***S**tatistica & **A**pplicazioni*
Vol. III, n. speciale 1, 2005
pp. 117-124

Un modello di regressione fuzzy per la valutazione della soddisfazione

Rosaria Romano[§]

Francesco Palumbo[‡]

Summary: *Customer Satisfaction Analysis is mostly based on the study of the deviations between customer expectations and customer perceptions on the quality of the product/service. According to this definition, the paper considers the CS in its own nature of interval value and proposes to apply fuzzy regression models for the CS estimation. The main assumption of the model is that the CS depends essentially on the perceived and expected satisfaction levels. The model capabilities are tested on a real dataset demonstrating how the fuzzy approach is promising in the CS analysis.*

Keywords: *customer satisfaction, codifica fuzzy, regressione fuzzy.*

1. Introduzione

Aziende e organizzazioni economiche, in generale, sono sempre più consapevoli che la "conoscenza" della propria clientela è una esigenza primaria per sopravvivere in un mercato sempre più concorrenziale. Il giudizio del cliente sulla qualità del prodotto/servizio garantisce all'azienda un'indicazione di quelle che sono le esigenze del mercato. Inoltre, il cliente soddisfatto è sicuramente più propenso al riacquisto e diventa anche promotore dei prodotti aziendali; laddove, un cliente insoddisfatto tenderebbe, invece, a danneggiare l'immagine dell'azienda.

La *Customer satisfaction* (CS), la cui traduzione nella lingua italiana (sebbene poco utilizzata) è *soddisfazione del cliente*, è una grandezza la cui definizione e misurazione è materia di studio ampia e controversa. Ancora più complessa è la misurazione della CS per i servizi, in quanto, trattandosi di beni intangibili, risulta difficile definire specifiche di produzione in base alle quali

[§] Dipartimento di Matematica e Statistica – Università degli Studi di Napoli "Federico II" – via Cintia, 80126 NAPOLI (e-mail: rosaroma@unina.it).
[‡] Dipartimento di Istituzioni Economiche e Finanziarie – Università di Macerata – via Crescimbeni, 20, 62100 MACERATA (e-mail: francesco.palumbo@unimc.it).

valutare il livello di qualità effettivo. Altre caratteristiche (eterogeneità, contestualità, interattività) contribuiscono a rendere più complessa la valutazione della qualità dei servizi rispetto a quella dei prodotti (Lewis e Booms, 1983). In tale contesto risulta difficile giungere ad una definizione univoca della CS. Si riscontrano, piuttosto, vari approcci fra i quali è possibile identificare due particolari orientamenti: l'uno mira a una definizione della misura in senso assoluto; l'altro, invece, si fonda sul convincimento che la soddisfazione si determina rispetto alle aspettative dei clienti.

L'altro aspetto rilevante riguarda l'effettiva possibilità di misurare (in senso assoluto o in senso relativo) il grado di soddisfazione dei clienti. Trova sempre maggiori consensi la teoria in base alla quale la soddisfazione è una grandezza che non può essere direttamente misurata (Fornell, 1992) se non come risultante di altre variabili manifeste. L'impossibilità di misurare direttamente una grandezza non è certamente una prerogativa delle scienze economiche e sociali. In fisica, per esempio, la *velocità* è un quantità che, pur avendo una sua precisa ed univoca definizione, non può essere direttamente misurata. La velocità viene determinata attraverso il rapporto spazio/tempo.

