

ALAIN BACCINI

ÉLISABETH LAMBERT

**Algorithmes et programme de segmentation
aux moindres carrés**

RAIRO. Recherche opérationnelle, tome 20, n° 4 (1986),
p. 355-368

http://www.numdam.org/item?id=RO_1986__20_4_355_0

© AFCET, 1986, tous droits réservés.

L'accès aux archives de la revue « RAIRO. Recherche opérationnelle » implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques
<http://www.numdam.org/>

ALGORITHMES ET PROGRAMME DE SEGMENTATION AUX MOINDRES CARRÉS (*)

par Alain BACCINI et Élisabeth LAMBERT ⁽¹⁾

Résumé. — *Après avoir rappelé le principe des méthodes de segmentation les plus courantes, nous présentons dans cet article les algorithmes utilisés pour réaliser un programme de segmentation aux moindres carrés conforme, pour l'essentiel, aux normes MODULAD et traitant les cas d'un critère nominal et d'un critère ordinal. Un exemple illustratif est présenté.*

Mots clés : Segmentation aux moindres carrés; critère nominal; critère ordinal; normes MODULAD.

Abstract. — *Principles of most current segmentation methods are first recalled. Then, some algorithms and a program (somewhat framed according to the norms set by MODULAD) for least squares segmentation are proposed; this program enables to deal with either the case of a nominal dependent variable or of an ordinal one. An illustrative example is given.*

Keywords: Least squares segmentation; nominal dependent variable; ordinal dependent variable; MODULAD norms.

1. INTRODUCTION

Diverses méthodes de segmentation ont été développées, en Statistique, depuis une vingtaine d'années. Si l'utilisation de la seule technique de segmentation dans le traitement statistique de données peut paraître limitée (la méthode opère dans le cadre un peu rigide des dichotomies d'une seule variable à chaque étape), son intérêt, si elle est utilisée conjointement à d'autres méthodes (analyse des correspondances multiples, typologie, modélisation, ...), est évident, en particulier pour hiérarchiser les variables dites explicatives.

(*) Reçu janvier 1986.

(²) Laboratoire de Statistique et Probabilités, U.A.-C.N.R.S. n° 745, Université Paul-Sabatier, 118, route de Narbonne, 31062 Toulouse Cedex.

Aussi, est-il regrettable de constater que les principaux logiciels d'Analyse des Données actuellement disponibles sur les systèmes informatiques universitaires (SICLA, SPAD, ADDAD, EDA, ...) ne comportent pas de programme de segmentation.

Pour pallier cette lacune, nous proposons ici un programme de segmentation aux moindres carrés traitant le cas d'un critère nominal et celui d'un critère ordinal. Le programme, conçu en mode conversationnel, a été écrit, pour l'essentiel, en conformité avec les normes MODULAD (novembre 1983) et est facilement portable sur différents systèmes informatiques.

Après avoir rappelé le principe des méthodes de segmentation au paragraphe 2, nous décrivons algorithmes et programme au paragraphe 3 et traitons un exemple au paragraphe 4.

2. RAPPELS SUR LES MÉTHODES DE SEGMENTATION

2.1. Généralités

L'objet des méthodes de segmentation est d'expliquer une variable statistique Y (appelée variable à expliquer ou *critère*) au moyen d'un ensemble de p variables $X^1, \dots, X^j, \dots, X^p$ (appelées variables explicatives ou *facteurs*).

Les variables X^j sont toujours supposées qualitatives, ou du moins « catégorielles » (certaines peuvent être des variables quantitatives regroupées en classes). La variable Y est de nature quelconque, les méthodes de segmentation différant selon cette nature.

On dispose *a priori* d'un échantillon Ω de n individus sur lesquels ont été observées les $p+1$ variables considérées; ces individus seront affectés de poids identiques ($1/n$), la généralisation au cas de poids quelconques étant immédiate.

Le principe des méthodes de segmentation est de réaliser des dichotomies successives de Ω , chacune induite par une dichotomie des modalités de l'une des variables X^j ; on obtient ainsi ce que l'on appelle un *arbre de segmentation dichotomique*. La sélection d'une dichotomie se fait en maximisant un indice de liaison avec Y .

Les principaux travaux sur la segmentation sont dus à Belson (1959, [5]), Morgan et Sonquist (1963, [17]), Vo-Khac et Nghiem (1968, [18]), Bouroche et Tenenhaus (1970, [7]), Hughes (1970, [11]) et Baccini (1975, [2]); on peut en trouver une synthèse dans [1]. Récemment, Meunier, Baufays et Rassin (1985, [15]) ont également proposé de nouveaux critères de sélection des dichotomies.

