REVUE FRANÇAISE D'INFORMATIQUE ET DE RECHERCHE OPÉRATIONNELLE. SÉRIE VERTE

FRANÇOIS WASSERVOGEL

J. LECHAT

Une méthode de classification d'items quantitatifs

Revue française d'informatique et de recherche opérationnelle. Série verte, tome 5, n° V2 (1971), p. 31-58

<http://www.numdam.org/item?id=RO_1971__5_2_31_0>

R.I.R.O.
(5e année, V-2, 1971, 31-58)

UNE METHODE DE CLASSIFICATION D'ITEMS QUANTITATIFS

par François WASSERVOGEL (1) et J. LECHAT (2)

Résumé. — *Il est extrêmement fréquent que des ensembles nombreux soient décrits par une suite de qualités s'ajoutant ou non à une description quantitative, il est alors utile de procéder à des comparaisons ou à des rangements des objets ou des qualités qui les définissent. Les auteurs ont montré qu'il suffit que quelques hypothèses simples soient vérifiées par l'ensemble objets-qualités pour qu'il soit possible de construire un classement dual des objets et des qualités. On montre en effet que la possession d'une qualité induit souvent la possession d'un ensemble de qualités. Si, de plus, cette induction est cohérente pour l'ensemble des objets, on peut mettre en évidence une structure hiérarchique des qualités.*

AVANT-PROPOS

L'analyse des ensembles nombreux et multidimensionnels fait l'objet de nombreuses études dont le but est d'exhiber leurs structures sous-jacentes. Le professeur Benzecri et les chercheurs de son laboratoire, en choisissant le parti d'aborder l'analyse des données sous l'aspect le plus général, ont fourni à la Statistique une base théorique très élaborée. La puissance des ordinateurs leur a permis ensuite de mettre au point un remarquable ensemble de techniques descriptives. Bien d'autres chercheurs ont largement contribué à faire progresser l'Analyse des Données en proposant des outils répondant à des objectifs particuliers. Citons l'analyse des correspondances, en composantes principales, l'analyse discriminante, la classification automatique, la régression multiple linéaire ou non... Mais la plupart de ces techniques ne posent aucune hypothèse *a priori* sur la distribution et la structure stochastique des données.

L'objet de cette note est tout autre : on compare le résultat d'une observation à une structure particulière construite à partir de certaines hypothèses

(1) Ingénieur économiste à la Régie Renault, professeur à l'Institut de Statistique de l'Université de Paris.
(2) Informaticien à la Régie Renault.

que le contexte de l'observation rend plausibles. La conclusion se limitera simplement à retenir ou rejeter la structure de référence.

Il nous est en effet apparu au cours de nombreuses analyses de données qualitatives réalisées à partir de l'analyse factorielle des correspondances qu'une même structure se retrouvait fréquemment. Il s'agit d'une structure dans laquelle la possession d'une qualité donnée implique la possession d'un ensemble de qualités. Nous avons alors appelé cette structure : la Structure Hiérarchique.

Les structures hiérarchiques ont été étudiées depuis fort longtemps par Guttman, puis présentées par Matalon dans un ouvrage *L'Analyse Hiérarchique* cité en référence. Mais le présent article propose d'une part, un modèle de construction systématique de structure hiérarchique et d'autre part, un test statistique basé sur la comparaison d'une distribution de structures simulées autour d'une structure de référence donnée et de la structure observée.

Lerman dans « Les Bases de la Classification Automatique » use du même procédé pour déterminer la valeur d'une classification. Mais alors que M. Lerman réalise les comparaisons au travers de préordonnances calculées à partir du tableau des données, nous avons choisi d'effectuer directement les comparaisons sur les tableaux de données.

Il est clair que si notre test se révèle significatif, notre conclusion se limitera à la négation de l'existence d'une structure hiérarchique sans pour autant proposer une structure alternative. Néanmoins, il est important de rappeler que les structures hiérarchiques s'observent fréquemment (Étude de qualités non indépendantes par exemple) et qu'elles peuvent masquer des structures secondaires. Enfin, l'intérêt de la structure hiérarchique, lorsqu'elle existe, est de permettre la classification ordinale des items qualificatifs d'une part et des individus d'autre part.

INTRODUCTION

La relation d'ordre est l'une des Relations les plus naturelles. L'esprit humain tend à classer toute collection d'objets à laquelle il lui arrive d'être confronté. Par définition la pensée est unidimensionnelle alors que la réalité appartient à un espace de dimension infinie. Pour cerner cette réalité et selon l'optique retenue, l'homme définira une classification : le premier, le dernier; le meilleur, le pire.

Mais si les classifications ont un support rationnel lorsque les grandeurs étudiées sont quantitatives, il n'en est pas de même si l'on s'intéresse à des propriétés ou des qualités possédées par des individus.

L'objet de cet article est de proposer une classification d'items qualitatifs sans jamais faire appel à une *valorisation quantifiable* de chacun d'eux. Pour qu'une telle classification ait un sens, il faut que des hypothèses simples soient vérifiées. Comme nous le verrons elles le sont fréquemment.

La classification proposée est purement statistique et repose sur la comparaison des suites de propriétés possédées par les individus. Il est donc toujours possible après avoir établi une telle classification de confronter les résultats avec une classification issue d'une échelle de valeurs des propriétés ou des qualités. On abordera successivement les trois parties suivantes.

A) Exposé du problème, la Relation d'ordre sur les individus dont se déduit celle sur les qualités.

B) Exemple d'un cas concret.

C) Le programme de calcul.

A. PREMIERE PARTIE

A.1. Exposé du problème

Les données sont :

— une collection de p objets,

— un ensemble de q qualités que peuvent posséder les objets,

— un tableau à double entrée de p lignes et q colonnes indiquant si l'objet i possède ou non la qualité j.

Le but de cette étude est de rechercher s'il est possible de créer, à partir de la seule connaissance des possessions, une classification des objets et des qualités. Il est donc clair que si une telle classification existe elle n'engage que les objets et qualités étudiés, que les mêmes qualités analysées sur d'autres objets pourraient conduire à des classifications différentes. En renonçant à associer aux qualités une échelle de valeur, on se limite à une approche quantitative qui peut dans certains cas se révéler tout à fait insuffisante. Le lecteur est donc averti que les résultats obtenus sont de nature purement statistique et que la solution a les avantages de ses inconvénients : seule la possession ou non d'une qualité est déterminante.

Hypothèse fondamentale

Les propriétés retenues doivent permettre, au regard de l'observateur, de définir une relation de préférence sur les objets :

> L'objet A sera préféré à B, si et seulement si, l'objet A possède entre autres les propriétés de B.

Cette hypothèse interdit donc l'étude simultanée de qualités et de défauts, de compliments et de reproches, de points forts et de points faibles. Il faut donc que l'observateur attache à chaque propriété étudiée une valeur dans le *même sens*.

La relation de préférence énoncée ci-dessus sera dite *totale* si de tout couple d'objets A et B ne possédant pas les mêmes propriétés on sait dire lequel est *préférable*. Dans le cas où A et B posséderaient exactement les mêmes propriétés, l'observateur se déclarerait *indifférent* à l'égard de A et de B.

Une conséquence de l'existence d'une relation de préférence totale

Si A possède une propriété que ne possède pas B, alors A est préférable à B. En effet, la relation de préférence étant totale, ou bien A possède les qualités de B, ou bien B possède les qualités de A. Ici A possède une qualité Q que ne possède pas B, A possède donc les qualités de B et au moins la qualité Q en plus. A est donc préférable à B.

Si la relation de préférence est *totale*, il existe une relation de préférence sur les propriétés qui se déduit de celle des objets.

Relation de préférence sur les propriétés

Si la relation sur les objets est *totale*, on dira que la propriété P est préférable à la propriété Q si cette dernière est possédée par un objet qui ne possède pas la propriété P.

Cette définition repose sur l'idée que si un objet A possède une propriété qu'un objet B ne possède pas, alors A est préféré à B et ce « grâce » à la propriété considérée. Cette propriété sera alors considérée comme préférable à *toutes* les propriétés de l'objet B.

Corollaires. Si la relation de préférence sur les objets est *totale* :

— Si l'objet A possède plus de propriétés que l'objet B, alors A est préférable à B.

— Si la propriété P est possédée par *moins* d'objets que la propriété Q, alors P est préférable à Q.

Classification des objets et classification des propriétés

Nous venons de voir que si la relation de préférence sur les objets est *totale*, il suffit de connaître le nombre de propriétés possédées par un objet ou le nombre d'objets possédant une propriété pour trancher entre deux objets ou deux propriétés.

On classera alors les objets en les rangeant selon le nombre décroissant de propriétés possédées et on classera les propriétés, par mesure de cohérence, également selon le nombre décroissant d'objets qui les possèdent.

Enfin pour que le rang des objets varie de 1 à p (nombre d'objets) et que le rang des propriétés varie de 1 à q (nombre de propriétés), on attribuera aux *ex-aequo* des rangs différents.

Ainsi la relation de préférence sur les objets nous a permis de construire deux classifications, l'une des objets et l'autre des propriétés. Nous allons maintenant donner de chacune d'elles une interprétation.

La classification des objets traduit la possession d'un nombre plus ou moins grand de propriétés et peut donc être interprétée en valeur, l'objet de rang 1 possède le plus grand nombre de propriétés et est préféré à tous les autres objets. La classification est donc cardinale et attribue à chaque individu une valeur.

La classification des propriétés a une signification très différente, elle traduit le niveau d'apparition des propriétés sur les individus de mieux en mieux classés. Le classement des propriétés est donc strictement ordinal et ne permet pas la comparaison des propriétés entre elles, mais il exprime le fait qu'un objet possédant la propriété P possède *aussi* toutes les propriétés moins bien classées.

