

Alcune Considerazioni sull'Analisi Multidimensionale dei dati nella Valutazione dei Servizi di Day Surgery

Luigi D'Ambra

Dipartimento di Matematica e Statistica.
Università di Napoli Federico II. e-mail: dambra@unina.it

Pietro Amenta

Dipartimento di Scienze Economiche, Matematico Statistiche, Economico Aziendale, Ambientale. Università di Lecce. e-mail: amenta@economia.unile.it

Michele Gallo

Dipartimento di Matematica e Statistica.
Università di Napoli Federico II. e-mail: micgallo@unina.it

Abstract: The evaluation of the quality on the Day Surgery service is an actual problem. The conceptual models proposed in literature allow to collected data of the different nature: ordinal, compositional, and so on. The multidimensional analysis of these data gives more problems and difficulties. Aim of this paper is to think about the mentioned problems as well as to suggest suitable multidimensional methods. Finally, the appropriateness of the inclusion the external information in a customers satisfaction framework is given.

Keywords: Valutazione dei servizi di day surgery, Analisi multidimensionale dei dati, dati ordinali e composizionali, Analisi non lineare dei dati, Informazioni esterne.

1. Introduzione

Nell'ambito dei servizi sanitari, assumono un ruolo sempre più importante le pratiche del Day Surgery. Tali servizi, come quelli di ricovero ordinario, sono diagnostici e/o terapeutici, tuttavia, si differenziano non tanto per le procedure tecniche seguite, quanto per l'erogazione in regime di ricovero breve: esso dura al massimo una giornata. La peculiarità di tali servizi, molto più che per quelli erogati in regime di ricovero ordinario, li rende per molti aspetti comparabili agli altri servizi pubblici e privati. Il management delle aziende sanitarie, quindi, deve calibrare la propria strategia aziendale sull'obiettivo di massimizzare la soddisfazione degli utenti/clienti.

Lo studio di tali problematiche ha portato alla formulazione di diversi modelli

concettuali per la valutazione della soddisfazione tanto di chi usufruisce direttamente del servizio (*Paziente*) quanto di chi a diverso titolo ne usufruisce indirettamente (*Istituzioni, Collettività, etc.*). Tuttavia, bisogna osservare che tali modelli d'ascolto permettono di raccogliere informazioni solo guardando il servizio dalla prospettiva di chi lo riceve, ma gli utenti, spesso, non conoscono né i problemi tecnologici né il livello di risorse adoperate dall'azienda erogante. Inoltre, l'opinione degli utenti è per sua natura soggettiva e, quindi, può essere non confermata dall'oggettiva del processo d'erogazione (tempo medio d'attesa, numero d'interventi ...). Per avere una visione completa sul livello qualitativo dei servizi è insufficiente raccogliere le opinioni degli utenti/clienti, un buon sistema di monitoraggio deve caratterizzarsi per la presenza sia d'indicatori oggettivi che di quelli soggettivi. I primi rappresentano una misura oggettiva della qualità dei servizi. I secondi, invece, danno una "misurazione" della soddisfazione degli utenti, la quale è una valutazione condizionata psicologicamente poiché riflette sia elementi cognitivi che emozionali (Oliver, 1993). Nel caso di una valutazione negativa degli utenti/consumatori e di una buona qualità del processo d'erogazione del servizio, evidenziata da indicatori oggettivi, dovrebbe portare il management ad intraprendere un'azione migliorativa soprattutto in termini di comunicazione/informazione (i fattori che probabilmente influiscono sulla insoddisfazione degli utenti/clienti sarebbero, in tal caso, quelli cognitivi ed emozionali e non di processo).

Senza perdere di vista la parzialità della misurazione basata sui soli indicatori soggettivi, l'insoddisfazione dell'utente rappresenta il problema principale del management aziendale e questo indipendentemente se imputabile alla bassa qualità del servizio erogato o all'influenza degli elementi cognitivi emozionali sulla valutazione. Del resto, il livello qualitativo percepito dal utente/cliente è stato definito da Zanella (1999) "livello qualitativo reale del servizio", mentre le caratteristiche oggettive possono essere relegate al mero ruolo di "qualità virtuale".

Nel presente lavoro, per valutare la qualità dei servizi, si approfondiranno essenzialmente le problematiche riguardanti l'analisi dei dati di customer satisfaction (CS) e, tra i tanti modelli proposti in letteratura, sono stati considerati quelli per la valutazione indiretta del livello di soddisfazione (Parasuraman, Zeithaml e Berry, 1991; Cronin e Taylor, 1992). In particolare, sia il modello Servqual Revisionato che Servperf sono costituiti da 22 items raggruppati a loro volta in 5 dimensioni, tuttavia, col primo si raccolgono dati sulle aspettative *X*, sulle percezioni *Y*, e sull'importanza attribuita alle diverse dimensioni *C*, mentre col secondo i dati sulle aspettative non sono raccolti. La differenza essenziale tra i due modelli è che per il Servqual la "overall satisfaction" è data dalla somma degli scostamenti tra percezione ed aspettative, eventualmente ponderate con l'importanza attribuita alle singole dimensioni; mentre per il Servperf la "overall satisfaction" è data semplicemente dalla somma di tutte le aspettative, anche in questo caso eventualmente ponderate per l'importanza attribuita alle dimensioni. Seguendo tali modelli concettuali e, tenuto conto dei molteplici aspetti che concorrono alla formazione di un giudizio, è indispensabile utilizzare tecniche di

analisi statistica che rispettino la natura della grandezza da valutare. In effetti, Parasuraman, Zeithaml e Berry (1991) definiscono la qualità una grandezza multidimensionale, quindi, raggruppare le valutazioni espresse dai diversi utenti/clienti è un processo complicato che richiede l'utilizzo di tecniche statistiche multidimensionali.

