

METODI DI QUANTIFICAZIONE PER LE VARIABILI QUALITATIVE ORDINALI

Daniela Nappo, M.G. Grassia

*Dipartimento di Matematica e Statistica, Università “Federico II” di Napoli
email: daniela.nappo@unina.it, mgrassia@unina.it*

Riassunto

In letteratura è ampiamente affrontato il problema della quantificazione delle variabili qualitative ordinali, e diverse sono le proposte che spaziano da una semplice ricodifica, ad una trasformazione integrale della scala di misura.

Il problema della quantificazione di suddette variabili è inquadrato nell’ambito dei Modelli ad Equazioni Strutturali (MES), e nello specifico, si fa riferimento all’algoritmo del PLS-PM, classicamente utilizzato per lo studio delle reazioni tra variabili manifeste e latenti di un modello.

Saranno presentate e discusse le proposte presenti in letteratura baste su procedure di quantificazione ottimali.

1. INTRODUZIONE

Le indagini statistiche si differenziano per la natura dell’obiettivo e per il tipo di variabili che si intende misurare, nonché per il metodo di svolgimento dell’indagine stessa. Di conseguenza si distingue tra le tecniche di analisi, tra metodi per lo studio delle variabili quantitative e metodi per lo studio delle variabili qualitative. L’obiettivo del ricercatore statistico è quello di poter trarre delle conclusioni e di poter “supportare delle decisioni” con i risultati ottenuti, identificando una qualche relazione tra le variabili, che se, di natura quantitativa, consiste nello studiare relazioni di causa-effetto, se di tipo qualitativo, relazioni di associazione.

L’ultimo secolo è stato caratterizzato dallo sviluppo di metodologie che consentono di studiare, simultaneamente, sia le relazioni tra variabili quantitative, sia tra le qualitative, ma soprattutto l’interazione tra le due tipologie di variabili, come ad esempio l’Analisi Fattoriale Multipla (AFM), per la quale è stata creata un’estensione al trattamento di tabelle miste di dati (Analisi Fattoriale Multipla per Dati Misti, Pagés 2004), o l’Analisi in Componenti Principali (ACP) che attraverso gli Alternating Least Squares è stato possibile estendere al caso di variabili

qualitative. I due metodi sopra citati pongono lo sviluppo della metodologia su due basi differenti: il primo pondera le variabili qualitative, mentre il secondo opera una quantificazione delle variabili, attraverso una procedura di optimal scaling. Un interrogativo che, a nostro avviso, è necessario porsi, riguarda la ragione che spinge un ricercatore ad operare una quantificazione delle variabili qualitative.

Ma soprattutto cosa significa quantificare una variabile qualitativa ordinale, e come interpretarla dopo la quantificazione?

La quantificazione, seppur contestabile, è un procedimento che naturalmente viene effettuato, ogni qualvolta si ha una variabile qualitativa ordinabile: ad una variabile qualitativa, espressa su di una scala ordinale, sono associati valori numerici interi equi distanziati, con la caratteristica che non possiedono le proprietà numeriche, proprie di una scala di rapporto, ma sono solamente ordinate tra loro. Normalmente si assume la continuità di tali variabili, che in realtà non esiste, rendendo più agevole l'applicazione di talune metodologie.

