

CAHIER DU LAMSADE

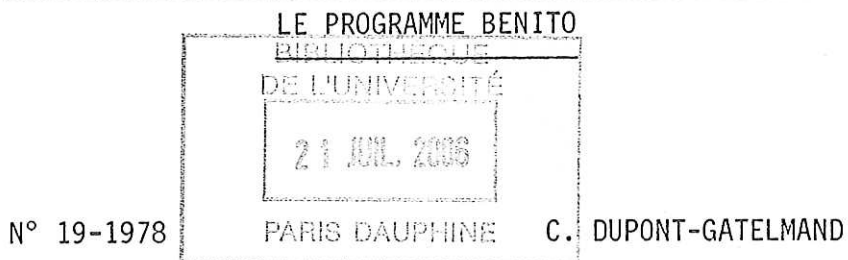
Laboratoire d'Analyse et Modélisation de Systèmes pour l'Aide à la Décision

(Université Paris IX Dauphine)

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



UN MODELE DE DECOMPOSITION ADDITIVE DES PREFERENCES :



Septembre 1978

Br. 1703-19

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
3. <u>CARACTERISTIQUES DU PROGRAMME</u>	7
4. <u>UTILISATION DU PROGRAMME</u>	9
4.1 Introduction des données	9
4.2 Lecture des résultats	12
4.2.1 Dans le cas général	12
4.2.2 Dans le cas de l'"option locale"	13
5. <u>EXEMPLE DE REFERENCE</u>	14
5.1 Les entrées sur l'exemple de référence	15
5.2 Les sorties sur l'exemple de référence	17
5.3 Autre exemple	19
ANNEXE 1 : ORGANISATION DES SOUS-PROGRAMMES LISTE ET FONCTIONS DES DIFFERENTS SOUS-PROGRAMMES	21
<u>REFERENCES</u>	24

AN ADDITIVE DECOMPOSITION MODEL IN PREFERENCE ANALYSIS :
THE PROGRAM BENITO (*)

ABSTRACT

This method belongs to the multiattribute attitude model class. Its specificity is to explain preferences data using the interactions of stimuli and judge characteristics.

In other words, the note y_{ij} given by the judge i on the stimulus j is approximated by the additive effects corresponding to the mix of the characteristics of j and the ones of i .

Therefore, the problem is to find this set of interactions weights, each one adjusting the global preference average according to the couple of characteristics it defines.

(*) BENITO is not an acronym : it is only a private joke.

UN MODELE DE DECOMPOSITION ADDITIVE DES PREFERENCES :
LE PROGRAMME BENITO

RESUME

BENITO (*) est une méthode d'analyse et d'explication des préférences ; elle s'inscrit dans la classe des modèles d'"Analyse Conjointe" (ou Conjoint Measurement). Toutefois, son intérêt particulier réside dans la prise en compte des caractéristiques propres à chaque individu dans "l'explication" des notes que chacun d'eux a attribué aux différents objets présentés.

On décompose ainsi la note y_{ij} donnée par le juge i à l'objet j en une somme de termes correspondant aux interactions entre les caractéristiques de i et celles de j .

Ainsi le problème est donc de trouver ce tableau de poids, chacun d'eux permettant de corriger la préférence globale moyenne selon les éléments décrivant le couple (i, j) auquel on s'intéresse.

(*) Le nom donné à ce programme est justifié par des raisons personnelles.

1. PRESENTATION DU PROBLEME

Ce modèle permet d'approcher chaque élément y_{ij} d'un tableau de préférences Y par une somme de coefficients dépendant simultanément des variables caractérisant les juges i ($\forall i \in [n]$) (*) et de celles caractérisant les objets classés j ($\forall j \in [p]$). Il détermine ainsi, par approximation au sens des moindres carrés, la note y_{ij} en corrigeant la préférence globale moyenne des n juges sur les objets par l'évaluation de coefficients dépendant des caractéristiques de i et de j .

1.1 Pour utiliser BENITO, il faut disposer en entrée, de trois tableaux Y , X_1 , X_2 dont les particularités sont décrites ci-dessous.

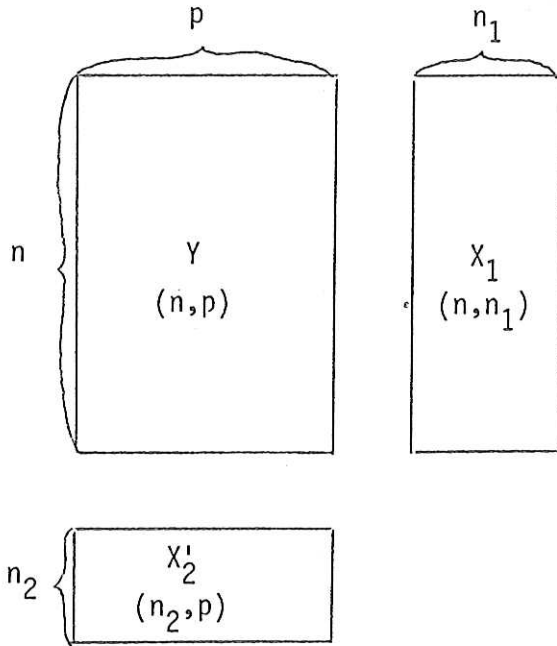
- La matrice Y de dimension (n, p) rassemble les n préordres de préférence émis par les individus sur les p objets. Un jugement est donc représenté par une ligne de la matrice.

