

Tempestività e qualità: aspetti concettuali e comportamenti di risposta nelle indagini via Internet. Un'applicazione ad un'indagine regionale sulle imprese.

Silvia Biffignandi, Monica Pratesi
Dipartimento di Matematica, Statistica, Informatica e Applicazioni
Via dei Caniana n.2, 2 I-24127 Bergamo
silvia@unibg.it

1. Introduzione

Il futuro delle indagini statistiche è sicuramente legato all'uso di Internet. La crescente importanza dei collegamenti in rete nel campo delle indagini è illustrata da un vivace dibattito nella letteratura corrente che affronta i vari aspetti del disegno delle indagini Web distinguendo le indagini basate su liste di indirizzi di posta elettronica, da quelle rivolte ad un insieme non identificato di utenti di Internet¹. La discussione sul ruolo della rete nella raccolta di dati statistici è tutt'ora aperta a contributi che definiscano compiutamente il quadro metodologico di riferimento e soprattutto individuino ed analizzino le dimensioni della qualità dei dati raccolti (Batagelj et al. 1998b; Couper et al., 1999; Clayton and Werking, 2000).

Senza entrare nel dibattito sulle dimensioni della qualità e sui diversi punti di vista che guidano le specifiche definizioni², ci limitiamo a sottolineare che la qualità dell'informazione statistica deriva sia dalla qualità del prodotto finale, i dati, sia dalla qualità del processo di produzione, il disegno dell'indagine statistica. La diffusione della cultura tecnologica e la penetrazione di Internet nella popolazione dei possibili rispondenti (famiglie o imprese) ha conseguenze su diverse dimensioni della qualità del disegno d'indagine. Fra queste la tempestività della raccolta dei dati è considerata uno dei vantaggi più competitivi di Internet nei confronti degli altri strumenti d'indagine. La tempestività è una caratteristica sia del prodotto finale sia del processo di produzione³. Nel primo caso, la tempestività ha a che vedere con la frequenza o la periodicità dei dati⁴; nel secondo caso, essa è connessa alle procedure d'indagine, in

¹ Per illustrazioni esaurienti della classificazione delle indagini via Internet si rimanda a Batagelj et al. 1998a, Couper 2000, Biffignandi e Pratesi 2000a, Bosnjak e Tuten, 2001.

² La letteratura al proposito è abbondante come citato in Biffignandi S. nel presente volume. Con riferimento alle definizioni adottate a livello europeo citiamo i lavori del "European LEG (Leadership Expert Group) on Quality" e la "International Conference on Quality in Official Statistics", Stoccolma, 14-15 maggio, 2001.

³ Tale classificazione è citata in Encyclopedia of Statistical Sciences, Quality Concept for Official Statistics (Elvers, Rosén, 1997).

⁴ La frequenza si riferisce a dati statistici prodotti secondo uno schema regolare; essa include la frequenza del tempo di riferimento (reference time frequency), vale a dire la periodicità del tempo di riferimento per dati

particolare è collegata alla lunghezza del periodo di rilevazione, alla velocità degli strumenti di raccolta e memorizzazione dei dati. Si dovrebbe osservare che l'accuratezza dei dati spesso entra in conflitto con la tempestività. Un periodo di rilevazione troppo breve può portare ad un incremento nel tasso di non risposta, un'eccessiva velocità nel processo di raccolta e memorizzazione può aumentare il rischio di errori ed imprecisioni.

L'uso di Internet insieme ai sistemi di intervista assistiti da calcolatore (Computer Assisted Web Interviewing , CAWI) apre ampi margini per l'aumento della velocità di indagine. La memorizzazione dei dati raccolti è infatti contestuale alla compilazione del questionario Web. Inoltre, controlli pre-impostati della sequenza logica della compilazione del questionario e del contenuto dei campi di risposta evitano in tempo reale il rischio di errori ed imprecisioni. In poche parole, la tempestività di comunicazione permessa da Internet potrebbe in teoria consentire la compressione della durata del periodo di rilevazione tenendo sotto controllo, al tempo stesso, almeno alcune fonti di possibili incoerenze ed incompatibilità nei dati raccolti. Ciononostante, in molte indagini via Internet, il periodo di rilevazione nel suo complesso è lungo, si parla spesso di mesi, i tempi individuali di risposta sono ben diversi da zero ed eterogenei tra i soggetti contattati.

