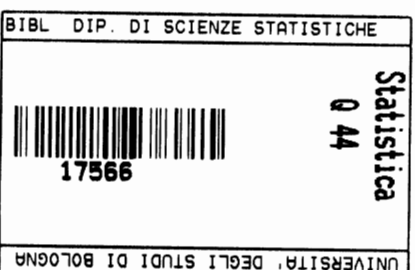


Cristina Biondi

La Statistica in Rete: un'indagine on-line
sul fenomeno e-commerce nel settore
italiano dell'editoria e stampa

Serie Ricerche 2002, n.4



Dipartimento di Scienze Statistiche "Paolo Fortunati"
Università degli studi di Bologna

Indice

- | | |
|---|--------|
| 1) Introduzione | pg. 5 |
| 2) Obiettivi e metodologia della ricerca | pg. 5 |
| 3) Analisi descrittiva dei principali risultati emersi dall'indagine | pg. 25 |
| 4) Differenze tra utenti che hanno risposto prima e dopo i solleciti | pg. 50 |
| 5) Analisi delle corrispondenze multiple | pg. 54 |
| 6) Considerazioni conclusive, problemi aperti e possibili sviluppi di ricerca | pg. 60 |
| 7) Riferimenti bibliografici | pg. 65 |

Finito di stampare nel mese di Novembre 2002
presso le Officine Grafiche Tecnoprint
Via del Legatore 3, Bologna

1. INTRODUZIONE

L'idea di progettare e realizzare un'indagine on-line è emersa col proposito di superare un limite di misurazione di un fenomeno telematico come l'e-commerce, che trova il suo luogo naturale di svolgimento esclusivamente attraverso la rete Internet. L'attuazione di questa ricerca, inoltre, si è dimostrata molto utile per verificare la validità e soprattutto il tasso di penetrazione di un'indagine di questo tipo, in modo da poterne trarre anche un eventuale futuro confronto con altre tipologie di rilevazione più tradizionali, sempre di tipo auto-amministrato. La struttura del lavoro è formata da una prima parte in cui vengono descritti i principali obiettivi dell'indagine, la metodologia di contatto impiegata e il collettivo di unità statistiche indagate con tutte le varie problematiche che lo riguardano. Nella seconda ci siamo poi addentrati in un'analisi piuttosto dettagliata dell'informazione emersa dalla compilazione dei questionari ed infine nell'ultima parte si è cercato di approfondire la conoscenza dei profili dei rispondenti utilizzando particolari tecniche di analisi statistica multivariata.

2. OBIETTIVI E METODOLOGIA DELLA RICERCA

2.1 - Obiettivi della ricerca

Questa ricerca nasce con l'obiettivo di studiare e monitorare le attività e il comportamento delle aziende italiane in Rete per cercare di capire come, in questi ultimi anni di esplosione della net-economy, queste hanno reagito ai fenomeni dell'e-commerce e del web marketing. I settori di attività economica ai quali avremmo potuto rivolgere l'attenzione erano diversi. In questa prima esperienza, abbiamo cercato di focalizzarci su uno in particolare di essi: quello dell'Editoria e Stampa, settore che ad una prima analisi, realizzata sull'archivio on-line di Infoimprese¹, ha manifestato di essere particolarmente attivo sul fronte del commercio elettronico.

¹Infoimprese è una società consortile di informatica delle Camere di Commercio Italiane che ha realizzato un archivio online (www.infoimprese.com), a consultazione gratuita, di tutte le aziende italiane registrate presso la C.C.I.A.A., suddivise per settore di attività economica.

Il progetto chiamato "LA STATISTICA IN RETE" ha preso corpo attraverso la realizzazione di un'indagine on-line, cioè un'indagine che prevedeva una metodologia di rilevazione attraverso la rete Internet, con l'impiego di due strumenti tipici del World Wide Web: indirizzi di posta elettronica, necessari per stabilire un contatto con le unità indagate, e sito web, per il posizionamento del questionario elettronico in Rete.

