



UNIVERSITÀ
CATTOLICA
del Sacro Cuore

Convegno nazionale

**Il valore della statistica per le imprese e
la società.**

Ricerche di opinione e di mercato

Milano, 14-15 febbraio 2013

Contributi spontanei

Sessioni:

14 Febbraio 15.30-16.30

15 Febbraio 10.00-11.00

15 Febbraio 11.30-12.30

(In ordine alfabetico di autore)



A statistical analysis of consumer behaviour in culture through the experience of reading in young people

Rossella Berni, Claudio Becagli and Silvia Ranfagni*¹

In this paper we focus on a survey data on reading that has involved young students (from 15 to 18 years old) of schools located in Florence. The aim of the study is to investigate the relation between the drivers impacting on reading motivations, the decisional processes and the correlated consumer behaviour. More precisely, we deal with personal, familiar and social aspects that generate specific consumer's choices by considering overall several characteristics such as the kinds of school, the cultural level of family, the way of spending free time and the relation with technology. To this end, a particular attention is devoted to the analysis of the development of reading, thanks to the new technologies, also considering the daily amount of time spent with them.

All the different drivers of reading are related with both the decisional processes answering to emotional, experiential and rational dimensions and with consumer behaviour expressing in terms of actions pre and post-buying, purchase place and consumption moments. The resulting attitude of consumer toward reading is studied through statistical models by highlighting consumer's preferences in terms of "read books" and by also including specific methodological problems due to the heterogeneity of the student. The questionnaire is supplied to a sample of about 900 students; the items included in the protocol of analysis have been defined through focus groups organized in separate groups and involving students, teachers and librarians.

Key-words: statistical modelling, consumer behaviour, cultural consumptions

¹ **Rossella Berni**, Department of Statistics "G.Parenti", University of Florence, Viale Morgagni 59, 50134 Florence, Italy. Tel. +39 0554237211; fax +39 0554223560; e-mail: berni@ds.unifi.it.

Claudio Becagli, Department of Business Sciences, University of Florence, Via delle Pandette 9, 50127 Florence, Italy. Tel. +39 0554374688; e-mail: claudio.becagli@unifi.it.

Silvia Ranfagni, Department of Business Sciences, University of Florence, Via delle Pandette 9, 50127 Florence, Italy. Tel. +39 0554374688; e-mail: silvia.ranfagni@unifi.it.

Livello d'informazione della cittadinanza fiorentina in merito alla realizzazione del passante dell'alta velocità ferroviaria

Bruno Bertaccini*, Giovanni Carta, Francesco Polverini*****

**Dipartimento di Statistica “G. Parenti”, Università degli Studi di Firenze;*

***Comune di Firenze; ***VALMON srl, Firenze*

Per avviare una puntuale e precisa messa in opera del Piano di comunicazione e informazione concernente la realizzazione dei lavori per l'Alta Velocità ferroviaria del nodo di Firenze, il Comitato di Garanzia per l'informazione e la comunicazione del Nodo Alta Velocità di Firenze ha ritenuto fondamentale commissionare un'indagine demoscopica locale finalizzata a comprendere lo stato di informazione dei cittadini fiorentini in merito alla realizzazione dell'opera. Con l'indagine si è voluto anche sondare il livello di preoccupazione della cittadinanza in merito ad alcuni aspetti legati alle fasi realizzative dell'opera nonché le opinioni che questa si è fatta in merito al rapporto costi/benefici derivanti dalla realizzazione dell'opera stessa. Le informazioni per raggiungere gli obiettivi sono state raccolte mediante la somministrazione di un questionario ad un campione casuale stratificato di residenti nel Comune di Firenze in età maggiorenne: è stata adottata una stratificazione per quartiere di residenza (5 quartieri), genere e fascia d'età (18 – 35; 36 – 65; oltre 65 anni), per un totale di 30 strati. L'indagine è stata condotta nei mesi di maggio e giugno 2012, in una fase in cui era previsto l'imminente avvio dei lavori di realizzazione del sottoattraversamento. Sono state realizzate 2446 interviste, di cui 1277 tramite somministrazione di questionario in formato elettronico con tecnologia CATI (*Computer Aided Telephone Interviewing*) / CAWI (*Computer Aided Web Interviewing*) e le rimanenti mediante interviste dirette condotte sul territorio comunale, dedicando particolare attenzione ai residenti nelle zone prospicienti al tracciato dei binari in fase di realizzazione, ovvero alle zone interessate dalle attività di trivellazione. Le variabili rilevate dal questionario sono in gran parte di tipo percentuale (es. percentuale di cittadini a conoscenza del progetto, di coloro che lo ritengono un intervento utile, ecc.); per queste variabili, la dimensione campionaria consente di contenere l'errore di stima intorno ai due punti percentuali (con probabilità pari a 0,95) nel caso di un campionamento di tipo casuale semplice; nello specifico pertanto l'effettivo errore di stima potrebbe essere addirittura inferiore a causa del particolare piano di campionamento adottato.

Data la delicatezza del tema oggetto del sondaggio (contenente notevoli risvolti di natura politica sia a livello locale che nazionale), la ricerca presenta alcuni spunti innovativi per quanto riguarda il disegno d'indagine adottato che sfrutta congiuntamente diverse tecniche di somministrazione del questionario (CATI, CAWI e PAPI) con lo scopo di rendere il più possibile simili la popolazione “target”, il “frame” e la popolazione campionata.
L'indagine è stata affidata a VALMON s.r.l., spin-off partecipato dell'Università degli Studi di Firenze, recentemente costituito.

La performance economica d'impresa: un'analisi longitudinale sul sistema imprese italiano

Matilde Bini*, Leopoldo Nascia*, Alessandro Zeli**

**Università Europea di Roma, **Istat, Roma*

Il ruolo degli imprenditori e dei manager, le capacità della forze lavoro, le attività in investimenti e le strategie di internazionalizzazione sono elementi fondamentali in un'impresa di beni e/o di servizi per il suo successo sul mercato. Questo insieme eterogeneo di competenze che viene definito in letteratura come "firms' managerial practise", è strettamente legato oltre che al capitale umano, fisico, alle capacità organizzative e imprenditoriali, anche al livello di produttività delle imprese, e gioca un ruolo sempre più importante nei mercati altamente competitivi. Lo scopo del presente lavoro è di analizzare mediante un modello multilivello longitudinale, la performance delle imprese italiane nel periodo 1989 – 2007 sulla base di alcune determinanti come le caratteristiche (capitale umano e fisico) delle strutture imprenditoriali, l'appartenenza ai settori, gli indicatori di innovazione, di capacità organizzative. Tutte queste variabili sono un patrimonio informativo di un unico e originale data base denominato Database Integrato (DBI), costituito come integrazione di più fonti ufficiali ISTAT sulle imprese: indagine censuaria sulle imprese (SCI), indagine campionaria sulle piccole e medie imprese (PMI), i bilanci delle società di capitale. Nel DBI le informazioni cross section sono ripetute nel tempo e si riferiscono a una grande parte delle imprese italiane con 20 addetti ed oltre nel periodo di riferimento; queste peculiarità rendono il DBI un potente strumento di analisi della dinamica temporale della performance delle imprese, confrontabile inoltre con sistemi imprenditoriali europei.

Customer satisfaction surveys also intended to gain information on the enterprise core business enlargement.

A case history in banking services.

Luigi Bollani*, Enrico Borgo, Anna Claudia Pellicelli*****

Dipartimento di Scienze Economico-Sociali e Matematico-Statistiche, Università di Torino; **Cassa di Risparmio di Asti; *Dipartimento di Management – Università di Torino*

The authors were involved for about twelve years in several customer satisfaction surveys referred to the same credit enterprise, located in the north west of Italy. This paper focuses on the way in which the bank collected client information using, in a conjoint way, customer satisfaction surveys and consumer behaviour/needs analyses.

A particular point of interest is also directed to the implementation of the management plan, based on the service quality and on the service extension following the client interests and needs. In this case a particular position is obtained by insurance and health assistance markets.

In the past twelve years, the most important customer surveys were carried out in four principal moments, separated from each other by three years (in 2003, 2006, 2009 and 2012).

In all these cases a sampling plan was designed, taking also care of the links among succeeding periods; different questionnaires were prepared and responses were collected by phone, mail or web. A statistical analysis was carried out using also multivariate techniques to jointly consider different aspects of marketing interest.

Looking back at the main results, a critical report of the implementation of marketing activities in the bank is written, also considering the use and the consequences of the marketing levers in different situations of market and products development.

The way to choose, determine and consolidate the marketing strategies is shown. The innovation policies of a local banking enterprise able to reach a leading position in the extension of services is also presented.

The good positioning on the market, the progressive development and the innovation process for a local enterprise have been greatly facilitated by the consideration of client needs and by the presence throughout the territory.

In particular it is analyzed the relevant role of the joint initiatives for customer satisfaction and for consumer needs analysis, also showing the bank's capability to be flexible and innovative. This capability may be shown in two main points:

- a) to observe and gradually adapt to client needs/requests
- b) to observe the client reactions to the proposal to satisfy new needs, using a bank channel present throughout the territory.

On the first point of view, the instruments to give importance to the client in core business services are shown; on the second point of view, the capability to reach new markets, like insurance and health assistance, is presented.

A survey design to evaluate financial risk behaviour according to individual values, life styles and personality characteristics

Luigi Bollani*, Guido Antonio Rossi*, Ivan Sciascia**

*Dipartimento di Scienze Economico-Sociali e Matematico-Statistiche, Università di Torino:

**Dipartimento di Biotecnologie Molecolari e Scienze per la Salute, Università di Torino

The survey foresees to ascertain the relationship between risk uncertainty attitudes and personality of decision makers. The survey design is organized in the following steps:

1. Questionnaires to be used for the survey. The questionnaire structure considers the following themes:

- a) Asking preference choices between/among financial lotteries. Multiple pairwise choices and full ranking choices are considered. More or less difficult choices are considered too and different amount of conventional money to be won (or loosed) is taken into account.
- b) Survey on human values and life styles.
- c) Determining personality characteristics.
- d) Socio-demographic informationd)

2. Target groups. Groups of students are considered. Other groups of people may also be investigated. For the risk attitude analysis a model is proposed. As utility function to evaluate risk attitude is used a third degree polynomial. This choice allows to transform expressed preferences into comparisons between means, variances and asymmetries, that is between first, second and third moments of the underlying probability distribution, and inequalities between the polynomial coefficients, unknown but in any case determined up to a linear affine transform, and then identified through their ratios (that is considering variables the last two ones: relative weights of second and third moment with respect to the first), up to identifying a limited plane region where those ratios fall.

That allows an easy comparison between different decision makers and even the same decision maker in different epochs, making it immediate to denote if there is attitude affinity by superposing the aforesaid regions and examining their intersection.

A multivariate analysis is carried out to interpret and synthesize each other sections of the questionnaire. For these purpose a principal component analysis is performed on Zuckerman scale values and on a transformation of Rokeach scale values. A cluster analysis, based on pca dimensions, is also performed to distinguish different groups according to their life styles and values. The theory of the Lüscher color test is used to group people with respect to some aspects of their personality. Finally a multiple correspondence analysis is performed to observe the different themes of the analysis all together.

