

---

## Importance des critères de choix de l'huile d'olive. La méthode Best Worst

*How important are choice criteria for olive oil? Applying the Best-Worst method to a study in France and Tunisia*

**Sihem Dekhili et Lucie Sirieix**

---



### Édition électronique

URL : <http://journals.openedition.org/economierurale/3727>

DOI : 10.4000/economierurale.3727

ISSN : 2105-2581

### Éditeur

Société Française d'Économie Rurale (SFER)

### Édition imprimée

Date de publication : 30 novembre 2012

Pagination : 119-131

ISSN : 0013-0559

### Référence électronique

Sihem Dekhili et Lucie Sirieix, « Importance des critères de choix de l'huile d'olive. La méthode Best Worst », *Économie rurale* [En ligne], 332 | novembre-décembre 2012, mis en ligne le 30 novembre 2014, consulté le 15 novembre 2019. URL : <http://journals.openedition.org/economierurale/3727> ; DOI : 10.4000/economierurale.3727

---

# Importance des critères de choix de l'huile d'olive

## La méthode Best Worst

**Sihem DEKHILI** • *Humans and Management in Society, HuManiS (EA 1347), École de Management de Strasbourg - Université de Strasbourg*  
sihem.dekhili@em-strasbourg.eu

**Lucie SIRIEIX** • *UMR 1110 MOISA, Montpellier*  
sirieix@supagro.inra.fr

La mesure des critères de choix des consommateurs a toujours constitué un champ d'intérêt important pour les chercheurs et les praticiens, notamment en marketing international. Elle permet aux décideurs de bien comprendre ce que les consommateurs espèrent, désirent, aiment et n'aiment pas, et d'adapter en conséquence leur offre de produits selon les marchés.

Dans la plupart des études qui portent sur les critères de choix, ou qui cherchent à hiérarchiser les attributs de produits qui sont pris en compte par les consommateurs, la collecte des données met en œuvre des échelles d'évaluation de type Likert ou des échelles sémantiques bipolaires, 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. Lorsqu'il s'agit par exemple de comparer les priorités de choix entre des groupes appariés, comme dans le cas des études interculturelles, la comparaison des échelles de mesure peut poser problème (Bartikowski *et al.*, 2006).

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) développée notamment par Louvière et Woodworth (1990).

À titre d'illustration, nous présentons une étude cherchant à déterminer quelle est l'im-

portance relative de différents critères de choix d'une huile d'olive par des consommateurs Français et Tunisiens. Pour les opérateurs de la filière oléicole dans ces deux pays, cette approche permettra de juger notamment de la pertinence de la mise en valeur de l'origine des produits comme critère de qualité ; cette stratégie, qui a fait ses preuves dans d'autres secteurs comme le vin ou les fromages en particulier en France, reste peu familière en Tunisie.

Après avoir comparé les méthodes classiques d'évaluation et rappelé leurs limites, nous présentons la méthode *Best-Worst* et ses apports en l'illustrant par le cas de notre étude. Enfin, l'article évalue les résultats obtenus, les limites et les conditions d'utilisation de cette méthode.

### Comparaisons entre différentes mesures d'importance des critères de choix

Les méthodes d'enquêtes que nous comparons s'appuient explicitement ou non sur le modèle classique d'attributs de Rosenberg (1956) et Fishbein (1967). Il s'agit en général de mesurer l'importance des attributs qui influencent les choix à partir des déclarations d'individus repérées sur des échelles du type « pas important-très important » ou « d'accord-pas d'accord ». Ces échelles psychométriques ont l'avantage d'être faciles à administrer et sont devenues familières aux répondants. En associant l'importance de l'attribut avec sa performance, le chercheur est supposé pouvoir prédire les choix des consommateurs. Beaucoup d'auteurs ont souligné les limites de cette

approche, et les critiques portent à la fois sur la qualité des données collectées mais aussi sur les capacités analytiques de ces échelles (Chandon et Bartikowski, 2004). Par exemple, 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 (Goodman *et al.*, 2005). Ce type d'échelle peut également être soumis à un biais culturel, certains répondants pouvant avoir une tendance particulière à « sur-noter » ou « sous-noter ». Enfin, d'autres auteurs ont mis en évidence le manque de validité prédictive des modèles de Rosenberg et Fishbein (Wilkie et Pessemier, 1973) construits à partir de cette combinaison importance/performance.