- La matrice X_1 de dimension (n, N_1) avec $n > N_1$ sert à caractériser les individus et N_1 désigne le nombre de variables considérées pour les définir. Elle est transformée en un tableau disjonctif complet et devient alors de dimension (n, n_1) avec $n > n_1$ où n_1 est le nombre total de modalités possibles de X_1 sur l'ensemble des variables qualitatives.

- La matrice X_2 de dimension (p, N_2) avec $p > N_2$ caractérise les p objets. Comme X_1 , elle est transformée en tableau disjonctif complet et sa dimension est alors (p, n_2) : n_2 correspond au nombre maximum de modalités de X_2 sur l'ensemble des N_2 variables qualitatives.

Afin de mieux "visualiser" les données, on peut les représenter de la manière suivante :

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



Remarque importante

Dans Y , les données se présentent sous forme d'évaluations sur des échelles. L'utilisateur doit savoir que tout jugement recueilli sous forme de rang est assimilé à une échelle cardinale. De plus, il ne faut pas oublier que les variables composant X_1 ou X_2 sont qualitatives. Toute variable quantitative sera donc préalablement "découpée" pour être transformée en variable qualitative à modalités.

1.2 Formulation du problème

Il s'agit de trouver T et $\tilde{t}$ tels que :

$$(n_1, n_2) \quad (n_2, 1)$$

$$Y \underset{\approx}{=} X_1 T + 1 \tilde{t}' X_2'$$

$$(n, p) \quad (n, n_1) \quad (n_1, n_2) \quad (n_2, p) \quad (n, 1) \quad (1, n_2) \quad (n_2, p)$$

où " $\underset{\approx}{=}$ " désigne l'approximation au sens des moindres carrés

ou, ce qui est équivalent :

Il s'agit de trouver T^* tel que :

$$Y \approx \begin{matrix} X_1^* & T^* & X_2' \\ (n,p) & (n,n_1+1) & (n_1+1,n_2) & (n_2,p) \end{matrix} \quad (P_1)$$

avec $X_1^* = \begin{pmatrix} X_1 & | & 1 \\ (n,n_1+1) & (n,n_1) & (n,1) \end{pmatrix}$ et $T^* = \begin{pmatrix} T \\ \tilde{t} \end{pmatrix} \begin{matrix} n_1 \\ 1 \end{matrix}$

1.3 Interprétation des matrices T et $\tilde{t}$

- $\tilde{t}$ est la préférence globale moyenne des n juges sur les n_2 caractéristiques des p objets.

- T est la matrice des coefficients d'influence qui tient compte des caractéristiques de l'individu qui juge et de celles de l'objet classé pour corriger $\tilde{t}$.

Avec ce modèle, la note y_{ij} donnée par le juge i à l'objet j est approximée par une note moyenne $\tilde{t}$ calculée sur l'ensemble des juges et corrigée par une somme de coefficients - ceux de la matrice T - correspondant simultanément aux caractéristiques de i et à celles de j.

$$y_{ij} = \sum_{s=1}^{n_1+1} \sum_{j=1}^{n_2} t_{sk}^* x_{1ij} x_{2jk}$$

$$y_{ij} = \sum_{k=1}^{n_2} \left[t_k x_{2jk} + \sum_{s=1}^{n_1} t_{sk} x_{1is} \right]$$

2. SCHEMA DE RESOLUTION

Les étapes sont les suivantes :

On décompose les matrices X_1^* et X_2 sous la forme :

$$\left\{ \begin{array}{l} X_1^* = \begin{matrix} U_1 & \bar{\beta}_1 & V_1' \\ (n, r_1) & (r_1, r_1) & (r_1, n_1+1) \end{matrix} \\ \text{où } r_1 = R_g X_1^* \\ \text{avec } \left\{ \begin{array}{l} U_1' U_1 = I_{r_1} \\ V_1' V_1 = I_{r_1} \end{array} \right\} \end{array} \right.$$

$$\left\{ \begin{array}{l} X_2 = \begin{matrix} U_2 & \bar{\beta}_2 & V_2' \\ (p, r_2) & (r_2, r_2) & (r_2, n_2) \end{matrix} \\ \text{où } r_2 = R_g X_2 \\ \text{avec } \left\{ \begin{array}{l} U_2' U_2 = I_{r_2} \\ V_2' V_2 = I_{r_2} \end{array} \right\} \end{array} \right.$$

Les matrices $\bar{\beta}_1$ et $\bar{\beta}_2$ sont diagonales. Elles contiennent respectivement les valeurs propres positives de $X_1' X_1$ et $X_2' X_2$.