Un exemple simple illustrera ces deux aspects : les individus sont des étudiants et les propriétés des qualités. Si les qualités possédées par les étudiants définissent une relation de préférence *totale* sur les étudiants, alors on pourra exhiber une classification des étudiants d'où l'on déduira la classification des propriétés.

On aura par exemple :

Rang 1 : esprit d'initiative.
Rang 2 : faculté de synthèse.
Rang 3 : rapidité de lecture.
Rang 4 : bonne mémoire.
Rang 5 : faculté de concentration.
. . . . .

On peut en déduire qu'il y a moins de lecteurs rapides que d'étudiants doués d'une bonne mémoire, mais lorsque l'on observe chez un étudiant la faculté de lire vite on peut en déduire qu'il a aussi une bonne mémoire. On voit qu'en aucun cas la lecture rapide est plus *importante* que la possession d'une bonne mémoire mais que la connaissance de la première de ces qualités nous fournit plus d'information sur l'individu que la seconde.

On peut proposer un second exemple. On présente à un échantillon d'individus une série d'affiches publicitaires en leur demandant d'attribuer à chacune d'elles les qualités qu'elles méritent. Notons au passage que les qualités peuvent être issues d'une liste donnée et fermée ou bien que la question posée peut être parfaitement ouverte :

1er cas : Parmi les qualités figurant dans la liste jointe, quelles sont celles que vous attribueriez à l'affiche A, puis l'affiche B, etc.

2e cas : Quelles sont les qualificatifs élogieux que vous employeriez à l'égard de l'affiche A, B, etc.

Si le dépouillement des réponses doit se faire de manières très différentes pour chacun des deux cas, l'esprit reste le même et l'on se trouvera confronté à un tableau indiquant si la qualité j a été ou non attribuée à l'affiche i.

Supposons que la relation de préférence soit totale. L'analyse des résultats permet alors de classer les affiches par ordre de préférence mais la « valeur ajoutée » réside dans le classement des qualités qui représente un *ordre de perception* des qualités par les individus, un ordre de liaison des qualités entre elles traduisant le fait que la *reconnaissance* de la qualité P implique celle de la qualité Q. Les résultats de cette analyse pourront alors être valablement complétés par une série de tests psychologiques au niveau de la signification de chacune des qualités.

Conclusion du paragraphe 1

La *structure emboîtée* que nous venons de décrire s'observe très fréquemment dès lors que l'on s'intéresse à un ensemble d'objets et un ensemble de qualités qui sont soumis au jugement et à l'appréciation d'individus. La structure psychologique des uns et des autres conduit en effet à cette forme de classification unidimensionnelle. Il en va tout autrement, c'est évident, de propriétés ou de qualités objectives que peuvent posséder des individus. Exemple : yeux bleus, cheveux blonds... Mais en tout état de cause la mise en évidence d'une structure emboîtée est un apport essentiel dans l'analyse des données.

Dans le second chapitre on s'éloignera des hypothèses théoriques pour constater qu'une structure emboîtée *parfaite* ne s'observe « jamais » et qu'il est nécessaire de faire appel à la statistique pour dire si l'on ne s'écarte pas « trop » de la perfection.

A.2. La relation de préférence et la statistique

Tout ce qui a été dit dans le § 1 repose sur l'observation d'une *structure emboîtée parfaite*. Mais il est clair qu'une telle structure n'existe qu'en théorie et que tous les cas concrets étudiés s'écartent plus ou moins de la « perfection ». L'apport d'une approche statistique réside dans l'analyse de la *structure observée* face à la *structure parfaite* et permet de préciser si l'écart est significatif ou non. En tout état de cause si la relation de préférence est *presque* totale les classifications des objets et des qualités ne seront que *presque* satisfaisantes.

Il faut néanmoins ajouter que la classification des qualités ou des propriétés « souffre » beaucoup moins que celle des objets puisque par définition elle traduit une structure collective des propriétés alors que l'interprétation individuelle du rang des objets peut être mise largement en défaut. Deux objets dont les ensembles des qualités ne seraient tels que l'un des ensembles contienne l'autre, ne sont pas absolument comparables et échappent donc à toute classification. Par contre la définition du rang d'une qualité, niveau d'apparition, repose sur l'ensemble des objets et n'est pas remise en question lorsque quelques objets échappent à la règle collective.

Rappel. La donnée de base est un tableau rectangulaire à p lignes et q colonnes rempli de 1 ou de 0. A l'intersection de la ligne i et de la colonne j on trouve 1 ou 0 selon que l'objet i possède ou non la qualité j.

Définitions

Marge des objets : c'est la suite des nombres de qualités possédées par les p objets.

Marge des qualités : c'est la suite des nombres d'objets qui possèdent les q qualités.

On remarquera que la marge des objets est la suite des sommes de chaque ligne et que la marge des qualités est la suite des sommes de chaque colonne du tableau à p lignes et q colonnes.

Marge ordonnée : marge dont la suite des nombres est ordonnée par valeurs décroissantes.

Structure observée : tableau des données observées.

Structure emboîtée parfaite : tableau dans lequel la relation de préférence est *totale*. A partir de la structure observée, on peut construire une structure emboîtée parfaite qui *respecte la marge des objets*. Pour simplifier, on appellera cette structure : structure déduite. On remarquera que si l'on a pour seule contrainte le respect de la marge des objets la construction de la structure déduite ne présente aucune difficulté.

Structure aléatoire : il s'agit d'un tableau dont *la marge des objets est donnée* et dont on remplirait les lignes « au hasard ». Pratiquement cela revient à considérer un objet et le nombre de qualités qu'il possède et de lui attribuer les qualités issues d'un tirage sans remise dans une urne contenant toutes les qualités. Il est alors facile de montrer que les espérances des nombres composant la marge des qualités sont égales (égales au nombre moyen d'objets possédant chaque qualité).

Structure aléatoire parfaite : (abus de langage destiné à simplifier la présentation). Cette structure respectant la marge des objets conduit à une marge des qualités composée de nombres égaux (entiers ou fractionnaires).

Le principe de l'analyse statistique

— La *structure emboîtée parfaite* et la *structure aléatoire parfaite* sont les deux extrêmes d'une structure qui peut évoluer continuement de l'une à l'autre.

— L'approche statistique repose sur le *respect* de la marge *observée des objets* et sur la *seule* considération des marges des qualités.

— On comparera la *structure observée* à la *structure emboîtée parfaite* et à *la structure aléatoire* et on déterminera quelle est la probabilité que la structure observée soit le résultat d'un tirage aléatoire. Si cette probabilité est inférieure à 5 % (1) on rejettera cette hypothèse et l'on admettra que la structure observée est une structure emboîtée « presque parfaite ». On pourra alors retenir le classement des objets comme étant celui des objets rangés par valeurs décroissantes du nombre de qualités possédées et pour classement des qualités celui des qualités rangées par valeurs croissantes du nombre d'objets les possédant.

(1) $\alpha = 5\%$ représente le seuil du test.

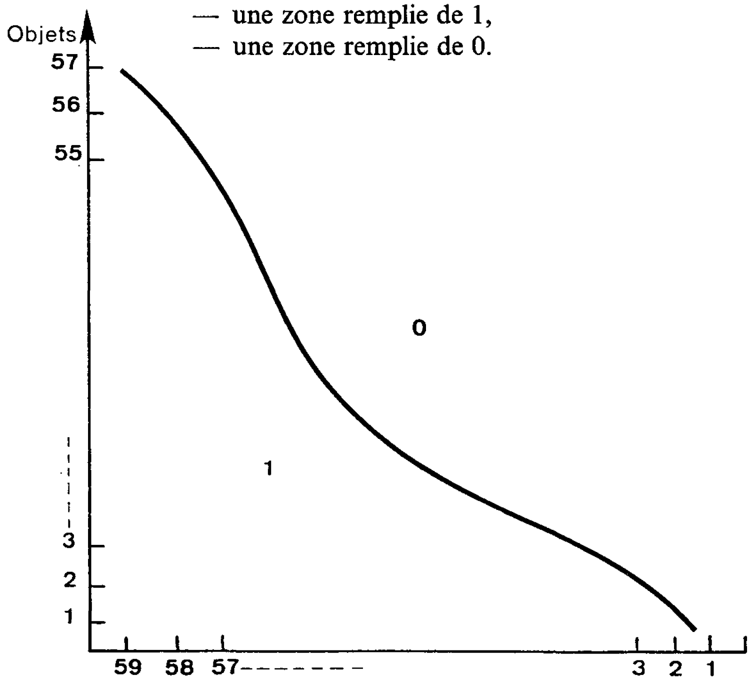
Dans le cas où la probabilité citée plus haut est supérieure à 5 % on s'abstiendra de toute conclusion opératoire.

A.2.1. Visualisation des diverses structures

a) On peut fournir une représentation simple des structures en ordonnant les lignes et colonnes des différents tableaux.

La *structure emboîtée parfaite* conduit alors à un tableau où deux zones distinctes apparaissent :

— une zone remplie de 1,
— une zone remplie de 0.



