UN NUOVO APPROCCIO DI CONJOINT ANALYSIS: STIMA DI PIÙ FUNZIONI DI RISPOSTA

Amedeo De Luca

Università Cattolica del Sacro Cuore di Milano – Dipartimento di Scienze Statistiche – amedeo.deluca@unicatt.it

Riassunto

La metodologia classica della Conjoint Analysis (COA), "metrica" e "non metrica", conduce, per ciascun intervistato, alla stima di un set di funzioni di utilità parziale o coefficienti separati (part worth), ciascuna delle quali è associata ad un attributo del prodotto da lanciare sul mercato.

Nell'approccio tradizionale, aggregando le valutazioni dei singoli rispondenti, appartenenti al campione osservato, si giunge ad una funzione di risposta aggregata (Hagerty, 1985).

L'approccio alla COA che qui si propone – di tipo "non metrico" – consente di stimare a livello aggregato non una, ma molteplici funzioni di risposta, tante quante sono le categorie di giudizio di valutazione globale o di overall.

Dette funzioni si ottengono collegando la categoria di overall – espressa su scala di punteggio di tipo "ordinato" – alle modalità degli attributi del prodotto tramite un modello di regressione lineare multivariata generalizzata (la quale evita, tra l'altro, il ricorso alla preliminare e complessa trasformazione di scala, operata – tipicamente – con la regressione monotona di Kruskal, che rende "metrica" la scala "ordinata" su cui è espresso l'overall).

Le funzioni di risposta così stimate consentono di valutare ed interpretare l'importanza di ciascuna modalità degli attributi di un prodotto sulle diverse categorie di giudizio di valutazione globale, in modo analitico e puntuale, saggiando la correttezza dei risultati del modello nel suo insieme anche per via empirica, tramite riscontro della coerenza tra i coefficienti delle diverse funzioni di risposta.

Del modello proposto si dà un'applicazione, con riferimento ad una nuova polizza assicurativa vita index-linked.

1. INTRODUZIONE

In termini semplificati, la Conjoint Analysis (COA) si basa sui giudizi di gradimento globale (*overall evaluation*) espressi, in forma di punteggio su scala graduata o di graduatoria di preferenza, da un campione di intervistati, su vari profili di un prodotto (bene o servizio) nuovo o innovato. Detti profili differiscono per definizione – in modo sistematico – per le modalità di qualche carattere o fattore, che sono state variate secondo un preciso schema sperimentale (ad esempio, secondo un piano di 2^h combinazioni, se gli h fattori sono considerati a 2 modalità), per definire le diverse unità sperimentali.

Il punteggio di gradimento, o il grado di preferenza globale, può ritenersi espresso su una scala *ordinale* e si tratta di studiare, mediante un modello di regressione lineare su variabili *dummy*, gli effetti dei fattori sistematici (attributi) sul grado di desiderabilità o di preferenza del nuovo prodotto.

Il quadro di riferimento è, pertanto, costituito dal problema dello studio di un "carattere qualitativo ordinato" (giudizio di *overall* Y) "in funzione" di altri caratteri qualitativi "esplicativi" X (attributi o fattori del prodotto). Come è noto, risulta in corrispondenza essenziale, quale premessa all'interpretazione ed analisi statistica dei risultati della COA, stabilire la scala sulla quale si intendono rappresentate le modalità del carattere ordinato in esame.

Al riguardo è abbastanza consueto individuare le modalità di un carattere qualitativo ordinato mediante la tecnica del differenziale semantico, che porta ad una graduazione delle percezioni tra due situazioni qualitative estreme (ad esempio, "sgradito", "molto gradito") e ad assegnare alle K classi o categorie di valutazione, così definite, dei valori interi, ad esempio da 1 ad K.

Ora, l'utilizzazione appropriata dei valori numerici (punteggi) ottenuti mediante scale del tipo anzidetto può presentare delle difficoltà: talvolta si assumono i punteggi come se fossero valori su una scala metrica *ad intervalli*, ciò che non è in generale vero.

Non sarebbe, pertanto, corretto risolvere il problema in studio utilizzando l'usuale tecnica dell'analisi di regressione, applicandola ai valori di punteggio ottenuti (ad esempio, 1, 2, 3) per Y in corrispondenza delle combinazioni delle modalità degli M fattori $X_1, X_2, \ldots, X_M$ considerati, come se detti valori fossero espressi su una scala ad intervalli, secondo la cosiddetta COA "metrica" (la quale pure viene usualmente applicata dagli operativi, anche se con evidenti limitazioni di carattere teorico ed interpretativo).

Nella presente proposta si è evitato di assegnare dei valori numerici alle modalità dell'*overall*, quali sarebbero, ad esempio, $\{1, 2, 3\}$, per la risposta Y_k ($k = 1, 2, 3$) di gradimento complessivo di un profilo di prodotto, i quali – anche se si ritiene di essere in grado di poter stabilire per l'*overall* una scala ordinale – sono valori convenzionali, in quanto definiti a meno di una trasformazione monotona strettamente crescente (Teorema di rappresentazione su una scala ordinale, si veda Krantz, 1971, Cap. I, p. 15), ciò che preclude di poterli trattare con le consuete tecniche di analisi della regressione, giustificate per valori su scale ad intervalli o di rapporti.

Valori numerici, di una scala *di rapporti*, sono invece ipotizzati – nella presente proposta – per le probabilità p_k, k = 1, 2, 3: grado di desiderabilità complessiva espresso dal campione di rispondenti per i vari profili completi di prodotto. Si tratta delle probabilità che il soggetto effettui la scelta di una delle situazioni oggettive (cioè, le diverse combinazioni di modalità degli attributi del prodotto) ritenute possibili e distinguibili.

Questo con il vantaggio dell'impiego, per esprimere un determinato grado di preferenza, delle probabilità p_k come "risposta media", che non richiede trasformazioni di scala per rendere "metrica" la scala di gradimento.

1.2 LA NUOVA INTERPRETAZIONE DEL MODELLO DI CONJOINT ANALYSIS

Il modello interpretativo di COA che viene illustrato in questa nota si presenta piuttosto complesso dal punto di vista matematico-statistico. In corrispondenza, nella speranza di renderne più agevole la lettura, la metodologia è presentata in stretto collegamento con un caso concreto, che ovviamente potrà costituire spunto per altre applicazioni.

