

CAHIER DU LAMSADE

Laboratoire d'Analyse et Modélisation de Systèmes pour l'Aide à la Décision

(Université Paris IX Dauphine)

Equipe de Recherche Associée au C.N.R.S. N° 656

UNE METHODE DE CODAGE ADAPTATIF EN TYPOLOGIE :

LE PROGRAMME CATY

N° 20-1978

C. DUPONT-GATELMAND

Septembre 1978

SES4 3719

80[860

<541583>



SOMMAIRE

	<u>Pages</u>
<u>ABSTRACT</u>	I
<u>RESUME</u>	II
1. <u>PRESENTATION DU PROBLEME</u>	1
2. <u>SCHEMA DE RESOLUTION</u>	4
2.1 Le processus interne	4
2.1.1 Recherche d'un codage intermédiaire indépendant du type de variable considérée	4
2.1.2 Recherche d'un nouveau codage fonction de la nature des différentes variables	5
2.2 Le processus externe	6
3. <u>UTILISATION DU PROGRAMME</u>	8
3.1 Caractéristiques du programme	8
3.2 Introduction des données	8
3.3 Lecture des résultats	13
4. <u>EXEMPLE DE REFERENCE</u>	14
4.1 Les entrées sur l'exemple de référence	14
4.2 Les sorties sur l'exemple de référence	16
ANNEXE I : BORDEREAU D'ENTREE DES DONNEES	18
ANNEXE II : EXEMPLE DE SORTIES	20
ANNEXE III : ORGANISATION DES SOUS-PROGRAMMES	25
1. Sous-programme HASARD	25
2. Sous-programme TRI	25
3. Sous-programme ORDO	25
4. Sous-programme QUANTA	25
5. Sous-programme INVMAT	25
<u>REFERENCES</u>	26

AN OPTIMAL SCALING METHOD IN CLUSTER ANALYSIS :
THE PROGRAM CATY

ABSTRACT

A method (*) is presented which extends a non hierarchical clustering analysis to the situation where the variables may be measured at a variety of levels (nominal, ordinal, or interval).

The method, which is purely descriptive, scales the observations on each variable, within the restriction imposed by the variable's measurement characteristics, so that a specific criterion W is minimum. W is stated to measure the partitioning of the individuals set I into natural and homogeneous groups and to find their most representative elements.

In other words, the problem is to determine at the same time, the partition of I , its representative samplings and the optimal scaling of the data such that W is minimized.

(*) For further information on this method, the reader is invited to consult my paper : "Une méthode de classification automatique sur variables hétérogènes", Revue de Statistiques Appliquées, n° 2, 1979.

UNE METHODE DE CODAGE ADAPTATIF EN TYPOLOGIE :
LE PROGRAMME CATY

RESUME

CATY est une méthode (*) de classification non hiérarchique sur variables hétérogènes (qualitative nominale, ordinale ou quantitative). Elle est purement descriptive et permet la constitution de groupes homogènes dans la population étudiée en respectant la nature de chacune des variables. La procédure de détermination des classes est alternative à l'instar de l'algorithme des Nuées Dynamiques.

Ainsi, le problème est donc de trouver simultanément la partition de l'ensemble I des individus, ses éléments représentatifs et le codage optimal des données au sens d'un certain critère W qui est minimisé.

(*) Pour une information approfondie sur cette méthode, se reporter à mon article "Une méthode de classification sur variables hétérogènes", Revue de Statistiques Appliquées, n° 2, 1979.

1. PRESENTATION DU PROBLEME

La plupart des méthodes d'analyse des données ont été élaborées pour utiliser un type particulier de données (qualitatives, ordinales ou quantitatives). Or, dans la réalité, le praticien désire souvent utiliser conjointement, à l'aide d'une seule méthode, des données à la fois qualitatives (situation de famille, sexe, catégories socio-professionnelles, etc.), ordinales (préférences pour des produits, attitudes à l'égard de phénomènes sociaux, etc.) et quantitatives (revenu, chiffre d'affaires, etc.).

Afin d'étudier et de traduire de telles observations, il est indispensable de considérer le problème à la base et de s'interroger sur le codage des données, inévitable pont entre celles-ci et une méthode de traitement n'acceptant qu'un seul type homogène de variable. Déjà abordé en analyse des correspondances et en analyse canonique, ce problème de codage n'a été que peu étudié en classification automatique. Pour cette raison, nous avons développé une méthode de typologie avec codage adaptatif.

Le problème peut se poser de la façon suivante : trouver simultanément la partition de l'ensemble d'individus et le codage optimal des données au sens d'un certain critère W mesurant l'adéquation des individus aux différentes classes et respectant les contraintes de nature des données (qualitatives, ordinales et quantitatives).