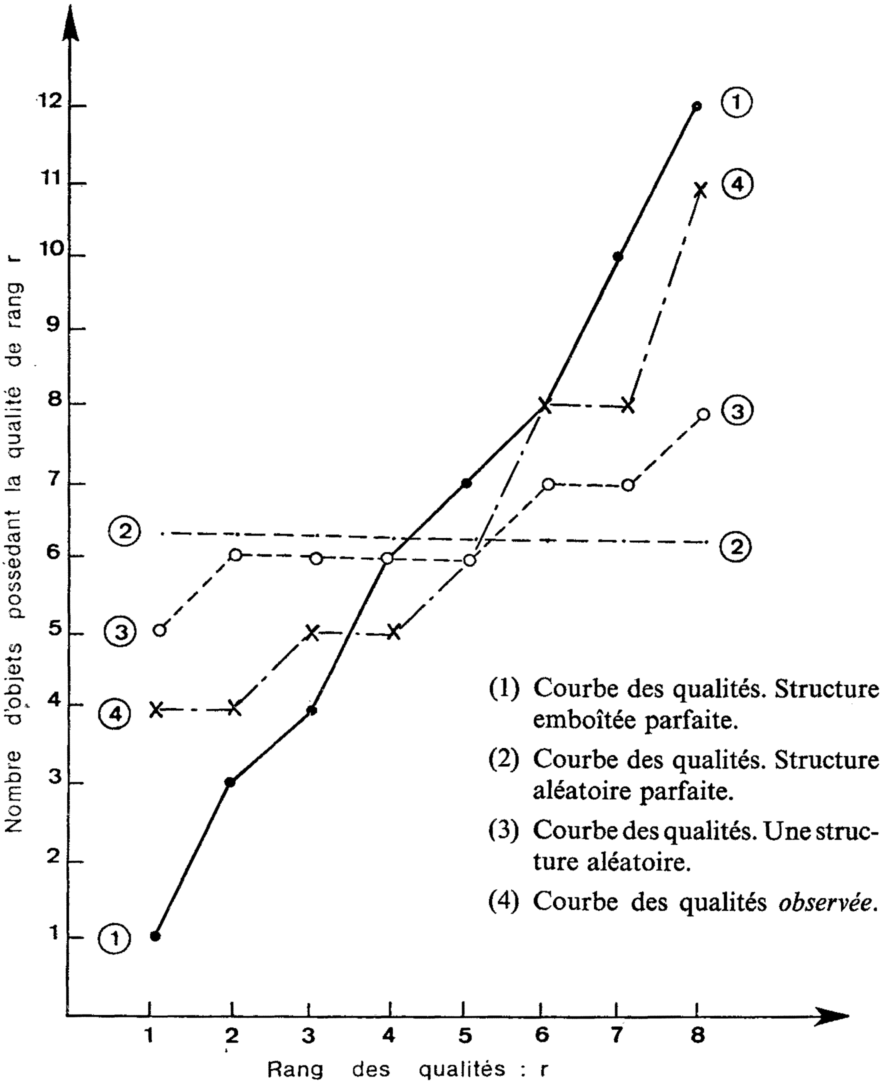
La *structure observée* s'écarte plus ou moins de cette situation particulière.

b) Un autre graphique permet plus simplement encore de se faire une idée de la qualité de la structure observée. On range les qualités selon le nombre croissant d'objets qui les possèdent.

On porte sur deux axes les rangs des qualités en abscisse et le nombre d'objets qui les possèdent en ordonnée.

A partir de la structure emboîtée parfaite déduite de la marge des objets on obtient une première suite de points. La structure observée en fournit une seconde. Enfin sur le même graphique on portera une structure aléatoire et la structure aléatoire parfaite.

On peut alors juger de la position de la structure observée par rapport à la structure emboîtée parfaite et à la structure aléatoire.



p = 12 objets q = 8 qualités

Marge observée des objets (donnée de base du problème)

	Rang :	1	2	3	4	5	6	7	8	9	10	11	12
Nombre de : qualités possédées		8	7	7	6	5	5	4	3	2	2	1	1

A.2.2. Le test statistique

Il repose précisément sur la proximité de la courbe observée des qualités par rapport à la courbe relative à la structure emboîtée parfaite et à la courbe relative à une structure aléatoire. On *simule* alors un grand nombre de structures aléatoires afin de construire l'histogramme des distances de chacune des structures aléatoires à la courbe de structure emboîtée parfaite d'une part et de la courbe de structure aléatoire parfaite d'autre part. Enfin en calculant les distances correspondantes de la courbe de la structure observée on situera cette dernière sur chacun des deux histogrammes. Cette dernière opération permettra de retenir ou non l'hypothèse selon laquelle la courbe observée aurait pu être obtenue par simulation.

Notation

N_0^r, N_a^r, N_e^r, N_j^r le nombre d'objets possédant la propriété de rang r, respectivement dans la structure observée, aléatoire parfaite, emboîtée et la J^e simulation, et en retenant une formule de distance très simple.

$$D = \sqrt{\sum_{r=1}^{q} (N_0^r - N_e^r)^2}$$

$$D_j = \sqrt{\sum_{r=1}^{q} (N_j^r - N_e^r)^2}$$

$$d = \sqrt{\sum_{r=1}^{q} (N_0^r - N_a^r)^2}$$

$$d_j = \sqrt{\sum_{r=1}^{q} (N_j^r - N_a^r)^2}$$

B. DEUXIEME PARTIE

Exemple concret d'application

Le caractère confidentiel de l'étude que nous avons abordée ne nous permet pas de préciser la nature des objets et des propriétés étudiées. Retenons simplement qu'il s'agit de biens de consommation durable analysés au travers de qualités qui leur ont été attribuées par un important échantillon d'individus. Notre point de départ sera un tableau à double entrée permettant de savoir si la qualité j a été ou non attribuée à l'objet i. A l'intersection de la ligne i et de la colonne j on trouvera un 1 ou un zéro selon que l'objet i possède ou non la qualité j. Nous n'entrerons pas dans le détail des intermédiaires qui conduisent à ce tableau.

Ainsi 57 objets possédant tout ou partie des 59 qualités retenues ont été étudiés. Dans toute la suite nous présentons les résultats dans l'ordre dans lequel ils ont été mis en évidence afin de respecter la démarche que nous avons suivie et l'enchaînement logique qui nous a amenés à élaborer le test que nous présentons.

B.1. Première phase d'étude

Nous avons commencé par analyser le tableau (objet $\times$ qualité) par une technique d'analyse factorielle qui nous a conduit à admettre que la structure des objets « s'étirait » essentiellement selon le *nombre* de qualités possédées.

Ce résultat est suffisamment étonnant pour paraître trivial. Après réflexion et analyse, force était de constater que le nombre de qualités possédées était effectivement un élément fondamental de la connaissance que l'on pouvait avoir des objets. Nous avons alors été « tout naturellement » conduits à émettre une hypothèse de *hiérarchie des qualités.* En effet, comment admettre que le *nombre* de qualités possédées est porteur d'une information importante quant à la structure de l'association objet-qualité sans supposer que *l'additivité* des qualités a un sens.

Or si on n'attribue pas de valeurs aux qualités possédées par les objets, il semble que l'additivité des qualités n'a de sens que si l'on admet une structure emboîtée dans laquelle lorsqu'un objet A possède plus de qualités qu'un objet B alors il possède entre autres qualités celles de B. On peut ainsi interpréter le *nombre* de qualités comme un indicateur de niveau hiérarchique de qualité puisque si l'hypothèse est vérifiée il est certain que de deux objets c'est celui qui possède le plus grand nombre de qualités qui est *préférable.* Dans la suite, nous nous attacherons à justifier l'hypothèse émise en décrivant une à une les étapes de l'analyse.

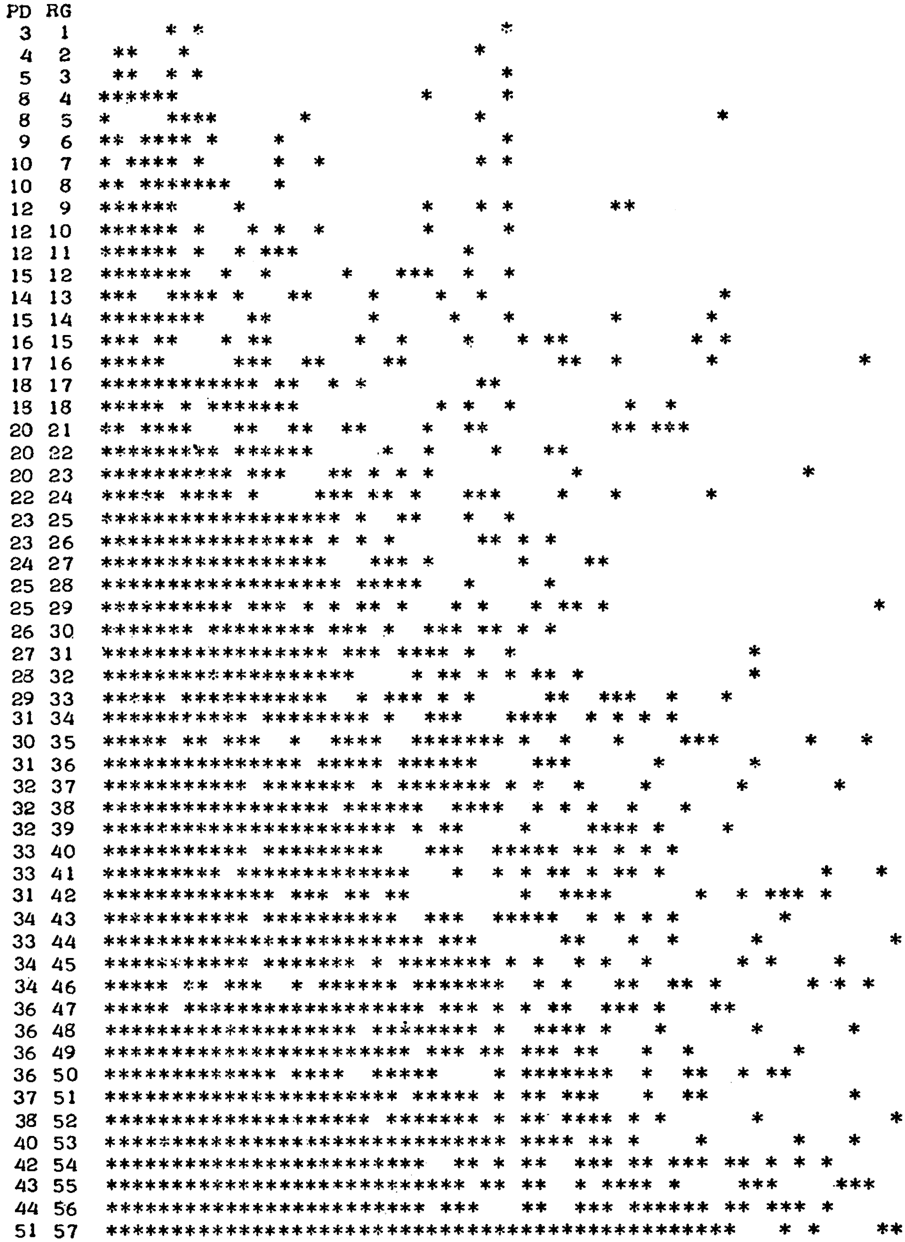
B.2. Seconde phase. — Description graphique de la structure objet-qualité.

