

CAHIER DU LAMSADE

Laboratoire d'Analyse et Modélisation de Systèmes pour l'Aide à la Décision
(Université de Paris-Dauphine)

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

AMELIORATION DE LA METHODE UTA PAR INTRODUCTION D'UNE DOUBLE FONCTION D'ERREURS

CAHIER N° 49
octobre 1983

J. SISKOS
D. YANNAKOPOULOS

SOMMAIRE

	<u>Pages</u>
ABSTRACT	I
RESUME	II
1. INTRODUCTION	1
2. UTA AMELIOREE VERSUS UTA	3
3. UN EXEMPLE NUMERIQUE	5
4. UNE EXPERIMENTATION ET CONCLUSION	7
REFERENCES	12

IMPROVING THE UTA ORDINAL REGRESSION METHOD
BY INTRODUCTION OF A DOUBLE ERROR FUNCTION

ABSTRACT

This paper presents an improved version of the UTA method performing an ordinal regression analysis using more powerful linear programming formulations. The ordinal variable to be analysed is a pre-order relation whereas the independent variables are criteria, i.e. quantitative and/or qualitative monotone variables. The method is illustrated on a simple numerical example. Finally, experimental results are given, showing, upon three distinct indicators, the superiority of the adjustments obtained with the new method.

Key-words : Ordinal regression ; Additive utility ; Multicriteria analysis.

AMELIORATION DE LA METHODE UTA
PAR INTRODUCTION D'UNE DOUBLE FONCTION D'ERREURS

RESUME

Ce cahier présente une version améliorée de la méthode UTA effectuant une analyse de régression ordinale à l'aide de formulations de programmation linéaire différentes. La variable ordinale à analyser est un préordre de préférence et les variables indépendantes des critères multiples, c'est-à-dire des variables quantitatives et/ou qualitatives. La méthode est illustrée sur un exemple numérique simple. Finalement, on montre la supériorité des ajustements obtenus à l'aide de la nouvelle méthode en utilisant trois indicateurs distincts.

Mots-clés : Régression ordinale ; Utilité additive ; Analyse multicritère.

1. INTRODUCTION

Le problème de régression ordinale, tel qu'il est traité par la méthode UTA (cf. Jacquet-Lagrèze et Siskos [2]), s'explicit de la façon suivante : soit $(\succ, \sim)$ une structure de préférences sous forme de préordre donnée sur un ensemble d'objets ou actions où $\succ$ désigne la préférence et $\sim$ l'indifférence ; on cherche à ajuster des fonctions d'utilité additives, sur critères multiples, de façon que la structure préférentielle résultant de ces utilités soit aussi compatible que possible avec la structure initiale.

Appelons $A = \{a, b, c, \dots\}$ l'ensemble des actions sur lequel est donnée la structure $(\succ, \sim)$. Soit $g_1, g_2, \dots, g_n$ une famille de n critères d'évaluation ; chacun est défini ici sous la forme d'une application réelle monotone $g_i : A \rightarrow [g_{i*}, g_i^*] \subseteq \mathbb{R}$ de façon à ce que $g_i(a)$ représente l'évaluation de l'action a sur le critère g_i . Les valeurs $g_{i*} = \min_{a \in A} g_i(a)$, $g_i^* = \max_{a \in A} g_i(a)$ représentent respectivement le niveau le moins et le plus préféré du critère.

Au niveau d'un seul critère, les préférences s'explicitent de la façon suivante :

$$a \succ b \iff g_i(a) > g_i(b) \quad (1)$$

$$a \sim b \iff g_i(a) = g_i(b), \quad (2)$$

ce qui signifie que chaque critère définit sur l'ensemble A une relation de préordre $(\succ, \sim)$.

Une fonction d'utilité en présence de certitude est une application réelle $u : \bigtimes_{i=1}^n [g_{i*}, g_i^*] \rightarrow \mathbb{R}$ définie par les relations :

$$a \succ b \iff u[g(a)] > u[g(b)] \quad (3)$$

$$a \sim b \iff u[g(a)] = u[g(b)] \quad (4)$$

avec $\underline{g}(a) = [g_1(a), g_2(a), \dots, g_n(a)]$ le profil des évaluations de l'action a sur les n critères.