Opinione sulla qualità complessiva di una prestazione

Egidio Cascini

Associazione Italiana Cultura Qualità
e.cascini@alice.it

Il lavoro propone un *indicatore di opinione* sulla qualità complessiva di una prestazione erogata. Si ammette che una prestazione possa essere identificata da n caratteristiche, alla i -esima delle quali può essere attribuito un peso di importanza, γ_i , variabile nell'intervallo 0 - 1, e un voto v_i , variabile nello stesso intervallo. Si ammette che i valori numerici γ_i, v_i , possono essere stabiliti dal fruitore della prestazione, mediante un questionario specifico composto da n domande. L'indicatore è determinato, operando sulle grandezze γ_i, v_i in modo appropriato; più precisamente, vengono costruiti due vettori ad n dimensioni, diciamo *ideale* e *reale*, con componenti i -esime, rispettivamente uguali a γ_i e $\gamma_i v_i$. Sono, quindi, determinate tre grandezze:

$$c_1 = \sqrt{\frac{\gamma_1^2 v_1^2 + \gamma_2^2 v_2^2 + \dots + \gamma_i^2 v_i^2 + \dots + \gamma_n^2 v_n^2}{\gamma_1^2 + \gamma_2^2 + \dots + \gamma_i^2 + \dots + \gamma_n^2}}$$

$$I_q = \frac{\gamma_1^2 v_1 + \gamma_2^2 v_2 + \dots + \gamma_i^2 v_i + \dots + \gamma_n^2 v_n}{\gamma_1^2 + \gamma_2^2 + \dots + \gamma_i^2 + \dots + \gamma_n^2}$$

$$c_2 = 1 - \frac{1}{\vartheta_{max}} \arccos \left(\frac{I_q}{c_1} \right),$$

con

$$\vartheta_{max} = \arccos \frac{\gamma_{min}}{\sqrt{\gamma_1^2 + \gamma_2^2 + \dots + \gamma_i^2 + \dots + \gamma_n^2}}$$

che indicano, nell'ordine: *valutazione complessiva del servizio erogato*, espressa in frazioni di unità, indipendentemente dai pesi assegnati alle singole caratteristiche, *opinione complessiva del servizio erogato*, espressa in frazioni di unità, che tiene conto dei pesi assegnanti a ciascuna caratteristica, *grado di scostamento dai pesi*, espresso in frazioni di unità. Si presenta un modello analogico meccanico che spiega come l'indicatore I_q rappresenti realmente l'*opinione* sulla qualità erogata espressa dal fruitore del servizio. Inoltre, si mostra, attraverso un esempio numerico reale, la possibilità di allestire un piano di miglioramento, in base ai valori c_1, c_2 .

Studio della domanda di energia fotovoltaica: percezione e driver decisionali dell’utente finale

Paola Cerchiello

Università di Pavia

L’Unione Europea ha definito una strategia di riduzione autonoma delle emissioni nocive del 20% entro il 2020, finalità formalizzata nell’ambito della direttiva 2009/28/CE, con specifici indirizzi relativi alle fonti rinnovabili. Considerando i consumi di energia dei 4 settori (industria, trasporto, riscaldamento, elettricità), la produzione di energia elettrica dovrà provenire per almeno il 35-40% da fonti rinnovabili. In Italia, gli impianti fotovoltaici installati e connessi alla rete (grid-connected) hanno raggiunto la capacità complessiva di 16.77 GWp (dicembre 2012), rispetto a 1 GW nel 2010. La ricerca scientifica di base ha dato un contributo fondamentale sul tema degli impianti fotovoltaici, negli aspetti tecnologici e di miglioramento di prestazioni ed efficienza. In base alla nostra conoscenza tuttavia non è stato mai indagato, in termini metodologicamente rigorosi, il cosiddetto utente finale ovvero la domanda. Se gli impianti fotovoltaici e le energie rinnovabili in generale, potranno vincere la sfida del mercato, ciò dipende, non solo dagli aspetti tecnologici e di policy nazionale/europea ma anche dalla capacità di diffusione informativa del tema delle energie rinnovabili e nello specifico del fotovoltaico. Quanto l’utente finale (in particolare il cliente residenziale) è informato e messo in condizioni di poter valutare il costo-opportunità di un impianto fotovoltaico? E’ del tutto evidente che la tematica non è di facile e immediata comprensione per un utente medio, gli aspetti tecnici da valutare sono svariati e l’assenza di una seppur minima conoscenza di base può incidere sfavorevolmente sulla decisione finale. In questo contesto risulta fondamentale eseguire una indagine statistica su base campionaria sia degli utenti che hanno deciso di installare un impianto sia di quelli che non ne posseggono uno, al fine di una valutazione dei fattori di rischio-opportunità che possono condurre ad un esito positivo. Tale studio si basa sulla costruzione di un opportuno questionario di valutazione differenziato in base all’utente (possessore/non possessore impianto) e sulla successiva analisi quantitativa con opportune metodologie statistiche tipiche dell’approccio score-card sviluppato nell’ambito dell’operational risk-management. Lo scopo del lavoro è altresì l’impiego di indici statistici sintetici per l’analisi di dati ordinali, recentemente sviluppati in contesti di risk management (Cerchiello e Giudici 2012(a), Cerchiello e Giudici 2012(b)) che permettano non solo la semplice e immediata comunicazione delle risultanze emerse dall’indagine ma anche di monitorare periodicamente la percezione e il sentimento dell’utente finale in linea con le finalità tipiche dello scorecard approach.

Bibliografia Essenziale

- P.Cerchiello, P. Giudici: Fuzzy methods for variable selection in operational risk management. In Journal of Operational Risk, Vol. 7, No. 4, pp-1-17. 2012(a)
- P.Cerchiello, P. Giudici: On the distribution of functionals of discrete ordinal variables, Statistics and Probability Letters, 82 (2012) 2044–2049 2012(b)

Social Media & Statistica: la nuova frontiera del sondaggio

Andrea Ceron*, Luigi Curini*, Stefano Maria Jacus*, Giuseppe Porro**

**Università di Milano-Bicocca, **Università dell'Insubria*

L'uso dei social media a fini di indagine e previsione del "sentimento" espresso dalla collettività è un fenomeno sempre più diffuso, specie in ambito politico ed elettorale. Il confronto tra metodi statistici di analisi vede due principali contrapposizioni: a)i metodi unsupervised, ampiamente automatizzati e spesso fondati sull'utilizzo di dizionari ontologici, e i metodi supervised, che prevedono un intervento più invasivo del ricercatore; b) metodi che si fondano sull'analisi individuale e metodi che adottano un approccio aggregato. In questo lavoro viene esposta una metodologia supervised di analisi aggregata che consente una ricostruzione approfondita e originale dell'opinione pubblica. Il metodo è applicato all'indagine previsiva dei risultati elettorali in due casi assai diversi: le Presidenziali americane del 2012 e i due turni delle primarie del centro-sinistra. Oltre ad illustrare i vantaggi competitivi del metodo, verranno mostrati in chiave comparativa diversi metodi di Sentiment Analysis di uso più comune.

Pensare globale, parlare locale

Paolo Cordella, Diana Moriniello, Carlo Borghi
SKIM, P.O. Box 29044, 3001 GA Rotterdam, The Netherlands

Per avere successo, una marca o un prodotto hanno bisogno di attirare l'attenzione del consumatore in modo positivo e accattivante. I *claims* sono affermazioni che contengono benefici funzionali ed emotivi di un prodotto e informazioni che servono a convincere il consumatore della rilevanza del prodotto e del bisogno che esso soddisfa. Il valore percepito del prodotto, soprattutto nella fase di lancio, sarà quindi influenzato dalla "bontà" dei *claims*. I *claim studies* sono una tipologia di ricerche di mercato che ha come obiettivo quello di testare l'efficacia dei *claims*. Tra le varie metodologie di esecuzione possibili, l'articolo fa uso del metodo MaxDiff che permette di classificare i *claims* in base alla loro attrattività, chiedendo ai consumatori di compiere dei trade-off tra diversi *claims*. Ciò si traduce in risultati più robusti rispetto a metodi tradizionali come il rating o il ranking, perché si giunge alla stima della "distanza" relativa dell'attrattività tra diverse affermazioni. La procedura di raccolta dei dati si basa inoltre su un modello "discrete choice" che fornisce risultati liberi da vincoli e bias causati dall'uso di scale e punteggi.

Nell'articolo viene presentata l'analisi di 32 studi realizzati da SKIM nel corso degli ultimi 10 anni, per un totale di 854 *claims* testati in diverse categorie di prodotto. L'obiettivo è di identificare le caratteristiche che rendono un *claim* efficace. L'articolo si propone quindi una "quantificazione" dell'efficacia dei *claims*, implementando una meta-analisi nella quale sono studiati i fattori che influenzano l'attrattiva di ogni *claim*. Il risultato è la stesura di una serie di raccomandazioni a carattere generale che permettono una "strategia di comunicazione globale" basata su un metodo scientifico di individuazione delle caratteristiche ottimali che un *claim* deve avere.

L'articolo si propone infine di analizzare uno scenario tipico delle realtà multinazionali: come quantificare un *claim* vincente in diverse lingue e diversi contesti regionali?

In un sistema dove cresce la commercializzazione delle stesse marche a livello internazionale, è importante che le aziende seguano una strategia di comunicazione e posizionamento dei loro prodotti a livello globale. Da un lato le strategie aziendali, basate sul "pensare globale", rendono necessario l'utilizzo di tecniche di analisi aggregata per ottenere risultati generalizzabili in differenti contesti; tuttavia, in tale sistema, numerose iniziative e prodotti falliscono perché non riescono a essere percepiti a livello locale. Diviene dunque fondamentale una considerazione e conoscenza del pensiero che si sviluppa sul territorio. Punto critico per il successo del prodotto è il "parlare locale".

La seconda parte dell'articolo è quindi di tipo esplorativo e si pone l'obiettivo di descrivere due problematiche legate al vincolo linguistico della comunicazione internazionale:

- Da una parte, i contenuti dei *claims* potrebbero non suscitare le stesse reazioni in contesti differenti. La sensibilità verso i simboli, le dimensioni emozionali e persuasive e le altre caratteristiche dei messaggi sono influenzate dalle norme culturali dei singoli paesi.
- Dall'altra, per denotare adeguatamente il significato del messaggio normalmente è necessario ricorrere alla lingua del paese target e la traduzione richiede una forte conoscenza della cultura affinché i richiami culturali, simbolici, evocativi del messaggio non vadano persi o deformati.

Tracciando alcune prospettive di ricerca future, analizzando 4 *claim studies* effettuati in Italia, la domanda centrale diventa quindi come riuscire a fare fronte a queste problematiche e presentare in modo adeguatamente accattivante la proposta comunicativa facendone comprendere esattamente il contenuto.

Importance and quality of service quality dimensions: A model based approach

Federica Cugnata and Silvia Salini
DEMM, Università degli Studi di Milano

The importance that users or customers attach to various services and products is an important part of customer satisfaction surveys. It can be assessed by an analysis of customer satisfaction. In some cases the level of importance is asked explicitly in the survey questionnaire, in other cases it is derived using statistical models.

If the questionnaire includes a question on overall satisfaction, then it is possible to measure the impact each satisfaction levels from individual dimensions on overall satisfaction. There are in the literature some proposals for linking quality and importance. The objective is to understand what are the dimensions with high importance but low perceived quality. These dimensions are primary candidates for focused improvement initiatives. For an analysis of such decision drivers see Kenett and Salini (Chapter 2 on *Modern Analysis of Customer Surveys: with applications using R*, Wiley, 2012). See also Ferrari and Barbiero, 2012; Ferrari and Salini, 2011.