Afin de disposer d'un support visuel de l'hypothèse émise dans la première phase, on représente le tableau objets-qualités en ordonnant les lignes et les colonnes. Les objets sont ordonnés selon le nombre décroissant de qualités qu'ils possèdent et les qualités selon le nombre croissant d'objets qui les possèdent.

— Si l'hypothèse est strictement vérifiée, on doit alors observer deux zones distinctes : l'une ne contenant que des 1, l'autre que des zéros.

— Dans le cas concret que nous avons traité on observe en fait trois zones :

- Zone 1 : à quelques rares exceptions près parfaitement explicables par la nature des données, cette zone ne contient que des 1.
- Zone 3 : ne contenant que des zéros.
- Zone 2 : zone intermédiaire entre les zones 1 et 3 dans laquelle il apparaît une répartition de 0 et de 1 (dans le graphique joint les 1 ont été remplacés par des étoiles * et les 0 par des blancs).



Graphique du tableau observé lignes et colonnes ordonnées

RG : rang des objets
PD : marge des objets

L'existence même de la zone intermédiaire (zone 2) montre que la structure objets-qualités du cas étudié s'écarte de la structure de hiérarchie parfaite. Le test statistique mis en œuvre a pour but de préciser si cet écart est significatif ou non. En d'autres termes si la perfection n'est pas atteinte on cherche à déterminer si les résultats observés permettent néanmoins de retenir une hiérarchie des qualités et partant, des objets.

Une analyse beaucoup plus fine sur la nature même des qualités nous a permis de scinder l'ensemble des qualités en deux groupes, deux objets possédant le même nombre de qualités se distinguant par la répartition de leurs qualités selon ces deux groupes. Ainsi se trouve expliquée la zone intermédiaire. Notre objet se limitant à la description de la hiérarchie, nous n'irons pas plus avant dans la description de ce nouvel aspect.

B.3. Le test statistique

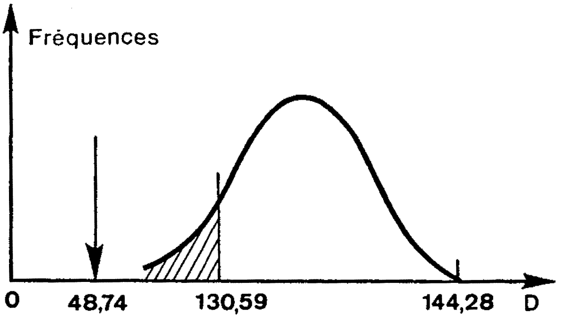
La donnée de base est le tableau ordonné de correspondance entre les objets et les qualités. De ce tableau on retient la marge des objets et la marge des qualités. La marge de l'objet i est le nombre de qualités qu'il possède, on appellera marge des objets la suite ordonnée de ces nombres. La marge des qualités se définit de la même façon.

En respectant la *marge des objets* on construira d'une part la marge des qualités correspondant à la hiérarchie parfaite et d'autre part, par simulation, un échantillon de marges correspondant à un remplissage au hasard du tableau.

Le test statistique repose alors sur la comparaison de la marge observée des qualités aux marges de la hiérarchie parfaite et des simulations, et son but est de décider si la marge observée est plus proche de la marge de hiérarchie parfaite ou des marges dues au hasard.

B.4. Les résultats du test statistique

Le nombre d'objets (57) et le nombre de qualités (59) étant élevés, il a été nécessaire de réaliser 1 000 simulations. Les deux tests mis en œuvre se sont révélés hautement significatifs et conduisent au rejet de l'hypothèse selon laquelle le tableau observé pourrait être dû au hasard.



B.4.1. Test 1 : Histogramme des distances des marges simulées à la marge de hiérarchie parfaite. C'est un histogramme sensiblement normal.

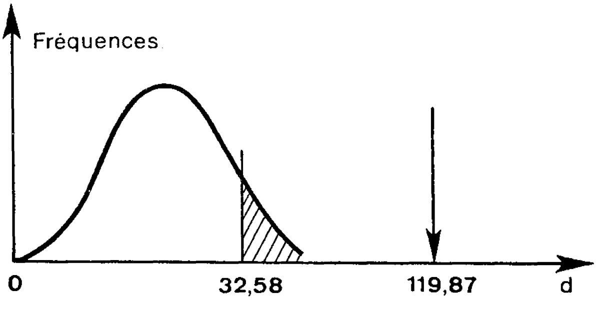
Intervalle d'acceptation à 95 % : [130,59, 144,28].

Distance observée : 48,74.

B.4.2. Test 2 : Histogramme des distances des marges simulées à la marge aléatoire parfaite.

Intervalle d'acceptation à 95 % : [0, 32,58].

Distance observée : 119,87.



C. TROISIEME PARTIE : LE PROGRAMME

Mode d'utilisation du programme Simul

Rappel sur la position du problème

Il s'agit essentiellement d'étudier comment se répartit une population de « qualités » sur une population « d'objets ». Une telle répartition se représente géométriquement par un tableau à double entrée dont les cases sont remplies de 0 et de 1. Pour fixer les idées, nous dirons que la case $P(i,j)$ contient la ou $P(i,j) = 1$, si l'objet qui est à la ligne i possède la qualité qui est à 1, colonne j; $P(i,j)$ vaudra 0 dans le cas contraire. On connaît bien ces tableaux qui établissent une correspondance « ensembliste » entre deux populations. En effet, un tel tableau peut représenter, par exemple, l'équipement de ménages en appareils électriques, ou encore la composition de différents corps complexes en éléments simples, la présence ou l'absence de certaines espèces végétales sur des parcelles de terrains, la présence à divers postes d'équipes qui se remplacent par roulement. Les cas très divers et l'on conçoit que les structures de ces tableaux soient très variés (nous entendons par « structure » la distribution des « 1 » dans le tableau).

La reconnaissance générale de ces structures est le domaine de prédilection de certaines techniques telles que l'analyse factorielle des correspondances. Notre objet n'est pas de donner une méthode de reconnaissance valable pour de multiples structures, méthode dont la puissance suppose une certaine complexité, mais nous proposons un moyen simple de savoir si un tableau possède ou non une certaine structure particulière qui a été nommée « hiérarchie parfaite ». Cette structure a été présentée en détails dans les pages précédentes et nous rappelons seulement une brève définition. Supposons donné un tableau de correspondance entre des « objets » et des « qualités ». Ces objets et ces qualités ont des « marges » : nombre de qualités possédées pour un objet i (marges sur les objets), nombre d'objets possédant la qualité j (marges sur les qualités). Nous dirons que les qualités sont parfaitement hiérarchisées si un objet i qui possède la qualité j, possède aussi toute qualité de marge supérieure ou égale à la marge de j, et ceci quelque soit l'indice i. Autrement dit si un objet possède une qualité, il possède aussi toutes les qualités plus fréquentes. On a vu que cette propriété observée sur les objets se transmet aux qualités, et réciproquement.

Le test proposé compare les marges sur les qualités « observées », d'une part aux marges correspondant à une hiérarchie parfaite et, d'autre part, aux marges provenant d'un tableau simulé par tirage au hasard, étant bien entendu que dans les trois cas les marges sur les objets sont les mêmes.

Pour effectuer ces comparaisons, nous construisons un indice de distance entre les marges, que nous appellerons « distance » pour abréger. Cette distance est à peu de chose près un écart quadratique entre les diverses marges. Mais avant d'expliciter ces distances, précisons comment les diverses marges ont été obtenues.

— Les marges observées sur les qualités sont évidemment une donnée du problème.

— Les marges sur les qualités correspondant à une hiérarchie parfaite sont obtenues de la façon indiquée par l'exemple suivant :

$$\begin{array}{|cccccc|c} 1 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 & 3 \\ 1 & 1 & 1 & 0 & 0 & 0 & 3 \\ 1 & 1 & 1 & 1 & 0 & 0 & 4 \\ 1 & 1 & 1 & 1 & 1 & 0 & 5 \end{array} \left.\vphantom{\begin{array}{c}1\\3\\3\\4\\5\end{array}}\right\} \text{marges observées sur les } \textit{objets}.$$

On ordonne les marges sur les objets.

