

Utilizzo di modelli causali per la Metropolitana di Roma

Giovanni Monaco, Barbara Ferraro, Fabio Marotta
CFI Group

Abstract: Conoscere la soddisfazione generata dal servizio erogato è un momento fondamentale della relazione tra l'Azienda e i suoi clienti. Ottenere un quadro dettagliato però è un esercizio complesso; esso infatti deriva da molteplici percezioni, riconducibili ad altrettanti elementi misurabili del processo di erogazione, non sempre e non completamente razionalizzabili dal fruitore. Nel caso di fenomeni complessi e articolati si deve procedere all'implementazione dei modelli di equazioni strutturali il cui obiettivo è quello di fornire una rappresentazione semplificata e grafica dei processi reali, tenendo conto non solo della molteplicità delle cause che agiscono su una variabile dipendente, ma anche delle connessioni esistenti fra le diverse cause. Viene qui offerta una panoramica delle principali caratteristiche di questo tipo di sistema, utilizzato da CFI Group nel caso delle rilevazioni di Customer Satisfaction condotte per Metropolitana di Roma SpA. Vengono presi in esame anche degli indicatori statistici utilizzati per definire l'affidabilità e la validità del modello proposto.

Parole chiave/Keywords: PLS, equazioni strutturali, modelli causali, variabili latenti, analisi multivariate.

1. Introduzione

Il cliente è sempre più esigente e consapevole che può esercitare il suo "potere" nei confronti delle aziende con delle scelte comportamentali, in quanto ogni scelta (uso-acquisto) è conseguenza della soddisfazione percepita da esperienze precedenti. Alle imprese occorre perciò individuare quali indicatori compongono le variabili "latenti" della soddisfazione, quali di queste influiscono maggiormente sulla soddisfazione, e quanto questa influisca a sua volta sui comportamenti futuri, sia che si tratti della ripetizione dell'acquisto o della propensione a parlare favorevolmente dell'esperienza fatta (promotori-apostoli). Statisticamente significa studiare le relazioni di causa-effetto che intercorrono tra le variabili della qualità, che influiscono sulla customer satisfaction e sulla loyalty, ottenendo per ciascuna di esse misure affidabili.

Nel caso di fenomeni complessi e articolati si deve quindi procedere all'implementazione di modelli il cui obiettivo è quello di fornire una rappresentazione semplificata e grafica dei processi reali, tenendo conto non solo della molteplicità delle cause che agiscono su una variabile dipendente, ma anche delle connessioni esistenti fra le diverse cause.

2. I modelli causali e le equazioni strutturali

Metodi come la correlazione o la regressione semplice esaminano la relazione esistente tra due variabili assumendo che il sistema non sia influenzato da altre variabili oltre a quelle considerate nell'analisi. In situazioni reali questa ipotesi non è realistica.

L'analisi di regressione multipla risolve in parte questi problemi perché analizza la relazione che esiste tra alcune variabili (es. vetture, stazioni, affidabilità) e la variabile dipendente (ad esempio "la soddisfazione complessiva"). Trattandosi di un test parametrico è però necessario che vengano soddisfatte alcune assunzioni. Ad esempio:

- sulle variabili → distribuzione normale, assenza di collinearità tra le variabili indipendenti, assenza di errori di misurazione;
- sui residui del modello → media nulla, omoschedasticità, incorrelazione, distribuzione normale.

Tra tutti la multicollinearità è un assunto particolarmente restrittivo soprattutto nel campo delle indagini di Customer Satisfaction, dove sono frequenti le forti interrelazioni tra le aree che costituiscono la qualità percepita del servizio e la valutazione complessiva dello stesso.

Laddove infatti i fenomeni da studiare sono più complessi è necessario procedere all'implementazione dei modelli di equazioni strutturali.

Questi modelli cercano di rispondere a due importanti quesiti: il primo è quello della misurazione e nasce dal fatto che conoscere lo stato di soddisfazione verso un servizio è un aspetto molto complesso, perché deriva da molteplici percezioni che, pur essendo riconducibili ad elementi misurabili del servizio, non sono completamente razionalizzabili; il secondo interrogativo riguarda il problema della causalità, dove si cerca il nesso causale esistente tra le variabili latenti (ad esempio quanto lo stato delle vetture o delle stazioni influenzano la soddisfazione del cliente nei confronti del servizio di trasporto).

Complessivamente i modelli ad equazioni strutturali sono distinti in due grandi famiglie: quelli *covariance-based*, di cui fanno parte tutti i metodi di tipo LISREL (Linear Structural Relationship) e quelli *components-based* ed il più noto è quello basato sul PLS (Partial Least Squares).

