UNE PROCÉDURE DE RÉDUCTION DU NOMBRE DE PAIRES EN ANALYSE CONJOINTE

Salwa BENAMMOU [1], Gilbert SAPORTA [2], Besma SOUISSI [3]

RÉSUMÉ

En analyse conjointe on utilise souvent des comparaisons par paires pour le recueil des préférences des consommateurs. La recherche des paires s'effectue à l'aide de plans d'expériences mais se trouve handicapée par l'explosion combinatoire du nombre de paires de produits dès que le nombre d'attributs et de modalités est grand.

Nous proposons ici une procédure simple qui réduit considérablement le nombre de paires candidates lorsque le nombre de produits devient grand.

Mots-clés : Analyse conjointe, comparaisons par paires

ABSTRACT

In conjoint analysis, paired comparison is often used to collect consumer's preferences. The research of pairs is done using experimental designs, but it is handicapped by the combinatory explosion of the number of pairs of products as soon as the number of attributes and categories is large.

We propose in this paper a simple procedure which reduces considerably the number of candidate pairs when the number of products becomes large.

Keywords : Conjoint analysis, paired comparison

1. Introduction

Introduite en recherche marketing au début des années 1970 (Green et Rao 1971), l'analyse conjointe est considérée comme une technique de mesure des arbitrages dans le choix de produits de consommation. Son évolution dans la recherche marketing est bien documentée par plusieurs auteurs (Green et Srinivasan 1978, 1990, Cattin et Wittink 1989,1982 ainsi que Carrol et Green 1995).

1. Département de Méthodes quantitatives, Faculté de Droit et des Sciences Economiques et Politiques de Sousse (Tunisie). E-mail : saloua.benammou@fdseps.rnu.tn

2. Chaire de Statistique Appliquée et CEDRIC, Conservatoire National des Arts et Métiers, Paris. E-mail : saporta@cnam.fr

3. Institut Supérieur de Gestion de Sousse (Tunisie). E-mail : besma.swissi@yahoo.fr

L'analyse conjointe est une méthodologie complète comportant trois phases. La première est une phase de recueil des observations (généralement par interview en face à face). Cette phase est basée sur un choix de combinaisons de modalités de variables déterminé par application de la technique des plans d'expériences. La deuxième est une phase de traitement destinée à estimer les paramètres. La dernière est une phase de simulation.

Lorsque chaque consommateur est amené à classer par ordre de préférence un ensemble de produits, l'analyse conjointe est une forme particulière du modèle linéaire général (Benammou *et al.* 2003). Elle cherche à expliquer une variable qualitative ordinale en fonction de plusieurs variables qualitatives à modalités ordonnées ou non.

Nous nous intéressons ici à la phase de recueil des données en analyse conjointe et plus particulièrement à la méthode des comparaisons par paires.

2. Rappels sur l'analyse conjointe et le modèle de régression linéaire

Nous donnons d'abord un rappel succinct sur les différentes méthodes de recueil des données dont la méthode des comparaisons par paires. Nous donnerons ensuite le modèle de régression linéaire pour cette dernière.

L'analyse conjointe est une méthode d'analyse des données qui cherche à expliquer une variable de préférence à l'aide de plusieurs variables explicatives qualitatives à modalités ordonnées ou non (Droesbeke *et al.* 1997), selon un modèle de choix supposé compensatoire, en admettant un certains nombre d'hypotèses telles que :

- le consommateur associe des valeurs subjectives (utilités partielles qui reflètent son propre système de valeurs) à chacune des modalités des variables explicatives
- le modèle est additif : l'utilité totale d'un produit est la somme des utilités partielles de ses caractéristiques. Introduire des termes d'interaction dans le modèle engendrant une forte augmentation du nombre des paramètres à estimer n'est en général pas retenu
- le consommateur choisit le produit qui lui procure l'utilité totale la plus élevée.

2.1. Méthodes de recueil de données en analyse conjointe

Nous nous intéressons ici à la première phase de l'analyse conjointe, où la tâche confiée au répondant dépend de la méthode de collecte des données fixée dans l'étude. Il peut s'agir de l'approche « Trade-off », de méthodes « CBC » ou des profils complets.

L'approche « Trade-off », dite encore « comparaison par paire d'attributs », consiste à comparer les attributs pris deux à deux et présentés sous la forme d'une matrice croisant deux sous-ensembles de modalités. La question posée au consommateur est simple, celui-ci ne juge que sur deux attributs en classant par ordre de préférence les cases du tableau croisé. Lorsqu'il y a p attributs, on doit présenter plusieurs tableaux croisés : la comparaison entre les attributs x_1 et x_2 par exemple donne un système d'utilités pour x_1 et x_2. La comparaison des attributs x_1 et x_3 donne alors des utilités différentes pour x_1. On a donc autant de systèmes d'utilités que de comparaisons entre attributs. La méthode a été abandonnée vu l'incohérence des estimations des utilités qu'elle induit (Faivre & Pioche 1976 et Kuhfeld 2005).

Dans la méthode des profils complets, on présente au consommateur un ensemble de produits obtenus par combinaison des modalités de tous les attributs considérés dans l'étude. On lui demande soit de classer ces produits, soit de les comparer deux à deux.

En cas de classement exhaustif, le répondant doit fournir un ordre de préférence des produits proposés. L'ensemble des produits proposés est un sous-ensemble de l'ensemble de tous les produits possédant des propriétés d'optimalité (plan fractionnaire orthogonal, D-optimal).

Les méthodes CBC ou Choice-Based Conjoint Analysis, attirent de plus en plus l'attention en recherche marketing puisqu'elles proposent un processus de choix proche de la réalité. Le consommateur choisit un produit parmi un ensemble, avec en outre une possibilité de non-choix. Lorsque l'on accepte d'agréger les résultats (utilités globales et non plus individu par individu) les méthodes CBC donnent aussi la possibilité d'étudier les interactions entre les attributs (Gustafsson *et al.* 2003).

