \forall (i,j) \in I \times J.$$

Inversement notons que si et seulement si certains f_{ij} s'écartent de p_{ij} nous aurons alors des cases éclairées.

Remarquons que le rendu visuel des e_{ij} est pondéré par les valeurs de p_{IJ} . En effet, les écarts sont d'autant plus perceptibles que la surface de la maille est grande. Ainsi les valeurs p_{ij} sont des éléments de pondération.

Notons que e_{ij} peut prendre des valeurs négatives. Pour tenir compte de cette éventualité, nous avons programmé une représentation graphique pouvant avoir deux états :

- Dans le premier état seules les cases associées à une valeur positive de e_{ij} sont éclairées. Les autres restent éteintes.

Ici plus une case est éclairée plus est mise en évidence une association positive (attraction) entre les modalités de la ligne et de la colonne correspondante.

- Inversement, dans le deuxième état, seules sont éclairées les mailles associées à une valeur de e_{ij} . Une case fortement éclairée indique une association négative (répulsion) entre la ligne et la colonne correspondante.

Nous allons voir maintenant que la construction graphique que nous venons d'exhiber est proche d'une représentation visuelle décomposant la valeur du critère dit du χ^2 d'indépendance.

En effet le χ^2 associé au tableau N s'écrit :

$$\begin{aligned} \chi^2 &= \sum_{i \in I} \sum_{j \in J} \frac{(n_{ij} - np_{ij})^2}{np_{ij}} \\ &= n \sum_{i \in I} \sum_{j \in J} p_{ij} \left(\frac{f_{ij}}{p_{ij}} - 1 \right)^2 \\ &= n \sum_{i \in I} \sum_{j \in J} p_{ij} c_{ij}^2 \end{aligned}$$

En reportant dans notre représentation c_{ij}^2 sous forme d'intensité lumineuse, nous faisons apparaître les écarts entre f_{ij} et p_{ij} qui contribuent le plus à la détermination de la valeur du χ^2 .

Notons que pour une même valeur de c_{ij}^2 l'œil sera plus sensible aux valeurs dessinées dans des mailles de tailles importantes.

Signalons également que sur le plan graphique c_{ij}^2 donne un rendu visuel voisin de $|c_{ij}|$. Seul le niveau du contraste de la représentation visuelle est modifié. L'œil perçoit avec c_{ij}^2 une représentation équivalente mais plus contrastée.

Toujours en poursuivant notre logique nous pouvons dessiner les lignes du tableau par des bandes, non plus de hauteurs variables, mais de hauteurs égales. Nous avons alors une construction propice à la comparaison des profils (distribution de fréquences conditionnelles).

En effet si nous prenons la distance du χ^2 comme critère de comparaison entre profils, on peut écrire :

$$\begin{aligned} d_{\chi^2}(i, i')^2 &= \sum_{j \in J} \frac{1}{f_{.j}} \left(\frac{f_{ij}}{f_i} - \frac{f_{i'j}}{f_{i'}} \right)^2 \\ &= \sum_{j \in J} f_{.j} \left(\frac{f_{ij}}{f_i \cdot f_{.j}} - \frac{f_{i'j}}{f_{i'} \cdot f_{.j}} \right)^2 \\ &= \sum_{j \in J} f_{.j} \left(\frac{f_{ij}}{p_{ij}} - \frac{f_{i'j}}{p_{i'j}} \right)^2 \\ &= \sum_{j \in J} f_{.j} (c_{ij} - c_{i'j})^2 \end{aligned}$$

Là encore le coefficient c_{ij} joue le rôle. Il permet de comparer deux profils lignes entre eux. En effet en affectant de nouveau à la maille associée à (i, j) une luminescence proportionnelle à c_{ij} , on perçoit alors visuellement les différences pondérées de lumière comme critère de la distance distributionnelle entre deux profils. Le système de pondération, ici, n'est plus p_{ij} mais $f_{.j}$, la loi des fréquences marginales des colonnes. En effet, la surface des mailles associées à (i, j) et à (i', j) est proportionnelle à $f_{.j}$. Par construction cette perception visuelle respecte la nature de la distance du χ^2 .

