

CAHIER DE ReCHERCHE

M A I 9 1

■ N° 16

CLASSIFICATION DICHOTOMIQUE DESCENDANTE



Sébastien Lion

CREDOC



CAHIER DE RECHERCHE

CLASSIFICATION DICHOTOMIQUE DESCENDANTE

Sébastien LION

Cette recherche a été réalisée dans le Département de Prospective de la Consommation, dirigé par Saadi Lahlou.

Mai 1991

SOMMAIRE

INTRODUCTION GENERALE	1
I - POSITION DU PROBLEME.....	3
I.1. Hypothèses de base	3
I.2. Classification de premier rang et itération du processus	3
I.3. Interprétation des résultats	6
I.4. Algorithme de calcul.....	7
II - ENRICHISSEMENT DU MODELE.....	9
II.1. Nature des variables caractéristiques.....	9
II.1.1. Calcul d'un indicateur d'écart des moyennes global	9
II.1.2. La transformation des modalités en variables binaires	12
Construction d'un tableau logique.....	12
Bimodalisation par la variance expliquée	13
II.2. Construction du test.....	14
II.2.1. Tester la significativité de la coupure en deux groupes	14
II.2.2. Garder des sous-groupes de taille significative	16
II.2.3. Construire un test de sélection optimal.....	16
Construction d'un test local	16
Hypothèse d'un test global	17
III - FONDEMENTS SOCIOLOGIQUES DU MODELE.....	19
III.1. Prospective et comportement des individus	19
III.2. Classification Descendante Dichotomique et Formalisation en Relativité Complète	21

IV - APPLICATION A LA GRANDE DISTRIBUTION.....	23
IV.1. Excédent brut d'exploitation / (Chiffre d'affaires + subventions).....	24
IV.1.1. Description analytique de la construction de l'arborescence	24
IV.1.2. Utilisation de l'arbre par un banquier.....	28
IV.2. Excédent brut d'exploitation / (C A + subventions) sans les variables comptables	29
 CONCLUSION.....	 31
 BIBLIOGRAPHIE.....	 32
 ANNEXES.....	 33

INTRODUCTION GENERALE

La méthode proposée résulte d'un travail mené au Crédoc conjointement par Philippe Moati et Sébastien Lion, du département Prospective de la consommation dirigé par Saadi Lahlou, dans le cadre d'une étude sur la grande distribution. Ce travail a bénéficié de discussions avec François Gardes (Crédoc, ENS Cachan). Par ailleurs, l'optimisation informatique de l'algorithme doit beaucoup à Lé Cong, responsable du centre informatique du Crédoc, qui a permis l'application pratique du modèle théorique.

Cette méthode a été mise au point à l'origine pour construire un outil d'aide à la décision financière facilement intelligible pour un non-statisticien. En l'occurrence, il s'agissait de porter un jugement sur la qualité de la gestion d'une grande surface, compte tenu des caractéristiques objectives de celle-ci : surface, nombre de salariés, nombre de caisses, etc...

Pour cela, nous disposons au départ d'un fichier comportant des variables continues et des variables discrètes bimodales (oui ou non) pour 439 supermarchés après suppression des individus mal renseignés. Parmi ces variables, certaines sont financières (taux de marge, taux de profit, ...), d'autres correspondent à la description physique des grandes surfaces (parking, surface, ...).

Ainsi, nous considérons la valeur d'un indicateur financier pour un échantillon significatif de grandes surfaces et nous cherchons à classer les gestionnaires en «bons» ou «mauvais» selon les caractéristiques matérielles des points de vente dont ils ont la charge. Cette classification est dichotomique. En effet, notre échantillon est partitionné en deux groupes «bons» et «mauvais» relativement à une variable caractéristique. Ces deux groupes seront à leur tour partitionnés en «bons» et «mauvais» relativement à une seconde variable caractéristique, etc....

Cependant, l'outil créé nous a paru avoir un intérêt plus général, car il est applicable aux cas où l'on cherche à déterminer des variables explicatives sans avoir à postuler que leur caractère explicatif est identique pour tout l'échantillon. La méthode est conçue pour être appliquée dans le cadre général défini par la formalisation en relativité complète (FRC), qui considère qu'il n'existe que des lois locales essentiellement liées à la définition de leur

domaine d'application. La méthode est donc faite précisément pour travailler sur des échantillons d'objets ayant des modèles de fonctionnement hétérogènes, et vise à désamorcer certains problèmes d'agrégation qui entravent la modélisation économétrique.

Après avoir modélisé le problème de base dans sa formulation la plus simple (partie I), nous étudierons les enrichissements possibles et les limites du modèle initial (partie II). Nous montrerons alors comment la méthode s'insère dans une conception particulière du réel. Enfin, nous présenterons un exemple d'application, celui-là même qui fut le support premier de la méthode (partie III).

I- POSITION DU PROBLEME

L'objet de cette partie est de présenter la méthode de la manière la plus générale possible. Nous serons donc amenés à simplifier au maximum les hypothèses initiales de notre modèle. La seconde partie permettra de revenir sur les enrichissements possibles du modèle, une fois que le processus aura été bien compris.

I.1. Hypothèses de base

Nous disposons d'une population de n individus dont on connaît la valeur pour $k+1$ variables que nous considérerons tout d'abord comme exclusivement continues : 1 variable principale X (ou endogène) et k variables caractéristiques Y_i (ou explicatives). La variable principale est celle qui servira à juger si les individus sont «bons» ou «mauvais». Elle est donc notre variable de décision.

Les k variables caractéristiques correspondent à des réalités objectives et mesurables : taille, poids, surface, effectifs, etc... Nous supposons bien évidemment qu'il existe une relation théorique entre les variables retenues et la variable principale. En revanche, nous n'avons aucune exigence a priori sur le degré de corrélation entre les variables caractéristiques et la variable principale. Pour simplifier, il importe seulement de supposer qu'il y ait un lien statistique, sans qu'il soit pour autant nécessaire de le quantifier. Notre méthode est avant tout exploratoire.

I.2. Classification de premier rang et itération du processus

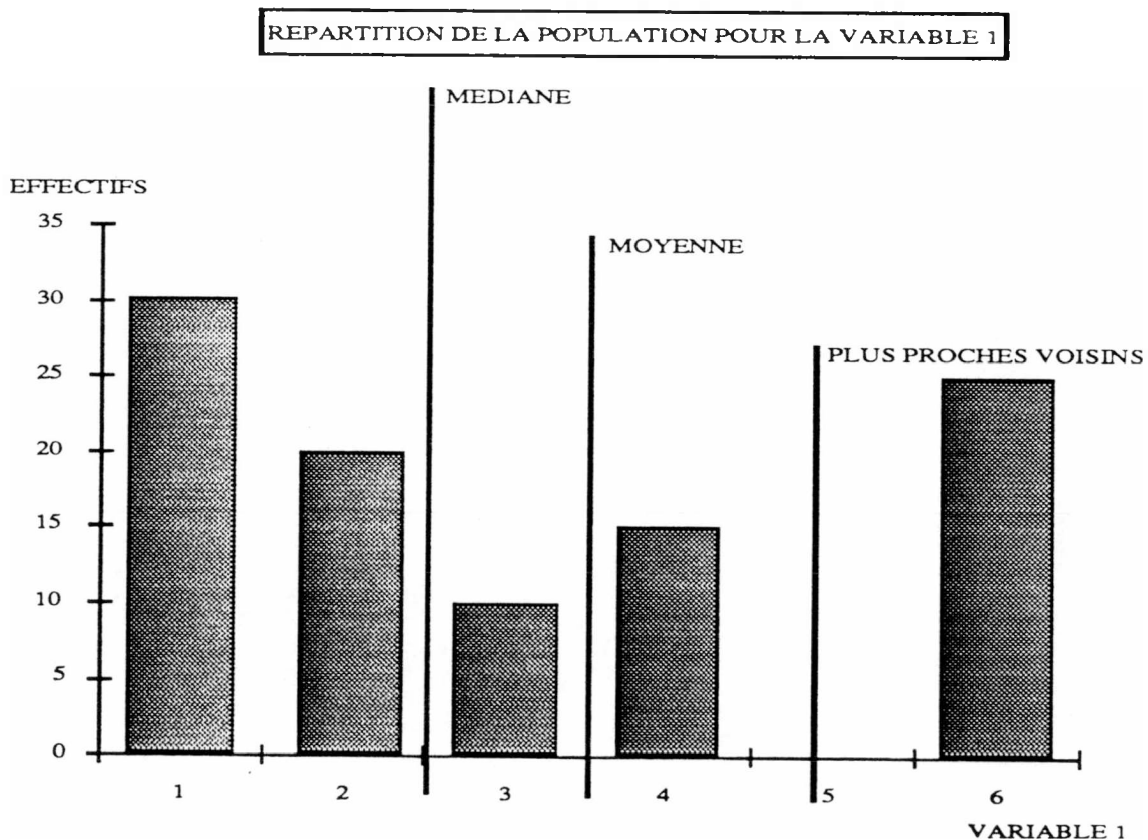
Nous considérons la répartition de notre population pour une variable explicative i . Nous cherchons à la scinder en deux groupes les plus homogènes possibles. Dans un espace de dimension élevée, plusieurs procédures sont possibles.

Nous pouvons faire glisser un hyperplan sur le premier axe factoriel tel qu'il isole deux groupes pour lesquels le rapport variance intergroupes sur variance intragroupes soit

maximal. Nous nous situons alors dans une logique de classification descendante. Nous pouvons également réaliser une classification avec la méthode des plus proches voisins selon une logique ascendante.

Quelle que soit la logique adoptée, le but est simple : isoler les deux groupes les plus homogènes pour la variable explicative considérée. Le fait de ne retenir ici qu'une seule variable rend en fait trivial le problème puisque cette analyse est univariée. Pourquoi alors ne pas recourir à des indicateurs de centralité de type médiane ou moyenne ?

La réponse peut être illustrée graphiquement. La moyenne ou la médiane risquent en fait de partitionner la population en deux groupes hétérogènes. Considérons la répartition suivante (que nous avons discrétisée par souci de clarté) et les groupes issus du calcul de la médiane, de la moyenne ou de la méthode des plus proches voisins. Les droites verticales tracées en gras séparent ici la population en deux groupes.



Le critère de la moyenne conduit à considérer les deux groupes [1, 3] et [4, 6].

Le critère de la médiane conduit à considérer les deux groupes [1, 2] et [3, 6].

Le critère des plus proches voisins conduit à considérer les deux groupes [1, 4] et [5].

Seul le critère des plus proches voisins nous permet ici d'isoler les deux populations les plus homogènes. Nous l'adopterons pour la suite de notre analyse.

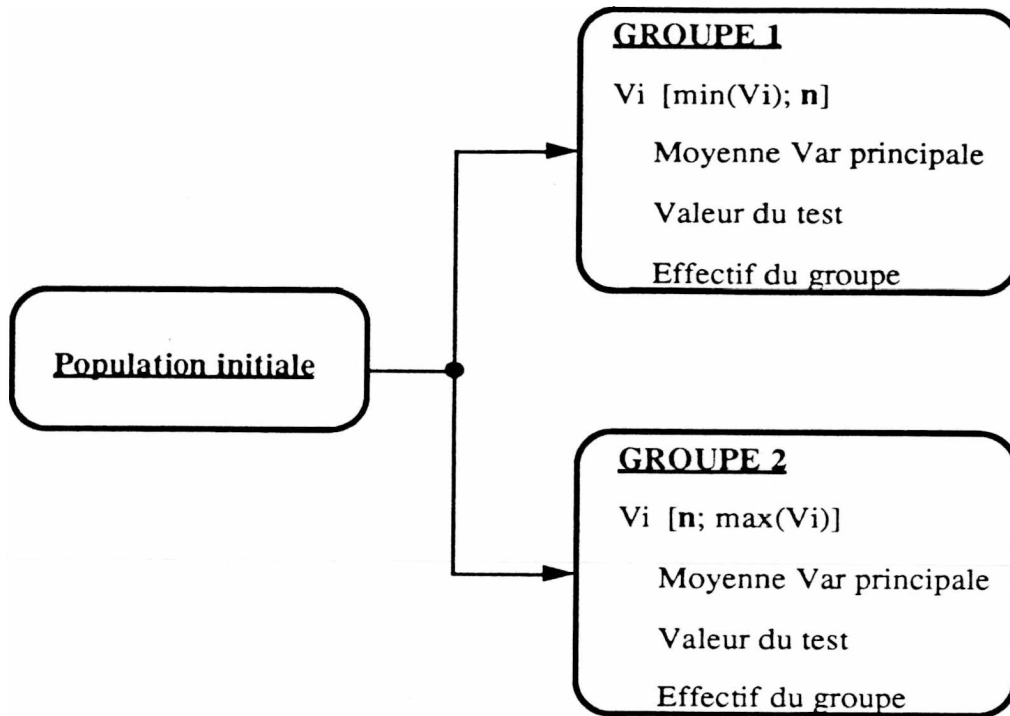
Lorsque ces groupes sont définis, nous pouvons pour chacun d'eux calculer la moyenne et l'écart-type de la variable principale. Munis de ces indicateurs de centralité et de dispersion ainsi que des effectifs de chacun des groupes, nous pouvons alors calculer un test d'écart de moyenne *relativement à la variable principale*, dont la première utilité sera bien évidemment de voir si la partition en deux classes est significativement discriminante pour la variable qui nous intéresse. Nous attirons l'attention du lecteur sur le fait que le test d'écart des moyennes porte sur une autre variable que celle qui a servi à séparer notre population en deux groupes.