La comparaison par paires de produits, introduite par Thurstone en 1927, est une méthode alternative à celle des scénarios complets. Elle consiste à présenter à l'interviewé plusieurs paires de produits (ou scénarios) afin de les « juger », et exige qu'il choisisse un produit parmi deux formant une paire et ayant des caractéristiques différentes. Le jugement se fait pour chacune des paires qui lui a été présentée (Carroll, J.D. & Green, P.E 1995). Les jugements peuvent prendre différentes formes selon l'étude. Il peut s'agir :

– du choix d'un des produits

– de la préférence pour un produit selon une échelle

La méthode de comparaison par paires de produits, présente l'avantage que l'interviewé n'est pas obligé de classer tous les produits qu'on lui présente. Il n'a pas à mémoriser tous les attributs de chaque produit pour le comparer avec tous les autres. Le seul inconvénient étant le nombre de paires qui doit être très supérieur à celui des scénarios pour obtenir une efficacité comparable.

2.2. Le modèle linéaire appliqué aux comparaisons par paires

Soit un produit t décrit par p attributs. On confondra dans ce qui suit le produit et son vecteur de description binaire (forme disjonctive complète) à $\sum_{i=1}^{p} m_i$ composantes, m_i étant le nombre de modalités du $i^{ème}$ attribut.

Si β est le vecteur des utilités partielles des $\sum_{i=1}^{p} m_i$ modalités, le score du produit t s'obtient par combinaison linéaire :

$$S(t) = \mu + t'\beta$$

Pour chaque paire de produits s et t comparés, on suppose que la préférence est fonction de la différence entre les utilités :

$$S(t) - S(s) = (t - s)'\beta$$

Notons Y la variable valant 1 si l'interviewé a choisi le premier produit et 2 si l'interviewé a choisi le deuxième produit de la paire. D'autres codages sont possibles et équivalents comme 1 et 0.

Pour N comparaisons par paires on doit donc expliquer un vecteur Y à N composantes par un modèle du type :

$$Y = \phi(F'\beta) + Z$$

où ϕ est une fonction vectorielle de lien, F' la matrice dont les lignes sont les différences du type $(t - s)'$ et Z le vecteur des erreurs aléatoires associé au modèle.

$$F' = \begin{bmatrix} 0 & 0 & \cdots & -1 & 1 \\ 1 & -1 & \cdots & 0 & 1 \\ \vdots & & \ddots & & \vdots \\ -1 & 0 & \cdots & 1 & 0 \end{bmatrix}$$

Comme les indicatrices des $\sum_{i=1}^{p} m_i$ modalités ne sont pas linéairement indépendantes, il existe une infinité de paramétrisations possibles des utilités. En général on impose des contraintes de somme nulle aux utilités de chaque attribut, ce qui revient en réalité à travailler avec une matrice F' dont le nombre de colonnes est égal à $\sum_{i=1}^{p} m_i - p$.

3. Recherche de «plans» d'expérience

Le nombre de paires doit être au moins égal au nombre de paramètres linéairement indépendants du modèle qui vaut $\sum_{i=1}^{p} m_i - p$.

La matrice F, c'est-à-dire les paires, doit être choisie de telle sorte que les estimations des utilités soient de variance minimale, mais le critère dépend de la fonction ϕ. Si le lien est logistique on devrait utiliser la X-optimalité (Gauchi 1999) au lieu de la D-optimalité qui convient aux modèles linéaires (Chaloner 1984).

Cette approche étant très complexe, nous utiliserons quand même la D-optimalité.

On recherchera parmi l'ensemble des paires candidates un sous-ensemble de taille q fixée maximisant le déterminant de (F'F). Or dans le cas de comparaison par paires, si n est le nombre total de produits possibles qui vaut $n = \prod_{i=1}^{p} m_i$, le nombre de paires candidates qui vaut $\frac{n \times (n-1)}{2}$ est très élevé, ce qui rend la recherche délicate. Nous proposons dans ce qui suit une méthode permettant de réduire très substantiellement le nombre de paires candidates.

4. Réduction du nombre de paires

Nous proposons ici une procédure simple qui permet de réduire considérablement le nombre de paires candidates sans perte d'information ni souci d'orthogonalité. Nous commençons par donner deux exemples illustratifs où nous détaillons la procédure de réduction proposée, avant de passer au cas général.

Le premier exemple correspond à une version standard de la procédure alors que le second correspond à une version heuristique de cette même procédure. La version heuristique commence par chercher un plan optimal factoriel fractionnaire, avant d'appliquer la procédure standard.

4.1. Exemple de trois variables dichotomiques

Considérons le cas ou les produits sont des chaussures caractérisées par trois variables qualitatives à deux modalités chacune qui sont la couleur (noire ou marron), le matériau (cuir ou daim) et la semelle (à talon ou plate).

Le nombre de combinaisons possibles de toutes les modalités et égal à $2^3 = 8$ d'où le tableau n°1.

TABLEAU 1. — Produits possibles (cas de l'exemple de trois variables dichotomiques)

A	*noire*	*cuir*	*à talon*
B	*noire*	*cuir*	*plate*
C	*noire*	*daim*	*à talon*
D	*noire*	*daim*	*plate*
E	*marron*	*cuir*	*à talon*
E	*marron*	*cuir*	*plate*
G	*marron*	*daim*	*à talon*
H	*marron*	*daim*	*plate*

TABLEAU 2. — Codage des produits (cas de l'exemple de trois variables dichotomiques)

A	1 0	1 0	1 0
B	1 0	1 0	0 1
C	1 0	0 1	1 0
D	1 0	0 1	0 1
E	0 1	1 0	1 0
E	0 1	1 0	0 1
G	0 1	0 1	1 0
H	0 1	0 1	0 1

Le codage de ces produits sous forme disjonctive complète est donné par le tableau n°2.

À partir des produits codés, nous construisons toutes les paires possibles, soit $\frac{8 \times 7}{2} = 28$ paires, par différence entre les éléments des deux produits qui forment la paire.

On représente l'information apportée par une paire de produits par un jeu de différences entre les attributs formant les scénarios de la paire.

Une différence nulle sur un attribut veut dire que les deux produits ont la même modalité et une différence égale à 1 ou (-1) veut dire que les deux produits prennent des modalités différentes sur cet attribut. Les paires obtenues, dans le cas de cet exemple, sont données par le tableau n°3.

