

Apports de la méthode « Best Worst » à l'analyse des critères de choix des consommateurs. Application au cas de l'huile d'olive en France et en Tunisie

Dekhili, Cohen, D'Hauteville, Sirieix

Résumé

Un objectif courant des études de marché, plus particulièrement lorsqu'il s'agit d'études interculturelles, consiste à comparer l'importance des attributs déterminants du choix entre sous-groupes de consommateurs. Plusieurs types de mesures peuvent être mises en œuvre : échelles, classement ou comparaisons par paires. Comment augmenter la comparabilité des données afin d'identifier les priorités des consommateurs ? Cet article rappelle les limites des méthodes classiques d'évaluation, et présente une application de la méthode Best-Worst au cas de l'huile d'olive, effectuée auprès de consommateurs de cultures différentes (Français et Tunisiens). Cette méthode permet de réduire certains biais, et de hiérarchiser facilement les attributs de choix d'un produit, mais n'est pas exempte de quelques critiques.

A contribution of the “Best Worst” method to the analysis of consumers choice criteria. An empirical application on olive oil consumers in Tunisia and France

Abstract

One frequent objective of market research, and particularly in cross cultural contexts, consists in comparing those attributes that determine the choice of different sub groups. Practitioners may resort to different types of importance scales, or use ranking methods, or even pair comparisons. How to increase the comparability between data in order to identify consumer's priorities ? This article shows the limits of traditional methods of evaluation that are largely discussed in the marketing literature, and presents an application of the Best-Worst scaling in the case of olive oil. The study was carried out in two different countries (France and Tunisia). The use of this method reduces the effect of respondent's culture on choice, and makes easy product's attributes ranking, but is not void from critics.

Sihem DEKHILI

- **Fonction** : Doctorante

- **Institution** : Montpellier Supagro, UMR MOISA

- **Adresse postale** : 2, Place Pierre Viala 34 060 Montpellier Cedex 1

- **Adresse E-mail** : dekhili@supagro.inra.fr

- **Centres d'intérêt** : Comportement du consommateur et méthodes de choix.

Eli COHEN

- **Fonction :** Professeur

- **Institution :** Guilford Glazer School of Business and Management
Department of Hotel and Tourism Management, Ben Gurion University
of The Negev

- **Adresse postale :** Ben Gurion University of The Negev

P.O.Box 653, Beer Sheva 84105

- **Adresse E-mail :** elico@bgu.ac.il

François D'HAUTEVILLE

- **Fonction :** Professeur
- **Institution :** Montpellier Supagro, UMR MOISA
- **Adresse E-mail :** hautevil@supagro.inra.fr
- **Centres d'intérêt :** Comportement du consommateur, qualité perçue.

Lucie SIRIEIX

- **Fonction** : Professeur

- **Institution** : Montpellier Supagro, UMR MOISA

- **Adresse E-mail** : sirieix@supagro.inra.fr

- **Centres d'intérêt** : Marketing agro-alimentaire, comportement du consommateur.

Introduction

La mesure des critères de choix des consommateurs a toujours constitué un champ d'intérêt important pour les chercheurs. Elle permet aux managers de bien comprendre ce que les consommateurs espèrent, désirent, aiment et n'aiment pas.

Dans le cas de collecte de données par enquêtes, la plupart des études utilisent des échelles d'évaluation de type Likert ou sémantiques, ou encore des méthodes de classement à modalités. Ces méthodes présentent cependant un certain nombre d'inconvénients. Par exemple, elles ne permettent pas toujours une discrimination nette entre les attributs. Ces derniers ne sont pas mesurés les uns relativement aux autres, ce qui rend difficiles les comparaisons entre les différentes mesures. Dans le cas d'études interculturelles, la comparaison des échelles de mesure peut poser problème (1).

Pour dépasser certaines de ces limites, nous nous intéressons dans cet article à la méthode Best-Worst, (ou encore méthode de la différence maximum, Maxdiff) qui permet des comparaisons intra et inter-sujets, et qui a été conseillée par plusieurs chercheurs dans le cas des études interculturelles. En outre, elle se distingue par la facilité de la tâche demandée aux répondants et une grande simplicité au niveau des analyses des données.

Pour illustrer cette méthode, nous proposons de l'appliquer à l'évaluation de l'importance de l'origine géographique dans le choix du produit pour des consommateurs français et tunisiens, dans le cas de l'huile d'olive.

Après un rappel des limites des méthodes classiques de mesure de l'importance des attributs, que la méthode Best-Worst permet de contourner, nous présentons l'application de cette méthode au cas proposé. Puis, les résultats seront synthétisés et discutés avant la présentation des conclusions de cette recherche et des limites de cette méthode.