Il questionario è stato costruito con l'obiettivo di acquisire maggiori informazioni sul tipo di attività telematica delle aziende di questo settore ed in parte per confermarne altre già rilevate in precedenza dal database della Camera di Commercio. Il questionario in tutto si componeva di 35 item, per la maggior parte di tipo qualitativo, per i quali si è cercato di alleggerire al massimo il carico di lavoro richiesto al rispondente. In particolare venivano richieste informazioni connesse al tipo di attività economica che queste aziende svolgono attraverso la rete Internet, che riguardano sostanzialmente i seguenti aspetti:

- le principali caratteristiche delle aziende che hanno deciso di operare su questo nuovo mercato virtuale;
- le principali motivazioni che hanno influito sulla scelta strategica di realizzare un sito web aziendale;
- identificazione e possibilità di quantificare il gap tra le aspettative delle aziende, nei confronti di questo nuovo mezzo di comunicazione e vendita con i risultati effettivamente ottenuti, sia in termini di notorietà del proprio prodotto/marca che in termini di veri e propri incrementi delle vendite;
- punti di forza e di debolezza delle iniziative legate alle attività di commercio elettronico e della rete Internet in generale.

L'intento è stato quello di costruire un questionario abbastanza lineare che ricalcasse in molte parti le sembianze di un questionario cartaceo, in modo da non disorientare eventuali rispondenti inesperti o comunque cercando di rimuovere il più possibile l'effetto del questionario elettronico sul successo dell'indagine. Proprio per questo motivo, l'intero questionario è stato costruito su un'unica pagina web, al termine della quale veniva richiesto al rispondente di confermare l'invio dei dati, azione questa che generava automaticamente l'apertura di un'ulteriore

pagina web, di ringraziamenti per la collaborazione assicurata. Per incoraggiare la partecipazione degli utenti contattati si è pensato inoltre di non impostare un meccanismo di riconoscimento tramite username e password, semplicemente, all'interno di ogni messaggio e-mail inviato, veniva evidenziato l'indirizzo web della pagina sulla quale era posizionato il questionario, impostato sotto forma di collegamento ipertestuale in modo tale da facilitare ulteriormente l'eventuale accesso ai potenziali rispondenti.

2.2 - *Durata dell'indagine e metodologia di contatto*

La lista di riferimento adottata per l'identificazione delle unità di interesse, è stata estratta dall'archivio di Infoimprese, il quale è a consultazione gratuita e visualizzabile direttamente on-line al sito www.infoimprese.com. Tale archivio contiene al suo interno informazioni di base su tutte le aziende italiane registrate presso la C.C.I.A.A., suddivise per settore di attività economica. Inoltre, a tutte le aziende in esso registrate, è consentito di rilasciare specifiche informazioni sulla propria attività, il proprio sito internet ed altri eventuali indirizzi o riferimenti tramite i quali poter essere contattate.

Questo database è stato il punto di partenza per la creazione di un elenco più ristretto di aziende "attivamente presenti su Internet" e appartenenti al settore dell'editoria e stampa (E.S). Il termine "attivamente presenti" in questo contesto, viene utilizzato per identificare aziende che, al momento dell'indagine risultano aver attivato un proprio sito internet e possiedono almeno un indirizzo di posta elettronica.

Prima di passare all'analisi più dettagliata della metodologia di contatto, è necessario soffermarci sulle scelte strategiche effettuate nella scelta del collettivo da contattare. Infatti, pur essendo in presenza di una lista di riferimento, verosimilmente completa ed aggiornata, della popolazione oggetto di indagine, si è preferito effettuare un censimento di tali aziende piuttosto che un vero e proprio campionamento. Questo per due motivi: la contenuta numerosità della popolazione di aziende oggetto di analisi, che nel complesso non supera le 1.000 unità e i bassi costi di contatto e di rilevazione consentiti dall'utilizzo della posta elettronica.

In totale, quindi, le aziende individuate ai fini della rilevazione sono state 981. Ad ognuna di esse è stato rivolto l'invito a partecipare



all'indagine rispondendo ad un questionario elettronico posizionato sul sito Internet della Facoltà di Scienze Statistiche (Bologna- sede di Rimini). L'indagine si è svolta su un periodo totale di 21 giorni: è iniziata il 4 febbraio 2002 ed è terminata il 24 dello stesso mese. In realtà possiamo dire che il periodo temporale effettivo è stato ancora minore, esattamente 20 giorni, in quanto per un'intera giornata, a causa di problemi tecnici di funzionalità del server, la pagina del questionario non risultava accessibile.