Ces échelles souffrent aussi de défauts plus conceptuels. L'on peut s'interroger par exemple sur l'hypothèse d'équidistance entre les échelons d'une échelle d'intervalles (Bartikowski *et al.*, *op. cit.*). Il est alors difficile d'interpréter l'ampleur des différences entre deux observations (par exemple, 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 essayé de s'affranchir des limites de cette échelle largement popularisée à partir des travaux de Rensis Likert et cherché d'autres voies et méthodes plus performantes. Nous allons d'abord examiner brièvement ces différentes méthodes selon le double point de vue de la collecte des données et des possibilités d'analyses, avant de présenter celle qui nous intéresse particulièrement ici, l'analyse « Best-Worst ». On peut classer les méthodes en deux groupes : dans un premier groupe, celles qui cherchent à obtenir un score indépendant pour chacun des attributs, l'objectif étant que les mesures obtenues s'affranchissent des biais liés aux caractéristiques de chaque individu, tout en conservant des capacités discriminantes.

Ces méthodes reposent finalement sur la notation. Un deuxième groupe rassemble des méthodes qui font plutôt appel à des tâches de tri, de classement ou de choix.

### 1. Méthodes utilisant la notation

Une première voie consiste à libérer le répondant des limites imposées par la taille de l'échelle de notation en lui proposant des notations libres ou « *unbounded ratings* » (Marder, 1997). Le répondant est ainsi libre d'attribuer des scores négatifs et de calibrer ses scores selon sa propre échelle de référence. Une manière courante de procéder est de proposer aux répondants un segment non gradué et borné par deux expressions opposées (pas d'accord du tout – tout à fait d'accord). Pour certains auteurs (Chandon et Bartikowski, *op. cit.*), le temps d'administration des questionnaires est deux fois plus long que pour l'échelle de Likert en 5 ou 7 points, surtout lorsque les répondants sont invités à mettre des évaluations négatives. Paradoxalement, et de l'aveu même de son auteur (Marder, *op. cit.*) la capacité de discrimination entre attributs et la validité prédictive de la méthode ne sont pas significativement meilleures. Elle impose évidemment une démarche d'harmonisation des échelles.

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 », Lodge, 1981). L'hypothèse sous-jacente est que l'individu procède de manière comparative et hiérarchique par rapport à ce qu'il connaît déjà. Cette approche séduisante fait cependant aussi l'hypothèse que les répondants sont capables de multiplier ou de diviser avec la même facilité, ce qui n'est pas démontré. Elle est également sensible aux effets d'ordre de présentation des attributs et à leur nombre, ce qui nécessite la mise en œuvre d'un plan

expérimental équilibré. Le résultat est que la durée d'administration serait deux fois plus élevée que celle de l'échelle classique de Likert, mais en revanche, elle témoignerait d'une meilleure validité prédictive et de bonnes capacités de discrimination entre individus (Chandon et Bartikowski, *op. cit.*).

## 2. Méthodes s'appuyant sur un choix

Avec ces méthodes, il s'agit de 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 des opinions en plaçant les répondants dans une situation de choix (Finn et Louviere, 1992). Toutefois, ces auteurs notent que ce type de méthode de classement par paires ne permet de prendre en compte qu'un nombre restreint d'attributs à considérer, et reste réservé aux études de moins de dix stimuli.

La méthode de choix discrets, fondée sur la mesure de l'utilité, permet de dépasser cette contrainte du nombre d'items (Finn et Louviere, *op. cit.*). Elle consiste à proposer à des répondants des tâches de choix successives selon un plan d'expérience mettant en concurrence les modalités des attributs faisant l'objet de l'étude. Ces choix révèlent les préférences relatives d'un individu pour certains attributs. La méthode des choix discrets présente des qualités supérieures en termes de discrimination entre attributs et en termes de validité prédictive. Bien menée, elle permet de calculer des parts de marché théoriques. Mais elle présente deux limites importantes. La première est celle du coût élevé de sa mise en œuvre et de la nécessité de disposer de logiciels sophistiqués. La seconde est que cette méthode pose des difficultés d'interprétation des données, particulièrement lorsqu'il s'agit de comparaisons d'utilités issues d'expérimentations différentes (Goodman *et al.*, 2005). Enfin, si le nombre des attributs et de leurs modalités est élevé elle nécessite la mise en œuvre d'échantillons importants.

Enfin, une variante classique et souvent utilisée 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 utilisable dès lors que l'on dépasse un certain nombre d'attributs. Stephenson (1953) a proposé une amélioration qui consiste à placer chaque item dans un nombre prédéterminé de catégories, allant du plus important au moins important, mais en imposant un maximum d'items dans chaque catégorie (tri forcé ou « Qsort »). À notre connaissance, cette piste n'a pas fait l'objet de travaux publiés.

## 3. Présentation de la méthode Best-Worst

Cette méthode, appelée aussi « Max Diff », fait partie des méthodes dans lesquelles les répondants sont appelés à faire un choix. Le principe est simple : après avoir établi une liste d'attributs considérés comme pertinents par rapport à une situation d'achat ou de consommation d'un produit, il s'agit de présenter à des répondants des listes restreintes d'attributs parmi lesquels le répondant doit choisir celui qui est le plus important pour lui, et celui qui est le moins important. Ces listes sont constituées et présentées selon un plan d'expérience combinatoire en blocs incomplet équilibrés (BIE) (Rao, 1960) permettant de s'assurer que chaque attribut est comparé le même nombre de fois à chacun des autres attributs. Le BEI est une extension du carré latin, dans lequel les  $n$  modalités à analyser sont disposées en  $n$  rangs et  $n$  colonnes, de manière à ce que chaque rangée et chaque colonne présentent chaque modalité dans une position différente. Chaque rangée peut être considérée comme un ensemble de choix (un « bloc »).