Classicamente questa procedura è seguita nel PLS-PM, che come ben noto, consente di stimare modelli in cui sono espresse una serie di equazioni, come ad esempio nel caso della valutazione della Customer Satisfaction. Gli indicatori usati, in tale contesto, misurano il grado di disaccordo o di accordo dei soggetti rispetto ad una serie di item, utilizzando una scala di misura composta di 10 valori (1 corrisponde al massimo disaccordo, mentre 10 al massimo accordo con la tematica proposta nell'item). Il problema si pone maggiormente allorquando si hanno delle variabili qualitative nominali: in questo caso non è possibile fare una trasformazione su scala, in quanto non è consentito stabilire alcun ordinamento tra le categorie delle variabili. In tutti quei casi, in cui l'obiettivo è la stima delle relazioni causali ed è necessario il loro inserimento nel modello, per ottenere risultati più completi ed esaustivi, è necessario procedere con una quantificazione diversa. La quantificazione, nel caso di variabili ordinali, produce un cambiamento di scala, o meglio un' "estensione" della scala dei valori assunti dalla variabile originaria, consentendo l'applicazione delle metodologie proprie dell'analisi quantitativa; nel caso di variabili nominali associa alle categorie della variabile, mediante una procedura ottimale, dei valori che godono delle proprietà numeriche, ed in quanto tali, assoggettabili alle tecniche statistiche quantitative. In particolare nel caso di variabili ordinali, vi sono delle tecniche di quantificazione che mantengono l'ordinamento originario delle categorie anche dopo il processo di quantificazione. L'interpretazione dei risultati ottenuti si diversifica a seconda della tecnica utilizzata per la quantificazione. Nei prossimi paragrafi saranno brevemente presentate le scale di misura, per poi procedere con la discussione di diverse tecniche di quantificazione, alcune delle quali sviluppate nel contesto del PLS-PM.

2. LE SCALE DI MISURA

La conoscenza di un fenomeno viene ottenuta attraverso la sua misurazione, ovvero mediante l'assegnazione di valori numerici (linguaggio matematico) alle caratteristiche, proprietà ed attributi di oggetti, secondo regole predefinite.

Tuttavia, non esistono procedure ben definite per la misurazione di grandezze non tangibili come l'estetica, il gusto, l'odore, il comportamento, la soddisfazione, la percezione della qualità ecc. [Finkelstein, 1982].

Per questa ragione il concetto di misurazione ha subito nel tempo numerose evoluzioni e ridefinizioni, fino a comprendere anche le procedure di misura nel campo delle scienze sociali e cognitive, dove tali tecniche sono utilizzate per misurare gli atteggiamenti degli individui.

Le variabili prodotte dalla tecnica delle scale non possono essere, però, considerate pienamente cardinali, perché scaturiscono da dimensioni sottostanti immaginate come proprietà continue non misurabili, da qui il nome di variabili quasi-cardinali.

Nell'ambito dei Modelli ad Equazioni Strutturali, ed in particolare nel PLS-PM, si ha la necessità di misurare l'attitudine rispetto a fenomeni complessi, non osservabili in quanto intangibili, e definiti "multidimensionali", tanto da essere innato il ricorso all'utilizzo di scale di misura, come quella di Likert, che ben si presta a misurare concetti multidimensionali.

La scala di Likert è composta da una serie di affermazioni (le domande di un questionario) per ognuna delle quali l'intervistato deve dire se, e in che misura, è d'accordo, attribuendo un punteggio complessivo attraverso la somma di punteggi attribuiti a singole prove (scala additiva).

I vantaggi della scala Likert consistono nella sua semplicità e applicabilità, mentre i suoi svantaggi sono il fatto che i suoi elementi vengono trattati come scale cardinali pur essendo ordinali (a parziale autonomia semantica), la mancata riproducibilità (dal punteggio della scala non è possibile risalire alle risposte delle singole domande) e il fatto che il punteggio finale non rappresenta una variabile cardinale.

Un'alternativa è costituita dalla scala di Guttman che consiste in una successione di elementi, aventi difficoltà crescente, in modo che chi ha risposto affermativamente ad una certa domanda deve aver risposto affermativamente anche a quelle che la precedono nella scala di difficoltà. In questo modo, se gli elementi della scala sono perfettamente scalati, solo alcune sequenze di risposte sono possibili; inoltre dal risultato finale è possibile risalire alle risposte date dal soggetto ai singoli elementi della scala (riproducibilità). Questa tecnica prevede solo elementi dicotomici, contraddistinti con i valori 0 e 1.