La méthode UTA cherche à estimer des utilités additives du type :

$$u(\underline{g}) = u_1(g_1) + u_2(g_2) + \dots + u_n(g_n) \quad (5)$$

avec

$$u_i(g_{i*}) = 0 \quad \forall i \quad (6)$$

et

$$u_1(g_1^*) + u_2(g_2^*) + \dots + u_n(g_n^*) = 1. \quad (7)$$

Les relations (6) et (7) consistent à normaliser la fonction d'utilité totale u entre 0 et 1.

Dans UTA ([2]), il existe une seule fonction d'erreur $\sigma : A \rightarrow [0, 1]$ où $\sigma(a) \geq 0$ l'écart éventuel entre l'utilité $u[\underline{g}(a)]$ estimée et l'utilité qu'il conviendrait d'attribuer à l'action a pour que celle-ci puisse récupérer son rang dans le préordre (voir figure 1). Néanmoins, cette fonction d'erreurs n'est pas suffisante pour minimiser complètement la dispersion des points autour de la courbe de régression. Le problème est posé notamment pour les points situés à droite de la courbe pour lesquels il faudrait soustraire une erreur plutôt que d'en ajouter aux autres.

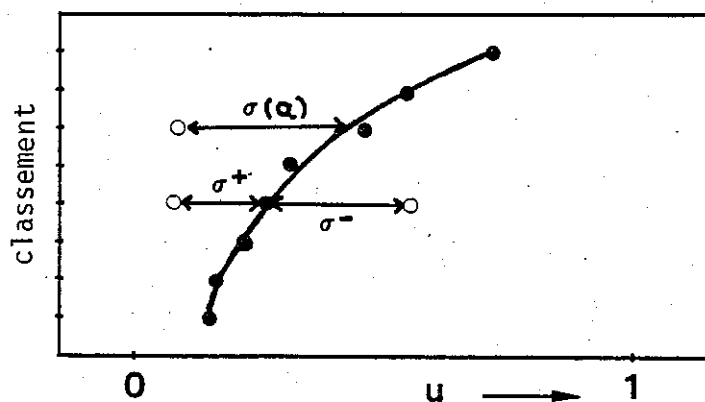


Figure 1 : Utilité vs préordre en régression ordinale

Dans la version d'UTA développée dans ce cahier, nous proposons une fonction d'erreurs composée de deux erreurs potentielles par action. Cela permettrait de mieux stabiliser la position des points autour de la courbe (figure 1). Ainsi, l'utilité d'une action $a \in A$ sera remplacée dans les relations (3)-(4) par : $u[\underline{g}(a)] + \sigma^+(a) - \sigma^-(a)$, $\sigma^+(a) \geq 0$, $\sigma^-(a) \geq 0$.

Dans la section suivante, nous présentons succinctement le modèle UTA ainsi que les améliorations introduites. La section 3 permet une comparaison des deux modèles à l'aide d'un exemple numérique à cinq actions et trois critères. Dans la dernière section, nous montrons expérimentalement la supériorité du nouveau modèle sur l'ancien en utilisant trois indicateurs différents.

2. UTA AMELIOREE VERSUS UTA

UTA utilise une formulation de programmation linéaire pour estimer les utilités marginales u_i sous les conditions (5)-(7). La technique d'estimation présuppose la discrétisation de chaque intervalle de variation des critères

$$[g_{i*}, g_i^*] = [g_{i*} \equiv g_i^1, g_i^2, \dots, g_i^{\alpha_i} \equiv g_i^*] \quad (8)$$

et l'introduction de contraintes $u_i(g_i^{j+1}) \geq u_i(g_i^j)$, $\forall i$ afin de préserver la monotonie des critères. Le nombre de points équidistants α_i peut se calculer par algorithme dans la mesure de l'information disponible ; pour les critères quantitatifs, nous employons la technique de l'interpolation linéaire.

Suivant les prescriptions (3)-(7), l'ancien modèle UTA se schématise en quatre étapes :

1) Exprimer, dans l'ordre imposé par le préordre initial $(\succ, \sim)$, les utilités des actions $u[\underline{g}(a)]$, $a \in A$ en fonction des utilités marginales $u_i(g_i^j)$.