Nous réitérons alors le même processus pour chacune des k variables explicatives. A l'issue de cette première étape, nous obtenons alors *pour chacune de ces variables* :

- deux sous-groupes de notre population initiale avec leurs effectifs respectifs ;
- les extrema des deux groupes pour la variable caractéristique Y_i ;
- la moyenne et l'écart-type de la variable principale X pour chacun des deux groupes ;
- la valeur d'un test d'écart de moyenne relatif à la variable principale X .

Il est donc possible de trier les variables explicatives en fonction de ce dernier test. La variable pour laquelle ce test est le plus élevé est donc celle qui discrimine au mieux la population initiale relativement à la variable principale.

Nous retiendrons alors une première variable explicative et la partition en deux groupes qui résulte de ce choix. Ainsi se dessine le premier nœud de notre arborescence. Par exemple, nous aurons :



où n est la valeur de la variable i qui scinde la population initiale en deux et V_i la i ème variable explicative.

Pour chacun des deux groupes, nous réitérons le processus jusqu'à ce que la taille du groupe ou que la valeur du test d'écart de moyenne ne soient plus significatives (ces seuils peuvent être plus ou moins stricts selon les exigences de l'utilisateur).

1.3. Interprétation des résultats

En fin d'analyse nous obtenons donc un arbre. Le premier nœud correspond, nous venons de le voir, à la variable caractéristique la plus discriminante relativement à la variable principale. Nous pouvons donc classer notre population initiale en «bons» ou «mauvais» selon différents degrés de précision qui correspondent aux nœuds successifs de notre arbre.

Cette méthode permet selon nous de rendre compte au mieux des logiques qui sous-tendent les décisions des agents, leurs résultats ou leur motivations. En effet, cette méthode repose sur des facteurs explicatifs locaux et hiérarchisés. La logique d'un groupe en fin d'arborescence dépend d'un enchaînement de conditions sur les variables successivement sélectionnées. Ces dernières constitueront l'ensemble des conditions déterminantes pour porter un jugement sur la valeur de la variable principale dans les groupes finaux.

L'intérêt de la méthode repose sur le caractère local de la classification. En effet, pour chacun des deux premiers groupes sélectionnés par le choix d'une variable caractéristique (celle qui aura satisfait au test précédemment défini), nous allons réitérer le processus de choix. La variable qui sera choisie alors pour scinder le premier des groupes peut tout à fait être différente de celle qui sera choisie pour le second des groupes.

Par exemple, supposons que nous nous intéressions à la pression artérielle chez les mammifères. S'il ressort, ce qui est une hypothèse gratuite, que la variable caractéristique qui permet de séparer au mieux notre population en deux soit la taille, nous aurons donc les petits et les grands qui formeront deux classes. Si nous considérons alors les petits mammifères, la seconde variable pourrait être la taille du cœur. Si au contraire nous nous intéressons aux grands mammifères, la seconde variable pourrait être la distance du cœur au cerveau.

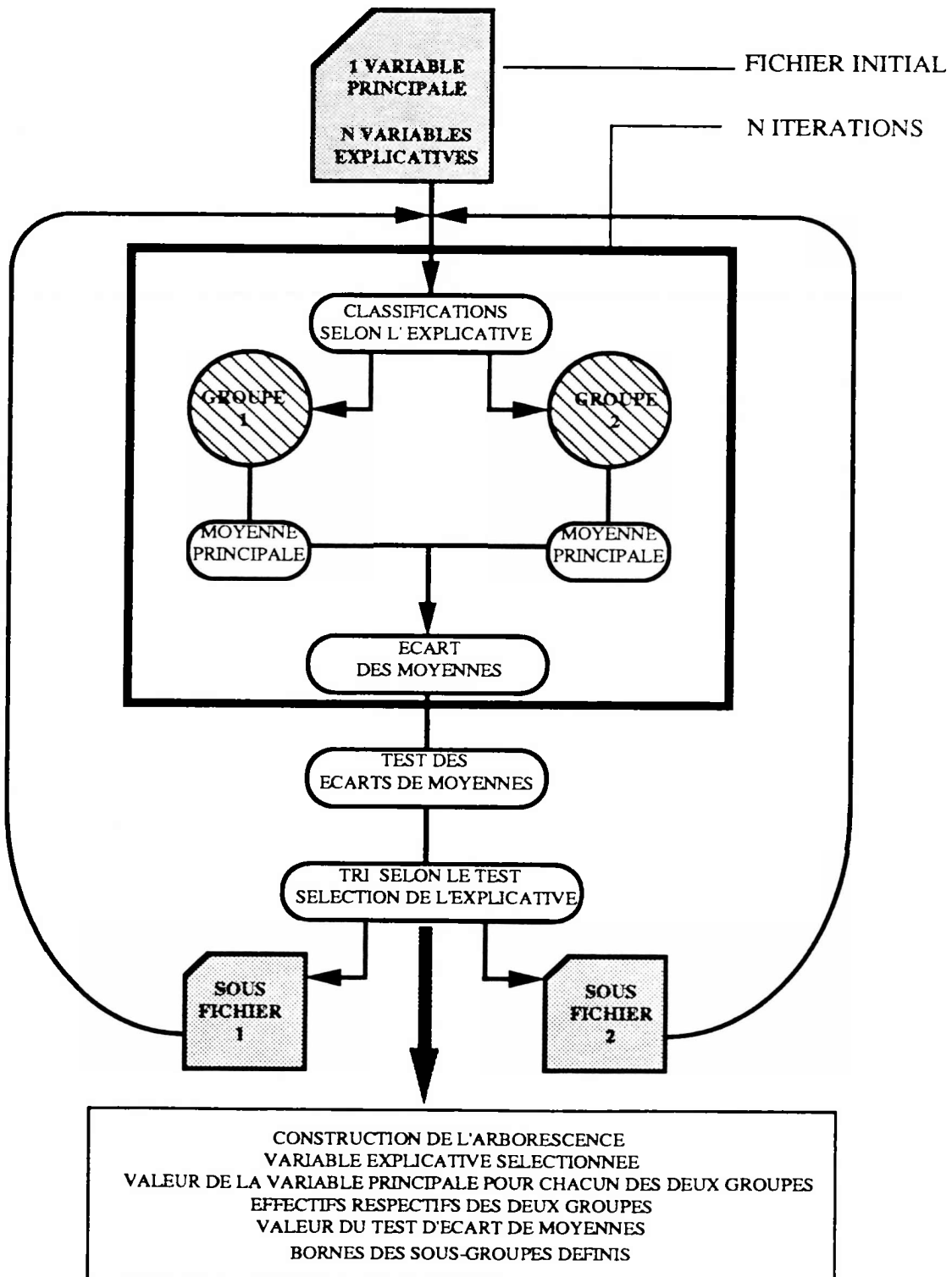
Les variables sélectionnées peuvent donc être différentes pour chacun des sous-groupes. Nous obtenons donc, pour chaque groupe final, la liste des variables hiérarchisées qui définissent celui-ci. Dès lors, il existe une et une seule affectation possible d'un individu à un groupe. L'ensemble des groupes finaux constitue alors une partition de l'échantillon de départ.

Un nouvel individu sera donc classé dans un de ces groupes finaux en fonction de réponses à des questions successives portant sur les variables de l'arbre. Pour reprendre notre exemple sur les mammifères, une girafe sera classée d'abord parmi les grands mammifères (réponse à la question : «quelle est sa taille ? ») puis parmi les mammifères dont la distance du cœur au cerveau est supérieure à 2 mètres (réponse à la question : «quelle est la distance du cœur au cerveau ? »).

I.4. Algorithme de calcul

Le schéma suivant vise à donner une vision globale de la méthode pour faciliter sa compréhension. Le souci de synthétiser la méthode nous a conduit à la simplifier au maximum, notamment en ce qui concerne l'ensemble des différents tests qui déterminent le critère de choix de la meilleure variable explicative locale.

Algorithme de construction de l'arborescence



II - ENRICHISSEMENT DU MODELE

II.1. Nature des variables caractéristiques

Nous avons exposé la méthode en considérant des variables continues. Cette contrainte provient du caractère dichotomique du processus. Par extension, nous pouvons admettre les variables bimodales : oui/non, homme/femme ou pour/contre. Ceci constitue une limite en l'état de nos recherches. La plupart des variables discrètes comportent un nombre de modalités supérieur à deux. La simple présence des modalités «ne sait pas» ou «autres» suffit souvent à exclure la variable du cadre d'analyse.

La généralisation de notre modèle à ces variables discrètes nécessite de concevoir une procédure qui ne soit plus dichotomique. Le problème qui se pose alors est celui du choix de la variable caractéristique qui permet d'obtenir les groupes les plus distincts relativement aux valeurs de la variable principale.

Dans le cas de deux groupes, un test d'écart de moyennes pondéré par la somme des écarts-types suffit pour notre tri des différentes variables explicatives. En revanche, dès que le nombre de groupes dépasse deux, nous sommes privés de cet outil. Il faudrait dès lors recourir à d'autres techniques. Nous proposons quelques voies de recherche possible.

La généralisation du modèle peut consister en l'élaboration d'un critère de sélection de la variable explicative qui permette de considérer trois modalités ou plus. Elle peut également conserver son caractère proprement dichotomique par une transformation du fichier de base en variables bimodales.

II.1.1. Calcul d'un indicateur d'écart des moyennes global

Pour chaque modalité, on calcule l'écart de moyennes de la variable principale pondéré par la somme des écarts types avec l'ensemble des autres modalités. En effet, cette pondération permet de prendre en compte la dispersion autour de la moyenne pour chacun des deux groupes. Deux groupes avec des moyennes peu éloignées mais avec une très faible dispersion seront mieux discriminés que deux autres groupes avec des moyennes peut-être plus éloignées l'une de l'autre mais avec une dispersion plus forte.

On choisit alors le plus faible de ces écarts de moyennes. Pour n modalités, nous aurons alors n valeurs de cet écart de moyennes minimal. Il suffit alors d'en prendre la moyenne pour obtenir notre indicateur de tri.

Par exemple, nous disposons de deux variables A et B.

A possède 3 modalités A1, A2, A3.

B possède 4 modalités B1, B2, B3, B4.

Le calcul des moyennes et des écarts-types de la variable principale pour chacune des modalités est le suivant :

	A1	A2	A3	B1	B2	B3	B4
Moyenne	2	4	4,5	3	7	1	6
Ecart-type	0,5	0,8	1,2	0,2	2	0,1	1,5

Nous noterons : $A1/A2$ l'écart des moyennes pondéré par la somme des écart-types pour les modalités A1 et A2.

$A1(-)$ la valeur minimale de cet indicateur pour les couples formés avec A1.

$A(.)$ la valeur moyenne des $Ai(-)$.

Nous obtenons : $A1/A2 = 1,54$

$A1/A3 = 1,47$

$A2/A3 = 0,25$

$B1/B2 = 1,82$

$B1/B3 = 6,67$

$B1/B4 = 1,76$

$B2/B3 = 2,86$

$B2/B4 = 0,285$

$B3/B4 = 3,125$

$A1(-) = 1,47$

$A2(-) = 0,25$

$A3(-) = 0,25$

$B1(-) = 1,76$

$B2(-) = 0,285$

$B3(-) = 2,86$

$B4(-) = 0,285$

$A(.) = 0,66$

$B(.) = 1,3$

Nous retiendrons dans ce cas la variable B qui scinde notre population en 4 sous-groupes. Cet exemple ne prend pas en compte la taille des différents sous-groupes. Il faudrait bien entendu inclure ce critère dans le processus.

Les effectifs doivent intervenir dans le calcul de la moyenne des écarts de moyennes minimaux entre modalités. Une solution envisageable consiste à donner une importance aux écarts d'autant plus grande que les sous-groupes ont des effectifs nombreux.

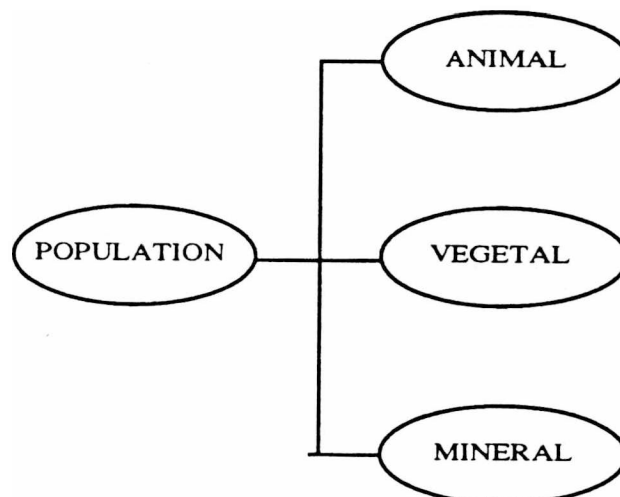
On considérera donc : $\left(\frac{1}{\text{effectif } A_1}\right) A_1(-)$

plutôt que : $A_1(-)$

Une telle méthode pose de nombreux problèmes. Elle semble notamment très lourde à mettre en œuvre. Cette limite, quoique contingente, est importante.

Plus fondamentale, sans doute, est la perte du caractère dichotomique du processus. La scission en trois groupes ou plus risque d'occulter une bonne part de l'information. En effet, un groupe peut se distinguer fortement des autres sans pour autant que l'ensemble des sous-groupes soient significativement distincts du groupe dont ils émanent. Le calcul tel que nous l'avons défini risque d'avoir des effets pervers, comme l'illustre l'exemple suivant :

Nous nous intéressons à la longévité pour une population dont on connaît par ailleurs les caractéristiques suivantes : le type (végétal, minéral et animal), la taille, etc... Supposons que la première variable sélectionnée soit le type ; nous aurons donc :



Or, c'est peut-être la modalité «minéral» qui se distingue des deux autres. La longévité chez les «non minéraux» peut s'expliquer d'abord par une autre variable. Ceci nous amène tout naturellement à proposer une autre voie de recherche...