Deux cas se présentent :

- des paires différentes conduisent à des lignes du tableau identiques (lignes 1 et 14 par exemple)
- des paires différentes conduisent à des lignes du tableau dont les signes sont de l'une des formes de la figure n°1. Ces lignes seraient identiques par permutation des signes de certains éléments sur une même ligne.

En permutant par exemple le signe du troisième élément sur la première ligne avec celui du quatrième, la première et la deuxième ligne seraient identiques. Ces permutations sont tout à fait légitimes vu que la numérotation des modalités d'une même variable est arbitraire.

TABLEAU 3. — Codage des paires (cas de l'exemple de trois variables dichotomiques)

*ligne*1	$A\ B$	0 0	0 0	1 −1
	$A\ C$	0 0	1 −1	0 0
	$A\ D$	0 0	1 −1	1 −1
	$A\ E$	1 −1	0 0	0 0
	$A\ F$	1 −1	0 0	1 −1
	$A\ G$	1 −1	1 −1	0 0
*ligne*7	$A\ H$	1 −1	1 −1	1 −1
	$B\ C$	0 0	1 −1	−1 1
	$B\ D$	0 0	1 −1	0 0
	$B\ E$	1 −1	0 0	−1 1
	$B\ F$	1 −1	0 0	0 0
	$B\ G$	1 −1	1 −1	−1 1
	$B\ H$	1 −1	1 −1	0 0
*ligne*14	$C\ D$	0 0	0 0	1 −1
	$C\ E$	1 −1	−1 1	0 0
	$C\ F$	1 −1	−1 1	1 −1
	$C\ G$	1 −1	0 0	0 0
	$C\ H$	1 −1	0 0	1 −1
	$D\ E$	1 −1	−1 1	−1 1
	$D\ F$	1 −1	−1 1	0 0
*ligne*21	$D\ G$	1 −1	0 0	−1 1
	$D\ H$	1 −1	0 0	0 0
	$E\ F$	0 0	0 0	1 −1
	$E\ G$	0 0	1 −1	0 0
	$E\ H$	0 0	1 −1	1 −1
	$F\ G$	0 0	1 −1	−1 1
	$F\ H$	0 0	1 −1	0 0
*ligne*28	$G\ H$	0 0	0 0	1 −1

Dans les deux cas ces paires seront dites équivalentes.

+	−	+	−	+	−
+	−	−	+	+	−
+	−	+	−	−	+
+	−	−	+	−	+

FIG 1. — Certaines formes des lignes de codage des paires.

Sept classes d'équivalence se forment comme suit :

- Le premier groupe formé des paires AC, BD, EG et FH qui sont toutes identiques et de la forme de la première ligne du tableau n°4.
- Le deuxième groupe formé des paires AB, CD, EF et GH qui sont toutes identiques et de la forme de la deuxième ligne du tableau n°4.
- Le troisième groupe formé des paires AE, BF, CG, et DH qui sont toutes identiques et de la forme de la troisième ligne du tableau n°4.
- Le quatrième groupe formé des paires AF, BE, CH, et DG qui sont toutes équivalentes et de la forme de la quatrième ou cinquième ligne du tableau n°4.
- Le cinquième groupe formé des paires AD, EH, BC, et FG qui sont équivalentes deux à deux et de la forme de la sixième ou septième ligne du tableau n°4.
- Le sixième groupe formé des paires AG, BH, CE, et DF qui sont équivalentes deux à deux et de la forme de la huitième ou de la neuvième ligne du tableau n°4.
- Le septième et dernier groupe est formé par les paires AH, BG, CF et DE qui sont toutes différentes mais équivalentes et de la forme de l'une des quatre dernières lignes du tableau n°4.

Dans chacun des 7 groupes de paires nous gardons une seule paire et nous éliminons les 21 paires restantes.

Nous pouvons prendre par exemple les paires décrites par le tableau n°5 (la première paire de chaque groupe)

Le nombre final de paires est égal au quart du nombre initial, soit une réduction de 75%. Cette réduction peut être encore améliorée, en recherchant les paires qui ne diffèrent que sur un seul attribut.

La procédure consiste à regarder les caractéristiques qui évaluent chaque paire et repérer les éléments non nuls. Un « 0 » veut dire que les produits ont le même niveau pour l'une des caractéristiques, ce qui veut dire qu'il n'y a pas de différence pour le consommateur entre ces deux produits dus à cette caractéristique.

En effet considérons par exemple la paire de produits « chaussures noire, en cuir et à talon » et « chaussures noire, en cuir et plate ». Lors du choix du produit dans la paire, la seule caractéristique qui va intervenir est la semelle : à talon/plate, les autres caractéristiques étant identiques. L'évaluation de cette

TABLEAU 4. — Formes des différents groupes des paires (cas de l'exemple de trois variables dichotomiques)

$A\ C$	0	0	1	−1	0	0
$A\ B$	0	0	0	0	1	−1
$A\ E$	1	−1	0	0	0	0
$A\ F$	1	−1	0	0	1	−1
$B\ E$	1	−1	0	0	−1	1
$A\ D$	0	0	1	−1	1	−1
$B\ C$	0	0	1	−1	−1	1
$A\ G$	1	−1	1	−1	0	0
$C\ E$	1	−1	−1	1	0	0
$A\ H$	1	−1	1	−1	1	−1
$B\ G$	1	−1	1	−1	−1	1
$C\ F$	1	−1	−1	1	1	−1
$D\ E$	1	−1	−1	1	−1	1

TABLEAU 5. — Paires obtenues à la première réduction (cas de l'exemple de trois variables dichotomiques)

$A\ B$	0	0	0	0	1	−1
$A\ C$	0	0	1	−1	0	0
$A\ D$	0	0	1	−1	1	−1
$A\ E$	1	−1	0	0	0	0
$A\ F$	1	−1	0	0	1	−1
$A\ G$	1	−1	1	−1	0	0
$A\ H$	1	−1	1	−1	1	−1

paire ne donnera pas de relation entre attributs, mais entre deux niveaux d'un même attribut. Or l'information cherchée est la relation entre différents attributs, donc nous pouvons supprimer les paires qui ne diffèrent que sur un seul attribut.