Mérites comparés et limites des mesures d'importance des attributs

Les mesures d'importance des attributs à partir des déclarations d'individus repérées sur des échelles du type « pas important- très important » ont l'avantage d'être faciles à administrer. Elles sont familières aux répondants, et permettent de mettre en œuvre assez commodément le modèle classique d'attributs de Rosenberg (15) et Fishbein (8). En associant l'importance de l'attribut avec sa performance, on est supposé pouvoir prédire la préférence des consommateurs. Beaucoup d'auteurs ont souligné les limites de cette approche, et les critiques sont multiples. En l'absence de choix alternatif, les répondants ont tendance à favoriser les

réponses positives (tout est important), ou un regroupement vers le milieu de l'échelle, de sorte que les résultats sont parfois peu discriminants, ce qui est gênant lorsque l'objectif est de segmenter les marchés (12 ; 9). Ce type d'échelle peut également être soumis à un biais culturel, certains répondants pouvant avoir une tendance particulière à « sur ou sous-noter ». Enfin, d'autres auteurs ont mis en évidence le manque de validité prédictive des modèles de Rosenberg et Fishbein (18) construits à partir de cette combinaison importance/performance.

On reproche aussi à ces échelles des défauts plus conceptuels. On peut s'interroger par exemple sur l'hypothèse d'équidistance entre les échelons d'une échelle d'intervalle (2). Il est alors difficile d'interpréter l'ampleur des différences entre deux observations (une moyenne de 4 signifie-t-elle que l'importance accordée à un attribut est deux fois supérieure à une moyenne de 2 ?).

Les chercheurs et les praticiens ont donc depuis longtemps essayé de s'affranchir de ces limites et de chercher d'autres voies.

Une première voie consiste à libérer le répondant des limites imposées par la taille de l'échelle en lui proposant des notations libres ou « unbounded ratings » (11). Une autre approche pour lever la contrainte des intervalles de l'échelle consiste à proposer au répondant d'évaluer les attributs par référence à la note donnée au premier attribut, en multipliant ou divisant cette note selon l'importance des différences perçues entre attributs (« magnitude estimation ») (10). Cette approche séduisante fait cependant l'hypothèse que les répondants sont capables de multiplier ou de diviser avec la même facilité, ce qui n'est pas démontré.

Une autre voie consiste au contraire à placer les répondants en situation de comparaison ou de choix des attributs. La méthode des comparaisons par paires a été longtemps utilisée pour mesurer la dispersion discriminante des opinions (7), en plaçant les répondants dans une situation de choix. Toutefois, ces méthodes de classement par paires présentent la limite du nombre restreint des attributs à considérer, et restent réservées aux études de moins de dix stimuli (7). La méthode dite « best/worst », ou méthode de la différence maximum (maxdiff) permet de s'affranchir de cette limite (7). La méthode de choix discrets, basée sur la mesure de l'utilité, permet elle aussi de dépasser cette contrainte du nombre des items (7). En revanche, la mesure de l'utilité présente deux limites. En plus du coût élevé et de la nécessité de logiciels sophistiqués, cette méthode pose des difficultés d'interprétation des données,

particulièrement lorsqu'il s'agit de comparaisons d'utilités issues d'expérimentations différentes (9).

Enfin, une variante classique et souvent pratiquée de la méthode de choix consiste à faire classer les attributs par ordre d'importance. La méthode présente l'avantage d'une utilisation unique de chaque point de l'échelle, ce qui n'est pas le cas des échelles d'importance. Toutefois, cette méthode souffre des biais de l'effet d'ordre et n'est guère praticable dès lors que l'on dépasse un certain nombre d'attributs. Stephenson (16) a proposé une solution qui consiste à placer chaque item dans un nombre pré-déterminé de catégories, allant du plus important au moins important, mais en imposant un maximum d'items dans chaque catégorie (tri forcé ou « Qsort »).

Dans un article récent, Chrzan et Golovashkina (3) procèdent à une évaluation comparative de six méthodes de mesure de l'importance déclarée des attributs, au regard de quatre critères qui sont la durée d'administration, la capacité à discriminer les attributs, la capacité à segmenter les groupes d'individus, et la validité prédictive. Nous présentons une synthèse de ces évaluations dans le tableau n°1.

Tableau n°1 : Comparaison de six méthodes de mesure d'importance des attributs

	Modalités des tests	avantages	inconvénients	Durée d'administration	Discrimination entre attributs	Discrimination entre individus	Validité prédictive
Mesure directe des scores d'importance Rosenberg 1956, (Fishbein, 1967)	Scores sur une échelle en 5 points « important/pas important.	Simplicité d'administration	Peu discriminant Manque de validité prédictive	courte	moyenne	bonne	médiocre
Estimation comparative des grandeurs (magnitude estimation) (Lodge, 1981)	A partir du score attribué à un attribut pris au hasard, évaluer les autres par comparaison.	Pas de contrainte sur la taille à donner à l'importance	Sollicite des capacités des répondants à manier des quotients et des multiplications Forte sensibilité aux effets d'ordre.	2 fois plus long que la mesure directe	plus faible	bonne	bonne
Echelle libre (Marder, 1997)	Attribuer un score négatif ou positif sans limite formelle.	Pas de contrainte sur la taille à donner à l'importance		2 fois plus long que la mesure directe	plus faible	bonne	médiocre
Somme constante des scores	Répartir un nombre de points prédéterminé entre les attributs à mesurer	Oblige le répondant à arbitrer ses choix.	Difficile à utiliser si les attributs sont nombreux (>10)	2,3 fois plus long que la mesure directe	plus faible	bonne	bonne
Tri forcé (Q sort) (Stephenson, 1953)	Répartir les attributs sur une échelle de score dont on a défini a priori une répartition « normale »	Permet de classer un grand nombre d'attributs dans un nombre limité de catégories, oblige le répondant à arbitrer		2 fois plus long que la mesure directe	Plus forte	bonne	bonne
Différence maximum ou choix des extrêmes « best worst » (Finn & Louvière, 1992)	Parmi une collection d'attributs, choisir celui qui est le plus important et celui qui est le moins important	Permet de calculer l'utilité relative de chaque attribut pour chaque répondant.		4,5 fois plus long que la mesure directe	Plus forte	bonne	Bonne (la meilleure)