Tali tecniche, in genere, sono sviluppate associando ad un concetto (nel nostro caso di soddisfazione) un "costrutto concettuale" e, al contempo, sono stati definiti dei processi di misurazione atti a collegarlo a variabili osservabili o indicatori (es. "latent trait models"). Nell'utilizzo delle tecniche bisogna considerare, inoltre, la natura dei dati. Le valutazioni sulle percezioni ed eventualmente sulle aspettative sono espresse mediante delle scale di misura ordinale, conseguentemente, senza una preventiva quantificazione, anche il calcolo della sola media è un'operazione contestabile. L'importanza attribuita alle dimensioni, inoltre, deve essere considerata come un dato di tipo compositazionale: ogni utente/cliente è vincolato ad attribuire alla somma delle 5 dimensioni un peso pari ad 1.

Obiettivo del presente lavoro è di analizzare in modo critico le tecniche statistiche multidimensionale nella valutazione dei servizi di Day Surgery, suggerendo anche approcci metodologici alternativi che tengono conto anche di eventuali informazioni aggiuntive disponibili sui fruitori dei servizi e sulle procedure o risorse utilizzate dalla struttura nella fase di erogazione del servizio. Nella figura 1 sono sintetizzate le diverse strategie utilizzabili per risolvere i problemi connessi all'analisi multidimensionale dei dati ordinali, compositazionali e misti.

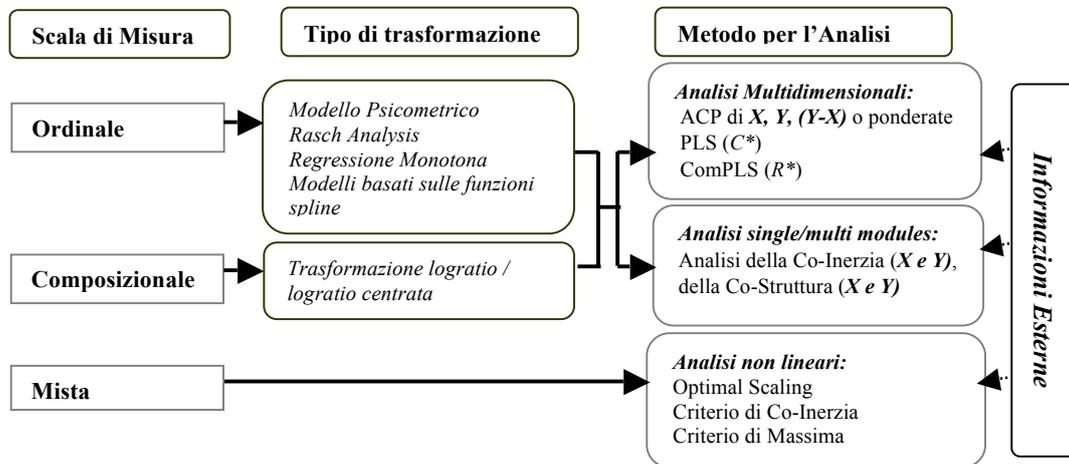


Fig. 1: Il Pretrattamento dei dati e tecniche statistiche multidimensionale.

In particolare, nel secondo paragrafo, sono richiamate le tecniche per la quantificazione dei dati ordinali. Le principali tecniche multidimensionali di co-inerzia e co-struttura per i dati sulla valutazione dei servizi di Day Surgery sono illustrate nel paragrafo 3. Le problematiche inerenti l'analisi dei dati compositazionali e come includere le informazioni esterne nell'analisi dei dati di CS sono date nei paragrafi 4 e 5. Nelle conclusioni, infine, sono descritti i possibili

sviluppi successivi.

2. Quantificazione dei dati ordinali

L'analisi dei dati ordinali comporta il problema della loro non confrontabilità "semantic differential" (Green e Tull, 1988). Questo aspetto, frequentemente, è trascurato nella applicazione delle tecniche multidimensionale: i dati ordinali sono spesso analizzati alla stregua di quelli quantitativi. Per procedere ad un'analisi che rispetti la natura del dato ordinale è necessaria una loro trasformazione in misure lineari e quantitative "calibrate" lungo l'intero arco dei numeri reali, inoltre, tale trasformazione deve evitare un "effetto compressione" (Wright e Linacre, 1989). Diverse sono le tecniche che consentono di attenuare le limitazioni all'elaborazione statistica dei dati ordinali, in particolare, quelle più utilizzate sono l'approccio "psicometrico" di Thurstone (Jones, 1986; Zanella 1999), il "Rating Scale Models" della Rasch Analysis (Wright, Masters, 1982), la regressione monotona di Kruskal (1965).