La première étape consiste à décider du nombre des critères de choix ou attributs qui seront comparés entre eux, puis à

élaborer le plan incomplet en fonction du nombre de ces attributs et du nombre de modalités à faire figurer dans chaque tâche de choix. En pratique, il convient d'arbitrer entre le nombre des modalités et la liste des choix présentés au répondant. L'expérience suggère qu'il ne faut pas dépasser 20 tâches de choix, sous peine de démotiver les répondants, et qu'il faut pour chaque tâche limiter le nombre des modalités à 4 ou 6, ce dernier chiffre étant un maximum à ne pas dépasser, surtout si le libellé des modalités est long ou complexe (Remaud et Lockshin, 2009).

Par exemple si l'on a 13 attributs, on peut adopter un BEI de type (13, 4, 4, 1) : les 13 attributs (critères de choix supposés) permettent de construire 13 tableaux (blocs), comportant chacun 4 attributs. Chacun d'eux apparaît le même nombre de fois (4 fois). Et chaque paire d'attributs apparaît le même nombre de fois (1 fois).

Cohen (2003) a comparé les résultats issus de la méthode *Best-Worst* avec ceux obtenus avec des échelles de type Likert, d'une part, et avec des comparaisons par paires, d'autre part. En termes de durée d'administration, il constate qu'en moyenne les évaluations par les méthodes de compa-

raisons par paires et *Best-Worst* sont trois fois plus longues à accomplir que les évaluations par les échelles de type Likert. Il explique cette différence de durée par une plus grande implication et réflexion dans le cas des méthodes de comparaisons par paires et *Best-Worst*.

En termes de capacité discriminante, il conclut au moyen d'un test d'égalité de moyennes des notes des attributs que la méthode *Best-Worst* est la plus performante des trois méthodes, tandis que l'échelle d'évaluation de type Likert est la moins discriminante. Dans le même sens, Goodman *et al.* (*op. cit.*) mesurent l'évaluation de l'importance des attributs explicatifs du choix d'un vin, au moyen de l'échelle d'évaluation de type Likert et de la méthode *Best-Worst*.

Cohen (*op. cit.*) vérifie également que l'échelle d'évaluation de type Likert est la moins performante des trois méthodes pour discriminer entre les individus. La méthode des comparaisons par paires semble être légèrement plus discriminante que la méthode *Best-Worst*. Enfin il conclut que la méthode *Best-Worst* est la plus stable puisqu'elle montre la fiabilité test-retest la plus satisfaisante (97 % contre 88 % dans le cas

Tableau 1. Comparaison de méthodes de mesure d'importance des attributs

		Collecte des données			Analyse des données		
		Administrations	Fiabilité des réponses	Durée administration	Discriminations entre attributs	Discriminations entre individus	Validité prédictive
Notation	Mesure directe des scores d'importance (Fishbein, <i>op. cit.</i> )	Très facile	Moyenne	Courte	Moyenne	Bonne	Médiocre
	Estimation comparative des grandeurs (magnitude estimation) (Lodge, <i>op. cit.</i> )	Complexe	Moyenne	Moyenne	Plus faible	Bonne	Bonne
	Échelle libre (Marder, <i>op. cit.</i> )	Facile	Moyenne	Moyenne	Plus faible	Bonne	Médiocre
	Somme constante des scores	Facile	Moyenne	moyenne	Plus faible	Bonne	Bonne
Arbitrage	Tri forcé (Q sort) variante du classement (Stephenson, <i>op. cit.</i> )	Complexe	Bonne	Moyenne	Plus forte	Bonne	Bonne
	Classement par paires	Facile	Bonne	Courte	Très forte	Bonne	
	Choix discrets (Finn et Louvière, <i>op. cit.</i> )	Complexe	Très bonne	longue	Plus forte	Peu aisée	Bonne
	Différence maximum ou choix des extrêmes « best worst » ( <i>ibid.</i> )	Facile	Très bonne	Longue	Plus forte	Bonne	Bonne (la meilleure)

Tableau 2. Caractéristiques socio-démographiques des répondants par pays

		Tunisie (n = 122)	France (n = 123)	Différences entre les deux échantillons	
				$\chi^2$	Sig (p)
Sexe	Homme	93	41	45,48	< 0,001
	Femmes	29	82		
Age	18 à 44 ans	67	49	5,59	< 0,05
	45 et plus 67	55	74		
Niveau d'éducation	Secondaire	27	56	14,97	< 0,001
	Universitaire	95	67		
Revenu	Faible et moyen*	59	85	10,88	< 0,01
	Haut*	63	38		

Notes : \* Faible et moyen revenu : Tunisie : < 1 000 dinars tunisiens, France : 3 693 euros par mois.