Per adattare le tradizionali linee guida della metodologia di indagine al contesto delle indagini Web, è, a nostro avviso, necessario studiare il ruolo del tempo nel comportamento di risposta dei soggetti coinvolti in un'indagine via Internet. Questo lavoro focalizza l'attenzione proprio su questo, studiando l'effetto del piano dei solleciti nelle indagini via Internet sulle imprese. La pianificazione dei solleciti è cruciale per aumentare i tassi di risposta nelle indagini faccia a faccia, telefoniche e postali; ed è ragionevole che il suo ruolo sia fondamentale anche nel caso di indagini Web condotte su soggetti identificati tramite una lista di indirizzi di posta elettronica.⁵ La struttura del piano dei solleciti si ripercuote generalmente sulla lunghezza del periodo di rilevazione; si tratta, allora, di un aspetto sicuramente da curare per ottenere un processo di indagine tempestivo. Nel lavoro, ipotizziamo che la durata individuale del periodo di rilevazione ed il comportamento dei rispondenti durante il periodo di rilevazione siano influenzati dalle loro caratteristiche individuali ed usiamo quest'ultime per la specificazione di modelli di durata per il rischio di rispondere senza sollecito o dopo uno o più solleciti.

Le principali ragioni che, a nostro avviso, rendono necessaria l'analisi di sopravvivenza nel processo di risposta nelle indagini via Internet sono illustrate nel paragrafo 2. Il

pubblicati, la frequenza di diffusione (dissemination frequency), vale a dire la periodicità con cui i dati statistici sono resi pubblici.

⁵ Notiamo che in alcune indagini Web - basate o meno su di una lista di indirizzi di posta elettronica - nel corso delle quali il questionario da compilare è stato inviato un'unica volta, il basso tasso di risposta ottenuto è stato interpretato come un fallimento di Internet come strumento di raccolta dati. Esperimenti condotti con piani di sollecito hanno provato che possono essere raggiunti gli stessi tassi di risposta ottenuti con analoghe indagini postali; per quello che riguarda la tempestività si può mostrare che la maggior parte delle risposte sono state ottenute più velocemente che nelle indagini postali (Schaefer and Dillman, 1998, Dillman, 1998). Metha e Sividas (1995) e Smith (1997) hanno confrontato indagini Web con un solo contatto con indagini Web con più di un contatto; entrambi i lavori hanno mostrato guadagni nei tassi di risposta.

paragrafo 3 è dedicato alla specificazione di un modello a rischi concorrenti per studiare gli effetti del piano dei solleciti. Nel paragrafo 4 sono illustrati i principali passi del processo produttivo dell'indagine da noi condotta e sui cui dati è stato applicato il modello proposto. Nello stesso paragrafo sono presentati anche i risultati dell'applicazione del modello, dai quali emergono considerazioni e spunti per il trattamento della tempestività nelle indagini via Internet.

2. Il ruolo del tempo nel processo di risposta delle indagini via Internet

Le indagini via Internet sono condotte con vari criteri. Con riferimento alla dimensione temporale del comportamento dei rispondenti, la distinzione principale è connessa alla strategia di contatto adottata. La classificazione di base fa riferimento alla distinzione tra contatto raggiunto attraverso un invio personalizzato del messaggio di invito e contatto raggiunto indirettamente attraverso la "navigazione". In tal caso il contatto è raggiunto quando l'auto-intervista è proposta al navigatore prima di entrare nel sito o prima della sua registrazione ad un gruppo. In questo lavoro restringiamo l'attenzione alla strategia di contatto personalizzata in presenza di piano dei solleciti, in particolare ci riferiamo ad una procedura di contatto personalizzata via posta elettronica. In questo caso, sono possibili due diverse procedure di accesso al questionario: il questionario può essere inviato in allegato al messaggio di invito via posta elettronica oppure nell'invito può essere trasmesso l'indirizzo Internet a cui il rispondente può compilare il questionario Web. Le nostre analisi empiriche si riferiscono a quest'ultimo caso.