L'applicazione di riferimento riguarda un'indagine condotta – sulla base di un piano sperimentale – su un campione di 100 funzionari di compagnie assicurative, con riguardo al gradimento verso un'ipotetica nuova polizza vita *index-linked*.

I giudizi di gradimento globale o *overall* (Y) su vari profili completi del prodotto sono espressi sulle seguenti tre classi: "non gradito", "gradito" e "molto gradito". Dette classi costituiscono le modalità di una variabile dipendente *ordinale*.

Per collegare il livello di gradimento globale Y_k, k = 1, 2, 3 (espresso su una scala di punteggio, di tipo *ordinato*, sui vari profili del nuovo prodotto) con le modalità di quattro attributi o fattori sperimentali (X_1 = *durata della polizza*, con due livelli o modalità: "5 anni", "8 anni"; X_2 = *taglio minimo*, con due livelli: "2.500 euro", "5.000 euro"; X_3 = *indice di borsa*, con tre livelli: "Comit", "Dow Jones", "Nikkei"; X_4 = *prestazione alla scadenza*, con due livelli: "capitale", "rendita vitalizia"), il vettore riassuntivo delle probabilità di scelta di una delle tre anzidette K = 3 categorie ordinate di giudizio viene interpretato tramite un modello di regressione lineare multipla multivariata di variabili indicatrici, con regressori costituiti da variabili dicotomiche, descriventi le modalità dei fattori sperimentali.

L'interpretazione probabilistica di scelta di una data categoria di *overall* – che così ne consegue – evita il ricorso a preliminari ed impegnative trasformazioni della scala *ordinata* (sulla quale è espresso il gradimento globale) in una scala metrica *ad intervalli*, secondo l'approccio tradizionale della COA "non metrica" (Green, 1988), con il quale tale trasformazione è effettuata tramite la regressione monotona di Kruskal (1964), basata su metodi iterativi (MONANOVA; De Luca, 2006).

2. IL MODELLO DI REGRESSIONE MULTIPLA MULTIVARIATA GENERALIZZATA VINCOLATA SU VARIABILI INDICATRICI PER LA STIMA DI PIÙ FUNZIONI AGGREGATE DI RISPOSTA NELLA CONJOINT ANALYSIS

Traendo spunto da uno studio inerente alla regressione lineare multipla multivariata su variabili indicatrici (De Luca *et alii*, 2004), nell'approccio che qui si propone la variabile di risposta (*overall*) è descritta come funzione di variabili indicatrici di natura dicotomica binaria (1, 0). Esso riguarda la COA basata sugli *scores* di "gradimento" (*rating*) o "utilità" (*utility*) assegnati da un campione di intervistati su S differenti profili di un prodotto.

Detti profili corrispondono alle combinazioni dei livelli di $M = 4$ fattori pari, rispettivamente, a: 2, 2, 3, 2, le quali formano un piano *fattoriale completo* di 24 stimoli sperimentali (con 100 "replicazioni" per combinazione, tante quanti sono gli intervistati).

Si intendono qui studiare gli effetti dei fattori sistematici sul livello di gradimento del prodotto mediante un modello di regressione lineare multipla multivariata (Goldberger, 1964).

Con la parametrazione adottata nel modello proposto – senza interazioni per semplicità – gli effetti dei fattori esprimono le variazioni delle probabilità p_{ks} di scelta della categaria di *overall* k-ima conseguenti alle differenti combinazione s (s =1, 2, …, S, con S = 24 nell'applicazione esemplificativa che viene qui presentata) di livelli di X_1, X_2, X_3, X_4 rispetto ad una prescelta classe di riferimento (*baseline*).

Per stimare dette probabilità p_{ks} si fa ricorso ad un modello aggregato di utilità, atto a descrivere la struttura di *preferenza comune* di un gruppo omogeneo di rispondenti. Le valutazioni dei soggetti su ciascuno stimolo sperimentale (di tipo *full-profile*) sono considerate delle osservazioni ripetute sulla stessa combinazione. Ipotesi ammissibile (secondo Moore, 1980, p. 517), essendo nell'applicazione i rispondenti dei funzionari assicurativi, valutatori esperti e, quindi, *omogenei* tra di loro.

Per stimare la relazione tra la variabile dipendente Y_k ed M variabili indipendenti qualitative (fattori X), con livelli l = 1, 2, …, l_m (con m = 1, 2, …, M), le K = 3 categorie di *overall* sono espresse nelle presente proposta (v. Tab. 1*a*) in funzione di tre variabili indicatrici binarie (Y_k, k = 1, 2, 3), che assumono valore 1 se il rispondente esprime un gradimento di classe k – se k è la categoria considerata – valore 0 nel caso contrario (De Luca, 2000); i livelli di ciascun attributo del prodotto sono descritti tramite variabili indicatrici Z_l, con l = 1, 2, per il primo, secondo e quarto fattore, e l = 1, 2, 3 per il terzo fattore.

I punteggi di gradimento assegnati dai rispondenti vengono considerati nel loro insieme (*pooled model*) e – novità dell'approccio qui presentato – per ciascun livello o categoria di *overall* Y_k viene stimato un *set di funzioni* di utilità parziali (*part worth*).

Tab. 1a: Codifica binaria disgiuntiva completa delle categorie di giudizio globale (Y_k)

Variabili indicatrici / Categoria di overall (Y_k)	Y_1	Y_2	Y_3
k = 1 "non gradito"	1	0	0
k = 2 "gradito"	0	1	0
k = 3 "molto gradito"	0	0	1

Per giungere alla definizione del modello che stabilisce le relazioni tra le categorie della variabile dipendente Y e le variabili indipendenti X_1, X_2, X_3 e X_4, si fa ricorso, anche per queste ultime – come è del resto tipico della COA – alla codifica binaria disgiuntiva completa.

A ciascuno degli M fattori qualitativi indipendenti X si associa un numero di variabili quantitative dicotomiche binarie $Z^{(m)}$ pari al numero di livelli, come di seguito descritto (Tabb. 1*b* ÷ 1*e*).