Source : A partir des résultats de Chrzan K., et Golovashkina N., 2006, An Empirical test of six stated importance measures, *International Journal of Market Research*, 48, n°6.

Dans la partie qui va suivre, nous allons appliquer la méthode Best-Worst au cas de notre étude sur l'huile d'olive.

Application : Une étude interculturelle sur les critères de choix de l'huile d'olive

Notre objectif est de déterminer si les notions d'origines (région, pays) sont des critères de choix importants chez les consommateurs, dans le cas de l'huile d'olive, en comparant deux populations différentes (Tunisie et France), tant sur le plan de la fréquence et des modes de d'achat de ce produit que sur la familiarité avec le concept d'appellation d'origine.

Un travail exploratoire déjà réalisé auprès de consommateurs français et tunisiens nous a permis de définir les critères de choix du produit (6). Nous avons retenu 13 attributs qui seront considérés dans la méthode Best-Worst (le pays d'origine, la région d'origine, le prix, la marque, la variété d'olives, le producteur/vendeur, le goût, la mention « vierge extra », la couleur, l'aspect, l'emballage, les signes officiels de qualité et les médailles, la mention « biologique »).

Un plan combinatoire en blocs incomplets équilibrés (14 ; 13) de type (13,13, 4, 4, 1) a été adopté afin de répartir les attributs en plusieurs groupes de choix qui seront présentés aux consommateurs. Ainsi les 13 attributs ont permis de construire 13 tableaux (blocs), comportant chacun 4 critères. Chaque attribut apparaît le même nombre de fois, soit 4 fois. Et chaque paire d'attributs apparaît le même nombre de fois, soit 1 fois.

Les consommateurs doivent choisir dans chaque tableau, qui représente l'une des treize tâches de choix, le critère qu'ils jugent le plus important dans leur choix d'huile d'olive, et le critère le moins important. Le tableau n°2 montre un exemple d'une tâche de choix.

Tableau n°2 : Exemple d'une tâche de choix

<i>Le moins important</i>	<i>Critère de choix</i>	<i>Le plus important</i>
<input type="checkbox"/>	1 <i>Le prix</i>	<input type="checkbox"/>
×	2 <i>Le producteur/vendeur</i>	<input type="checkbox"/>
<input type="checkbox"/>	3 <i>L'emballage</i>	<input type="checkbox"/>
<input type="checkbox"/>	4 <i>L'aspect</i>	×

Le niveau d'importance de chaque critère de choix est le résultat de la différence entre le nombre de fois où le critère a été choisi comme étant le plus important (Best) et le nombre de fois où il a été considéré comme étant le moins important (Worst).

Le niveau d'importance dépend du nombre des répondants et de la fréquence d'apparition de chaque attribut dans les ensembles de choix. Il est donc pertinent de considérer un score

standardisé afin de pouvoir comparer des groupes de répondants qui diffèrent en termes d'effectifs (7).

$$\text{Score standardisé} = \frac{\text{Niveau d'importance (Best-Worst)}}{4 n}$$

Où n : le nombre des répondants

4 : la fréquence d'apparition de chaque attribut dans les ensembles de choix

128 consommateurs tunisois et 123 parisiens ont été interrogés durant l'année 2006. En Tunisie, nous avons privilégié les acheteurs qui fréquentent les « nouveaux » lieux d'achats (supermarchés, magasins spécialisés...) qui offrent des produits conditionnés. Ces consommateurs ont généralement un niveau de revenu supérieur à la moyenne. Une deuxième condition de recrutement était d'avoir un bon niveau de français afin de s'assurer d'une bonne compréhension des questionnaires.

Ces exigences d'échantillonnage ont induit des différences socio-démographiques entre les répondants des deux nationalités (Voir tableau n°3).

Après dépouillement, 123 questionnaires français et 122 tunisiens se sont avérés utilisables pour l'analyse Best-Worst.