\* Revenu élevé: Tunisie :  $\geq$  1 000 dinars tunisiens, France : 3 693 euros par mois.

de la méthode des comparaisons par paires, et 85 % dans le cas des échelles de Likert).

Nous présentons une synthèse des évaluations de ces méthodes dans le *tableau 1*, en prenant en compte les critères liés à la qualité et à la facilité d'obtention des données et les critères liés au pouvoir discriminant et prédictif des analyses.

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

Les choix alimentaires sont des marqueurs des différences culturelles qui se reflètent dans la variété des dispositifs institutionnels ou privés qui encadrent l'offre et la demande de qualité alimentaire. Comme nous l'avons souligné, le choix d'étudier le secteur de l'huile d'olive est notamment lié aux interrogations des responsables de la filière quant à l'opportunité et la manière de promouvoir des politiques d'appellations d'origine, en France comme en Tunisie. Notre objectif est donc de déterminer quelle est l'importance relative des critères de choix pour les consommateurs d'huile d'olive, au sein de deux populations différentes (Tunisie et France), tant sur le plan de la fréquence et des modes d'achat de ce produit que sur la familiarité avec le concept d'appellation d'origine.

Nous avons mené dans un premier temps

une étude exploratoire afin d'identifier une liste exhaustive des critères de choix de l'huile d'olive (Dekhili et d'Hauteville, 2006). Deux enquêtes ont été réalisées : la première a consisté en des entretiens téléphoniques auprès de 60 consommateurs en France et en Tunisie (30 répondants dans chaque pays). La deuxième a mobilisé la méthode Delphi (15 experts ont été interrogés à deux reprises dans chaque pays). Les deux approches ont conduit à des résultats convergents. Nous avons ainsi retenu 13 attributs qui seront considérés dans la construction du plan d'expérience (pays d'origine, région d'origine, variété d'olives, goût, couleur, aspect, mention « extra-vierge », mention « biologique », prix, marque, emballage, producteur/ vendeur, signes officiels de qualité et médailles). Il s'agit d'un plan de type 13, 4, 4, 1, les treize critères apparaissant en blocs incomplets de quatre critères, chaque critère apparaissant quatre fois et étant comparé aux douze autres une fois et une seule. Cohen (2009), dans une recherche sur la consommation du vin, détaille la procédure suivie pour la construction du plan d'expérience.

Les tâches de choix sont présentées sous forme de tableaux successifs (voir *annexe 1*). Les répondants doivent choisir dans chaque tableau le critère qu'ils jugent le plus important dans leur choix, et le critère le moins important.

128 consommateurs tunisiens et 123 parisiens ont été interrogés au cours l'année 2006. Après dépouillement, 123 questionnaires français et 122 Tunisiens se sont avérés utilisables pour l'analyse *Best-Worst* (tableau 2).

Comme le montre le tableau 2 les deux échantillons sont de structure différente : plus de femmes en France et plus d'hommes en Tunisie, une moyenne d'âge supérieure en France, et une cible plus aisée en Tunisie (niveaux d'études et de revenus supérieurs) alors qu'en France l'achat d'huile d'olive en bouteille est plus courant. Nous avons donc analysé les interactions entre le pays du répondant et la composition de l'échantillon. L'analyse sur les variables sexe, âge, niveau d'éducation et revenu a été menée de façon indépendante pour chaque échantillon.

### Résultats

Le niveau d'importance de chaque critère de choix est mesuré par 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). Marley et Louvière (2005) montrent que lorsque le plan expérimental est complet, cette différence est une représentation dont l'échelle de valeur est à 95 % aussi correcte que le modèle logit multinomial construit à partir des mêmes données.

Pour chaque attribut, on mesure la différence des cumuls de « best » et de « worst ». Dans notre situation, la différence maximum B-W que peut atteindre un attribut est + 4 ou - 4. 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 entre groupes de répondants qui diffèrent en termes d'effectifs.

$$\text{Score standardisé} = \frac{\text{score (Best-Worst)}}{4n}$$

Où :  $n$  = le nombre des répondants  
 4 = la fréquence d'apparition de chaque attribut dans les ensembles de choix.