Le problème à résoudre est alors :

$$\text{Min } \| Y - \begin{matrix} U_1 & \bar{\beta}_1 & V_1' & T^* & V_2 & \bar{\beta}_2 & U_2' \\ T^* & (n, p) & (n, r_1) & (r_1, r_1) & (r_1, n_1+1) & (n_1+1, n_2) & (n_2, r_2) & (r_2, r_2) & (r_2, p) \end{matrix} \|^2$$

On pose :

$$\theta = \bar{\beta}_1 V_1' T^* V_2 \bar{\beta}_2.$$

L'obtention de la solution passe ensuite par les deux propositions suivantes :

* Proposition 1

Si les matrices X_1 et X_2 sont orthonormales, c'est-à-dire

$$\left\{ \begin{array}{l} X_1^{*'} X_1^* = I_{n_1+1} \\ X_2' X_2 = I_{n_2} \end{array} \right\}$$

une solution du problème (P_1) :

$$\min_{T^*} \|Y - \hat{Y}\|^2 = \min_{T^*} \|Y - X_1^* T^* X_2'\|^2$$

$$\text{est } T^* = X_1^{*'} Y X_2.$$

* Proposition 2

Soit (P_1) le problème : $\min_{T^*} \|Y - U_1 \bar{\beta}_1 V_1' T^* V_2 \bar{\beta}_2 U_2'\|^2$

Soit (P_2) le problème : $\min_{\theta} \|Y - U_1 \theta U_2'\|^2$

Si θ est solution de (P_2) , alors T^* est tel que

$$\theta = \bar{\beta}_1 V_1' T^* V_2 \bar{\beta}_2 \text{ est solution de } (P_1).$$

On démontre ainsi que la solution du problème (P_1) est telle que :

$$\bar{\beta}_1 V_1' T^* V_2 \bar{\beta}_2 = U_1' Y U_2 \quad (1)$$

Ce système est résolu en deux étapes.

$$\text{On pose } \left\{ \begin{array}{l} W_1 = \bar{\beta}_1 V_1' \\ (r_1, n_1 + 1) \\ W_2 = V_2 \bar{\beta}_2 \\ (n_2, r_2) \end{array} \right.$$

Dans une première étape, on résoud le système

$$U_1' Y U_2 = [\bar{W}_1 \ T^*] W_2$$

ce qui revient à chercher une matrice $\bar{W}_2$ de dimension (r_2, n_2) telle a

$$\bar{W}_2' W_2 = I_{r_2}$$

Dans une seconde étape, on résoud le système

$$[U_1^T \ Y \ U_2^T] W_2^- = W_1^T T^*$$

ce qui revient à chercher une matrice W_1^- de dimension $(n_1 + 1, r_1)$ telle que :

$$W_1 W_1^- = I_{r_1}$$

Il existe un ensemble de solutions admissibles vérifiant (1). Parmi celles-ci, on choisit, pour des raisons d'interprétation, celle dont les lignes et les colonnes sont au mieux centrées par bloc de variables. Autrement dit, à la résolution du premier système, on choisit parmi l'ensemble des solutions celle qui, pour chacun des r_2 systèmes, minimise la fonction suivante :

$$f(a_1, \dots, a_i, \dots, a_{n_2}) = \sum_{j=1}^{N_2} \frac{1}{M(j)} \left(\sum_{q=A}^B a_q \right)^2$$

où $\begin{cases} M(j) \text{ désigne le nombre de modalités de la variable } j \\ A = \sum_{s=1}^{j-1} M(s) + 1 \\ B = \sum_{s=1}^j M(s) \end{cases}$

sous r_2 contraintes ;

$$g_k(a_1, \dots, a_i, \dots, a_{n_2}) = [\bar{0}, \dots, \underset{\substack{\uparrow \\ k^{\text{ème}} \text{ place}}}{1}, \dots, \bar{0}] \quad \forall k \in [r_2]$$

De même, à la résolution du second système, on choisit parmi l'ensemble des solutions celle qui, pour chacun des r_1 systèmes, minimise la fonction suivante :

$$f(b_1, \dots, b_i, \dots, b_{n_1+1}) = \sum_{j=1}^{N_1+1} \frac{1}{M(j)} \left(\sum_{q=A}^B b_q \right)^2$$

sous r_1 contraintes :

$$g_k(b_1, \dots, b_i, \dots, b_{n_1+1}) = [\bar{0}, \dots, \underset{\substack{\uparrow \\ k^{\text{ème}} \text{ place}}}{1}, \dots, \bar{0}] \quad \forall k \in [r_1]$$

3. CARACTERISTIQUES DU PROGRAMME

① Une approche locale du problème (p_1) est réalisable

Il devient alors possible de trouver simultanément :

- la partition P de l'ensemble I des individus en k classes ;
- et les matrices locales des coefficients d'influence T^1 pour chaque classe P_1 de $P \quad \forall 1 \in \overline{[k]}$ telles que :

$$\hat{Y}^1_{(n_1, p)} = \begin{matrix} X^{*1} & T^{*1} & X_2 \\ (n_1, n_1+1) & (n_1+1, n_2) & (n_2, p) \end{matrix} \text{ avec } n_1 = |P_1|$$

donne la meilleure approximation de Y^1 au sens des moindres carrés, ceci $\forall 1 \in \overline{[k]}$.