Signalons aussi différents travaux sur une approche quelque peu différente de la segmentation visant à s'affranchir du cadre des dichotomies : Dehédin (1975, [9]), Kass (1980, [12]) et Meunier, Diday et Rasson (1985, [14]). Bien que n'envisageant que l'approche classique dans le programme proposé ici, nous pensons que ces dernières approches sont intéressantes et amenées à se développer.

2.2. Définition d'une étape de segmentation

Lorsque la partition dichotomique optimale de Ω , notée $\{\Omega_1, \Omega_2\}$, a été obtenue, la procédure est réitérée sur Ω_1 , puis sur Ω_2 , et ainsi de suite, jusqu'à satisfaire une règle d'arrêt fixée *a priori*.

Si l'on note I soit l'ensemble Ω , soit l'un quelconque de ses sous-ensembles obtenu par application de cette procédure, la recherche de la dichotomie optimale de I est toujours la même : pour chaque variable X^j , on cherche la dichotomie de ses modalités qui maximise l'indice de liaison avec Y .

Notant X l'une quelconque des variables explicatives X^j , on appelle *étape de segmentation* la recherche sur I de la dichotomie des modalités de X maximisant l'indice de liaison considéré.

Dans la suite de cet article, les procédures intervenant seront le plus souvent décrites pour une étape quelconque de segmentation.

Remarque : Si X comporte m modalités, il existe $2^{m-1} - 1$ dichotomies de ces modalités; si X est nominale (à modalités non ordonnées), chacune de ces dichotomies doit être considérée; par contre, si X est ordinale (à modalités ordonnées), ou quantitative regroupée en m classes, on ne considère en général que les $m - 1$ dichotomies correspondant à des regroupements de modalités (ou de classes) consécutives; cette dernière façon de procéder est très naturelle et, de plus, elle est cohérente avec le traitement d'un critère ordinal (voir [1]).

2.3. Les méthodes de segmentation aux moindres carrés

2.3(a) Définition

Ces méthodes consistent, à chaque étape, à considérer $F = \mathbb{R}^{\text{card}(I)}$ muni de la métrique euclidienne classique, ainsi que les sous-ensembles (définis ci-après) F_x et H_y respectivement associés à X et à Y . La définition de H_y dépend de la nature de Y (elle sera donnée dans chaque cas particulier), celle de F_x est la suivante :

soit $X^{(l)}$ la variable indicatrice de la l -ième modalité de X ($l = 1, \dots, m$), définie sur I ; on peut lui associer un vecteur unique x de F muni de la base

canonique, considérer le sous-espace vectoriel F'_x de F engendré par les x^i et définir F_x comme étant la supplémentaire orthogonale à $\mathbf{1}$ dans F'_x ($\mathbf{1}$ désigne le vecteur de F dont toutes les coordonnées valent 1, et l'on a : $\sum_{i=1}^m x^i = \mathbf{1}$).

A toute dichotomie $\{L_1, L_2\}$ de l'ensemble L des modalités de X correspond un ensemble de vecteurs x de F'_x du type

$$x = a \sum_{i \in L_1} x^i + b \sum_{i \in L_2} x^i, \quad a \neq b,$$

appelés vecteurs de Bernoulli.

Une étape de segmentation aux moindres carrés est alors la recherche du vecteur de Bernoulli de F_x normé et de distance minimale à H_y (voir [1]).

2.3(b) Cas d'un critère quantitatif

Si la variable à expliquer Y est quantitative, soit Z la variable centrée associée à Y sur I ; H_y est alors le sous-espace vectoriel de dimension 1 de F engendré par Z . La méthode de segmentation ainsi définie revient à chercher, à chaque étape, la dichotomie des modalités de X qui maximise le rapport de corrélation avec Y (voir [1]) et constitue la *méthode A.I.D.* (Automatic Interaction Detection) proposée par Morgan et Sonquist dans [17]. Une généralisation de cette méthode au cas d'un critère multidimensionnel a été proposée par Bouroche et Tenenhaus dans [7].

Remarque : Nous ne proposons pas ici de programme pour la méthode A.I.D.; nous envisageons de le réaliser ultérieurement. Nous avons préféré mettre d'abord au point le programme traitant le cas d'un critère qualitatif, ce cas étant plus fréquent dans notre pratique et le cas quantitatif pouvant s'y ramener après découpage en classes du critère.