Le tableau n°6 nous donne seulement les caractéristiques où les produits sont différents.

Les trois paires : AB, AC et AE ne différent que sur un seul attribut, donc elles peuvent être supprimées. Sur les 28 paires initiales il ne nous reste plus que 4 paires : AD, AF, AG, et AH, la réduction est donc de 85.7%.

Pour un interviewé, il est assez aisé d'évaluer 4 paires et de choisir dans chaque paire le produit qu'il préfère. Ses réponses sont donc plus fiables que dans le cas où il devait choisir 28 produits à partir de 28 paires et l'étude des importances des attributs se fait plus facilement.

TABLEAU 6. — Paires différentes sur un seul attribut (cas de l'exemple de trois variables dichotomiques)

A B			*à talon plate*
A C		*cuir daim*	
A D		*cuir daim*	*à talon plate*
A E	*noire marron*		
A F	*noire marron*		*à talon plate*
A G	*noire marron*	*cuir daim*	
A H	*noire marron*	*cuir daim*	*à talon plate*

Remarque 4.1.1. — Quatre est le nombre minimal de paires à présenter à un individu si on veut par la suite estimer les préférences associées au modèle.

4.2. Exemple de cinq variables dichotomiques

À l'exemple précédent nous rajoutons deux variables dichotomiques supplémentaires : le modèle (Boots ou Chaussures) et la forme (Arrondi ou Pointu). Le plan factoriel complet associé serait de $32 = 2^5$ scénarios ce qui engendre un nombre de paires candidates égal à $\dfrac{(32 \times 31)}{2}$ soit 496 paires.

Plutôt que de rechercher des paires directement dans cet ensemble, une heuristique consiste à les rechercher en réduisant tout d'abord le nombre de scénarios au moyen d'un plan fractionnaire orthogonal (comme dans le cas de classement de scénarios complets).

Le tableau n°7 est une des fractions orthogonales au $\frac{1}{4}$ du plan complet.

Nous appliquons par la suite la procédure utilisée dans l'exemple précédent pour réduire le nombre de paires.

TABLEAU 7. — Plan fractionnaire $\frac{1}{4}$ (cas de l'exemple de cinq variables dichotomiques)

N °scénario	*Couleur*	*Matériau*	*Semelle*	*Modèle*	*Forme*
1	*noire*	*cuir*	*plate*	*chaussures*	*pointue*
2	*marron*	*cuir*	*plate*	*boots*	*arrondie*
3	*noire*	*cuir*	*à talon*	*boots*	*pointue*
4	*marron*	*cuir*	*à talon*	*chaussures*	*arrondie*
5	*noire*	*daim*	*plate*	*chaussures*	*arrondie*
6	*marron*	*daim*	*plate*	*boots*	*pointue*
7	*noire*	*daim*	*à talon*	*boots*	*arrondie*
8	*marron*	*daim*	*à talon*	*chaussures*	*pointue*

Le plan fractionnaire $\frac{1}{4}$ du tableau n°7 donne 28 paires candidates (au lieu des 496 initiales, soit $94,4\%$ de réduction). Ce nombre est encore trop élevé, nous utilisons donc la procédure décrite dans l'exemple précédent pour le réduire plus.

Le groupement des paires égales en valeur absolue nous donne 7 groupes de 4 paires chacun. En gardant une paire par groupe nous obtenons les paires (1,2), (1,3), (1,4), (1,5), (1,6), (1,7), et (1,8) comme l'indique le tableau n°8.

TABLEAU 8. — Paires obtenues à la première réduction (cas de l'exemple de cinq variables dichotomiques)

1	2	*noire*	*marron*	*cuir*	*cuir*	*plate*	*plate*	*chaussures*	*boots*	*pointue*	*arrondie*
1	3	*noire*	*noire*	*cuir*	*cuir*	*plate*	*à talon*	*chaussures*	*boots*	*pointue*	*pointue*
1	4	*noire*	*marron*	*cuir*	*cuir*	*plate*	*à talon*	*chaussures*	*chaussures*	*pointue*	*arrondie*
1	5	*noire*	*noire*	*cuir*	*daim*	*plate*	*plate*	*chaussures*	*chaussures*	*pointue*	*arrondie*
1	6	*noire*	*marron*	*cuir*	*daim*	*plate*	*plate*	*chaussures*	*boots*	*pointue*	*pointue*
1	7	*noire*	*noire*	*cuir*	*daim*	*plate*	*à talon*	*chaussures*	*boots*	*pointue*	*arrondie*
1	8	*noire*	*marron*	*cuir*	*daim*	*plate*	*à talon*	*chaussures*	*chaussures*	*pointue*	*pointue*

Aucune des paires présentées n'évalue un seul attribut, cela vient du fait que le plan fractionnaire utilisé pour la construction des paires est orthogonal.

La réduction ici est de 75% (passage de 28 à 7 paires), soit une réduction totale de $98,6\%$ (passage de 496 à 7 paires).

Remarque 4.2.1. — Nous avons effectivement utilisé la procédure mise au point pour réduire le nombre de paires dans cet exemple des chaussures. L'exemple n'est pas détaillé ici, car trop long et un peu hors du cadre de cet article.

Remarque 4.2.2. — Dans certains cas on peut utiliser des plans en blocs incomplets équilibrés pour la réduction préliminaire.

4.3. Cas d'un nombre quelconque de variables dichotomiques

D'une manière générale et pour un nombre quelconque de variables dichotomiques nous adoptons la procédure standard, ce qui nous donne les résultats suivants :

TABLEAU 9. — Récapitulatif des résultats de la procédure de réduction (cas de variables dichotomiques)

Nombre de variables : k	2	3	4	5	6	7	8
Nombre de produits	4	8	16	32	64	128	256
Nombre initial de paires	6	28	120	496	2016	8128	32640
Nombre de paires (première réduction)	3	7	15	31	63	127	255
Nombre de paires (deuxième réduction)	1	4	11	26	57	120	247
Réduction (en%)	83,3	85,7	90,8	94,7	97,2	98,5	99,2

Nous remarquons que le nombre de produits restant après la deuxième réduction dépend uniquement du nombre de variables. Soit p_k ce nombre, nous avons alors la relation suivante :

$$p_k = 2^k - (1+k), \qquad k > 1 \tag{1}$$

En effet :

à la première étape nous obtenons, par regroupement des paires égales en valeur absolues $(n-1)$ groupes de $\frac{n}{2}$ éléments chacun, ou n désigne le nombre de produits $(n = 2^k)$.