Tab. 1b-e: Codifica binaria disgiuntiva completa dei livelli dei fattori "Durata polizza", "Taglio minimo", "Indice di borsa" e "Prestazione alla scadenza" della polizza vita

b)

Fattori e livelli	Variabili d*ummy*	
Durata polizza	$Z_1^{(1)}$	$Z_2^{(1)}$
5 anni	1	0
8 anni	0	1

c)

Fattori e livelli	Variabili d*ummy*	
Taglio minimo	$Z_1^{(2)}$	$Z_2^{(2)}$
2500 euro	1	0
5000 euro	0	1

d)

Fattori e livelli	Variabili d*ummy*		
Indice di borsa	$Z_1^{(3)}$	$Z_2^{(3)}$	$Z_3^{(3)}$
Comit	1	0	0
Dow Jones	0	1	0
Nikkei	0	0	1

e)

Fattori e livelli	Variabili d*ummy*	
Prestazione alla scadenza	$Z_1^{(4)}$	$Z_2^{(4)}$
capitale	1	0
rendita vitalizia.	0	1

2.1 PRIMA FASE DELLA PROCEDURA DI CALCOLO: STIMA DEI PARAMETRI DELLE REGRESSIONI UNIVARIATE

Nel modello di regressione multipla multivariata qui configurato le K equazioni di regressione univariate lineari su variabili indicatrici risultano collegate tra di loro, essendo le risposte sull'*overall* Y_k mutuamente esclusive ed esaustive; pertanto la K-ima equazione può essere ricavata dalle restanti $q = K - 1$ equazioni.

Si consideri per la generica k-ima categoria della variabile di gradimento di *overall* Y il modello di regressione *univariata* senza intercetta che stabilisce una relazione lineare tra la variabile Y_k, indicatore della k-ma categoria, e le variabili indipendenti X_1, X_2, X_3 e X_4:

$$\mathbf{Y}_k = \mathbf{Z}\delta_k + \mathbf{E}_k, \qquad k = 1, 2, \tag{1}$$

dove:

$\mathbf{Y}_k = [Y_{k1}, Y_{k2}, \ldots, Y_{kJ}]'$ è il vettore colonna, di dimensioni $SJ\times 1$, con generica osservazione (categoria di *overall*) Y_{ksj}, $k = 1, 2, \ldots K$; $s = 1, 2, \ldots, S$; $j = 1, 2, \ldots, J$, sulla variabile dipendente k-ima associata allo stimolo s ed al rispondente j;

$\mathbf{Z}$ matrice *fissa* del piano sperimentale di dimensioni $SJ \times \sum_{m=1}^{M} l_m$ (composta da J sotto-matrici $\mathbf{Z}_j$ di variabili indicatrici associate alle S combinazioni del piano; ciascuna $\mathbf{Z}_j$ è di dimensioni $S \times \sum_{m=1}^{M} l_m$, pari a 24×9 nell'applicazione), con elementi $z_{ls}^{(m)}$ (variabile indicatrice relativa alla modalità "l" del fattore "m" nello stimolo "s");

$\delta_k = \left[\delta_{k1}^{(1)}, \delta_{k2}^{(1)}; \delta_{k1}^{(2)}, \delta_{k2}^{(2)}; \delta_{k1}^{(3)}, \delta_{k2}^{(3)}, \delta_{k3}^{(3)}; \delta_{k1}^{(4)}, \delta_{k2}^{(4)}\right]$ vettore colonna di dimensioni $\sum_{m=1}^{M} l_m \times 1$, con l_m indicante il numero di livelli del fattore m-imo (di dimensioni 9×1, nel caso qui considerato) contenente i parametri incogniti $\delta_{kl}^{(m)}$, $k = 1, 2, \ldots, q$; $m = 1, 2, \ldots, M$; $l = 1, 2, \ldots, l_m$, associati alle variabili indicatrici delle classi l di ogni attributo;

$\mathbf{E}_k = [E_{k11}, E_{k21}, \ldots, E_{kS1,} E_{k12}, E_{k22}, \ldots, E_{kS2,} \ldots, E_{k1J}, E_{k2J}, \ldots, E_{kSJ}]'$ vettore colonna $SJ\times 1$. E_{ksj} rappresenta l' "errore di osservazione" relativo all's-imo stimolo e alla j-ima unità statistica (Y_k assume valori 0 o 1). Le E_{kj} sono Ueteroschedastiche (Johnston, 1966; pp. 244 - 248) e assunte indipendenti.

La forma algebrica del modello *con intercetta*, riparametrato in seguito alla soppressione – per i quattro fattori[1] – della colonna corrispondente alla prima classe, é:

$$y_{ks} = \tilde{c}_k + \sum_{m=1}^{M} \sum_{l=2}^{l_m} \tilde{\delta}_{kl}^{(m)} \tilde{z}_{ls}^{(m)} + E_{ks}, \qquad k = 1, \dots, q;\ s = 1, 2, \dots, S, \tag{2}$$

dove: $\tilde{c}_k$ corrisponde alla media condizionata $M(Y_{k|\tilde{z}_1^{(1)},\tilde{z}_1^{(2)},\tilde{z}_1^{(3)},\tilde{z}_1^{(4)}}) = \tilde{\delta}_{k1}$ media dei casi con valore 1 per la variabile Y_k attinenti alla classe di riferimento); $\tilde{\delta}_{kl}^{(m)}$, $\tilde{z}_{ls}^{(m)}$ e E_{ks} hanno un significato omologo ai corrispondenti termini della (1).

2.2 SECONDA FASE DELLA PROCEDURA DI CALCOLO: STIMA DEI PARAMETRI DELLE REGRESSIONI UNIVARIATE IN FORMA COMPATTA

La matrice $\tilde{\mathbf{Z}}$ del piano degli esperimenti sottostante la (2) è di dimensioni $JS \times \left(1 + \sum_{m=1}^{M} l_m - M\right)$, pari a 2400×6 nell'applicazione.