3.4. Exploration interactive

Les graphismes que nous venons de présenter, s'ils portent en eux les reflets d'une réalité que nous cherchons à appréhender, sont cependant des images statiques.

Les transformations que l'on peut imprimer dynamiquement à ces graphiques, sont autant de moyens de révéler différents aspects de leur structure.

Pour disposer des fonctionnalités permettant de réaliser ces animations, nous avons programmé plusieurs commandes spécifiques. L'usage de la "souris" pour déclencher et désigner ces commandes a été choisi. Ceci a permis d'atteindre un bon niveau de convivialité dans l'interaction homme-machine. Les commandes de base actuellement disponibles sont :

- Changement de couleur d'une ligne ou d'une colonne,
- Effacement d'une ligne ou d'une colonne,
- Déplacement d'une ligne ou d'une colonne,

- Modification du contraste des couleurs,
- Zoom,
- Rotation d'un nuage de points.

Nous avons, par ailleurs, construit un ensemble d'écrans regroupant chacun plusieurs graphiques :

- Le premier (figure 3, page 63) permet d'étudier la ressemblance des distributions conditionnelles (en ligne et en colonne) ainsi que la décomposition des termes du χ^2 . Dans cet écran, les changements de couleurs permettent de mettre en évidence des structures typologiques sur les lignes ou les colonnes permet la structuration de la matrice des données.
- Le deuxième écran (figure 4, page 63) permet de mettre en relation les résultats graphiques d'une analyse des correspondances avec les deux matrices visualisant chacune les ressemblances entre les lignes et entre les colonnes. Là encore la possibilité de modifier la couleur, soit des points figurant sur les graphiques, soit des éléments des tableaux, donne un pouvoir illustratif puissant aux représentations factorielles.

On associe alors, d'une manière naturelle et très pédagogique, les concepts qui sous-tendent la double représentation graphique des points (lignes et colonnes) obtenus par l'analyse des correspondances. Le concept de distance est évident et celui des pondérations des points parfaitement perceptible par la taille donnée aux points.

Ce dernier écran présente un exemple d'association d'un modèle mathématique connu, avec un système interactif d'exploration. L'approche interactive ne fait qu'illustrer le modèle théorique en permettant à l'utilisateur de percevoir par sa propre intuition les éléments constitutifs de la théorie statistique sous-jacente. Ainsi, la manipulation des fonctionnalités informatiques, permet d'explorer par l'image des scénarios dont la justification se trouve dans le modèle théorique de l'analyse des correspondances.

Le travail informatique sous-jacent à l'élaboration des fonctionnalités interactives présentées ici est très important. L'élaboration d'objets informatiques adaptés et le niveau d'abstraction que nous avons atteint sont sans nul doute la partie la plus avancée de notre travail.

Chaque graphique est un élément d'une structure chaînée arborescente. Les "feuilles" de cette arborescence représentent les différents graphismes que nous avons étudiés plus haut. Les nœuds de l'arborescence symbolisent les "fenêtres" d'affichage à différente échelle de grossissement. La représentation interne de chaque graphique est un objet informatique structuré. Les éléments de cette structure sont aussi bien les paramètres de position, que les pointeurs sur les tableaux de données, que les pointeurs sur les fonctions graphiques applicables suivant les différentes commandes. Si le présent exposé n'est pas le lieu pour présenter ce travail informatique très technique, il n'en demeure pas moins que la rapidité des réponses obtenues pour l'exécution des commandes et la facilité d'utilisation apportée par la "souris" montrent que nous sommes sur la bonne voie.

Nous pensons disposer maintenant d'un environnement de développement approprié à la mise au point de programmes d'exploration interactive d'objets ou de méthodes statistiques.