In this study, a questionnaire with explicit variables of importance, for each dimension, and a question on overall satisfaction is considered. The first step will be to verify whether the ranked order of importance obtained by estimating a CUB model (Iannario and Piccolo, 2012) on overall satisfaction using the various dimensions as covariates is consistent with the order of declared importance. Then the estimated importance and the perceived quality, both obtained with CUB model, will be jointly analyzed. This approach will be compared with those reported in the literature.

References

- Ferrari P.A. and Barbiero A. Nonlinear Principal Component Analysis. *Modern Analysis of Customer Surveys: with applications using R*, Kenett and Salini (eds). John Wiley and Sons: Chichester: UK, 2012.
- Ferrari P.A. and Salini S. Complementary use of Rasch models and Nonlinear principal components analysis in the assessment of the opinion of europeans about utilities. *Journal of classification*, 28:53–69, 2011.
- Iannario M. and Piccolo D. CUB Models: Statistical Methods and Empirical Evidence. *Modern Analysis of Customer Surveys: with applications using R*, Kenett and Salini (eds). John Wiley and Sons: Chichester: UK, 2012.
- Kenett R.S. and Salini S. The ABC annual customer satisfaction survey. *Modern Analysis of Customer Surveys: with applications using R*, Kenett and Salini (eds). John Wiley and Sons: Chichester: UK, 2012.

Multilevel models IN Meta-Analysis. An application ON PRE-election Polls Accuracy

Rosario D'Agata, Venera Tomaselli

Department of Political and Social Sciences, University of Catania

In pre-election polls, more than in other surveys, the accuracy of predicted outcome is verifiable. Comparing the expected result of a party with the actual electoral performance, in fact, it is possible to compute a measure of poll predictive accuracy. Some scholars proposed a measure whereby the accuracy of electoral forecast is assessable. Employing such measure, in this paper, we present a meta-analysis of electoral poll accuracy aiming to explain the differences among the polls to identify the potential sources of biases affecting forecasts. In traditional meta-analysis, researchers collect a set of independently conducted studies, with the same object, focusing on each outcome of them.

Following an approach based on random-effects models, we decompose the variance of poll outcomes into two components: the *within* poll variance, as sampling variance, and the *between* poll variance due to the differences across the poll features. In order to explain the differences among poll accuracy measures we carry out a meta-analysis study specifying a ‘special’ multilevel regression model. Unlike the standard multilevel model, in fact, we do not have raw data of interviewed subjects as *1-level* units but only summary statistics varying across polls.

Transforming poll outcome in accuracy measure, we standardize all results making them comparable, controlling for effect size. The accuracy measure, so, can be used as dependent variable in a hierarchical model where every outcome is affected by a specific sampling error assumed normally distributed and with a known variance. This variance can be viewed as a ‘*1-level*’ variance in a nested structure where the variance among the polls represents the ‘*2-level*’ variance. By, the means of intra-class correlation coefficient, then, is possible estimate the proportion of total variance due to the single poll.

Multilevel regression models, furthermore, allow including in the model poll features as explanatory variables in order to correlate such the poll features with accuracy measure to explain the variance.

The paper shows the results of meta-analysis regarding a large body of pre-election polls, carried out in different type of elections in Italy and published in the official Department of Ministry Council from 2001 to 2012.

La valutazione dei servizi nel settore orafo attraverso il Mystery shopping e le reti Bayesiane

Daniela D'Ambrosio*, Floriana Marino, Alfonso Piscitelli*, Giancarlo Ragozini***

*Dipartimento di Scienze Politiche, Università di Napoli Federico II, Napoli

**Il Tarì scpa, Marcianise (CE)

Il settore orafo campano è particolarmente importante sia per l'economia nazionale sia per quella locale. Esso è legato ad una lunga tradizione che ha portato alla nascita di due distretti orafi ed un polo orafo (il consorzio Tarì) che attraverso opportune strategie di marketing hanno promosso la sistematizzazione di una filiera radicata sul territorio, per migliorarne la competitività a livello nazionale e internazionale. Allo stesso tempo, la commercializzazione del gioiello ha subito ulteriori trasformazioni, con la diffusione di tipologie di distribuzione al dettaglio completamente diversi dalle usuali gioiellerie, con l'insediarsi di punti vendita di grandi dimensioni nei centri commerciali. In questi ultimi, il tradizionale rapporto tra gioielliere e cliente si modifica radicalmente divenendo più impersonale e frettoloso. Centrale diviene, quindi, l'orientamento al cliente e la gestione del servizio di vendita (Carcano *et al.*, 2007), in cui l'incontro di servizio rappresenta uno dei momenti più critici. L'interesse per questa problematica, non ancora studiata, ha posto la necessità di individuare strumenti adeguati per rilevare la qualità delle prestazioni del personale di contatto e per utilizzare tali informazioni come base al fine di migliorare il servizio stesso. Per far fronte a tali esigenze si è ricorso alla combinazione della tecnica del *mystery shopping* per la rilevazione dei dati e delle reti bayesiane per determinare i fattori competitivi su cui puntare. Il *mystery shopping* è stato utilizzato per la rilevazione delle informazioni sulle prestazioni di servizio e il relativo grado di soddisfazione (Wilson, 2001), in 45 gioiellerie situate in alcuni centri commerciali campani. L'analisi dei dati raccolti è stata condotta attraverso le reti bayesiane (Lauritzen, 1996) per identificare gli elementi chiave che influenzano la soddisfazione del cliente e definire azioni di miglioramento. Nella definizione del modello sono state selezionate solo le caratteristiche con una maggiore capacità discriminante, attraverso l'algoritmo di classificazione proposto da Chow e Liu (1968). In un secondo momento, le relazioni individuate sono state utilizzate come vincoli di partenza per la costruzione della rete definitiva attraverso l'algoritmo NPC – *Necessary Path Condition* – (Steck, 2001), che ha permesso di individuare il modello migliore, dati vincoli e parametri definiti. Nello specifico, le reti bayesiane, hanno individuato come variabili chiave della soddisfazione per il servizio la capacità di fornire un'assistenza premurosa e personalizzata, capacità di fornire informazioni dettagliate sul prodotto e tempestività nell'accoglienza. Infine, sono state condotte delle simulazioni attraverso la rete bayesiana stimata, al fine di individuare le potenziali strategie di miglioramento.

Bibliografia

- Carcano L., Catalani A., Varacca Capello P., (2006), *Il gioiello italiano ad una svolta. Dalla crisi alla costruzione di nuove opportunità*, Franco Angeli, Milano.
- C. Chow, C. Liu (1968). Approximating discrete probability distributions with dependence trees. *IEEE Transactions on Information Theory*, 14, pp. 462-67.
- Lauritzen, S.L. (1996), *Graphical Models*, Oxford University Press, Oxford.
- Steck H., (2001), *Constraint-Based Structural Learning in Bayesian Networks using Finite Data*, PhD thesis, Institut fur Informatik der Technischen Universität München.
- Wilson A., (2001), *Mystery shopping: Using deception to measure service performance*, Psychology Marketing, 18, pp. 721-34.

La valutazione della qualità di un evento enogastronomico: strategie di promozione del territorio

Veronica Distefano, Alessandra Spennato

Università del Salento

La misurazione della qualità di un evento enogastronomico, risulta essere uno strumento conoscitivo essenziale nell'ambito delle decisioni di *marketing* territoriale, nonché un canale di promozione dei prodotti tipici agroalimentari locali.

Nel presente lavoro, dopo alcuni cenni sul concetto di turismo enogastronomico, saranno descritti i principali risultati conseguiti nell'ambito dell'indagine campionaria di *customer satisfaction* riguardante un particolare evento enogastronomico che si svolge annualmente, durante la stagione estiva, in un Comune della Provincia di Lecce.

Particolare attenzione è stata rivolta nel valutare le opinioni, le abitudini ed il livello di soddisfazione dei visitatori, distinti tra visitatori non residenti e visitatori residenti nel territorio di riferimento. La parte successiva del lavoro, descrive i risultati ottenuti mediante opportune tecniche statistiche, al fine di valutare, i principali fattori che influenzano le abitudini e le scelte motivazionali di partecipazione all'evento. Infine, mediante l'analisi di regressione logistica è stato possibile valutare le variabili che determinano le scelte di acquisto del consumatore.

A conclusione del lavoro i risultati, provenienti dall'indagine campionaria, sono stati utilizzati per individuare, mediante l'analisi S.W.O.T, i punti di forza e di debolezza della manifestazione al fine di potenziare il livello di soddisfazione dei visitatori, incentivando nel contempo lo sviluppo culturale ed economico del territorio.

Keywords: livello di soddisfazione, analisi S.W.O.T, test di indipendenza, regressione logistica.

A Bagged Clustering approach for market segmentation

Pierpaolo D'Urso*, Marta Disegna**, Riccardo Massari*

*Sapienza, University of Rome;

**School of Economics and Management, Free University of Bolzano

Market segmentation consists in subdividing a market into smaller and homogeneous groups, according to specific characteristics of customers. Indeed, markets and individuals who make up such markets are not homogeneous and, hence, no single market offering will satisfy all individuals. Therefore, the market segmentation can provide important information to the marketers in order to direct marketing efforts toward the groups of customers that are resulting more economically significant.

Cluster analysis methods, which can be broadly divided into partitioning and hierarchical methods, represent the most used market segmentation technique in the literature.

Each clustering method conducts a multivariate description of the data, grouping units based on their similarity. Unfortunately, as emphasized by many researchers, no absolutely “correct” method to segment exists in the literature, since the underlying relationships among units have different structures, depending on data at hand, and the researcher must find the best segmentation method to capture this hidden structure.

Beyond more traditional clustering methods, recently, the Bagged Clustering (BC) algorithm, based on the Bagging (bootstrap aggregating) procedure, has been introduced.

In its initial formulation BC combines sequentially partitioning and hierarchical clustering methods, to overcome some limitations of both the two procedures. In particular, first a partitioning method, namely the k -means algorithm, is applied to B bootstrap samples generated from the data set. Then a hierarchical method is applied to the results of the partitioning steps. The main advantage of this algorithm is that it is not necessary to impose the number of clusters in advance.

The aim of this work is to propose a new segmentation technique based on the BC algorithm.

In the partitioning step, the Fuzzy C -Medoids Clustering (FCMdC) algorithm is adopted in place of a crisp clustering algorithm proposed in the original BC algorithm. In the market segmentation, it is reasonable to associate one unit to more than one cluster. Then $B \times C$ medoids and the membership degrees of each unit to the different clusters are obtained. The second step consists in running a hierarchical clustering algorithm on the $B \times C$ medoids for determining a set of partitions of medoids. The best partition is obtained investigating properly the dendrogram. The membership degrees of the original units to each cluster are obtained as the maximum membership degree of each unit amongst all medoids in the cluster, instead of considering the minimum distance of each unit to all medoids, as in the traditional BC. This allows to fuzzy allocate each unit to all clusters, with different membership degrees.

In order to show the effectiveness of the procedure, two empirical applications on tourism data are finally provided.

La grande rete della distribuzione

Giovanni Marco Esposito*, Alfonso Piscitelli, Giancarlo Ragozini****

*Consorzio Di.Al., Marcianise (CE); **Università degli Studi di Napoli “Federico II”

Il settore commerciale italiano ha subito negli ultimi anni un periodo di forte cambiamento, identificato nella letteratura di riferimento come “*rivoluzione commerciale*” (Fornari, 2005). Il riferimento al termine rivoluzione è determinato dalla volontà di evidenziare l’entità e la rapidità del cambiamento, contrariamente ad una normale evoluzione che identifica, invece, un trend decisamente più lento, graduale, sistematico.