L'interprétation des matrices T^1 et $\tilde{t}^1$ est semblable à celle que nous avons déjà donnée. Mais ATTENTION, cette option locale est très délicate à utiliser en raison du nombre de coefficients inconnus par classe : on risque de déduire du modèle des relations entre variables qui n'existent pas. L'utilisateur doit donc prendre conscience que l'option locale ne peut présenter un intérêt que sur des données très particulières.

② Cette version du programme n'admet pas l'éventualité de données manquantes dans la matrice des préférences.

③ Ce modèle est totalement ADDITIF, autrement dit : tous les coefficients de T et $\tilde{t}$ sont sommables et chaque somme a une signification.

Avec les dimensions actuelles du programme, on peut traiter des exemples où au maximum :

- la taille de la population étudiée est de 460 individus ;
- le nombre d'objets classés est de 24 ;
- les caractéristiques des juges sont au nombre de 12, soit au plus 26 modalités ;
- les caractéristiques des objets sont au nombre de 2, soit au plus 10 modalités.

Remarque

Deux versions ont été écrites :

- une version CII IRIS 80 ;
- une version IBM 370.

Pour de telles dimensions, la taille mémoire demandée est assez importante (130 K sur IRIS 80, 320 K sur IBM 370) et le temps machine est d'environ 1' C.P.U.

4. UTILISATION DU PROGRAMME

4.1 Introduction des données

* 1ère carte : carte de paramètres généraux

En format 6 I 2, on écrit successivement : NOBJ, N1, N2, NCLASS, NTIR, ITERMX avec :

NOBJ = nombre d'objets classés

N1 = nombre total de caractéristiques - sujets pris en compte

N2 = nombre total de caractéristiques - objets pris en compte

NCLASS = nombre de classes désiré

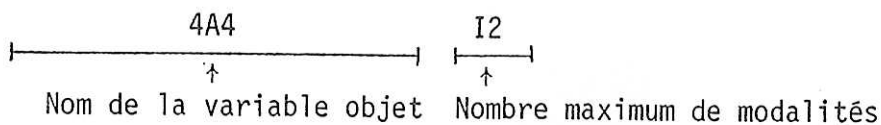
NTIR = nombre de tirages au hasard

ITERMX = nombre d'itérations maximum pour un tirage.

Remarque : Les trois derniers paramètres NCLASS, NTIR et ITERMX doivent prendre la valeur 1 dans le cas général. Ils sont différents de celle-ci si l'option locale du programme désire être utilisée.

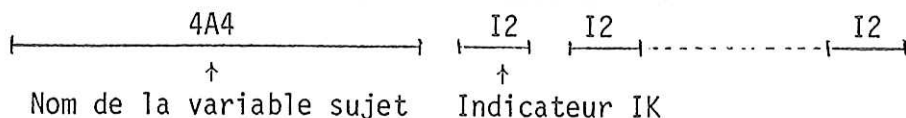
* N₂ cartes suivantes : elles servent à décrire les N₂ variables-objets. Elles sont toutes établies selon le schéma suivant :

- sur 16 caractères alphanumériques, on donne le nom de la variable ;
- puis, en format I 2, on écrit le nombre maximum de modalités correspondant à la variable en question.



* N₁ cartes suivantes : elles servent à décrire les N₁ variables-sujets. Elles sont toutes établies sur le schéma suivant :

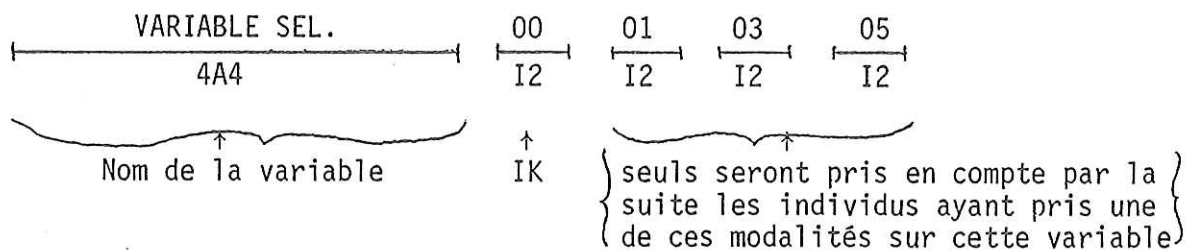
- sur 16 caractères alphanumériques, on donne le nom de la variable ;
- puis, en format I 2, on détermine toutes ses particularités selon qu'il s'agit d'une variable de sélection ou d'une variable d'analyse.