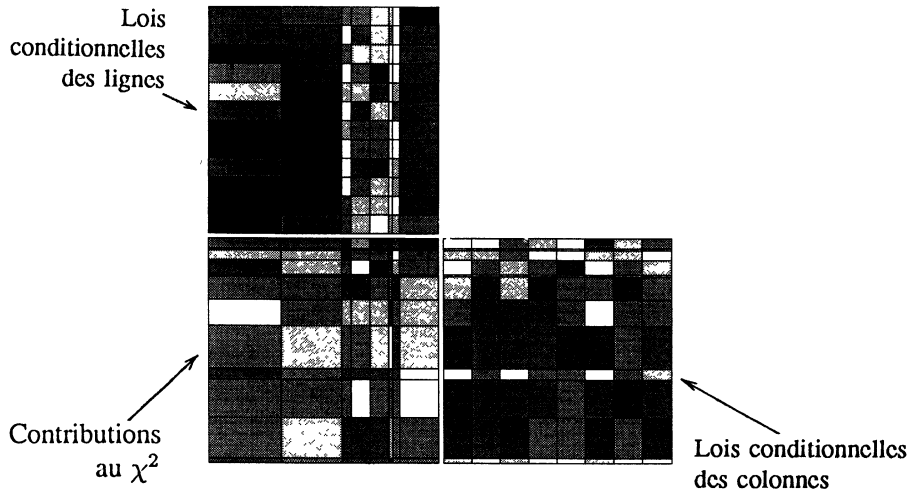


FIGURE 3

Représentation du contenu de l'un des écrans de travail

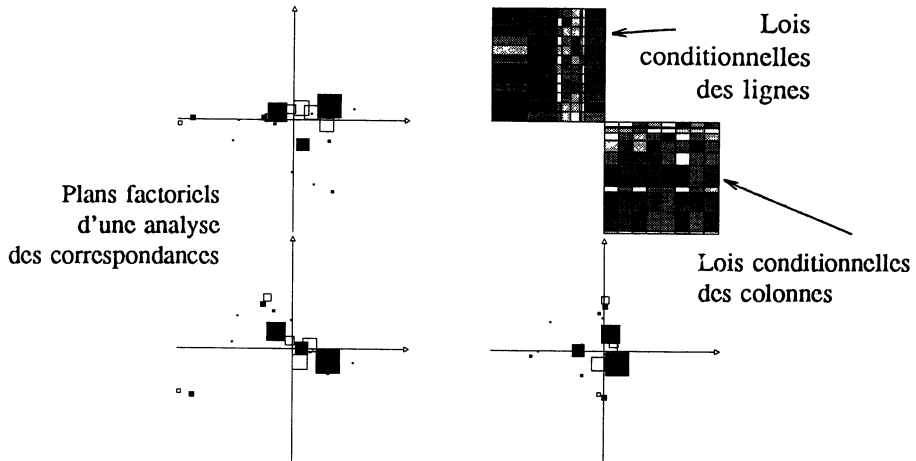


FIGURE 4

Représentation du contenu de l'un des écrans de travail

Les exemples présentés ici ne sont pas spécifiques. D'autres graphismes et d'autres fonctionnalités peuvent être introduits maintenant dans notre environnement sans difficulté majeure.

4. Conclusion

Si la richesse de l'exploration interactive ne peut être rendue par une rédaction manuscrite, il n'en demeure pas moins que les démonstrations du programme que nous avons réalisé, démontrent l'intérêt d'une telle approche.

Beaucoup des connaissances statistiques que nous avons acquises de façon abstraite par ailleurs, trouvent ici un support visuel au pouvoir illustratif puissant.

Le champ d'investigation qui s'ouvre à nous est très large et prometteur. Il peut, à terme, poser en d'autres termes l'approche d'un jeu de données statistiques.