Un'ulteriore tecnica di quantificazione segue il principio che ogni individuo ha una propria soggettiva scala di valutazione, conseguentemente si verifica la non ricercata situazione che la qualità di un servizio è valutata utilizzando di volta in volta un diverso "misurando": tanti quanti sono gli utenti/clienti. È chiaro che procedendo ad una quantificazione dei dati originali, sarebbe preferibile eliminare dalla valutazione del servizio le differenze imputabili al "misurando". Gallo (2002) ha proposto una nuova procedura, la quale determina una quantificazione dei dati ordinali eliminando tali differenze. Il criterio adottato è quello della minimizzazione della funzione di perdita $n^{-1} \sum_{i=1}^n tr(\mathbf{A} - \mathbf{B}_i \zeta_i \varphi_i') (\mathbf{A} - \mathbf{B}_i \zeta_i \varphi_i')$, con i vincoli $\hat{\mathbf{1}}_s \mathbf{A} = \mathbf{0}_p$ and $\mathbf{A}' \mathbf{A} = p \mathbf{I}$ e $\Psi_i = \zeta_i \varphi_i'$, dove $\hat{\mathbf{1}}_s$ è un vettore unitario di dimensione s , $\mathbf{0}_p$ è un vettore nullo di dimensione p , \mathbf{A} è una matrice di dimensione (p, s) , φ_i' è un vettore score, ζ_i è il vettore delle quantificazioni dalle singole categorie per ogni singolo soggetto, \mathbf{B}_i è la matrice B-spline ottenuta dalle valutazioni espresse dall'individuo *i-esimo* sui diversi items. La quantificazione dei giudizi espressi dall'*i-esimo* utente/cliente è data da $\mathbf{B}_i \zeta_i$. Facendo uso dell'algorithm Alternative Least Squares (ALS), la procedura permette di eliminare la diversa propensione degli utenti/clienti a preferire alcune modalità di risposta e questo indipendentemente dalla valutazione espressa per ogni specifico item.

La caratteristica che accomuna le diverse tecniche di quantificazione dei dati ordinali è l'utilizzo della frequenza con cui sono scelte le diverse modalità della scala ordinale. Tali frequenze sono utilizzate per poter passare da una scala di misura più povera (scala ordinale) ad una più informativa (scala ad intervalli) grazie appunto alle informazioni contenute nella frequenza con le quali le diverse modalità sono scelte dagli utenti/clienti.

3. Analisi statistica multidimensionale della qualità dei servizi attesi e percepiti

Molti metodi per la valutazione della qualità dei servizi sanitari, concentrano la loro attenzione sulla qualità percepita o sul “gap” tra qualità attesa e percepita, utilizzando gli strumenti dell’analisi multidimensionale dei dati. Tuttavia, in tal modo non si considera come la valutazione espressa sulla prima influenzi la seconda. La stessa analisi della differenza può essere inquadrata in un ambito più generale, in quanto componente dell’analisi della co-struttura di tabelle totalmente sovrapponibili di Torre e Chessel (1995). Infatti, le matrici \mathbf{X} ed \mathbf{Y} associate alla quantificazione dei dati di soddisfazione sulla qualità attesa e percepita, sono considerate “totalmente sovrapponibili”, in quanto descrivono gli stessi individui osservati sulle stesse variabili risultando idonee per l’analisi di co-struttura. Tale approccio può essere considerato un caso particolare dell’analisi della co-inerzia (Chessel e Mercier 1993). Infatti, siano $(\mathbf{X}_{n,p}, \mathbf{Q}_X, \mathbf{D}_n)$ e $(\mathbf{Y}_{n,q}, \mathbf{Q}_Y, \mathbf{D}_n)$ due triplete statistiche sovrapponibili (stessi individui con diverse variabili), \mathbf{D}_n matrice diagonale dei pesi e \mathbf{Q}_X (p, p), \mathbf{Q}_Y (q, q) matrici definite positive. L’analisi di co-inerzia, prendendo le mosse dall’Inter-Battery Analysis (Tucker, 1958) il cui criterio è alla base anche del PLS, ricerca gli assi di massima covarianza (*co-inerzia*) fra le coordinate delle proiezioni delle righe delle due matrici: $\mathbf{XQ}_X\mathbf{a}$ e $\mathbf{YQ}_Y\mathbf{b}$ con \mathbf{a} e \mathbf{b} vettori di coefficienti. Tale analisi massimizza, in modo sequenziale $cov(\mathbf{XQ}_X\mathbf{a}, \mathbf{YQ}_Y\mathbf{b})$, con i vincoli $\|\mathbf{a}\|_{\mathbf{Q}_X} = \|\mathbf{b}\|_{\mathbf{Q}_Y} = 1$; criterio esteso al caso K matrici (D’Ambra, Sabatier e Amenta, 1998). Le soluzioni sono gli autovettori associati allo stesso autovalore λ delle matrici $\mathbf{Y}'\mathbf{D}_n\mathbf{XQ}_X\mathbf{X}'\mathbf{D}_n\mathbf{YQ}_Y$ ed $\mathbf{X}'\mathbf{D}_n\mathbf{YQ}_Y\mathbf{Y}'\mathbf{D}_n\mathbf{XQ}_X$. L’analisi della co-struttura è equivalente all’ACP di $\mathbf{Y}'\mathbf{D}_n\mathbf{X}$. Partendo dallo stesso criterio, l’analisi di co-struttura ricerca delle variabili di sintesi che tengono conto del legame tra le matrici e dell’inerzia delle stesse. Siano $(\mathbf{X}, \mathbf{Q}, \mathbf{D}_n)$ e $(\mathbf{Y}, \mathbf{Q}, \mathbf{D}_n)$ due triplete totalmente sovrapponibili che ci autorizzano a considerare le unità statistiche delle tabelle come elementi di uno stesso spazio, con un asse di co-inerzia comune. Si ricercano, allora, gli assi di massima covarianza (co-struttura): $\max_{\mathbf{a}} \langle \mathbf{XQa}, \mathbf{YQa} \rangle_{\mathbf{D}_n}$ con il vincolo $\|\mathbf{a}\|_{\mathbf{Q}} = 1$. La soluzione è l’autovettore associato al massimo autovalore della matrice $\mathbf{Q}^{1/2}(\mathbf{Y}'\mathbf{D}_n\mathbf{X} + \mathbf{X}'\mathbf{D}_n\mathbf{Y})\mathbf{Q}^{1/2}$. La totale sovrapponibilità di \mathbf{X} e \mathbf{Y} ci porta a considerare anche $(\mathbf{X} - \mathbf{Y}, \mathbf{Q}, \mathbf{D}_n)$. L’analisi della co-struttura evidenzia gli elementi di rassomiglianza, mentre l’ACP di $\mathbf{X} - \mathbf{Y}$ evidenzia lo scarto tra gli individui. Siano \mathbf{I}_X , \mathbf{I}_Y e \mathbf{I}_D le inerzie associate, rispettivamente, all’ACP di \mathbf{X} , \mathbf{Y} e $\mathbf{X} - \mathbf{Y}$, l’Analisi della co-struttura e quella della differenza sono legate dalla seguente relazione $2\text{tr}(\mathbf{X}'\mathbf{D}_n\mathbf{YQ}) = \mathbf{I}_X + \mathbf{I}_Y - \mathbf{I}_D$. Un’estensione dell’analisi di Torre e Chessel nel caso di questionari somministrati più volte (*multi-modulo*) agli stessi individui, che si presenta per alcuni tipi di servizi di day surgery i cui