I due approcci si distinguono sulla base del diverso obiettivo perseguito:

- nei metodi di stima *covariance-based* i parametri del modello vengono individuati in modo da ottenere una ricostruzione della matrice di varianze-covarianze tra le variabili manifeste il più possibile vicina a quella osservata;
- nei metodi cosiddetti *component-based*, invece, l'obiettivo è quello di stimare le variabili latenti in modo che siano il più rappresentative possibili del proprio gruppo di variabili manifeste e che, contemporaneamente, riescano a spiegare al meglio le relazioni individuate nel modello strutturale.

CFI GROUP per le analisi di customer satisfaction utilizza l'approccio PLS (o meglio, un sviluppo dell'algoritmo NIPALS proposto da Claes Fornell nel 1994).

Con questo metodo si vuole ridurre il più possibile la probabilità di errori statistici e aumentare la precisione della rilevazione perché le valutazioni dei clienti su aspettative, qualità, valore, soddisfazione e comportamenti futuri non vengono misurate attraverso le

singole domande dirette rivolte agli intervistati, ma calcolando Variabili Latenti¹ che “spiegano” la varianza in comune di più indicatori singoli afferenti allo stesso fenomeno.

3. La costruzione di modelli stabili

La stabilità del modello causale è un requisito fondamentale ed imprescindibile per CFI Group, soprattutto nel caso di monitoraggi che si ripetono a distanza di pochi mesi e, quindi, presuppongono un'elevata precisione ed affidabilità delle misurazioni.

Per CFI Group affinché un modello di Customer Satisfaction sia stabile deve rispettare pienamente i criteri di *convergent validity* e *discriminant validity*. In altre parole si ha un buon modello se si riscontra un'elevata correlazione fra gli item associati alla misura di un costrutto latente (*convergent validity*) e, nel contempo, si evince una debole correlazione fra gli item che costituiscono una Latente “A” e quelli utilizzati per esprimere una differente Variabile Latente “B” (*discriminant validity*). In estrema sintesi, le Variabili Latenti sono costruite in modo corretto se esprimono al loro interno un solo concetto (unidimensionalità) e, se riescono a spiegare una variabile dipendente senza sovrapporre più concetti (riduzione della multicollinearità).

Per verificare questi due criteri CFI utilizza dei parametri statistici molto precisi in base ai quali si può decidere quali modifiche strutturali apportare al modello in esame. In questo modo il modello finalizzato sarà statisticamente il più corretto e, di conseguenza, più stabile nel tempo.

Il primo passo è quello di verificare il concetto di validità del costrutto latente e quindi il criterio di *convergent validity*.

Le Variabili Manifeste rientrano nel calcolo dello stesso indice (Variabile Latente) quando risultano più correlate con le variabili che concorrono alla formazione dell'indice che con altre variabili misurate. Ciò significa che le Manifeste hanno una parte rilevante di “varianza in comune” e che l'indice risultante è unidimensionale². Una condizione necessaria, ma non sufficiente, per la validità di misure è che queste siano *sicure, affidabili* (J. Paul Peter 1979).

La cosiddetta *reliability* è il grado con cui le misure sono libere da errori in modo da produrre risultati consistenti e coerenti. Nella ricerca, il termine *reliability* indica, infatti, “ripetibilità”. Una misura è considerata affidabile quando ci dà lo stesso risultato più e più volte (assumendo che ciò che stiamo misurando non sta cambiando).

I parametri che CFI utilizza per testare la *convergent validity* sono la *communality* e/o i *loading*, l' α di Cronbach e gli *autovalori* della Latente.

Una volta verificato il primo criterio si prosegue verificando il secondo (*discriminant validity*) che viene soddisfatto se ogni Variabile Latente è più correlata con la Variabile Latente da spiegare (Es: Soddisfazione) che con le altre Variabili Latenti del modello. In caso contrario

¹ Il metodo PLS costruisce nuove variabili esplicative ortogonali tra loro, chiamate variabili latenti, usate come predittori delle variabili di risposta. L'obiettivo è quello di creare un numero di latenti, tali da catturare la massima informazione dalle variabili esplicative, in modo che siano le più utili a spiegare le variabili dipendenti e quindi il fenomeno oggetto di studio. Le componenti oltre ad essere predittori della Y, modellano anche la matrice X. Tali variabili latenti sono stimate come combinazione lineare tra la matrice delle variabili originali X e la matrice dei coefficienti chiamati “weights” W. Le latenti ricostruiscono la matrice X, minimizzandone i residui.