II.1.2. La transformation des modalités en variables binaires

Deux transformations sur les données initiales sont envisageables. La première est simple par sa construction. Elle consiste à coder les variables discrètes de manière à se reporter à un tableau logique. La seconde utilise le critère de la variance expliquée.

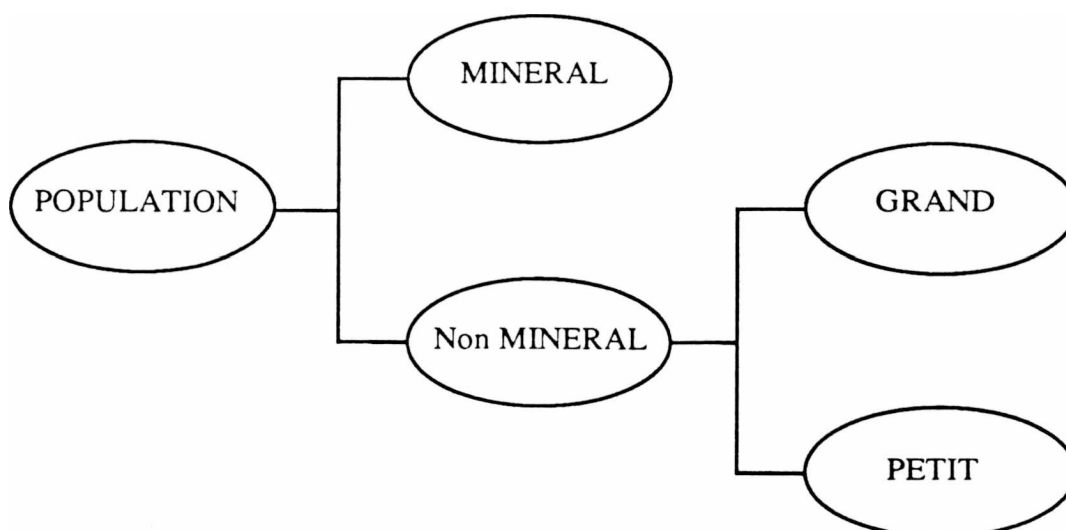
Construction d'un tableau logique

Nous venons d'esquisser une première résolution alternative de notre modèle élargi aux variables discrètes. En effet, il nous suffit de traiter notre tableau statistique initial pour transformer celui-ci en tableau logique. Chacune des modalités peut correspondre à une variable du type «présence/absence» ou «oui/non». Alors, l'algorithme précédemment défini reste utilisable.

Ainsi, si nous reprenons l'exemple précédent, nous aurons 3 variables bimodales :

- Végétal : oui ou non.
- Animal : oui ou non.
- Minéral : oui ou non.

La structure même de notre arbre sera changée :



Quelque soit la véracité de l'arborescence construite ici, il est clair que nous évitons maintenant l'écueil préalable. Nous ne risquons plus ici de choisir une partition qui comporterait des classes trop semblables relativement à la variable principale. Auparavant, en effet, une des modalités pouvait être très significativement différente des autres et, par conséquent, déterminer pour une grande part la valeur du critère de sélection. Or, les autres modalités peuvent être peu distinctes l'une de l'autre. Si nous reprenons l'exemple des variables A et B, nous avons sélectionné B mais, si B3 se distingue fortement des autres, B2 et B4 sont fort proches l'une de l'autre. Nous ne risquons plus ce travers avec la construction de variables bimodales.

Bimodalisation par la variance expliquée

Nous disposons de variables plurimodales. Nous cherchons alors à nous ramener à la meilleure partition en deux classes de notre population. Celles-ci devront contenir chacune les modalités ayant la plus forte homogénéité, relativement à notre variable principale.

Pour ce faire, nous pouvons utiliser le critère de la variance expliquée¹ de préférence à l'écart des moyennes. Le premier indice possède en effet la bonne propriété de tenir compte des effectifs relatifs des sous-groupes. Si nous considérons une variable qui possède k modalités, nous cherchons quelle partition de ces k modalités en 2 classes maximise la variance expliquée.

Si nos modalités sont ordonnées logiquement (par exemple pour l'âge : enfants, adolescents, jeunes adultes, etc...), nous allons définir les deux classes par glissement successif sur les modalités. Ainsi, pour une variable caractéristique disposant de 4 modalités, nous calculerons la variance expliquée de la variable principale pour la classe constituée de la modalité 1 contre la classe composée des 3 autres modalités. Puis, nous recommencerons avec la classe composée des 2 premières modalités contre celle composée des 2 autres. Nous obtenons alors 3 valeurs de la variance expliquée. Nous choisirons alors la partition en deux classes qui maximise l'indice de la variance expliquée.

¹ La variance expliquée est donnée par le calcul suivant :

$$\sum_i n_i (\bar{x}_i - \bar{X})^2$$

où : i correspond à l'une des 2 modalités d'une variable caractéristique,
 n_i est le nombre des individus possédant la modalité i ,
 \bar{x}_i est la moyenne de la variable principale pour la modalité i de la variable caractéristique,
 \bar{X} est la moyenne de la variable principale sur la population totale.

Si nos modalités ne sont pas ordonnées de manière logique (par exemple : *minéral*, *végétal*, *animal*), la méthode reste la même à une exception près, le nombre des partitions possibles est plus vaste. Nous devons considérer la classe composée des modalités 1 et 4 contre celle réunissant 2 et 3 par exemple. Ainsi, pour une variable caractéristique ayant k modalités, le nombre de partitions en deux classes possibles est de :

$$k-1 \text{ si les modalités sont ordonnées et } \sum_{i=1}^{E\left(\frac{K}{2}\right)} C_K^i \text{ si les modalités ne sont pas ordonnées.}$$

Il est donc possible de ramener l'ensemble des variables plurimodales à des variables bimodales ayant la meilleure homogénéité possible relativement à l'indice de la variance expliquée.

Plutôt que de généraliser la méthode aux variables, il semble donc plus pertinent d'adapter le codage des variables à la méthode.

II.2. Construction du test

La richesse comme les limites de la méthode reposent sur la construction du test de sélection de la variable séparant au mieux un groupe à chaque nœud de notre arborescence. Nous avons exposé la méthode de manière à en esquisser les grandes lignes. Nous pouvons maintenant développer plus avant la construction pratique du processus.

La sélection de la variable à retenir pour séparer notre groupe en deux répond à plusieurs objectifs :

II.2.1. Tester la significativité de la coupure en deux groupes

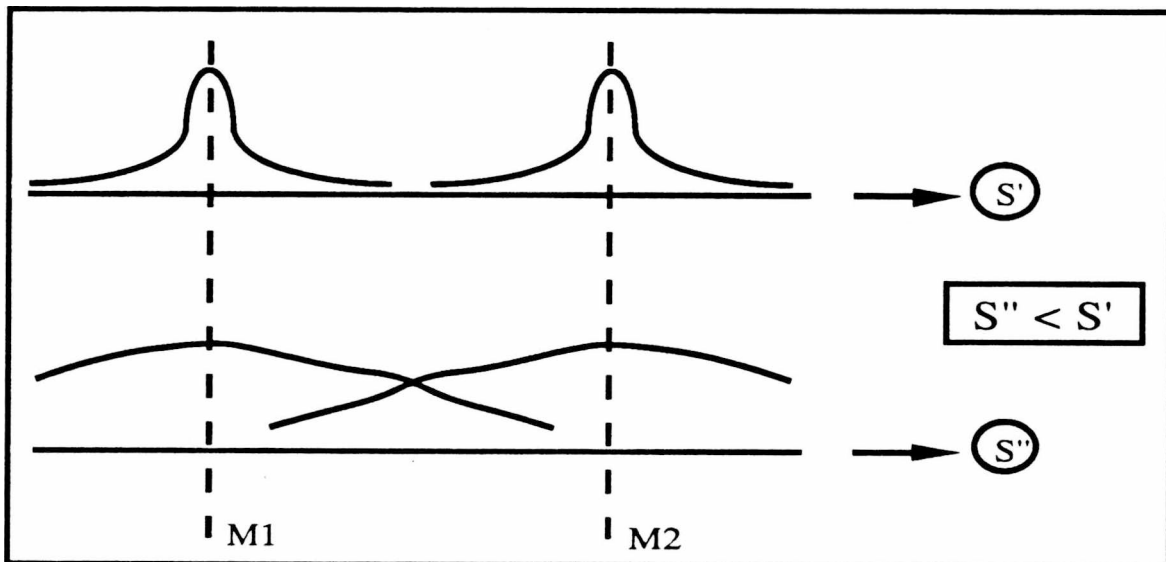
Cet objectif est le plus fondamental. Lors de notre présentation générale, nous avons choisi un critère simple en choisissant de calculer la statistique :

$$S = \frac{|\text{moyenne}(A) - \text{moyenne}(A^{\circ})|}{\text{écart-type } A + \text{écart-type } A^{\circ}}$$

où A° est le complémentaire de A .

Nous l'avons envisagé comme critère minimal de sélection. En fait, il ne suffit pas de calculer cette statistique pour l'ensemble des variables et de les trier selon celle-ci pour choisir la meilleure. Il faut d'abord s'assurer que cette statistique est significative à un seuil que nous fixons a priori, par exemple 95%.

Par ailleurs, la statistique S permet d'arbitrer entre deux variables qui scindent l'échantillon en deux groupes ayant un écart de moyennes identique. Il est en effet souhaitable de retenir, à écart de moyennes égal, la variable pour laquelle la dispersion autour de la moyenne est la plus faible. Les deux groupes sont dans ce cas mieux discriminés comme l'illustre le schéma suivant :



Ce test est double. Il faut d'abord s'assurer que S est significativement non nulle ; c'est-à-dire que la moyenne de A est significativement différente de celle de A° . Ceci constitue l'hypothèse faible. Ensuite, il faut s'assurer que la moyenne de A et la moyenne de A° sont toutes deux significativement différentes de la moyenne du groupe avant partition. C'est ce que nous appellerons l'hypothèse forte.

En effet, si l'hypothèse forte est vérifiée, l'hypothèse faible le sera également. La réciproque n'est pas vraie. Si l'on désire s'imposer des contraintes fortes, on choisira de satisfaire à l'hypothèse forte. On choisira l'hypothèse faible si on prend l'option inverse.

II.2.2. Garder des sous-groupes de taille significative

Les valeurs extrêmes introduisent un biais dans la méthode des plus proches voisins. Une classification en deux classes a souvent la mauvaise propriété, comme nous l'avons expérimenté, d'établir une classe d'individus atypiques. En l'occurrence, nous pouvons obtenir une partition en deux classes avec la quasi totalité des effectifs de la population dans l'une des classes et deux ou trois individus dans l'autre.

Pour remédier à ce problème, plusieurs méthodes sont possibles. Nous avons choisi d'éliminer d'abord les individus les plus atypiques pour l'ensemble des variables et de les analyser à part. Ensuite, dans le cas de variables continues, il convient de ne considérer, pour établir la coupure, qu'un échantillon central compris entre le cinquième et le quatre-vingt-quinzième centile par exemple. Enfin, après avoir réintroduit les centiles initialement écartés, nous calculons l'écart de moyennes pour la variable principale entre ces deux groupes.

En l'état actuel de nos recherches, nous avons laissé le choix du seuil à l'utilisateur. Soit *min* ce seuil, nous avons donc continué la procédure dichotomique à condition que l'effectif du groupe à scinder soit au moins égal à deux fois ce seuil. En effet, dans le cas contraire, l'un des deux groupes issus de la coupure aurait un effectif inférieur à ce seuil.

Une fois la procédure dichotomique exécutée pour l'ensemble des variables caractéristiques, lorsque la taille du groupe le permet a priori, on élimine de la sélection les variables qui ont donné un groupe d'effectif inférieur au seuil. Cette contrainte vient doubler celle du test d'écart de moyenne.

La sélection de la variable caractéristique résulte de ce jeu de contraintes. La variable retenue sera, parmi les variables qui isolent deux groupes de taille licite, celle pour laquelle l'écart de moyennes sur la variable principale est le plus élevé.

II.2.3. Construire un test de sélection optimal

Construction d'un test local

Nous venons de voir la façon dont il convient de gérer les deux types de contraintes de notre modèle. Si nous en connaissons les deux arguments, il reste encore à trouver un critère de sélection des variables optimal.

Ceci constitue l'un des axes de recherche qu'il faudrait approfondir. Jusqu'ici, nous avons choisi un critère arbitraire pour la taille des groupes, à peu près une vingtaine d'individus.