2) Définir une chaîne sur le préordre allant de la tête à la queue en écrivant, pour chaque paire (a, b) d'actions consécutives, les expressions analytiques :

$$\Delta(a, b) \equiv u[\underline{g}(a)] - u[\underline{g}(b)] + \sigma(a) - \sigma(b). \quad (9)$$

Le nombre de ces expressions est égal au nombre d'actions moins 1.

3) Résoudre le dual du programme linéaire :

$$\text{Minimiser } F = \sum_{a \in A} \sigma(a)$$

sous les contraintes (d'après l'étape 2)

$$\Delta(a, b) \geq \delta \quad \text{si } a \succ b$$

$$\Delta(a, b) = 0 \quad \text{si } a \sim b$$

$$u_i(g_i^{j+1}) - u_i(g_i^j) \geq 0 \quad \forall i \text{ et } j$$

$$\sum_i u_i(g_i^*) = 1$$

$$u_i(g_{i*}) = 0, u_i(g_i^j) \geq 0, \sigma(a) \geq 0 \quad \forall a \in A, \forall i \text{ et } j$$

δ une petite valeur positive.

4) Tester l'existence de solutions optimales multiples. En cas de dégénérescence, trouver celles des solutions qui maximisent et/ou minimisent $u_i(g_i^*)$ pour chaque i .

Les modifications intégrées dans le nouveau modèle sont, par étape, les suivantes :

1) Les contraintes de monotonie des critères sont prises en compte dans les transformations de variables

$$w_{ij} = u_i(g_i^{j+1}) - u_i(g_i^j) \geq 0 \quad \forall i \text{ et } j. \quad (10)$$

Les utilités $u[\underline{g}(a)]$ deviennent donc fonctions de w_{ij} , c'est-à-dire, comme $u_i(g_i^1) = 0$, on a, pour $j > 1$:

$$u_i(g_i^j) = \sum_{k=1}^{j-1} w_{ik}. \quad (11)$$

2) Introduction d'une double fonction d'erreurs : poser pour chaque paire d'actions consécutives (a, b) dans la chaîne

$$\Delta(a, b) \equiv u[g(a)] - u[g(b)] + \sigma^+(a) - \sigma^-(a) - \sigma^+(b) + \sigma^-(b). \quad (12)$$

3) Résoudre le primal du programme linéaire :

$$\text{Minimiser } F = \sum_{a \in A} [\sigma^+(a) + \sigma^-(a)]$$

sous les contraintes

$$\Delta(a, b) \geq \delta \quad \text{si } a \succ b$$

$$\Delta(a, b) = 0 \quad \text{si } a \sim b$$

$$\sum_i \sum_j w_{ij} = 1$$

$$w_{ij} \geq 0, \sigma^+(a) \geq 0, \sigma^-(a) \geq 0 \quad \forall a \in A, \forall i \text{ et } j$$

δ une petite valeur positive.

4) Sans changement.

Remarque : Cette formulation s'apparente à celles développées en "goal programming" ([1]). Il est donc facile de prouver qu'à l'optimum on aura $\sigma^+(a) \cdot \sigma^-(a) = 0, \forall a \in A$, c'est-à-dire au moins l'une des erreurs nulle.

3. UN EXEMPLE NUMERIQUE

Considérons le cas d'un individu dont on veut analyser les choix de modes de transport domicile-lieu de travail pendant les heures d'affluence. Cet individu s'intéressant uniquement aux trois critères (1) prix (en Francs), (2) temps du trajet (en minutes) et (3) confort (possibilités de s'asseoir), attribue aux cinq modes possibles le classement (pré-ordre) suivant : RER $\succ$ (METRO 1ère $\sim$ METRO 2e) $\succ$ BUS $\sim$ TAXI (cf. Tableau 1). Pour le critère confort, une échelle qualitative a été utilisée : 0 : aucune chance de s'asseoir, + : peu de chances de s'asseoir, ++ : grandes chances de trouver une place assise, +++ : place assise certaine.