2.3(c) Cas d'un critère nominal

Si Y est qualitative à modalités non ordonnées, notons F_y le sous-espace de F défini, à partir de Y , de façon analogue à F_x [voir 2.3(a)] et prenons $H_y = F_y$. La dichotomie des modalités de X retenue à chaque étape est celle qui maximise le χ^2 de contingence avec Y ; la méthode de segmentation associée est alors la *méthode E.L.I.S.E.E.* (Exploration des Liaisons et Interactions par Segmentation d'un Ensemble Expérimental) proposée par Cellard, Labbé et Savitsky dans [8].

2.3(d) Cas d'un critère ordinal

Lorsque Y est qualitative à modalités ordonnées, considérons le sous-espace F_y défini comme en 2.3(c) et notons C_1 (resp. C_2) le cône convexe fermé de F constitué des combinaisons linéaires à coefficients croissants (resp. décroissants), au sens large, des indicatrices des modalités de Y ; on prend dans ce cas :

$$H_y = F_y \cap (C_1 \cup C_2).$$

La méthode de segmentation associée a été proposée par Baccini dans [1]; il a montré que la dichotomie des modalités de X que l'on sélectionne est celle qui, croisée avec Y , maximise un indice de liaison dérivé du χ^2 ; cet indice est lié à la quantité $\bar{\chi}^2$ définie, dans un contexte différent, dans [4].

3. PROGRAMME ET ALGORITHMES PROPOSÉS ⁽²⁾

3.1. Description générale du programme

Le programme réalisé permet d'effectuer la segmentation aux moindres carrés dans le cas d'un critère nominal et dans celui d'un critère ordinal. Après la lecture des données (qui doivent être mises dans un fichier), ses principales phases sont : la génération des codes des dichotomies associées à chaque facteur; la recherche, à chaque étape, de la dichotomie optimale de la sous-population considérée et la segmentation de cette sous-population selon la dichotomie retenue; l'édition des résultats, une fois la procédure terminée. Les algorithmes permettant de réaliser ces différentes phases sont détaillés dans ce paragraphe.

On notera qu'il est nécessaire de fixer *a priori* une règle d'arrêt de la procédure, opérant sur chaque branche de l'arbre de segmentation. Le programme réalisé n'effectue plus la segmentation d'une sous-population lorsque son effectif est inférieur à 5% de l'effectif de la population initiale, cette règle pouvant être modifiée par l'utilisateur.

Le programme a été conçu en mode conversationnel, en particulier pour la construction et la modification du fichier des paramètres. Dans son écriture, nous nous sommes efforcés de rester le plus près possible des normes MODULAD (du 30 novembre 1983). Compte tenu du matériel dont nous disposions

(²) Le programme peut être obtenu en s'adressant aux auteurs.

(à savoir : un micro-ordinateur Victor S1 avec le logiciel Fortran MS-DOS 3.04 puis 3.13), nous avons été obligés d'apporter les modifications suivantes aux normes MODULAD : pour le choix des paramètres en mode conversationnel, pour les variables alphanumériques (titre, nom des variables explicatives, à expliquer, ...), nous avons introduit des variables de type caractère; dans les formats, les chaînes de caractères sont encadrées par des apostrophes et non définies par des constantes Hollerith; de même, pour lire ou créer les fichiers paramètres et de données, nous avons eu recours aux instructions d'entrée-sortie : « open » et « close »; enfin, pour l'édition des résultats sous forme de tableaux, nous avons construit des formats variables, ce qui a nécessité l'emploi des instructions « data » et « equivalence ».

Nous pensons néanmoins que ce programme reste facilement portable sur différents systèmes informatiques. Actuellement, il fonctionne sur un micro-ordinateur Victor S1 (sur lequel a été traité l'exemple présenté au paragraphe 4) et sur le D.P.S.8 (C.I.I.-H.B.) du Centre Interuniversitaire de Calcul de Toulouse; une version est en cours de mise au point sur l'I.B.M. 3081-K du Centre National Universitaire Sud de Calcul de Montpellier.