Le codage n'est plus fixé une fois pour toutes avant toute méthode d'analyse comme cela était fait jusqu'à présent, au risque d'une importante perte d'information.

Pour utiliser CATY, il faut disposer en entrée d'un tableau X dont les particularités sont les suivantes :

X est de dimension $(NSUJ, NVAR)$ où $NSUJ$ désigne le nombre d'individus considérés et $NVAR$ le nombre total de variables caractérisant ceux-ci (on pose $NVAR = p$).

Ces variables peuvent être de trois natures :

- qualitatives nominales : elles sont "codées a priori", de façon absolument quelconque ;
- qualitatives ordinales : elles sont "codées a priori" de façon absolument quelconque ;
- quantitatives : elles sont prises en compte sans aucune modification préalable.

Remarque importante : l'utilisation de CATY est essentiellement justifiée par des données hétérogènes pour lesquelles l'approche typologique présente un intérêt certain.

Formulation du problème

En désignant par :

$$G_{(k,p)} = \begin{pmatrix} G_1 \\ \vdots \\ G_s \\ \vdots \\ G_k \end{pmatrix} \quad \begin{array}{l} \text{la matrice des noyaux des classes de la parti-} \\ \text{tion } P \\ \text{avec } G_s \subset \mathbb{R}^p \quad \forall s \in \overline{[k]} \quad (*) \end{array}$$

$\mathcal{IP}_k$ l'ensemble des partitions en k classes de I

$\mathcal{IG}_k$ l'ensemble des k -uplets admissibles des noyaux.

Le problème est alors de minimiser W application définie sur $\mathbb{R}^p \times \mathcal{IG}_k \times \mathcal{IP}_k$ et à valeurs dans $\mathbb{R}^+$ telle que :

(*) Notation : $\overline{[k]} = \{1, \dots, k\}$.

$$W(c, G, P) = \sum_{l=1}^k \sum_{i \in [P_l]} \|G_l - c_i\|^2$$

où c_i représente le vecteur codage associé à l'individu i . Il s'agit donc de :

$$\begin{aligned} & \text{Min } W(c, G, P) \\ & \left\{ \begin{array}{l} c \in \mathbb{R}^p \\ G \in IG_k \\ P \in IP_k \end{array} \right. \end{aligned}$$

2. SCHEMA DE RESOLUTION

La solution est obtenue par l'articulation itérative de deux processus : "le processus interne" qui permet l'obtention d'un nouveau codage à partition fixée et le "processus externe" qui permet la constitution d'une nouvelle partition à partir de la solution issue du processus interne.

2.1 Le processus interne

Chaque phase de ce processus peut être décomposée en deux étapes :

- recherche d'un codage intermédiaire indépendant du type de variable considérée ;
- recherche d'un nouveau codage fonction de la nature des différentes variables.

2.1.1 Recherche d'un codage intermédiaire indépendant du type de variable considérée

Notons : A^0 la matrice des codages intermédiaires
 (n, p)
 G^0 la matrice des centres de gravité des k classes
 (k, p) de la partition P^0

$\Phi^0 = \{Y^0 g \mid g \in \mathbb{R}^k\}$ le sous-espace vectoriel engendré par les indicatrices associées à P^0 .

$D_n = \begin{pmatrix} p_1 & & 0 \\ & \ddots & \\ 0 & & p_n \end{pmatrix}$ la matrice des poids des individus.

$A^0 = Y^0 G^0$ est solution du problème :

$$\min_G \| \bar{C}^0 - Y^0 G \|^2$$

Remarque : On utilise la propriété de "conservation du centrage" apportée par le projecteur $T_0 = Y^0 (Y^{0T} D_n Y^0)^{-1} Y^{0T}$, D_n projecteur orthogonal de $\bar{c}_j$ sur Φ^0 .

2.1.2 Recherche d'un nouveau codage fonction de la nature des différentes variables

* Si x_j est une variable qualitative nominale, son codage c_j est un élément du sous-espace vectoriel W_j :

$$W_j = \{c_j \in \mathbb{R}^n / c_j = X_j b_j, b_j \in \mathbb{R}^{m_j}\}$$

où X_j est la matrice des indicatrices associée à la variable j .

On s'intéresse au codage centré $c_j \in W_j^-$ avec $W_j = 1 \oplus W_j^-$.

$\bar{c}_j^1$ est alors solution du problème :

$$\min_{\bar{c}_j \in W_j^- \cap S} \|\bar{c}_j - a_j^0\|^2$$

où $S = \{y \mid \|y\|^2 = 1\}$.