trattamenti devono essere ripetuti più volte, è stata proposta da D'Ambra e Amenta (2000b). Siano $(\mathbf{X}, \mathbf{Q}, \mathbf{D}_n)$ e $(\mathbf{Y}_h, \mathbf{Q}, \mathbf{D}_n)$, rispettivamente, triplette statistiche associate alle aspettative e alle percezioni rilevate al termine del h -esimo trattamento in day surgery ($h = 1, \dots, H$). Siano $\mathbf{Y}' = [\mathbf{Y}'_1 | \dots | \mathbf{Y}'_H]$ e $\mathbf{X}' = [\mathbf{X}' | \dots | \mathbf{X}']$ ($nH \times p$), matrici ottenute giustapponendo per riga le H matrici \mathbf{Y}_h e \mathbf{X} con $\mathbf{D} = \text{diag}(\mathbf{D}_n, \dots, \mathbf{D}_n)$ matrice dei pesi. L'analisi della co-struttura di tabelle "multi-modulo" totalmente sovrapponibili consiste nella ricerca del vettore dei coefficienti $\mathbf{a}_{(p,1)}$ che massimizza la somma della covarianze fra $\mathbf{Y}_h \mathbf{Q} \mathbf{a}$ e

$$\mathbf{X} \mathbf{Q} \mathbf{a} : \max \sum_{h=1}^H \langle \mathbf{X} \mathbf{Q} \mathbf{a}, \mathbf{Y}_h \mathbf{Q} \mathbf{a} \rangle_{\mathbf{D}_n} = \max \langle \mathbf{X} \mathbf{Q} \mathbf{a}, \mathbf{Y} \mathbf{Q} \mathbf{a} \rangle_{\mathbf{D}} \quad \text{con} \quad \mathbf{a}' \mathbf{Q} \mathbf{a} = 1.$$

Parimenti alla co-struttura, l'obiettivo è di evidenziare la rassomiglianza fra \mathbf{Y}_h ed \mathbf{X} così come le singole strutture. L'analisi delle differenze $(\mathbf{X} - \mathbf{Y}, \mathbf{Q}, \mathbf{D})$ "multi-modulo" risulta legata a quella della co-struttura "multi-modulo": $(1/H) \mathbf{I}_{\text{DF}} = [(1/H) \sum_{h=1}^H \mathbf{I}_{\mathbf{Y}_h}] + \mathbf{I}_{\mathbf{X}} - 2 \text{tr}(\mathbf{X}' \mathbf{D}_n \mathbf{Y} \mathbf{Q})$ dove $\mathbf{I}_{\mathbf{Y}_h}$, $\mathbf{I}_{\mathbf{X}}$ e \mathbf{I}_{DF} sono, le inerzie associate alle triplette $(\mathbf{Y}_h, \mathbf{Q}, \mathbf{D}_n)$, $(\mathbf{X}, \mathbf{Q}, \mathbf{D}_n)$ e $(\mathbf{X} - \mathbf{Y}, \mathbf{Q}, \mathbf{D})$. Una generalizzazione di tipo tensoriale al caso di H rilevazioni, ciascuno all'inizio ed alla fine di ogni modulo (*full multi modules*) è stata proposta da Amenta e Sarnacchiaro (2001). L'analisi della co-struttura di tabelle "full multi-modules" totalmente sovrapponibili consiste nella ricerca del vettore dei coefficienti \mathbf{a} tale che: $\max \sum_{h=1}^H \langle \mathbf{X}_h \mathbf{Q} \mathbf{a}, \mathbf{Y}_h \mathbf{Q} \mathbf{a} \rangle_{\mathbf{D}_n}$ con il vincolo $\mathbf{a}' \mathbf{Q} \mathbf{a} = 1$.