Tableau n°3 : Caractéristiques socio-démographiques des répondants par pays

		Tunisie	France
Nombre de répondants		122	123
Sexe	Homme	93	41
	Femmes	29	82
Age	18 à 44 ans	67	49
	45 et plus	55	74
Niveau d'éducation	Secondaire	27	56
	Universitaire	95	67
Revenu	Faible et moyen*	59	85
	Haut*	63	38

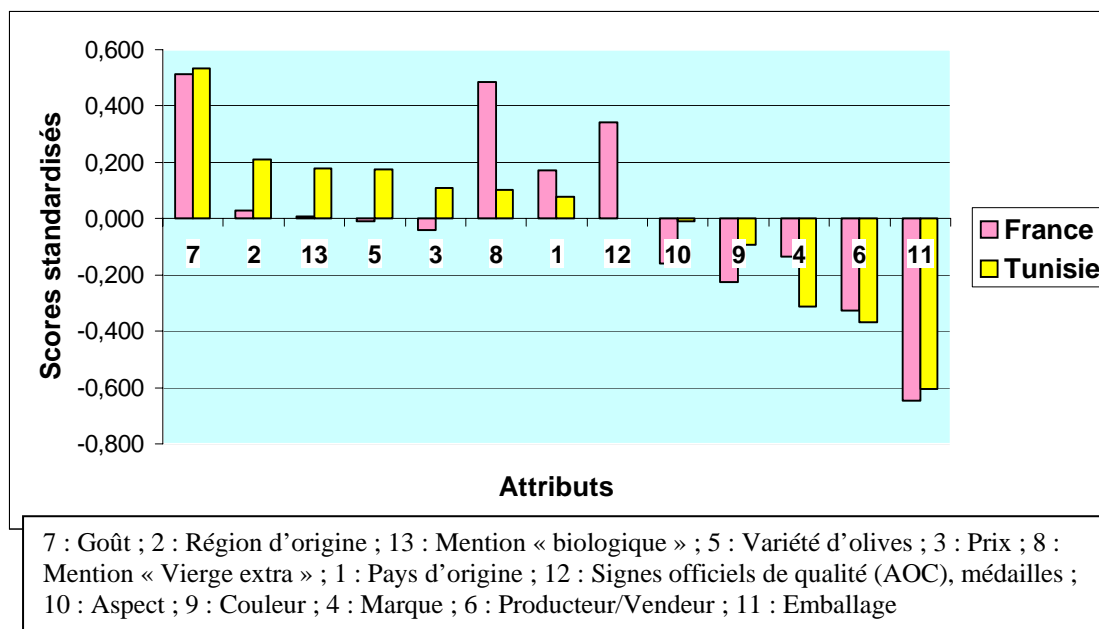
* Faible et moyen revenu : <1000Dinars Tunisiens/3693 Euros

* Revenu élevé: >=1000Dinars Tunisiens/3693 Euros

Résultats

Les résultats de l'analyse des données sont synthétisés dans la figure n°1.

Figure n°1 : Critères de choix de l'huile d'olive en France et en Tunisie



Les résultats montrent que le goût est le critère déclaré le plus important dans le choix de l'huile d'olive, aussi bien chez les Tunisiens que les Français. C'est une réponse relativement attendue, s'agissant d'un produit alimentaire à forte vocation gustative.

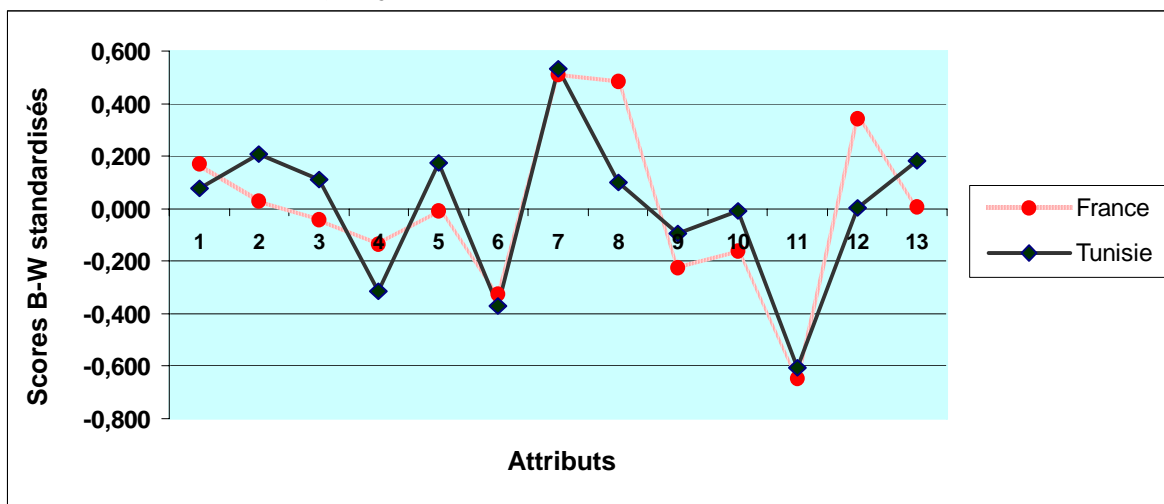
La figure 1 montre que l'origine géographique de la production, opérationnalisée par le pays, la région et la variété d'olives, est classée parmi les critères les plus importants dans les deux pays, à l'exception de la variété d'olives qui présente un score neutre chez les consommateurs français (score standardisé: -0,01) alors que ce score est positif chez les consommateurs tunisiens. Le critère « pays d'origine » est plus important que le critère « région » dans le cas français. En revanche, les Tunisiens accordent plus d'importance à la région d'origine et à la variété d'olives qu'au pays. Ce résultat peut-être expliqué en partie par la nature locale de l'offre d'huile d'olive en Tunisie. Les consommateurs tunisiens ne sont pas habitués à consommer des huiles étrangères.