I problemi della scala di Guttman consistono nel fatto che il punteggio finale è ancora una variabile ordinale; si tratta di una tecnica applicabile solo ad atteggiamenti ben definiti e scalabili, pertanto non si presta allo studio della Customer Satisfaction, dove è necessario spaziare tra le opinioni degli individui intervistati.

Una scala che nettamente si discosta dalle precedenti per la sua natura probabilistica è quella di Rash, con la quale si suppone che la probabilità di dare una certa risposta ad un dato elemento non è solo 0 o 1, ma si colloca tra questi due estremi.

Questa impostazione presuppone un modello di relazione tra posizione del soggetto sul continuum e la probabilità di risposta ad un determinato elemento della scala, che viene chiamata traccia. La traccia è una curva ad "S", esprimibile attraverso la funzione logistica, che descrive la probabilità di rispondere affermativamente ad un certo elemento a seconda della posizione dell'individuo sul continuum sottostante. I vantaggi di questo modello sono due: esso è una descrizione molto più adeguata ai reali meccanismi che generano le risposte rispetto al modello deterministico e le variabili prodotte da questo modello sono variabili cardinali.

3. L'APPROCCIO OPTIMAL SCALING

In un contesto multivariato, l'optimal scaling può essere considerato come una tecnica che consente di assegnare valori numerici alle categorie di variabili qualitative, attraverso la ricerca dei valori di scala, tali da rendere massima la correlazione tra le variabili. Lo scaling di variabili categoriali (nominali, ordinali) ha lo scopo di assegnare un punteggio (scaling) ai soggetti o individui (score) e/o alle variabili o item (categorie), in modo da trasformare le variabili originali in variabili numeriche, secondo un criterio ottimale (optimal).

Sono molteplici le tecniche statistiche che possono essere considerate di optimal scaling, differenziandosi a seconda se sono metodi ricavati da modello con consistenza interna, metodi ricavati da una funzione obiettivo, o metodi in cui simultaneamente si ottimizzano le quantificazioni e i parametri del modello definito.

In particolare, appartengono all'ultima categoria di metodi di optimal scaling, gli ALternating Least Squares, che, mediante la definizione di una funzione obiettivo espressa rispetto a più parametri, consentono di ottenere le migliori quantificazioni e i parametri ottimali del modello.

3.1 GLI ALTERNATING LEAST SQUARES

Gli Alternating Least Squares (ALS) sono algoritmi nati con l'obiettivo di estendere le tecniche, sviluppate per lo studio delle relazioni tra variabili quantitative,

nell'ambito dell'analisi multivariata, ai dati di natura qualitativa. Le tecniche, quali l'Analisi in Componenti Principali, il Modello di Regressione, l'Analisi delle Correlazioni Canoniche, sono state estese allo studio delle relazioni tra variabili qualitative e tra blocchi di matrici miste. Ciò ha consentito il pieno utilizzo delle informazioni espresse dalle variabili qualitative, implicando un arricchimento dei risultati ottenuti in termini di maggiore informazione ricavata dalle analisi effettuate sull'osservato. Gli algoritmi, che appartengono alla categoria degli ALS, si basano sulla minimizzazione di una loss function, definita rispetto a due parametri: uno rappresenta il parametro del modello da sviluppare, mentre l'altro è il vettore di scaling per le variabili qualitative. La loss function si diversifica, pertanto, in base all'obiettivo del metodo utilizzato.

Gli ALS sono caratterizzati dall'alternanza di due passi: nel primo si ottimizza il parametro di scaling, mantenendo fisso l'altro, nel secondo, invece, fissando il parametro di scaling ottenuto al passo precedente, si minimizza la funzione rispetto all'altro parametro. I passi sono alternati fino alla convergenza. I risultati che si ottengono sono equivalenti a quelli delle classiche tecniche di Analisi Multivariata, con l'aggiunta di un'informazione data dalle rappresentazioni grafiche delle quantificazioni, per ciascuna variabile qualitativa. La distinzione tra i diversi algoritmi è dovuta non solo alla tipologia del metodo sviluppato, ma anche alla tipologia di variabili su cui sono applicabili tali algoritmi.