En ne gardant qu'un seul élément par groupe, il ne reste que $(n - 1)$ paires.

À la deuxième étape, nous recherchons les paires qui n'évaluent qu'un seul attribut, et il y en a exactement k paires.

En supprimant les paires n'évaluant qu'un seul attribut, il nous reste $(n-1-k)$ paires candidates, ce qui donne (1).

Il est évident que le cas extrême ou $k = 1$ (un seul produit) donne un nombre de paires nul.

La réduction du nombre de paires dépasse les 90% dès que le nombre de variables dépasse 3.

La relation (1) peut aisément s'écrire sous forme d'une relation de récurrence entre p_k et p_{k-1}

$$\begin{cases} p_k = 2 \times p_{k-1} + (k-1), & k > 1 \\ p_1 = 0 \end{cases}$$

4.4. Cas de variables à trois modalités

Considérons maintenant les cas de produits issus de variables non dichotomiques, et particulièrement, de variables à trois modalités.

4.4.1. Cas de deux variables à 3 modalités

Dans le cas où les produits sont issus de deux variables avec trois modalités chacune, le nombre total de produits est alors égal à $3^2 = 9$ ce qui donne le plan complet décrit par le tableau n°10.

TABLEAU 10. — Plan complet (cas de deux variables à trois modalités)

P_1	1 0 0	1 0 0
P_2	1 0 0	0 1 0
P_3	1 0 0	0 0 1
P_4	0 1 0	1 0 0
P_5	0 1 0	0 1 0
P_6	0 1 0	0 0 1
P_7	0 0 1	1 0 0
P_8	0 0 1	0 1 0
P_9	0 0 1	0 0 1

Le nombre total de paires est alors de $\frac{(9 \times (9-1))}{2}$, soit 36 paires ce qui donne les paires décrites par le tableau n°11.

TABLEAU 11. — Codage des paires (cas de deux variables à trois modalités)

P_1 P_2	0 0 0	1 −1 0	P_3 P_7	1 0 −1	−1 0 1
P_1 P_3	0 0 0	1 0 −1	P_3 P_8	1 0 −1	0 −1 1
P_1 P_4	1 −1 0	0 0 0	P_3 P_9	1 0 −1	0 0 0
P_1 P_5	1 −1 0	1 −1 0	P_4 P_5	0 0 0	1 −1 0
P_1 P_6	1 −1 0	1 0 −1	P_4 P_6	0 0 0	1 0 −1
P_1 P_7	1 0 −1	0 0 0	P_4 P_7	0 1 −1	0 0 0
P_1 P_8	1 0 −1	1 −1 0	P_4 P_8	0 1 −1	1 −1 0
P_1 P_9	1 0 −1	1 0 −1	P_4 P_9	0 1 −1	1 0 −1
P_2 P_3	0 0 0	0 1 −1	P_5 P_6	0 0 0	0 1 −1
P_2 P_4	1 −1 0	−1 1 0	P_5 P_7	0 1 −1	−1 1 0
P_2 P_5	1 −1 0	0 0 0	P_5 P_8	0 1 −1	0 0 0
P_2 P_6	1 −1 0	0 1 −1	P_5 P_9	0 1 −1	0 1 −1
P_2 P_7	1 0 −1	−1 1 0	P_6 P_7	0 1 −1	−1 0 1
P_2 P_8	1 0 −1	0 0 0	P_6 P_8	0 1 −1	0 −1 1
P_2 P_9	1 0 −1	0 1 −1	P_6 P_9	0 1 −1	0 0 0
P_3 P_4	1 −1 0	−1 0 1	P_7 P_8	0 0 0	1 −1 0
P_3 P_5	1 −1 0	0 −1 1	P_7 P_9	0 0 0	1 0 −1
P_3 P_6	1 −1 0	0 0 0	P_8 P_9	0 0 0	0 1 −1

En regroupant les paires égales en valeur absolue nous obtenons 6 groupes formés de 3 paires chacun et 9 groupes formés de 2 paires chacun (soit deux types de groupes de paires ici) décrits par le tableau n°12.

TABLEAU 12. — Groupement des paires et formes des groupes (cas de deux variables à trois modalités)