Tableau 3. Distinction entre l'effet pays et celui des variables démographiques sur l'évaluation des critères de choix

Attributs	Effet Pays des répondants	Effet variables socio-démographiques		Effets d'interactions	
		Type de variable	Sig	Type d'interaction	Sig
Pays d'origine	n.s.				
Région d'origine	0,009				
Prix	n.s.	Niveau d'éducation	0,037		
Marque	n.s.	Âge	0,032		
Variété d'olives	0,013			Pays*Sexe Pays*Âge Pays*Revenu	0,048 0,000 0,040
Producteur/Vendeur	n.s.			Pays*Niveau d'éducation Pays*Revenu	0,017 0,003
Goût	n.s.			Pays*Âge	0,024
Mention « Vierge extra »	0,000			Pays*Âge	0,010
Couleur	n.s.				
Aspect	n.s.				
Emballage	n.s.				
Signes officiels de qualité, médailles	0,000				
Mention « biologique »	0,015				

Note : Seuls les effets significatifs des variables sociodémographiques et de leur interaction avec le pays des répondants sont reportés dans ce tableau.

Tableau 4. Importance accordée aux critères de choix montrant un effet d'interaction significatif (pays d'origine\*variables sociodémographiques)

Attributs	Caractéristiques socio-démographiques		Score B-W standardisé	
			France	Tunisie
Variété d'olives	Sexe	Homme	-0,09	0,20
		Femme	0,03	0,08
	Âge	18-44 ans	0,08	0,10
> 44 ans		-0,07	0,27	
Producteur/Vendeur	Revenu	Faible et moyen	-0,03	0,24
		Haut	0,04	0,12
Goût	Études	Secondaires	-0,35	-0,23
		Universitaires	-0,31	-0,41
Mention « Vierge extra »	Revenu	Faible et moyen	-0,28	-0,41
		Haut	-0,43	-0,33
Goût	Âge	18-44 ans	0,64	0,52
		> 44 ans	0,43	0,55
Mention « Vierge extra »	Âge	18-44 ans	0,29	0,11
		> 44 ans	0,61	0,09

Tableau 5. Hiérarchie des critères de choix selon les tranches d'âge (France et Tunisie), établie sur la base des scores B-W standardisés

Attributs	France			Tunisie		
	Total	18-44 ans	> 44 ans	Total	18-44 ans	> 44 ans
	123 répondants	49 répondants	74 répondants	122 répondants	67 répondants	55 répondants
Goût	100,0	100,0	100,0	100,0	100,0	100,0
Mention « vierge extra »	55,0	30,9	88,2	28,3	32,2	23,5
SOQ, médailles	41,6	27,1	59,4	23,4	30,5	16,0
Pays d'origine	33,5	26,1	40,1	28,4	32,1	23,6
Région d'origine	22,7	16,7	28,1	36,6	40,7	31,1
Mention « biologique »	21,6	16,5	25,9	34,2	48,1	23,6
Variété d'olives	20,7	22,0	19,7	37,6	34,0	42,3
Prix	19,3	19,2	19,2	29,1	30,3	26,8
Marque	16,3	10,1	22,3	12,2	11,4	12,6
Aspect	14,8	15,4	14,3	22,9	25,3	19,5
Producteur/vendeur	10,4	8,1	12,3	11,0	13,3	8,3
Couleur	9,3	10,5	8,3	17,2	17,5	16,2
Emballage	4,7	4,6	4,7	5,9	6,4	5,1

*Note* : Le critère « goût » est considéré comme une référence (score=100)

En raison des différences constatées entre les deux échantillons (français et tunisien) en termes de caractéristiques socio-démographiques, nous avons conduit une analyse de variance (ANOVA) afin de distinguer entre l'effet pays et celui de ces variables sur l'évaluation des critères de choix du produit. Le tableau 3 synthétise les résultats de l'analyse ANOVA.

Concernant les attributs liés 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 de la mention « vierge extra », signes officiels de qualité, médailles et la mention « biologique ». Les Français accordent plus d'attention aux deux premiers attributs, alors que les Tunisiens considèrent plus le dernier dans leur choix.

Les attributs variété d'olives, producteur/vendeur, goût et mention « vierge extra » présentant un effet d'interaction « pays d'origine\*variables sociodémographiques » significatif. Nous rapportons dans le *tableau 4* la variation de leur importance selon le pays et les caractéristiques socio-démographiques des répondants.

En ce qui concerne les variables de segmentation de ces groupes, l'âge semble être celle qui différencie le plus les critères de choix, à la fois à l'intérieur des groupes mais aussi entre les deux groupes (*tableau 5*).