Per la valutazione della *soddisfazione* in termini assoluti, piuttosto che della sua diretta misurazione, sono stati proposti in letteratura numerosi modelli statistici dove la soddisfazione è la *dimensione* latente (Tenenhaus *et al.* 2005). Secondo tali modelli, definiti *modelli causali*, la misurazione della CS richiede un'analisi approfondita delle relazioni strutturali tra variabili latenti, relative alla soddisfazione del cliente, e variabili manifeste, che esprimono il livello di soddisfazione del cliente su specifici attributi del servizio, espresso mediante scale di punteggi. La stima del modello può seguire vari approcci tra cui l'approccio PLS (Partial Least Squares, Wold 1982), che stima le variabili latenti attraverso un sistema ricorsivo basato sul metodo dei Minimi Quadrati, e l'approccio LISREL (LInear Structural RELationship, Jöreskog 1970), basato sul metodo di Massima Verosimiglianza. I due approcci seguono specifici obiettivi: il PLS è orientato alla predizione delle variabili manifeste e latenti, il LISREL alla stima dei parametri. Il PLS è, pertanto, più indicato per misurare la CS, essendo il LISREL essenzialmente incentrato a confermare la teoria del processo decisionale del consumatore.

Altri studiosi hanno concentrato la propria attenzione sulla definizione di soddisfazione come una misura relativa. In questa ottica sono stati proposti modelli di analisi che valutano la soddisfazione attraverso gli scostamenti fra il livello di qualità percepito e quello atteso (Zeithaml *et al.*, 1991). Il sistema ServQual, ad esempio, è uno strumento di indagine rigido e schematico composto di due sezioni, aspettative e percezioni, che fornisce una misura relativa della CS. Ogni sezione è composta di una serie di proposizioni relative a ciascuna delle 5 dimensioni della qualità del servizio (aspetti tangibili, affidabilità, capacità di risposta, capacità di rassicurazione, empatia). Gli scostamenti tra queste due grandezze forniscono una misura della soddisfazione del cliente e, quindi, della qualità del prodotto/servizio.

I modelli causali e l'approccio ServQual presentano indubbiamente notevoli capacità di analisi che ne giustificano l'ampia diffusione. La struttura rigida del ServQual, se da un lato non soddisfa le diverse esigenze di analisi dall'altro ne garantisce la comparabilità. I modelli causali, invece, seppure più flessibili, forniscono una misurazione della CS esclusivamente in termini assoluti.

In un'ottica assolutamente innovativa, basandosi su una particolare codifica, Lauro *et al*. (2001), prima, Amato e Palumbo (2004) e Grassia *et al*. (2004), in un secondo momento, propongono di codificare i valori connessi alla percezione ed alle aspettative in un'unica struttura numerica: i dati ad intervallo. Si tenga presente che i dati ad intervallo sono anche una particolare tipologia di dati *fuzzy*.

Obiettivo del presente lavoro è quello di valutare la possibilità di adattare modelli di regressione per dati *fuzzy*, già proposti in letteratura, alla codifica di Amato e Palumbo. L'ipotesi su cui si fonda la proposta è che la qualità percepita e la qualità attesa del servizio sono fattori latenti responsabili della determinazione del grado di soddisfazione del consumatore. In virtù di ciò, il modello proposto ha come variabile dipendente la "soddisfazione", codificata come un dato ad intervallo, e come variabili esplicative i cosiddetti *driver* della soddisfazione. Nella proposta, tali driver sono ottenuti a partire da gruppi di variabili manifeste attraverso combinazioni di tipo lineare, come normalmente si fa per i modelli causali.

Il lavoro si articola in base allo schema seguente: nel paragrafo successivo vengono esposti i concetti di base della *teoria fuzzy*; nel terzo paragrafo è descritto un modello di regressione per dati *fuzzy* e, infine, un'applicazione del modello è presentata e discussa nel quarto paragrafo.

2. Notazioni e definizioni

La scelta di voler codificare le variabili manifeste e/o latenti sotto forma di deviazioni fra attese e percezioni rende necessaria l'introduzione dei concetti e del formalismo essenziale della *teoria fuzzy*.

Definiamo un *intervallo* come un insieme chiuso e compatto che indicheremo con la notazione $[x]$; dove $[x]$ si riferisce ad una coppia ordinata di valori $[x] \equiv [\underline{x}, \overline{x}]$, con $\underline{x} \leq \overline{x}$. Si osservi che un dato a intervallo può essere rappresentato anche in termini di centro $c = (\overline{x} + \underline{x})/2$ e raggio $r = (\overline{x} - \underline{x})/2$.