De même, le projecteur $\tilde{T}_j = X_j (X_j^t D_n X_j)^{-1} X_j^t D_n$, projecteur D_n orthogonal de a_j^0 sur W_j conserve le centrage.

* Si x_j est une variable qualitative ordinale, son codage c_j est un élément du cône convexe C_j de $\mathbb{R}^n$ engendré par la matrice d'indicatrices associée à la variable j . C_j est un cône polyédral de $\mathbb{R}^n$.

$$C_j = \{c_j \in \mathbb{R}^n / c_j = X_j b_j, b_j \in \mathcal{C}_j\}$$

où $\mathcal{C}_j = \{b_j \in \mathbb{R}^{m_j} / b_j^1 \leq \dots \leq b_j^k \leq \dots \leq b_j^{m_j}\}$.

On s'intéresse au codage centré $c_j \in C_j^-$ avec $C_j = 1 \oplus C_j^-$.

$\bar{c}_j^1$ est alors solution du problème

$$\min_{\bar{c}_j \in \mathbb{C}_j / S} \|\bar{c}_j - a_j^0\|^2$$

Quand la deuxième étape a été réalisée, on cherche à nouveau la matrice des codages intermédiaires et on obtient le couple solution $(\bar{c}_j^*, g_j^*)$ à la convergence du processus interne.

On montre que :

① le codage $\bar{c}_j^*$ est vecteur propre associé à la plus grande valeur propre de la matrice $(\bar{T}_j \circ T)$.

② pour chaque variable qualitative nominale x_j , le processus interne est en fait comparable à une analyse canonique entre le tableau des indicatrices associées à x_j et celui des indicatrices associées à la partition P^0 .

2.2 Le processus externe

Il est mis en oeuvre au moyen de la fonction d'identification f définie sur $\mathbb{R}^p \times \mathbb{G}_k$ et à valeurs dans $\mathbb{P}_k$.

$$f(c, G) = P$$

$$\text{avec } G = \begin{pmatrix} G_1 \\ \vdots \\ G_1 \\ \vdots \\ G_k \end{pmatrix} \text{ et } P = (P_1, \dots, P_1, \dots, P_k)$$

Elle permet de définir les classes $P_l \quad \forall l \in]k]$ de la nouvelle partition :

$$P_l = \{i \in I / \|G_l - c_i\|^2 \leq \|G_{l'} - c_i\|^2 \quad \forall l' \neq l\}$$

La fonction sert à construire la matrice d'indicatrices associée à la nouvelle partition. Les k classes sont construites de manière à être plus homogènes qu'à l'itération précédente.

Le "processus externe" définit alors la suite des partitions $P_1, P_2, \dots, P_N$.

Les deux "processus" sont ainsi étroitement imbriqués. Ils convergent vers un minimum local. Les codages obtenus pour les différentes variables respectent leur nature.

Remarque 1 : les variables initialement quantitatives peuvent soit être conservées sous leur première forme, soit être transformées par régression polynômiale dont le degré est choisi a priori.

Remarque 2 : la constitution du processus interne a l'avantage et l'intérêt de déterminer, pour les variables qualitatives nominales, des codages indépendants des codes fixés a priori avant l'analyse. Cette caractéristique n'est que partiellement vérifiée pour les variables qualitatives ordinales.

3. UTILISATION DU PROGRAMME

3.1 Caractéristiques du programme

On désigne par :

NSUJ = le nombre total d'individus considérés

NPAR = le nombre total de variables les caractérisant

NCLASS = le nombre de classes de la partition P

IM = le degré du polynôme de transformation des variables quantitatives + 1

MODMAX = le nombre de modalités maximum pour les variables qualitatives.

La version actuelle du programme est établie pour permettre la classification :

- d'une population de 300 individus
- caractérisés par 20 variables "hétérogènes"
- en un maximum de 6 classes.

Les variables quantitatives peuvent être transformées par des polynômes d'ordre maximal égal à 5. Les variables qualitatives peuvent avoir un nombre de modalités inférieur ou égal à 8.

3.2 Introduction des données (cf. annexe 1)

1ère carte : Carte de paramètres généraux

en format 20 I 4 (cadré à droite), on écrit successivement :

NSUJ, NPAR, NCLASS, NTIR, LECCLA, INOM, ITERMX, MODMAX, LEC, IM, ISOR, IEDIT, LNEU, MODMOX.