La decisione di realizzare l'indagine in un periodo di tempo di sole tre settimane è nata dall'idea di poter verificare un'ipotesi a priori sui vantaggi acquisibili con l'implementazione di indagini on-line. Tale ipotesi è stata proposta nell'ottica di un possibile confronto tra questa metodologia di indagine e altre sempre di tipo auto-amministrato che utilizzano però tecniche di contatto più tradizionali, come ad esempio i sondaggi postali. L'idea si basa sostanzialmente sul fatto che a parità di tassi di risposta il periodo di indagine si riduce notevolmente, e viceversa, a parità di tempi, con un'indagine on-line si è in grado di ottenere tassi di risposta molto migliori.

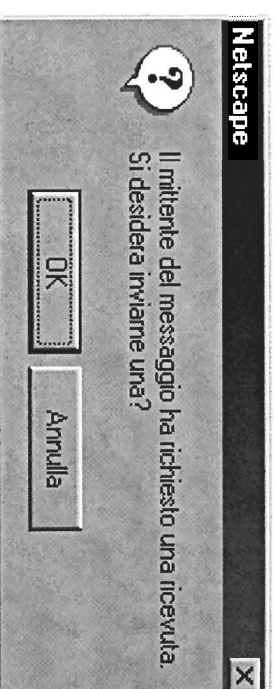
Basandoci, inoltre, sulla considerazione che una e-mail diventa "vecchia" in tempi più rapidi di una lettera ordinaria, si è deciso di pianificare due solleciti a distanze di tempo abbastanza ravvicinate tra di loro.

Durante la mattinata del giorno 4 è stato inviato, al collettivo di aziende indagate, per e-mail un primo messaggio di invito, contenente al suo interno, oltre all'informativa della ricerca in atto, anche un *link*² alla pagina web del questionario e un indirizzo di posta elettronica al quale potersi rivolgere in caso di chiarimenti o richiesta di ulteriori informazioni sulla ricerca. Nelle due settimane successive sono stati poi pianificati e realizzati due solleciti, uno ad ogni inizio settimana, eliminando di volta in volta le unità statistiche che avevano manifestato una risposta, sia positiva che negativa, all'indagine.

Solo dalla seconda settimana è stato impostato un meccanismo di ricevuata di ritorno, a conferma di un'avvenuta lettura del messaggio di sollecito inviato. Tale ricevuata ha permesso di ottenere una preliminare informazione sul numero di contatti avvenuti anche se, come abbiamo

² Collegamento ipertestuale alla pagina web su cui è stato posizionato il questionario elettronico, che se cliccato permette di visualizzare direttamente la pagina.

potuto appurare in seguito, un'informazione di questo tipo non può essere ritenuta del tutto affidabile. Infatti, i problemi a cui va incontro sono diversi: il primo tra essi è il fatto che il ricercatore non riesce ad avere il pieno controllo su questa fase, dato che è il rispondente stesso a decidere se confermare o meno l'invio di una ricevuata di ritorno nel momento in cui, aprendo la posta elettronica, gli compare a video un messaggio di questo tipo:



Ipotizzando quindi che, non tutti coloro che hanno ricevuto e letto il messaggio abbiano anche confermato l'invio di una ricevuata di ritorno, con questo metodo non conosceremo mai con certezza il numero effettivo di utenti con i quali siamo realmente riusciti a stabilire un contatto. Inoltre, grazie anche ad ulteriori approfondimenti sul campo, abbiamo potuto renderci conto che tale meccanismo non funziona sempre ed in uguale modo per tutti i programmi di gestione di posta elettronica. Talvolta il server ricevente è programmato per spedire automaticamente un messaggio di ritorno ogniqualvolta ne riceva uno, senza verificare che l'indirizzo sia effettivamente attivo o che l'utente interessato abbia almeno aperto l'e-mail.

Proprio per questi motivi nella fase successiva di elaborazione dei tassi è stato necessario completare e talvolta sostituire questa informazione con quella più affidabile archiviata nei *file di log*, che sono file di sistema del server che permettono di registrare tutte le azioni compiute da chiunque si colleghi alle pagine web contenute sul server e quindi anche alla pagina web del questionario elettronico.