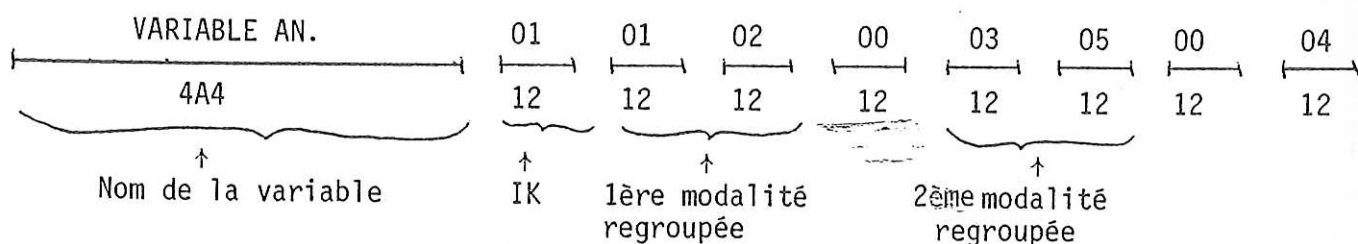
L'indicateur IK établit la nature de la variable.

- Si IK = 0, la variable lue est une variable de sélection.

Exemple :



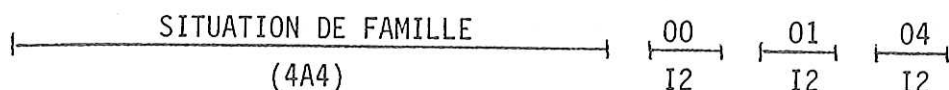
- Si IK = 1, la variable lue est une variable d'analyse. On peut regrouper certaines modalités si on le désire. Exemple :



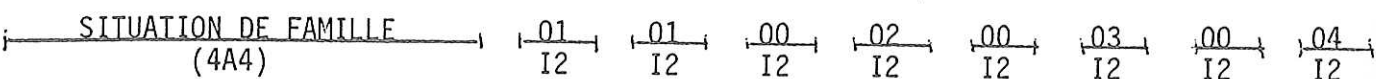
Remarque : L'utilisation d'une variable de sélection autorise la segmentation de la population sur certaines de ses modalités. Exemple : considérons la variable "situation de famille" dont les modalités sont les suivantes :

- | | | |
|---|-----------------|---|
| { | 1 - Célibataire | } |
| | 2 - Marié | |
| | 3 - Veuf | |
| | 4 - Divorcé | |

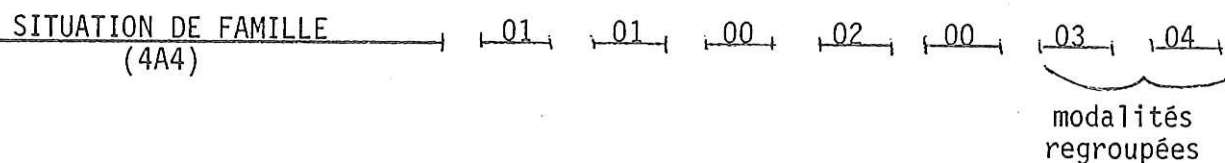
Supposons que l'on désire étudier l'influence d'autres variables sur une population exclusive d'individus célibataires et divorcés. On écrira alors :



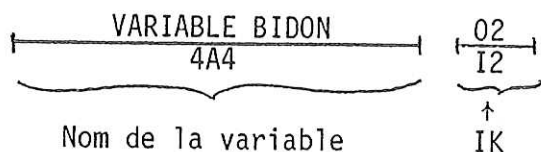
En revanche, la prise en compte de cette variable comme une variable d'analyse avec toutes ses modalités se traduira par :



Si l'on regroupe les veufs et divorcés ensemble, on aura :



* Carte suivante : elle indique que la lecture des variables sujets est terminée. Elle ne correspond à aucune variable réelle. Elle a la forme suivante :



L'indicateur IK vaut 2.

* Carte suivante : carte de format de lecture des données. Pour chaque individu, on doit lire :

N1 variables caractéristiques et NOBJ notes établies sur les objets

* NSUJ cartes suivantes : où NSUJ désigne le nombre d'individus. Sur chaque carte, on lit les N1 caractéristiques d'un individu et les NOBJ notes qu'il a données aux objets qu'on lui a présentés.

* Carte suivante : on écrit un "0" en 1ère colonne. Elle indique que la carte correspondant au dernier individu a été lue.

* NOBJ cartes suivantes : où NOBJ désigne le nombre d'objets. Sur chaque carte, on lit les N2 caractéristiques d'un objet en format I2.

4.2 Lecture des résultats

. On peut lire en premier lieu un rappel des données. On a ainsi une récapitulation des caractéristiques des variables objets et des variables sujets (variables de sélection ou d'analyse).