Notre test est, de plus, un test local et paramétrable de la forme :

$$\text{Test}_{\text{loc}} = F(\text{seuil}_{\text{test écart moyennes}}, \text{seuil}_{\text{effectif du groupe}})$$

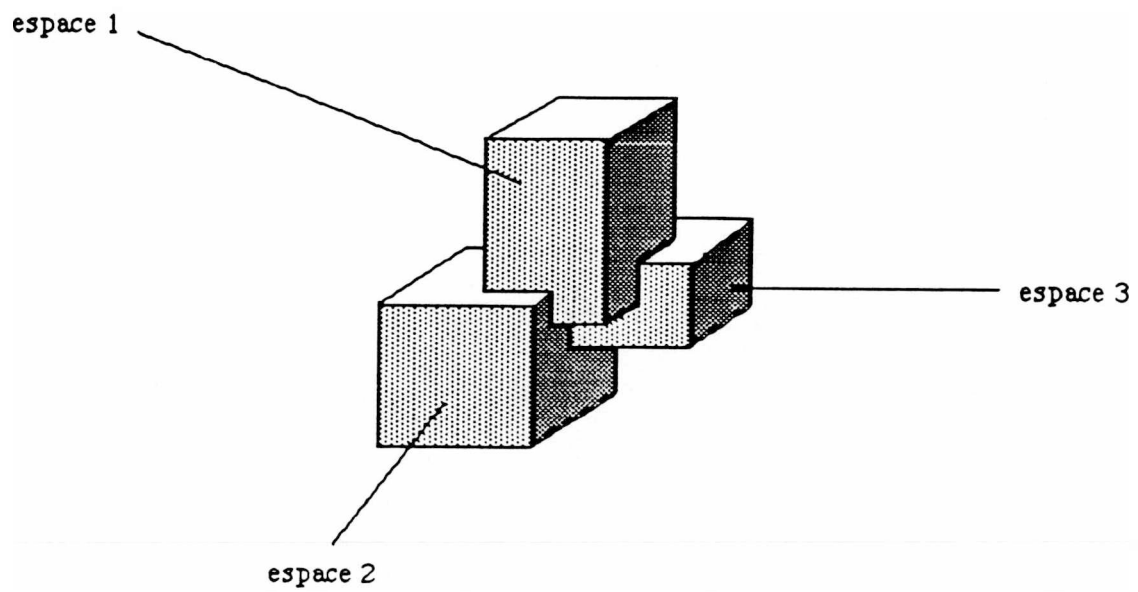
Nous avons laissé le choix à l'utilisateur de fixer chacun des deux seuils. Nous serons sans doute amenés ultérieurement à intégrer une procédure automatisée de sélection optimale des contraintes.

Hypothèse d'un test global

Existe-t-il un moyen de tester la «performance» de notre arborescence ? S'il est possible de tester la validité des segmentations pour chacun des nœuds pris séparément, nous ne disposons pas d'une statistique globale de notre modèle. En effet, le choix des variables peut différer à chaque segmentation. Chaque couple de groupes finaux est donc issu d'un choix particulier de variables. Nous avons donc pour chacun de ces couples, un espace de variables différent des autres. Par conséquent, il est impossible de calculer l'inertie totale du modèle et a fortiori de bâtir un indicateur sur celui-ci.

C'est pourquoi, nous n'émettons que l'hypothèse d'un test global de ce modèle. Est-il possible par exemple de bâtir un tel test à partir des différents tests locaux d'écart de moyennes ? Si oui, quel sens lui donner ?

La structure arborescente de notre modèle nous ôte les moyens habituels d'estimation. Tout se passe comme si, lors d'une analyse factorielle, nous scindions notre population selon sa répartition sur le premier axe factoriel et que le second axe soit différent pour chacune des deux sous-populations ainsi définies. Chacun des espaces correspondant à un couple de groupes finaux contient au moins une dimension commune avec les autres espaces. Il existe donc des points de rencontre entre tous les espaces finaux issus de l'analyse.



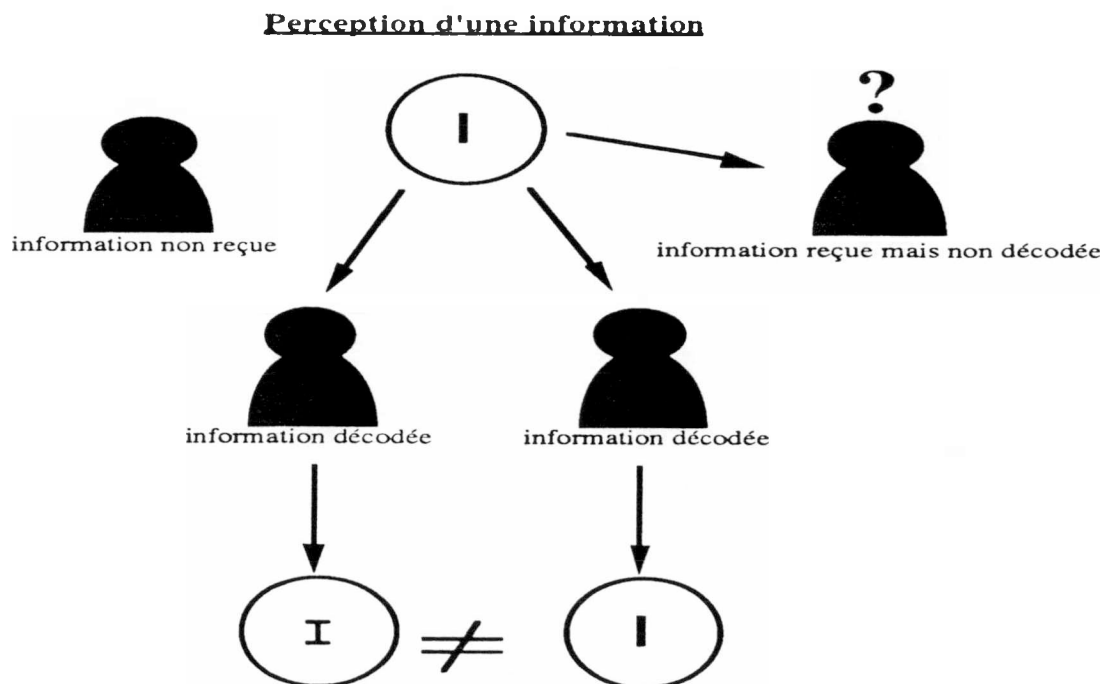
Ce schéma a pour but de souligner que chacun des espaces finaux de l'arbre est issu d'un tronc commun. Mais ces espaces ne se confondent pas. L'enchaînement des segmentations singularise, en fin de compte, des logiques comportementales différentes. Le schéma ne doit cependant pas donner à penser que ces espaces sont des sous-espaces d'un même espace. Les espaces 1, 2 et 3 ne possèdent que quelques dimensions communes, situées en haut d'arborescence.

III - FONDEMENTS SOCIOLOGIQUES DU MODELE

Nous voudrions montrer ici comment la méthode s'inscrit dans l'optique du formalisme de la F.R.C¹.

III.1. Prospective et comportement des individus

A l'origine de la démarche statistique que nous avons développée, réside l'idée selon laquelle la logique sous-jacente à des comportements sociaux, économiques ou politiques n'est pas globale. Chaque agent économique, chaque acteur social ne réagit pas de la même façon face à un ensemble de stimuli. Ils sont les observateurs d'un monde qu'ils se représentent, chacun isolément. En ce sens, il n'est pas besoin de supposer un réel perçu par l'ensemble des agents. Nous pouvons nous passer du postulat de réalité, comme le suggère Lahlou dans son formalisme, et considérer seulement les diverses représentations des agents afin de juger si elles sont plus ou moins proches. Ils peuvent ne pas prendre en compte l'information, ne pas la décoder ou encore la décoder différemment.



¹ La F.R.C. ou Formalisation en Relativité Complète est développée par Saadi Lahlou dans *Eléments de formalisation pour une théorie de l'évolution des systèmes : la système-compatibilité*, Cahier de recherche N°4, Crédoc, Paris, octobre 1990, 186 pages.

Nous prendrons ici l'exemple de l'univers boursier dans le but d'illustrer chacun de ces points. Ainsi, le comportement d'un intervenant détenteur de titres dans l'industrie audiovisuelle ne sera pas affecté par l'annonce d'une sécheresse, contrairement à celui qui détiendrait des valeurs céréalières. Dans ce cas, il y a prise en compte de l'information de façon sélective.

La connaissance par un agent du recours accru à des «junk bonds» par une entreprise dont il détient des actions peut être prise en compte sans être décodée ou analysée. L'agent peut ne pas savoir ce que sont ces junk bonds et ne pas juger cette information comme pertinente.

Enfin, deux agents peuvent détenir des titres pétroliers, connaître l'existence de tensions militaires dans le Golfe persique et spéculer, l'un à la hausse, l'autre à la baisse. L'information, connue, est alors décodée différemment.

Cet exemple nous amène à penser qu'il existe bien des logiques différentes pour les agents. Il existe des variables pertinentes localement et d'autres non. Enfin, il existe un grand nombre de cas où la prétention à rendre compte globalement des logiques d'un marché en fonction d'un nombre donné de variables masque les logiques véritables de sous-groupes plus homogènes. Une démarche plus analytique est plus coûteuse mais les progrès en matière d'informatique nous permettent de définir la logique d'un marché en fonction de la logique des agents qui s'y rencontrent et qui, par conséquent, le constituent.

Ceci demande de retourner une formulation souvent faite qui consiste à confondre l'essence de l'objet avec son existence. Ainsi, il est commun de dire que les agents se rencontrent sur le marché des capitaux. Ceci sous-entend que le marché précède la rencontre des agents. Or, sans cette rencontre, le marché n'est rien. Il est donc souvent dangereux de chercher la logique d'un marché pré-défini. Il conviendrait beaucoup mieux d'analyser le comportement des groupes d'agents qui, en fin de compte, définissent le terme même de marché. Une telle approche permet de mieux comprendre le marché en étudiant les fondements de sa genèse.

La CDD en l'occurrence, permet d'isoler ces marchés puisqu'elle isole des groupes d'individus homogènes en fin d'arborescence. Marcel Mauss a montré le caractère structurant de l'échange dans l'émergence de la société¹. De même, tout objet social, économique ou

¹ Mauss M. - "Essai sur le don. Formes et raisons de l'échange dans les sociétés archaïques", in *Sociologie et Anthropologie*, P.U.F., Paris, pp 143-273.

politique peut se définir par une relation, un lien, entre différents individus. Ainsi, tout objet est une relation et réciproquement toute relation est un objet. La rencontre des agents est un marché et inversement le marché est une rencontre d'agents.

Par ailleurs, la C.D.D. sélectionne des variables pertinentes et les hiérarchise. Cette sélection peut alors constituer une étape prospective, préalable à une analyse économétrique sur chacun des marchés mis en évidence. L'intérêt d'une telle méthode est qu'elle permet l'analyse des rationalités procédurales, et, d'une manière générale, prend en compte l'asymétrie des variables critiques dans un modèle de choix.

III.2. Classification Descendante Dichotomique et Formalisation en Relativité Complète.

L'intérêt prospectiviste nous amène à mettre en parallèle la C.D.D. et la F.R.C.. En effet, la C.D.D. peut s'interpréter comme une mise en application pratique des concepts définis dans la F.R.C.. La procédure mise en œuvre a pour fonction première, comme nous venons de le montrer, d'isoler les différents objets issus de l'arborescence : marchés, représentations sociales, comportement, performance, symptôme, etc...

Le principe de la C.D.D. permet de séparer, d'un même mouvement, à la fois des groupes d'individus et les variables pertinentes pour expliquer une caractéristique particulière de ces individus. Dit autrement : soit un observateur A (le modélisateur) qui observe au moyen d'un U-langage composé de $k+1$ variables un ensemble (O_i) d'objets. A essaie de prévoir, pour un O_j quelconque, la $(k+1)$ ième variable, à l'aide des k autres variables. La C.D.D. lui permet de déterminer des lois locales, ou plutôt les variables qui permettent de construire les lois locales, c'est-à-dire :

- (1) des combinaisons de variables ordonnées selon leur pouvoir explicatif (les signes des U-langages locaux).
- (2) des catégories d'objets sur lesquels ces lois sont valables (le domaine de validité, ou étendue sur laquelle les lois *locales* sont applicables).

On retrouve ici la circularité caractéristique de la F.R.C., en ce sens que (2) est défini en termes des mêmes variables que celles qui sont utilisées pour (1). Nous avons ici une illustration concrète de ce que signifie «U-langage local» dans le domaine de l'analyse économique, c'est-à-dire l'ensemble des objets qui sont pertinents pour la description et l'explication, à l'intérieur d'un langage d'observation préalable (les $k+1$ variables). On

remarquera que la sélection des sous-groupes et des variables qui les déterminent est concomitante, ce que le lecteur habitué à la F.R.C. considérera comme épistémologiquement sain.

Nous verrons ainsi que, dans le cas des points de vente, les signes localement pertinents pour expliquer le ratio EBE/CA peuvent différer suivant les différents sous-groupes. Pour notre échantillon initial, il existe un univers des possibles composés des variables que l'analyste aura cru bon de renseigner. Chacun des groupes mis en évidence dépendra d'une sélection de ces variables qui isole la partie de l'univers des agents pertinente pour la modélisation. L'ensemble des variables retenues pour l'analyse constitue pour nous les signes de l'alphabet, c'est-à-dire l'ensemble des signes qui définissent l'univers initial. L'arborescence nous donne alors une série de lois locales qui constituent, par leur combinaison, des parties pertinentes de l'univers pour chacun des groupes finaux (du moins, du point de vue du modélisateur).

Après que ces U-langages ont été déterminés, il est possible de vérifier si une combinaison donnée (hors échantillon initial) est «système-compatible» avec l'un des groupes finaux. Si l'agent respecte en effet l'ensemble des règles locales qui conduisent à l'affecter dans un groupe, il obéit sans doute, en première approximation, à la même logique d'évolution que les autres individus du groupe puisqu'on a supposé l'existence d'un lien entre les variables observées et la variable principale. Il est associé à un même alphabet, à un même U-langage et probablement aux mêmes règles. Reprenons notre exemple sur le marché boursier pour illustrer ceci.

Supposons que nous avons mis en évidence certains marchés grâce à notre arborescence en considérant l'évolution des actions négociées. Nous disposons d'une nouvelle action, dont nous aimerions savoir à quel type de marché mis en évidence elle se rapproche le plus. Nous cherchons alors avec quel groupe elle est le plus «système-compatible». Une fois ce groupe déterminé, nous connaissons la logique d'évolution la plus probable de l'action considérée puisque nous connaissons la combinaison licite des variables auxquelles elle est censée réagir. L'économétrie peut alors aider à quantifier l'effet de chacune de ces variables sur l'évolution du cours de l'action.