All'interno di ogni messaggio e-mail, sia il primo che i successivi due solleciti, il potenziale rispondente è invitato a "cliccare" sul link, che gli

viene seganato, per accedere direttamente alla pagina contenente il questionario elettronico. All'interno di tale pagina la prima schermata mostra chiaramente il logo della Facoltà di Scienze Statistiche, con relativo riferimento al progettista dell'indagine e, come incentivo alla partecipazione, la possibilità offerta ad ogni rispondente di ricevere gratuitamente una sintesi dei principali risultati emersi dall'indagine.

Il primo invio ha messo subito in evidenza una situazione già preventivata di precaria affidabilità di alcuni indirizzi e-mail delle aziende indagate. Infatti, alcuni di essi hanno dato chiara evidenza di non essere raggiungibili al momento dell'indagine; fenomeno questo che abbiamo ricondotto sostanzialmente a due motivi:

- tali indirizzi di posta elettronica non erano corretti;
- oppure risultavano irraggiungibili per cause tecniche legate alla cattiva funzionalità del server.

La prima reazione è stata quella di riprovare a stabilire un contatto nelle giornate successive, facendo precedere a questa fase un accurato controllo della corrispondenza tra indirizzo e-mail a nostra disposizione, estrapolato dal sito di Infoimprese, e quello riportato invece all'interno del sito web dell'azienda stessa. Si riportano di seguito alcune statistiche riassuntive che mostrano la situazione iniziale del collettivo di aziende indagate e la sua successiva evoluzione e riduzione relativa alla prima settimana.

TABELLA 1: *Evoluzione dei contatti durante la prima settimana di indagine*

Statistiche riassuntive	n	%
Numerosità del database di origine	981	
E-mail tornate indietro perché non raggiungibili	138	14.07%
E-mail apparentemente recuperate	63	6.42%
Ulteriore ricaduta di contatto	15	1.53%
Totale e-mail recuperate	48	4.89%
Totale e-mail eliminate dal database di origine	90	9.17%
Numerosità finale del database al netto delle cadute	891	90.83%

È interessante notare che inizialmente la "caduta" dei contatti appariva molto alta, ben il 14% del totale; la fase di ricontatto ha permesso poi di sanare un buon 4.89% delle e-mail risultate inizialmente irraggiungibili (probabilmente la maggior parte di quelle che avevano riscontrato problemi tecnici del server di posta), facendo emergere quindi un tasso netto di caduta del 9,17% che comprende gli indirizzi e-mail risultati irrecuperabili.

Questi ultimi sono stati quindi definitivamente eliminati dal database di origine. La causa di questo elevato tasso di caduta degli indirizzi può essere ricondotta alla qualità della lista utilizzata per la costruzione del database, ed in particolare al fatto che dal momento della pubblicazione di tali indirizzi sul sito di Infoimprese al momento del lancio dell'indagine, l'indirizzo di posta elettronica e talvolta anche il sito Internet di alcune aziende aveva subito cambiamenti o non risultava più attivo. La numerosità iniziale del collettivo di aziende si è quindi ridotta da 981 unità a 891, nuovo valore a cui abbiamo fatto riferimento per il calcolo del tasso di risposta complessivo dell'indagine.

Va precisato inoltre che anche nelle due settimane successive alla prima, esattamente quelle in cui sono stati inviati i solleciti, si sono verificate altre cadute, in numero molto minore, come era lecito attendersi, dovute per lo più a problemi tecnici del server ricevente. Per queste ultime si è quindi deciso di non eliminare le corrispondenti unità statistiche dal database, in quanto, non avevano dato alcuna evidenza di mancato contatto durante la prima settimana di indagine, abbiamo dato per certo che un contatto fosse già effettivamente avvenuto.

Nella seguente tabella riportiamo i tassi di caduta verificatisi nella seconda e nella terza settimana, calcolati rispetto al totale degli invii fatti per ogni rispettiva settimana.