Tra i diversi algoritmi che fanno parte della famiglia degli ALS annoveriamo il PRINCALS [de Leeuw et al. 1980] e il MORALS [Young et al 1976], i quali sono applicabili a qualsiasi tipologia di dato, ma sviluppano due analisi diverse: il primo consente di ottenere una sintesi degli indicatori nell'ottica dell'ACP, ma con variabili qualitative attive, mentre il secondo sviluppa un modello di regressione multipla con variabili o solo qualitative, o solo quantitative o miste.

Il PRINCALS ha come obiettivo di ottenere la migliore quantificazione per le variabili qualitative e la migliore approssimazione della matrice iniziale (l'algoritmo procede con la quantificazione unicamente per le variabili dichiarate non numeriche, per poi procedere con la sintesi globale e la definizione dei fattori includendo tutte le variabili). Si minimizza, pertanto, una funzione di perdita uguale alla somma dei quadrati delle differenze tra la matrice di approssimazione (cioè la matrice più prossima a quella originaria) e la matrice iniziale in codifica disgiuntiva completa, moltiplicata per il vettore di quantificazione:

$$\sigma_j(X, Y) = \frac{1}{m} \sum_{j=1}^m SSQ(X - G_j Y_j)$$

dove $AVE(q_j) = 0$ and $SSQ(q_j) = 1$.

Il PRINCALS determina il vettore di optimal scaling per le variabili qualitative, v_{j_s} , e la matrice X , di dimensioni $n \times p$ ridotte rispetto alla matrice iniziale, che meglio approssima la matrice dei dati osservati, minimizzando la loss function. La soluzione del problema si basa sulla scomposizione in valori singolari della quantità $G_j(G_j'G_j)^{-1/2}$.

L'algoritmo è iterato fino alla convergenza, che si ottiene quando è raggiunto il minimo della funzione. Inoltre la convergenza è assicurata, in quanto la quantità residua diminuisce ad ogni passo [de Leeuw, 1973].

L'algoritmo MORALS consente di stimare un modello di regressione in presenza di variabili qualitative e quantitative. È basato sulla minimizzazione di una funzione quadratica rispetto a tre parametri: il vettore di scaling y_j , il vettore t delle categorie delle variabili e il vettore β dei parametri di regressione:

$$\min_{\beta, t, y_j} SSQ(G_z t - G_y \beta)$$

dove G_y e G_{zt} sono, rispettivamente, la matrice indicatore delle k variabili indipendenti e le variabili dipendenti. Gli obiettivi di questa tecnica sono, pertanto, ottenere la migliore quantificazione delle variabili qualitative e i coefficienti di regressione ottimali.

Ottenuta la prima stima dei vettori z_{os} e x_{os} (os è l'acronimo di optimal scaling), i parametri della regressione multipla sono calcolati ottimizzando la loss function rispetto a β , fissando, cioè, il vettore di scaling ottenuto al passo precedente, e usando come matrice, la matrice nuova ottenuta, al passo precedente, dal prodotto della matrice indicatore per il vettore di scaling. I due passi vengono iterati fino alla convergenza, ovvero, come nel caso precedente, quando si raggiunge il minimo della loss function. Entrambe le procedure impongono vincoli sulla funzione di quantificazione, in particolare per dati ordinali le quantificazioni sono vincolate ad appartenere al cono convesso generato dalle possibili soluzioni.

Attraverso l'utilizzo di questi due algoritmi è pertanto possibile ottenere due stime diverse degli indicatori qualitativi: il primo ci consente di ottenere una sintesi nell'ottica di un'Analisi in Componenti Principali, mentre il secondo ci consente di stimare i legami causali tra le variabili, sia esse qualitative sia esse quantitative.