4. Analisi statistica multidimensionale delle "dimensioni"

Nell'analisi dei dati di CS, spesso l'importanza attribuita alle dimensioni è utilizzata semplicemente come sistema di ponderazione delle valutazioni espresse dagli utenti/clienti sulle percezioni o sullo scostamento percezioni meno aspettative. Tali dati, tuttavia, possono essere analizzati anche direttamente. Vari studi hanno evidenziato, infatti, l'interrelazione esistente tra l'importanza attribuita alle diverse dimensioni (Parasuraman, Zeithman, e Berry; 1994). L'utilizzo di tecniche statistiche multidimensionali può evidenziare la natura di tali interrelazioni e come l'importanza attribuita alle dimensioni influisce sulla "overall satisfaction".

I dati sull'importanza attribuita alle dimensioni, tuttavia, presentano una scala di rapporto di tipo compositivo. Una matrice di dati compositivi \mathbf{C} (n, w) ha tutti elementi positivi ed ogni riga presenta il vincolo di somma unitaria $\mathbf{C} \hat{\mathbf{1}}_w = \hat{\mathbf{1}}_n$, dove $\hat{\mathbf{1}}_w$ e $\hat{\mathbf{1}}_n$ sono vettori unitari rispettivamente di dimensione w e n .

Il vincolo di somma unitaria per ogni riga di \mathbf{C} implica alcuni problemi e difficoltà nell'interpretazione dei risultati (Aitchison, 1986). In particolare, ponendoci in un contesto multidimensionale, tali difficoltà sono legate alla struttura della matrice di covarianza ed al relativo "bias" negativo per la matrice di correlazione. Infatti, la matrice di covarianza associata a \mathbf{C} , così come quella di

correlazione, presenta la caratteristica che ogni riga contiene almeno una covarianza (o correlazione) negativa, quindi, l'usuale interpretazione "correlazione pari a zero non vi è associazione" va vista con prudenza. Un altro problema è l'effetto curvatura che tali dati presentano. Per superare le difficoltà nell'interpretazione ed il problema della curvatura, Aitchison (1986) propone diverse trasformazioni basate sulla combinazione lineare di "logratio". La motivazione è che nello studio dei dati di composizione si è interessati alle valutazioni relative più che ai valori assoluti. A tal fine, la trasformazione "centred logratio" $c_{ij}^* = \log(c_{ij} / \sqrt{c_{i1} \times \dots \times c_{iw}})$ ($i = 1, \dots, n$, $j = 1, \dots, w$) si rivela appropriata per l'analisi multidimensionale.

Tuttavia, le trasformazioni basate sui logaritmi non sono sempre le più performanti per risolvere i problemi di curvatura, inoltre, esse richiedono il rispetto dell'ipotesi che ogni elemento della matrice \mathbf{C} deve essere strettamente positivo: $c_{ij} > 0 \quad \forall i, j$. Alternativamente, alle trasformazioni basate sui logaritmi, è possibile utilizzare delle trasformazioni spline (Gallo, 2001). La matrice dei dati composizionali trasformata \mathbf{R}^* è ottenuta applicando una particolare funzione polinomiale a tratti e l'operatore di proiezione ortogonale

$\mathbf{P}_w^\perp = \left(\mathbf{I}_w - \frac{1}{w} \hat{\mathbf{1}}_w \hat{\mathbf{1}}_w' \right)$, cosicché ogni riga e colonna della matrice \mathbf{R}^* ha somma

uguale a zero. In particolare, la matrice \mathbf{R}^* è data da $[\mathbf{B}^1 \boldsymbol{\alpha}^1 | \dots | \mathbf{B}^w \boldsymbol{\alpha}^w]$, dove le \mathbf{B}^w è la matrice di B-spline, relativa alla w -esima variabile, e gli $\boldsymbol{\alpha}^j$ sono i vettori dei coefficienti delle funzioni spline.

Un'analisi sulla struttura dell'importanza attribuita alle dimensioni può aversi con la "Logconstrant ACP" (Aitchison, 1986) equivalente a massimizzare $\mathbf{a}'\mathbf{C}^*\mathbf{C}^*\mathbf{a}$ con il vincolo $\mathbf{a}'\mathbf{a} = 1$. Similmente, può essere effettuata un'ACP "centered logratio" della matrice \mathbf{C}^* o sulla matrice \mathbf{R}^* .