Une méthodologie plus explicite devra ultérieurement être développée pour lier les trois éléments précédemment énoncés : C.D.D., F.R.C. et économétrie. Nous laissons ce travail à un rapport ultérieur car une telle méthodologie dépasse le strict cadre de l'exposé de la méthode.

IV - APPLICATION A LA GRANDE DISTRIBUTION

Nous disposons d'un échantillon de plus de 400 supermarchés. Nous cherchons alors à déterminer pour quelques indicateurs financiers la performance de ceux-ci, eu égard à leurs caractéristiques. Ici, les seuils ont été fixés à 95% pour le test d'écart des moyennes et à 18 individus comme effectif minimal des sous-groupes.

Nous disposons pour tester la méthode de 60 variables explicatives, tirées du bilan comptable et physique des supermarchés. En voici la liste :

Variables physiques :

Surface (en m²)
 Nombre de caisses
 Nombre d'employés
 Nombre de places de parking
 Equipement en caisses à lecture optique
 Nombre de caisses / m²
 Nombre d'employés / m²
 Nombre de places de parking / m²
 Commerces annexes
 Année d'ouverture
 Nombre d'hypermarchés dans le département
 Nombre de supermarchés dans le département
 Densité d'hypermarchés dans le département (nombre de m²/habitant)
 Densité de supermarchés dans le département (nombre de m²/habitant)

Variables comptables :

Chiffre d'affaires / m² (en kf)
 Valeur ajoutée / nombre d'employés (en kf)
 Taux de croissance du chiffre d'affaires en 1988 (en %)
 Frais de personnel / chiffre d'affaires (en %)
 Charges externes / chiffre d'affaires (en %)
 Frais de personnel / nombre d'employés (en kf)
 Charges financières / chiffre affaires (en %)
 Intérêts / chiffre d'affaires (en %)
 Produits financiers / chiffre d'affaires (en %)
 Rotation des stocks (en jours)
 Crédit fournisseurs (en jours)
 Crédit clients (en jours)
 Capitaux propres / actif immobilisé (en %)
 Dette financière / actif circulant (en %)
 Dette fournisseurs / passif (en %)
 Chiffre d'affaires / actif (en %)

Résultat net / actif (en %)

Excédent brut d'exploitation / actif (en %)

Stocks / actif (en %)

Ressources durables / emplois stables (en %)

Capitaux propres / ressources stables (en %)

Dettes financières / ressources durables (en %)

Capitaux propres / actif (en %)

Amortissement / immobilisations brutes (en %)

Fond de roulement / actif circulant (en %)

Actif circulant / dette à court terme (en %)

(actif circulant net - stock net) / dette à court terme (en %)

(dettes + effets à porter à l'escompte) / chiffre d'affaires (en %)

Dette financière / capacité d'autofinancement avant répartition

Capacité d'autofinancement / (chiffre d'affaires net + subventions d'exploitation)

Couverture du chiffre d'affaire par le fond de roulement (en jours)

Couverture du chiffre d'affaires par le besoin de fond de roulement (en jours)

Valeur ajoutée / actif circulant + effets portés à l'escompte

Valeur ajoutée / immobilisations corporelles et incorporelles brutes

Valeur ajoutée / actif + effets portés à l'escompte

Excédent brut d'exploitation / (chiffre d'affaires + subvention d'exploitation) (en %)

Résultat courant avant impôts / (chiffre d'affaires + subventions d'exploitation) (en %)

Résultat courant avant impôts / capitaux propres nets (en %)

Résultat net / (chiffre d'affaires + subventions d'exploitation) (en %)

Résultat net / capitaux propres nets (en %)

Taux de marque (en %)

Part des salariés (en %)

Valeur ajoutée / chiffre d'affaires (en %)

Part de l'Etat (en %)

Part des prêteurs (en %)

Part de l'autofinancement (en %)

IV.1. Excédent brut d'exploitation / (Chiffre d'affaires + subventions)

IV.1.1. Description analytique de la construction de l'arborescence

Cette méthode impose de ne travailler que sur des magasins renseignés sur l'ensemble des variables retenues pour l'analyse. Après élimination des magasins présentant des données manquantes, l'analyse des déterminants de la rentabilité économique (qui est mesurée par le ratio **EBE / (CA + subventions)**) porte sur 439 supermarchés. Le niveau moyen de rentabilité économique s'établit à 1,74 %. Les variables explicatives retenues ont été les suivantes :

Parmi l'ensemble des décompositions en deux groupes qui ont été effectuées avec chacune des variables explicatives (selon la méthode des plus proches voisins), c'est celle fondée sur le **nombre de places de parking** offertes à la clientèle qui maximise l'écart de rentabilité (pondéré par les écarts-types).

Ainsi, les magasins disposant de plus de 420 places de parking affichent une rentabilité moyenne de 2,74 %, contre seulement 1,70 % pour les autres. Notre échantillon ne comprend que 20 magasins disposant de plus de 420 places de parking. Il est donc impossible de poursuivre l'analyse des déterminants de la rentabilité pour ces derniers.

Pour les 419 magasins qui possèdent un parking de 420 places ou moins, on re-décompose en deux groupes pour chacune des variables explicatives restantes (toutes les variables initiales sauf le nombre de places de parking). C'est la décomposition réalisée à partir de l'importance du **crédit fournisseurs** qui maximise l'écart de rentabilité entre les magasins des deux groupes.

Ainsi, pour les magasins disposant d'un parking de 420 places ou moins, le fait de bénéficier d'un crédit-fournisseurs de plus de 58 jours s'accompagne d'une rentabilité de 0,29 %, très inférieure à celle des magasins ne bénéficiant que d'un crédit fournisseurs inférieur ou égal à 58 jours (1,79 %).

Ne disposant que de 26 magasins dotés d'un parking de 420 places ou moins et bénéficiant d'un crédit fournisseurs de plus de 58 jours, on ne peut poursuivre l'analyse.

Pour les 393 magasins dotés d'un parking de 420 places ou moins et caractérisés par un crédit fournisseurs de 58 jours ou moins c'est la **part des frais de personnel** dans le chiffre d'affaires qui apparaît comme la variable la plus explicative des écarts de rentabilité. Logiquement, les supermarchés ayant des charges de personnel représentant plus de 10,7 % de leur chiffre d'affaires affichent une rentabilité économique nettement inférieure à celle des magasins caractérisés par une part des frais de personnel dans le chiffre d'affaires inférieure à ce seuil (0,72 % contre 1,88 %).

La poursuite de l'analyse indique que, pour les magasins dotés de moins de 421 places de parking, bénéficiant d'un crédit fournisseurs de moins de 58 jours et caractérisés par des frais de personnel représentant moins de 10,7 % du chiffre d'affaires, le **nombre de caisses par m²** de surface de vente a une influence négative sur la rentabilité économique : lorsque ce ratio est inférieur à 0,008, la rentabilité économique moyenne est de 1,91 %, alors qu'elle est de 1,55 % dans le cas contraire.

Nous ne pouvons poursuivre l'analyse que pour les supermarchés dont le nombre de caisses par m² est inférieur à 0,008. Parmi les supermarchés offrant entre 0 et 420 places de

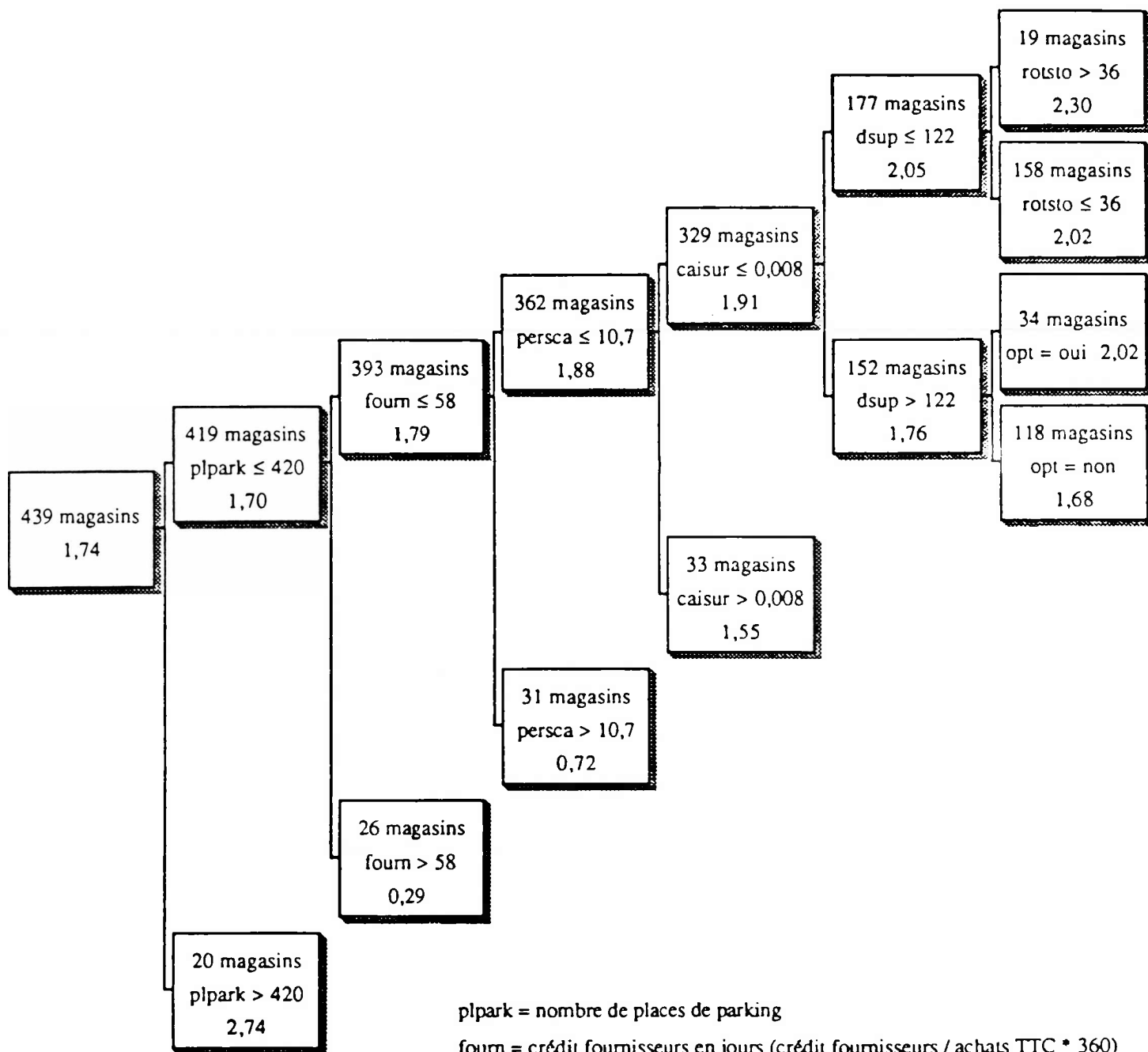
parking, avec un crédit fournisseurs de moins de 58 jours, des frais de personnel inférieurs à 10,7 % du chiffre d'affaires et un nombre de caisses par m² ne dépassant pas 0,008, ceux qui sont implantés dans un département dont la **densité en supermarchés** est inférieure ou égale à 122 m² par habitant jouissent d'une rentabilité moyenne de 2,05 % alors que les autres ne peuvent atteindre en moyenne que 1,76 % en raison sans doute de l'intensité de la pression concurrentielle. Les deux sous-groupes que nous obtenons ainsi sont suffisamment fournis pour que l'on puisse poursuivre l'analyse dans les deux directions.

Pour les magasins ayant les caractéristiques énoncées précédemment et situés dans un département à densité en supermarchés inférieure à 122 m² par habitant, c'est la **rotation des stocks** qui semble avoir la plus forte influence sur la rentabilité. Lorsque les stocks dépassent les 36 jours, la rentabilité moyenne s'établit à 2,30 %, alors qu'elle est de 2,02 % dans le cas contraire.

Pour les magasins affichant toutes les caractéristiques précitées et situés dans un département à densité de supermarchés supérieure à 122 m² par habitant, c'est le fait de disposer de **caisses à lecture optique** qui agit le plus fortement sur la rentabilité moyenne. Alors que les magasins équipés en scanner atteignent une rentabilité économique de 2,02 %, les autres ne se situent qu'à 1,68 %.

On ne peut aller plus loin dans l'analyse, soit que le nombre de magasins à l'intérieur des sous-groupes ne permette pas d'atteindre une nouvelle décomposition avec au moins 18 magasins dans chacun des deux sous-groupes qui en résultent, soit qu'aucune nouvelle variable n'offre une décomposition respectant ce seuil pour chacun des deux sous-groupes en résultant, ou alors avec un écart de rentabilité non significatif. L'arbre final est donc le suivant :

Excédent brut d'exploitation / (chiffres d'affaires + subventions)
(avec variables comptables)



plpark = nombre de places de parking

fourm = crédit fournisseurs en jours (crédit fournisseurs / achats TTC * 360)

persca = frais de personnel / chiffre d'affaires (en %)

caisur = nombre de caisses / surface de vente

dsup = nombre de m² de supermarchés par habitant dans le département

rotsto = rotation des stocks en jours (marchandises / (achats de marchandises + variation de stock de marchandises) * 360)

opt = équipement des caisses en scanners (oui/non)

IV.1.2. Utilisation de l'arbre par un banquier

L'utilisation par le banquier du schéma arborescent résultant de cette analyse est extrêmement simple. En présence d'un responsable de magasin dont l'analyste souhaite étudier le niveau de rentabilité par rapport aux normes de la profession, il suffit d'interroger le responsable du magasin selon l'ordre suggéré par le graphique.