Il comportamento del rispondente dopo il messaggio elettronico di invito è quindi, dal nostro punto di vista, la nuova prospettiva per la definizione del periodo d'indagine. L'intero periodo di rilevazione è il periodo tra il primo messaggio di invito e l'ultima auto-intervista. La lunghezza individuale del periodo d'indagine dipende da molti fattori ed è anche il risultato della reattività del soggetto contattato al piano stabilito dei solleciti.

Data la risposta del soggetto, l'attenzione è sulla lunghezza del periodo T_i tra il messaggio di invito e l'auto-intervista. I candidati a partecipare all'indagine possono completare il questionario in pochi giorni dal primo messaggio di invito o possono completare dopo alcuni messaggi di sollecito, alla fine del periodo d'indagine. In questo contesto, è importante esplorare la reattività dei rispondenti ai messaggi inviati da chi organizza l'indagine: la comunicazione con l'azienda avviene on-line ed è possibile mandare molti messaggi di sollecito (stimoli) per incoraggiare e promuovere la partecipazione all'indagine.

L'effetto dei messaggi di sollecito può essere esplorato attraverso l'analisi di sopravvivenza delle risposte (Cox and Oakes, 1984). In tal caso, l'evento di interesse è l'auto-intervista dopo il sollecito e lo scopo è determinare come l'occorrenza ed il momento della risposta dipenda da alcune covariate. Questo processo nelle indagini Internet basate su lista di indirizzi di posta elettronica e nelle indagini Web ha tre

caratteristiche difficili da trattare con i metodi statistici convenzionali: l'importanza del momento della risposta, la censura e la presenza di variabili dipendenti dal tempo.

L'analisi del comportamento dei rispondenti potrebbe essere effettuata tramite una regressione logistica con variabile dipendente dicotomica: risposta o non risposta dopo uno stimolo; comunque questo tipo di analisi ignorerebbe il tempo di risposta. E' naturale ipotizzare che i soggetti che rispondono subito abbiano, in media, una propensione più elevata a completare l'auto-intervista di coloro che non partecipano fino alla terza o quarta settimana del periodo di indagine dopo essere stati sollecitati più volte.

Una soluzione al problema potrebbe essere quella di considerare il tempo prima dell'auto-intervista come variabile dipendente in un convenzionale modello di regressione. In tal caso, comunque il problema dei casi censurati, vale a dire dei casi il cui tempo di risposta supererebbe il limite stabilito per la rilevazione, rimarrebbe aperto. Il problema è complesso. Una prima ipotesi semplificatrice potrebbe essere che i casi censurati siano i membri della popolazione obiettivo che non reagiscono prima della fine del periodo d'indagine. Per adattare un modello convenzionale di regressione ai dati potremmo scartare questi dati. Purtroppo, però, nelle indagini Web i tassi di non risposta sono rilevanti⁶.

Con i metodi convenzionali non è chiaro come debbano essere inserite nel modello le variabili che dipendono dal tempo e questo vale sia per il modello di regressione logistica per il verificarsi della risposta sia per il modello di regressione lineare per il tempo prima della risposta. Tali covariate sono di solito le caratteristiche individuali del rispondente (nel caso d'indagini su una popolazione di aziende pensiamo alla forma giuridica, alla dimensione dell'azienda, al settore di attività economica) ed anche variabili che riflettono le decisioni prese nel pianificare i contatti (nel caso di indagini basate su liste di indirizzi di posta elettronica: numero e momento dei messaggi di sollecito). Alcune di queste covariate rimangono costanti durante il periodo di rilevazione, altre potrebbero cambiare in qualsiasi momento durante il follow up del caso.