### **Intérêt et limites de la méthode Discriminer et hiérarchiser**

Certains contextes d'études nécessitent que l'on puisse à la fois discriminer entre les attributs des produits qui déterminent les choix, mais aussi quantifier les écarts d'importance entre les attributs, en s'affranchissant des biais méthodologiques liés aux échelles de mesure. Il faut donc disposer d'instruments qui favorisent ces deux objectifs, et le choix de la méthode *Best-Worst* paraît bien adapté. Le traitement de l'exemple que nous avons donné nous a permis de comparer assez efficacement le poids des attributs de choix d'une huile d'olive entre Tunisiens et Français, et nous a permis de vérifier que les attributs se hiérarchisent différemment selon les deux groupes nationaux, et à l'intérieur de ces groupes, selon les classes d'âge. 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. Par exemple, les consommateurs

Français sont plus sensibles à l'appellation d'origine que les consommateurs en Tunisie, où la réputation des signes officiels de qualité est encore à construire. De même, les Français (surtout les plus âgés) ont adopté le signal « vierge extra » comme un critère de choix presque aussi important que le goût, et cet attribut est également valorisé par les jeunes Tunisiens. À l'opposé, la variété d'olives est plus importante pour les consommateurs Tunisiens que pour les consommateurs Français. Cependant, les différences ne sont pas seulement interculturelles : par exemple en France les consommateurs plus jeunes se montrent plus sensibles à la variété d'olives que les consommateurs plus âgés. La méthode *Best-Worst* se révèle donc être un outil intéressant pour segmenter de façon fine les marchés. Les différences entre les deux échantillons peuvent cependant également être imputées à leur structure différente ; par ailleurs les échantillons n'étant pas des échantillons représentatifs des populations, nos résultats ne peuvent pas être généralisés et doivent plutôt être considérés comme les premiers éléments pouvant servir de base à une enquête de plus grande ampleur avec des échantillons représentatifs.

Le nombre de tâches de choix (treize) dans notre enquête était assez élevé, ce qui peut être un frein à l'application de cette méthode impliquante pour les répondants. Cependant la quasi-totalité des questionnaires s'est avérée exploitable. Les échantillons de répondants mis en œuvre sont relativement restreints, de l'ordre de la centaine, et le coût de mise en œuvre est donc très comparable avec des études plus classiques.

Notre expérience montre que les répondants comprennent facilement la tâche demandée, l'enquêteur doit veiller cependant à ce que toutes les tâches de choix soient renseignées par le répondant, au risque de ne pouvoir utiliser le questionnaire, c'est donc un point important auquel devra être attentif

le client de l'étude. Contrairement aux échelles de mesure classiques, le tableau de données ne peut pas comporter de valeurs manquantes.

Si la méthode Best-Worst présente de nombreux avantages qui justifient son utilisation, elle ne résout cependant pas tous les problèmes.

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. Dans l'exemple donné par notre étude, il est peu réaliste de croire que la mention du producteur/vendeur, ou que l'emballage n'influence pas le choix d'une huile d'olive par les consommateurs. 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 est donc sensible à l'hétérogénéité des attributs qui sont pris en considération dans l'analyse. La même prudence d'interprétation que pour d'autres études déclaratives doit donc prévaloir. Finalement, la méthode recèle les faiblesses inhérentes aux modèles d'attributs de type compensatoire qui mettent en comparaison des attributs appartenant à des catégories en réalité peu comparables entre elles. Les praticiens devront donc être particulièrement attentifs à cet inconvénient : dans notre étude, l'attribut « goût » aurait peut-être dû être évité.

La méthode partage également avec toutes les autres méthodes d'analyse de données déclaratives effectuées dans un contexte interculturel la double difficulté du choix des items pertinents dans des cultures différentes et de la traduction du libellé des items. Dans notre étude, nous avons contourné la difficulté de traduction en sélectionnant des acheteurs tunisiens francophones. Mais ce n'est pas évidemment une situation généralisable. Par ailleurs l'huile d'olive est achetée en Tunisie pour

ses propriétés sanitaires ou comme produit de beauté, ce qui n'est guère le cas en France. La liste proposée reflétait peut-être mal cette finalité particulière du produit liée à la culture Tunisienne.

Finalement, on notera dans les remarques précédentes qu'il ne s'agit pas de limites propres à la méthode elle-même, mais à toute méthode qui s'appuie sur l'hypothèse d'un arbitrage entre attributs par le consommateur lors du choix.

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 (notamment Marder, *op. cit.*) concerne le temps nécessaire aux répondants pour compléter le questionnaire. Notre expérience ne confirme pas cependant que ce temps nécessaire soit une limite réelle. En effet, les répondants ne semblent pas avoir été gênés ; il n'y a pratiquement aucun questionnaire non rempli. Cependant ce temps passé (20 à 30 minutes) peut constituer une limite forte s'il correspond à une difficulté perçue de la tâche trop importante, avec le 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.

Il conviendra donc de prévenir les répondants que ces répétitions font partie de la méthode, et il est recommandé que les instructions soient rappelées à chaque répondant au début de la tâche. La présence de l'enquêteur lors de l'administration des questionnaires sera une précaution utile.