Definito $X=\{x_i\}$ (i=1,…,n) un *universo del discorso*, ovvero l'insieme di tutti i possibili elementi relativi ad un concetto, un insieme *fuzzy* A in un universo del discorso $X=\{x_i\}$, con $A \subseteq X$, è definito dalla coppia $\{\mu_A(x_i), x_i\}$, $x_i \in X$, dove μ_A: $X \rightarrow [0,1]$ è la funzione di appartenenza (o funzione di possibilità) di A e $\mu_A(x_i) = h$ è definito grado di appartenenza di x_i in A.

Quindi h = 0 vuol dire che $x_i \notin A$, h = 1 vuol dire che $x_i \in A$, mentre $0 < h < 1$ indica che x_i appartiene parzialmente o in modo *fuzzy* ad A (Zadeh, 1965).

Un *numero fuzzy*, in particolare, è definito come un insieme *fuzzy* $A \subset \mathbb{R}$ con le seguenti proprietà:

a) $\mu_A(x_i) = 1$, per almeno un $x_i \in \mathbb{R}$;

b) $\mu_A[\lambda x_1 + (1-\lambda)x_2] \geq \mu_A(x_1) \wedge \mu_A(x_2)$, per x_1 e $x_2 \in \mathbb{R}$ e $\lambda \in [0,1]$.

I numeri *fuzzy* maggiormente utilizzati, per la semplicità della codifica, sono i numeri *fuzzy* simmetrici. Un numero *fuzzy* simmetrico è generalmente definito con la notazione $A = (m, \alpha)_L$ dove m ed α sono, rispettivamente, il *centro* e lo *spread* del numero *fuzzy* A con funzione di appartenenza $L(x)$. La seguente funzione di possibilità definisce alcuni dei più diffusi numeri *fuzzy* simmetrici:

$$L(x; m, \alpha) = 1 - \left| \frac{x - m}{\alpha} \right|^q \tag{1}$$

Al variare di q la funzione di appartenenza si modifica: per $q = 1$ si ottiene la funzione *triangolare*, per $q = 2$ la funzione *parabolica*. Si osservi che, se consideriamo $q = \infty$, essendo $|x - m|/\alpha < 1$ per $m - \alpha \leq x \leq m + \alpha$, la quantità in (1) è sempre uguale ad 1; questa considerazione ci consente di connotare i dati ad intervallo come una particolare tipologia di numero *fuzzy*.

Partendo dalla proposta di Lauro *et al.* (2001), nel presente lavoro, la soddisfazione verrà codificata come un dato ad intervallo, i cui estremi sono funzioni delle attese e delle percezioni. Nella proposta originaria l'estremo inferiore è definito dalla soddisfazione attesa (A), mentre l'estremo superiore dalla soddisfazione percepita (P). Il segmento che unisce i due estremi è definito *intervallo di soddisfazione.* Il cliente è: soddisfatto se le percezioni sono maggiori delle attese $(P - A) > 0$; insoddisfatto in caso contrario $(P - A) < 0$ e indifferente se percezioni e attese coincidono $P = A$. Se le percezioni sono inferiori alle attese, per garantire il rispetto della condizione $\underline{x} \leq \overline{x}$, Lauro *et al.* (2001) propongono di moltiplicare gli estremi dell'intervallo per – 1. Tale trasformazione si è rilevata molto utile nell'approccio esplorativo di Grassia *et al.* (2004) permettendo di discriminare facilmente su piani fattoriali i soggetti (prevalentemente) soddisfatti, rispetto agli insoddisfatti.

Nell'ottica dei modelli causali, la codifica di Lauro *et al.* (2001), come messo in evidenza in Amato e Palumbo (2004), implica che il cambiamento di segno altera il livello assoluto di soddisfazione solo degli utenti insoddisfatti, introducendo così una variabilità non effettivamente presente nei dati.