3. Un modello a rischi concorrenti per l'effetto del piano dei solleciti.

Generalmente tutte le auto-interviste sono trattate in modo identico, senza distinguere se sono avvenute dopo nessun sollecito o dopo diversi solleciti. Per valutare l'effetto del piano di solleciti è essenziale distinguere le auto-interviste avvenute dopo un solo messaggio di sollecito da quelle ottenute dopo due o tre solleciti. Il rischio concorrente di reagire dopo il primo o gli altri solleciti può essere gestito sotto l'ipotesi che l'occorrenza di un tipo di evento rimuove il soggetto dal rischio di sperimentare tutti gli altri tipi di evento. I soggetti che completano l'auto-intervista dopo il primo messaggio

⁶ Una classificazione dei possibili esiti dell'invito a partecipare ad un'indagine via Internet si può trovare in Biffignandi, Pratesi (2000a).

di sollecito non sono più a rischio di completare l'auto-intervista dopo il secondo o il terzo (ed ultimo) messaggio. In altre parole, chi completa dopo il primo messaggio di sollecito al tempo t_1 avrebbe potuto completare più tardi al tempo t_2 dopo il secondo messaggio se la compilazione non fosse avvenuta in t_1 .

Indichiamo con T_i una variabile casuale che indica il tempo di auto-intervista per il soggetto i e con J_i quella che indica il tipo di auto-intervista (tipo 1: auto-intervista dopo un sollecito, tipo 2: auto-intervista dopo 2 solleciti, tipo 3: auto-intervista dopo 3 solleciti). Così $J_i = 2$ significa che il soggetto completa dopo il secondo messaggio di sollecito. La funzione di rischio (hazard) specifica del tipo di auto-intervista j al tempo t per l'azienda i , è definita da

$$h_{ij}(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t < T_i < t + \Delta t, J_i = j | T_i = t)}{\Delta t} \text{ per } j = 1, 2, 3. \quad (1)$$

La dipendenza del rischio da eventuali variabili esplicative può essere modellata tramite un modello a rischi proporzionali (general proportional hazard model di Cox) per ciascuno dei tre tipi di auto-intervista

$$\log(h_{ij}(t)) = \alpha_j(t) + \beta_j x_i(t) \text{ per } j = 1, 2, 3 \quad (2)$$

dove $x_i(t)$ è un vettore di variabili esplicative, alcune delle quali possono variare nel tempo, ed il vettore dei coefficienti β è indicizzato per segnalare che l'effetto delle covariate può essere diverso per i diversi tipi di auto-intervista. Anche la dipendenza della funzione di rischio dal tempo $\alpha(t)$ può variare tra i tipi di auto-intervista.

Nel modello precedente gli eventi che non sono quello di immediato interesse sono trattati come una forma di censura. Il meccanismo di censura non dovrebbe essere informativo pena la distorsione delle stime dei parametri. In altre parole, per l'adattamento del modello dobbiamo assumere che, condizionatamente alle covariate, quelle aziende che hanno un rischio particolarmente alto (o basso) di vivere un tipo di evento (per esempio completare il questionario dopo il primo sollecito) sperimentano una ridotta (o alta) probabilità di vivere gli altri tipi di evento.