4.2.1 Dans le cas général

On indique la valeur du critère par :

$$\text{SOMME} = \left[\frac{1}{\text{NSUJ} \times \text{NOBJ}} \sum_{i=1}^{\text{NSUJ}} \sum_{j=1}^{\text{NOBJ}} (y_{ij} - y_{ij}^*)^2 \right]^{1/2}$$

où $\left\{ \begin{array}{l} y_{ij} \text{ est la note donnée par l'individu } i \text{ à l'objet } j \\ y_{ij}^* \text{ est la note reconstituée par le modèle BENITO} \end{array} \right\}$

On détermine les tableaux T et $\tilde{t}$ qui apparaissent sous la forme suivante :

		<u>Variables-objets</u>		
		1ère variable	2ème variable	N ₂ ème variable
<u>Tableaux-sujets</u>	$\tilde{t}$			
	1ère var.			
	2e var.			
	N ₁ ème var.			

En marge des tableaux sont données les fréquences correspondant à chaque modalité des différentes variables-sujets.

4.2.2 Dans le cas de l'"option locale"

- On indique comme précédemment la valeur du critère.
- On détermine la composition de chaque classe l de la partition P pour tout $l \in [k]$ ainsi que les k tableaux T^l et $\tilde{T}^l$ correspondants.

5. EXEMPLE DE REFERENCE

On considère 6 individus A, B, C, D, E, F et 4 objets $0_1, 0_2, 0_3, 0_4$. Les individus sont caractérisés par deux variables de respectivement 2 et 3 modalités. Les objets sont repérés par 2 variables de 2 modalités chacune.

Les matrices X_1 et X_2 sous forme disjonctive complète se présentent ainsi :

	1ère Var. Suj.		2ème Var. Suj.		
	mod 1	mod 2	mod 1	mod 2	mod 3
A	1	0	1	0	0
B	1	0	0	1	0
C	1	0	0	0	1
D	0	1	1	0	0
E	0	1	0	1	0
F	0	1	0	0	1

	1ère Var. Objet		2ème Var. Objet	
	mod 1	mod 2	mod 1	mod 2
0_1	1	0	1	0
0_2	1	0	0	1
0_3	0	1	1	0
0_4	0	1	0	1

$$n_1 = 5$$

$$n_2 = 4$$

et le tableau des préférences Y est le suivant :

	0_1	0_2	0_3	0_4
A	+4	+6	-6	-4
B	+5	-1	+1	-5
C	0	-2	+2	0
D	-2	+4	-4	+2
E	-1	-3	+3	+1
F	-6	-4	+4	+6

5.1 Les entrées sur l'exemple de référence

Dans l'exemple, on a :

$$\left\{ \begin{array}{l} \text{NOBJ} = 4 \\ \text{N1} = 2 \\ \text{N2} = 2 \end{array} \right.$$

L'option locale n'est pas prise :

$$\left\{ \begin{array}{l} \text{NCLASS} = 1 \\ \text{NTIR} = 1 \\ \text{ITERMX} = 1 \end{array} \right.$$

Les deux variables-sujets sont des variables d'analyse : l'indicateur IK vaut donc 01. Il n'y a pas de variable de sélection. Le bordereau d'entrée des données établit l'ensemble des cartes nécessaires à l'exécution du programme BENITO sur l'exemple choisi.

PROGRAMME "FORTRAN"