I nuovi assetti delle aziende commerciali hanno messo in crisi il classico modello distributivo favorendo lo sviluppo di una multicanalità distributiva in grado di condizionare le preferenze della domanda sviluppando un proprio marketing di “*insegna*”.

In questo contesto di cambiamento è stata effettuata un’indagine sulle relazioni intercorrenti fra le aziende associate ad un consorzio di distribuzione alimentare, in particolare delle bevande, al fine di analizzare la configurazione delle relazioni fra le aziende, le strategie di collaborazione e competizione all’interno del consorzio, i ruoli assunti da vari partner, la presenza di cluster territoriali e di interesse che possono condizionare il funzionamento dell’intero sistema della distribuzione.

Il consorzio considerato è formato da 110 aziende di distribuzione tradizionale di bevande localizzate nel sud Italia. Nello specifico, si tratta di imprese omogenee con caratteristiche strutturali e organizzative differenti, che aderiscono al consorzio per molteplici ragioni: migliorare la logistica, spuntare prezzi migliori, stringere accordi commerciali, collaborare per contrastare la crisi, etc.

La rilevazione dei dati è stata effettuata attraverso la somministrazione di un questionario costruito *ad hoc* (in grado di rilevare informazioni: sulle strutture organizzative delle imprese associate; sulle strategie di comunicazione adottate; sulla *performance* economica/informativa/relazionale del Consorzio). Il questionario è stato somministrato attraverso la rete informativa del consorzio, sotto forma di questionario elettronico. Particolare attenzione è stata posta nella rilevazione dei dati relazionali, ovvero dei legami di tipo economico e informativo presenti fra le imprese. Ogni legame è stato verificato e validato attraverso una serie di controlli incrociati.

L’analisi dei dati è stata condotta su due livelli. Una prima analisi ha messo in luce le strutture organizzative delle aziende associate al consorzio, dedicando particolare attenzione ai *feedback* dei canali informativi e delle *performance* economiche, organizzative, informative e relazionali del consorzio. Il secondo livello di analisi ha riguardato i dati relazionali, analizzati attraverso gli strumenti della *social network analysis* (Wasserman e Faust, 1994). Sono state presi in considerazione gli indici di rete usuali per poter analizzare la struttura delle reti e i ruoli di ciascun partner. In particolare è stata messo in atto una analisi di tipo *core/periphery* per mettere in luce i *broker* centrali e i *player* periferici. L’analisi di rete ha anche messo in evidenza una differenza fra la rete formata dalle relazioni economiche quella formata dalle relazioni di tipo informative tra le aziende associate (che entrano in gioco ad esempio nel caso dell’elezione degli organi collegiali/sindacali. Inoltre, la *network analysis* ha messo in luce le diverse le logiche decisionali delle imprese e descritto l’attivazioni dei diversi *ego-networks* in casi di mobilitazioni specifiche.

Questo lavoro di ricerca ha evidenziato come il *networking*, nei suoi aspetti più sociali, sia entrato a far parte nella vita dell’uomo quanto nella nuova azienda “umana”: capace di scegliere le relazioni, di investire in esse e di agire su nuove dimensioni, ben oltre la sfera della pura razionalità economica.

Conjoint metrica per somministrazione online

Fausta Faini, Luca Massaron
Carlo Erminero & Co srl, Milano

Il tema: gli studi di conjoint analysis hanno raggiunto un elevato livello di sofisticazione, senza peraltro superare completamente i problemi legati alla raccolta delle informazioni. L'approccio proposto nell'articolo recupera l'impostazione generale del metodo metrico classico, ma lo semplifica nel numero degli esperimenti da sottoporre ai rispondenti, sia ricorrendo al metodo di ricampionamento jackknife sia all'exchange algorithm di Federov. Sono presentati casi di applicazioni concrete nell'ambito della messa a punto di nuovi concetti di prodotto e di offerta in mercati differenziati.

Sviluppo: negli ultimi 30 anni gli studi di conjoint analysis si sono particolarmente sofisticati, nel modello decisionale trattato (da metrico a choice based), nel disegno degli esperimenti (da fattoriale ad adattivo), nel modello di stima delle utilità (dalla semplice regressione lineare ai modelli gerarchici bayesiani).

Recentemente, sono stati recuperati e modificati alcuni fra i modelli più semplici di conjoint analysis (tipicamente modello metrico, disegno fattoriale degli esperimenti, stima con regressione lineare) per poterli applicare sistematicamente al testing dei concept online (vedi ad esempio ma non solo: HR. Moskowitz, A. Gofman et alii, "Rapid, inexpensive, actionable concept generation and optimization: the use and promise of self-authoring conjoint analysis for the food service industry", *Food Service Technology*, Volume 1, Issue 3-4: 149–167, 2001).

Per quanto i modelli di conjoint analysis di tipo metrico con disegno fattoriale degli esperimenti offrano indubbi vantaggi nelle attività di implementazione e stima, nonché consentano di poter disporre di stime di utilità a livello individuale, il loro limite rimane comunque nella somministrazione, particolarmente onerosa per i rispondenti, con conseguente minore qualità ed affidabilità del dato raccolto e delle conseguenti stime ottenute, in modo particolare nelle rilevazioni auto-compilate via Internet. Infatti, al fine di mantenere elevato il realismo dell'esperimento e rendere credibile per i rispondenti il processo di *trade-off* fra le diverse caratteristiche in gioco, è necessario che queste ultime siano in numero significativo sia in termini di numero di variabili coinvolte che di livelli delle variabili. A ciò corrispondono esperimenti onerosi per numero di configurazioni da valutare, suscettibili quindi di generare risposte standardizzate con un conseguente basso coefficiente di determinazione individuale.

Nel presente articolo viene illustrato un nuovo approccio alla conjoint analysis metrica, in cui pur confermando l'impianto generale del metodo metrico classico, esso viene decisamente semplificato nel numero degli esperimenti da sottoporre ai rispondenti ricorrendo sia al metodo di ricampionamento jackknife (con lo scopo di escludere la valutazione di cartellini hold-out da parte dei rispondenti) ed all'exchange algorithm di Federov al fine di generare disegni degli esperimenti ridotti in numero di cartellini (idoneo quindi alla somministrazione da questionario via Internet), quasi ottimi (non perfettamente ortogonali e bilanciati, ma D-ottimi) ed efficaci (nella corretta stima delle utilità individuali).

Il nuovo approccio è stato applicato in numerosi studi, in modo particolare relativi alla messa a punto di nuovi concetti di prodotto o di offerta in mercati differenziati: dal largo consumo all'editoria, dalla moda all'ITC. Nell'articolo viene fornita dimostrazione della correttezza ed efficacia dell'approccio non solo su base teorica ma anche empirica, grazie ad una simulazione Montecarlo ed all'esposizione di numerosi casi di applicazioni concrete.

Dall'intelligenza artificiale alle ricerche di mercato: l'algoritmo Random Forest

Fausta Faini, Luca Massaron

Carlo Erminero & Co srl, Milano

Il tema: in anni recenti nell'ambito del data mining sono stati sviluppati numerosi e particolarmente efficaci algoritmi di previsione. Tali algoritmi richiedendo per il loro buon funzionamento grandi moli di dati, difficilmente trovano utilizzo nelle ricerche di mercato che operano usualmente su campioni ristretti. Nella nostra esperienza, particolarmente negli studi di Customer Satisfaction, abbiamo verificato come l'algoritmo di apprendimento automatico ideato da Leo Breiman, Random Forests, possa essere utilizzato anche su campioni di dimensioni ridotte per imputare, velocemente e con buona approssimazione, i dati mancanti e per effettuare data fusion fra campioni indipendenti.

Sviluppo: negli anni recenti, nell'ambito dell'intelligenza artificiale e dell'apprendimento automatico (data mining), sono stati sviluppati numerosi e particolarmente efficaci algoritmi di previsione. Tali algoritmi, di natura computazionale e non parametrica, richiedono per il loro buon funzionamento significative moli di dati e non trovano facile utilizzo nel campo delle ricerche di mercato, che operano usualmente su campioni ristretti.

Empiricamente, in numerosi studi realizzati, abbiamo rilevato come l'algoritmo di apprendimento automatico ideato da Leo Breiman, Random Forests, possa essere utilizzato anche nelle ricerche di mercato al fine di imputare velocemente e con buona approssimazione i dati mancanti e per effettuare data fusion fra campioni indipendenti.

Questo lo rende specialmente utile nelle ricerche di customer satisfaction per integrare i numerosi casi mancanti con tecniche meno penalizzanti delle classiche cancellazioni dei dati listwise o pairwise o della semplice imputazione per media (che ha lo svantaggio di comprimere significativamente la varianza della variabile), senza dover ricorrere ad imputazioni attraverso regressione lineare, con conseguente elevato dispendio di tempo ed attenzione del ricercatore nella diagnostica del modello.

Risulta anche molto utile nelle evenienze in cui sia necessario fondere insieme due campioni indipendenti laddove solo una parte delle variabili è comune. Le Random Forest, infatti, sono modelli molto più complessi del modello lineare, e sono in grado di intercettare anche interazioni complesse fra variabili, mantenendo generalizzabili le previsioni.

L'algoritmo Random Forests consiste in una combinazione di alberi di previsione tale che ciascun albero appartiene alla medesima distribuzione pur essendo diverso sulla base di un processo di generazione randomizzata.

Si tratta di un algoritmo particolarmente versatile, potendo operare con variabili di qualsiasi tipo e scala. Nell'implementazione da noi utilizzata in linguaggio Python (un linguaggio per la programmazione ad oggetti), l'algoritmo si è dimostrato molto veloce nei calcoli anche in presenza di dataset con numerosi casi e variabili.

Nel presente lavoro sono investigate, dandone dimostrazione matematico-statistica, le caratteristiche dell'algoritmo che lo rendono idoneo all'utilizzo nelle ricerche di mercato e viene dimostrata, grazie ad una simulazione Montecarlo, la sua funzionalità nell'imputazione dei dati mancanti e nel *data fusion* fra campioni indipendenti, formalizzando criteri e limiti di utilizzo.

The impact of exposure to pre-election polls on voting behaviour

Giancarlo Gasperoni, Debora Mantovani

Dept. Political and Social Sciences, Alma Mater Studiorum-Università di Bologna

This paper reports a subset of findings drawn from a national research project – entitled “Electoral Choice: Voters’ Heuristic Strategies and Information Processing” – aiming to identify the strategies that Italian voters enact in order to combine political information originating from exposure to election campaigns and information processed during social interaction, in the wider context of the Italian political system. Recently new fields of research have focused on socio-cognitive factors that affect voting choices and information search processes enacted to formulate judgments via cognitive shortcuts (or “heuristics”). In particular, this project extends the scope of the innovative voting decision model developed by Richard R. Lau and David P. Redlawsk and adapts it to the Italian context. The voting decision model is operationalized via a “dynamic information board” simulating election campaigns tailored to observe information research strategies in which voters engage. This technique employs a controlled-environment, on-line simulation, endeavouring to reproduce a complex, realistic environment, in which the information that the voter can access changes over time. This paper, in particular, focuses on voters’ perception of pre-election polls and on polls’ role as a heuristic influencing voter choice. Opinion polls produce information flows – originating among citizens and directed to public authorities – that are more extensive, detailed and constant over time than those generated by actual elections (which are comparatively infrequent and feature generic semantic content). Surveys create information flows that are also directed toward citizenry via the diffusion of poll findings by the media. Since mass media are the major source of political information, they contribute to a great extent to the determination of the most salient issues and thus can affect public opinion. This has fed fear and reservations about the role of polls in the political life of contemporary society. A special status is reserved for polls concerning voting behaviour, especially when they attempt to record voting intentions and predict election results. Pre-election polls perform a legitimate political information function vis-à-vis voters: they convey the election “supply” and make known, within limits, the corresponding demand (preferences and tendencies among the citizenry). Pre-election polls can comprise an “epistemic source”, providing the foundation for the particular type of heuristics based on the amount of support a candidate appears to enjoy, and influence individual voting behaviour. The paper concerns, in particular, the use of the **opinion poll heuristic** and addresses three specific topics: A. attitudes towards opinion poll in terms of reliability and usefulness; B. actual access to opinion poll flow items; C. effects on final voting decisions of “forced” exposition to polls on voting intentions. The first two topics were addressed employing data concerning 895 study participants; the third point enjoys a smaller empirical base, with a sample size of 188. On the whole, voters participating in the simulated election campaign displayed somewhat negative attitudes towards pre-election polls. During the actual campaign simulation, pre-election polls were not used extensively as an information source by voters: on average, only one flow item out of four was accessed; over 40% of participants ignored the polls; participants with higher levels of education and interest in politics resorted to this heuristic more than others. As regards the part of the study in which voters were “forcibly” exposed to polls that reported unfavourable predictions for their preferred candidates, one-tenth of voters switched votes. Vote switching appeared to be more widespread among groups of voters who should have been – in light of their reported attitudes towards them – comparatively less receptive to pre-election polls.