² I valori delle Variabili Latenti sono calcolati come “media ponderata” dei punteggi assegnati dai clienti intervistati alle Variabili Manifeste. La metodologia CFI usa il PLS come algoritmo di calcolo dei pesi.

ci si aspetta un'elevata multicollinearità e, conseguentemente, la presenza di impatti³ non significativi.

Per verificare questo criterio si utilizza un indicatore detto Condition Number.

Per poter monitorare con facilità la presenza o assenza di multicollinearità è necessario esaminare la matrice di correlazione tra tutte le Variabili Latenti e la variabile da spiegare (es. Soddisfazione) in modo da capire quali potrebbero creare instabilità e quindi porre rimedio con una path-analysis ad-hoc.

3. L'esempio Metropolitana di Roma SpA

Si vedranno, ora, i passi salienti del processo seguito per costruire dei modelli stabili per le ferrovie concesse di Metropolitana di Roma SpA.

Le linee oggetto di studio sono:

- Linea Roma-Pantano;
- Linea Roma-Lido;

Partendo dal modello base (Figura 1) i risultati delle due linee vengono analizzati singolarmente per verificare se i criteri di convergent validity e di discriminant validity sono soddisfatti.

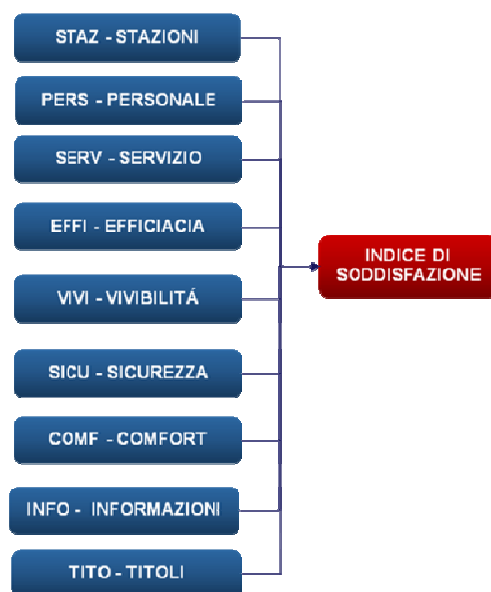


Figura 1

³ In termini statistico-econometrici l'impatto è il coefficiente di regressione (standardizzato e riportato in scala 0-100) di ciascuna delle rette interpolanti le distribuzioni delle coppie di valori assegnati dai rispondenti alle componenti della qualità (Stazioni, Personale, Sicurezza, ecc.) e alla Soddisfazione. Misura "l'elasticità" della soddisfazione alla componente della qualità in analisi, a parità di tutte le altre componenti.

In termini interpretativi l'impatto è una misura della forza di ciascuna componente della qualità nel determinare variazioni nella soddisfazione, ovvero è una stima delle aspettative, del cliente su quella componente. L'impatto consente anche di comprendere l'utilità marginale del miglioramento di ciascuna componente della qualità percepita. Tale valore, essendo "estrapolato" e non direttamente richiesto, consente di evitare le distorsioni derivanti dal richiedere un giudizio diretto sull'importanza.

La Roma-Pantano rispetta il criterio di convergent validity ma presenta problemi di multicollinearità che vengono risolti eseguendo una path-analysis.

Osservando la matrice di correlazione (Tabella1) si può notare infatti che molte Variabili Latenti sono più correlate tra di loro di quanto non lo siano con la Soddisfazione (variabile dipendente): PERS, EFFI, VIVI, COMF e INFO. Inoltre gli impatti³ ottenuti non sono significativi. Il modello della Figura 1 non ha quindi la struttura ottimale per spiegare in modo affidabile i risultati della linea oggetto di studio.

	staz	pers	servi	effi	vivi	sicu	comf	info	tito	csi
staz	1.00	0.63	0.56	0.44	0.62	0.58	0.58	0.50	0.28	0.61
pers	0.63	1.00	0.64	0.68	0.61	0.58	0.61	0.67	0.36	0.54
servi	0.56	0.64	1.00	0.67	0.48	0.52	0.57	0.65	0.43	0.59
effi	0.44	0.68	0.67	1.00	0.44	0.47	0.61	0.60	0.42	0.45
vivi	0.62	0.61	0.48	0.44	1.00	0.61	0.70	0.56	0.29	0.51
sicu	0.58	0.58	0.52	0.47	0.61	1.00	0.65	0.64	0.30	0.56
comf	0.58	0.61	0.57	0.61	0.70	0.65	1.00	0.63	0.40	0.55
info	0.50	0.67	0.65	0.60	0.56	0.64	0.63	1.00	0.45	0.54
tito	0.28	0.36	0.43	0.42	0.29	0.30	0.40	0.45	1.00	0.52
csi	0.61	0.54	0.59	0.45	0.51	0.56	0.55	0.54	0.52	1.00