En col. 1.4	NSUJ	= le nombre d'individus considérés
col. 5.8	NPAR	= le nombre de variables les caractérisant
col. 9.12	NCLASS	= le nombre de classes désiré
col. 13.16	NTIR	= le nombre de tirages au hasard
col. 17.20	LECCLA	= 0 si tirage au hasard) ≠ 0 si partition a priori) au début du processus
col. 21.24	INOM	= 0 quand on n'attribue pas de nom aux individus ≠ 0 sinon
col. 25.28	ITERMX	= nombre d'itérations maximum
col. 29.32	MODMAX	= nombre de modalités maximum pour les variables qualitatives
col. 33.36	LEC	= unité de lecture des données Par défaut, LEC = lecteur de cartes
col. 37.40	IM	= nombre de paramètres à estimer pour les variables quantitatives (ou degré du polynôme de transformation + 1)
col. 41.44	ISOR	= unité de sortie pour variables codées + classes d'affectation Par défaut, ISOR = imprimante
col. 45.48	IEDIT	= nombre d'individus à imprimer en sortie (sert de contrôle)
col. 49.52	LNEU	= fichier contenant le numéro des individus et leur classe d'affectation. A utiliser dans le cas d'une partition a priori
col. 53.56	MODMOX	= valeur maximale des modalités des variables qualitatives.

2ème carte : Carte indiquant le format de lecture des données (FOMAT)

Cette carte permet d'inclure en donnée le format de lecture des caractéristiques des individus et évite la compilation systématique du programme principal lorsqu'un nouveau jeu de données se présente.

col. 1 à 4 (cadré à droite)

on indique le nombre de mots du format ; c'est le nombre de caractéristiques alphanumériques du format divisé par 4 et arrondi à l'entier supérieur.

col. 5 et suivantes

on écrit le format proprement dit. Celui-ci commence par une parenthèse ouverte (en col. 5) et se termine par une parenthèse fermée.

Exemple : 3 (3 4 X , 1 2 F 1 . 0)
 ↓
 $\frac{12}{4} = 3$ 12 caractères

Deux cas peuvent être traités dans le programme selon que les éléments de la population étudiée sont ou non repérés par un nom ou un signe particulier. Cette différenciation possible est traduite dans le format.

Exemple : 3 (I 2 , 1 2 F 4 . 2)
 ↓
 Nom ou
 signe permettant
 d'identifier les
 éléments
 Se lit en format entier.

Dans le cas contraire, on aurait par exemple :

 2 (1 2 F 4 . 2)

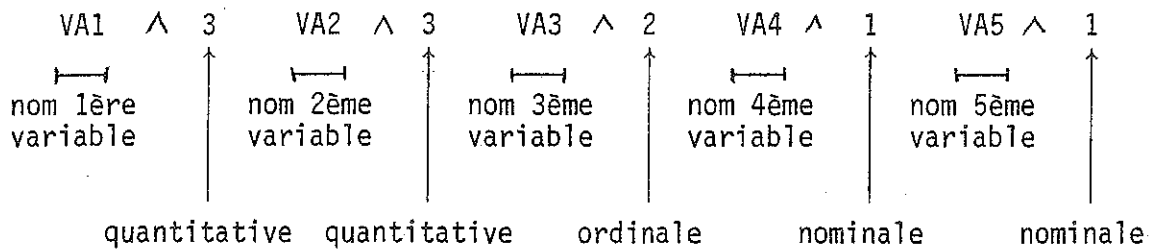
et le programme affecte alors automatiquement un numéro d'ordre aux éléments initialement non identifiables.

3ème carte : Lecture des noms des variables et de leur nature (sur une ou plusieurs cartes) en format (16(A4 I1)).

Par carte, on peut donner au maximum les noms et natures de 16 variables différentes. Le nom est donné sur 4 caractères alphanumériques et la nature est déterminée par un entier occupant le caractère suivant.

- On note le type -
- qualitatif nominal par : 1
 - qualitatif ordinal par : 2
 - quantitatif par : 3

Exemple : Prenons le cas de 5 variables dont les deux premières sont quantitatives, la troisième ordinale et les deux dernières nominales.