Le $q = 2$ equazioni (2), in forma compatta si possono indicare nel modo seguente:

$$\mathbf{Y}^* = \tilde{\mathbf{Z}}^* \tilde{\boldsymbol{\delta}}^* + \mathbf{E}^*, \tag{3}$$

con:

- $\mathbf{Y}^*$ vettore composto (VEC) da $q = 2$ vettori colonna $JS \times 1$, ognuno dei quali contiene le osservazioni della variabile dipendente (per $k = 1$ e $k = 2$) per ciascuno dei $J = 100$ rispondenti sugli S stimoli;
- $\tilde{\mathbf{Z}}^*$ matrice diagonale composta *quadrata*, contenente $q \times q$ sottomatrici di dimensioni $SJ \times \left(1 + \sum_{m=1}^{M} l_m - M\right)$, delle quali le q sottomatrici $\tilde{\mathbf{Z}}$ collocate sulla diagonale principale (uguali tra di loro) presentano in colonna le variabili indicatrici indipendenti corrispondenti alle diverse equazioni, mentre le restanti sottomatrici sono composte da elementi nulli;
- $\boldsymbol{\delta}^*$ vettore composto (VEC) di q vettori colonna dei coefficienti di regressione $\tilde{\delta}_k$, ciascuno di dimensioni $\left(1 + \sum_{m=1}^{M} l_m - M\right) \times 1$;

---

[1] *Non avendo la matrice Z caratteristica completa vengono posti eguali a zero M coefficienti del modello (corrispondenti alla prima categoria degli M fattori) per ottenere la stimabilità dei parametri (De Luca, 2000).*

$\mathbf{E}^*$ vettore composto (VEC) di q vettori colonna degli errori $\mathbf{E}_k$.

Con riferimento alla (3), interpretata in termini probabilistici ("risposta media"), la condizione di *disuguaglianza*, cui deve soggiacere un valore di probabilità, è: $0 \leq p_{ks} \leq 1$; tale condizione si traduce[2] (De Luca, 2001) – indicando con $\tilde{\mathbf{z}}_s'$ il vettore riga delle variabili indicatrici del generico stimolo sperimentale della matrice $\tilde{\mathbf{Z}}$ – nel vincolo:

$$0 \leq \tilde{\mathbf{z}}_s' \, \tilde{\delta}_k \leq 1, \quad k = 1, 2, \ldots, q \,;\, s = 1, 2, \ldots, S. \tag{4}$$

La (4) impone, per la stima dei parametri del modello (2), il ricorso al metodo dei *minimi quadrati vincolati* (Goldberger, 1964) ed alla *Programmazione Quadratica* (PQ; Gill *et alii*, 1981) su ciascuna equazione univariata.

2.3 TERZA FASE DELLA PROCEDURA DI CALCOLO: IL MODELLO DI REGRESSIONE MULTIVARIATA GENERALIZZATA VINCOLATA

Per giungere, infine, alla stima dei parametri del modello multivariato (3) necessita considerare le covarianze fra le Y_k, $k = 1, 2, \ldots, q$ (Kotz, Johnson, 1985; pp. 659-660).

Ritenendo che per ogni rispondente j ($j = 1, 2, \ldots, J$) la valutazione Y_{kj} sia descritta da una osservazione multinomiale (De Luca *et alii*, 2004) con q componenti e che sussista l'indipendenza stocastica al variare di j, la matrice di varianze-covarianze del modello risulta la seguente:

$$\mathbf{\Phi} = \left[\begin{array}{ccc|ccc|ccc}
p_{111}(1-p_{111}) & \cdots & 0 & -p_{111}p_{211} & \cdots & 0 & \dots -p_{111}p_{q11} & \cdots & 0 \\
\vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\
0 & \cdots & p_{1sj}(1-p_{1sj}) & 0 & \cdots & -p_{1SJ}p_{2SJ} & 0 & \cdots & -p_{1SJ}p_{2SJ} \\
\hline
-p_{211}p_{111} & \cdots & 0 & p_{211}(1-p_{211}) & \cdots & 0 & \dots -p_{211}p_{q11} & \cdots & 0 \\
\vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\
0 & \cdots & -p_{2SJ}p_{1SJ} & 0 & \cdots & p_{2SJ}(1-p_{2SJ}) & 0 & \cdots & -p_{2SJ}p_{qSJ} \\
\hline
-p_{q11}p_{111} & \cdots & 0 & -p_{q11}p_{211} & \cdots & 0 & \dots p_{q11}(1-p_{q11}) & \cdots & 0 \\
\vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\
0 & \cdots & -p_{qSJ}p_{1SJ} & 0 & \cdots & -p_{qSJ}p_{2SJ} & 0 & \cdots & p_{qSJ}(1-p_{qSJ})
\end{array}\right]$$

---

[2] *Siccome la variabile dipendente Y_k assume solo modalità 0 o 1, il modello di regressione lineare non garantisce che il valore di probabilità, stimato per una data combinazione di livelli dei fattori, sia compreso tra 0 e 1, potendosi ottenere con una funzione lineare valori di probabilità esterni all'intervallo 0 – 1, diversamente dalla regressione logistica (il cui impiego nell'ambito della* COA *è attualmente al nostro studio).*

Le stime degli elementi della matrice $\boldsymbol{\Phi}$, non singolare, sono calcolate sulla base delle stime $\hat{p}_{ksj}$ (k = 1, 2, …, q), ottenute applicando il metodo dei *Minimi Quadrati Generalizzati* (MQG; Goldberger, 1964) alle singole equazioni del modello stimate nella seconda fase della procedura di calcolo.

Per la stima del modello di regressione multivariata, effettuata con il metodo dei MQG, si minimizza, sotto i vincoli (4), e sotto la condizione: $\sum_{k=1}^{3} \tilde{\mathbf{z}}'_j \tilde{\delta}_k = 1$, la seguente espressione:

$$F = \left(\mathbf{Y}^* - \tilde{\mathbf{Z}}^* \tilde{\boldsymbol{\delta}}^*\right)' \hat{\Phi}^{-1} \left(\mathbf{Y}^* - \tilde{\mathbf{Z}}^* \tilde{\boldsymbol{\delta}}^*\right), \tag{5}$$

dove la matrice $\hat{\boldsymbol{\Phi}}^{-1}$ è l'inversa della matrice $\hat{\boldsymbol{\Phi}}$ (una stima della matrice Φ).

La (5), unita ai vincoli (4) ed alla condizione $\sum_{k=1}^{3} \tilde{\mathbf{z}}'_j \tilde{\delta}_k = 1$, richiede il ricorso alla PQ (Gill *et alii*., 1981).

3. APPLICAZIONE DEL MODELLO PROPOSTO: STIMA DI PIÙ FUNZIONE DI RISPOSTA NELLA CONJOINT ANALYSIS

Il modello è stato applicato alle 24 valutazioni globali (Y_k) individuali espresse sui 24 stimoli sperimentali – parzialmente casualizzati[3] (v. Tab. 3) – dal campione ragionato di 100 funzionari di compagnie di assicurazione, su una scala graduata ordinale di K = 3 categorie: "non gradito", "gradito", "molto gradito", come descritto nel § 1.2. In Tab. 2 sono riportati gli stimoli sperimentali secondo l'ordine *standard* del piano fattoriale completo.