Références

- [1] M. AERTS, and A. HOLTBERG, (1987). – *XploRe a computing Environment for Exploratory Regression and Density Estimation Methods*, Rheinische Friedrich Wilhelms Universität Bonn, Bonn, West Germany.
- [2] N. BARBARY, et J.L. CHANDRY, (1987). – “Essais Logiciels : MacSpin. (Tournez nuages!)”, *L'Ordinateur Individuel*, 94, 79-81.
- [3] R.A. BECKER, and J.M. CHAMBERS, (1984). – *S : An Interactive Environment for Data Analysis and Graphics*, Monterey, CA : Wadsworth/Brooks Cole.
- [4] R.A. BECKER, and W.S. CLEVELAND, (1987). – “Brushing Scatterplots”, *Technometrics*, 29, 127-142.
- [5] K. BERK, and D.R. COOK, (1987). – “Statistical Computing Software Reviews - MacSpin (Release 1.1)”, *The American Statistician*, Vol.41, N° 3, 233-236.
- [6] J. BERTIN, (1967). – *Sémiologie graphique*, Mouton-Gautier Villars, Paris.
- [7] G. CAVAUX, (1984). – “Réorganisation et représentation visuelle d'une matrice de données numériques : Un algorithme itératif”, *Revue de Statistique Appliquée*, vol XXXII, N° 4, 5-23.
- [8] W.S. CLEVELAND, (1987). – “Research in Statistical Graphics”, *Journal of the American Statistical Association*, Vol. 82, N° 398, 419-423.
- [9] D.L., DONOHO, P.J., HUBER, E., RAMOS, and H.M. THOMA, (1986). – “The Man-Machine-Graphics Interface for Statistical Data Analysis”, in *Statistical image Processing and Graphics*, eds. E.J. Weqman and D.J. DePriest, Statistics : Textbooks and Monographs, Vol 72, Marcel-Dekker Inc, New-York.
- [10] A.W., DONOHO, M. DONOHO, and M. GASKO, (1985), “MACSPINTM”, *Graphical Data Analysis Software*, Austin, Texas : D² Software.
- [11] C. GENEST, and P.E.J. GREEN, (1987). – “A graphical display of association in two-way contingency tables”, *The Statistician*, N° 36, 371-380.
- [12] J.A., HARTIGAN, and B. KLEINER, (1981). – “Mosaics for contingency tables”, in *Computer Science and Statistics : Proceedings of the 13th Symposium on the Interface*, eds. William, F.Eddy, Springer Verlag, New-York.
- [13] J.A., HARTIGAN, and B. KLEINER, (1984). – “A mosaic of television ratings”, *The American Statistician*, Vol 38, 32-35.
- [14] J. HINDE, (1988). – “Statistical Software Reviews (MacSpin. Graphical Data Software : Version 1.1, 1986)”, *Applied Statistics*, Vol 37, N° 1, 124-126.
- [15] A. KAY, (1984). – “Les logiciels”, *Pour la Science*, N° 85, 14-22.

- [16] L. LEBART, et A. MORINEAU, (1981), "Statistical significance criteria in multiple-choice data reduction and visualisation", *Psychometric Society Meeting CHAPEL HILL*, Rapport technique du CEPREMAP, N° 8127, Paris.
- [17] J.A. MCDONALD, (1982). – "Orion I : Interactive Computer Graphics in Statistics", technical report, Stanford University, Dept. of Statistics.
- [18] W. STUETZLE, (1986). – "Design and Implementation of Plot Windows", in *Proceedings of the Statistical Computing Section, American Statistical Association*, 32-40. Addison-Wesley Publishing Compagny.
- [19] W. STUETZLE, (1987). – "Plot Windows", *Journal of the American Statistical Association*, Vol. 82, N° 398, 419-423.
- [20] J.W. TUKEY, (1977). – *Exploratory Data Analysis* Addison-Wesley Publishing Compagny.