Interdisciplinary or multidisciplinary for the research in statistics?

M. Gabriella Grassia*, Emma Zavarone**

*Università di Napoli “Federico II”, **IULM; Milano

In this paper, we assess cross disciplinary of research produced by the Italian Academic Statisticians and by Social Statisticians combining text mining and bibliometrics techniques. Interdisciplinary research would be supported and encouraged to solve new statistical challenges. The term interdisciplinary tends to be tacitly understood by researchers, without shared definition. We adopt the definition suggested by Porter et al. (2007) interdisciplinary research requires an integration of concepts, theories, techniques and/or data from two or more bodies of specialized knowledge. Multidisciplinary research may incorporate elements of other specialized knowledges, but without interdisciplinary synthesis (Wagner et al., 2011) which includes more than single parts. Analysis of cross disciplinary improves traditional indicators assessing and quantifying interdisciplinary research. Indicators of different disciplinary describe heterogeneity of a bibliometric set obtained starting from predefined categories. Network coherence indicators are constructed to measure the intensity of similarity relations within a bibliometric set using a bottom-up approach, which reveals the structural consistency of the publications network (Rafols and Meyer, 2010). We incorporate both points of view and show how to improve on existing text-based and bibliometric methods. We propose an hybrid clustering procedure based on Fisher _s inverse chi-square method for integrating textual content and citation information. Given clustered papers, it's possible to evaluate ISI subject categories as descriptive labels for statistical documents, and to address individual researchers interdisciplinary based on monthly Italian Statisticians papers, for ten years from 2001 to 2011. Also the data gathered on the research of Social Statisticians will be analyzed.

References

- Porter, A., Cohen, A., Roessner, J. D., & Perreault, M.. (2007) Measuring researcher interdisciplinarity. *Scientometrics*, 72, 117–147
- Rafols, I. and Meyer, M. (2010) Diversity and Network Coherence as indicators of interdisciplinarity: case studies in bionanoscience. *Scientometrics* 82(2), 263-287
- Wagner, C. S., Roessner, J. D., Bobb, K., Klein, J. T., Boyack, K. W., Keyton, J., Rafols, I., & Borner, K. (2011) Approaches to understanding and measuring interdisciplinary scientific research (IDR): A review of the literature. *Journal of Informetrics*, 5(1), 14–26

Modelling seasonality in innovation diffusion

Mariangela Guidolin, Renato Guseo

Department of Statistical Sciences, University of Padova,

The ability to forecast new product growth is particularly important for innovative firms that compete in the marketplace. Innovation diffusion research aims to describe the evolution of a new product in the market by modelling its life-cycle. There is a quite long tradition in this field: wide research has been produced to capture several phenomena visible in sales data. A particular effort has always been devoted to extending the structure of the basic and most known Bass model (Bass, 1969), by taking into account price dynamics, competition, targeted or pulsing advertising strategies, network externalities, consumer heterogeneity and technological generations, arguing that these may help explain turning points in life-cycle such as take-off, saddle, technological substitution. Indeed, a typical focus of diffusion models is to provide an efficient description and interpretation of the mean trajectory of a life-cycle, so that a seasonal pattern in sales data has not been often studied in an explicit way. However, many new products exhibit a very strong seasonal behaviour, which may deserve a specific modelling, both for producing better forecasts in the short term and for better explaining special market dynamics. Seasonal behaviour is typically visible when data are collected with monthly or quarterly frequencies and some works on diffusion concentrated on the issue of temporal aggregation of data: indeed, the aggregation of monthly or quarterly data to obtain a smoother shape of sales would result in a loss of information to avoid. Moreover, since product life-cycles are increasingly shortening due to high competition, it becomes more and more necessary to have short-term projections, by accurately analyzing not only the trend of sales but also their oscillations within the year, which in some cases may be very strong. In fact, the sales pattern of many technological devices exhibits strong seasonal oscillations, which may be partly explained by choices of consumers and partly by the business strategies of the producer. By considering seasonality as a deterministic component to be estimated jointly with the trend through Nonlinear Least Squares methods, we develop two extensions of the Guseo and Guidolin model (2009) that enable to describe simultaneously trend and seasonality. These models are based on two different but equally reasonable approaches: in one case we consider a simple additive decomposition of a time series and design a model where seasonality is directly added to the trend and jointly estimated with it; in the other we design a more complex structure, miming that of a Generalized Bass model (Bass et al., 1994) and embed a seasonal perturbation within the evolutionary trend, so that seasonality directly operates on it. In both cases the seasonal amplitude is modulated by trend components. Testing the models to two applied cases, the life-cycle of a pharmaceutical drug and the diffusion of a technological device produced by Apple, we find that the more complex modelling option is more suitable to the case of the pharmaceutical drug, while the additive solution performs better in the other case. Both models are quite simple to implement and to interpret from a business point of view.

Keywords: Seasonality, Time Series Decomposition, Guseo-Guidolin model, Generalized Bass model, Nonlinear Least Squares

References

- Bass, F.M. (1969) A new product growth model for consumer durables, *Management Science*, 15, 215-227.
- Bass, F.M., Krishnan, T.V., Jain, D.C. (1994) Why the Bass model fits without decision variables, *Marketing Science*, 13, 203-223.
- Guseo, R., Guidolin, M. (2009) Modelling a Dynamic Market Potential: A Class of Automata Networks for Diffusion of Innovations, *Technological Forecasting and Social Change*, 76(6): 806-820

Social media e sentiment analysis: una applicazione in ambito televisivo

Caterina Liberati, Paolo Mariani, Federico Neri

Università di Milano-Bicocca

Il lavoro descrive uno studio di Sentiment Analysis condotto su circa 1.000 post inerenti le reti televisive e collezionati su Facebook. In particolare vengono comparati i sentiment per la rete pubblica Rai1 e quelli per la rete privata La7. Lo studio utilizza inoltre dati provenienti dall'Osservatorio di Pavia ed Auditel.

Il Web rappresenta lo spazio virtuale dove esprimere e condividere le proprie opinioni e influenza gli aspetti del nostro vivere come quelli dipendenti dalle attività di marketing o di comunicazione. I social media quali Facebook e Twitter dominano lo spazio del marketing digitale e molte società offrono studi di sentimento, di opinione e di reputazione che ormai sono consolidati e fructi. Il sapere chi parla a proposito di cosa, quando e in che modo, si presenta giorno dopo giorno come un ambito di ricerca ricco e ancora da esplorare.

Il lavoro descrive un studio di Sentiment Analysis condotto su circa 1.000 post inerenti le reti televisive e collezionati su Facebook. In particolare vengono comparati i sentiment sui programmi di informazione per la rete pubblica Rai1 e quelli per la rete privata La7. Lo studio utilizza inoltre dati provenienti dall'Osservatorio di Pavia ed Auditel. Dopo avere riportato lo stato dell'arte della Sentiment Analysis vengono descritte le componenti del Sentiment and Knowledge Mining System:

- il crowler utilizzato per collezionare i testi da Internet/Intranet o database;
- il motore semantico che identifica le informazioni rilevanti contenute nei testi attraverso le relazioni semantiche fino all'analisi sentimento,
- il motore di ricerca basato sul linguaggio naturale
- il traduttore dei testi,
- il motore di geo-referenziazione che offre una rappresentazione spaziale e interattiva dei documenti,
- il sistema di classificazione supervisionato e non supervisionato.

Il monitoraggio dei social media verrà riportato attraverso strumenti di navigazione del testo, quali lo spazio dei concetti, le mappe tematiche, le proiezioni delle regole semantiche, cruscotti di consenso e relativa distribuzione temporale. Proprio su questa ultima verrà proposta una applicazione legata all'analisi dinamica di spazi di sintesi dell'informazione, aprendo la discussione sulla contaminazione di tecniche fortemente utilizzate in azienda con i nuovi orizzonti aperti dalle rinnovate tecniche di collezione e trattamento dei testi.

Prospettive di studio per l'analisi predittiva dei contenuti e dello sviluppo delle idee all'interno di una market research online community (MROC)

E. Lillini R. *, A. Costa **, P. Perelli P. **, A. Giovenali A. ** , Staff di Nextplora **

*Scuola di Dottorato in Scienze Sociali, Università degli Studi di Milano–Bicocca;

** Nextplora Idee & Opinioni S.p.A., Milano

L'analisi delle conversazioni su tematiche di sviluppo di nuovi prodotti e servizi e/o sulle caratteristiche innovative, di migliore utilizzo, di bisogni che essi possono soddisfare, ecc., è uno strumento che, con l'utilizzo delle tecnologie proprie del *Web 2.0*, si sta rilevando sempre più utile per le finalità delle ricerche di mercato.

La possibilità, infatti, di coinvolgere via rete (su *social network, forum, chat* dedicate, ecc.) un numero di partecipanti sempre maggiore e, grazie alla progressiva riduzione del *digital divide*, sempre più stratificato da un punto di vista demografico e socio-economico rende possibile analizzare i testi prodotti in tali ambienti non più solamente con il metodo etnografico-interpretativo, ma anche con l'approccio statistico proprio dell'analisi del teso e del contenuto, permettendo classificazioni e proiezioni probabilistiche dei risultati.

In più, le caratteristiche con cui si sviluppano e sempre maggiore sono conservate le conversazioni in ambiente web permettono anche uno sviluppo più agevole di analisi che definiscono la struttura di rete che gestisce le interazioni fra i partecipanti, permettendo, almeno potenzialmente, di ricavare considerazioni sulle modalità di costruzione condivisa delle informazioni e, nel caso degli oggetti di conversazione sopra citata, i *desiderata* di coloro che poi saranno acquirenti/consumatori/utilizzatori.

Obiettivo del presente lavoro è l'analisi dei contenuti e delle strutture di rete che emergono dal forum dell'agenzia Nextplora Idee & Opinioni S.p.A. di Milano, relativamente allo sviluppo di idee e innovazioni per il miglioramento di alcune sfere della vita quotidiana, al fine di verificare come essi possano essere utilizzati in maniera valida e attendibile per le finalità di cui sopra.

I testi delle conversazioni da analizzare e le persone che hanno prodotto tali testi provengono dal social network delle idee <http://ideeopinioni.nextplora.com/> pubblicate nel 2012.