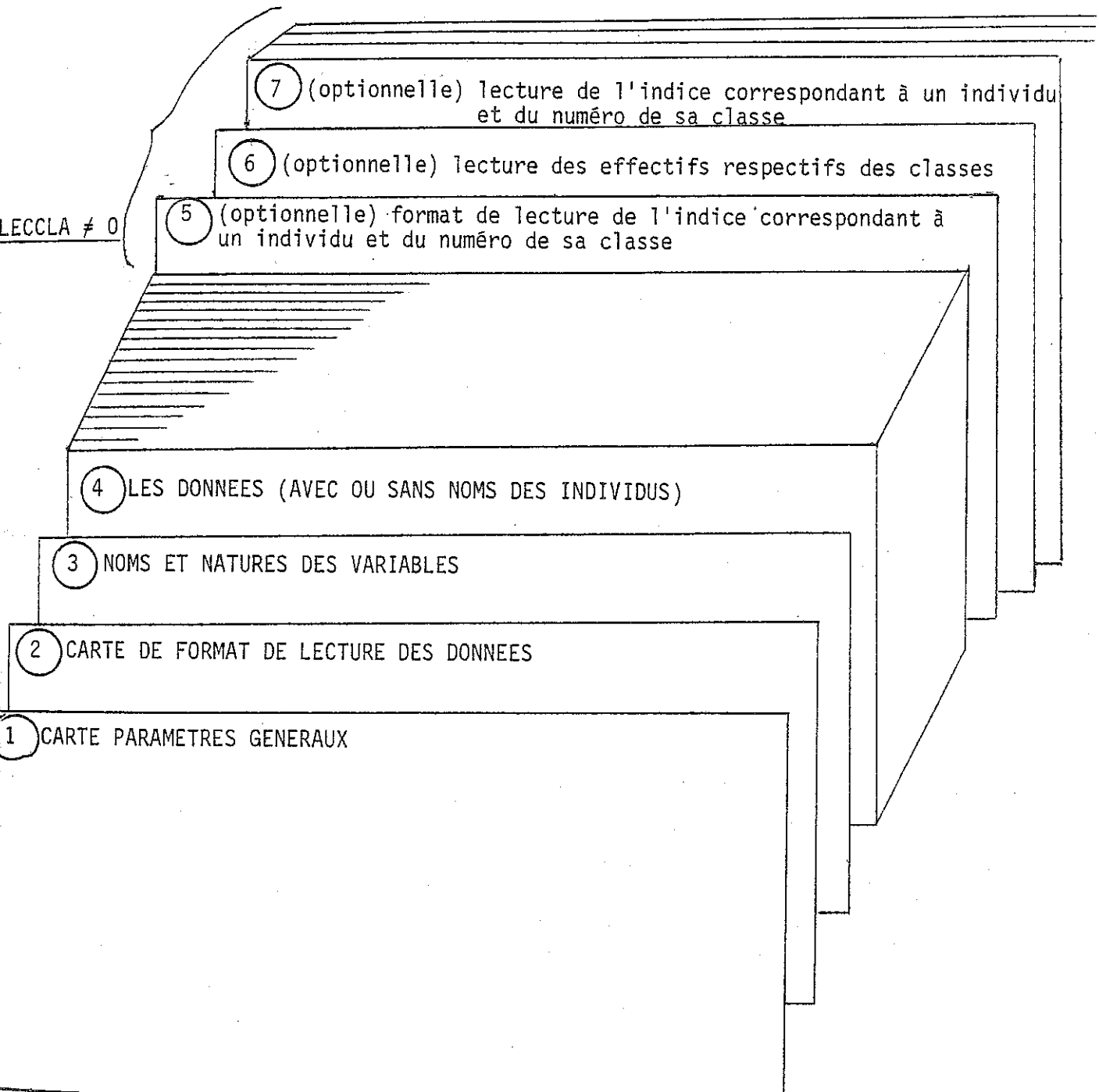
Si LECCLA $\neq 0$, il est nécessaire de décrire la partition choisie a priori et par conséquent d'ajouter les trois cartes suivantes :

4ème carte (optionnelle) : elle donne le format de lecture FMT qui permet de lire l'indice correspondant à chaque individu ainsi que le numéro de la classe à laquelle il appartient.

5ème carte (optionnelle) : en format (20I4), on lit les effectifs respectifs des classes constituant la partition de départ.

6ème carte (optionnelle) : en format FMT (déterminé par la 4ème carte), on lit, pour chaque individu, l'indice correspondant ainsi que le numéro de la classe à laquelle il appartient.

D'une manière schématique, le paquet de cartes de données nécessaire à l'utilisation de la méthode CATY doit se présenter sous la forme suivante :



3.3 Lecture des résultats

On lit, en premier lieu, un rappel des données ; on a ainsi une récapitulation concernant :

- le format de lecture des données ;
- les noms et natures des variables ;
- une description des premiers individus et de leurs caractéristiques pour contrôler la prise en compte des données.

Pour chaque tirage au hasard, on a :

- la description de la partition établie a priori ;
 - la valeur du critère correspondant à chacune des itérations ;
 - la description de la partition à la convergence du critère ;
- chaque classe est explicitée par les noms ou les numéros repérant les individus la composant ;
- la transformation des codages de chaque variable qualitative.