Dans le cas présent, on commence par demander la dimension du parking. Admettons que le magasin n'en possède pas. On se trouve alors dans l'embranchement "parking inférieur ou égal à 420 places". On demande ensuite la durée du crédit-fournisseurs. Le responsable du magasin annonce 50 jours. On prend l'embranchement "crédit-fournisseurs inférieur à 58 jours". A la question de la part des frais de personnel dans le chiffre d'affaires, on obtient la réponse 8 % ; on se situe alors dans l'embranchement "frais de personnel inférieurs ou égaux à 10,7 % du chiffre d'affaires". On interroge ensuite le responsable de magasin sur son nombre de caisses par m² ; il répond 0,006 et l'on poursuit le cheminement en passant par l'embranchement "nombre de caisses par m² inférieur ou égal à 0,008". On sait que le département dans lequel est situé le magasin dispose de 135 m² de supermarchés par habitant ; on prend l'embranchement "densité en supermarchés supérieure à 122 m² par habitant". Enfin, le magasin n'est pas équipé en scanner ; on est donc dans l'embranchement terminal "lecture optique = non". L'analyste sait alors que les supermarchés présentant les mêmes caractéristiques que le magasin qu'il étudie (sur les variables apparaissant dans l'analyse) affichent une rentabilité moyenne de 1,68 %. La comparaison du niveau effectif de rentabilité économique du magasin étudié permet de le situer de manière fine en-dessous ou au-dessus des normes de la profession. Cet écart s'explique par les autres variables qui n'ont pas été prises en compte par l'analyse, comme le pouvoir d'achat et l'intensité de la concurrence à l'intérieur de la zone de chalandise, les effets liés à l'enseigne, à la qualité du manager ...

Cette méthode permet aussi de hiérarchiser les facteurs clés de succès. Ceux-ci ne sont pas nécessairement les mêmes pour tous les types de magasins. Pour l'étude de la rentabilité économique, cette méthode révèle l'importance du fait de détenir un parking, de grande dimension si possible. Elle aboutit au résultat a priori surprenant que, pour les magasins ne disposant pas d'un grand parking, un crédit fournisseur long va de pair avec une faible rentabilité. Sans doute ne faut-il pas voir ici une stricte relation de cause à effet ; les magasins les moins rentables s'efforcent peut-être de se rattraper dans leur négociation avec les fournisseurs. Des frais de personnel élevés dépriment la rentabilité, ce qui signifie que l'effet

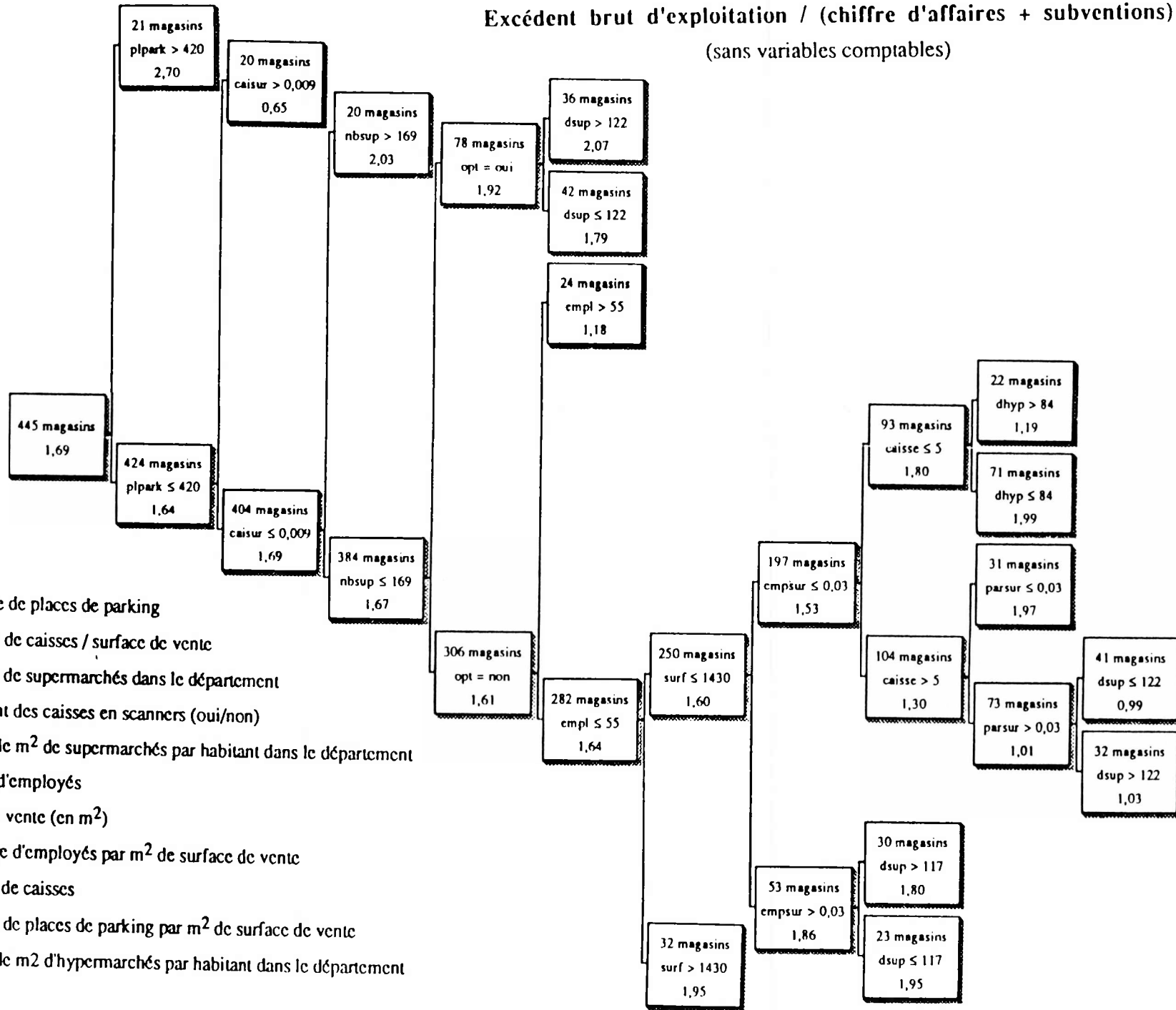
coût l'emporte sur l'effet différenciation (service rendu à la clientèle), tout au moins pour les magasins dotés d'un parking de 420 places ou moins et bénéficiant d'un crédit-fournisseurs inférieur ou égal à 58 jours. Dans le même ordre d'idée, un nombre de caisses important par rapport à la surface du magasin a un effet négatif sur la rentabilité ; là aussi, l'effet coût est plus fort que l'effet différenciation. Toutefois, on ne peut pas dire a priori qu'il s'agit d'une vérité de portée générale ; elle ne s'applique qu'aux magasins dont les caractéristiques les ont mené à ce niveau de l'arborescence. Avec les mêmes réserves, l'intensité de la pression concurrentielle dans le département (mesurée par le nombre de m² de supermarchés par habitant) influence le niveau de la rentabilité dans le sens conforme à l'intuition. Dans les départements où la pression concurrentielle est faible, des stocks importants agissent positivement sur la rentabilité. La faible densité de supermarchés encourage vraisemblablement la concurrence hors-prix. Des rayons bien achalandés ont sans doute un effet attractif permettant de vendre plus cher. Par contre, dans les départements à forte pression concurrentielle, l'efficacité retirée de l'équipement des caisses en scanner se retrouve dans une rentabilité supérieure.

IV.2. Excédent brut d'exploitation / (C A + subventions) sans les variables comptables

Nous avons également étudié les déterminants de la rentabilité économique des supermarchés en éliminant toutes les variables comptables, afin d'offrir un outil d'évaluation des projets d'ouverture.

Nous avons obtenu l'arbre suivant :

Excédent brut d'exploitation / (chiffre d'affaires + subventions)
(sans variables comptables)



- plpark = nombre de places de parking
- caisur = nombre de caisses / surface de vente
- nbsup = nombre de supermarchés dans le département
- opt = équipement des caisses en scanners (oui/non)
- dsup = nombre de m² de supermarchés par habitant dans le département
- empl = nombre d'employés
- surf = surface de vente (en m²)
- empsur = nombre d'employés par m² de surface de vente
- caisse = nombre de caisses
- parsur = nombre de places de parking par m² de surface de vente
- dhyp = nombre de m² d'hypermarchés par habitant dans le département

CONCLUSION

La méthode que nous avons proposée reste encore à l'état de proposition. Elle nécessite plusieurs améliorations qui permettront de la *généraliser à l'ensemble des données discrétisées*. Pour l'avoir testée sur des variables continues, nous pouvons juger néanmoins des développements futurs qu'elle requiert.

Tout d'abord, il conviendrait de construire un *test global* qui permette de juger de la qualité de notre arborescence. Actuellement, en effet, nous ne disposons que d'un test local (à chaque niveau de l'arbre) pour juger du caractère licite de la césure dichotomique. Ensuite, il semble utile de s'interroger sur la pertinence d'un *seuil minimal* quant aux effectifs des groupes. Ne pourrait-on pas en effet omettre ce seuil quitte à rallonger la procédure? Nous aurions ainsi des groupes statistiquement marginaux mais réels. Notre arborescence comporterait ainsi des branches majeures et d'autres mineures. Ces deux axes de recherche pourront faire l'objet d'un travail ultérieur.

L'intérêt d'une méthodologie comme la Classification Descendante Dichotomique réside principalement dans le *caractère local* de la sélection des variables de classification. Nous pouvons ainsi explorer les processus de détermination des choix des agents et définir les différents «marchés» en même temps que leur logique. Notre démarche est donc simultanément constitutive et descriptive. Nous ne pouvons donc pas nous tromper d'objet dans notre analyse, ces derniers apparaissant tout naturellement au fur et à mesure que l'arbre se construit.

Par ailleurs, les marchés sont considérés dans leur *nature relative*. Un marché, dans notre conception, n'est que la conjugaison des représentations, actions, caractéristiques ou des aspirations de plusieurs agents. Aussi, la C.D.D. peut à terme constituer une phase exploratoire, tant pour l'économie (comme le démontre notre exemple d'application) que pour l'étude des consommations ou des aspirations des français. Cet outil s'annonce donc d'un intérêt manifeste pour l'ensemble des départements du Crédoc. La prise en compte des impératifs de chacun permettra, de surcroît, d'améliorer l'outil.

BIBLIOGRAPHIE

Benzécri J.P. et coll. - *L'Analyse des données*. Dunod, Paris, 1973.

Benzécri J.P. et coll. - *Pratique de l'Analyse des données*. Dunod, Paris, 1981.

Brasseur A., P. Debreu P. et Lemel Y. - *Typologie des loisirs*. Les Collections de l'INSEE, N°292, série M, Paris, mars 1979.

Cohen L. et Holliday M. - *Statistic for Social Scientists*. Harper & Row, London, 1982, 382 pages.

Lahlou S. - *Eléments de formalisation pour une théorie de l'évolution des systèmes : la système-compatibilité*, Cahier de recherche n°4, Crédoc, Paris, octobre 1990, 186 pages.

Mauss M. - "Essai sur le don. Formes et raisons de l'échange dans les sociétés archaïques", in *Sociologie et Anthropologie*, P.U.F., Paris, pp 143-273.

Moati P. - *Les performances dans la grande distribution alimentaire*, Rapport Crédoc, Paris, novembre 1990, 173 pages.

Tassi P. - *Méthodes statistiques*. Economica, Paris, 1985, 389 pages.

Volle M. - *Analyse des données*. Economica, Paris, 1985, 324 pages.

ANNEXES

Nous donnons en annexe quelques éléments des listings issus des procédures rédigées en langage SAS. Ces listings ne correspondent pas aux arbres que nous avons présentés dans le corps du rapport. En effet, lors de l'analyse sur les performances dans la grande distribution alimentaire, la présentation n'avait pas encore fait l'objet d'une optimisation.

Les listings suivants concernent donc une autre enquête effectuée sur l'industrie française des équipements automobiles. La variable principale est ici TCA, définie comme le taux de croissance du chiffre d'affaires, les autres variables explicatives sont quant à elles des variables financières. Il importe surtout ici de donner une vision des sorties du programme et non pas d'interpréter les résultats.

Le lecteur trouvera d'abord les Résultats des itérations puis le début du Détail des itérations. Pour chaque variable explicative, seront donnés :

- le nom de la variable.