4. Il caso di studio

4.1 Il processo di raccolta dei dati

Il modello proposto è stato applicato ai risultati di un'indagine via Internet basata su una lista di indirizzi di posta elettronica, effettuata presso circa 2500 imprese lombarde, che operano nel settore manifatturiero e delle costruzioni, situate nelle provincie di Bergamo, (usata per l'indagine pilota) Brescia, Lecco, Cremona, Mantova e Varese. Si

tratta dell'indagine su "Comunicazione tecnologica e relazioni tra imprese", le cui caratteristiche, metodologie di rilevazione e risultati sono descritte in Biffignandi, Pratesi (2000a). Qui di seguito richiamiamo brevemente alcuni aspetti essenziali al fine della valutazione della portata pratica e teorica delle analisi effettuate:

- a) l'indagine ha ottenuto un elevato grado di partecipazione. Dopo il buon esito dell'indagine pilota sulle circa 500 aziende della provincia di Bergamo, l'indagine Internet è stata inizialmente avviata su una lista di indirizzi di 2047 aziende nelle restanti 5 provincie. Il numero di aziende si è subito ridotto a 1790 per la verificata impossibilità di contattare alcune aziende a causa della qualità della lista (il 12% degli indirizzi di posta elettronica è risultato errato). Durante le operazioni di indagine sulle 1790 imprese, si è avuto successo nel contattare le aziende in 559 casi. Per quello che riguarda i rimanenti 884 casi non si è raggiunta nessuna sicura prova dell'avvenuto contatto durante il periodo di rilevazione⁷. Tracce dell'avvenuto contatto con l'azienda possono essere desunte dall'analisi dei log file degli accessi al questionario da parte di aziende che non hanno poi concluso l'auto-intervista. La mancanza di visite al questionario è stata considerata come assenza di reazione all'invito, vale a dire come segnale di mancanza di contatto con l'azienda. Sono state ottenute 442 interviste complete (IC) e 117 rifiuti espliciti (RE). Calcolando i tassi di contatto e risposta secondo i criteri proposti in Biffignandi and Pratesi (2000a, 2000b) i risultati sono tasso lordo di contatto = 54,7% (1120/2047); tasso netto di contatto = 56,8% (1120/1790); tasso di risposta dato il contatto 39,5%(442/1120), tasso generale di risposta = 21,9% (442/2047)⁸;
- b) la rilevazione è stata effettuata tramite un questionario semplice, con prevalenza di modalità di risposta qualitative e la richiesta di fornire alcune percentuali⁹. Anche la procedura di accesso al questionario era semplice. Ciò ci consente di poter analizzare i risultati, sotto ipotesi di assenza di distorsioni indotte dalla difficoltà del modello di rilevazione adottato;
- c) il processo di raccolta dei dati è stato articolato su fasi ben strutturate¹⁰, la cui qualità di implementazione e monitoraggio è stata ampiamente curata e controllata

⁷ Per esempio, alcune aziende potrebbero non essere state raggiunte poiché non consultano frequentemente la posta elettronica, altre perché hanno cessato nel frattempo la loro attività.

⁸ Il numero dei contatti è 1120: esso è ottenuto sommando le interviste completate (442), i rifiuti espliciti (117) e le visite al questionario (561). Nessuna visita al questionario è stato considerato come un indicatore grezzo di mancata reazione all'invito: questo è avvenuto in 323 casi.

⁹ Le aziende sono state intervistate sull'uso del commercio elettronico, sulla collaborazione con altre imprese e/o sulla loro appartenenza a gruppi, sui mercati e sull'occupazione. In tutto il questionario risultava composto da 8 pagine (6 pagine per la rilevazione di 40 variabili, 1 pagina di benvenuto e 1 pagina finale).

¹⁰ Il primo invito a partecipare all'indagine è stato inviato a ciascuna azienda della lista all'inizio del periodo di rilevazione (lettera di presentazione per l'indagine). Per promuovere la partecipazione sono stati offerti incentivi (rapporto d'indagine ed elaborazioni personalizzate dei risultati) il cui ruolo al fine di ridurre il tasso di non risposta è stato argomento di molti contributi in letteratura (Batagelj et al., 1998). Ciascuna azienda ha ricevuto il suo indirizzo per la compilazione; non erano richiesti né codice identificativo, né password. Ciascun accesso al questionario da parte delle aziende invitate è stato registrato, così come l'ora ed il giorno in cui questo avveniva. Durante il periodo di rilevazione sono stati inviati tre messaggi di sollecito. Tali messaggi sono stati inviati settimanalmente. Il primo dopo 14 giorni di follow up. Dopo la fine del periodo di rilevazione (3 mesi dopo l'invio del messaggio di invito) è stato inviato un quarto messaggio di sollecito. Questo ultimo messaggio suggeriva all'azienda di compilare il questionario oppure di rifiutare esplicitamente di partecipare. L'obiettivo principale di quest'ultimo messaggio era capire quali fossero le