On peut ainsi lire :

VARIABLE NUMERO N SA NATURE :

<u>MODALITES</u>	<u>VALEURS CENTREES REDUITES</u>	<u>CODAGE OBTENU</u>
1	x	c_1
2	x	c_2
$\vdots$	$\vdots$	$\vdots$
M	x	c_M

Pour les variables quantitatives, on donne les coefficients de la régression polynômiale si celle-ci a été demandée.

4. EXEMPLE DE REFERENCE

Cet exemple porte sur des données provenant d'un laboratoire de Rhône-Poulenc.

Nature des données : 275 malades ont été examinés très attentivement et des dossiers fort complets ont été constitués : l'information recueillie est conséquente... 250 variables ont été prises en compte. Toutefois, sur les conseils du médecin responsable de l'étude, on en a sélectionné 14 qui lui semblaient particulièrement intéressantes pour une tentative de classification des malades. Elles nous renseignent sur les caractéristiques de la maladie (diagnostic, symptômes de début) et sur l'effet produit par l'absorption du produit T (résultat, tolérance). On observe notamment l'évolution de certains troubles après la prise du médicament (variables 3 à 11).

Les natures des variables sont diverses :

- les variables 1 à 10 sont qualitatives ordinales ;
- les variables 11 à 13 sont qualitatives nominales ;
- la variable 14 est quantitative ; elle est exprimée en milligrammes.

Le problème : On cherche à former des groupes homogènes de malades qui auraient éventuellement des réactions analogues à la prise du médicament T compte tenu des troubles spécifiques qui les caractérisent. Ceci permettrait peut être d'établir des contre-indications à l'utilisation de T si l'on découvrait des tolérances mauvaises pour certaines sous-populations.

Les variables sont par nature hétérogènes : on aurait pu, en découpant la variable quantitative, utiliser la méthode des Nuées Dynamiques en prenant une distance du χ^2 sur modalités, mais alors, on aurait perdu l'information ordinale apportée par les 10 premières variables. On applique donc la méthode CATY "avec codage adaptatif".

4.1 Les entrées sur l'exemple de référence (voir bordereau des données)

Les données fournies par le laboratoire sont les suivantes :

{ NSUJ = 275
NPAR = 14
MODMAX = 4
MODMOX = 4

Par ailleurs, on choisit a priori :

- d'obtenir une partition en 3 classes :
{ NCLASS = 3
- de réaliser 4 tirages aléatoires pour initialiser le processus :
{ NTIR = 4
LECCLA = 0
- de limiter à 20 itérations chaque processus
{ ITERMX = 20
- de ne pas transformer les variables quantitatives
{ IM = 1

Enfin, on donne les paramètres permettant la lecture des fichiers et l'écriture des résultats

{ LEC = 5
ISOR = 40
LNEU = 50

Afin de vérifier si le fichier d'entrée a bien été lu, on peut faire écrire un certain nombre de lignes correspondant aux premiers individus

{ IEDIT = 10

4.2 Les sorties sur l'exemple de référence

Analyse des résultats obtenus par une typologie avec codage adaptatif sur variables hétérogènes

Nous sommes partis d'un tirage au hasard sur la population déterminant ainsi une partition à trois classes.

Nous avons examiné 4 tirages au hasard et nous avons choisi la solution (C^*, G^*, P^*) qui donnait un critère W minimum. Le programme donne à la fois la description des classes (repérage des individus) et le codage des modalités des différentes variables.

Analyse de la partition P^* : On l'obtient par des tableaux de contingence réalisés entre la partition P^* et chacune des variables. On décrit ainsi les trois classes avec leurs caractéristiques respectives sur les variables considérées. On remarque que les trois classes sont bien différenciées par le résultat de l'expérience et la tolérance au médicament :

- la classe 1 rassemble des individus sur lesquels le médicament T n'a eu aucun effet bénéfique puisqu'ils conservent les mêmes troubles qu'avant la prise du médicament T . De plus, ils l'ont mal supporté ;
- la classe 2 est une classe intermédiaire : la tolérance est bonne mais les individus composant cette classe n'étaient atteints que de troubles très légers. On constate toutefois que le médicament n'induit aucun déséquilibre supplémentaire ;
- la classe 3 comprend des malades plus sérieux et pour lesquels le médicament T est le remède miracle.

Une analyse détaillée des 36 dossiers médicaux caractérisant la première classe permettrait sans doute au médecin d'établir certaines contre-indications à la prise du médicament T .