La méthode *Best-Worst* pose enfin 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 impor-

tants ou non pertinents. Ainsi, nous l'avons évoqué, le répondant sera sommé de choisir entre le goût et l'emballage, même si pour lui le choix ne se présente pas de cette manière-là. Cette limite rejoint en partie celle qui est faite à propos des méthodes d'analyse conjointe dans laquelle certains choix peuvent s'avérer aberrants. Elle est cependant compensée par le fait que le nombre de situations de choix qui est proposé offre une diversité dans laquelle la totalité des répondants doit pouvoir se retrouver. Contrairement à l'analyse conjointe, les attributs ne sont pas présentés dans une logique de construction d'offre.

\* \*  
\*

À partir de la littérature, nous avons comparé dans cet article les performances de la méthode *Best-Worst*, encore appelée « Maxdiff », à d'autres méthodes d'enquêtes, fondées soit sur des notations sur des échelles numériques, soit sur des comparaisons entre alternatives. Nous avons utilisé cette méthode pour déterminer si les notions d'origine (pays d'origine, région d'origine, variété d'olives) constituaient des critères

importants du choix de l'huile d'olive chez les consommateurs en France et en Tunisie.

Nos résultats montrent qu'il existe des alternatives aux mesures classiques obtenues à partir d'échelles de Likert notamment pour améliorer le pouvoir discriminant des résultats et contourner les biais liés à l'appréciation différente des intervalles proposés par ces échelles. 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 tout particulièrement 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.

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. ■

---

 RÉFÉRENCES BIBLIOGRAPHIQUES

- Bartikowski B., Chandon J.-L., Gierl H. (2006). Calibration internationale des échelles sémantiques. *Décisions Marketing*, n° 43-44, p. 207-219.
- Chandon J.-L., Bartikowski B. (2004). Une échelle ordinaire permettant de classer les répondants en « satisfait », « indifférent » et « insatisfait ». *Recherche et Applications en Marketing*, n° 19, p. 39-51.
- Cohen E. (2009). Applying Best-Worst Scaling to Wine Marketing. *International Journal of Wine Business Research*, vol. 21, n° 1, p. 8-23.
- Cohen S. (2003). *Maximum Difference Scaling: Improved Measures of Importance and Preference for Segmentation*. Sawtooth Software Conference Proceeding, Sequim (WA).
- Dekhili S., d'Hauteville F. (2006). *Les dimensions perçues de l'image de la région d'origine. Cas de l'huile d'olive*. 22<sup>e</sup> Congrès de l'Association Française de Marketing (AFM), Nantes, 11-12 Mai, 19 p.
- Finn A., Louviere J.-J. (1992). Determining the Appropriate Response to Evidence of Public Concern: The Case of Food Safety. *Journal of Public Policy and Marketing*, n° 11, p. 12-25.
- Fishbein M. (1967). *A behaviour theory approach to the relations between beliefs about an object and the attitude towards the object*. Readings in attitude theory and measurement, New York: Wiley, p. 389-399.
- Goodman S., Lockshin L., Cohen E. (2005). *Best-Worst Scaling: A simple Method to Determine Drinks and Wine Style Preference*. 2<sup>nd</sup> International Wine Marketing and Business Conference, Sonoma (USA).
- Lodge M. (1981). *Magnitude scaling: quantitative measurement of opinions*. Sage University paper on quantitative applications in social sciences, Beverly Hills, Sage Publications, p. 7-25.
- Louviere J.-J., Woodworth G.-G. (1990). *Best Worst Scaling: a model for largest Difference Judgements*. Working paper, faculty of Business, University of Alberta.
- Marder E. (1997). *The laws of Choice: predicting consumer behaviour*. N.-Y., the Free Press.
- Marley A.-A.-J., Louvière J.-J. (2005). Some probabilistic Models of Best, Worst and Best Worst choices. *Journal of mathematical psychology*, n° 49, p. 464-480.
- Rao P.-V. (1960). *The dual of balanced incomplete block design*. The Annals of Mathematical Statistics, p. 779-785.
- Remaud H., Lockshin L. (2009). Building brand salience for commodity-based wine regions. *International Journal of Wine Business research*, n° 21.
- Rosenberg M.-J. (1956). Cognitive structure and attitudinal affect. *The Journal of Abnormal and Social Psychology*, n° 53, p. 367-372.
- Stephenson W. (1953). *The study of behavior: the Q-Technique and its methodology*. Chicago, University of Chicago Press.
- Wilkie W.-L., Pessemier E.-A. (1973). Issues in marketing's use of multi attribute attitude models. *Journal of Marketing research*, n° 10, p. 428-441.