3.2. Génération de l'ensemble des dichotomies associées à un facteur

3.2(a) Cas d'un facteur nominal

Pour un facteur à m modalités ($m \geq 2$), les $2^{m-1} - 1$ dichotomies à considérer sont définies par les colonnes d'un tableau $m \times (2^{m-1} - 1)$ constitué de 0 et de 1. Le programme génère le tableau nécessaire en utilisant une procédure pas-à-pas déduite de l'égalité $2^{m-1} - 1 = 2(2^{m-2} - 1) + 1$: le tableau d'ordre m s'obtient par duplication du tableau d'ordre $m-1$, une première fois en lui rajoutant une ligne de 0, une seconde fois en lui rajoutant une ligne de 1, enfin par adjonction aux deux premiers tableaux juxtaposés d'une colonne comportant 0 dans ses $m-1$ premières lignes et 1 en dernière ligne :

Tableau
d'ordre 2

0
1

Tableau
d'ordre 3

0	0	0
1	1	0
0	1	1

Tableau
d'ordre 4

0	0	0	0	0	0	0
1	1	0	1	1	0	0
0	1	1	0	1	1	0
0	0	0	1	1	1	1

3.2(b) Cas d'un facteur ordinal

En procédant de façon analogue, le programme génère un tableau $m \times (m-1)$ de 0 et de 1 que l'on peut obtenir, par exemple, en supprimant la

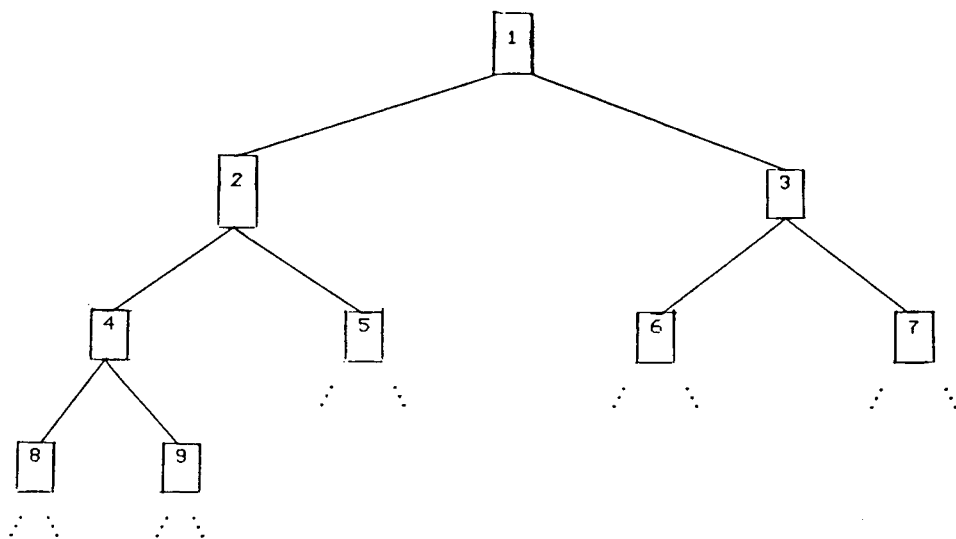
dernière colonne de la matrice carrée d'ordre m dont la diagonale principale et le triangle supérieur ne comportent que des 0 et dont le triangle inférieur ne comporte que des 1.

3.3. Algorithme de segmentation

3.3(a) Cas d'un critère nominal

Avant de définir la dichotomie optimale d'une sous-population donnée, il est nécessaire de repérer la position de celle-ci dans l'arbre de segmentation, afin de connaître la liste exacte des variables explicatives restant à considérer. En effet, si l'on considère une sous-population définie par une seule modalité d'une variable explicative, cette variable doit être exclue des variables explicatives considérées dans la suite de la procédure sur la partie de l'arbre issue de cette sous-population; de la même manière, si l'on considère une sous-population définie par les deux premières modalités d'une variable explicative à plus de deux modalités, cette variable sera considérée comme ayant seulement deux modalités dans la suite de la procédure; et ainsi de suite...

Le repérage d'une sous-population se fait de manière très simple, en utilisant, par exemple, la numérotation suivante des sous-populations successives :



En faisant alors une succession de divisions entières par 2 du numéro de la sous-population considérée, jusqu'à obtenir 1, on reconstitue immédiatement

le chemin qui, depuis la population initiale, permet d'obtenir cette sous-population.

La conservation en mémoire, pour chaque sous-population, de la variable explicative et de la (des) modalité(s) la définissant permet ainsi, chaque fois que nécessaire, de disposer de la liste exacte des variables explicatives à considérer.