---

[3] *La casualizzazione – considerando il calo d'attenzione di un rispondente dopo la valutazione di un certo numero di profili del prodotto – mira ad evitare che venga penalizzato l'interesse verso i profili collocati alla fine dell'elenco di un piano standard. La casualizzazione parziale o ristretta consente di far variare – nella successione degli stimoli sperimentali – il meno possibile i livelli di quei fattori il cui frequente cambiamento disturberebbe l'operativa di valutazione del rispondente (De Luca, 2006; p. 360).*

Tab. 2: Piano fattoriale completo secondo l'ordine standard delle combinazioni sperimentali delle modalità dei quattro attributi di una polizza vita index-linked.

Profilo di prodotto (s)	**Durata della polizza in anni** (X_1)	**Taglio minimo in euro** (X_2)	**Indice di borsa** (X_3)	**Prestazione alla scadenza** (X_4)
1	5	2500	Nikkei	capitale
2	5	2500	Nikkei	rendita
3	5	2500	Comit	capitale
4	5	2500	Comit	rendita
5	5	2500	Dow Jones	capitale
6	5	2500	Dow Jones	rendita
7	5	5000	Nikkei	capitale
8	5	5000	Nikkei	rendita
9	5	5000	Comit	capitale
10	5	5000	Comit	rendita
11	5	5000	Dow Jones	capitale
12	5	5000	Dow Jones	rendita
13	8	2500	Nikkei	capitale
14	8	2500	Nikkei	rendita
15	8	2500	Comit	capitale
16	8	2500	Comit	rendita
17	8	2500	Dow Jones	capitale
18	8	2500	Dow Jones	rendita
19	8	5000	Nikkei	capitale
20	8	5000	Nikkei	rendita
21	8	5000	Comit	capitale
22	8	5000	Comit	rendita
23	8	5000	Dow Jones	capitale
24	8	5000	Dow Jones	rendita

Nell'Appendice 2 si riporta uno stralcio esemplificativo del *file* di input dell'applicazione, in forma di codifica binaria disgiuntiva completa, sia dei giudizi di *overall* che degli attributi della polizza vita *index-linked;* l'ordine degli stimoli sperimentali è quello della somministrazione casualizzata riportato in Tab. 3.

Tab. 3: Piano fattoriale completo *parzialmente casualizzato* delle combinazioni sperimentali delle modalità dei quattro attributi di una polizza vita index-linked.

Profilo di prodotto (s)	**Durata della polizza in anni** (X_1)	**Taglio minimo in euro** (X_2)	**Indice di borsa** (X_3)	**Prestazione alla scadenza** (X_4)
1	5	2500	Nikkei	capitale
16	5	5000	Comit	capitale
9	8	5000	Dow Jones	capitale
18	8	5000	Comit	capitale
12	5	5000	Dow Jones	capitale
19	8	2500	Nikkei	capitale
10	8	2500	Comit	capitale
24	8	2500	Dow Jones	capitale
2	5	2500	Comit	capitale
15	5	5000	Nikkei	capitale
5	8	5000	Nikkei	capitale
17	5	2500	Dow Jones	capitale
6	8	5000	Comit	rendita
21	8	2500	Dow Jones	rendita
7	5	5000	Comit	rendita
23	5	2500	Nikkei	rendita
8	5	5000	Dow Jones	rendita
14	5	2500	Comit	rendita
11	5	2500	Dow Jones	rendita
20	8	5000	Nikkei	rendita
4	8	2500	Nikkei	rendita
13	8	5000	Dow Jones	rendita
3	8	2500	Comit	rendita
22	5	5000	Nikkei	rendita

Gli stimatori (*part worth*) calcolati con la procedura *Constrained Non Linear Regression* - CNLR di programmazione quadratica dell'SPSS (2002), sono riportati in Tab. 4, nella quale tra parentesi sono indicate le stime degli errori standard.

Tab. 4. Stimatori e relativi errori standard delle tre funzioni aggregate di risposta ottenuti con il modello di conjoint analysis proposto.

Categoria di *overall*	Stimatori della prima equazione		Categoria di *overall*	*Stimatori della seconda equazione*		*Categoria di overall*	*Stimatori della terza equazione*	
	$\tilde{c}_1$	0,4297 (0,0436)		$\tilde{c}_2$	0,3664 (0,0458)		$\tilde{c}_3$	0,2039 (0,0346)
	$\tilde{\delta}^{(1)}_{12}$	0,0243 (0,0316)		$\tilde{\delta}^{(2)}_{22}$	-0,0093 (0,0346)		$\tilde{\delta}^{(1)}_{32}$	-0,0150 (0,0245)
"non gradito"	$\tilde{\delta}^{(2)}_{12}$	-0,0063 (0,0316)	*"gradito"*	$\tilde{\delta}^{(3)}_{22}$	0,0036 (0,0346)	*"molto gradito"*	$\tilde{\delta}^{(2)}_{32}$	0,0028 (0,0245)
	$\tilde{\delta}^{(3)}_{12}$	-0,3321 (0,0424)		$\tilde{\delta}^{(3)}_{22}$	0,2788 (0,0469)		$\tilde{\delta}^{(3)}_{32}$	0,0533 (0,0316)
	$\tilde{\delta}^{(3)}_{13}$	-0,3396 (0,0424)		$\tilde{\delta}^{(3)}_{22}$	0,2563 (0,0469)		$\tilde{\delta}^{(3)}_{33}$	0,0833 (0,0917)
	$\tilde{\delta}^{(4)}_{12}$	0,1956 (0,0332)		$\tilde{\delta}^{(4)}_{22}$	-0,0959 (0,0374)		$\tilde{\delta}^{(4)}_{32}$	-0,00977 (0,0265)

Nelle Figg. 1$a) \div c$) sono riportate le funzioni di utilità dei quattro attributi del prodotto relative al giudizio di *overall* per le classi: "non gradito", "gradito", "molto gradito", rispettivamente.