10, 8556

PROBLEME		PROGRAMMEUR		DATE		PAGE		IDENTIFICATION	
TYPE		N°		7		8		9	
1	2	1	1	1	1	1	1	1	1
2	2	1	1	1	1	1	1	1	1
3	2	1	1	1	1	1	1	1	1
4	2	1	1	1	1	1	1	1	1
5	2	1	1	1	1	1	1	1	1
6	2	1	1	1	1	1	1	1	1
7	2	1	1	1	1	1	1	1	1
8	2	1	1	1	1	1	1	1	1
9	2	1	1	1	1	1	1	1	1
10	2	1	1	1	1	1	1	1	1
11	2	1	1	1	1	1	1	1	1
12	2	1	1	1	1	1	1	1	1
13	2	1	1	1	1	1	1	1	1
14	2	1	1	1	1	1	1	1	1
15	2	1	1	1	1	1	1	1	1
16	2	1	1	1	1	1	1	1	1
17	2	1	1	1	1	1	1	1	1
18	2	1	1	1	1	1	1	1	1
19	2	1	1	1	1	1	1	1	1
20	2	1	1	1	1	1	1	1	1
21	2	1	1	1	1	1	1	1	1
22	2	1	1	1	1	1	1	1	1
23	2	1	1	1	1	1	1	1	1
24	2	1	1	1	1	1	1	1	1
25	2	1	1	1	1	1	1	1	1
26	2	1	1	1	1	1	1	1	1
27	2	1	1	1	1	1	1	1	1
28	2	1	1	1	1	1	1	1	1
29	2	1	1	1	1	1	1	1	1
30	2	1	1	1	1	1	1	1	1
31	2	1	1	1	1	1	1	1	1
32	2	1	1	1	1	1	1	1	1
33	2	1	1	1	1	1	1	1	1
34	2	1	1	1	1	1	1	1	1
35	2	1	1	1	1	1	1	1	1
36	2	1	1	1	1	1	1	1	1
37	2	1	1	1	1	1	1	1	1
38	2	1	1	1	1	1	1	1	1
39	2	1	1	1	1	1	1	1	1
40	2	1	1	1	1	1	1	1	1
41	2	1	1	1	1	1	1	1	1
42	2	1	1	1	1	1	1	1	1
43	2	1	1	1	1	1	1	1	1
44	2	1	1	1	1	1	1	1	1
45	2	1	1	1	1	1	1	1	1
46	2	1	1	1	1	1	1	1	1
47	2	1	1	1	1	1	1	1	1
48	2	1	1	1	1	1	1	1	1
49	2	1	1	1	1	1	1	1	1
50	2	1	1	1	1	1	1	1	1
51	2	1	1	1	1	1	1	1	1
52	2	1	1	1	1	1	1	1	1
53	2	1	1	1	1	1	1	1	1
54	2	1	1	1	1	1	1	1	1
55	2	1	1	1	1	1	1	1	1
56	2	1	1	1	1	1	1	1	1
57	2	1	1	1	1	1	1	1	1
58	2	1	1	1	1	1	1	1	1
59	2	1	1	1	1	1	1	1	1
60	2	1	1	1	1	1	1	1	1
61	2	1	1	1	1	1	1	1	1
62	2	1	1	1	1	1	1	1	1
63	2	1	1	1	1	1	1	1	1
64	2	1	1	1	1	1	1	1	1
65	2	1	1	1	1	1	1	1	1
66	2	1	1	1	1	1	1	1	1
67	2	1	1	1	1	1	1	1	1
68	2	1	1	1	1	1	1	1	1
69	2	1	1	1	1	1	1	1	1
70	2	1	1	1	1	1	1	1	1
71	2	1	1	1	1	1	1	1	1
72	2	1	1	1	1	1	1	1	1
73	2	1	1	1	1	1	1	1	1
74	2	1	1	1	1	1	1	1	1
75	2	1	1	1	1	1	1	1	1
76	2	1	1	1	1	1	1	1	1
77	2	1	1	1	1	1	1	1	1
78	2	1	1	1	1	1	1	1	1
79	2	1	1	1	1	1	1	1	1
80	2	1	1	1	1	1	1	1	1
81	2	1	1	1	1	1	1	1	1
82	2	1	1	1	1	1	1	1	1
83	2	1	1	1	1	1	1	1	1
84	2	1	1	1	1	1	1	1	1
85	2	1	1	1	1	1	1	1	1
86	2	1	1	1	1	1	1	1	1
87	2	1	1	1	1	1	1	1	1
88	2	1	1	1	1	1	1	1	1
89	2	1	1	1	1	1	1	1	1
90	2	1	1	1	1	1	1	1	1
91	2	1	1	1	1	1	1	1	1
92	2	1	1	1	1	1	1	1	1
93	2	1	1	1	1	1	1	1	1
94	2	1	1	1	1	1	1	1	1
95	2	1	1	1	1	1	1	1	1
96	2	1	1	1	1	1	1	1	1
97	2	1	1	1	1	1	1	1	1
98	2	1	1	1	1	1	1	1	1
99	2	1	1	1	1	1	1	1	1
100	2	1	1	1	1	1	1	1	1

5.2 Les sorties sur l'exemple de référence

Le tableau obtenu est le suivant :

180-SLP1- IRIA 23*0
180-SLP1- IRIA 23*0

5.3 Autre exemple

Un exemple réel a été traité dans [5] sur des données recueillies par la SOFRES.

Il s'agissait d'évaluer la valeur qualitative spécifique conférée par un support de presse aux produits dont la publicité lui est confiée. Cette valeur a été recherchée pour trois dimensions qualitatives pouvant constituer des thèmes sous-jacents à certains messages publicitaires :

- le thème de la nouveauté ;
- le thème de la personnalisation (c'est-à-dire de l'expression de soi dans la nouveauté)
- le thème du choix judicieux dans la nouveauté.

Six produits ont été sélectionnés et quatre supports ont été considérés de manière à définir vingt quatre couples produits x supports, chacun représentant une publicité qui aurait pu paraître pour le produit concerné dans le support correspondant.

Les produits pris en compte étaient les suivants :

- tenue de loisir
- modèle de lingerie féminine
- living, séjour
- cuisine
- séjour de vacances
- appartement en multipropriété pour les vacances d'hiver ou d'été.

Trois supports appartenaient à la presse féminine tandis que le dernier était un hebdomadaire d'information générale.

Les questions posées étaient les suivantes :

- existe-t-il une valeur ajoutée propre au support indépendamment des produits publicités ?
- comment les supports se hiérarchisent-ils par rapport à cette valeur ajoutée ?
- existe-t-il une valeur ajoutée produit indépendante du support et surtout y a-t-il des produits dont l'image aurait tout intérêt à bénéficier d'un apport du titre pour appuyer le message publicitaire ?