A chaque étape de segmentation, on considère donc toutes les dichotomies associées à la variable explicative considérée et pour chacune d'elle, on construit la table de contingence $2 \times s$ (s est le nombre de modalités du critère); on calcule alors le χ^2 correspondant et l'on retient ensuite la dichotomie donnant le χ^2 maximum.

3.3(b) *Cas d'un critère ordinal*

La procédure est analogue à la précédente, mais, pour chaque dichotomie considérée, on calcul un indice de liaison avec Y (ce n'est plus le χ^2) tenant compte du caractère ordonné des modalités de Y . Sa mise en œuvre nécessite l'utilisation d'un algorithme de régression isotone.

Nous décrivons ci-dessous ce qu'est la régression isotone, ainsi que la mise en œuvre du calcul de l'indice envisagé.

Régression isotone : on appelle régression isotone de q nombres réels a_1, \dots, a_q par rapport aux poids p_1, \dots, p_q , les q nombres réels b_1, \dots, b_q minimisant la quantité $\sum_{i=1}^q p_i(a_i - b_i)^2$, sous les contraintes $b_1 \leq b_2 \leq \dots \leq b_q$.

Pour mettre en œuvre cette régression, nous avons utilisé l'algorithme « Up and Down Blocks » proposé dans [4] en se référant à [13].

Calcul de l'indice de liaison : dans une étape de segmentation, appelons Z une dichotomie quelconque des modalités de la variable explicative X considérée; notons n_{lk} ($l = 1, 2; k = 1, \dots, s$) le nombre d'individus présentant, dans la population considérée à l'étape où l'on se place, la modalité l de Z (c'est-à-dire l'une des modalités de X dont celle-ci est la réunion) et la modalité k de Y ; notons encore $n_{.k}$ la somme $n_{1k} + n_{2k}$, et ω_k le rapport $n_{1k}/n_{.k}$. Désignons par $(\alpha_k)_{k=1, \dots, s}$ (resp. $(\beta_k)_{k=1, \dots, s}$) la régression isotone de la suite $(\omega_k)_{k=1, \dots, s}$ (resp. $(-\omega_k)_{k=1, \dots, s}$) affectée des poids $(n_{.k})_{k=1, \dots, s}$, par T_α (resp. T_β) la variable obtenue en regroupant les modalités de Y correspondant à des valeurs égales de α_k (resp. β_k) et par T_z celle de ces deux variables dont le χ^2 de contingence calculé avec Z est le plus grand; l'indice de liaison entre Z et Y calculé dans ce cas est alors le χ^2 entre Z et T_z .

3.4. Édition des résultats

Le programme réalise l'édition des résultats de deux façons distinctes et successives.

3.4(a) Édition détaillée

Les différents niveaux de segmentation sont présentés successivement; pour chaque niveau est fournie la liste des sous-populations à segmenter, avec leur numéro et leur effectif; pour chacune de ces sous-populations est ensuite indiqué le facteur réalisant la dichotomie optimale, la valeur du χ^2 , et un descriptif des deux sous-populations engendrées : numéros, modalités du facteur les définissant, effectifs, pourcentages d'individus dans chaque modalité du critère.

3.4(b) Édition sous forme d'arbre

L'arbre de segmentation est présenté de gauche (population initiale) à droite, sur le nombre de pages nécessaires, en fonction du nombre de sous-populations définies. Le nombre de niveaux de segmentation représenté ici est limité à 5 sur D.P.S. 8 et à 4 sur Victor S1 (à cause de la limitation à 128 d'une variable caractère). Au sein de l'arbre, chaque sous-population est représentée par un « pavé », de 10 lignes sur 20 colonnes, contenant toutes les informations utiles.

4. EXEMPLE

4.1. Présentation

Une importante société toulousaine, effectuant de multiples transactions commerciales avec ses clients (fourniture de documentation, rapports techniques, devis, journées d'initiation technique, soirées, etc.), souhaitait savoir si ces transactions avaient une influence sur le montant des commandes des clients et, si oui, quelles étaient les plus importantes.

Un échantillon au hasard de 1259 clients a été constitué; pour chaque client, on a enregistré d'une part le nombre de transactions commerciales effectuées avec la société considérée, pour la période 1978-1982, dans 10 groupes de transactions définis *a priori* (il s'agit des 10 variables explicatives), et d'autre part la classe du montant des commandes pour l'année 1983 (variable à expliquer). Chaque groupe de transactions comporte différents niveaux (de 3 à 6) correspondant au nombre de transactions effectuées, avec certains regroupements (par exemple, pour le premier groupe, les trois niveaux

correspondent à 0,1 et 2,3 ou 4 transactions); on a donc affaire ici à 10 variables ordinales. La variable « montant des commandes » comporte six classes et sera également considérée comme ordinale.