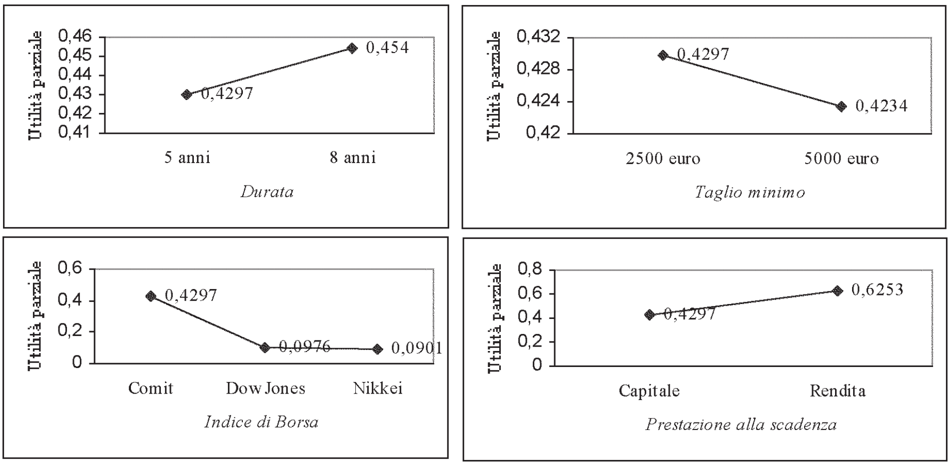

Fig. 1a: Funzioni di utilità parziale degli attributi della Polizza vita per la 1ª equazione: (categoria di *overall*: "non gradito").

I coefficienti (*part worth*) riportati in tali figure sono ottenuti dai valori di Tab. 4 cumulando per ciascun attributo gli effetti (stimatori) *relativi* o *incrementali* rispetto alla categoria di riferimento (*baseline*) $\tilde{c}_k$, con $k = 1, 2, 3$, rispettivamente.

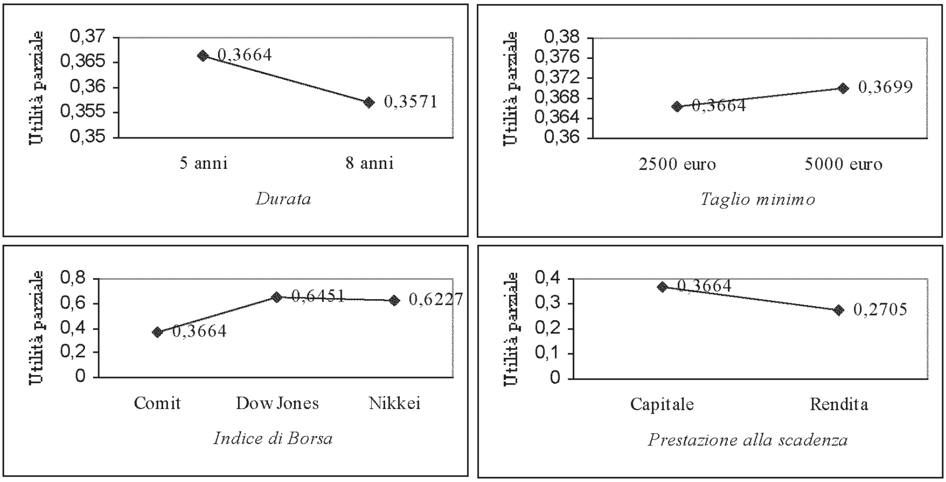

Fig. 1b: Funzioni di utilità parziale degli attributi della Polizza vita per la 2ª equazione (categoria di *overall*: "gradito").

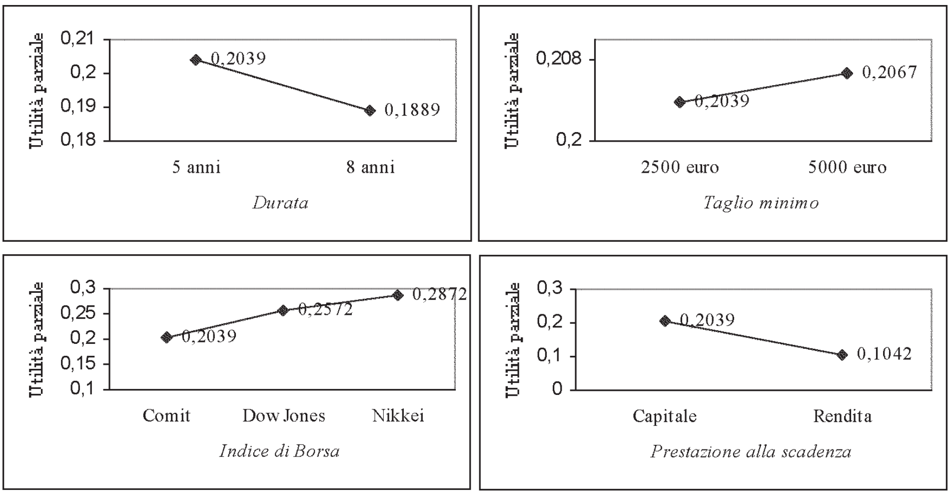

Fig. 1c: Funzioni di utilità parziale degli attributi della Polizza vita per la 3ª equazione (categoria di overall: "molto gradito").

Ad esempio, in Fig. 1*a*) il coefficiente di utilità parziale 0,454 inerente alla *durata della polizza* di “8 anni” è ottenuto dalla relazione seguente: $\tilde{c}_1 + \tilde{\delta}_{12}^{(1)} =$ 0,4297 + 0,0243 = 0,454; il coefficiente 0,4234, associato alla modalità “5.000 euro”, del fattore *taglio minimo*, è ricavato dalla relazione: $\tilde{c}_1 + \tilde{\delta}_{12}^{(2)} = 0{,}4297$-0,0063 = 0,4234.

In modo analogo sono ottenuti gli altri coefficienti riportati nelle Figg. 1*a*)÷*c*).

Dalle Figg. 1*a*)÷*c*) si rileva che le funzioni di utilità parziale relative alla 1ª equazione, associate alla categoria di valutazione globale “non gradito”, hanno andamenti di senso opposto (in termini di incremento/decremento dei coefficienti di utilità rispetto alla categoria baseline) a quelli corrispondenti alla seconda e terza equazione (associate, rispettivamente, alla categorie di giudizio “gradito” e “molto gradito”). Ciò consente – pregio dell’approccio proposto – di accertare la coerenza dei risultati del modello attraverso il confronto dei coefficienti di utilità dei livelli dei singoli fattori corrispondenti alle diverse funzioni di risposta.