A la ligne i, en partant de la colonne de gauche on remplit consécutivement de 1 un nombre de cases égal à la valeur de la marge de l'objet i. Les marges sur les colonnes de ce tableau sont les marges cherchées. Ce calcul montre aussi qu'à des marges données sur les objets il correspond une hiérarchie parfaite unique.

— Les marges sur les qualités obtenues par simulations : Le processus est analogue au précédent, mais au lieu de remplir les cases consécutivement, pour chaque ligne on tire au hasard les indices des colonnes où on mettra un 1. Ces tirages au hasard sont faits par le programme « RANDU » d'I.B.M.

A ces différentes marges nous associons des « courbes » (terminologie du listing ci-joint) qui ne sont rien d'autre que les histogrammes des marges. La « courbe de hasard pur » est la droite horizontale dont l'ordonnée est égale à la fréquence moyenne d'apparition des qualités.

Le calcul des distances s'explicite alors très simplement : si $M_1(J)$ et $M_2(J)$ sont deux marges relatives à la qualité J, si NQ est le nombre de qualités, nous posons :

$$D\,12 = \left(\sum_{J=1}^{NQ} M_1(J) - M_2(J)^2\right)^{1/2}$$

au coefficient $1/NQ$ près on a bien un écart quadratique moyen.

A chaque simulation on calcule la distance des marges simulées aux marges de hiérarchie parfaite et la distance de ces mêmes marges au « hasard pur ». Ces distances sont mises en mémoire, puis on trace les histogrammes de leurs distributions.

On calcule également les distances des marges observées aux marges de hiérarchie parfaite et au hasard pur.

La comparaison de ces distances avec les distributions de distances correspondantes obtenues par simulation constitue le test proprement dit.

Le programme

— La terminologie utilisée est indiquée dans les cartes commentaires qui sont en tête du programme (cf. listing ci-joint).

— Les cartes de données sont les suivantes :

* 1re carte, lue selon le FORMAT 1, i.e. en 2014.
  - — Nombre d'objets.
  - — Nombre de qualités.
  - — Nombre de simulations.
  - — Pourcentage d'erreur admissible dans le rejet de l'hypothèse « la structure est due au hasard ».
  - — Un nombre pair quelconque servant à initialiser le programme de tirage de nombres au hasard. Ce nombre devra être modifié à chaque passage du programme en machine.

* 2e carte et suivantes, lues selon le FORMAT 1,
  — les marges sur les objets.
* Cartes suivantes, lues selon le FORMAT 1,
  — les marges *observées* sur les qualités, ordonnées par valeurs décroissantes.

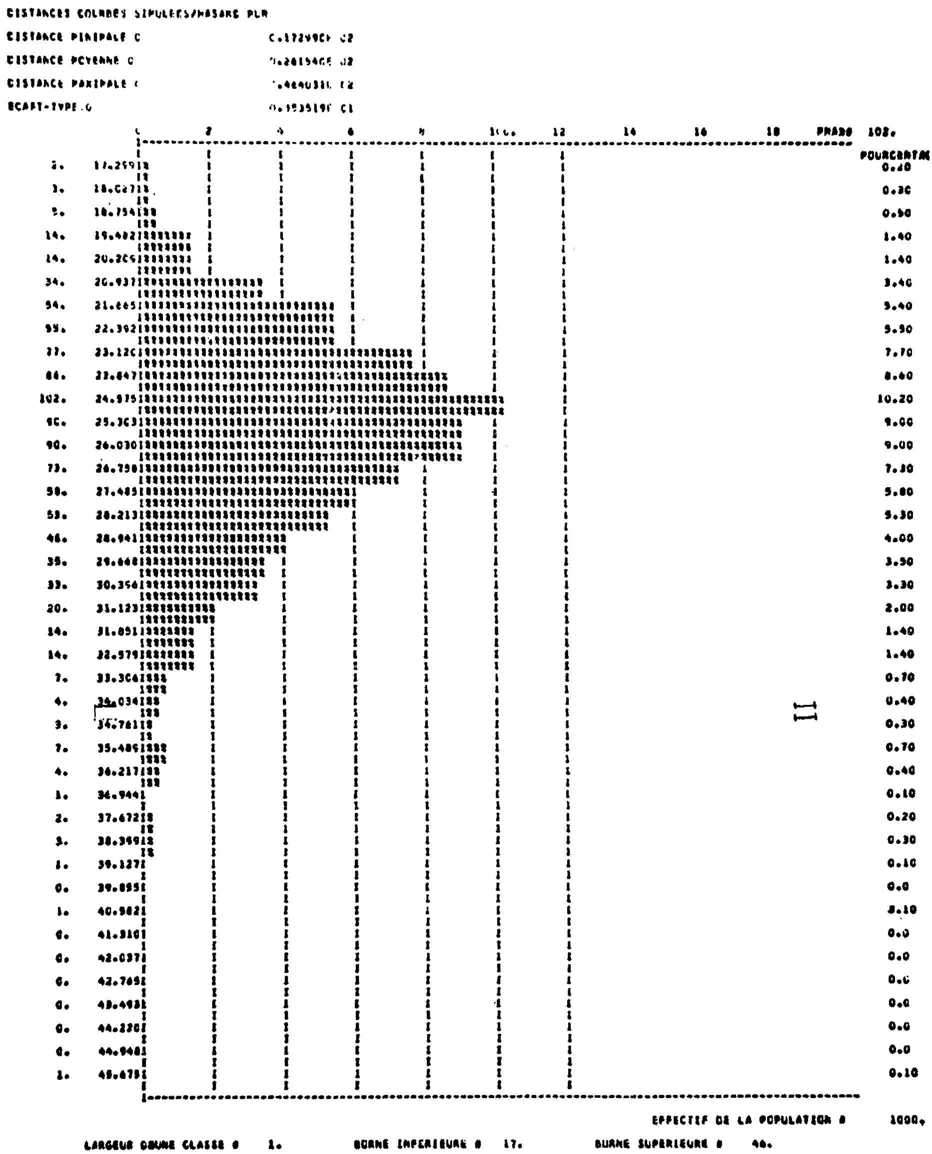
— Le programme édite :
  — Les dimensions du tableau et le nombre de simulations demandées.
  — Le taux de remplissage du tableau.
  — Les marges sur les qualités correspondant à une hiérarchie parfaite construite à partir des marges sur les objets.
  — Les distances : marges observées/hasard pur,
  marges observées/marges de hiérarchie parfaite,
  marges de hiérarchie parfaite/hasard pur,
  les paramètres et histogrammes des distributions des distances marges simulées/hiérarchies parfaites et marges simulées/hasard pur.

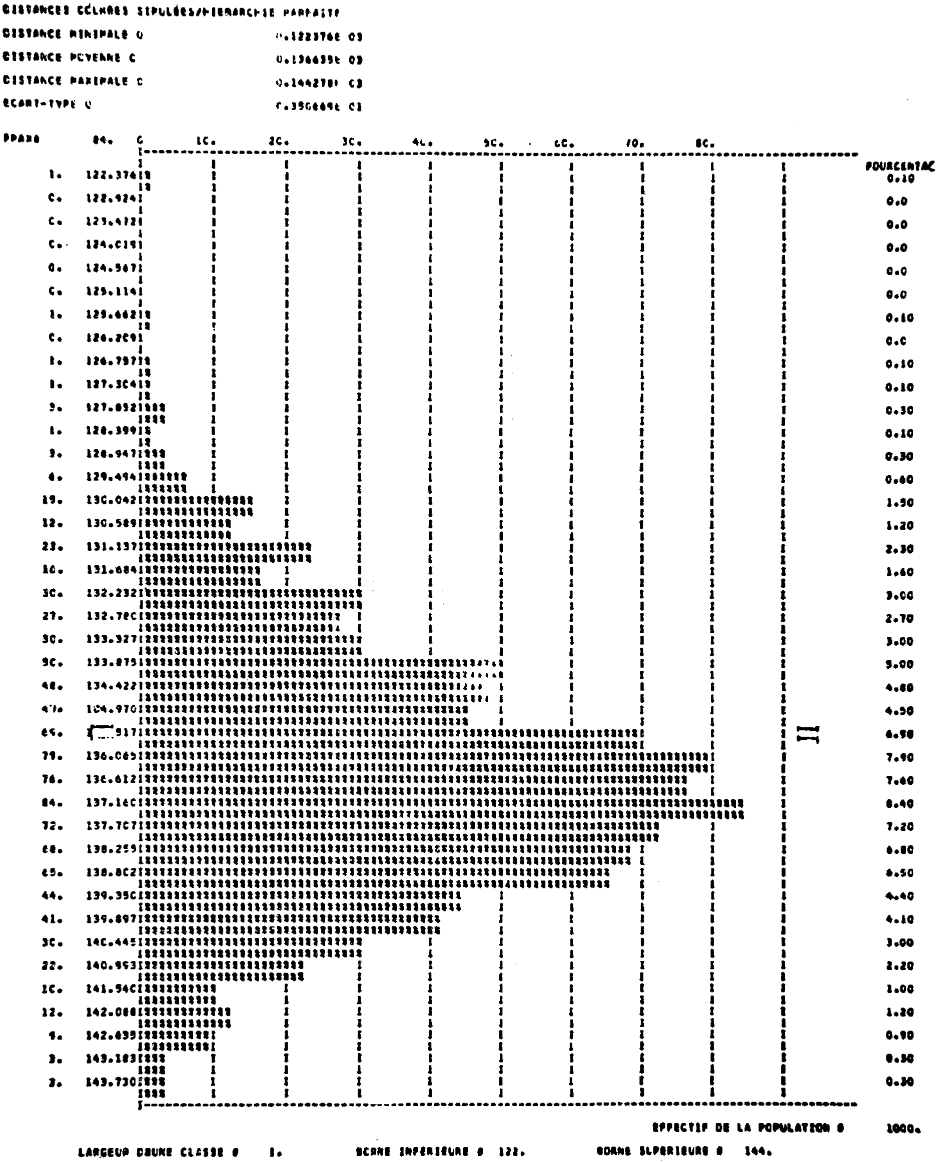
— Les résultats des comparaisons entre les distances.