Tabella 1

La soluzione proposta è quella di creare due sub-modelli (per il Servizio e per la Sicurezza) riducendo in questo modo il numero di variabili che influiscono direttamente sulla Soddisfazione. Queste modifiche non tolgono nessuna informazione dato che oltre agli impatti diretti di STAZ, SERVI, SICU e TITO sulla soddisfazione si avranno anche gli impatti totali di EFFI e INFO (mediati da SERVI) e di PERS, COMF e VIVI (mediati da SICU) sulla soddisfazione.

Qui di seguito la matrice di correlazione (tabella 2) ottenuta dopo la path-analysis:

	staz	servi	sicu	tito	csi
staz	1.00	0.56	0.58	0.28	0.61
servi	0.56	1.00	0.52	0.43	0.59
sicu	0.58	0.52	1.00	0.30	0.56
tito	0.28	0.43	0.30	1.00	0.52
csi	0.61	0.59	0.56	0.52	1.00

Tabella 2

Il modello che ne risulta (Figura 2) è stabile e tutti gli impatti sono significativi al 95% delle probabilità.

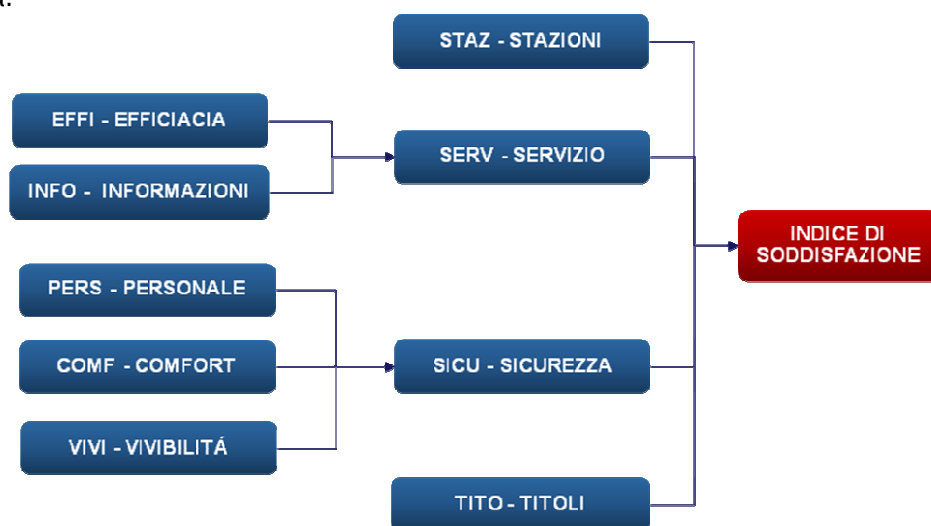


Figura 2

Applicando lo stesso modello ottenuto con la linea Roma-Pantano alla Roma-Lido non si ottengono risultati ugualmente stabili.

Per la linea Roma-Lido è impossibile utilizzare tutte le Variabili Manifeste come è stato fatto per la linea precedente. A differenza della linea Roma-Pantano infatti il criterio di convergent validity non è rispettato quindi si procede all'isolamento delle Variabili Manifeste che potrebbero portare ad un costrutto Latente poco corretto⁴. Inoltre, anche in questo caso è stata effettuata una path-analysis per ridurre la multicollinearità.

Dalla matrice di correlazione si evince che le variabili più problematiche, con impatti non statisticamente significativi, sono:

- STAZ
- EFFI
- SICU

Il problema si risolve costruendo di un sub-modello della Latente Vivibilità.

La nuova matrice di correlazione presenta ancora delle Variabili Latenti più correlate tra loro che con la Soddisfazione ma questa situazione non crea particolari problemi di multicollinearità (condition number inferiore alla soglia limite di 30).

Il modello ottenuto garantisce stabilità e impatti statisticamente significativi al 95% (Figura 3).