Il vantaggio di utilizzare queste tecniche, rispetto ad altre, risiede nel fatto che esse sono algoritmi non parametrici, mantengono la medesima filosofia delle tecniche di cui ne sono l'estensione (a parte per il caso di MORALS che ottiene i parametri di regressione senza ipotesi distribuzionali, scostandosi nettamente dall'approccio OLS), e pertanto costituiscono un valido supporto per tutti quei casi

in cui si ha una matrice di dati misti, su cui è supposto un modello, e non perfette (multicollinearità, dati mancanti). Come abbiamo potuto vedere attraverso gli ALS si riesce a dare risposta a due distinte esigenze simultaneamente: la prima è quella di poter quantificare, e la seconda è quella di poter stimare un modello o analizzare la struttura delle relazioni tra i concetti latenti.

4. UNA PROPOSTA PER LA STIMA DELLE VARIABILI QUALITATIVE NEL PLS-PM

Il PLS-PM è una metodologia che, come noto, consente di stimare un modello fattoriale caratterizzato dalla definizione di una serie di legami causali tra variabili latenti (concetti non direttamente misurabili) e tra queste ultime e le variabili manifeste, di cui ne sono la misurazione empirica.

La procedura consiste nello sviluppo di regressioni semplici e multiple, sia tra i diversi costrutti latenti, sia in ciascun blocco di variabili latenti, ottenendo come risultato, la stima degli impatti dei costrutti latenti sull'outcome del modello, la stima dei legami all'interno di ciascun blocco, tra la latente e le variabili manifeste, ad essa legate, e una stima finale delle variabili latenti per ciascun individuo. L'assenza delle ipotesi distribuzionali, la possibilità di lavorare con valori mancanti, la possibilità di stimare il modello in presenza di multicollinearità, la possibilità di stimare il modello con matrici dove il numero di individui è inferiore a quello delle variabili, ha reso questa tecnica molto applicabile ai problemi reali e estremamente versatile.

Appare evidente, tuttavia, che tale metodologia è nata per lo studio delle relazioni tra variabili quantitative e non si presta bene allo studio di relazioni in presenza di variabili qualitative.

Tuttavia nei casi in cui è necessario introdurre indicatori qualitativi, essi vengono inseriti in codifica disgiuntiva completa, e successivamente considerati come indicatori quantitativi, applicando le consuete tecniche quantitative previste dall'algoritmo per la stima del modello (il modello di regressione semplice o multiplo, mediante la metodologia OLS).

Attuare questo tipo di trasformazione non consente, sempre, di poter ottenere dei risultati validi ai fini interpretativi ed esplicativi dei legami stimati. Può, infatti, accadere che gli indicatori qualitativi inseriti debbano essere eliminati dal modello in quanto presentano coefficienti di regressione prossimi allo zero, o intervalli di confidenza non escludenti lo zero; inoltre la loro scala di misura è estremamente piccola, tanto da influenzare la stima finale dei costrutti latenti e gli indici di validità del modello, potrebbero assumere valori molto bassi.

Nonostante sia un problema reale e nonostante sia sempre più frequente il ricorso a tale metodologia per lo studio di fenomeni sociali, caratterizzati da una forte presenza di tali indicatori, in letteratura esistono pochissime proposte, atte a risolvere questo problema di stima.

Una valida alternativa alla post codifica è data da Jakobowicz e Derquenne [2007], che basano il loro metodo su di una trasformazione iniziale delle variabili, attraverso una quantificazione “ottimale”, rispetto ad un criterio.

Gli autori modificano il primo passo dell’algoritmo in base alla natura delle variabili manifeste (nominali o ordinali), introducendo il concetto di variabile di riferimento, che rappresenta la stima iniziale della variabile latente.

La variabile di riferimento è una variabile manifesta, che si suppone possa meglio rappresentare il concetto latente, associata a uno qualsiasi dei blocchi latenti, presenti nel modello, legato, da un nesso causale, al *j*-esimo blocco latente. Il vettore dei pesi iniziali sarà uguale a (Lohmöller dimostra che qualsiasi sia il vettore iniziale utilizzato, l’algoritmo del PLS-PM converge):

$$w_{jh}^0 = \text{cov}(x_{jh}, x_{i1})$$

dove x_{i1} è la variabile di riferimento scelta tra i blocchi associati al blocco *j*.