TABELLA 2: *Tassi di caduta delle e-mail nelle settimane successive alla prima*

Seconda settimana	
Solleciti inviati	776 Tx
Caduta	6 0.77%
Terza settimana	
Solleciti inviati	693 Tx
Caduta	4 0.58%

2.3 - Il questionario

Il questionario elettronico non può essere improvvisato, ma richiede una rigorosa progettazione e verifica (Pierchala, Manners 1996). Le diverse fasi, necessarie per la sua costruzione, devono essere coordinate e implementate attraverso il contributo di diverse competenze sia dal punto di vista statistico che informatico. In generale, si procede quindi facendone una prima bozza su carta o su computer e cominciando poi la sua implementazione solo nel momento in cui si ha una chiara idea, sia del tipo di domande da porre che delle modalità di risposta da utilizzare. Il questionario elettronico offre la possibilità di definire, al suo interno, oggetti diversi da abbinare a modalità di risposta diverse e questo è sicuramente un punto di forza; tale meccanismo rappresenta infatti un controllo di primo livello sulla qualità dei dati prodotti che si innesca al momento della compilazione.

La progettazione di un questionario di questo tipo, oltre ad avere una serie di implicazioni tecniche non indifferenti, porta con sé il vantaggio di eliminare quasi completamente l'interazione tra intervistatore e intervistato e di riconsiderare quindi molti degli aspetti legati alla gestione e ai costi delle risorse umane. Viene a mancare quasi completamente la ricerca di un contatto diretto con le unità indagate, ma assume un ruolo fondamentale l'eventuale fase di assistenza durante la compilazione, su richiesta del rispondente.

Può essere prevista un'assistenza di tipo telefonico, ad esempio nel caso in cui si ha a che fare con utenti non sempre esperti di comunicazione telematica; in generale, però, la tecnica di contatto preferita ed anche più spesso utilizzata per indagini di questo tipo è quella che si avvale di messaggi e-mail.

Per renderci conto di quali siano realmente le fasi salienti della realizzazione di un'indagine integrata tramite e-mail e web, è possibile schematizzarla in senso cronologico come segue:

- * analisi preliminare e studio di fattibilità dell'indagine;
- * predisposizione del disegno campionario;
- * progettazione del questionario;
- * indagine pilota e taratura dello strumento di rilevazione;

- * invio della lettera di invito a partecipare all'indagine tramite messaggio e-mail;
- * invio di solleciti sempre tramite messaggi e-mail;
- * raccolta e ricodifica dei dati;
- * controlli;
- * nuovo contatto con alcune unità del collettivo indagato per eventuali integrazioni imputabili ad errori dei rispondenti non correggibili dai controlli implementati;
- * elaborazione ed analisi dei dati;
- * presentazione dei risultati della ricerca.

Un ruolo fondamentale è ricoperto dai diversi tipi di controlli implementabili attraverso la realizzazione di questionari elettronici su web. È ben noto che per indagini tradizionali, terminata la fase di raccolta dei dati, inizia quella di data-entry e quindi anche quella di individuazione degli errori commessi dai rispondenti durante la compilazione e, ove possibile, la loro correzione secondo un prefissato schema di compatibilità. Tutto questo richiederà all'organizzatore dell'indagine un enorme sforzo sia in termini di tempo sia a livello progettuale.

Non deve essere trascurato inoltre il fatto che qualunque correzione venga fatta questa avrà sempre una probabilità non nulla di essere ancora una volta errata, anche se l'intervento del ricercatore si è basato su ragionevoli presupposti per effettuare un'adeguata correzione.

Un primo tipo di controlli implementabili per un'indagine su web, è quello già citato precedentemente, che permette di definire caselle di risposta diverse a seconda che le domande siano di tipo "chiuso" e "a risposta singola" oppure "a risposta multipla". Altri tipi di controlli sono quelli che permettono l'adozione di particolari domande filtro, attraverso le quali è possibile impostare un salto, più o meno forzato, ad altra domanda o sezione del questionario per tutti quegli utenti che selezionano una certa modalità di risposta piuttosto che un'altra.

Non bisogna però dimenticare che si sta "costruendo" un questionario: se sicuramente gli strumenti informativi permettono potenzialità superiori, la logica di strutturazione, il coordinamento tra gli argomenti