Les deux attributs Producteur/Vendeur et Emballage sont les moins importants dans le choix de l'huile d'olive des deux échantillons, avec des scores très comparables.

Pour valider les différences observées entre Français et Tunisiens, nous avons effectué une analyse de profils. Il s'agit d'une application particulière de l'analyse de variance (MANOVA et ANOVA) qui vise à comparer des profils de réponses entre des groupes (17). Elle s'effectue à l'aide de trois tests (test de différence de niveau des profils, test d'horizontalité des profils, test de parallélisme des profils).

Pour conduire l'analyse de profils, nous avons considéré les 13 attributs de choix et les deux groupes de consommateurs (Tunisiens et Français) (Voir figure n°2).

Figure n°2 : Les profils des attributs qui déterminent le choix de l'huile d'olive chez les consommateurs tunisiens et français



Le test de parallélisme constitue le test principal de l'analyse de profils. Il permet de déterminer si les données provenant des deux groupes de participants présentent la même évolution. Autrement dit, il s'agit de vérifier si les profils ont la même « forme » pour chaque groupe.

Dans le cas de notre étude, ce test s'est avéré significatif (F de l'interaction attribut*pays = 8,135 ; $p=0.000$), suggérant que les profils ne sont pas parallèles. Le profil des attributs pour les Français (dispersion=0.951) montre une plus grande dispersion que dans le cas tunisien (dispersion=0.855). Ce résultat permet de déduire que les critères de choix de l'huile d'olive sont plus contrastés chez les Français que les Tunisiens.

Le test d'horizontalité des profils permet de vérifier si le niveau moyen d'évaluation, pour un groupe donné, varie d'une mesure à l'autre. Il s'agit de tester l'effet du facteur « mesure ». Ce test porte sur des valeurs de différences entre des paires de mesures, et est généralement conduit dans le cas de non parallélisme entre les profils des deux groupes.

Appliqué à nos données, ce test s'est avéré significatif (F du facteur « attribut » = 103,410; $p=0.000$), indiquant un effet « attribut », au sein du même groupe. C'est-à-dire que la différence entre attributs n'est pas la même pour toutes les paires d'attributs. Comme exemple, nous pouvons souligner que la différence entre l'évaluation du goût et celle de la région d'origine n'est pas la même que la différence entre l'évaluation de la région d'origine et celle de la mention « biologique ».

Enfin, le test de différence de niveau vérifie l'hypothèse d'une différence entre les profils des groupes, reflétée par une comparaison de la moyenne globale de toutes les mesures, entre les groupes considérés.

Ce test est non significatif dans le cas de notre étude (F du facteur « pays » = 0.000, $p=1.000$), soulignant que la moyenne globale de l'ensemble des attributs dans le cas des consommateurs français est équivalente à celle des Tunisiens.

Pour déterminer si les différences observées dans la figure 1 entre les deux populations en terme d'évaluation des différents attributs sont significatives ou non, un modèle linéaire général a été conduit. Les résultats sont synthétisés dans le tableau n°4.

Tableau n°4: Différences d'évaluation des attributs de choix de l'huile d'olive entre les Français et les Tunisiens.

Attributs	Significativité
1 : Pays d'origine	n.s
2 : Région d'origine	**
3 : Prix	*
4 : Marque	**
5 : Variété d'olive	***
6 : Producteur/Vendeur	n.s
7 : Goût	n.s
8 : Mention « Vierge extra »	***
9 : Couleur	**
10 : Aspect	**
11 : Emballage	n.s
12 : Signes officiels de qualité (AOC), médailles	***
13 : Mention « biologique »	*

* : $p<0.05$; ** : $p<0.01$; *** : $p<0.001$

Concernant les attributs reliés à l'origine, les résultats montrent des différences significatives entre les deux populations dans l'évaluation de la région d'origine et de la variété d'olives (les Tunisiens valorisent mieux les deux attributs), mais pas en ce qui concerne le pays d'origine. Pour le reste des attributs, des différences significatives entre les deux groupes ont été

constatées dans le cas du prix, mention « biologique », couleur, aspect, marque, mention « vierge extra », et signes officiels de qualité (AOC), médailles. Les Tunisiens accordent plus d'attention aux quatre premiers attributs, alors que les Français considèrent plus les trois derniers dans leur choix.

Nous allons maintenant nous intéresser à la validation statistique de la hiérarchie des attributs au sein de chaque population. Pour ce faire, des tests d'analyse de variance (ANOVA), complétés par le test Tukey ont été utilisés.