4. GRADO DI BONTÀ DEL MODELLO DI CONJOINT ANALYSIS STIMATO

Con riguardo all’indice di adattamento del modello (valutato come proposto in De Luca *et alii*, 2004, per il caso di regressione lineare multipla multivariata su variabili indicatrici) si rileva che la percentuale di varianza spiegata dallo stesso è pari a 45,89%: valore da ritenere elevato se si tiene presente che trattasi di modello di regressione su variabili *dummy*.

Della bontà del modello si ha conferma empirica, del resto, osservando i risultati di Tab. 5, la quale riporta il confronto tra le frequenze *osservate* (f_{ks}) ed i corrispondenti valori *teorici* di probabilità $\left(\hat{p}_{ks}\right)$ stimati con il modello di regressione lineare multipla multivariata generalizzata vincolata. In detta tabella si può apprezzare il notevole accostamento dei valori osservati a quelli teorici e, quindi, l’elevata capacità interpolante del modello.

5. CONCLUSIONI ED INTERPRETAZIONE DEI COEFFICIENTI DEL MODELLO IN CHIAVE DI MARKETING OPERATIVO

I coefficienti (*part worth*) delle funzioni di utilità parziale (ovvero, gli “effetti” relativi), inerenti alle tre equazioni (cioè, alle tre categorie dell’*overall*) riportati in Tab. 4 consentono di individuare i livelli/fattori rilevanti del prodotto, che contribuiscono ad incrementare/decrementare i valori di $p_k\,(k = 1, 2, 3)$, ovvero,

Tab. 5. Confronto tra frequenze osservate (f_{ks}) e valori di probabilità ($\hat{p}_{ks}$) stimati con il modello proposto, per le categorie di overall: "non gradito", "gradito","molto gradito"

	Frequenze e probabilità / Numero di profilo	Categoria di *overall:* "non gradito" (Y_1)		Categoria di *overall:* "gradito" (Y_2)		Categoria di *overall:* "molto gradito" (Y_3)	
	s	f_{1s}	$\hat{p}_{ks}$	f_{2s}	$\hat{p}_{ks}$	f_{3s}	$\hat{p}_{ks}$
	1	0.45	0.43	0.45	0.37	0.10	0.20
	2	0.10	0.09	0.61	0.65	0.29	0.26
	3	0.08	0.11	0.60	0.62	0.32	0.28
	4	0.10	0.12	0.60	0.64	0.30	0.24
profilo più gradito $\rightarrow$	*5*	0.05	0.08	0.67	0.63	0.28	0.29
	6	0.52	0.45	0.34	0.36	0.14	0.19
	7	0.08	0.12	0.54	0.64	0.38	0.24
	8	0.12	0.11	0.48	0.61	0.40	0.27
	9	0.10	0.10	0.5	0.65	0.40	0.26
	10	0.50	0.42	0.43	0.37	0.07	0.21
	11	0.51	0.45	0.41	0.36	0.08	0.19
	12	0.10	0.09	0.55	0.62	0.35	0.29
	13	0.35	0.31	0.54	0.54	0.11	0.15
	14	0.31	0.31	0.54	0.52	0.15	0.17
	14	0.31	0.29	0.60	0.55	0.09	0.16
	16	0.59	0.63	0.26	0.27	0.15	0.10
	17	0.32	0.28	0.57	0.53	0.11	0.19
	18	0.27	0.29	0.62	0.55	0.11	0.16
	19	0.29	0.29	0.53	0.53	0.18	0.19
	20	0.65	0.64	0.22	0.26	0.13	0.09
profilo meno gradito $\rightarrow$	21	0.63	0.65	0.33	0.26	0.04	0.09
	22	0.34	0.30	0.52	0.52	0.14	0.18
	23	0.33	0.32	0.59	0.54	0.08	0.14
	24	0.63	0.62	0.24	0.27	0.13	0.11

la probabilità di scegliere una delle tre categorie di *overall*: "non gradito", "gradito", "molto gradito".

Più specificamente, con riferimento alla probabilità di scelta della categoria dell'overall ($Y_{k=1}$: "non gradito"), relativa alla prima equazione, i coefficienti che ne fanno incrementare il valore (Fig. 1*a)* corrispondono alle modalità: *durata 8 anni* del primo fattore, *taglio di 2.500 €* del secondo fattore, *indice Comit* del terzo fattore, *rendita vitalizia* del quarto fattore. Tutte queste modalità sono vissute, perciò, “negativamente” dai valutatori/potenziali acquirenti del nuovo prodotto, dando esse luogo ad un giudizio di valutazione globale negativo (“non gradito”).

Nella seconda e terza equazione del modello, relative a giudizi positivi di gradimento globale della polizza, si rileva una situazione di preferenza esattamente opposta a quella della prima equazione (verificabilità dei risultati del modello proposto, ovvero, assenza di discrepanza tra i risultati delle diverse funzioni aggregate di risposta).

Più specificamente, si rileva che nella seconda equazione contribuiscono ad elevare la probabilità di scelta della categoria $Y_{k=2}$ = “gradito”, con riguardo al terzo fattore (“Indice di borsa”), principalmente la seconda modalità (indice *Dow Jones*, con un incremento del coefficiente di utilità parziale del 76% rispetto a quello della classe di riferimento “Comit”) e la terza (indice *Nikkei*, con un incremento del coefficiente di utilità del 70%), oltre che la prima modalità del quarto fattore (*capitale*, con un coefficiente di utilità superiore del 35% a quello corrispondente alla classe di riferimento “rendita”).

Nella terza equazione, la probabilità di scelta della categoria $Y_{k=3}$ = “molto gradito”, si muove – in rapporto ai vari fattori – in modo analogo a quello relativo alla seconda equazione (pur se con valori dei coefficienti di utilità parziale meno elevati in valore assoluto), con la particolarità che l’indice *Nikkei* (con un coefficiente di utilità relativa pari a 0,2872) si rivela più importante rispetto all’indice *Dow Jones* (0,2572) sul grado di desiderabilità globale del prodotto; inoltre la modalità *capitale* del quarto fattore assume ancor più importanza (confrontata con quella relativa alla seconda equazione) rispetto alla modalità di riferimento “rendita”, con un incremento del coefficiente di utilità del 96%.