Gli autori, individuata la variabile di riferimento, propongono una serie di metodologie statistiche, note in letteratura, che si differenziano in base al tipo di variabile di riferimento x_{i1} e di variabile manifesta x_{jh} .

In particolare se la variabile di riferimento è numerica e la manifesta è categorica, sarà utilizzato un modello ANOVA, viceversa si sceglierà un modello logistico. Se entrambe le variabili sono categoriche sarà usato il modello logistico, mentre se la variabile di riferimento è nominale con *r* categorie e la manifesta è numerica, una regressione logistica politomica; se la variabile di riferimento è nominale e la manifesta è categorica sarà utilizzato un modello logistico.

In tal modo solo il primo step dell’algoritmo è modificato, in modo tale che quando entrambe le variabili sono quantitative, l’algoritmo non cambia.

Tuttavia la proposta di Jakobowicz e Derquenne, anche se appare maggiormente in sintonia con la problematica affrontata, a nostro parere non è in linea con la filosofia del PLS-PM, in quanto utilizzano metodi statistici per ottenere una codifica a priori delle variabili latenti, senza tener conto del ruolo predittivo che esse hanno nei confronti della variabile latente a cui sono associate, e soprattutto se i metodi basati su ipotesi distribuzionali, che nel caso si hanno matrici di dati incomplete, possono comportare dei problemi per la stima.

5. CONCLUSIONI

L'articolo è una breve rassegna delle metodologie usate per la codifica ed il trattamento delle variabili qualitative, inquadrando il problema nel contesto dei Modelli ad Equazioni Strutturali, ed in particolare nell'ambito del PLS-PM.

Per quanto concerne le scale di misura è ancora aperto e discusso il dibattito circa il loro utilizzo e la loro efficacia nella rilevazione delle attitudini dei soggetti. Un problema ancora non risolto riguarda, inoltre, il numero di "livelli" da utilizzare per la misurazione degli atteggiamenti: per default si usano dieci livelli, ma è corretto e perché utilizzare una scala di dieci valori? Perché non usarne una di tre valori o di cinque? E soprattutto, come poter stabilire scientificamente il numero di livelli? Sono domande a cui in letteratura si cerca di fornire una corretta risposta [Cox, 1980].

Tuttavia negli studi, come quelli che valutano la Customer Satisfaction, è consueto l'utilizzo di scale di dieci valori, in quanto le stime dei parametri del modello sono più stabili e significative, e le variabili avrebbero, in tal modo, un range di variazione più elevato, rendendo più giusta, anche se non corretta, l'assunzione di continuità.

In riferimento all'approccio Path Modeling, è necessario sottolineare il numero esiguo di proposte alternative alla consueta trasformazione di scala che viene adoperata, che consentano di stimare un generico modello in presenza di variabili qualitative.

Sembra più corretto poter effettuare una quantificazione interattiva, piuttosto che effettuarne una a priori senza tenere in considerazione i legami variabili presenti nel modello.

Diverse sono ancora le questioni aperte e che hanno bisogno di una risposta, nell'ambito dello studio del mondo qualitativo, che sempre più spesso, oggi, sconfina con quello quantitativo, per le sue innumerevoli caratteristiche ed apporto di conoscenza. Il tentativo di estendere le attuali tecniche in uso, per lo studio quantitativo delle relazioni tra le variabili, al caso qualitativo, indica che il ruolo delle variabili qualitative non è più quello di determinare una classificazione a priori tra i soggetti, ma sta assumendo un carattere deterministico per lo studio dei legami causali, e si è maggiormente interessati allo studio dell'impatto che queste variabili possono avere su altri anche di natura quantitativa.