Tableau n°5 : Hiérarchisation des critères de choix de l'huile d'olive chez les Tunisiens

	Attributs	Attributs équivalents $p < 0.05$ (Tukey HSD)						
7	Goût	×						
2	Région d'origine		×					
13	Mention « biologique »		×	×				
5	Variété d'olives		×	×	×			
8	Mention « Vierge extra »		×	×	×			
3	Prix		×	×	×			
1	Pays d'origine		×	×	×	×		
12	Signes officiels de qualité (AOC), médailles			×	×	×		
10	Aspect				×	×		
9	Couleur					×		
4	Marque						×	
6	Producteur/Vendeur						×	
11	Emballage							×

Tableau n°6 : Hiérarchisation des critères de choix de l'huile d'olive chez les Français

	Attributs	Attributs équivalents $p < 0.05$ (Tukey HSD)						
7	Goût	×						
8	Mention « Vierge extra »	×						
12	Signes officiels de qualité (AOC), médailles	×	×					
1	Pays d'origine		×	×				
2	Région d'origine			×	×			
13	Mention « biologique »			×	×	×		
5	Variété d'olives			×	×	×		
3	Prix				×	×	×	
4	Marque				×	×	×	
10	Aspect					×	×	×
9	Couleur						×	×
6	Producteur/Vendeur							×
11	Emballage							×

Les tableaux 5 et 6 mettent en évidence que la hiérarchisation des attributs diffère selon le pays. Sept groupes d'attributs dans le cas tunisien, dans lesquelles les moyennes ne sont pas significativement différentes l'une de l'autre, et huit dans le cas français ont été recensés. Nous pouvons conclure à une meilleure discrimination des attributs chez les consommateurs français.

Si nous reprenons les trois attributs d'origine qui nous intéressent (pays, région, variété d'olives), nous remarquons qu'ils sont classés dans le même groupe quelle que soit la nationalité des répondants. La différence entre ces trois attributs n'est pas significative pour les répondants des deux nationalités.

A ce stade, nous posons la question de savoir si des critères socio-démographiques comme le sexe, l'âge, le niveau d'éducation et le revenu interviennent dans la hiérarchie des attributs.

Pour répondre à cette question, nous avons adopté un modèle linéaire général au sein de chaque population, en considérant les segments suivants ; hommes/femmes, les plus âgés/les plus jeunes ; les consommateurs ayant un niveau universitaire d'éducation/ceux ayant un niveau secondaire, les plus aisés/les moins aisés (Voir tableau n°3).

Seules les différences significatives dans l'évaluation des attributs entre segments (quel que soit le critère de segmentation) ont été reportées dans le tableau n°7.

Tableau n°7 : Différences significatives entre segments (sexe, tranche d'âge, niveaux d'éducation, et niveau de revenu) en terme d'évaluation des attributs, au sein de chaque population.

Segments	Attributs	France	Tunisie
		Significativité	Significativité
Sexe	Goût	*	
Age	Marque	*	
	Variété d'olives	*	*
	Goût	**	
	Mention « vierge extra »	***	
	Couleur	*	
	Aspect	*	
Niveau d'éducation	Prix	*	
	Mention « vierge extra »	*	

* : $p < 0.05$; ** : $p < 0.01$; *** : $p < 0.001$

Parmi les quatre critères socio-démographiques considérés dans la segmentation, le revenu semble ne pas avoir d'effet sur l'évaluation de l'importance des attributs de choix de l'huile d'olive aussi bien chez les Français que les Tunisiens.

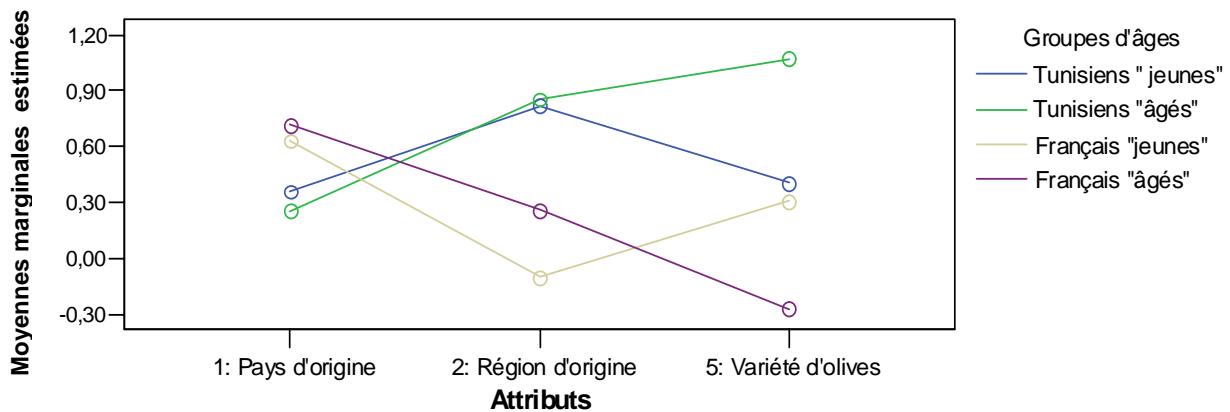
Les résultats montrent plus d'hétérogénéité au sein de la population des consommateurs français. Ainsi, plusieurs différences ont été recensés sur la base des différences d'âge, niveau

d'éducation, et sexe. En revanche, dans le cas tunisien, seul l'âge semble influencer l'évaluation de la variété d'olives.

L'âge semble constituer le critère de segmentation le plus pertinent. En effet, la majorité des différences constatées dans le cas français (marque, variété d'olives, goût, mention « vierge extra », couleur et aspect) est due aux différences d'âges entre les consommateurs.