Tableau 1 : Problème d'analyse du classement de modes de transport

Modes de transport	Rangs	Prix (Francs)	Temps (minutes)	Confort (qualitatif)
RER	1	3	10	+
METRO 1ère	2	4	20	++
METRO 2e	2	2	20	0
BUS	3	6	40	0
TAXI	4	30	30	+++

La première étape d'UTA consiste à expliciter les utilités des cinq actions. Nous avons retenu, pour le calcul analytique des utilités, les échelles discrètes suivantes :

$$[g_{1*}, g_1^*] = [30, 16, 2]$$

$$[g_{2*}, g_2^*] = [40, 30, 20, 10]$$

$$[g_{3*}, g_3^*] = [0, +, ++, +++]$$

d'où, par interpolation linéaire pour le critère g_1 , on trouve :

$$u[g(\text{RER})] = .07 u_1(16) + .93 u_1(2) + u_2(10) + u_3(+)$$

$$u[g(\text{METRO 1})] = .14 u_1(16) + .86 u_1(2) + u_2(20) + u_3(++)$$

$$u[g(\text{METRO 2})] = u_1(2) + u_2(20) + u_3(0) = u_1(2) + u_2(20)$$

$$\begin{aligned} u[g(\text{BUS})] &= .29 u_1(16) + .71 u_1(2) + u_2(40) + u_3(+)) \\ &= .29 u_1(16) + .71 u_1(2) \end{aligned}$$

$$u[g(\text{TAXI})] = u_1(30) + u_2(30) + u_3(+++) = u_2(30) + u_3(+++).$$

Ensuite, en remplaçant les $u_i(g_i^j)$ par les w_{ij} (cf. relations 10-11) :

$$u[g(\text{RER})] = w_{11} + .93 w_{12} + w_{21} + w_{22} + w_{23} + w_{31}$$

$$u[g(\text{METRO 1})] = w_{11} + .86 w_{12} + w_{21} + w_{22} + w_{31} + w_{32}$$

$$u[g(\text{METRO 2})] = w_{11} + w_{12} + w_{21} + w_{22}$$

$$u[g(\text{BUS})] = w_{11} + .71 w_{12}$$

$$u[g(\text{TAXI})] = w_{21} + w_{31} + w_{32} + w_{33}.$$

Finalement, en passant par l'étape 2 de comparaison des actions par paires (RER $\succ$ METRO 1 $\sim$ METRO 2 $\succ$ BUS $\succ$ TAXI), on arrive bien au programme linéaire du nouveau modèle UTA (δ est pris égal à .1) :

w_{11}	w_{12}	w_{21}	w_{22}	w_{23}	w_{31}	w_{32}	w_{33}	variables				σ^+	et	σ^-	Signe	Second membre	
0	.07	0	0	1	0	-1	0	1	-1	-1	1				$\geq$	.1	
0	-.14	0	0	0	1	1	0		1	-1	-1	1			=	0	
0	.29	1	1	0	0	0	0			1	-1	-1	1		$\geq$	.1	
1	.71	-1	0	0	-1	-1	-1				1	-1	-1	1	$\geq$	.1	
1	1	1	1	1	1	1	1								=	1	
									1	1	1	1	1	1	1	1	F

A partir de la solution optimale du programme linéaire, on trouve :

$u_1(2) = .5$, $u_1(16) = 0$, $u_1(30) = 0$, $u_2(10) = .2$, $u_2(20) = .1$, $u_2(30) = .1$, $u_2(40) = 0$, $u_3(0) = 0$, $u_3(+)=0$, $u_3(++)=0$, $u_3(+++)=.3$ et $F^* = \min F = 0$ mais cette solution n'est pas unique. L'ancienne version donne $u_1(2) = .56$, $u_1(16) = 0$, $u_1(30) = 0$, $u_2(10) = .14$, $u_2(20) = 0$, $u_2(30) = 0$, $u_2(40) = 0$, $u_3(0) = 0$, $u_3(+)=0$, $u_3(++)=.08$, $u_3(+++)=.3$ et $F^* = 0$.