Pertanto è lecito quantificare, ma è necessario farlo in modo da poter rendere quanto più possibile continue le variabili trasformate, ottenendo delle stime significative per le metodologie sviluppate.

RIFERIMENTI BIBLIOGRAFICI

- E.P.COX (1980) 'The optimal number of response alternatives for a scale: a review', *Journal of Marketing Research*, **17**(4), 407-422.
- J. DE LEEUW (1980), 'HOMALS and PRINCALS some generalizations of Principal Components Analysis', *Data Analysis and Informatics*, E. Diday et al. (eds.), North-Holland Publishing Company, pp. 231-241.
- DE LEEUW, J., YOUNG, F. W., TAKANE, Y. (1976) 'Additive Structure in Qualitative Data: An Alternating Least Squares Method With Optimal Scaling Features', *Psychometrika*, **41**, 471-504.
- H. HWANG, Y. TAKANE (2004), 'Generalized structured component analysis', *Psychometrika*, **69** (1), pp. 81-99.
- E. JAKOBOWICZ, C. DERQUENNE (2007), 'A modified PLS Path Modeling algorithm handling reflective categorical variables and a new model building strategy', *Computational Statistics and Data Analysis*, **51** (8), pp. 3666-3678.
- J. B. KRUSKAL, R.N. SHEPARD (1974), 'A nonmetric variety of linear factor analysis', *Psychometrika*, **39** (2), pp. 123-157.
- A. GIFI (1981), *Nonlinear multivariate analysis*, Department of data Theory, University of Leiden, The Netherlands.
- LAURO C., NAPPO D., GRASSIA M.G., MIELE R. (2008), 'Methods of quantification for qualitative manifest variables in PLS-PM', Relazione invitata all' International Conference on Multivariate Statistical Modelling & High Dimensional Data Mining, 19-23 Giugno, Kayseri, Turchia.
- P.G. LOVAGLIO (1997), 'Un Algoritmo di regressione multipla con dati categoriali', *Quaderni di Statistica e Matematica Applicata alle Scienze Economico-Sociali*, **19** (3), pp. 281-291, Trento.
- P.G. LOVAGLIO (2002), 'La stima di Variabili Latenti da variabili osservate miste', *Statistica*, Bologna, **LXII** (2), pp. 203-213.
- B. MUTHÉN (1984), 'A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators', *Psychometrika*, **49** (1), pp. 115-132.
- G. SAPORTA (1990), 'Simultaneous analysis of qualitative and quantitative data', Proceedings of the 35th Scientific Meeting of the Italian Statistical Society, pp. 63-72
- M. TENENHAUS, F.W. YOUNG (1985), 'An analysis and synthesis of multiple correspondence analysis, optimal scaling, dual scaling, homogeneity analysis and other methods for quantifying categorical multivariate data', *Psychometrika*, **50** (1), pp. 91-119.
- WOLD H. (1985) 'Partial Least Square's', *Encyclopedia of Statistical Sciences*, John Wiley & Sons, New York, S. Kotz & N.L. Johnson (eds.), **6**, 581-591.
- F. W. YOUNG, J. DE LEEUW, Y. TAKANE (1976), 'Regression with qualitative and quantitative variables: an Alternating Least Square method with optimal scaling features', *Psychometrika*, **41**(4), pp. 505-529.
- F.W. YOUNG (1981), 'Quantitative analysis of qualitative data', *Psychometrika*, **46** (4), pp. 357-388.

METHODS OF QUANTIFICATION FOR THE QUALITATIVE VARIABLES IN THE PLS-PM

Summary

In the literature is largely addressed the problem of quantification of qualitative variables, and different proposals, ranging from a simple re-up to a transformation of the full scale of measurement of variables. In this work we discuss some techniques of quantification, used in the Structural Equations (SEM). The SEM allow simultaneous estimation of several causal links between different variables, in particular, we treat the problem of referring to the Partial Least Squares Path Modeling (PLS-PM) for which we present some proposals made in Litterature, based on algorithms of optimal quantification.