L'analisi del testo e dei contenuti si basa sulla ricerca delle co-occurenze, delle concatenazioni testuali, delle clusterizzazioni testuali tematiche tramite analisi delle corrispondenze multiple e cluster analysis sviluppata con tecniche gerarchiche, con l'ausilio del software per l'analisi testuale T-Lab

L'analisi delle strutture di rete è mirata ad individuare nodi primari, legami primarie e secondari, frequenza di interazione fra individui e contenuti, tramite metodologia e software Gephi.

Risultati: Lo studio è attualmente in corso.

Conclusioni attese

- Individuare le modalità di interazioni sotse allo sviluppo e diffusione di una nuova idea nella sua dimensione sistemica (co-creazione del gruppo) e nel suo sviluppo diacronico
- Definire il contributo dei ruoli e le posizioni di rete degli attori del network
- Individuare i gruppi di interesse che esprimono maggiore appeal per una nuova idea connotandoli secondo le loro caratteristiche identitarie (es. demografiche, socio-economiche)

L'Index of Consumer Sentiment: una tradizione da rifondare

Giorgio Marbach

Ateneo Telematico "Universitas Mercatorum"

Negli anni più recenti è aumentata l'importanza attribuita ai cosiddetti indici di fiducia. Essi traggono tutti origine da un unico antenato, il Consumer Sentiment Index realizzato dall'Università di Michigan a partire dal 1946.

In questo studio se ne ricostruiscono le origini, si descrivono le indagini che ne sono derivate in Italia e in numerosi altri paesi, riportando per il nostro i risultati che vari Istituti e l'Istat forniscono per gli anni più recenti.

Si rileva che il *sentiment* non è concettualmente definito, cosicché il clima di attese che se ne ricava, in quanto differenza tra valutazioni positive e negative sia del recente passato che del futuro a breve-medio termine, ha esclusivamente valenza pragmatica ed operativa.

A seguito dell'esame di pubblicazioni di George Katona ed Eva Müller si è potuto rilevare che gli *item* dell'ICS costituiscono soltanto la prima sezione del modello proposto dai due autori. Inoltre alcune semplificazioni nella elaborazione assunte anche dalla Michigan University sono correttamente segnalate da Katona e Müller, ma con riferimento esclusivamente alle proprie ricerche, senza mai menzionare l'ICS. A quest'ultimo alcuni autori moderni riconoscono però un ruolo positivo qualora venga incluso in modelli econometrici soprattutto in presenza di brusche variazioni. Inoltre indici che all'ICS si ispirano segnalano in numerosi Paesi del mondo la percezione differenziata del clima di recessione.

Una vasta letteratura internazionale individua le improprietà di una impostazione cristallizzatasi nel tempo, laddove l'ICS fornisce *performances* diverse in qualità di previsore dei consumi nelle fasi di turbolenza e di shock ed in quelle «normali». Alcuni studi che hanno preso in considerazione otto tra i Paesi più evoluti confermano inoltre che non è accoglibile un unico modello universalmente valido ed immutabile nel tempo.

Il campionamento telefonico, poi, determina a sua volta distorsioni, individuate dallo scrivente con riferimento all'Italia. Una proposta di più idonea impostazione conclude il lavoro.

Dimensionalità del costrutto di fiducia nella marca dei beni durevoli

Cristiana Martini, Chiara Arseni, Giovanna Galli

Università di Modena e Reggio Emilia

Negli studi di marketing la fiducia nella marca ha assunto un'importanza sempre crescente, in quanto elemento in grado di determinare relazioni durature e positive con il consumatore; la fiducia si distingue come elemento alla base del successo dell'impresa e come fonte di vantaggio competitivo. Numerosi sono stati i tentativi di concettualizzazione del costrutto di fiducia nella marca, ma la definizione di tale costrutto e delle sue dimensioni è ancor oggi uno dei temi più dibattuti.

Nel presente lavoro si è analizzata la fiducia nella marca in relazione ai beni durevoli; ciò rappresenta un elemento di novità, poiché in letteratura si è posta maggiore attenzione ai beni ad alta frequenza di acquisto. I beni scelti come oggetto di analisi sono stati il computer portatile e la lavatrice. Si sono inizialmente ipotizzate 5 dimensioni alla base della fiducia nella marca: competenza, benevolenza, empatia, lealtà e reputazione; al fine di rappresentare tali dimensioni si sono utilizzati 26 item tratti dalla letteratura sul brand trust. Il questionario includeva anche domande volte a rilevare la soddisfazione, la fedeltà, l'intenzione al riacquisto e alcune caratteristiche del comportamento d'acquisto, oltre alle caratteristiche socio-demografiche. I questionari sono stati somministrati attraverso il web a 400 rispondenti.

Dopo alcune preliminari analisi di affidabilità, che hanno condotto all'eliminazione di 8 item dal comportamento instabile, sono emerse due sole dimensioni principali: l'affidabilità della marca, che comprende indicatori di competenza e reputazione, e le intenzioni della marca, che racchiudono indicatori di benevolenza, empatia e lealtà; la dimensione di affidabilità risulta nettamente prevalente nel determinare la fiducia complessiva. L'indicatore di fiducia così ottenuto è stato sottoposto a validazione, confermando le relazioni teoriche che attribuiscono alla fiducia un ruolo di mediazione fra la soddisfazione e la fedeltà. Inoltre, si osserva che il livello di complessità del processo di acquisto (rischio percepito, rilevanza del bene per l'individuo) influenza le dinamiche relazionali tra consumatore e marca, inducendo un livello di fiducia maggiore; anche una maggiore esperienza personale nella categoria di prodotto si associa a livelli più alti di fiducia.

A taxonomy of higher-education effectiveness indicators for decision-making

M. Cristiana Martini*, Luigi Fabbris**

**Università di Modena e Reggio Emilia; **Università di Padova*

This paper is aimed at devising a taxonomy of indicators that can be employed to assess the external effectiveness of graduates' education in order to enhance decision making at universities and in the broader society. We assume that graduates can provide conscious information on external effectiveness cheaply and effortlessly, and that university decision-makers, employers and concerned representatives of the civil society are the main stakeholders of the higher education system.

A set of indicators related to external effectiveness were selected and their relevance and reliability were analysed. The data were drawn from a longitudinal computer-assisted telephone survey carried out from 2008 to 2011 by the University of Padua on a sample of graduates; since the concept of effectiveness is particularly meaningful when referred to an aggregate level (e.g. study programmes, faculties, universities), data were analysed at the study programme level.

A second survey was conducted via a CAWI (web-based) Delphi-type procedure on sets of university stakeholders, in order to describe the differences in the definition of external effectiveness among different categories of stakeholders and to suggest further possible indicators for future studies.

Our analyses highlighted the existence of three main dimensions of effectiveness indicators; two of them are directly connected with the marketability of degrees, the first referring to the employability, and the second to the job quality, while a third dimension collects all the facets which are independent on what happened after graduation (fidelisation, overall satisfaction with the university experience and the study contents). As for the simple indicators, some proved to be unsuitable effectiveness indicators (e.g. the rate of study continuation and, somehow unexpectedly, the average job satisfaction level), while some were partly ill-suited in their specific effectiveness dimension (e.g., the average satisfaction for the achieved competences, or the rate of graduates who define their level of professional specialisation as adequate). Moreover, data collected one year or three years after graduation did not show different performances as indicators of external effectiveness.

Data fusion e marketing: un'applicazione nel settore farmaceutico

Paolo Mariani, Maria Cristina Vittori

Università di Milano-Bicocca

La crescente necessità di informazioni utili e utilizzabili dalle imprese per prendere decisioni e il lungo periodo di crisi economica attraversato dal nostro paese ha reso indispensabile un intervento in termini di razionalizzazione dei costi e delle risorse anche in campo statistico. L'uso di fonti amministrative di dati, in sostituzione o accanto a quelle di tipo statistico, in relazione alla loro disponibilità e all'utilità, ha favorito la riduzione delle spese e dei tempi di sviluppo delle indagini. Allo stesso tempo, la digitalizzazione dei sistemi di gestione, ormai quasi completamente informatizzati, ha facilitato l'armonizzazione tra le diverse tipologie di dati. Inoltre, l'elevata numerosità di archivi amministrativi e statistici, presenti sul territorio nazionale, ha stimolato l'utilizzo di tecniche di data fusion per favorire la creazione di basi di dati sì più grandi e complesse ma anche meglio strutturate e complete. Questo permette di soddisfare a pieno le necessità informative aziendali, sempre più specifiche e articolate, non solo legate alla popolazione complessiva oggetto d'indagine o a suoi sottoraggruppamenti omogenei, ma anche direttamente al singolo individuo. Il mercato farmaceutico, in particolare, offre una buona opportunità di sfruttamento delle tecniche di data fusion su dati di entrambe le tipologie – nel nostro caso provenienti dal database amministrativo dei medici di base italiani e da un'indagine postale sugli stessi svolta nell'anno 2005 -. Lo scopo è quello di ottenere informazioni il più possibile corrette e verosimili sulle potenzialità di prescrizione, e quindi di vendita/consumo, di un generico prodotto farmaceutico, attraverso l'utilizzo di tecniche di data mining (cluster analysis) e la somiglianza (criterio della minima distanza) tra record di fonti differenti sulla base di alcune delle variabili comuni rilevate - quelle considerate più influenti sulla previsione -. Lo studio prende come figura centrale quella del medico di base, principale responsabile dei consumi di farmaci nel mercato (soprattutto per quanto riguarda i prodotti con obbligo di prescrizione), e, a partire dai dati in nostro possesso, adeguatamente elaborati e organizzati, cerca di valutare l'atteggiamento del singolo medico nei confronti del prodotto in analisi. L'attenzione è, di conseguenza, rivolta alle diversità di comportamento prescrittivo legate alla soggettività della persona e alle sue caratteristiche professionali (area geografica di svolgimento dell'attività, numero di assistiti) per poter attribuire, anche in assenza di informazione diretta, un valore di prescrizione a tutti i medici sul territorio, sfruttando la somiglianza tra soggetti. La scelta del contesto farmaceutico e in particolare del soggetto "medico di base" è stata compiuta tenendo conto dei cambiamenti che sta affrontando il settore sulla base del Piano E-Gov 2012 per le Pubbliche Amministrazioni. Infatti, l'informatizzazione completa prevista del settore sanitario, sia per quanto riguarda le informazioni sui pazienti che l'attività prescrittiva dei medici, favorisce, da un lato, una più efficiente comunicazione tra tutti i soggetti coinvolti e, dall'altro, la costruzione di archivi di dati immediatamente utilizzabili a fini statistici e di controllo. L'elaborato ha lo scopo di mostrare come sia possibile sfruttare il potenziale informativo di grandi basi di dati statistiche e amministrative attraverso l'integrazione delle stesse (utilizzando le tecniche di data mining e di data fusion) e la successiva analisi del database unico finale ricavato, al fine di offrire un utile supporto alle decisioni aziendali di pianificazione strategica e operativa dell'attività. Lo studio è stato compiuto nell'ambito farmaceutico, utilizzando il database dei medici di base italiani e i dati rilevati con un'indagine postale sugli stessi soggetti e ha lo scopo di mostrare l'utilità dell'utilizzo del data fusion per elaborare una previsione, il più possibile accurata, sul potenziale di prescrizione/vendita di un prodotto farmaceutico, in modo che i risultati possano essere d'aiuto nell'ambiente aziendale.