— Le tableau du nombre de fois où la marge d'une qualité de rang donné a pris une valeur donnée (au cours des simulations). Ce tableau permet d'établir les histogrammes des marges en fonction du rang des qualités.

— Les sorties de l'exemple concret exposé en B.

```
NCMERE C&OBJETS ETLDIES C          57

NOMBRE DE CUALITES ETUDIEES O      59

NOMBRE DE SIMLLATICNS C          1000


LE TABLEAL ETLDIE CCMPCRTE  3363 CASES DCNT  1413 SONT REMPLIES, SOIT UN TAUX DE REMPLISSAGE DE 42.02 POUP CENT.

MCYENNE D&APPARITICN D&UNE CUALITE C      23.9491

MARGES SUR LES QUALITES CORRESPCNCANT A LA HIERARCHIE PARFAITE ASSOCIEE AUX MARGES DONNEES SUR LES OBJETS
     57    57    57    56    55    54    54    54    52    51    49    49    46    46    45    43    42    41    38    37
     34    34    33    31    30    28    28    27    24    24    23    22    17    14    11    11     7     6     5     5
      4     4     3     2     1     1     1     1     1     1     1     0     0     0     0     0     0     0     0


                         DISTANCE CCURBE CBSERVEE/HASARC PUR #                  120.

                         DISTANCE COURBE OBSERVEE/HIERARCHIE PARFAITE #                  49.

                         DISTANCE HIERARCHIE PARFAITE/HASARC PUR #                160.


----- TEST NC 1
POSITION CE LA CCURBE CBSERVEE PAR RAPPORT AUX CISTANCES DES COURBES SIMULEES A LA COURBE DE HIERARCHIE PARFAITE
        LIMITES DE L&INTERVALLE C&ACCEPTATICN CU HASARC CE SEUIL AU PLUS EGAL A 95 POUR CENT O %   130.59,   144.28#
        DISTANCE CE LA COURBE OBSERVEE A LA CCURBE C&HIERARCHIE PARFAITE C    48.74

TEST SIGNIFICATIF C REJET CU HASARC


----- TEST NC 2
POSITICN DE LA CCLRBE CBSERVEE PAR RAPPORT AUX DISTANCES DES COURBES SIMULEES A LA COURBE DE HASARD PUR
        LIMITES CE L&INTERVALLE C&ACCEPTATICN CU HASARC DE SEUIL AU PLUS EGAL A 95 POUR CENT O % O,   32.58#
        DISTANCE CE LA COURBE CBSERVEE A LA CCURBE CE HASARC PUR O   119.87

TEST SIGNIFICATIF O REJET CU HASARC
```


POURCENTAGE
EFFECTIF DE LA POPULATION = 1000.
LARGEUR D'UNE CLASSE = 1.
BORNE INFERIEURE = 17.
BORNE SUPERIEURE = 46.



N.B. Le listing ci-joint contient des symboles qui paraissent erronés et qui en rendent la lecture assez ardue. Ces fautes d'impression proviennent de l'imprimante du centre de calcul où le problème a été traité, cette imprimante ne possédant pas la chaîne de caractères correspondant au code EBCDIC des IBM 360. La table suivante permettra d'établir la correspondance entre les caractères imprimés et les caractères corrects.