Alla luce di quanto sopra rilevato si ritiene che il modello ed il metodo di analisi qui proposto, nell’ambito della COA, oltre che offrire sul piano metodologico il vantaggio dell’impiego delle probabilità p_{ks} come “risposta media” per esprimere un determinato livello di gradimento (che non richiede trasformazioni di scala per rendere “metrica” la scala delle preferenze), dà la possibilità di considerare sul piano operativo – rispetto all’approccio classico, basato su *una* sola funzione aggregata di risposta – *più funzioni* aggregate di risposta, le quali consentono di svolgere una più approfondita analisi dell’entità e della direzione (in senso algebrico) degli effetti dei livelli dei singoli fattori rispetto alle classi graduate di gradimento globale del nuovo prodotto. Elementi conoscitivi, questi, di rilevante importanza per le decisioni di marketing-mix del prodotto da lanciare sul mercato.

APPENDICE

Variabili indicatrici (o *dummy*) corrispondenti alla variabile qualitativa di valutazione complessiva (*overall*) Y di un intervistato e variabili indicatrici relative agli attributi del prodotto (X_1, X_2, X_3, X_4), riferite al piano sperimentale fattoriale completo parzialmente casualizzato di S = 24 stimoli sperimentali (s = 1, 2, …, 24)

s:	Y_1	Y_2	Y_3	Z_{11}	Z_{12}	Z_{21}	Z_{22}	Z_{31}	Z_{32}	Z_{33}	Z_{41}	Z_{42}
1	0	1	0	1	0	1	0	0	0	1	1	0
16	1	0	0	1	0	0	1	1	0	0	1	0
9	0	1	0	0	1	0	1	0	1	0	1	0
18	1	0	0	0	1	0	1	1	0	0	1	0
12	0	0	1	1	0	0	1	0	1	0	1	0
19	0	1	0	0	1	1	0	0	0	1	1	0
10	0	1	0	0	1	1	0	0	0	1	1	0
24	0	1	0	0	1	1	0	0	1	0	1	0
2	1	0	0	1	0	1	0	1	0	0	1	0
15	0	1	0	1	0	0	1	0	0	1	1	0
5	0	1	0	0	1	0	1	0	0	1	1	0
17	0	0	1	1	0	1	0	0	1	0	1	0
6	1	0	0	0	1	0	1	1	0	0	0	1
21	1	0	0	0	1	1	0	0	1	0	0	1
7	1	0	0	1	0	0	1	1	0	0	0	1
23	0	0	1	1	0	1	0	0	0	1	0	1
8	0	1	0	1	0	0	1	0	1	0	0	1
14	1	0	0	1	0	1	0	1	0	0	0	1
11	1	0	0	1	0	1	0	0	1	0	0	1
20	0	1	0	0	1	0	1	0	0	1	0	1
4	1	0	0	0	1	1	0	0	0	1	0	1
13	1	0	0	0	1	0	1	0	1	0	0	1
3	1	0	0	0	1	1	0	1	0	0	0	1
22	0	0	1	1	0	0	1	0	0	1	0	1

Ringraziamenti

Si ringrazia la dott.ssa Sara Ciapparelli della società *Gfk – Marketing Service Italia*, per la collaborazione prestata nell’elaborazione del modello in ambiente SPSS.

RIFERIMENTI BIBLIOGRAFICI

DE LUCA, A. (2000), Un modello di misurazione della customer satisfaction con codifica binaria additiva dei predittori ordinali, in A. Zanella (a cura di) *Valutazione della qualità e customer satisfaction: il ruolo della Statistica*, Vita e Pensiero, Milano, pp. 275-290.

DE LUCA A. (2001), Un modello di valutazione della qualità percepita con variabile risposta politomica, in *Processi e metodi statistici di valutazione*, Convegno Intermedio della Società Italiana di Statistica, Università di Roma, 4-6 giugno, pp. 87-90.

DE LUCA A., ZANELLA A., CANTALUPPI G. (2004), Un modello di valutazione della customer satisfaction con variabile risposta politomica, *Statistica Applicata*, vol. 16, n. 4, 563-598.

DE LUCA A. (2006), *Le applicazioni dei metodi statistici alle analisi di mercato – manuale di ricerche per il marketing*, F. Angeli, Milano, 5ª ediz.

GILL P.E., MURRAY W., WRIGTH M.H. (1981), *Pratical optimization*, Academic Press, San Francisco.

GREEN P. E., TULL D. S., ALBAUM G. (1988), *Research for marketing decisions,* Fifth Ed., Prentice-Hall, New York, 1988.

GOLDBERGER A.S. (1964), *Econometric theory*, John Wiley & Sons, Inc., New York.

JOHNSTON J. (2000), *Econometrica*, F. Angeli, Milano.

KOTZ S., JOHNSON N. L. (1985), *Encyclopedia of statistical sciences*, Volume 5, J. Wiley & Sons, Nw York.

KRANTZ D. H., LUCE R.D., SUPPES P., TVERSKY A. (1971), *Foundations of measurements,* Academic Press, New York, vol. 1.

KRUSKAL J. B. (1964), Nonmetric multidimensional scaling: a numerical method, *Psychometrika,* 29, 2, pp. 113-129.

MOORE W. L. (1980), Levels of aggregation in conjoint analysis: an empirical comparison, *Journal of Marketing Research,* 17, 516-523.

SADOCCHI S. (1981), *Manuale di analisi statistica multivariata*, F. Angeli, Milano.

SPSS Inc. (2002), *Constrained Non Linear Regression- CNLR*, SPSS Inc., Chicago.

SUITS D.B. (1959), Use of dummy variables in regression equations, *Journal of the American Statistical Association,* 52, 548-551.

A NEW APPROACH TO CONJOINT ANALYSIS TO ESTIMATE MORE THAN ONE RESPONSE FUNCTION

Summary

In the Conjoint Analysis (CA) model proposed here – an extension of the traditional CA – the polytomous response variable (i.e. evaluation of the overall desirability of alternative product profiles) is described by a sequence of binary variables. To link the categories of overall evaluation to the factors levels, we adopt – at the aggregate level – a multivariate linear regression model, based on a main effects experimental design. The model provides several overall desirability functions (aggregated part-worths sets), as many as the overall ordered categories are, unlike the traditional metric and non metric CA, which gives only one response function. We provide an application of the model and an interpretation of the main interactive effects.

Keywords: *Aggregate level analyses, Conjoint analysis, Multivariate regression, Ordinal response.*