Amato e Palumbo (2004) considerano la soddisfazione percepita come centro dell'intervallo, mentre il raggio è funzione dell'intervallo di soddisfazione e del massimo scostamento osservabile. In particolare, il raggio è dato da:

$$r = |A - P| / \max[L_u - A; A - L_l] \tag{2}$$

dove A è l'attesa, P la percezione ed L_l e L_u sono rispettivamente il limite inferiore e superiore della scala di preferenze. Così facendo si perde il segno dello scostamento, che però può essere recuperato in fase di rappresentazione dei risultati attraverso un'opportuna ricostruzione degli intervalli.

Il presente contributo partendo dalla codifica di Amato e Palumbo (2004), in cui gli autori utilizzano numeri *fuzzy* rettangolari ($q = \infty$), propone di codificare la CS come numero fuzzy triangolare simmetrico, ovvero, assumendo ($q = 1$).

3. Il modello di regressione fuzzy per la valutazione della soddisfazione

In questo paragrafo viene introdotto un modello per valutare la CS codificata come numero *fuzzy*. Diversi modelli per dati *fuzzy* sono stati proposti in letteratura. In questo lavoro si è scelto di utilizzare il modello di Tanaka e Watada (1987) le cui caratteristiche bene si adattano alle esigenze di un'analisi di CS (numero di osservazioni limitato e inferiore al numero di variabili, multicollinearità). Il modello, infatti, non si basa su ipotesi restrittive di alcun tipo se non quella della linearità della relazione tra le variabili. In particolare, si considera il modello in cui la variabile di risposta $Y_i=\{y_i, e_i\}$ (i=1,…,n) è un numero *fuzzy* triangolare simmetrico con centro e raggio rispettivamente pari a y_i ed e_i. Il modello è definito dalla seguente funzione lineare fuzzy:

$$\tilde{Y} = \boldsymbol{X}\tilde{B} \tag{3}$$

dove la generica riga della matrice $X_i = [X_{i0},\dots, X_{ij},\dots, X_{ip}]$ ($i = 1,\dots,n$) è un vettore di variabili indipendenti puntuali e $\tilde{B} = [\tilde{B}_o, \dots, \tilde{B}_j, \dots, \tilde{B}_p]^T$ è un vettore di coefficienti, con $\tilde{B}_j = \{c_j, a_j\}$ numero *fuzzy* triangolare simmetrico. Le stime derivano dalla risoluzione di un *problema di programmazione lineare* in cui la funzione obiettivo è minimizzare l'incertezza delle stime, che equivale a minimizzare l'ampiezza dei coefficienti $\tilde{B}_j$, con il vincolo che il *fuzzy set* delle stime includa il *fuzzy set* dei dati osservati, scelto un livello h:

$$\textit{minimizzare} \quad \sum_{j=0}^{p}\left(\sum_{i=1}^{n} a_j \left|X_{ij}\right|\right) \tag{4}$$

$$\textit{con vincoli:} \quad \sum_{j=0}^{p} c_j X_{ij} + (1-h)\sum_{j=0}^{p} a_j \left|X_{ij}\right| \geq y_i + (1-h)e_i, \forall i = 1,\dots,n$$

$$\sum_{j=0}^{p} c_j X_{ij} - (1-h)\sum_{j=0}^{p} a_j \left|X_{ij}\right| \leq y_i - (1-h)e_i, \forall i = 1,\dots,n$$

$$a_j \geq 0,\ c_j \in \mathrm{R},\ X_{i0} = 1,\ 0 \leq h \leq 1, i = 1,\dots,n;\ j = 1,\dots, p.$$

Il coefficiente h è una misura del grado di “possibilità” delle stime ed assume un ruolo assimilabile a quello del livello di confidenza nell’inferenza classica: all’aumentare del livello di possibilità aumenta l’ampiezza dell'intervallo dei coefficienti così come all’aumentare del livello di confidenza aumenta l’ampiezza dell'intervallo delle stime.