Approcci differenti, sempre nell'ambito del criterio di co-inerzia, sono quelli di utilizzare l'analisi della co-struttura tensoriale di tabelle "full multi-modules" totalmente sovrapponibili (Amenta e Sarnacchiaro, 2001) ed il PLS (Wold, 1966). Nel caso dei dati composizionali, il PLS può essere utilizzato per effettuare un'analisi dell'inerzia sulla matrice dei dati centred logratio (o \mathbf{R}^*) con il vincolo che le variabili latenti spieghino il più possibile il vettore della soddisfazione complessiva \mathbf{z} : $cov^2(\mathbf{z}, \mathbf{a}'\mathbf{C}^*)$ con i vincoli $\mathbf{a}'\mathbf{a} = 1$ e $\mathbf{a}'\hat{\mathbf{1}}_p = 0$.

5. Il ruolo delle informazioni esterne nella CS

La valutazione della CS nei servizi di Day Surgery può essere migliorata se si inseriscono nel modello eventuali informazioni esterne disponibili sugli utenti/clienti, quali le caratteristiche socio-culturali degli utenti, il numero e la tipologia di esperienze avute in altre strutture di Day Surgery, etc..

Nella costruzione degli indici sintetici di soddisfazione basati sulla combinazione

lineare degli indicatori, si ipotizza che il collettivo di utenti/clienti sia distribuito in modo casuale rispetto a queste ultime. Tale ipotesi viene meno se si identifica una struttura sugli utenti, in riferimento alle modalità di uno o più determinati fattori, tale partizione può portare ad influenzare il sistema di aspettative e di percezioni rendendo delicata l'attività di aggregazione delle informazioni raccolte da soggetti diversi con esperienze diverse. È possibile, inoltre, formulare delle ipotesi di non indipendenza delle componenti delle dimensioni. Tali aspetti assumono un ruolo d'informazioni esterne rispetto al sistema in studio e di differente natura: informazioni esterne quantificabili come vincoli lineari sugli items o sulle dimensioni ed informazioni esterne indotte da una partizione degli utenti, quest'ultimo per tener conto dei diversi stati iniziali degli utenti/clienti. Nella tabella 1, si può notare come sia possibile inserire le informazioni esterne nei diversi modelli trattati.

Tab. 1: Le informazioni esterne ed alcuni criteri per l'analisi multidimensionale dei dati di Day Surgey.

Metodologia	Criterio	Vincoli	Inserimento Inf. esterne
<i>Log-Constraint PLS</i> Aitchison (1986)	$cov^2(\mathbf{z}, \mathbf{a}'\mathbf{C}^*)$	$\mathbf{a}'\mathbf{a} = 1$	Si
<i>Centered Logratio PLS</i> D'Ambra, Amenta, Gallo (2001)	$cov^2(\mathbf{z}, \mathbf{a}'\mathbf{C}^*)$	$\mathbf{a}'\mathbf{a} = 1,$ $\mathbf{a}'\hat{\mathbf{1}}_p = 0$	Si
<i>ComPLS</i> Gallo (2001)	$cov^2(\mathbf{z}, \mathbf{a}'\mathbf{R}^*)$	$\mathbf{a}'\mathbf{a} = 1,$ $\mathbf{a}'\hat{\mathbf{1}}_p = 0$	Si
<i>Co-Inerzia</i> Chessel-Mercier (1993)	$\langle \mathbf{XQa}, \mathbf{YQb} \rangle_{D_n}$	$\mathbf{a}'\mathbf{a} = 1,$ $\mathbf{b}'\mathbf{b} = 1$	Si
<i>Co-Struttura (Single Module)</i> Torre-Chessel (1995)	$\langle \mathbf{XQa}, \mathbf{YQa} \rangle_{D_n}$	$\mathbf{a}'\mathbf{a} = 1$	Si
<i>Co-Struttura (Multi Modules)</i> D'Ambra-Amenta (2000b)	$\sum_{h=1}^H \langle \mathbf{XQa}, \mathbf{Y}_h \mathbf{Qa} \rangle_{D_n}$	$\mathbf{a}'\mathbf{a} = 1$	Si
<i>Co-Struttura (Full Multi Modules)</i> Amenta-Sarnacchiaro (2001)	$\sum_{h=1}^H \langle \mathbf{X}_h \mathbf{Qa}, \mathbf{Y}_h \mathbf{Qa} \rangle_{D_n}$	$\mathbf{a}'\mathbf{a} = 1$	Si

Il modello di Tarane e Shibayama (1991) decompone i dati in diverse componenti secondo le informazioni esterne. Sia $\mathbf{E}_{(n,s)}$ la matrice delle informazioni sulle unità e $\mathbf{F}_{(p,m)}$ quella sulle colonne con $\mathbf{P}_Z = \mathbf{E}(\mathbf{E}'\mathbf{E})^{-1}\mathbf{E}'$ il proiettore ortogonale. Le matrici \mathbf{E} ed \mathbf{F} possono essere di vario tipo (componenti unitarie, variabili indicatrici o continue). Ciascuna componente della matrice delle aspettative \mathbf{X} (similmente per le percezioni \mathbf{Y} o anche $\mathbf{X}-\mathbf{Y}$): $\mathbf{P}_E\mathbf{X}\mathbf{P}_F + \mathbf{P}_E^\perp\mathbf{X}\mathbf{P}_F + \mathbf{P}_E\mathbf{X}\mathbf{P}_F^\perp + \mathbf{P}_E^\perp\mathbf{X}\mathbf{P}_F^\perp$; ha un significato statistico: $\mathbf{P}_E\mathbf{X}\mathbf{P}_F$ indica l'effetto delle informazioni di riga e colonna; $\mathbf{P}_E^\perp\mathbf{X}\mathbf{P}_F$ quello delle informazioni di colonna al netto di quelle di riga; $\mathbf{P}_E\mathbf{X}\mathbf{P}_F^\perp$ quello delle informazioni di riga al netto di quelle di colonna e $\mathbf{P}_E^\perp\mathbf{X}\mathbf{P}_F^\perp$ la parte che non tiene

conto delle informazioni esterne. Decomposta la matrice \mathbf{X} , può essere interessante eseguire analisi multivariate su ciascun (o più) termine, per cogliere (o depurare) l'influenza delle informazioni esterne sulla valutazione della qualità del servizio erogato.