Ces données avaient déjà fait l'objet d'analyses statistiques élémentaires et d'une analyse des correspondances multiples; il était donc intéressant de les traiter par segmentation, pour confronter les résultats obtenus.

4.2. Les entrées (données initiales)

La première page de résultats fournie par le programme de segmentation est, en fait, un résumé des données traitées; dans l'exemple considéré, cette page est la suivante :

```

Societe toulousaine (1259 clients, 10 transactions, chiffre d'affaires)
=====

1259 individus

variable a expliquer y:
      y      6 modalites ordonnees
      modalites: 0 1 2 3 4 5

10 variables explicatives x:
1 - 01t      3 modalites ordonnees
      modalites: 0 1 2
2 - 02t      3 modalites ordonnees
      modalites: 0 1 2
3 - 03t      6 modalites ordonnees
      modalites: 0 1 2 3 4 5
4 - 04t      4 modalites ordonnees
      modalites: 0 1 2 3
5 - 05t      5 modalites ordonnees
      modalites: 0 1 2 3 4
6 - 06t      4 modalites ordonnees
      modalites: 0 1 2 3
7 - 07t      3 modalites ordonnees
      modalites: 0 1 2
8 - 08t      3 modalites ordonnees
      modalites: 0 1 2
9 - 09t      3 modalites ordonnees
      modalites: 0 1 2
10 - 10t     4 modalites ordonnees
      modalites: 0 1 2 3

- rangement de l'arborescence dichotomique de haut en bas par ordre
  decroissant de moyenne des valeurs de la var. a expliquer

- l'effectif minimum d'une sous-population pour lequel on arrete la segmentation
  d'une branche est de      62

- l'effectif minimum d'une sous-population pour lequel on rejette
  la dichotomie est de      7

- le nombre de niveaux est 5

```

4.3. Les sorties (résultats)

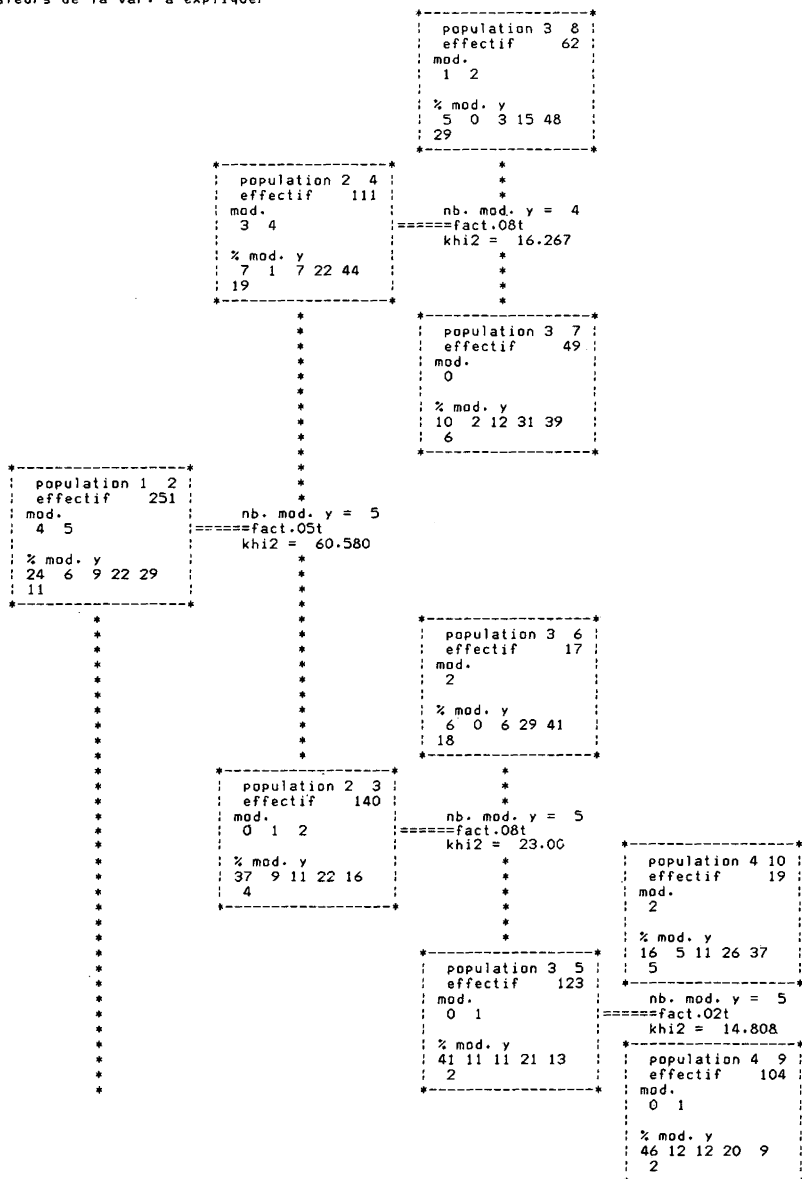
Nous présentons ici l'arbre de segmentation dichotomique, deuxième partie de l'édition des résultats fournie par le programme [voir 3.4(b)]; seuls les découpages (segmentations) correspondant à une valeur du χ^2 significative au niveau 1 % ont été conservés.