4. Il caso studio

I dati elaborati si riferiscono ad un’indagine effettuata per conto di una società di servizi per le Pubbliche Amministrazioni. Per valutare la soddisfazione, agli utenti è stato somministrato un questionario attraverso il quale sono stati rilevati i giudizi su 17 attributi del servizio erogato, espressi contestualmente su scala [1, 10]. L’indagine ha interessato un campione di 50 unità. Per vincoli di riservatezza non tutti i termini dell’indagine possono essere resi pubblici. La codifica del *gap* come valore a *intervallo*, ha permesso di ridurre la complessità del modello di CS; ciò è possibile, in quanto la complessità dell'informazione viene tradotta nella codifica stessa.

Il primo passo dell’analisi ha riguardato l’identificazione delle variabili “*qualità attesa”* e “*qualità percepita”*, quali fattori latenti delle rispettive variabili manifeste. Un’Analisi in Componenti Principali, rispettivamente sui due blocchi di variabili manifeste, ha messo in luce la natura unidimensionale dei blocchi stessi, pertanto, si sceglie la prima Componente Principale come unica variabile di sintesi dei rispettivi blocchi. La varianza spiegata dalla prima componente principale per ciascun blocco di variabili è rispettivamente 85,2% e 50,6%. La fase successiva dell’analisi ha riguardato la costruzione della variabile di risposta ad intervallo, la *soddisfazione globale*, in base alla espressione (2). Sulla base dei giudizi espressi in merito alla *soddisfazione rispetto all’ideale* e alla *soddisfazione rispetto alle attese* è stata costruita la variabile ad intervallo *gap di soddisfazione*.

Il modello stimato, al livello di possibilità $h = 0,5$ è il seguente:

$$Y = \{0,59;\ 2,47\} + \{0,23;\ 0,10\}\ X_1 + \{0,60;\ 0,05\}\ X_2 \tag{5}$$

Le variabili X_1 e X_2 sono, rispettivamente, la componente principale *qualità attesa* e la componente principale *qualità percepita*; i relativi coefficienti sono espressi utilizzando la consueta notazione in cui il primo e il secondo valore sono, rispettivamente, i centri e i raggi degli intervalli. Dall'analisi dei valori relativi ai centri dei coefficienti emerge che vi è una maggiore dipendenza del *gap* di soddisfazione rispetto alla componente *qualità percepita* piuttosto che non rispetto alle attese. La vaghezza del sistema, riflessa dall’ ampiezza dei raggi è, invece, maggiormente dovuta alla componente *qualità attesa*, a sottolineare l’incertezza dell'utente nell’esprimere un giudizio su un attributo aleatorio del servizio. Tale risultato è assolutamente coerente con l'impostazione del modello. Infatti, è più plausibile ritenere che la soddisfazione sia maggiormente influenzata dalla percezione piuttosto che dall'attesa. È altresì ragionevole ritenere che l'imprecisione dei coefficienti sia determinata dalle aspettative degli utenti.

I risultati del modello sono visualizzati nei seguenti grafici in cui sono riportati i *gap* di soddisfazione osservati e gli intervalli di regressione.

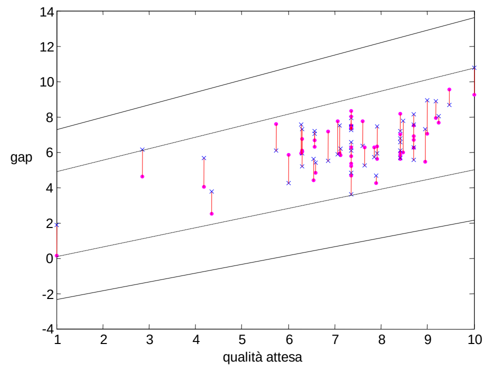

Figura 1. *Intervalli di regressione rispetto alla variabile indipendente "qualità attesa"*