La restitution numérique du préordre est, pour les deux modèles, comme suit : Ancienne version : RER $\rightarrow .6$, METRO 1ère $\rightarrow .56$, METRO 2e $\rightarrow .56$, BUS $\rightarrow .4$, TAXI $\rightarrow .3$; UTA amélioré : RER $\rightarrow .7$, METRO 1ère $\rightarrow .6$, METRO 2e $\rightarrow .6$, BUS $\rightarrow .5$, TAXI $\rightarrow .4$.

Une analyse post-optimale, au sens où elle a été définie par l'étape 4, donnerait un ensemble représentatif de solutions optimales dont on pourrait éventuellement retenir, comme solution unique, le centre de gravité (voir [3] pour une discussion sur ce point).

4. UNE EXPERIMENTATION ET CONCLUSION

Dans le but d'évaluer quantitativement les avantages offerts par la nouvelle version d'UTA vis-à-vis de l'ancienne, nous avons procédé à une

expérimentation des deux modèles sur des données simulées. En effet, nous avons considéré un ensemble de dix actions évaluées sur six critères (tableau 2) sur lequel nous avons généré des préordres aléatoires (de l'ordre d'une vingtaine). Les évaluations des actions ont été gardées inchangées et nous avons pris pour chaque essai $\delta = .05$. Les échelles des critères sont définies dans le sens : $4 \succ 3 \succ 2 \succ 1$.

Tableau 2 : Données de l'expérimentation

Actions		Critères				
1	1	4	2	3	4	4
2	3	3	3	2	3	2
3	1	4	2	2	4	4
4	2	3	1	3	2	3
5	3	2	2	1	2	2
6	2	4	2	2	4	3
7	4	1	3	2	1	2
8	3	4	2	2	4	3
9	4	1	3	1	2	1
10	4	1	2	1	1	1

Pour faciliter la comparaison, nous avons utilisé trois indicateurs : (1) le nombre d'itérations simplexe nécessaires pour parvenir à l'utilité optimale ; (2) le τ de Kendall entre le préordre initial et celui induit par l'utilité estimée et (3) le critère minimisé F^* (somme des σ) pris ici comme indicateur de dispersion des observations. Les résultats globaux apparaissent dans le tableau 3. Il en sort une nette supériorité en faveur de la nouvelle méthode, notamment pour ce qui concerne les indicateurs (1) et (3) ; celle-ci est respectivement de l'ordre de 48 % et 36 %. Au niveau du τ de Kendall, une amélioration de 10 % se dégage.

Afin d'évaluer l'ampleur de la différence de qualité entre les deux ajustements sur des problèmes de grande taille, nous avons testé la nouvelle version sur deux jeux de données provenant d'études concrètes sur

Tableau 3 : Résultats sommaires de l'expérimentation

Indicateur	UTA amélioré	UTA classique	Bilan
Nombre moyen d'itér. simplexe	11.1	21.5	baisse de 48 %
τ de Kendall moyen	.69	.63	croissance de 10 %
Indice de dispersion moyenne : $\bar{F}^*$	.34	.53	baisse de 36 %

lesquels nous disposons par ailleurs les résultats de l'ancienne version. Le premier jeu (pris dans [4], p. 198) avait 57 actions rangées dans sept classes et 6 critères qualitatifs ; l'ancienne analyse de régression avait fourni 128 itérations, τ de Kendall de .81 et indicateur de dispersion de 2.38 ; UTA amélioré obtient 129 itérations (+ 1 %), τ de Kendall de .86 (+ 5 % du fait du bornage du τ par l'unité) et indicateur de dispersion de 1.64 (- 31 %). Le schéma de régression pour les deux analyses est représenté dans la figure 2. Nous avons tracé, à la main, une courbe en