Pour approfondir l'étude des différences entre segments au sein de chaque population, nous avons conduit une analyse de profils sur les variables liées à l'origine géographique (pays d'origine, région d'origine, variété d'olives). Seule la segmentation sur l'âge des répondants a été considérée. Les résultats sont reportés dans la figure n°4.

Figure n°4 : Les profils des attributs sur l'origine dans les deux pays selon l'âge des consommateurs



L'analyse des résultats sur la base de l'âge des répondants, au sein de chaque population, montre une ressemblance entre le cas tunisien et le cas français. Le test de parallélisme est non significatif dans les deux cas (F de l'interaction attribut*pays=1,814 et 2,937 respectivement ; $p=0.167$ et 0.057 respectivement).

Vue la non significativité du test de parallélisme, il serait peu pertinent de conduire le test d'horizontalité. D'autres informations (Voir tableau n°8) peuvent être déduites de cette analyse.

Tableau n°8 : Informations issues de l'analyse des profils des attributs sur l'origine géographique selon l'âge dans les deux pays (Tunisie, France)

Mesures (En scores BW)	Jeunes		Agés	
	Tunisie	France	Tunisie	France
Niveau du profil	0.527	0.279	0.727	0.234
Dispersion du profil	0.196	0.254	0.315	0.336
Forme du profil :				
Point le plus élevé	Région (0.821)	Pays (0.633)	Variété (1.073)	Pays (0.716)
Point le plus bas	Pays (0.358)	Région (-0.102)	Pays (0.255)	Variété (-0.270)

Nous constatons que parmi les attributs faisant apparaître des différences significatives entre segments, la variété d'olives montre le plus de divergence. Ainsi, dans les deux pays, des différences en terme de l'importance accordée à cet attribut selon l'âge ont été observées, avec deux tendances opposées puisqu'en France ce sont les jeunes qui valorisent le plus cet attribut, alors qu'en Tunisie ce sont les plus âgés qui prennent plus en considération la variété d'olives dans le choix du produit.

Discussion : apports et limites de la méthode Best-Worst

Deux avantages principaux de la méthode Best-Worst nous paraissent importants à souligner à l'issue de cette recherche : la capacité à discriminer les attributs d'une part, l'absence de biais liés aux différences individuelles en matière d'évaluation des attributs d'autre part.

La méthode Best-Worst a permis de hiérarchiser les attributs de choix de l'huile d'olive, selon leur importance relative déclarée par les répondants français et tunisiens. La discrimination entre attributs est assez nette et le caractère significatif ou non des différences entre attributs a pu être mis en évidence à l'aide de tests statistiques simples (Tukey).

Par ailleurs, les personnes n'ayant pas à donner de notes aux attributs mais seulement à les choisir, il n'y a pas de possible sur-évaluation des notes par un groupe de répondants (groupe culturel ou groupe d'âge). La méthode Best-Worst a donc permis des comparaisons inter-groupes, à la fois entre deux populations de cultures différentes (France et Tunisie), et sur la base des caractéristiques démographiques, au sein d'une même population.

Si la méthode Best-Worst présente de nombreux avantages qui justifient son utilisation, elle ne résout cependant pas un certain nombre de problèmes inhérents à l'ensemble des méthodes d'analyse de données déclaratives centrées sur les critères de choix des produits. Par ailleurs, la méthode Best-Worst souffre de limites qui lui sont propres.

Comme toute méthode d'analyse de données déclaratives, la méthode Best-Worst ne traite par nature que de l'importance déclarée, et non de l'importance réelle des critères proposés. Par exemple, il est peu réaliste de croire que la mention du producteur/vendeur ou l'emballage n'influencent pas le choix des consommateurs, ou encore de conclure que le goût est un critère plus important que l'origine du produit. Il est par ailleurs très « rationnel » pour un consommateur de produit alimentaire de désigner le « goût » comme attribut le plus important, et de ne pas se déclarer influencé par des indicateurs comme l'emballage ou la publicité.

La méthode partage également avec toutes les autres méthodes d'analyse de données déclaratives la difficulté du choix d'un nombre restreint d'items, et dans le cas d'études inter-culturelles, du choix d'items pertinents dans des cultures différentes et de la traduction du libellé des items.

La méthode Best-Worst souffre également de limites qui lui sont propres, concernant le travail du répondant et la sélection proposée des critères.

Une limite soulignée dans plusieurs études (11) concerne le temps nécessaire aux répondants pour compléter le questionnaire. Il ne nous semble pas que ce temps nécessaire soit, dans l'absolu, une limite ; Dans le cas de notre étude, les répondants ne semblent pas avoir été gênés ; il n'y a pratiquement aucun questionnaire non rempli. Cependant ce temps passé peut constituer une limite très forte s'il correspond à une difficulté perçue de la tâche trop importante, risque d'engendrer de la lassitude face au caractère répétitif des questions posées, voire une réaction de rejet de la part de répondants qui se demandent pourquoi la même question leur est posée plusieurs fois avec une sélection tournante parmi les mêmes attributs.