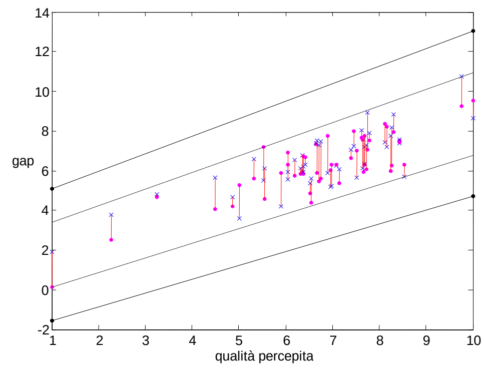

Figura 2. *Intervalli di regressione rispetto alla variabile indipendente "qualità percepita"*

È utile ricordare che la codifica adottata non consente di tenere conto del segno dell'intervallo di soddisfazione. Tuttavia ricorrendo ai dati originari è possibile recuperare tale informazione. Nei grafici si è utilizzato un diverso simbolo per indicare l'attesa '•' e la percezione '×'; in tal modo l'orientamento dell'intervallo permette di valutarne il segno. Infine, le rette continue si riferiscono alle stime al livello di "possibilità" h = 0,5 e le linee tratteggiate al livello h = 0.

5. Conclusioni

I risultati presentati mettono in evidenza le potenzialità dell'approccio. In particolare, la chiarezza della rappresentazione grafica consente di percepire immediatamente sia l'intensità della relazione fra i driver della soddisfazione e la soddisfazione, (guardando all'inclinazione del fascio di rette di regressione), sia, il grado di soddisfazione/insoddisfazione in termini relativi (guardando alle ampiezze e al verso degli intervalli).

La letteratura specialistica per l'analisi statistica di dati *fuzzy* offre numerosi e più articolati modelli rispetto al modello di Tanaka e Watada (1987) qui presentato. Si ritiene, pertanto, che nelle ricerche future, l'interesse debba essere rivolto alla verifica dell'applicabilità di nuovi modelli. Inoltre, in un'ottica di *modelli causali*, tale approccio può essere parte integrante di modelli con una struttura di relazioni più complessa.

Riferimenti Bibliografici

Amato S., Palumbo F. (2004). Multidimensional Gap Analysis. *Statistica Applicata*, **16**(3), 265-283.

Fornell C. (1992). A National Customer Satisfaction Barometer: The Swedish Experience. *Journal of Marketing*, **56**, 6-21.

Grassia G., Lauro N.C., Scepi G. (2004). L'analisi dei dati ad intervallo nell'ambito della qualità. *Data Mining e Analisi Simbolica.* Franco Angeli.

Jöreskog K.G. (1970). A general method for analysis of covariance structure. *Biometrika* **57**, 239-251.

Lauro N.C., Esposito Vinzi V., Scepi G. (2001). Visualizzazione e sintesi della Customer Satisfaction in termini di Analisi dei dati Simbolici. *Atti del Convegno Intermedio SIS: Processi e metodi statistici di valutazione*, Roma.

Lewis R.C., Booms B.H. (1983). The Marketing Aspects of Quality, in *Emerging Perspectives on Service Marketing,* a cura di Berry L. et al., 99-107.

Tanaka H., Watada J. (1987). Possibilistic linear systems and their application to the linear regression model. *Fuzzy Sets and Systems*, **27**, 275-289.

Tenenhaus M., Esposito Vinzi V., Chatelin Y., Lauro N.C. (2005). PLS path modelling. *Computational Statistics & Data Analysis*, **48**, 159-205.

Wold H. (1982). *Partial Least Squares*, in Kotz & Johnson (eds.), Encyclopaedia of Statistical Sciences, John Wiley & Sons, New York, 581-591.

Zadeh L. (1965). Fuzzy sets. *Information and Control,* **8**, 338-353.

Zeithaml V., Parasuraman A., Berry L. (1991). *Servire Qualità*. McGraw-Hill.