au lieu de :	lire :
#	= (égal)
%	((paranthèse ouverte)
□	) (paranthèse fermée)
∂	' (apostrophe)
o	: (deux points)

```
AN IV G LEVEL  18                SIMUL             DATE # 70276          02/01/19

       C****              TEST DE HIERARCHIE          23 JUIN 1970
       C
       C
       C      NOBJ # NOMBRE D∂OBJETS
       C      NQL # NOMBRE DE QUALITES
       C      COURBE NO 1 O COURBE DE LA HIERARCHIE PARFAITE %SE DEDUIT DES
       C      MARGES SUR LES OBJETS□
       C      COURBE NO 2 O COURBE OBSERVEE
       C      COURBE NO 3 O COURBE SIMULEE
       C      COURBE NO 4 O DROITE CORRESPONDANT AU HASARD TOTAL
       C      DISKL O DISTANCE DE LA COURBE K A LA COURBE L
       C      NOAQD%J□ O NOMBRE D∂OBJETS AYANT LA QUALITE J DONNE A PRIORI
       C      NOAQHP O MARGES CORRESPONDANT A UNE HIERARCHIE PARFAITE %COURBE 1□
       C      NOAQ%J□  O NOMBRE D∂OBJETS AYANT LA QUALITE J DANS UNE SIMULATION
       C      NQPO%I□  O NOMBRE DE QUALITES POSSEDEES PAR L∂OBJET I
       C      RQL%NQL□ O RANGS DES QUALITES
       C      ROBJ%NOBJ□ O RANGS DES OBJETS
       C      P%NOBJ*NQL□ O MATRICE DES DONNEES SIMULEE
       C      NSIM # NOMBRE DE SIMULATIONS
       C      NO EST UN NOMBRE PAIR QUELCONQUE DE MOINS DE QUATRE CHIFFRES
       C      NO EST UTILISE DANS LE TIRAGE DES NOMBRES AU HASARD.IL DOIT ETRE
       C      MODIFIE A CHAQUE EXPERIENCE
       C
       C      ISEUIL O POURCENTAGE D∂ERREUR ADMISSIBLE DANS LE REJET DU HASARD
              INTEGER P%60,60□,HIST%60,60□
              DIMENSION NQPO%60□,NOAQ%60□,NOAQD%60□,NOAQHP%60□
              DIMENSION DIS31%5000□,DIS34%5000□
              REAL*8    EPS,RH
              READ%5,1□ NOBJ,NQL,NSIM,ISEUIL,NO
              READ%5,1□ %NQPO%I□,I#1,NOBJ□
              READ%5,1□ %NOAQD%I□,I#1,NQL□
              SEUIL # FLOAT%ISEUIL□
              CSEUIL # 100.-SEUIL
              EPS # DFLOAT%NQL□-1.D-30
              NPT # 0
              DO 170 J#1,NQL
          170 NPT # NPT+NOAQD%J□
              XNPT # FLOAT%NPT□/FLOAT%NQL□
              NPT2 # NOBJ*NQL
              PCTNPT#%FLOAT%NPT□/FLOAT%NPT2□□*100.
              XNPT2 #%XNPT*XNPT□*NQL
              WRITE%6,50□
              WRITE%6,57□ NOBJ,NQL,NSIM
              WRITE%6,80□ NPT2,NPT,PCTNPT
              WRITE%6,81□ XNPT
              DO 260 J#1,NQL
              DO 261 JP#1,NOBJ
          261 HIST%J,JP□#0
          260 CONTINUE
       C
       C      CALCUL DES MARGES ∂QUALITES∂ D∂UNE HIERARCHIE PARFAITE ASSOCIEE AUX
       C      MARGES ∂OBJETS∂ DONNEES
              DO 280 J#1,NQL
          280 NOACHP%J□ # 0
              DO 270 J#1,NQL
              DO 271 JP#1,NOBJ
          271 P%JP,J□ # 0
          270 CONTINUE
```

```
AN IV G LEVEL  19                SIMUL              DATE # 70276          02/01/19
            DO 210 I#1,NCBJ
            MAXJ # NCPO%I¤
            DO 211 J#1,MAXJ
        211 P%I,J¤ # 1
        210 CONTINUE
            DO 220 J#1,NCL
            JP # NCL-J+1
            DO 221 I#1,NCBJ
        221 NCACHP%JP¤ # NCACHP%JP¤+P%I,JP¤
        220 CONTINUE
            WRITE%6,83¤ %NCACHP%J¤,J#1,NCL¤
      C
      C**** SIMULATIONS
      C
            DO 100 LSIM#1,NSIM
      C **  DEBUT D%UNE SIMULATION
      C     REMISE A ZERO DE P
            DO 110 I#1,NCL
            DO 111 J#1,NCBJ
        111 P%J,I¤ # 0
        110 CONTINUE
      C     REMPLISSAGE DE P PAR SIMULATION
            DO 120 I#1,NCBJ
      C     TIRAGE DE NCPO%I¤ NOMBRES AU HASARD COMPRIS ENTRE 1 ET NCL
      C     ET REMPLISSAGE DE P%I,.¤
            N # 0
            IX # %200*I+1¤*%2*LSIM+1¤+NO
       9000 N # N+1
       9001 CALL RANDU%IX,IY,YFL¤
            RH # YFL*EPS + 1.000
            IH # RH
            IX # IY
            IF%P%I,IH¤.EQ.1¤ GOTO 9001
            P%I,IH¤ # 1
            IF%N.LT.NCPO%I¤¤ GOTO 9000
        120 CONTINUE
      C     FIN DU REMPLISSAGE DE P
      C     REORDONNANCEMENT DES COLONNES DE P
      C     CALCUL DES MARGES NCAC%J¤ DES COLONNES
            DO 130 J#1,NCL
            NOAC%J¤ # 0
            DO 131 I#1,NCBJ
        131 NCAC%J¤ # NCAC%J¤+P%I,J¤
        130 CONTINUE
            JM # NCL
            DO 140 J1#1,NCL
            JM#JM-1
            DO 141 J#1,JM
            IF%NOAC%J+1¤.LT.NOAC%J¤¤ GOTO 142
            NTR#NCAC%J+1¤
            NCAC%J+1¤#NCAQ%J¤
            NCAC%J¤#NTR
        142 CONTINUE
        141 CONTINUE
        140 CONTINUE
      C
      C     HIST%J,K¤ O TABLEAUCONTENANT LE NOMBRE DE FOIS OU NCAC%J¤ # K
            DO 150 J#1,NCL
```

```
IV G LEVEL  18                 SIMUL              DATE # 70276           02/01/19
          K # NCAC%J¤
    150   HIST%J,K¤ # HIST%J,K¤+1
C         CALCUL DE LA DISTANCE DE LA COURBE SIMULEE A LA COURBE DE LA HIERAR-
C         CHIE PARFAITE
          DIS31%LSIM¤ # 0.
          DO 160 J#1,NCL
          X # FLOAT%NCAC%J¤-NCACHP%J¤¤
    160   DIS31%LSIM¤ # DIS31%LSIM¤+X*X
          DIS31%LSIM¤#  SQRT%DIS31%LSIM¤¤
C         CALCUL DE LA DISTANCE DE LA COURBE SIMULEE A LA DROITE X#E%NOAQD%J¤¤
          DIS34%LSIM¤ # 0.
          DO 180 J # 1,NCL
          X # FLOAT%NCAC%J¤¤-XNPT
    180   DIS34%LSIM¤ # DIS34%LSIM¤+X*X
          DIS34%LSIM¤#  SQRT%DIS34%LSIM¤¤
C****     FIN D@UNE SIMULATION
    100   CONTINUE
C         CALCUL DE LA DISTANCE DE LA COURBE OBSERVEE A LA COURBE DE LA HIE-
C         RARCHIE PARFAITE
          DIS21 # 0.
          DO 230 J#1,NCL
          X # FLOAT%NCACHP%J¤-NCAQD%J¤¤
    230   DIS21 # DIS21 + X*X
          DIS21 #  SQRT%DIS21¤
C         CALCUL DE LA DISTANCE DE LA COURBE OBSERVEE A LA DROITE DE HASARD PUR
          DIS24 # 0.
          DO 240 J#1,NCL
          X# FLOAT%NCAQD%J¤¤-XNPT
    240   DIS24 # DIS24 + X*X
          DIS24 #  SQRT%DIS24¤
C         CALCUL DE LA DISTANCE DE LA COURBE D@HIERARCHIE PARFAITE
C         A LA COURBE DE HASARD PUR
          DIS14 # 0.
          DO 250 J#1,NCL
          X # FLOAT%NCACHP%J¤¤-XNPT
    250   DIS14 # DIS14 + X*X
          DIS14 #  SQRT%DIS14¤
C         CALCUL DES PARAMETRES DE DIS31 ET DE DIS34
          C31INF # DIS31%1¤
          C31SUP # DIS31%1¤
          C31BAR # 0.
          C31SIG # 0.
          DO 190 L#1,NSIM
          IF%C31INF.GT.DIS31%L¤¤  C31INF # DIS31%L¤
          IF%C31SUP.LT.DIS31%L¤¤  C31SUP # DIS31%L¤
          C31SIG # DIS31%L¤*DIS31%L¤+C31SIG
    190   C31BAR # C31BAR+DIS31%L¤
          C31BAR # C31BAR/FLOAT%NSIM¤
          C31SIG # C31SIG/FLOAT%NSIM¤-C31BAR*C31BAR
          C31SIG # SQRT%C31SIG¤
          C34INF # DIS34%1¤
          C34SUP # DIS34%1¤
          C34BAR # 0.
          C34SIG # 0.
          DO 200 L#1,NSIM
          IF%C34INF.GT.DIS34%L¤¤ C34INF # DIS34%L¤
          IF%C34SUP.LT.DIS34%L¤¤ C34SUP # DIS34%L¤
          C34SIG # DIS34%L¤*DIS34%L¤+C34SIG
```

```
NN IV G LEVEL  18                    SIMUL              DATE # 70276          02/01/19
           200 C34BAR # C34BAR + DIS34%L¤
               C34BAR # C34BAR/FLOAT%NSIM¤
               C34SIG # C34SIG/FLOAT%NSIM¤-C34BAR*C34BAR
               C34SIG # SQRT%C34SIG¤
               WRITE%6,50¤
               WRITE%6,51¤
               WRITE%6,52¤ C31INF,C31BAR,C31SUP,C31SIG
               CALL HISTOG%DIS31,C31SUP,C31INF,S31,SEUIL,NSIM,0¤
               WRITE%6,53¤
               WRITE%6,52¤ C34INF,C34BAR,C34SUP,C34SIG
               CALL HISTOG%DIS34,C34SUP,C34INF,S34,SEUIL,NSIM,1¤
               WRITE%6,55¤ DIS24
               WRITE%6,54¤ DIS21
               WRITE%6,56¤ DIS14
               WRITE%6,86¤
               DO 400 JO#1,NQL
               WRITE%6,87¤ JQ,%HIST%JQ,JO¤,JO#1,NOBJ¤
           400 CONTINUE
               ISEUIL # 100-ISEUIL
               WRITE%6,62¤
               WRITE%6,63¤ ISEUIL,S31,C31SUP,DIS21
               IF%S31-DIS21¤1030,1030,1031
          1030 WRITE%6,64¤
               GO TO 1032
          1031 WRITE%6,65¤
          1032 CONTINUE
               WRITE%6,66¤
               WRITE%6,67¤ ISEUIL,S34,DIS24
               IF%S34-DIS24¤ 1040,1041,1041
          1040 WRITE%6,65¤
               GOTO 1042
          1041 WRITE%6,64¤
          1042 CONTINUE
             1 FORMAT%20I4¤
            50 FORMAT%1H1¤
            51 FORMAT%a1DISTANCES COURBES SIMULEES/HIERARCHIE PARFAITEa¤
            52 FORMAT%a0DISTANCE MINIMALE 0a,10X,E20.6/a0DISTANCE MOYENNE 0a,11X,
              1E20.6/a0DISTANCE MAXIMALE 0a,10X,E20.6/a0ECART-TYPE 0a,17X,E20.6/¤
            53 FORMAT%a0DISTANCES COURBES SIMULEES/HASARD PURa¤
            54 FORMAT%a0DISTANCE COURBE OBSERVEE/HIERARCHIE PARFAITE #a,F20.0¤
            55 FORMAT%a0DISTANCE COURBE OBSERVEE/HASARD PUR #a,F20.0¤