- | | |
|--------------|--|
| - nombre1 | effectif du sous-groupe 1. |
| - moyen1 | moyenne de la variable principale pour le sous-groupe 1. |
| - min1 | la valeur minimum des individus du sous-groupe 1. |
| - max1 | la valeur maximum des individus du sous-groupe 1. |
| - écart1 | l'écart-type de la variable principale pour sous-groupe 1. |
|
 | |
| - nombre2 | effectif du sous-groupe 2. |
| - moyen2 | moyenne de la variable principale pour le sous-groupe 2. |
| - min2 | la valeur minimum des individus du sous-groupe 2. |
| - max2 | la valeur maximum des individus du sous-groupe 2. |
| - écart2 | l'écart-type de la variable principale pour sous-groupe 2. |
|
 | |
| - indicateur | valeur de l'indicateur de tri |

SAS

11:49 WEDNESDAY, DECEMBER 5, 1990 1

RESULTATS DES ITERATIONS

ITERATION N° 1 :

variable STOAC88 nombre1 = 90 moyen1 = 12.30379 mini = 27.3272 maxi = 63.75056 ecart1 = 20.65622 nombre2 = 138 moyen2 =
18.33242 min2 = 1.514093 max2 = 26.08414 ecart2 = 26.10714 Indicateur = 0.1289178

ITERATION N° 2 classe 1 :

variable STOAC88 nombre1 = 70 moyen1 = 13.11069 mini = 27.3272 maxi = 41.90276 ecart1 = 21.90032 nombre2 = 20 moyen2 =
9.479633 min2 = 43.51549 max2 = 63.75056 ecart2 = 15.68658 Indicateur = 0.09660433

ITERATION N° 2 classe 2 :

variable PERSVA88 nombre1 = 102 moyen1 = 17.18157 mini = -99.2657 maxi = -61.821 ecart1 = 27.80671 nombre2 = 36
moyen2 = 21.59315 min2 = -60.8893 max2 = -33.1669 ecart2 = 20.53879 Indicateur = 0.09125111

ITERATION N° 3 classe 1 :

variable B126 nombre1 = 42 moyen1 = 16.17699 mini = 2.9 maxi = 40.03 ecart1 = 26.58233 nombre2 = 28 moyen2 = 8.511242
min2 = 43.64 max2 = 75.62 ecart2 = 10.77618 Indicateur = 0.2051941

ITERATION N° 4 classe 1 :

variable STOAC88 nombre1 = 22 moyen1 = 9.310466 mini = 34.50161 maxi = 41.90276 ecart1 = 15.95227 nombre2 = 24 moyen2 =
23.73016 min2 = 28.16705 max2 = 33.20088 ecart2 = 33.59916 Indicateur = 0.2910047

DETAILS DES ITERATIONS

ITERATION N° 1 :

```

variable B114 nombre1 = 226  moyen1 = 15.91793  min1   = 46.77  max1   = 591.68  ecart1  = 24.33742  nombre2 = 2  moyen2 = 19.88069

  min2   = 711.37  max2   = 746.77  ecart2  = 5.500538  Indicateur = 0
B114 = 746.77  B126 = 24.94  b111 = 0.03
B114 = 711.37  B126 = 59.85  b111 = 0.22

variable B126 nombre1 = 94  moyen1 = 12.84466  min1   = 41.86  max1   = 75.62  ecart1  = 21.92198  nombre2 = 134  moyen2 = 18.13296

  min2   = 2.9  max2   = 41.4  ecart2  = 25.58971  Indicateur = 0.1113051
variable b111 nombre1 = 205  moyen1 = 15.9745  min1   = -0.04  max1   = 0.11  ecart1  = 25.0641  nombre2 = 23  moyen2 = 15.75837

  min2   = 0.12  max2   = 0.26  ecart2  = 15.33783  Indicateur = 0.005349494
variable STDAC88 nombre1 = 90  moyen1 = 12.30379  min1   = 27.3272  max1   = 63.75056  ecart1  = 20.65622  nombre2 = 138  moyen2 =
18.33242  min2   = 1.514093  max2   = 26.08414  ecart2  = 26.10714  Indicateur = 0.1289178
variable PERSVA88 nombre1 = 217  moyen1 = 14.9978  min1   = -99.2857  max1   = -47.2014  ecart1  = 23.3405  nombre2 = 11  moyen2 =
34.79025  min2   = -45.9216  max2   = -33.1669  ecart2  = 33.99651  Indicateur = 0
variable RD88 nombre1 = 226  moyen1 = 16.02282  min1   = 0  max1   = 2.430207  ecart1  = 24.28194  nombre2 = 2  moyen2 = 8.028601

  min2   = 3.850203  max2   = 7.281866  ecart2  = 23.2714  Indicateur = 0
B114 = 183.05  B126 = 44.82  b111 = 0.18
B114 = 261.11  B126 = 49.09  b111 = 0.1

variable BREV88 nombre1 = 223  moyen1 = 16.28255  min1   = 0  max1   = 1.674822  ecart1  = 24.0701  nombre2 = 5  moyen2 = 1.241274

  min2   = 2.743056  max2   = 6.645337  ecart2  = 30.01934  Indicateur = 0
variable INSTEC88 nombre1 = 227  moyen1 = 16.019  min1   = 0  max1   = 173.5759  ecart1  = 24.26837  nombre2 = 1  moyen2 = 0.901386
7
  min2   = 304.3018  max2   = 304.3018  ecart2  = .  Indicateur = 0
B114 = 251.37  B126 = 40.71  b111 = 0.08

variable TXDX88 nombre1 = 219  moyen1 = 15.28305  min1   = 0  max1   = 80.10958  ecart1  = 22.98706  nombre2 = 9  moyen2 = 32.2475

  min2   = 89.83599  max2   = 99.98512  ecart2  = 44.23091  Indicateur = 0
variable SERV88 nombre1 = 222  moyen1 = 15.28351  min1   = 0  max1   = 18.88196  ecart1  = 22.20178  nombre2 = 6  moyen2 = 40.7124

  min2   = 25.02596  max2   = 56.60819  ecart2  = 64.21584  Indicateur = 0
variable ST88 nombre1 = 224  moyen1 = 15.86268  min1   = 0  max1   = 46.76685  ecart1  = 24.14372  nombre2 = 4  moyen2 = 20.99367

  min2   = 67.87818  max2   = 95.97765  ecart2  = 32.84638  Indicateur = 0

```

```
variable B166 nombre1 = 224 moyen1 = 15.83827 min1 = 3 max1 = 2500 ecart1 = 24.18131 nombre2 = 4 moyen2 = 22.36027
  min2 = 3000 max2 = 5757 ecart2 = 30.37469 Indicateur = 0
variable ROT88 nombre1 = 199 moyen1 = 15.48968 min1 = 56.68232 max1 = 233.0594 ecart1 = 22.45718 nombre2 = 29 moyen2 =
19.12996 min2 = 238.2435 max2 = 405.694 ecart2 = 34.40276 Indicateur = 0.06402198
variable B104 nombre1 = 224 moyen1 = 16.2066 min1 = 4.91 max1 = 128.88 ecart1 = 24.36277 nombre2 = 4 moyen2 = 1.734133
  min2 = 140.43 max2 = 213.02 ecart2 = 7.012457 Indicateur = 0
  ITERATION N° 2 classe 1
variable B114 nombre1 = 89 moyen1 = 12.17495 min1 = 98.71 max1 = 468.49 ecart1 = 20.73685 nombre2 = 1 moyen2 = 23.77015
  min2 = 746.77 max2 = 746.77 ecart2 = . Indicateur = 0
B114 = 746.77 B126 = 24.94 b111 = 0.05
variable B126 nombre1 = 88 moyen1 = 12.54997 min1 = 2.9 max1 = 63.09 ecart1 = 20.77153 nombre2 = 2 moyen2 = 1.471693
  min2 = 69.3 max2 = 75.62 ecart2 = 14.06282 Indicateur = 0
B114 = 197.87 B126 = 69.3 b111 = 0.1
B114 = 237.32 B126 = 75.62 b111 = 0.02
variable b111 nombre1 = 87 moyen1 = 12.43848 min1 = -0.03 max1 = 0.14 ecart1 = 20.82189 nombre2 = 3 moyen2 = 8.397593
  min2 = 0.18 max2 = 0.26 ecart2 = 17.91272 Indicateur = 0
B114 = 183.05 B126 = 44.82 b111 = 0.18
B114 = 285.3 B126 = 45.03 b111 = 0.26
B114 = 224.56 B126 = 46.59 b111 = 0.21
variable ST0AC88 nombre1 = 70 moyen1 = 13.11069 min1 = 27.3272 max1 = 41.90276 ecart1 = 21.90032 nombre2 = 20 moyen2 =
9.479633 min2 = 43.51549 max2 = 63.75056 ecart2 = 15.68658 Indicateur = 0.09660433
variable PERSVA88 nombre1 = 85 moyen1 = 10.25806 min1 = -93.1732 max1 = -50.6325 ecart1 = 16.30915 nombre2 = 5 moyen2 =
47.08121 min2 = -45.9216 max2 = -38.1534 ecart2 = 48.03222 Indicateur = 0
variable RD88 nombre1 = 88 moyen1 = 12.40095 min1 = 0 max1 = 1.583958 ecart1 = 20.73243 nombre2 = 2 moyen2 = 8.028601
  min2 = 3.850203 max2 = 7.281866 ecart2 = 23.2714 Indicateur = 0
B114 = 183.05 B126 = 44.82 b111 = 0.18
B114 = 261.11 B126 = 49.09 b111 = 0.1
variable BREV88 nombre1 = 88 moyen1 = 12.49735 min1 = 0 max1 = 1.674822 ecart1 = 20.61396 nombre2 = 2 moyen2 = 3.787045
  min2 = 2.917732 max2 = 4.067613 ecart2 = 29.26987 Indicateur = 0
```



```
B114 = 286.63 B126 = 51.02 b111 = 0.1
B114 = 261.11 B126 = 49.09 b111 = 0.1
variable INSTEC88 nombre1 = 74 moyen1 = 11.98333 min1 = 0 max1 = 56.27279 ecart1 = 20.97882 nombre2 = 16 moyen2 =
13.78591 min2 = 59.73052 max2 = 105.4112 ecart2 = 19.66996 Indicateur = 0
variable IXDX88 nombre1 = 81 moyen1 = 10.38913 min1 = 0 max1 = 42.33278 ecart1 = 15.60911 nombre2 = 9 moyen2 = 29.53567
min2 = 49.22111 max2 = 99.98512 ecart2 = 44.03628 Indicateur = 0
variable SERV88 nombre1 = 89 moyen1 = 12.44177 min1 = 0 max1 = 18.88196 ecart1 = 20.7315 nombre2 = 1 moyen2 = 0.02339729
min2 = 56.60819 max2 = 56.60819 ecart2 = . Indicateur = 0
B114 = 144.24 B126 = 33.74 b111 = 0.02
variable ST88 nombre1 = 89 moyen1 = 11.71763 min1 = 0 max1 = 22.39472 ecart1 = 20.00635 nombre2 = 1 moyen2 = 64.47219
min2 = 78.59953 max2 = 78.59953 ecart2 = . Indicateur = 0
B114 = 254.26 B126 = 29.25 b111 = 0.06
variable B166 nombre1 = 89 moyen1 = 12.28216 min1 = 4 max1 = 702 ecart1 = 20.77223 nombre2 = 1 moyen2 = 14.22914
min2 = 1337 max2 = 1337 ecart2 = . Indicateur = 0
B114 = 250.61 B126 = 59.85 b111 = 0.08
variable ROT88 nombre1 = 84 moyen1 = 12.74376 min1 = 59.69572 max1 = 265.8926 ecart1 = 21.07327 nombre2 = 6 moyen2 =
6.144166 min2 = 296.0254 max2 = 405.694 ecart2 = 13.2036 Indicateur = 0
variable B104 nombre1 = 86 moyen1 = 12.7954 min1 = 28.63 max1 = 128.88 ecart1 = 20.96476 nombre2 = 4 moyen2 = 1.734133
min2 = 140.43 max2 = 213.02 ecart2 = 7.012457 Indicateur = 0
ITERATION N° 2 classe 2
variable B114 nombre1 = 135 moyen1 = 17.96731 min1 = 46.77 max1 = 417.09 ecart1 = 26.19195 nombre2 = 3 moyen2 = 34.76218
min2 = 550.68 max2 = 711.37 ecart2 = 17.64318 Indicateur = 0
B114 = 591.68 B126 = 26.42 b111 = 0.04
B114 = 711.37 B126 = 59.85 b111 = 0.22
B114 = 550.68 B126 = 11.33 b111 = 0.03
variable B126 nombre1 = 126 moyen1 = 16.17939 min1 = 21.72 max1 = 68.66 ecart1 = 23.4895 nombre2 = 12 moyen2 = 40.93916
min2 = 11.33 max2 = 19.74 ecart2 = 40.10262 Indicateur = 0
variable b111 nombre1 = 121 moyen1 = 18.4396 min1 = -0.04 max1 = 0.11 ecart1 = 27.32104 nombre2 = 17 moyen2 = 17.56951
min2 = 0.12 max2 = 0.23 ecart2 = 15.39644 Indicateur = 0
variable STOAC88 nombre1 = 83 moyen1 = 18.37079 min1 = 15.45713 max1 = 26.08414 ecart1 = 23.64537 nombre2 = 55 moyen2 =
```

```

18.27451  min2  = 1.514093  max2  = 14.80865  ecart2  = 29.66792  Indicateur = 0.001805935
variable PERSVA88  nombre1 = 102  moyen1 = 17.18157  min1   = -99.2857  max1   = -61.821  ecart1  = 27.80671  nombre2 = 36  moyen2 =
21.59315  min2  = -60.8893  max2  = -33.1669  ecart2  = 20.53879  Indicateur = 0.09125111
variable RD88  nombre1 = 137  moyen1 = 18.10088  min1   = 0  max1   = 1.099394  ecart1  = 26.06035  nombre2 = 1  moyen2 = 50.05274
  min2  = 2.430207  max2  = 2.430207  ecart2  = .  Indicateur = 0
B114 = 258.35  B126 = 57.51  b111 = 0.23
variable BREV88  nombre1 = 136  moyen1 = 18.61592  min1   = 0  max1   = 2.743056  ecart1  = 25.80522  nombre2 = 2  moyen2 = -0.94563
9
  min2  = 5.124535  max2  = 6.645337  ecart2  = 52.20035  Indicateur = 0
B114 = 257.54  B126 = 32.52  b111 = 0.09
B114 = 221.67  B126 = 44.32  b111 = 0
variable INSTEC88  nombre1 = 137  moyen1 = 18.45965  min1   = 0  max1   = 173.5759  ecart1  = 26.15997  nombre2 = 1  moyen2 =
0.9013867  min2  = 304.3018  max2  = 304.3018  ecart2  = .  Indicateur = 0
B114 = 251.37  B126 = 40.71  b111 = 0.08
variable TXDX88  nombre1 = 123  moyen1 = 17.93485  min1   = 0  max1   = 38.31698  ecart1  = 26.71511  nombre2 = 15  moyen2 = 21.5924
6
  min2  = 41.65575  max2  = 99.75991  ecart2  = 20.91981  Indicateur = 0
variable SERV88  nombre1 = 133  moyen1 = 17.18513  min1   = 0  max1   = 11.91615  ecart1  = 23.01407  nombre2 = 5  moyen2 = 48.85021
  min2  = 25.02596  max2  = 42.70833  ecart2  = 68.24891  Indicateur = 0
variable ST88  nombre1 = 130  moyen1 = 17.9303  min1   = 0  max1   = 28.5109  ecart1  = 25.36764  nombre2 = 8  moyen2 = 24.86687
  min2  = 36.12144  max2  = 95.97765  ecart2  = 37.79743  Indicateur = 0
variable B166  nombre1 = 136  moyen1 = 18.46604  min1   = 3  max1   = 3500  ecart1  = 26.24298  nombre2 = 2  moyen2 = 9.245917
  min2  = 5000  max2  = 5757  ecart2  = 15.34898  Indicateur = 0
B114 = 303.2  B126 = 41.86  b111 = 0.14
B114 = 269.69  B126 = 43.13  b111 = 0.1
variable ROT88  nombre1 = 134  moyen1 = 17.53981  min1   = 56.68232  max1   = 282.3005  ecart1  = 22.96231  nombre2 = 4  moyen2 =
44.88459  min2  = 308.9975  max2  = 368.9482  ecart2  = 82.35284  Indicateur = 0
variable B104  nombre1 = 126  moyen1 = 18.11258  min1   = 4.91  max1   = 57.03  ecart1  = 26.92658  nombre2 = 12  moyen2 = 20.64068
  min2  = 58.89  max2  = 86.26  ecart2  = 15.59935  Indicateur = 0
ITERATION N° 3  classe 1
variable B114  nombre1 = 69  moyen1 = 12.95621  min1   = 98.71  max1   = 453.33  ecart1  = 22.02231  nombre2 = 1  moyen2 = 23.77015