Le informazioni esterne hanno ruoli differenti e, nell'ambito della valutazione della CS, generalmente, si è interessati ad eliminare i differenti sistemi di riferimento degli utenti/clienti così come cogliere gli effetti delle informazioni sia sulle componenti che sulle dimensioni. In effetti, includere le informazioni esterne all'interno del modello di valutazione, permette di cogliere il livello di soddisfazione dei diversi gruppi omogenei di utenti/clienti. Nella tabella 2 è data la matrice delle aspettative in presenza di informazioni esterne sulle righe e/o sulle colonne (similmente può essere ottenuta la matrice delle percezioni con informazioni esterne).

Tab. 2: La matrice delle aspettative in presenza di informazioni esterne.

Informazioni esterne	$\mathbf{E} = \mathbf{I}_n ; \mathbf{F} = \mathbf{I}_p$	$\mathbf{E} = \mathbf{I}_n ; \mathbf{F}$	$\mathbf{E}; \mathbf{F} = \mathbf{I}_p$	$\mathbf{E}; \mathbf{F}$
Matrice	$\mathbf{P}_E \mathbf{X} \mathbf{P}_F$	$\mathbf{P}_E \mathbf{X} \mathbf{P}_F$	$\mathbf{P}_E^\perp \mathbf{X} \mathbf{P}_F$	$\mathbf{P}_E^\perp \mathbf{X} \mathbf{P}_F$

6. Verso un'analisi non lineare dei dati basata sulla Co-Inerzia

Spesso le informazioni disponibili si riferiscono a variabili nominali, ordinali e continue. Differenti tecniche sono state proposte per l'analisi congiunta di variabili di tipo misto che consistono nella codifica delle variabili rispettando le loro strutture (nominale, ordinale). Ponendosi nell'ottica dell'Optimal Scaling, ad ogni variabile è associata un sottospazio di trasformazione non lineare per (variabili quantitative) o un sottospazio di quantificazione (variabili qualitative), nel primo caso si usano le variabili indicatrici, nel secondo le funzioni indicatrici di ordine con parametri positivi. In effetti, è sempre possibile definire un generico vettore \mathbf{t} del sottospazio delle trasformazioni $\mathbf{t} = \mathbf{T}\beta$ con $\beta > 0$ e dove le colonne di \mathbf{T} sono le basi del sottospazio. Le matrici \mathbf{T} sono o variabili indicatrici o funzioni indicatrici di ordine. Un criterio generale è basato su una funzione di perdita i cui parametri sono stimati con l'ALS. Un criterio diverso (Di Ciaccio, 1989) si basa sulla correlazione: $\max \text{corr}(\mathbf{T}\gamma; \sum \mathbf{V}\delta)$; dove $\mathbf{T}\gamma$ è la generica trasformazione della variabile di risposta e \mathbf{V} è la generica trasformazione delle variabili esplicative come definite precedentemente per rispettare la struttura delle variabili si perviene a Addals e Morals se le variabili sono di tipo qualsiasi. Il criterio della correlazione può essere utilizzato anche per estendere l'ACP all'analisi dei caratteri misti e si ritrova ACP non lineare Princals. Generalizzando lo stesso criterio, si ha un'estensione dell'Analisi Canonica Generalizzata per caratteri con livello di misura mista, nota come Overals (Van der Burg, De Leeuw e Verdegaaal, 1988). Le strategie basate sulla massimizzazione del coefficiente di correlazione conducono a soluzioni dove le correlazioni, per quanto importanti,

presentano rappresentazioni delle strutture deboli: il principale problema del criterio di correlazione (principio alla base dell'analisi canonica), la quale presenta oltre a difficoltà interpretative legate alla metrica di Mahalanobis, anche quella di spiegare la struttura comune ai due insiemi di dati senza considerare la variabilità di ognuno di essi. L'analisi di Chessel e Mercier, invece, massimizza $cov(\mathbf{XQ}_x\mathbf{a}, \mathbf{YQ}_y\mathbf{b}) = \text{corr}(\mathbf{XQ}_x\mathbf{a}, \mathbf{YQ}_y\mathbf{b})\sqrt{\text{var}(\mathbf{XQ}_x\mathbf{a})}\sqrt{\text{var}(\mathbf{YQ}_y\mathbf{b})}$ che evidenzia un compromesso tra il criterio della correlazione e quello della varianza. In effetti, la principale caratteristica di questo tipo d'analisi è di considerare simultaneamente le analisi d'inerzia e le relative relazioni tra le due matrici. Per la valutazione della CS è interessante considerare anche approcci non lineari basati sul criterio di co-inerzia e non su quello di correlazione. Ciò consentirebbe di ottenere delle analisi che permettano di studiare i legami associativi fra variabili di differente natura e di garantire un'alta capacità esplicativa.