$P_4\ P_7$	$0\ 1\ -1$	$0\ 0\ 0$	$P_1\ P_7$	$1\ 0\ -1$	$0\ 0\ 0$	$P_1\ P_4$	$1-1\ 0$	$0\ 0\ 0$
$P_5\ P_8$	$0\ 1\ -1$	$0\ 0\ 0$	$P_2\ P_8$	$1\ 0\ -1$	$0\ 0\ 0$	$P_2\ P_5$	$1-1\ 0$	$0\ 0\ 0$
$P_6\ P_9$	$0\ 1\ -1$	$0\ 0\ 0$	$P_3\ P_9$	$1\ 0\ -1$	$0\ 0\ 0$	$P_3\ P_6$	$1-1\ 0$	$0\ 0\ 0$
$P_2\ P_3$	$0\ 0\ 0$	$0\ 1\ -1$	$P_1\ P_3$	$0\ 0\ 0$	$1\ 0\ -1$	$P_1\ P_2$	$0\ 0\ 0$	$1\ -1\ 0$
$P_5\ P_6$	$0\ 0\ 0$	$0\ 1\ -1$	$P_4\ P_6$	$0\ 0\ 0$	$1\ 0\ -1$	$P_4\ P_5$	$0\ 0\ 0$	$1\ -1\ 0$
$P_8\ P_9$	$0\ 0\ 0$	$0\ 1\ -1$	$P_7\ P_9$	$0\ 0\ 0$	$1\ 0\ -1$	$P_7\ P_8$	$0\ 0\ 0$	$1\ -1\ 0$
$P_4\ P_8$	$0\ 1\ -1$	$1\ -1\ 0$	$P_2\ P_9$	$1\ 0\ -1$	$0\ 1\ -1$	$P_1\ P_9$	$1\ 0\ -1$	$1\ 0\ -1$
$P_5\ P_7$	$0\ 1\ -1$	$1\ -1\ 0$	$P_3\ P_8$	$1\ 0\ -1$	$0\ -1\ 1$	$P_3\ P_7$	$1\ 0\ -1$	$-1\ 0\ 1$
$P_1\ P_8$	$1\ 0\ -1$	$1\ -1\ 0$	$P_2\ P_6$	$1\ -1\ 0$	$0\ 1\ -1$	$P_1\ P_6$	$1\ -1\ 0$	$1\ 0\ -1$
$P_2\ P_7$	$1\ 0\ -1$	$-1\ 1\ 0$	$P_3\ P_5$	$1\ -1\ 0$	$0\ -1\ 1$	$P_3\ P_4$	$1\ -1\ 0$	$-1\ 0\ 1$
$P_1\ P_5$	$1\ -1\ 0$	$1\ -1\ 0$	$P_5\ P_9$	$0\ 1\ -1$	$0\ 1\ -1$	$P_4\ P_9$	$0\ 1\ -1$	$1\ 0\ -1$
$P_2\ P_4$	$1\ -1\ 0$	$-1\ 1\ 0$	$P_6\ P_8$	$0\ 1\ -1$	$0\ -1\ 1$	$P_6\ P_7$	$0\ 1\ -1$	$-1\ 0\ 1$

Nous gardons une paire de chaque groupe (la première par exemple), ce qui nous donne 15 paires (tableau n°13).

TABLEAU 13. — Paires obtenues à la première réduction (cas de deux variables à trois modalités)

$P_1\ P_2$	$0\ 0\ 0$	$1\ -1\ 0$
$P_1\ P_3$	$0\ 0\ 0$	$1\ 0\ -1$
$P_1\ P_4$	$1-1\ 0$	$0\ 0\ 0$
$P_1\ P_5$	$1-1\ 0$	$1\ -1\ 0$
$P_1\ P_6$	$1-1\ 0$	$1\ 0\ -1$
$P_1\ P_7$	$1\ 0\ -1$	$0\ 0\ 0$
$P_1\ P_8$	$1\ 0\ -1$	$1\ -1\ 0$
$P_1\ P_9$	$1\ 0\ -1$	$1\ 0\ -1$
$P_2\ P_3$	$0\ 0\ 0$	$0\ 1\ -1$
$P_2\ P_6$	$1-1\ 0$	$0\ 1\ -1$
$P_2\ P_9$	$1\ 0\ -1$	$0\ 1\ -1$
$P_4\ P_7$	$0\ 1\ -1$	$0\ 0\ 0$
$P_4\ P_8$	$0\ 1\ -1$	$1\ -1\ 0$
$P_4\ P_9$	$0\ 1\ -1$	$1\ 0\ -1$
$P_5\ P_9$	$0\ 1\ -1$	$0\ 1\ -1$

En supprimant les paires qui n'évaluent qu'un seul attribut nous obtenons les 9 paires décrites par le tableau n°14.

TABLEAU 14. — Paires obtenues à la deuxième réduction (cas de deux variables à trois modalités)

P_1 P_5	$1 -1 \quad 0$	$1 -1 \quad 0$
P_1 P_6	$1 -1 \quad 0$	$1 \quad 0 -1$
P_1 P_8	$1 \quad 0 -1$	$1 -1 \quad 0$
P_1 P_9	$1 \quad 0 -1$	$1 \quad 0 -1$
P_2 P_6	$1 -1 \quad 0$	$0 \quad 1 -1$
P_2 P_9	$1 \quad 0 -1$	$0 \quad 1 -1$
P_4 P_8	$0 \quad 1 -1$	$1 -1 \quad 0$
P_4 P_9	$0 \quad 1 -1$	$1 \quad 0 -1$
P_5 P_9	$0 \quad 1 -1$	$0 \quad 1 -1$

Nous remarquons que toutes les paires qui évaluent un seul attribut étaient initialement dans les groupes formés de trois paires, et donc que les paires retenus sont issue uniquement des groupes de deux paires.

4.4.2. Cas de plus de deux variables à 3 modalités

Dans le cas de **deux** variables à 3 modalités, nous avons obtenu 9 produits et 36 paires possibles. La première étape de la procédure de réduction nous donne 6 groupes de trois paires et 9 groupes de 2 paires (égales en valeurs absolue), soit **deux** types de groupes de paires. En retenant une paire par groupe il en reste 15 à la fin de la première étape de la procédure de réduction.

Dans le cas de **trois** variables à trois modalités, le nombre de produits étant égal à 27, le nombre de paires possibles est alors égal à 351. La première étape nous donne une décomposition en 9 groupes de 9 paires, 27 groupes de 6 paires et 27 groupes de 4 paires (égales en valeur absolue), soit ici **trois** types de groupes de paires. Nous retenons une paire par groupe soit 63 paires.

La deuxième étape nous donne 9 paires évaluant un seul attribut ; nous remarquons que ces paires proviennent toutes des groupes à 9 paires. Nous obtenons donc 54 paires en fin de procédure.

D'une manière plus générale et pour un nombre $\underline{k}$ quelconque de variables à trois modalités nous adoptons la même procédure de réduction du nombre de paires. Le nombre initial de paires se décompose en un nombre $\underline{k}$ de type de paires (égal au nombre de variables).