```

```

  min2 = 746.77 max2 = 746.77 ecart2 = . Indicateur = 0
B114 = 746.77 B126 = 24.94 b111 = 0.05
variable B126 nombre1 = 42 moyen1 = 16.17699 min1 = 2.9 max1 = 40.03 ecart1 = 26.58233 nombre2 = 28 moyen2 = 8.511242
  min2 = 43.64 max2 = 75.62 ecart2 = 10.77618 Indicateur = 0.2051941
variable b111 nombre1 = 67 moyen1 = 13.32172 min1 = -0.03 max1 = 0.14 ecart1 = 22.15056 nombre2 = 3 moyen2 = 8.397593
  min2 = 0.18 max2 = 0.26 ecart2 = 17.91272 Indicateur = 0
B114 = 183.05 B126 = 44.82 b111 = 0.18
B114 = 285.3 B126 = 45.03 b111 = 0.26
B114 = 224.56 B126 = 46.59 b111 = 0.21
variable STOAC88 nombre1 = 32 moyen1 = 8.629286 min1 = 34.50161 max1 = 41.90276 ecart1 = 13.74649 nombre2 = 38 moyen2 =
16.8845 min2 = 27.3272 max2 = 34.0573 ecart2 = 26.53518 Indicateur = 0.2049373
variable PERSVA88 nombre1 = 66 moyen1 = 11.19367 min1 = -92.2163 max1 = -50.6325 ecart1 = 17.42262 nombre2 = 4 moyen2 =
44.74149 min2 = -45.9216 max2 = -38.1534 ecart2 = 55.13284 Indicateur = 0
variable RD88 nombre1 = 68 moyen1 = 13.26016 min1 = 0 max1 = 1.583958 ecart1 = 22.02418 nombre2 = 2 moyen2 = 8.028601
  min2 = 3.850203 max2 = 7.281866 ecart2 = 23.2714 Indicateur = 0
B114 = 183.05 B126 = 44.82 b111 = 0.18
B114 = 261.11 B126 = 49.09 b111 = 0.1
variable BREV88 nombre1 = 68 moyen1 = 13.38491 min1 = 0 max1 = 1.674822 ecart1 = 21.87426 nombre2 = 2 moyen2 = 3.787045
  min2 = 2.917732 max2 = 4.067613 ecart2 = 29.26987 Indicateur = 0
B114 = 286.63 B126 = 51.02 b111 = 0.1
B114 = 261.11 B126 = 49.09 b111 = 0.1
variable INSTEC88 nombre1 = 63 moyen1 = 13.91573 min1 = 0 max1 = 78.85405 ecart1 = 22.39434 nombre2 = 7 moyen2 = 5.86531
3
  min2 = 86.79775 max2 = 105.4112 ecart2 = 16.29089 Indicateur = 0
variable TXDX88 nombre1 = 65 moyen1 = 12.11722 min1 = 0 max1 = 59.70306 ecart1 = 17.07565 nombre2 = 5 moyen2 = 26.02586
  min2 = 73.73086 max2 = 99.98512 ecart2 = 58.16983 Indicateur = 0
variable SERV88 nombre1 = 67 moyen1 = 13.6674 min1 = 0 max1 = 3.013515 ecart1 = 22.06929 nombre2 = 3 moyen2 = 0.6774976
  min2 = 6.667526 max2 = 8.773231 ecart2 = 15.23157 Indicateur = 0
B114 = 286.63 B126 = 51.02 b111 = 0.1
B114 = 165.61 B126 = 43.97 b111 = 0.08
B114 = 202.9 B126 = 26.15 b111 = 0.04

```

```

variable ST88 nombre1 = 69  moyen1 = 12.36632  min1  = 0  max1  = 6.210491  ecart1  = 21.14997  nombre2 = 1  moyen2 = 64.47219
  min2  = 78.59953  max2  = 78.59953  ecart2  = .  Indicateur = 0
B114 = 254.26  B126 = 29.25  b111 = 0.06
variable B166 nombre1 = 69  moyen1 = 13.09448  min1  = 4  max1  = 702  ecart1  = 22.06034  nombre2 = 1  moyen2 = 14.22914
  min2  = 1337  max2  = 1337  ecart2  = .  Indicateur = 0
B114 = 250.61  B126 = 59.85  b111 = 0.08
variable ROT88 nombre1 = 67  moyen1 = 13.05073  min1  = 59.69572  max1  = 265.8926  ecart1  = 22.29662  nombre2 = 3  moyen2 =
14.44985  min2  = 362.2505  max2  = 405.694  ecart2  = 11.77417  Indicateur = 0
B114 = 242.55  B126 = 11.98  b111 = 0.02
B114 = 216.29  B126 = 13.55  b111 = 0.01
B114 = 746.77  B126 = 24.94  b111 = 0.05
variable B104 nombre1 = 69  moyen1 = 13.19257  min1  = 28.63  max1  = 140.43  ecart1  = 22.04997  nombre2 = 1  moyen2 = 7.46114
  min2  = 213.02  max2  = 213.02  ecart2  = .  Indicateur = 0
B114 = 182.03  B126 = 55.17  b111 = 0.01
  ITERATION N° 3  classe 2
variable B114 nombre1 = 19  moyen1 = 7.008031  min1  = 131.88  max1  = 334.47  ecart1  = 11.4357  nombre2 = 1  moyen2 = 56.44007
  min2  = 468.49  max2  = 468.49  ecart2  = .  Indicateur = 0
B114 = 468.49  B126 = 15.02  b111 = 0.06
variable B126 nombre1 = 6  moyen1 = 11.84428  min1  = 13.51  max1  = 20.77  ecart1  = 24.57285  nombre2 = 14  moyen2 = 8.466213
  min2  = 29.77  max2  = 54.34  ecart2  = 11.12257  Indicateur = 0
variable b111 nombre1 = 17  moyen1 = 5.599987  min1  = -0.01  max1  = 0.04  ecart1  = 8.207592  nombre2 = 3  moyen2 = 31.46429
  min2  = 0.06  max2  = 0.1  ecart2  = 30.754  Indicateur = 0
B114 = 305.53  B126 = 29.93  b111 = 0.06
B114 = 468.49  B126 = 15.02  b111 = 0.06
B114 = 265.11  B126 = 53.57  b111 = 0.1
variable STOAC88 nombre1 = 17  moyen1 = 6.700679  min1  = 43.51549  max1  = 53.73034  ecart1  = 12.0254  nombre2 = 3  moyen2 =
25.22704  min2  = 57.32722  max2  = 63.75056  ecart2  = 27.26089  Indicateur = 0
B114 = 183.98  B126 = 43.38  b111 = 0.02
B114 = 151.51  B126 = 42.65  b111 = 0.02

```

B114 = 468.49 B126 = 15.02 b111 = 0.06
variable PERSVA88 nombre1 = 19 moyen1 = 7.008031 mini = -93.1732 maxi = -38.7822 ecart1 = 11.4357 nombre2 = 1 moyen2 = 56.44007 min2 = -43.0429 max2 = -43.0429 ecart2 = . Indicateur = 0
B114 = 468.49 B126 = 15.02 b111 = 0.06
variable RD88 variable = constante continuation a la variable suivante
variable BREV88 nombre1 = 18 moyen1 = 9.169948 mini = 0 maxi = 0.336558 ecart1 = 16.54463 nombre2 = 2 moyen2 = 12.2668 min2 = 0.6583139 max2 = 1.016315 ecart2 = 2.172622 Indicateur = 0
B114 = 194.94 B126 = 41.4 b111 = 0.04
B114 = 252.27 B126 = 43.86 b111 = 0.01
variable INSTEC88 nombre1 = 19 moyen1 = 9.79633 mini = 1.910828 maxi = 54.64404 ecart1 = 16.0506 nombre2 = 1 moyen2 = 3.462394 min2 = 80.60194 max2 = 80.60194 ecart2 = . Indicateur = 0
B114 = 195.57 B126 = 44.31 b111 = 0.02
variable TXDX88 nombre1 = 19 moyen1 = 7.008031 mini = 0 maxi = 34.3198 ecart1 = 11.4357 nombre2 = 1 moyen2 = 56.44007 min2 = 97.51982 max2 = 97.51982 ecart2 = . Indicateur = 0
B114 = 468.49 B126 = 15.02 b111 = 0.06
variable SERV88 nombre1 = 19 moyen1 = 9.97733 mini = 0 maxi = 18.88196 ecart1 = 15.95337 nombre2 = 1 moyen2 = 0.02339729 min2 = 56.60819 max2 = 56.60819 ecart2 = . Indicateur = 0
B114 = 144.24 B126 = 33.74 b111 = 0.02
variable ST88 nombre1 = 19 moyen1 = 7.008031 mini = 0 maxi = 1.011378 ecart1 = 11.4357 nombre2 = 1 moyen2 = 56.44007 min2 = 22.39472 max2 = 22.39472 ecart2 = . Indicateur = 0
B114 = 468.49 B126 = 15.02 b111 = 0.06
variable B166 nombre1 = 19 moyen1 = 10.13046 mini = 5 maxi = 158 ecart1 = 15.83657 nombre2 = 1 moyen2 = -2.88604 min2 = 600 max2 = 600 ecart2 = . Indicateur = 0
B114 = 265.11 B126 = 53.57 b111 = 0.1
variable ROT88 nombre1 = 17 moyen1 = 11.53395 mini = 104.9506 maxi = 262.9272 ecart1 = 15.84397 nombre2 = 3 moyen2 = -2.16152 min2 = 296.0254 max2 = 343.3059 ecart2 = 9.500169 Indicateur = 0
B114 = 137.61 B126 = 54.34 b111 = 0.02
B114 = 166.69 B126 = 37.3 b111 = 0.02
B114 = 176.4 B126 = 20.77 b111 = 0.04
variable B104 nombre1 = 18 moyen1 = 10.09139 mini = 52.26 maxi = 128.88 ecart1 = 16.46282 nombre2 = 2 moyen2 = 3.973807

CAHIER DE RECHERCHE

Récemment parus :

Le statut social d'assisté : étude des stratégies d'utilisation de l'action sociale (rédaction novembre 1987), par Michel Messu, N° 10, Février 1991.

Les opinions des Français fin 1989 et une comparaison 1981-1989, Enquête "Conditions de vie et Aspirations des Français", par Laurent Clerc, Olivier Martin, sous la direction de Georges Hatchuel, N° 11, Février 1991.

Enquête sur mémoires - Etudes et Recherches dans le secteur social : Lexicographie de la littérature grise en action sociale - Volume 1, par Séverine Binard, Michel Legros, N° 12, Mars 1991.

Construction d'un corpus et perte d'information en analyse lexicale : méthodes et pratiques, par Sébastien Lion, sous la direction de Saadi Lahlou, N° 13, Avril 1991.

Penser l'insertion - Méthodes et critères : Contribution à une analyse des critères de l'insertion dans les réseaux de prise en charge des jeunes en difficulté, par Michel Legros, N° 14, Avril 1991.

L'analyse propositionnelle du discours, par Michel Messu, N° 15, Mai 1991.

Président : Bernard SCHAEFER Directeur : Robert ROCHEFORT
142, rue du Chevaleret, 75013 PARIS - Tél. : (1) 40.77.85.00

CREDOC

Centre de recherche pour l'Étude et l'Observation des Conditions de Vie