ANNEXE I

BORDEREAU D'ENTREE DES DONNEES

(419E)

1951.5.2

16(94, 11)

Remarque[illegible]

ANNEXE II

EXEMPLE DE SORTIES

DESCRIPTION DE LA PARTITION

CLASSE NUMERO 1	13	20	32	38	39	42	44	53	55	56	58	61	68	92	93	108	113	114	120	120
	130	134	137	139	138	145	147	205	211	216	217	228	232	235	240	265				
CLASSE NUMERO 2	1	2	3	4	6	7	8	9	12	14	15	17	18	19	21	22	23	24	25	26
	28	29	30	31	33	35	36	47	48	54	57	59	60	62	63	64	65	69	70	71
	73	74	76	77	79	80	81	82	83	84	85	86	88	91	94	95	96	97	100	101
	103	104	105	106	109	116	117	119	121	122	123	124	126	127	131	135	147	152	156	159
	165	166	167	168	170	171	172	173	175	176	177	178	179	180	183	184	185	186	189	190
	191	192	193	194	196	199	200	201	202	203	204	206	207	208	209	210	212	213	215	219
	220	221	222	223	224	226	227	229	230	231	236	238	239	243	244	245	246	247	248	249
	250	251	252	253	256	257	258	259	260	263	264	266	267	268	269	270	271	272	273	
CLASSE NUMERO 3	5	10	11	16	27	34	37	40	41	43	45	46	49	50	51	52	66	67	72	75
	78	87	89	96	98	99	102	107	110	111	112	115	118	125	129	132	133	136	138	140
	141	142	143	144	145	146	148	149	150	151	153	154	155	157	158	160	161	162	163	164
	169	174	181	182	187	214	218	225	233	234	237	241	242	254	255	261	262	274	275	

TRANSFORMATION DES CODAGES

TABLE NUMERO 1 ORDINALE	MODALITES	CENT RED.	NOUVEAU CODAGE
	1.00	-1.58	-1.52
	2.00	-0.74	-1.52
	3.00	0.21	0.66
	4.00	1.15	0.66
TABLE NUMERO 2 ORDINALE	MODALITES	CENT RED.	NOUVEAU CODAGE
	1.00	-1.93	-2.29
	2.00	-0.01	0.44
	3.00	0.72	0.44
TABLE NUMERO 3 ORDINALE	MODALITES	CENT RED.	NOUVEAU CODAGE

1.00	-1.49	-0.93
2.00	-0.58	-0.93
3.00	0.33	-0.12
4.00	1.24	1.56

VARIABLE NUMERO 4 ORDINALE	MODALITES	CENT RED.	NOUVEAU CODAGE
	1.00	-0.70	-0.62
	2.00	0.38	-0.30
	3.00	1.66	1.98

VARIABLE NUMERO 5 ORDINALE	MODALITES	CENT RED.	NOUVEAU CODAGE
	1.00	-1.34	-0.57
	2.00	-0.45	-0.57
	3.00	0.43	-0.57
	4.00	1.31	1.76

VARIABLE NUMERO 6 ORDINALE	MODALITES	CENT RED.	NOUVEAU CODAGE
	1.00	-0.75	-0.35
	2.00	0.71	-0.35
	3.00	2.17	2.86

VARIABLE NUMERO 7 ORDINALE	MODALITES	CENT RED.	NOUVEAU CODAGE
	1.00	-1.18	-0.52
	2.00	-0.31	-0.52
	3.00	0.58	-0.52
	4.00	1.43	1.91

VARIABLE NUMERO 8 ORDINALE	MODALITES	CENT RED.	NOUVEAU CODAGE
	1.00	-0.68	-0.47
	2.00	0.84	0.07
	3.00	2.35	3.01

VARIABLE NUMERO 9 ORDINALE	MODALITES	CENT RED.	NOUVEAU CODAGE
	1.00	-1.93	-0.79
	2.00	-0.14	-0.79
	3.00	0.75	1.27
	4.00	1.64	1.27

VARIABLE NUMERO 10 ORDINALE	MODALITES	CENT RED.	NOUVEAU CODAGE
	1.00	-0.58	-0.52
	2.00	1.09	0.72
	3.00	2.76	3.26

BLE N°11 QUALITATIVE			
	MODALITES	CENT. RED.	NOUVEAU CODAGE
	1.00	-1.18	-0.44
	2.00	-0.26	-0.95
	3.00	0.66	-0.13
	4.00	1.57	2.45

BLE N°12 QUALITATIVE			
	MODALITES	CENT. RED.	NOUVEAU CODAGE
	1.00	-1.06	-0.11
	2.00	0.07	-1.67
	3.00	1.19	1.07

BLE N°13 QUALITATIVE			
	MODALITES	CENT. RED.	NOUVEAU CODAGE
	1.00	-0.85	-0.83
	2.00	0.07	1.26
	3.00	1.00	1.30
	4.00	1.92	-0.38

BLE N°14 QUANTITATIVE		COEF. REGRESSION	
.033	0.000		

ANNEXE III

ORGANISATION DES SOUS-PROGRAMMES

1 Sous-programme HASARD

Tirage au hasard d'une partition sur les NSUJ individus étudiés.