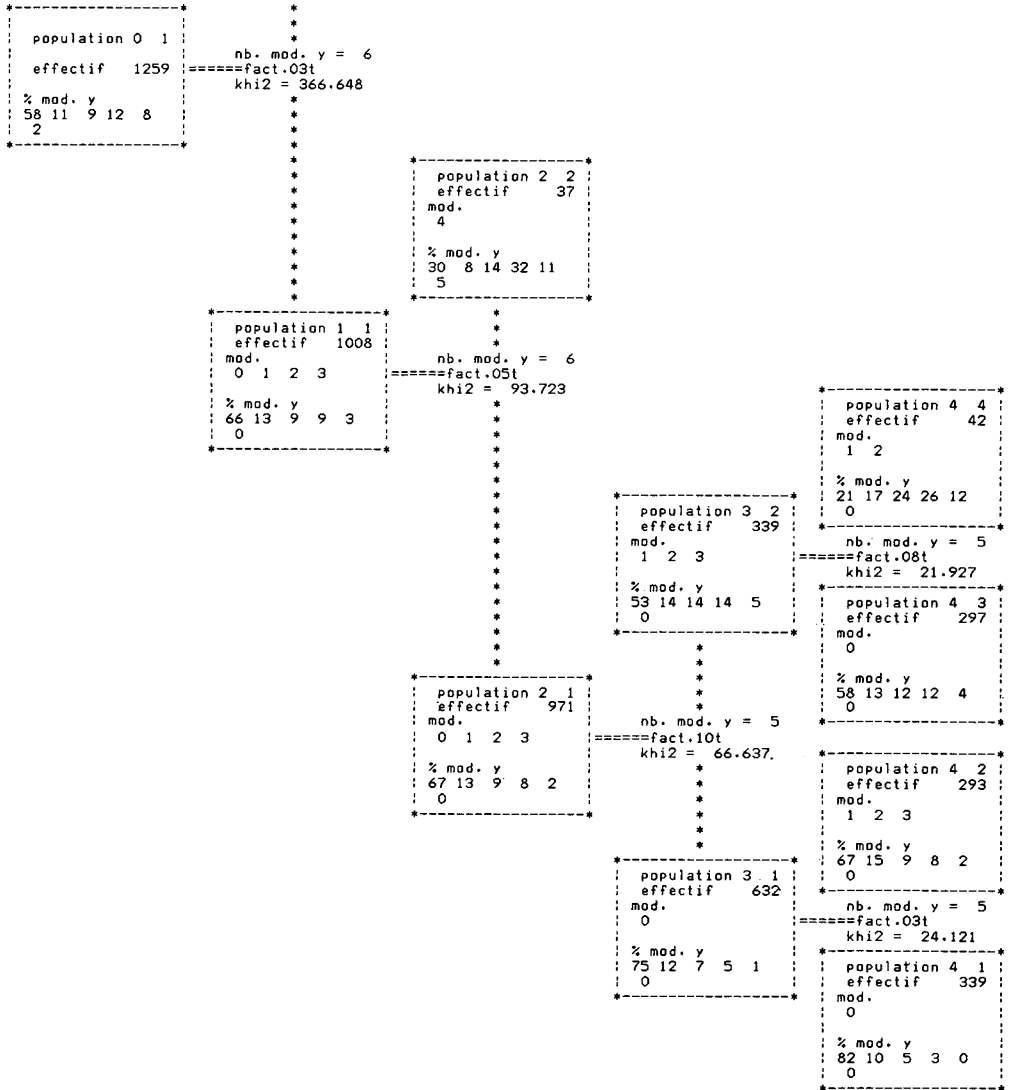
4.4. Commentaires

Les variables explicatives apparaissant comme les plus liées au montant des commandes sont, dans l'ordre, 03 t, 05 t, 10 t et 08 t. L'importance des trois premières avait déjà été mise en évidence par l'analyse des correspondances multiple, mais ce n'était pas le cas pour 08 t. Remarquons que ce dernier facteur n'apparaît qu'au troisième niveau, c'est-à-dire sous la forme d'interaction avec 03 t et 05 t; c'est précisément l'intérêt de la segmentation de mettre en évidence ce type d'interactions que les analyses de correspondances ne peuvent faire apparaître.

Societe toulousaine (1259 clients, 10 transactions, chiffre d'affaires)

- rangement de l'arborescence dichotomique de haut en bas par ordre décroissant de moyenne des valeurs de la var. à expliquer





BIBLIOGRAPHIE

1. A. BACCINI, *Aspect synthétique de la segmentation et traitement de variables qualitatives à modalités ordonnées*, Thèse de 3^e cycle, Université Paul-Sabatier, Toulouse, 1975.
2. A. BACCINI, *Segmentation sur variables à modalités ordonnées*, Note aux C.R.A.S., Paris, tome 281, série A, 1975, p. 249-252.
3. A. BACCINI et A. POUSSE, *Segmentation aux moindres carrés : un aspect synthétique*, *Revue de Statistique Appliquée*, vol. 23, n° 3, 1975, p. 17-35.
4. R. E. BARLOW, D. J. BARTHOLOMEW, J. M. BREMNER et H. D. BRUNK, *Statistical inference under Order Restriction*, Wiley, 1972.
5. W. A. BELSON, *Matching and Prediction on the Principle of Biological Classification*, *Applied Statistics*, vol. 8, 1959, p. 65-75.
6. P. BERTIER et J. M. BOUROCHE, *Analyse des données multidimensionnelles*, P.U.F., Paris, 1975.