ANNEXE

Description des 13 blocs de 4 attributs chacun

Tableau 1

Le moins important	Critère de choix	Le plus important
<input type="checkbox"/>	1 Le pays d'origine	<input type="checkbox"/>
<input type="checkbox"/>	2 La région d'origine	<input type="checkbox"/>
<input type="checkbox"/>	3 La marque	<input type="checkbox"/>
<input type="checkbox"/>	4 L'aspect	<input type="checkbox"/>

Tableau 2

Le moins important	Critère de choix	Le plus important
<input type="checkbox"/>	1 La région d'origine	<input type="checkbox"/>
<input type="checkbox"/>	2 Le prix	<input type="checkbox"/>
<input type="checkbox"/>	3 La variété d'olives	<input type="checkbox"/>
<input type="checkbox"/>	4 L'emballage	<input type="checkbox"/>

Tableau 3

Le moins important	Critère de choix	Le plus important
<input type="checkbox"/>	1 Le prix	<input type="checkbox"/>
<input type="checkbox"/>	2 La marque	<input type="checkbox"/>
<input type="checkbox"/>	3 Le producteur/vendeur	<input type="checkbox"/>
<input type="checkbox"/>	4 Les signes officiels de qualité (AOC), les médailles	<input type="checkbox"/>

Tableau 4

Le moins important	Critère de choix	Le plus important
<input type="checkbox"/>	1 La marque	<input type="checkbox"/>
<input type="checkbox"/>	2 La variété d'olives	<input type="checkbox"/>
<input type="checkbox"/>	3 Le goût	<input type="checkbox"/>
<input type="checkbox"/>	4 La mention « biologique »	<input type="checkbox"/>

Tableau 5

Le moins important	Critère de choix	Le plus important
<input type="checkbox"/>	1 La variété d'olives	<input type="checkbox"/>
<input type="checkbox"/>	2 Le producteur/vendeur	<input type="checkbox"/>
<input type="checkbox"/>	3 La mention « Vierge extra »	<input type="checkbox"/>
<input type="checkbox"/>	4 Le pays d'origine	<input type="checkbox"/>

Tableau 6

Le moins important	Critère de choix	Le plus important
<input type="checkbox"/>	1 Le producteur/vendeur	<input type="checkbox"/>
<input type="checkbox"/>	2 Le goût	<input type="checkbox"/>
<input type="checkbox"/>	3 La couleur	<input type="checkbox"/>
<input type="checkbox"/>	4 La région d'origine	<input type="checkbox"/>

Tableau 7

Le moins important	Critère de choix	Le plus important
<input type="checkbox"/>	1 Le goût	<input type="checkbox"/>
<input type="checkbox"/>	2 La mention « Vierge extra »	<input type="checkbox"/>
<input type="checkbox"/>	3 L'aspect	<input type="checkbox"/>
<input type="checkbox"/>	4 Le prix	<input type="checkbox"/>

Tableau 8

Le moins important	Critère de choix	Le plus important
<input type="checkbox"/>	1 La mention « Vierge extra »	<input type="checkbox"/>
<input type="checkbox"/>	2 La couleur	<input type="checkbox"/>
<input type="checkbox"/>	3 L'emballage	<input type="checkbox"/>
<input type="checkbox"/>	4 La marque	<input type="checkbox"/>

Tableau 9

Le moins important	Critère de choix	Le plus important
<input type="checkbox"/>	1 La couleur	<input type="checkbox"/>
<input type="checkbox"/>	2 L'aspect	<input type="checkbox"/>
<input type="checkbox"/>	3 Les signes officiels de qualité (AOC), les médailles	<input type="checkbox"/>
<input type="checkbox"/>	4 La variété d'olives	<input type="checkbox"/>

Tableau 10

Le moins important	Critère de choix	Le plus important
<input type="checkbox"/>	1 L'aspect	<input type="checkbox"/>
<input type="checkbox"/>	2 L'emballage	<input type="checkbox"/>
<input type="checkbox"/>	3 La mention « biologique »	<input type="checkbox"/>
<input type="checkbox"/>	4 Le producteur/vendeur	<input type="checkbox"/>

Tableau 11

Le moins important	Critère de choix	Le plus important
<input type="checkbox"/>	1 L'emballage	<input type="checkbox"/>
<input type="checkbox"/>	2 Les signes officiels de qualité (AOC), les médailles	<input type="checkbox"/>
<input type="checkbox"/>	3 Le pays d'origine	<input type="checkbox"/>
<input type="checkbox"/>	4 Le goût	<input type="checkbox"/>

Tableau 12

Le moins important	Critère de choix	Le plus important
<input type="checkbox"/>	1 Les signes officiels de qualité (AOC), les médailles	<input type="checkbox"/>
<input type="checkbox"/>	2 La mention « biologique »	<input type="checkbox"/>
<input type="checkbox"/>	3 La région d'origine	<input type="checkbox"/>
<input type="checkbox"/>	4 La mention « Vierge extra »	<input type="checkbox"/>

Tableau 13

Le moins important	Critère de choix	Le plus important
<input type="checkbox"/>	1 La mention « biologique »	<input type="checkbox"/>
<input type="checkbox"/>	2 Le pays d'origine	<input type="checkbox"/>
<input type="checkbox"/>	3 Le prix	<input type="checkbox"/>
<input type="checkbox"/>	4 La couleur	<input type="checkbox"/>