Par ailleurs, la méthode Best-Worst pose le problème d'un choix forcé sur deux items (le plus et le moins important) dans le cas où le répondant ne perçoit aucune différence d'importance entre les critères proposés, ou entre plusieurs critères jugés importants, ou encore entre plusieurs critères jugés peu importants.

Conclusions

Récemment, quelques chercheurs (9 ; 4 ; 5) ont exploré les potentialités de la méthode Best-Worst, qui permet de palier certains inconvénients de mesures d'attributs perçus, qu'il s'agisse d'échelles de likert, d'échelles sémantiques ou de classements. De fait, cette méthode présente des avantages dans le contexte d'études comparatives entre populations, ou d'études multiculturelles. S'appuyant sur une démarche dans laquelle le consommateur est invité à faire des choix, la méthode Best-Worst s'avère plus discriminante que les méthodes déclaratives classiques. En faisant des choix, le répondant est en effet peut-être plus impliqué dans sa réponse. Cette implication augmente la durée du questionnaire, sans toutefois présenter des difficultés pour le répondant.

Nous avons utilisé cette méthode pour déterminer si les notions d'origines (pays d'origine, région d'origine, variété d'olives) constituent des critères importants du choix de l'huile d'olive chez les consommateurs en France et en Tunisie.

Avec beaucoup de facilité, la méthode Best-Worst nous a permis de hiérarchiser les attributs de préférence de l'huile d'olive, et de comparer les résultats entre groupes différents. Les résultats issus de cette méthode sont par ailleurs cohérents avec des résultats antérieurs, obtenus par des méthodes qualitatives.

Compte tenu de la facilité de mise en œuvre, de sa capacité à hiérarchiser les attributs de choix des consommateurs, la méthode Best-Worst mérite donc d'être prise en compte dans les études consommateurs par lesquelles les chercheurs et décideurs cherchent à comprendre les arbitrages que font les consommateurs entre les attributs d'un produit, et à segmenter les marchés en fonction des avantages recherchés.

Références bibliographiques

- (1) Bartikowski B. Chandon J-L. et Gierl H. (2006), Calibration internationale des échelles sémantiques. *Décisions Marketing*, 43-44, juillet-décembre, 207-219.
- (2) Chandon J.-L. et Bartikowski B. (2004), Une échelle ordinale permettant de classer les répondants en "satisfait", "indifférent" et "insatisfait". *Recherche et Applications en Marketing*, 19, 39-51.
- (3) Chrzan K. et Golovashkina N. (2006), An empirical test of six stated importance measures, *International Journal of Market Research*, 48, n°6.
- (4) Cohen S. (2003), Maximum Difference Scaling: Improved Measures of Importance and Preference for Segmentation. *Sawtooth Software Conference Proceedings*, Sequim (WA).
- (5) Cohen S. et Neira L. (2003), Measuring preference for product benefits across countries. Overcoming scale usage bias with Maximum Difference Scaling. *ESOMAR. Latin America Conference Proceedings*, Amsterdam (Netherlands).
- (6) Dekhili S. (2005), Product origin as a quality cue and image as perceived by Tunisian and French consumers: the case of olive oil. *The European/EAAE PhD Workshop*, Wageningen University.
- (7) Finn A. et Louviere J.-J. (1992), Determining the Appropriate Response to Evidence of Public Concern: The Case of Food Safety, *Journal of Public Policy and Marketing*, 11, 12-25.
- (8) Fishbein M. (1967), A behaviour theory approach to the relations between beliefs about an object and the attitude towards the object. In : M. Fishbein (ed) *Readings in attitude theory and measurement*, New York : Wiley, pp. 389-399.
- (9) Goodman S. Lockshin L. et Cohen E. (2005), Best-Worst Scaling: A simple Method to Determine Drinks and Wine Style Preferences, *2nd International Wine Marketing and Business Conference*, Sonoma (USA).
- (10) Lodge M. (1981), Magnitude scaling: quantitative measurement of opinions. Sage University paper on quantitative applications in social sciences, 07-025, Beverly Hills, Sage Publications.
- (11) Marder E. (1997), *The laws of Choice : predicting consumer behavior*, N.Y., the Free Press
- (12) Myers J.H. (1996), *Segmentation and positioning for strategic marketing decisions*. Chicago, American Marketing Association.

- (13) Raghavarao D. et Federer W. T. (2003), Sufficient conditions for balanced incomplete block designs to be minimal fractional combinatorial treatment designs. *Biometrika* vol.90, n°2: p. 465-470.
- (14) Rao P. V. (1960), The dual of balanced incomplete block design, *The Annals of Mathematical Statistics*: p. 779-785.
- (15) Rosenberg M.J. (1956), Cognitive structure and attitudinal affect, *Journal of Abnormal and Social Psychology*, 53, 367-372.
- (16) Stephenson W. (1953), *The study of behavior : the Q-Technique and its methodology*. Chicago, University of Chicago Press
- (17) Tabachnik B. G. et Fidell L. S. (2001), *Using multivariate statistics*, MA, Allyn and Bacon.
- (18) Wilkie W.L. et Pessemier E.A. (1973), Issues in marketing's use of multi attribute attitude models. *Journal of Marketing research*, 10, pp. 428-441