            56 FORMAT%a0DISTANCE HIERARCHIE PARFAITE/HASARD PUR #a,F20.0¤
            57 FORMAT%27HONOMBRE DaOBJETS ETUDIES 0 ,4X,I6/
              130HONOMBRE DE QUALITES ETUDIEES 0 ,I6/
              224HONOMBRE DE SIMULATIONS 0,6X,I6///¤
            60 FORMAT%a1DISTANCES C.OBS/C.SIMa/%1X,10F12.6¤¤
            61 FORMAT%a0DISTANCES C.SIM/ESPERa/%1X,10F12.6¤¤
            62 FORMAT%a1-----  TEST NO 1a/a0POSITION DE LA COURBE OBSERVEE PARa,
              1a RAPPORT AUX DISTANCES DES COURBES SIMULEES A LA COURBE DEa,
              2a HIERARCHIE PARFAITEa¤
            63 FORMAT% 10X,aLIMITES DE LaaINTERVALLE DaaACCEPTATION DU HASARDa,
              1a DE SEUIL AU PLUS EGAL Aa,I3  ,a POUR CENT 0 %a,F9.2,a,a,F9.2,
              2a¤a/10X,aDISTANCE DE LA COURBE OBSERVEE A LA COURBE DaaHIERARa
              3aCHIE PARFAITE 0a,F9.2¤
            64 FORMAT% a0TEST NON SIGNIFICATIFa///¤
            65 FORMAT% a0TEST SIGNIFICATIF 0 REJET DU HASARDa///¤
            66 FORMAT%a0-----  TEST NO 2a/a0POSITION DE LA COURBE OBSERVEE PARa,
              1a RAPPORT AUX DISTANCES DES COURBES SIMULEES A LA COURBE DEa,
```

```
AN IV G LEVEL  18                 SIMUL               DATE # 70276           02/01/19

            2a HASARC PURa¤
         67 FORMAT%10X,aLIMITES DE Laa INTERVALLE Daa ACCEPTATION DU HASARDa,
           1a DE SEUIL AU PLUS EGAL Aa,I3  ,a POUR CENT 0 % 0,a,F8.2,a¤a/
           2 10X,aDISTANCE DE LA COURBE OBSERVEE A LA COURBE DE HASARD PUR 0a,
           3F9.2¤
         80 FORMAT%aOLE TABLEAU ETUDIE COMPORTEa,I6,a CASES DONTa,I6,a SONT RE
           1MPLIES, SOIT UN TAUX DE REMPLISSAGE DEa,F6.2,a POUR CENT.a ¤
         81 FORMAT%aOMOYENNE Daa APPARITION Daa UNE QUALITE 0a,F12.4¤
         83 FORMAT%aOMARGES SUR LES QUALITES CORRESPONDANT A LA HIERARCHIE PAR
           1FAITE ASSOCIEE AUX MARGES DONNEES SUR LES OBJETSa/%1X,20I6¤¤
         86 FORMAT%a1FREQUENCES DES MARGES EN FONCTION DU RANGa/a RANGa¤
         87 FORMAT%a0a,I4,10I12/%5X,10I12¤¤
            STOP
            END
```

```
AN IV G LEVEL  18                 RANDU               DATE # 70276           02/01/19

            SUBROUTINE RANDU%IX,IY,YFL¤
            IY#IX*65539
            IF%IY¤5,6,6
          5 IY#IY+2147483647+1
          6 YFL#IY
            YFL#YFL*.4656613E-9
            RETURN
            END
```

```
AN IV G LEVEL  18                HISTOG           DATE # 70276          02/01/19

      .  SUBROUTINE HISTOG%X,XSUP,XINF,XS,SEUIL,IMAX,ISL¤
C  DIMENSIONNER P,CG,PCT,D A NC#NMB DE CLASSES
C  KMAX#NMB DE LIGNES PAR CLASSE-1
         DATA LEC,IMP/5,6/
         DIMENSION X%1¤
         DIMENSION P%50¤,ICOLON%100¤,NCIR%100¤,CG%50¤,PCT%50¤,D%50¤
         KMAX # 1
         PAS#0.
         CD # XSUP
         DATA IREMP/1H%/
         DATA JBLANC/1H /,KOLONE/1HI/
         DO 20 I#1,100
         NOIR%I¤ # IREMP
      20 CONTINUE
         DO 21 I#1,10
         I10 # I*10
         I1 # %I-1¤*10
         DO 22 I2 #1,9
         I3 # I1+I2
         ICOLON%I3¤ # JBLANC
      22 CONTINUE
         ICOLON%I10¤ # KOLONE
      21 CONTINUE
     200 NC # 40
         PAS # XSUP-XINF+%XSUP+XINF¤*1.E-6
         PAS # PAS/FLOAT%NC¤
         NC10#NC+1
         CG%1¤ # XINF
         DO 500 IJ#2,NC10
     500 CG%IJ¤ # CG%IJ-1¤+PAS
         DO 501 IJ #1,NC10
     501 P%IJ¤ # 0.
         DO 502 IJ#1,IMAX
         XNL # %X%IJ¤-XINF¤/PAS + 1.
         NL # XNL
     502 P%NL¤ # P%NL¤+ 1.
         TP # 0.
         DO 26 N#1,NC
      26 TP # TP+P%N¤
         DO 27 N#1,NC
         PCT%N¤ # P%N¤*100./TP
      27 CONTINUE
         PMAX # P%1¤
         DO 2 N#2,NC
         IF%PMAX.LT.P%N¤¤  PMAX#P%N¤
       2 CONTINUE
         M2 # 0
       4 IF%%10.¤**M2.LE.PMAX.AND.PMAX.LT.%10.¤**%M2+1¤¤ GOTO 3
         M2 # M2+1
         GO TO 4
       3 PDIX # %10.¤**M2
         C # 0.
       5 IF%C*PDIX.LE.PMAX.AND.PMAX.LT.%C+1¤*PDIX¤  GOTO 6
         C # C+1.
         GOTO 5
       6 NCOEF # C
         C1 # C*PDIX
         GOTO %7,8,9,92,10,10,10,10,10¤,NCOEF
```

```
AN IV G LEVEL  18                HISTOG           DATE # 70276          02/01/19
             7 NCAD#PMAX/%PDIX/5.¤
               NCAD#%NCAD+1¤*10
               PCAD#50./PDIX
               WRITE%6,71¤ C1,PMAX
               GOTO11
             8 NCAD#60
               PCAD#20./PDIX
               C2 # %C-1¤*PDIX
               WRITE%6,81¤ C2,C1,PMAX
               GOTO 11
             9 NCAD#80
               PCAD#20./PDIX
               C2 # %C-2.¤*PDIX
               C3 # %C-1.¤*PDIX
               WRITE%6,91¤ C2,C3,C1,PMAX
               GOTO 11
            92 NCAD#100
               PCAD#20./PDIX
               C2 # %C-3.¤*PDIX
               C3 # %C-2.¤*PDIX
               C4 # %C-1.¤*PDIX
               WRITE%IMP,921¤C2,C3,C4,C1,PMAX
               GO TO 11
            10 NCAD#%NCOEF+1¤*10
               PCAD#10./PDIX
               DO 23 I#1,10
               Q # I
            23 D%I¤ # Q*PDIX
               WRITE%IMP,101¤ PMAX,%D%I¤,I#1,NCOEF¤
               GOTO 11
            11 WRITE%IMP,111¤
               WRITE%IMP,112¤ %ICOLON%J¤,J#1,NCAD¤
               WRITE%IMP,116¤
               DO 24 N#1,NC
               LC#P%N¤*PCAD
               LC1 # LC+1
               IF%LC1.EQ.1¤  GOTO 14
               WRITE%IMP,113¤P%N¤,CG%N¤,%NOIR%I¤,I#1,LC¤,%ICOLON%J¤,J#LC1,NCAD¤
               GOTO 16
            14 WRITE%IMP,113¤P%N¤,CG%N¤,%ICOLON%J¤,J#LC1,NCAD¤
            16 WRITE%IMP,271¤ PCT%N¤
               IF%LC1.EQ.1¤  GOTO 15
               DO 25 K#1,KMAX
               WRITE%6;114¤ %NOIR%I¤,I#1,LC¤,%ICOLON%J¤,J#LC1,NCAD¤
            25 CONTINUE
               GOTO 17
            15 CONTINUE
               DO 29 K#1,KMAX
               WRITE%IMP,115¤ %ICOLON%J¤,J#LC1,NCAD¤
            29 CONTINUE
            17 CONTINUE
            24 CONTINUE
               WRITE%6,111¤
               WRITE%6,12¤ TP
               WRITE%6,13¤ PAS,CG%1¤,CD
           202 CONTINUE
         C     CALCUL DES SEUILS D%ACCEPTATION
         C     CALCUL SUR DIS31 SI ISL#0
```

```
N IV G LEVEL  18                  HISTOG              DATE # 70276            02/01/19
       C      CALCUL SUR DIS34 SI ISL#1
       C      SEUIL O POURCENTAGE D@ACCEPTATION
              PCT2 # O.
              N#O
              IF%ISL¤ 1000,1000,1001
        1000  CONTINUE
        1010  N#N+1
              PCT2 # PCT2+PCT%N¤
              IF%PCT2-SEUIL¤ 1010,1012,1012
        1012  IF%N-1¤1013,1014,1013
        1014  N#N+1
        1013  XS # CG%N-1¤
              GOTO 1100
        1001  N # NC+1
        1020  N # N-1
              PCT2 # PCT2+PCT%N¤
              IF%PCT2-SEUIL¤ 1020,1022,1022
        1022  IF%N-NC¤ 1023,1024,1023
        1024  N # N-1
        1023  XS # CG%N+1¤
        1100  CONTINUE
              WRITE%6,999¤
           1  FORMAT%3F7.0,53X,F6.0¤
          12  FORMAT%1H0,90X,28HEFFECTIF DE LA POPULATION # ,F10.0¤
          13  FORMAT%1H0,10X,22HLARGEUR D@UNE CLASSE # ,F6.0,10X,18HBORNE INFERI
             1EURE # ,F6.0,10X,19HBORNE SUPERIEURE # ,F6.0¤
          71  FORMAT%19X,1H0,9X,1H2,9X,1H4,9X,1H6,9X,1H8,7X,F6.0,5X,2H12,8X,2H14
             1,8X,2H16,8X,2H18,5X,5HPMAX#,F6.0¤
          81  FORMAT%19X,1H0,17X,F6.0,14X,F6.0,5X,5HPMAX#,F6.0¤
          91  FORMAT%19X,1H0,17X,F6.0,14X,F6.0,14X,F6.0,5X,5HPMAX#,F6.0¤
         101  FORMAT%1H0,5HPMAX#,F10.0,3X,1H0,2X,10%F8.0,2X¤¤
         111  FORMAT%19X,1HI,20%5H-----¤¤
         112  FORMAT%19X,1HI,100A1¤
         113  FORMAT%1H ,F8.0,1X,F9.3,1HI,100A1¤
         114  FORMAT%19X,1HI,100A1¤
         115  FORMAT%19X,1HI,100A1¤
         116  FORMAT%1H+,120X,11HPOURCENTAGE¤
         203  FORMAT%1H1,24HHISTOGRAMME DU CARACTERE  ,I3¤
         271  FORMAT%1H+,122X,F5.2¤
         921  FORMAT%19X,1H0,3X,4%14X,F6.0¤,5X,5HPMAX#,F6.0¤
         999  FORMAT%1H1,3X¤
              RETURN
              END
```

BIBLIOGRAPHIE SOMMAIRE

J. P. BENZECRI, *Analyse des Données*. Cours professés à la Faculté des Sciences de Paris. Publications de l'I.S.U.P., 1969.

L. LEBARD et C. DENIAU, « Introduction à l'analyse factorielle », *Consommation*, vol. 15, n° 3, juillet 1969, p. 57-96.

L. LEBARD, « Introduction à l'analyse des données », *Consommation*, vol. 15, n° 4, octobre 1969, p. 65-86.

J. C. LERMAN, *Les bases de la classification automatique*, Gauthier-Villars, 1970.

B. MATALON, *L'analyse hiérarchique*, Gauthier-Villars, 1965.

D. F. MORRISON, *Multivariate Statistical Methods*, McGraw-Hill, 1967.

M. ROUX, *Algorithme pour construire une hiérarchie particulière*. Thèse de 3e cycle publiée par l'I.S.U.P., 1968.