⁴ Marotta F., Monaco G., Bianchi S. (2007) "La Costruzione di Indicatori Affidabili nei Modelli CFI di Customer Satisfaction: Il Caso MET.RO", Roma – Università Tor Vergata

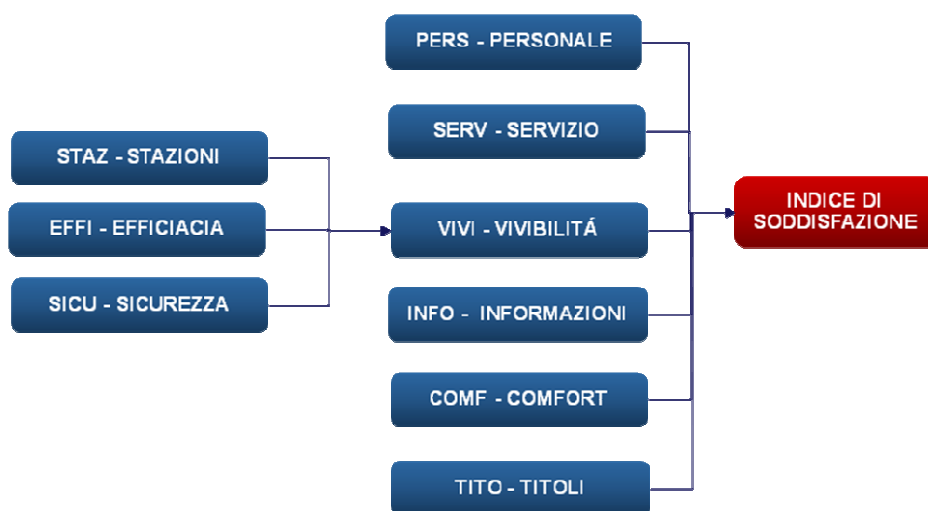


Figura 3

Queste linee sono molto diverse tra loro (per lo stato attuale delle carrozze e delle stazioni, la tratta di viaggio, la tipologia di viaggiatori) e questo sicuramente ha inciso sulla costruzione del modello di Customer Satisfaction. La peculiarità di ciascuna tratta non ha infatti permesso la validazione di una struttura di modello comune ma ha “costretto” alla determinazione di 2 modelli distinti.

Trattandosi di monitoraggio semestrale infatti, si è preferito privilegiare l'affidabilità ottenuta con i 2 modelli all'omogeneità offerta dal modello unico.

Bibliografia

Cronbach L.J. (1951), Coefficient alpha and the internal structure of tests, *Psychometrika*, 16, 297-334.

Fornell C. and F.L. Bookstein (1982), Two structural equation models: LISREL and PLS applied to consumer exit-voice theory, *Journal of Marketing Research*, 19: 440-452.

Fornell C., Larcker D.F. (1982), Evaluating Structural Equation Models with Unobservable Variables and Measurements Errors, *Journal of Marketing Research*, 18: 39-50.

Fornell C. and Cha J., 1994 "Partial Least Square", Richard Bagozzi (Ed.), *Advanced Methods of Marketing*, pp.52-78.

Lewis-Beck M., Bryman, A. Futing T. (2003), Partial Least Squares (PLS) Regression, *Encyclopedia of Social Sciences Research Methods*.

Lynd D. Bacon (1999), Using LISREL and PLS to Measure Customer Satisfaction, *Seventh Annual Sawtooth Software Conference*.

Lohmöller J.B. (1989), Latent Variable Path Modeling with Partial Least Squares. *Springer-Verlag*. New York

Jöreskog K.G., Wold H. (1982) The ML and PLS Techniques For Modeling with Latent Variables: Historical and Comparative Aspects, Wold and Jöreskog (Eds.), *Systems Under Indirect Observation: Causality, Structure, Prediction*, 1: 263-270. Amsterdam: North-Holland.

J. Paul Peter (1979), Reliability: A Review of Psychometric Basics and Recent Marketing Practices, *Journal of Marketing Research*, Vol.XVI pp.6-17.

J. Paul Peter (1981), Construct Validity: A Review of Basic Issues and Marketing Practices, *Journal of Marketing Research*, Vol.XVIII pp.133-145.

Richard P. Bagozzi, Youjae Yi, Surrenda Singh (1991), On use of structural equation models in experimental designs: two extentions, *Journal of Research in Marcheting*, 8: 125-140.

Roger J.L (1999), Reliability and Validity: Meaning and Measurement, *Annual Meeting of the Society for Academic Emergency Medicine (SAEM)*, Boston, Massachussets.

Wold H. (1985), Partial least squares, Kotz and Johnson (Eds), *Encyclopedia of Statistical Sciences*, 6:581-591.