Il comportamento dei rispondenti durante la compilazione di un questionario online: il caso della Rilevazione statistica su Ricerca e Sviluppo nelle imprese

Mauro Masselli, Alessandra Nuccitelli
Istat, Roma

Le opportunità tecnologiche offerte dalle indagini gestite via web permettono di monitorare con precisione il comportamento di ciascun rispondente durante la compilazione di un questionario elettronico. Ad esempio, i files di log generati in occasione di una rilevazione consentono di avere informazioni dettagliate e oggettive sul processo di risposta. L'analisi di tali informazioni può fornire utili indicazioni sulle eventuali criticità nella progettazione del questionario e, di riflesso, sugli accorgimenti da adottare per migliorare l'accuratezza delle risposte.

In tale contesto, a partire dai dati di log generati in occasione della Rilevazione statistica su Ricerca e Sviluppo nelle imprese sono costruiti e analizzati alcuni indicatori del comportamento dei rispondenti durante la compilazione del questionario relativo alla stessa rilevazione.

Il primo quesito del questionario consiste in una domanda filtro sull'attività di ricerca e sviluppo svolta o meno all'interno dell'impresa nel corso dell'anno di riferimento. A seconda della risposta fornita a tale quesito, viene evidenziato il percorso da seguire. Il questionario è stato progettato per essere compilato secondo la numerazione crescente dei quesiti; tuttavia, è ammessa la compilazione anche mediante selezione di singoli quesiti secondo una sequenza definita dall'impresa rispondente. Gli indicatori proposti per studiare il comportamento dei rispondenti assumono valori tanto più elevati quanto maggiori possono essere le difficoltà incontrate durante la compilazione (ad esempio, tempo lordo utilizzato per la compilazione, tempo di collegamento effettivo, numero di giorni in cui avviene il collegamento, numero di richieste di accesso, numero di salvataggi, lunghezza del percorso effettivamente compiuto in termini di numero di quesiti visitati, numero di volte in cui avviene un'inversione, ovvero una variazione rispetto all'ordine di percorrenza dei quesiti secondo la loro numerazione crescente, numero di volte in cui ogni quesito è visitato, numero di errori, ecc.). Tali indicatori sono analizzati secondo il tipo di percorso da seguire durante la compilazione del questionario; inoltre, sono integrati, a livello di impresa, con alcune informazioni statistiche (quali, ad esempio, numero medio di addetti e volume di affari) provenienti da ASIA, l'Archivio Statistico delle Imprese Attive, per capire se eventuali difficoltà nella compilazione siano in qualche modo collegabili a caratteristiche dimensionali dell'impresa rispondente.

A new tool for measuring customer satisfaction: The Anchoring Vignette approach

Omar Paccagnella
Statistics Department, University of Padua

In socio-economic surveys, collecting subjective evaluations of individual satisfaction, health status, living conditions or thoughts about certain aspects of the own life or society has become quite common. Researchers are then interested in comparing such self-reported assessments across countries, socio-economic groups or market segments. The main problem in such comparisons is that individuals in different countries or groups can interpret or understand the same question in different ways, because they use different scales to evaluate themselves. These interpersonal and inter-cultural differences in interpreting, understanding or using response categories for the same question is called differential item functioning (DIF). The approach of the "Anchoring Vignettes" has been recently introduced as a new and powerful way to detect the systematic differences in the use of the response scales within countries or socio-economic groups when evaluating themselves. This approach allows analysing self-reported ordinal survey responses taking into account individual differences in the interpretation of the survey questions. Each "vignette" depicts a hypothetical scenario, where a (hypothetical) person is described in particular situations or conditions. Respondents are asked to evaluate this individual, using the same scale adopted for self-rating. Thus, each respondent is provided with an "anchor" to adjust his/her self-evaluation. Consequently, these anchors can be used to make subjective assessments comparable across countries, socio-economic groups or market segments, because these subjective evaluations of all respondents are now reported to a common, DIF-free, scale. This approach relies on two assumptions. First, each individual uses the response categories for a question in the same way when providing a self-assessment as well as the evaluation of a vignette (response consistency). Second, the level of the variable represented in a vignette is perceived in the same way by all respondents apart from a random error (vignette equivalence); this implies that any difference across respondents on the vignette evaluations will be ascribed to DIF. The econometric specification of such a model is known as "chopit" (Compound Hierarchical Ordinal Probit) model and it mainly consists of two components modelling jointly self-assessments and vignette evaluations through some ordered probit models. DIF is modelled through variations in the thresholds. The thresholds determine the individual response scales and are allowed to vary with individual characteristics and across countries or groups. Despite the recent introduction in the literature, there are several and remarkable applications of this approach to various domains: from life satisfaction to job satisfaction, from work disability to self-reported health, etc. All findings support the ability of this approach to correct for DIF. The aim of this paper is twofold. On the one hand, it provides an overview of the anchoring vignette approach and its main features, showing how the marketing literature can be enriched by the application of this new methodology for the measurement of not-observable constructs, such as customer satisfaction, brand loyalty or product involvement. On the other hand, this approach is applied for measuring and comparing across different socio-economic groups the consumer satisfaction after the purchase of a smartphone, exploiting data collected through the VECS (Vignette Evaluation of Customer Satisfaction) module administrated within the LISS project. The LISS (Longitudinal Internet Studies for the Social sciences) panel is a new online panel of 5,000 Dutch households carried out by CentERdata. It focuses on fundamental longitudinal research and provides a laboratory for the development and testing of new, innovative research techniques.

Mixed mode survey su campioni probabilistici: un’alternativa alle indagini CATI e CAWI?

Teresio Poggio

Libera Università di Bolzano

Le indagini CATI orientate alla generalità della popolazione presentano problemi di copertura non più ignorabili. Si è accresciuta la quota delle famiglie italiane con solo telefoni cellulari, e non più un impianto telefonico fisso nell’abitazione. Tra chi possiede un telefono fisso, una quota importante di famiglie ha scelto inoltre di non comparire sugli elenchi telefonici e/o di iscriversi presso il Registro delle Opposizioni, rendendo il proprio numero di telefono inutilizzabile da chi svolge sondaggi e ricerche. Si stima che almeno il 40% della popolazione italiana adulta non sia coperta dalle indagini CATI, laddove – come è d’uso – si fa ricorso agli elenchi telefonici come lista di campionamento. Il dato si riferisce al 2009 e non considera gli ulteriori sviluppi in questa direzione, né le conseguenze dell’entrata a regime del Registro delle Opposizioni.

La diffusione dell’uso di internet nella società italiana ha proposto nell’ultimo decennio le indagini CAWI come una valida – e più economica – alternativa alle indagini CATI. Anche nel caso delle indagini via web, si pongono tuttavia seri problemi di copertura laddove l’interesse è sulla generalità della popolazione. Per quanto pervasiva, la diffusione di internet non ha infatti raggiunto allo stesso modo tutti gli strati della società italiana. Anche all’interno delle famiglie collegate in rete si rileva un livello di utilizzo differenziato – per età, genere e livello di istruzione – dello strumento internet. Sia nelle indagini CATI sia nelle indagini CAWI è prassi diffusa ricorrere a campioni non probabilistici assumendo una loro rappresentatività della popolazione di interesse. Si tratta tuttavia di un assunto non verificabile. Nel caso delle indagini CAWI rivolte alla generalità della popolazione, i meccanismi tipici di (auto)reclutamento dei rispondenti sul web sollevano ulteriori dubbi sull’effettiva rappresentatività dei campioni utilizzati.

Per le ragioni sopra esposte, è lecito nutrire dei dubbi sull’effettiva generalizzabilità dei risultati che emergono dalle indagini telefoniche e via web rivolte alla generalità della popolazione, così come sono tipicamente svolte. CATI e CAWI rimangono tuttavia modalità interessanti di raccolta dati, per la loro economicità (in termini di tempo, oltre che di denaro) e per la relativa semplicità di organizzazione e gestione logistica. Il paper discute la possibilità di utilizzare queste modalità di rilevazione in una prospettiva ‘mixed mode’, cercando di affrontare alcuni quesiti metodologici specifici:

- (a) È possibile – ed è sensato – separare le modalità di costruzione dei campioni e di contatto dei soggetti selezionati dalle modalità di raccolta dati, in modo da massimizzare la qualità dei dati raccolti (in termini di generalizzabilità) e minimizzare i costi della rilevazione?
- (b) È possibile affiancare a CATI e CAWI altre modalità di raccolta dati, più adatte a strati di popolazione che queste due modalità di indagine non riescono a raggiungere?
- (c) Quali sono le principali implicazioni metodologiche di un approccio mixed mode, rispetto alle questioni presentate?
- (d) Quali sono le principali implicazioni logistiche ed economiche di una scelta di questo tipo?
- (e) Quali le altre implicazioni per la qualità dei dati prodotti?

Nella discussione ci si avvale delle evidenze emerse dalla survey ‘Crescere a Trento: indagine sui servizi socio-educativi per la prima infanzia’. Si tratta di un’indagine campionaria, svolta a Trento nella Primavera 2010, su un campione probabilistico di 905 famiglie (delle quali 816 rispondenti), italiane e straniere. La survey è stata svolta con un approccio mixed mode CAWI + CATI + PAPI.

A test to assess the dynamic evolution of the preferences in marketing surveys

Angiola Pollastri, Francesco Porro

Dipartimento di Statistica e Metodi quantitativi, Università di Milano-Bicocca

In this paper we consider a two-stage hypothesis test, which has interesting applications in several situations related to the field of opinion research. Such test is based on the exact distribution of the absolute value of the minimum and on the exact distribution of the absolute value of the maximum of the components of a Bivariate Correlated Normal random variable. This test was proposed for the first time by Duncan [Miller, 1981], and it is recently improved by Pollastri [2008]. It needs sample sizes enough large, but nowadays this requirement is easily met in every opinion research.

In Pollastri [2012] the test is applied in the following situation. Let us suppose that there are two important products (or parties or opinions) indicated by A and B, respectively, and let C be the set of all the smallest remaining choices. The test can evaluate whether the percentage of preferences for A and the percentage of preferences for B are unchanged, increased or decreased after some event, for example a big investment in advertising. In other words, the test allows the comparison of the preference percentages at two different times.

In the present paper we describe a variant of the test, since we consider two samples in two different times. We analyse the case of two independent samples as well the case where the two samples are correlated.

An important example of the latter situation is panel data, where the same sample is examined in two different times.

In order to exhibit an application, a dataset regarding preferences of banks is considered.

Keywords: *trinomial distribution, Bivariate Correlated Normal, two-stage hypothesis test, panel data, Miller-Pollastri test.*

References

Miller, R. (1981). *Simultaneous Statistical Inference*. Springer, New York.

The Italian election polls over time: An analysis of the quality and completeness of the methodological aspects

Claudio Quintano, Antonella Rocca, Alfonso Carfora

Department of Statistics and Mathematics for Economic Research, University of Naples Parthenope

Election polls are an important source of political information both for politicians and voters, because they help to translate public preferences into policy, giving to decision makers impartial information about what the public wants, and helping public to understand where they fit into the political system.

In spite of their proliferation, due to the recent advances in technology (that have reduced the once formidable entry barriers) and to the increasingly insatiable appetite of cable network news, final users often can't be able to discriminating enough to know if a pollster, partisan or independent, is talented and conscientious and the real limitations of polls in this ever challenging environment.

When conducted through a scientific design, election polls represent the most successful political development of the last century, but a set of "best practices" or industry standards about their quality is likely to remain elusive. In Italy, the Authority for Communications Guarantees (AGCOM) fixed the information concerning the methodological aspects that should be provided for each poll.