Estratto da : S. Biffignandi, G. Gozzi (a cura di), Qualita' e Informazione Statistico Economica Territoriale: Aspetti del Processo di Formazione dei Dati e delle Metodologie di Analisi, F. Angeli. Milano, 2003.

(come descritto nel già citato lavoro Biffignandi, Pratesi 2000a). In particolare, l'attività di sollecito è stata ampiamente descritta e monitorata. Sono stati inviati tre messaggi di sollecito; dopo la fine del periodo di rilevazione (durato 3 mesi) si è inviato un quarto messaggio di sollecito che chiedeva all'azienda di compilare il questionario oppure rifiutare esplicitamente.

4.2 La funzione di sopravvivenza

Nell'applicazione le possibilità di uscita dallo studio per auto-intervista sono state classificate in tre categorie:

1. auto-intervista dopo il 1^o sollecito, inviato dopo 14 giorni dal messaggio di invito (tipo 1);
2. auto-intervista dopo il 2^o sollecito, inviato dopo 21 giorni dal messaggio di invito (tipo 2);
3. auto-intervista dopo il 3^o sollecito, inviato dopo 28 giorni dal messaggio di invito (tipo 3).

Le funzioni di sopravvivenza di ciascun gruppo sono stimate con il metodo delle tavole di sopravvivenza¹¹. Ciascuna funzione fornisce la probabilità di permanenza nell'indagine tra un sollecito e il successivo (sopravvivenza durante il periodo di rilevazione), in altre parole la probabilità di non completare l'auto-intervista, oltre il tempo t . L'origine del tempo (punto 0) è la data del primo messaggio di sollecito. Le date dei solleciti sono usate come *proxy* delle date di contatto. Il processo di auto-intervista per i tre tipi di evento è illustrato nella Figura 1¹².

caratteristiche delle imprese che non avevano reagito in alcun modo all'invito (vale a dire nessun accesso al questionario, nessun rifiuto esplicito, nessuna evidenza di errore nell'indirizzo di posta elettronica).

¹¹ Tutte le elaborazioni per la stima delle funzioni di sopravvivenza e per l'adattamento dei modelli a rischi concorrenti sono state effettuate usando le procedure per l'analisi di sopravvivenza implementate nel sistema SAS (Allison 1998).

¹² Se tutti i solleciti fossero stati inviati alla data fissata le funzioni di sopravvivenza per il secondo e terzo tipo di auto-intervista sarebbero partite da 1. Così non è poiché in 7 casi il secondo sollecito ed in 37 casi il terzo sollecito sono stati inviati prima della data stabilita a causa di un fraintendimento durante la pianificazione dell'indagine.

Estratto da : S. Biffignandi, G. Gozzi (a cura di), Qualita' e Informazione Statistico Economica Territoriale: Aspetti del Processo di Formazione dei Dati e delle Metodologie di Analisi, F. Angeli. Milano, 2003.