Notons P_k, le nombre initial de paires, P_k s'écrit sous la forme :

$$P_k = 3 \times k \times n_1 + 3^k \times (n_2 + \cdots + n_k).$$

Avec n_i pour $i = 1, \cdots, k$ le nombre de paires du ième type de paires. En fait ce nombre n'est pas très important puisqu'il n'intervient pas dans la suite, mais nous pouvons le donner à titre indicatif pour les cas de deux et trois variables :

Pour $k = 2$ on a : $P_2 = 6 \times (3) + 9 \times (2)$ donc $n_1 = 3$ et $n_2 = 2$

Pour $k = 3$ on a : $P_3 = 9 \times (9) + 27 \times (6) + 27 \times (4)$ donc $n_1 = 9$ $n_2 = 6$ et $n_3 = 4$

Les paires évaluant un seul attribut proviennent toujours d'un seul type de groupe de paires, celui de taille n_1, les autres types de groupes étant toujours au nombre de 3^k.

Pour un nombre de variables supérieur à 3 les résultats sont dans le tableau n°15.

TABLEAU 15. — Récapitulatif des résultats de la procédure de réduction (cas de variables à trois modalités)

Nombre de variables : k	2	3	4	5	6	7
Nombre de produits	9	27	81	243	729	2187
Nombre initial de paires	36	351	3240	29403	265356	2390391
Nombre de types de groupes de paires	2	3	4	5	6	7
Nombre de paires (première réduction)	15	63	255	987	3663	13143
Nombre final de paires : p_k	9	54	243	972	3645	13122
Réduction (en%)	75	84,6	92,5	96,7	98,6	99,5

Le nombre de produits restant après la deuxième réduction dépend uniquement du nombre de variables. Soit p_k ce nombre, nous avons alors la relation suivante :

$$p_k = 3^k \times (k-1), \qquad k > 1 \tag{2}$$

En effet, au départ nous avons exactement k types de groupes de paires. Nous gardons à la première étape une paire par groupe. Il reste 1 groupe du type 3k et (k-1) groupes de type 3^k soit donc $3 \times k + 3^k \times (k-1)$ groupes. À la deuxième étape de la procédure, nous supprimons les paires n'évaluent qu'un seul attribut soit $3k$ paires ce qui nous donne le résultat annoncé par la relation (2).

Nous remarquons que la réduction du nombre de paires dépasse les 90% dès que le nombre de variable dépasse 3.

La relation (2) peut aisément s'écrire sous forme d'une relation de récurrence entre p_k et p_{k-1}

$$\begin{cases} p_k = 3 \times p_{k-1} + 3^k, & k > 1 \\ \\ p_1 = 0 \end{cases}$$

4.4.3. Cas de mélanges de variables dichotomiques et de variables à 3 modalités

Considérons le cas de deux variables dont une dichotomique et l'autre à 3 modalités. Le nombre total de produits est égal à 6 et le nombre de paires associé est égal à 15. Le regroupement des paires égales en valeur absolue nous donne 6 groupes de deux paires et un groupe de 3 paires. La première réduction nous donne donc 7 paires, dont 4 qui évaluent un seul attribut.

Le nombre de paires restant est alors égal à 3 et la réduction est de 80%. Nous remarquons ici que le 3 est le nombre minimal de paires pour un modèle identifiable.

Le tableau n°16 nous donne quelques résultats pour certains mélanges de variables.

TABLEAU 16. — Quelques résultats pour certains mélanges de variables

Nombre de variales dichotomiques : k	1	2	3	4	1	2	2
Nombre de variables à 3 modalités : l	1	1	1	1	2	2	3
Nombre de produits	6	12	24	48	18	36	108
Nombre initial de paires	15	66	276	1128	153	630	5778
Nombre de paires (première réduction)	7	15	31	63	31	63	255
Nombre final de paires : p_{kl}	3	10	25	56	24	55	244
Réduction (en%)	80	84,8	90,9	95	84,3	91,3	95,7

Le nombre de paires pour un mélange de k variables à deux modalités et d'une seule variable à trois modalités est donné par la relation (3).

$$p_{k1} = 3\times(2^k - 1)+2^k-(1+k), \qquad k > 1 \tag{3}$$

Nous n'avons pas encore trouvé de formule générale pour p_{kl} pour des valeurs de l supérieures ou égales à 2.

5. Transformation des variables

Nous avons vu que l'application de la procédure de réduction proposée permet de réduire considérablement le nombre de paires candidates. Le nombre de paires obtenues dépend uniquement du nombre de variables et implicitement du nombre de modalités de chaque variable. Ceci donne à chaque fois une relation différente selon le nombre de modalité des variables (les relations (1), (2) et (3)). Pour remédier à ce problème nous pouvons ramener les variables ayant un nombre de modalités supérieur à 2 au cas dichotomique.

Les plans optimaux pour variables dichotomiques ont été beaucoup étudiés. Nous citons par exemple Shah et Sinha (1989), Atkinson et Donev (1992) ainsi que Burgess et Street (2003). L'utilisation dans notre cas de tels types de plans nous permet d'avoir le nombre de paires en fonction du nombre de variables uniquement par utilisation de la relation (1).

En effet l'utilisation d'une variable à k modalités est équivalente à celle de k variables dichotomiques, dans le cas du modèle additif simple, ce qui est le cas des comparaisons par paires.

Le plan complet issu de variables dichotomiques est légèrement différent du plan issu de variables quelconques. Une variable à k modalités sera codée avec (k-1) zéros et un seul 1, alors que k variables dichotomiques seraient codées avec k zéros et k uns. Ce genre de transformation engendre des produits fictifs supplémentaires faciles à repérer. Il suffit de recoder les produits du plan initial sous forme dichotomique et de les comparer aux produits du plan construit à l'aides des variables transformées.

La suppression des produits fictifs permet de retrouver le même plan dans les deux cas. Cette suppression engendre une étape supplémentaire préalable à la procédure de réduction proposée.

Prenons par exemple le cas de deux variables X_1 et X_2 à 3 modalités chacune, le plan complet sera de taille 9×6 (tableau n°10), la matrice d'expérience est formée de 2 blocs de 3 colonnes. Le recodage sous forme dichotomique du plan initial nous donne les produits décrits dans le tableau n°17.

TABLEAU 17. — Plan complet recodé sous forme dichotomique (cas de deux variables à trois modalités)

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

La transformation de X_1 et X_2 donne 6 variables dichotomiques. Le plan complet se présente sous forme d'un tableau disjonctif complet de dimension 64×12, soit 55 produits fictifs supplémentaires.