Les coefficients déterminés par le tableau T^* ont permis de répondre à ces interrogations en évaluant l'influence exercée par la lecture ou la non lecture de tel ou tel support selon les différents critères servant de messages publicitaires. De plus, on a tenu compte de variables telles que l'âge ou la profession du chef de famille afin de considérer les variations qu'elles pouvaient induire sur le préordre moyen des préférences.

ANNEXE 1
ORGANISATION DES SOUS-PROGRAMMES
LISTE ET FONCTIONS DES DIFFERENTS SOUS-PROGRAMMES

① Sous-programme SUITE

Il constitue la suite du programme principal et par conséquent est décrit en même temps que lui.

② Sous-programmes CODY et CODX

forment en fait un seul sous-programme dont le rôle est la construction de la matrice XF à partir de X. C'est là que l'on transforme les matrices X_1 et X_2 en tableaux disjonctifs complets.

Les arguments utilisés sont les suivants :

X est la matrice initiale $\left\{ \begin{array}{l} \text{à } N1 \text{ colonnes (les variables qualitatives)} \\ \text{et } NS \text{ lignes} \end{array} \right.$

XF est la matrice finale $\left\{ \begin{array}{l} \text{à } NS \text{ lignes} \\ \text{et } \left[\begin{array}{c} N1 \\ \Sigma \\ i=1 \end{array} \right] \text{ MODMAX}(i) \text{ colonnes} \end{array} \right.$

MODMAX(i) = nombre de modalités par variable

NS est le nombre de lignes

N1 en entrée est le nombre de colonnes de X

en sortie est le nombre de colonnes de XF

N1CONS en sortie désigne le nombre de colonnes de X

IM = 0 on remplit X2F (matrice de caractéristiques d'objets)

IM = 1 on remplit X1F (matrice de caractéristiques des sujets)

Elle a une colonne supplémentaire composée uniquement de "1"

③ Sous-programme SMITH (nom d'emprunt pour HOUSEHOLDER)

$X(N1, NS)$ est la matrice en entrée. Elle est décomposée de la façon suivante :

$$X(N1, NS) = \underbrace{XTR(N1, NVAL)}_{\substack{\text{a ses colon-} \\ \text{nes orthogo-} \\ \text{nales entre} \\ \text{elles}}} * \underbrace{XT(NVAL, NS)}_{\substack{\text{a ses lignes} \\ \text{normées et} \\ \text{orthogonales} \\ \text{entre elles}}}$$

NVAL désigne le rang de $X'X$

BTA, D, E, IK sont des matrices de travail.

④ Sous-programme SYMQR

Diagonalisation de la matrice symétrique $A(NA, NA)$. Les valeurs propres sont en sortie dans $VAL(NA)$. Les vecteurs propres sont en sortie dans A.

⑤ Sous-programme INVMAT

Résolution du système linéaire :

$$A(N1, N1) * X(N1, N2) = B(N1, N2)$$

En sortie la solution X est dans B.

JP est un tableau de travail.

⑥ Sous-programme HASARD

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

NCLASS = nombre de classes de la partition

KLASS(I) = indice de la classe de l'individu i

NSCLA(1, j) = nombre d'individus dans la classe j

IECLA(1, i, j) = indice du j^{ème} individu de la classe i.

Ces deux derniers tableaux ont une dimension supplémentaire correspondant à la "mémoire" d'une itération antérieure.

RANDU(NCLASS) = nombre entier aléatoire entre 1 et NCLASS.

REFERENCES

- [1] BOUROCHE J.M. - "Analyse des données en marketing". Monographies de l'AFCEP, Masson, 1977.
- [2] CARROLL J.D., GREEN P.E., CARMONE F.J. - "Super Ordinate Factorial Designs in the Analysis of Consumer Judgments". Handout Bell Laboratories, Murray Hill, New Jersey, 1975.
- [3] CARROLL J.D., GREEN P.E., CARMONE F.J. - "CANDELINC (Canonical Decomposition with Linear Constraints)". International Congress of Psychology in Paris, July 18-25, 1976.
- [4] CARROLL J.D., PRUZANSKY S. - "MULTILINC : MULTI-WAY CANDELINC". Meeting of the American Psychological Association in San Francisco (California), August 26-30, 1977.
- [5] DUPONT-GATELMAND C. - "Deux méthodes d'analyse des données qualitatives : typologie avec codage adaptatif et modèle de décomposition additive des préférences". Thèse de 3e Cycle, Université Paris IX, avril 1978.
- [6] GREEN P.E., WIND Y.E. - "Multiattribute Decisions in Marketing". The Dryden Press, 1973.
- [7] RAO C.R., MITRA S.K. - "Generalized Inverse of Matrices and its Applications". Wiley, New York, 1971.
- [8] SEARLE S.R. - "Linear Models". Wiley, 1971.