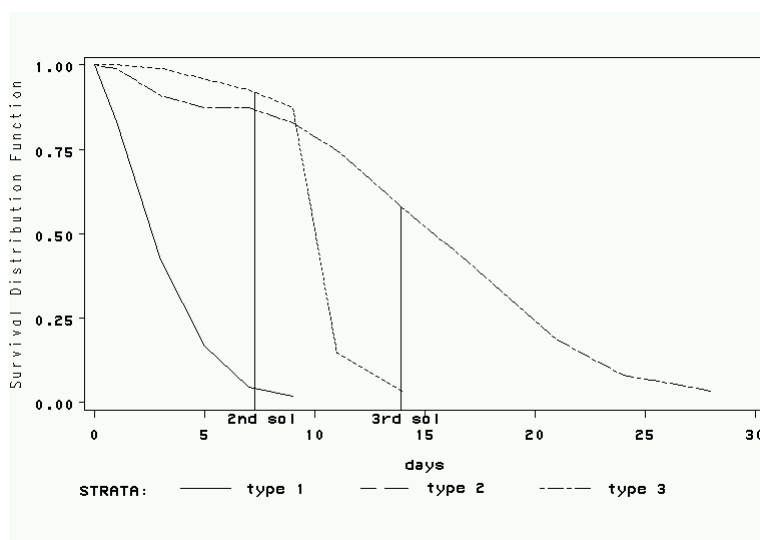


Fig. 1 - Funzioni di sopravvivenza per i tre tipi di auto-intervista

Le imprese hanno una reazione pronta al primo e al secondo messaggio di sollecito: la probabilità di sopravvivenza (permanenza nell'indagine) decresce repentinamente dopo due giorni dal primo e dal secondo sollecito. Le imprese sollecitate tre volte rimangono in attesa di uscire dalla rilevazione più a lungo delle altre e sembrano essere meno reattive allo stimolo del sollecito.

L'ipotesi che le tre funzioni di sopravvivenza siano uguali è rifiutata¹³, in altre parole i tre gruppi hanno diversi comportamenti di sopravvivenza all'interno del periodo di rilevazione. Mentre le risposte dopo 1 o 2 solleciti possono avere determinanti comuni, le risposte dopo 3 solleciti possono essere considerate un fenomeno distinto.

Esaminiamo ora i dati raccolti separatamente per provincia di localizzazione dell'impresa. Considerata la dimensione del campione e la struttura industriale nelle province, nonché le specificità dei sistemi industriali provinciali, si è proceduto ad elaborare singolarmente i dati raccolti a Bergamo, Brescia e Varese, mentre le province di Mantova, Cremona e Lecco sono considerate tutte insieme. I risultati già visti per l'intera popolazione studiata vengono confermati. Emergono, tuttavia, alcune differenziazioni a livello locale.

Nella provincia di Bergamo, vedi Figura 2, le imprese hanno una reazione pronta al primo messaggio di sollecito: la probabilità di sopravvivenza (permanenza nell'indagine) decresce repentinamente dopo due giorni dal primo sollecito. Le imprese sollecitate tre volte hanno una curva di sopravvivenza più bassa di quelle sollecitate per due volte: dai risultati pare, pertanto, che l'effetto del terzo sollecito sia più marcato di quello del secondo. Le imprese sollecitate due volte rimangono in attesa di uscire dalla

¹³ I test di eguaglianza delle funzioni di sopravvivenza nei tre gruppi sono significativi al livello $p < 0.0001$ (Log Rank: Chi quadro 323.0337 gf 2; Wilcoxon: Chi quadro 254.5206 gf 2).

Estratto da : S. Biffignandi, G. Gozzi (a cura di), Qualita' e Informazione Statistico Economica Territoriale: Aspetti del Processo di Formazione dei Dati e delle Metodologie di Analisi, F. Angeli. Milano, 2003.

rilevazione più a lungo delle altre e sembrano essere meno reattive allo stimolo del sollecito.