NCLASS = nombre de classes dans la partition

KCLASS (i) = indice de la classe de l'individu i

NSCLA (j) = nombre d'individus dans la classe j

RANDU(NCLASS) = tirage au hasard d'un nombre entier entre 1 et NCLASS.

2 Sous-programme TRI

Tri des individus pour l'édition des résultats. Les individus, rangés par classe, apparaissent dans le tableau KLATRI.

3 Sous-programme ORDO

Régression monotone effectuée sur les variables ordinales. Transformation de codage des variables ordinales.

4 Sous-programme QUANTA

Régression polynômiale de degré IM-1 pour les variables quantitatives.

5 Sous-programme INVMAT

Inversion de matrice servant dans QUANTA.

REFERENCES

- [1] BARLOW R.E., BARTHOLOMEW D.J., BREMNER J.M., BRUNK H.D. - "Statistical inference under order restrictions". Wiley, 1972.
- [2] BOUROCHE J.M. - "Développement de l'utilisation des méthodes d'analyse des données qualitatives". Note de travail n° 26, COREF.
- [3] CAILLIEZ F., PAGES J.P. - "Introduction à l'analyse des données". SMASH, 1976.
- [4] CAZES P., BAUMERDER A., BONNEFOUX S., PAGES J.P. - "Codage et analyse des tableaux logiques - Introduction à la pratique des variables qualitatives". Cahier n° 27, BUR0, 1977.
- [5] DE LEEUW J. - "A normalized cone regression approach to alternating least squares algorithm". University of Leiden, The Netherlands, 1976.
- [6] DIDAY E. - "La méthode des nuées dynamiques et la reconnaissance des formes". Cahiers de l'IRIA, 1970.
- [7] DROUET D'AUBIGNY G. - "Description statistique des données ordinales : analyse multidimensionnelle". Thèse de 3ème cycle, Université de Grenoble, 1975.
- [8] DUPONT-GATELMAND C. - "Deux méthodes d'analyse des données qualitatives : typologie avec codage adaptatif et modèle de décomposition additive des préférences". Thèse de 3ème cycle, Université Paris IX, avril 1978.
- [9] HAYASHI L. - "On the quantification of qualitative data from the mathematico-statistical point of view". Annals. Inst. Stat. Math 2, 1950.

- [10] KRUSKAL J.B. - "Analysis of factorial experiments by estimating monotone transformation of the data". Journal of the Royal Statistical Society, Series B, 27, 1965.
- [11] KRUSKAL J.B. - "Monotone regression, continuity and differentiability properties". Psychometrika, Tome 36, n° 1, 1971, p. 57-62.
- [12] LECHEVALLIER Y. - "Classification automatique optimale sous contrainte d'ordre total". Rapport de recherche n° 200, IRIA-LABORIA, 1976.
- [13] MASSON M. - "Analyse non linéaire des données". Compte rendu à l'Académie des Sciences de Paris, Tome 278, Série A, 1974, p. 803-806.
- [14] SAPORTA G. - "Liaison entre plusieurs ensembles de variables et codages de données qualitatives". Thèse de 3ème cycle, Université de Paris VI, 1975.
- [15] SAPORTA G. - "Le traitement de variables qualitatives par codage". Note de travail n° 11, COREF.
- [16] TENENHAUS M. - "Etude des programmes MORALS et CORALS de F.W. YOUNG, J. de LEEUW et Y. TAKANE". Note de travail n° 15, COREF.
- [17] TENENHAUS M. - "Analyse en composantes principales d'un ensemble de variables nominales et numériques". Note de travail n° 18, COREF.
- [18] YOUNG F.W., DE LEEUW J., TAKANE Y. - "How to use principals". Psychometrika Laboratory, 1975.
- [19] YOUNG F.W., DE LEEUW J., TAKANE Y. - "Regression with qualitative and quantitative variables : An Alternating Least Squares method with optimal scaling features". Psychometrika, 41, 1976.

- [20] YOUNG F.W., DE LEEUW J., TAKANE Y. - "Additive structure in qualitative Data : an Alternating Least Squares method with optimal scaling features (ADDALS)". Psychometrika, 41, 1976.
- [21] YOUNG F.W., DE LEEUW J., TAKANE Y. - "Non metric individual differences multidimensional scaling : an Alternating Least Squares method with optimal SCALing features (ALSCAL)". Psychometrika, 42, 1977.