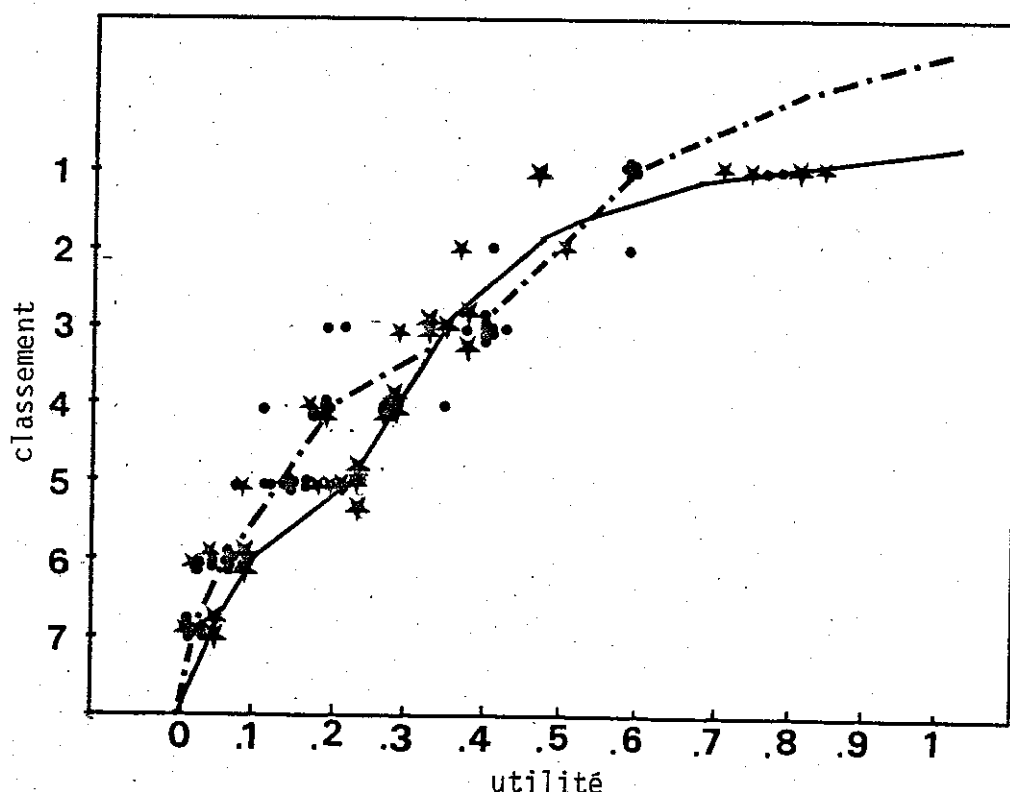


Figure 2 : Nouvelle version d'UTA versus l'ancienne sur le problème de gestion du personnel [4]. Les points et la courbe en pointillés se réfèrent à la nouvelle version.

continu pour l'ancienne version et une courbe en pointillés pour la nouvelle). Le deuxième jeu de données provient de [3], p. 127 et comporte 50 actions et 12 critères dont un seul quantitatif. Les améliorations apportées par la nouvelle version se récapitulent comme suit : nombre d'itérations : 135 vs 146 (- 8 %), τ : .49 vs .36 (+ 37 %) et dispersion : 1.11 vs 1.87 (- 41 %). Ces modifications s'interprètent par la déviation à droite de la courbe de régression dans la figure 3 (c'est pratiquement une droite). En règle générale, nous constatons un net resserrement des points autour des courbes obtenues par la nouvelle méthode.

UTA amélioré a été programmé en Basic et fonctionne en mode conversationnel sur micro-ordinateur ZENITH 100 alors que l'ancien UTA existe en FORTRAN IV sur système IBM 370/168.

Nos recherches actuelles sont orientées vers la création de systèmes interactifs d'aide à la décision dont le principe essentiel est de dégager des consensus sur des modélisations à partir d'un dialogue homme-machine. L'approche UTA, notamment sous sa nouvelle forme, semble être parfaitement adaptée à cette idée. Il a été constaté (cf. [3]) que la configuration "utilité versus décision extériorisée de la part des individus" peut devenir un support puissant d'interaction entre l'homme et le modèle d'homme.