7. J. M. BOUROCHE et M. TENENHAUS, *Quelques méthodes de segmentation*, R.A.I.R.O.-Rech. Op., vol. 4, n° 2, 1970, p. 29-42.
8. J. C. CELLARD, B. LABBE et G. SAVITSKY, *Le programme E.L.I.S.E.E., présentation et applications*, Métra, vol. 6, n° 3, 1967.
9. J. DEHEDIN, *Discrimination sur variables qualitatives*, Thèse de 3^e cycle, Université de Paris-VI, 1975.
10. A. FIELDING et C. A. O'MUIRCHARTAIGH, *Binary Segmentation in Survey Analysis with Particular Reference to AID*, *The Statistician*, vol. 26, n° 1, 1977, p. 17-28.
11. M. HUGUES, *Segmentation et typologie*, Bordas, Paris, 1970.
12. G. V. KASS, *An exploratory Technique for Investigating Large Quantities of Categorical Data*, *Applied Statistics*, vol. 29, n° 2, 1980, p. 119-127.
13. J. B. KRUSKAL, *Non metric multidimensional scaling: a numerical method*, *Psychometrika*, vol. 29, n° 2, 1964, p. 115-129.
14. P. MEUNIER, E. DIDAY et J. P. RASSON, *Méthode et algorithme de sélection typologique de paramètres*, R.A.I.R.O.-Rech. Op., vol. 19, n° 4, 1985, p. 351-373.
15. P. MEUNIER, P. BAUFAYS et J. P. RASSON, *Nouveau critère de segmentation pour des variables à expliquer qualitative ordinale et quantitative multidimensionnelle*, *Statistique et Analyse des données*, vol. 10, n° 3, 1985, p. 50-67.
16. MODULAD, *Bibliothèque fortran pour l'analyse des données : document de normalisation*, novembre 1983.
17. J. N. MORGAN et J. A. SONQUIST, *Problems in the analysis of Survey Data, and a Proposal*, J.A.S.A., vol. 58, 1963, p. 415-433.
18. K. VO-KHAC et P. T. NGHIEM, *Étude sur les aspects théoriques et pratiques de la segmentation aux moindres carrés*, R.A.I.R.O.-Rech. Op., vol. 2, n° 1, 1968, p. 77-90.