Riferimenti bibliografici

- Aitchison J. (1986) *The Statistical Analysis of Compositional Data*. Chapman and Hall.
- Amenta P., Sarnacchiaro P. (2001) Tensorial Co-Structure Analysis for the full Multi Modules Customer Satisfaction Evaluation. *Atti Riunione Scientifica del Gruppo di Classificazione dell'IFCS su "Classificazione e Analisi dei Dati"*, Palermo.
- Chessel D., Mercier P. (1993) *Couplage de Triplets Statistiques et Liaisons Espèces-Environnement*. In J.D. Lebreton & B. Asselain (eds.), *Biométrie et Environnement*.
- Cronin J. J., Taylor S. A. (1992) Measuring Service Quality: A Reexamination and Extension. *Journal of Marketing*, vol. 56, 3.
- D'Ambra L., Amenta P. (2000a) Un Approccio Statistico Multidimensionale per la Valutazione della Performance dei Servizi di Interfaccia con Informazioni Esterne. *Aicq Meeting*, Bologna.
- D'Ambra L., Amenta P. (2000b) Multidimensional Statistical Methods based on Co-Inertia for the 'multi-modules' Customer Satisfaction Evaluation. *Cimasi'2000 EHTP*, Morocco.
- D'Ambra L., Amenta P., Gallo M., F., Sarnacchiaro P. (1999) Multidimensional Statistical Methods based on Co-Inertia for the Customer Satisfaction Evaluation. *ICQ*, Thailandia.
- D'Ambra L., Sabatier R., Amenta P. (1998) Three-way Factorial Analysis: synthesis and new approaches. In *atti di convegno SIS Sorrento*, 39 p.229-240.
- D'Ambra L., Lombardo R., Amenta P. (2000) Multivariate Co-Inertia Analysis for Nominal and Ordinal Variables by PLS. *Submitted*.
- Di Ciaccio A. (1989) *Maximal Association Criteria and Optimal Scaling – a unified approach to the analysis of qualitative and quantitative data*. Tesi di dottorato. Università di Roma.
- Gallo M. (2001) Partial Least Squares for Compositional data. In *Atti del II International Symposium on PLS and Related Methods*. Capri, 153-162.
- Gallo M. (2002) Multidimensional Analysis of Customer Satisfaction Data: The Scaling Problems. *Presentato a DEINDE*. Torino.
- Gifi A. (1990) *Nonlinear Multivariate Analysis*. Wiley, Chicester.
- Green P.E., Tull D.S. (1988) *Research for Marketing Decisions*. Prentice Hall, New York.

- Jones L.V., (1986) *Psychological Scaling*. Encyclopaedia of Statistical Sciences, J. Wiley, N.Y
- Kruskal J.B. (1965) Analysis of Factorial Experiments by Estimating Monotone Transformations of the data. *Journal of the Royal Statistical Society, B*, vol. 27.
- Oliver R. L. (1993): *A conceptual model of service quality and service satisfaction: compatible goals, different concepts*. In *Advances in Service Marketing and Management: Research and Practice*, Vol. 2. JAI Press.
- Parasuraman A., Zeithaml V., Berry L. (1991) Refinement and reassessment of the Servqual scale, *Journal of Retailing*, vol. 67, 4.
- Parasuraman A., Zeithaml V., Berry L. (1994) Reassessment of Expectations as a Comparison Standard in Measuring Service Quality: Implications for Future Research, *Journal of Marketing*, vol. 58, 1.
- Sarnacchiaro P., Gallo M., D'Ambra L. (2001) Multidimensional Statistical Analysis as Tool for the Study of Customer Satisfaction Models. *Submitted*.
- Takane Y., Shibayama T. (1991) Principal Component Analysis with External Information on both Subjects and Variables. *Psychometrika*, 56, pp. 97-120.
- Torre F., Chessel D. (1995) Co-Structure de deux Tableaux Totalemt Appariés. *Revue de Statistique Appliquée*, XLIII (1), 109-121.
- Tucker L.R. (1958) An Inter-battery method of Factor Analysis. *Psychometrika*, vol. 23.
- Van der Burg E., De Leeuw J., Verdegaal R. (1988) Nonlinear Canonical Correlation with m sets of variables. *Psychometrika*, vol. 53, n. 2.
- Wold H. (1966) *Estimation of Principal Components and Related Models by Iterative Least Squares*. In *Multivariate Analysis*, ed. Krishnaiah. New York: Academic Press.
- Wright B.D., Masters G.N. (1982) Rating Scale Analysis, Rasch Measurement. *MESA*.
- Wright, B.D., Linacre, J.M. (1989) Observations are Always Ordinal: Measures, However, Must be Interval. *Archives of Physical Medicine and Rehabilitation*, vol. 70.
- Zanella A., (1999) A Statistical Model for the Analysis of Customer Satisfaction: some theoretical aspects. *Statistica*, vol. 59.

(1) *Il presente lavoro è stato svolto con il contributo COFIN 2001 (Prof. L. D'Ambra) e con il contributo COFIN 2001 (Prof. P. Amenta).*