Many Authors evaluated the accuracy of election polls in terms of a comparison between the poll estimates and the corresponding election results. This analysis is particularly suited to the polls published in the periods immediately preceding the election, because in the last days of electoral campaign significant changes on vote intentions can occur, especially where the share of undecided and of voters deciding in the last days of electoral campaign are high.

In this paper, taking into account that Italian legislation prohibits publishing election polls in the 15 days before election, which is a limitation for predictive analysis, an analysis based on their effective quality is made. For the first time, all the political polls, published in Italy since 2000, have been analyzed through statistical tools and models quality-oriented in terms of the completeness of the information provided and in terms of qualitative content of methodological aspects that have been made known. The analysis aims to identify good election polls and put them in relation with other aspects, as the type of polling company and the specific aim related to the poll, following the related performance over time, with the purpose to give to final users a real measure of their accuracy and credibility. In particular, an ordered logit model in which an ad hoc calculated index synthesizing qualitative and completeness requisites is regressed on some methodological characteristics of polls with the aim to identify the aspects related to the expected rigorous requirements. Moreover, to verify if in this last years an improvement in their quality is occurred, a comparison between the polls conducted in the period 2000-2009 (already analyzed in an our previous work) and those conducted later (2010-2012) is made through a synthesis of chi-square distances calculated between each poll and the ideal one (which is the poll satisfying all the accuracy and completeness requisites).

MICSI: Modelling Incentive Customer Satisfaction Index

Giampiero Remotti, Luigi Giampaoli, Andrea De Maria
GMR srl – GMC srl

La customer satisfaction è considerata da anni un obiettivo importante per aumentare la fidelizzazione e quindi il valore nel tempo della propria clientela.

Per questo motivo alcune organizzazioni strutturate sul territorio hanno adottato sistemi di misurazione oggettiva della CS e con essi dei modelli premianti della propria rete nel tentativo di migliorare la relazione con la clientela ed identificare le aree su cui concentrare gli sforzi futuri.

Se in linea di principio gli attori coinvolti in questo processo, la sede e la rete, condividono gli obiettivi, dal punto di vista pratico questo sistema non è immune da difficoltà che possono provocare tensione fra i soggetti coinvolti, soprattutto in quei settori, come per esempio l'automotive, dove la rete è spesso rappresentata da strutture con un elevato grado di complessità e potere contrattuale.

In questo caso l'utilizzo di strumenti di analisi statistica evoluta possono aiutare a identificare le aree di problematicità e aiutare a risolverle costruendo dei modelli di balancing che tengano conto di differenti peculiarità della clientela e permettano confronti più omogenei fra i diversi partner territoriali.

Nel presente lavoro viene presentato MICSI (Modelling Incentive Customer Satisfaction Index) modello che propone una soluzione finalizzata a migliorare il sistema di incentivazione di una azienda dell'automotive.

Nel caso specifico, il sistema originario presentava alcune peculiarità che tendevano a penalizzare le concessionarie più grosse, che sono anche quelle più importanti dal punto di vista del fatturato complessivo, in termini di minor numero di aziende premiate e di maggior distanza dai top-performant.

Seppur complesso nella sua formulazione, il sistema di incentivazione non considerava con sufficiente attenzione il profilo tipico della clientela di ciascuna concessionaria con il risultato di trascurare elementi importanti nella formulazione del giudizio di soddisfazione tali da non valorizzare a pieno lo sforzo dei dealer di maggior dimensione. MICSI ha permesso di evidenziare e correggere queste imprecisioni, suggerendo possibili alternative nella costruzione del modello di valutazione più allineate con le esigenze del produttore e dei singoli concessionari.

By choice or by force? The role of voters' age in the increase of abstention

Maria Francesca Romano*, Luana Russo**

*Scuola Superiore Sant'Anna, Pisa; **Université Lille 2

In Italy, voter abstention has been constantly increasing since 1979 (Russo, 2011), and the most recent Italian Parliamentary Elections (2008) saw the highest percentage of non-voters in Italian history (19.5%). Even if this might seem quite a low abstention rate in comparison with France (42.7% in 2012 – second round legislative), Spain (28% in 2011) or the UK (34.9% in 2010), it is quite high for Italy. From the beginning of the Italian Republican era (1946) until 1979, the abstention rate has remained at around 6-7%, and we can label this as inevitable abstention: there will always be a proportion of the voting population who cannot vote. A portion of voters, according to some the 20% of all who abstain (Manneheimer and Sani 2001), are prevent from voting by circumstances beyond one's control (invalid, sick, very old people, people resident elsewhere etc). In fact, even in countries where voting is mandatory (i.e. Belgium) it is possible to observe a small proportion (under 10%) of abstainers. This no-vote rate is unavoidable, and we can assume that the main reason for not voting is not down to choice. Because of the increasing of abstention by choice (Raniolo 2002, Tuorto 2006, Tuorto e Colloca 2010) very little attention was given to the *forced* abstention. It might be the case that with the ageing of population these form of abstention is increasing as well.

In this paper we want to analyse the electoral abstention phenomenon in Italy taking into account the age structure of the Italian population. The aim of our proposal is trying to understand how the rapid ageing of the Italian population can contribute to explain the overall quota of abstention. In fact in Italy we can observe two complementary dynamics: in the last twenty years the young age classes are decreasing whilst the elderly classes (over 80 years) are increasingly composed by women. Moreover, there is a large variability among areas with regard to these demographical aspects. We will test the hypothesis that the abstention (which is here indented in a strict sense: only voters not going to the ballots, excluding the non-valid vote) tends to be higher as the population of a given areas is older. We will test this hypothesis for five Parliamentary elections (1992-2008) by using ecological data at the municipal level and by employing as explanatory variable the demographic structure (age and gender). The results will be validated by using the electoral records regarding the percentage of abstention by age and gender of three municipalities (one for each of the main geographical areas: North, Centre and South).

Heterogeneity and Capture-Ricapture method in a closed population: A new parameterization to treat continuous covariates in a Poisson Log Linear Model

Giuseppe Rossi*, Pasquale Pepe*, Olivia Curzio*, Marco Marchi**

**Institute of Clinical Physiology, CNR and G. Monasterio Foundation;*

***Department of Statistics “G. Parenti”, University of Florence*

The capture-recapture method is used by epidemiologists to estimate the size of hidden populations using incomplete and overlapping lists of cases. In our study, we specifically considered closed discrete-time models with dependence between the sources and heterogeneity. The main objective of this work was to propose a parameterization that allows the Poisson Log Linear Model (LLM) to treat continuous covariates in their original measurement scale. The proposed model was applied to simulated and real capture-recapture data sets and compared with the Multinomial Conditional Logit Model (MCLM) used as a reference model. The results showed that the proposed model was comparable to the MCLM in dealing with continuous covariates. Furthermore, the proposed approach allowed us to easily analyze the capture-recapture data using commonly available software.

Keywords: Capture-recapture; Closed population; Continuous covariates; Log-Linear Model; McFadden parameterization; Multinomial Conditional Logit Model

La statistica e l'arricchimento qualitativo: due facce della stessa medaglia

Francesca Soldi
Dupuntozero Research

Qualcuno sostiene che gli artefatti tecnologici abbiano, nella società, la funzione che nelle reazioni chimiche hanno gli enzimi [Fraia 2011]: questa visione è per noi condivisibile anche in riferimento a quanto la stessa tecnologia abbia influenzato i processi di ricerca, scardinando quei trade off ormai giudicati da tempo come assunti indiscutibili. La diffusione e l'evoluzione del digitale ha, infatti, significato uno sviluppo anche nelle ricerche di mercato e di opinione obbligando gli studiosi a ripensare ai processi, alle tecniche e alla loro integrazione.

Agire all'interno delle ricerche di mercato o di opinione tramite i metodi di raccolta tradizionali implica un'attenta valutazione delle metodologie di rilevazione e analisi in base agli obiettivi che ci si è posti, assumendosi il rischio di perdere i vantaggi della soluzione scartata. Nella costruzione del modello dell'indagine, il ricercatore, fra le altre cose, si trova spesso a dover compiere una scelta a monte di tutto: scegliere se applicare un approccio qualitativo o quantitativo e, a cascata, metodi e tecniche di diversa natura e applicazione.

Questo fino ad oggi: le nuove tecnologie permettono, in alcuni casi, di evitare una scelta esclusiva a questo primo bivio, permettendo, invece, di percorrere sentieri intermedi che colleghino le due strade principali.

La tesi del paper è che il mondo digitale abbia permesso, proprio al pari di un enzima, di ampliare le possibilità di ricerca e di analisi offrendo un'ampia produzione di materiali "secondari" e la nascita di nuovi punti di contatto fra un approccio quantitativo e uno qualitativo.

Il digitale permette di partire da una mole di dati molto ampia per concentrarsi su aree di significato rilevanti e darne letture approfondite e a carattere interpretativo. La statistica, in questo percorso può avere un ruolo di assoluto rilievo: può accompagnare il processo di analisi sistematizzando, raggruppando e gestendo una grossa mole di dati, permettendo al ricercatore di individuare le aree di interesse per un'analisi qualitativa favorendo l'emergere dell'interpretazione più profonda.

In termini concreti, il paper mostrerà due strumenti atti a questo scopo: le analisi dei dati Open Source e il T@lk, tecnica proprietaria dell'istituto Duepuntozero Research.

Nel primo caso si tratta di utilizzare la statistica per descrivere il fenomeno di produzione di testi spontanei online e valutarne l'andamento nel tempo, il topic oggetto di discussione e il tono. In questo caso, impiegare strumenti di statistica descrittiva permetterà di individuare quelle aree "anomale" su cui applicare un approfondimento qualitativo.

Nel secondo caso, lo strumento T@lk, impiegato all'interno di questionari semi-strutturati, permette di coniugare analisi testuale quantitativa e qualitativa integrando i due approcci e portando a definire un'interpretazione completa del dato.

Le tecniche tipiche della ricerca qualitativa e quantitativa non si sommano, ma si mescolano nello strumento: proiettivi qualitativi vengono usati per stimolare le risposte; le analisi di statistica testuale permettono di classificare e sintetizzare le molte informazioni rilevate; le letture motivazionali sono infine impiegate per l'interpretazione dei nodi caldi di significato emersi nelle analisi quantitative.

Intendiamo portare un esempio per ciascuna metodologia con lo scopo di sottolineare le possibili integrazioni e interazioni fra gli aspetti qualitativi e gli aspetti quantitativi nella ricerca di mercato o di opinione.

Sustainability and textual network: analysis of the Italian blogs

Emma Zavarrone*, M. Gabriella Grassia, Davide Jabes*, Sandra De Francisci***

**IULM; Milano, **Università di Napoli “Federico II”*

The concept of sustainable development is something relatively new. The first definition goes back to 1987, and is provided by the Brundtland Commission. While the idea of sustainability seems relatively clear (the term comes from “to sustain”, and therefore implies the effort to develop within the carrying capacity of the system where it belongs, in this case the earth), it's also true that the measures are still too vague, and this lack of knowledge makes the concept of sustainability an highly problematic concept.

The purpose of this research is to carry out an initial exploratory analysis in order to investigate how the construct of sustainability on the web is configured, which are the keywords and which are the related constructs. The collected texts dealing with the issue of sustainability raised by the main Italian blogs during the first quarter of 2011. The network textual analysis is developed in two steps: i) the pre-processing blog contents and vocabulary definition (R package - tm), ii) defining adjacency matrix and performing two-mode network analysis (R package - sna) in order to obtain the network of bloggers and the network of keywords.