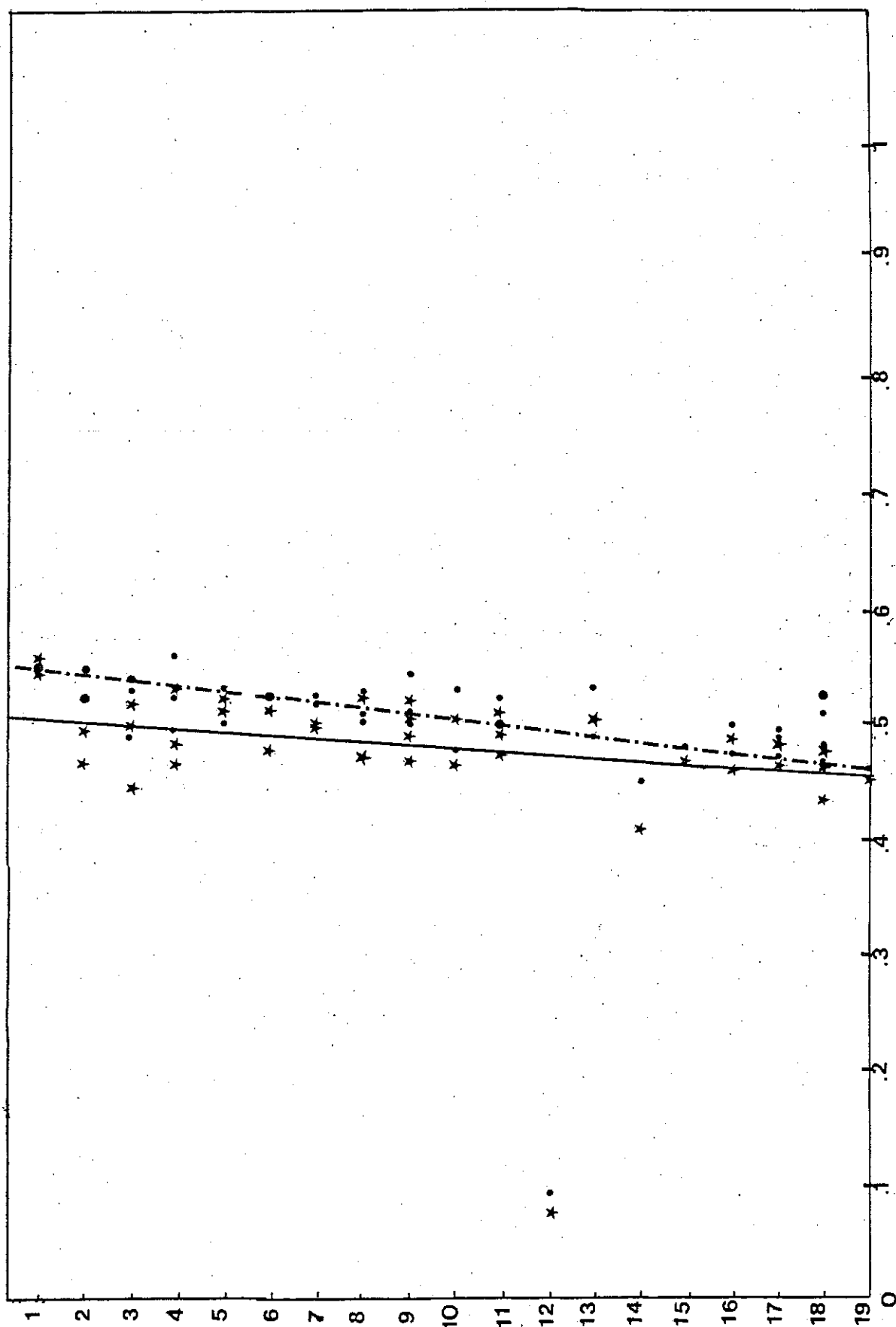


Figure 3 : Nouvelle version d'UTA versus l'ancienne sur le problème du choix de points de vente [3].

REFERENCES

- [1] A. Charnes et W.W. Cooper, Goal programming and multiple objective optimizations Part 1, European Journal of Operational Research, 1 (1977) 39-54.
- [2] E. Jacquet-Lagrèze et J. Siskos, Assessing a set of additive utility functions for multicriteria decision-making, the UTA method, European Journal of Operational Research 10 (1982) 151-164.
- [3] J. Siskos, Application de la méthode UTA I à un problème de sélection de points de vente mettant en jeu des critères multiples, RAIRO Recherche Opérationnelle 17 (1983) 121-136.
- [4] J. Siskos, La modélisation des préférences au moyen de fonctions d'utilité additives, Thèse 3e cycle, Université Pierre et Marie Curie, Paris (1979).