La comparaison des deux matrices d'expérience nous redonne les produits du plan initial et nous permet de supprimer les produits fictifs.

De manière plus générale un produit fictif est facile à repérer puisqu'il représente un produit irréalisable.

Si on considère par exemple la première variable X_1 qui est donc associée aux 6 premières colonnes du tableau disjonctif complet 64×12, un produit sera irréalisable dans une des 3 situations suivantes :

- Il prend 2 fois la première modalité d'une des 3 premières variables dichotomiques, ce qui signifie qu'il prend 2 modalités différentes de la variable X_1, ce qui est bien sûr impossible.
- Il prend 3 fois la première modalité des 3 premières variables dichotomiques, ce qui signifie qu'il prend les 3 modalités de X_1, ce qui est encore impossible.
- Il prend 3 fois la seconde modalité des 3 premières variables dichotomiques, ce qui signifie qu'il ne prend aucune modalité de X_1 ce qui est évidemment impossible (puisque le produit étant connu, il n'y a pas de non réponses).

Les produits réalisables doivent présenter exactement une modalité et une seule pour chacune des variables initiales (position des zéros et des 1).

Nous pouvons utiliser par la suite la procédure de réduction du nombre de paires dans le cas dichotomique.

La transformation de variables ayant un nombre quelconque de modalités en variables dichotomiques, permet donc l'utilisation de la procédure de réduction pour des mélanges de variables avec un nombre quelconque de modalités. L'automatisation de cette procédure peut se faire aisément moyennant une étape préliminaire de transformations des variables et de suppression des produits fictifs, et un logiciel approprié pour sa programmation. Elle peut aussi être combinée dans certains cas particuliers (cas de l'exemple des chaussures §4.2) avec l'utilisation de plans fractionnaires optimaux.

6. Conclusion

La méthode de comparaison par paires en analyse conjointe fréquemment utilisée pour le lancement de nouveaux produits, se trouve handicapée par l'explosion combinatoire du nombre de paires de produits candidates dès que le nombre de produits proposés dépasse la dizaine. Nous avons donné une procédure de réduction du nombre de paires illustrée par des exemples simples dans le cas dichotomique et le cas de produits issus de variables à 3 modalités. Cette procédure permet des réductions supérieures à 90% dés que le nombre de variables dépasse trois. Nous donnons aussi les formules de calcul du nombre de paires en fonction du nombre de variables. La procédure est généralisable à des variables avec un nombre quelconque de modalités, moyennant une transformation de ces dernières.

Références

ATKINSON A. C. et DONEV A. N. (1992), *Optimum experimental designs.* Clarendon Press.

BENAMMOU S., HARBI S. et SAPORTA G. (2003), Sur l'utilisation de l'analyse conjointe en cas de réponses incomplètes ou de non réponses, *Revue de Statistique Appliquée*, 51, 31- 55.

BURGESS L et STREET D.J. (2003), Optimal Designs for 2^k Choice Experiments, Communication in Statistics, vol.32, 2185-2206.

CARROLL J.D. et GREEN P.E (1995), Psychometric methods in marketing research : part I, conjoint analysis, *Journal of Marketing Research*, Vol XXXII, 385-391.

CATTIN P et WITTINK D.R. (1982), Commercial Use of Conjoint Analysis : A Survey, *Journal of Marketing*, Vol.46, 44-53.

CATTIN P et WITTINK D.R. (1989), Commercial Use of Conjoint Analysis : An Update, *Journal of Marketing*, Vol.53, 91-96.

CHALONER K. (1984), Optimal bayesian experimental design for linear models. *Ann. Stat*, 12, 1, 283-300.

COURCOUX P.H et SEMENOU M. (1997), Une méthode de segmentation pour l'analyse de données issues de comparaisons par paires, *Revue de Statistique Appliquée*, XLV, 59-69.

DROESBEKE J.J, FINE J. et SAPORTA G. (1997), Plans d'Expériences, *Applications à l'entreprise*, Editions Technip, Paris.

FAIVRE J.-Ph., PIOCHE A. (1976), L'Analyse des Décisions d'Achat par le Modèle Tradeoff : Problèmes Théoriques et Méthodologiques, *Revue Française du Marketing*, Cahiers 64-65.

GAUCHI J.P. (1999), Contribution à l'étude et au calcul de critères de plans d'expériences optimaux pour des modèles de régression non linéaire, thèse de doctorat du Conservatoire National des Arts et Métiers de Paris.

GRABHOF U., GROBMANN H et HOLLING H. (2004), Optimal designs for main effects in linear paired comparison models, *Journal of Statistical Planning and Inference*, 126, 361-376.

GREEN P.E et SRINIVASAN V. (1978), Conjoint Analysis in Consumer Research : Issues and Outlook, *Journal of Consumer Research*, 5 (September), 103-123.

GREEN P.E et SRINIVASAN V. (1990), Conjoint Analysis in Marketing : New Developments with Implications for Research and Practice, *Journal of Marketing* 54, (October), 3-19.

GREEN P.E et RAO V.R. (1971), Conjoint Measurement for Quantifying Judgmental Data, *Journal of Marketing Research*, 8 (August), 355-363.

GUSTAFSSON A., HERRMANN A., HUBER F.(2003), *Conjoint Measurement : Methods and Applications*, Third Edition, Springer.

KUHFELD W. F. (2005), *Marketing Research Methods in SAS : Experimental Design, Choice, Conjoint, and Graphical Techniques*, SES 9.1 Edition, SAS Institute Inc., Cary, NC, USA.

KUNG-E CH., ZHENG L., YUANQIONG W., STARR H. et TUROFF M. (2001), Thurstone's law of comparative judgment for group support, *Data Management and Decision Support*, Seventh Americas Conference on Information Systems, 241-244.

SHAH K.R, SINHA B.K. (1989) Theory of Optimal designs, *Lecture Notes in Statistics* n°54, Springer, Verlag.

THURSTONE L.L. (1927). A law of comparative judgment. *Psychological Review*, 34, 278-286.