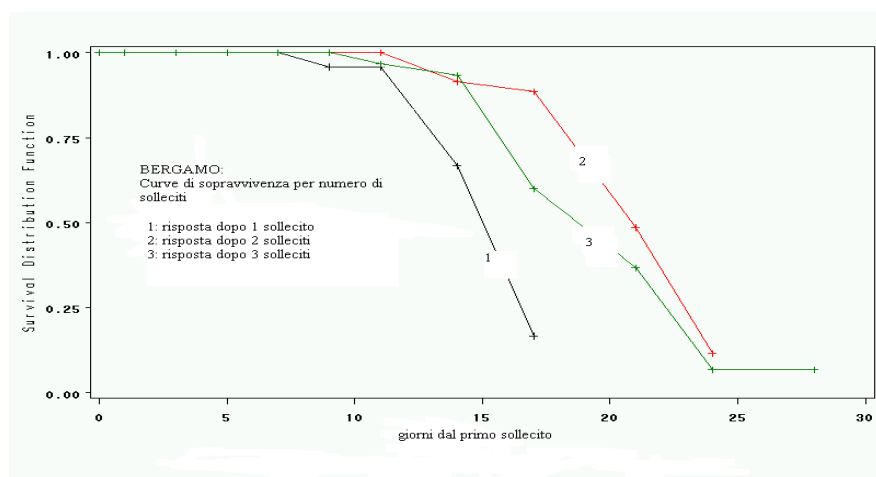


Fig. 2- Provincia di Bergamo: curve di sopravvivenza

L'evidenza cambia nelle altre province. Brescia e Varese (vedi Figura 3) si comportano come Bergamo per quello che riguarda l'effetto del primo sollecito; ribaltano, invece, l'importanza del secondo e del terzo. L'effetto del secondo sollecito è del tutto simile a quello del primo: la probabilità di sopravvivenza (permanenza nell'indagine) decresce repentinamente dopo due giorni dal primo sollecito e anche dal secondo. La curva di sopravvivenza è meno ripida per coloro che sono sollecitati 3 volte. Essi tendono a rimanere tra i non rispondenti più a lungo dei sollecitati 1 o 2 volte.

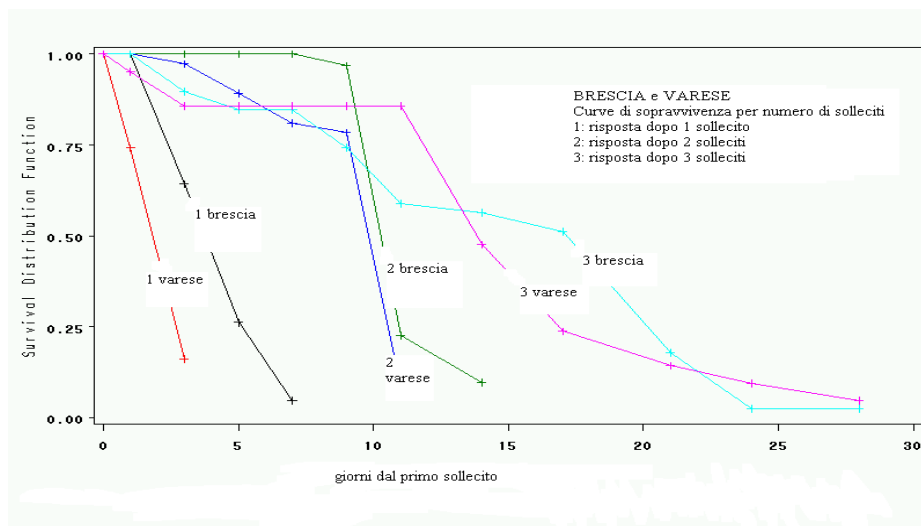


Fig. 3- Province di Brescia e provincia di Varese: curve di sopravvivenza

Tale comportamento è confermato nel gruppo che aggrega i risultati delle province di Cremona, Mantova e Lecco, considerate tutte insieme (vedi Figura 4).

In tutte le province, l'ipotesi che le tre funzioni di sopravvivenza siano uguali è rifiutata¹⁴, in altre parole i tre gruppi hanno diversi comportamenti di sopravvivenza all'interno del periodo di rilevazione anche leggendo i risultati per provincia.

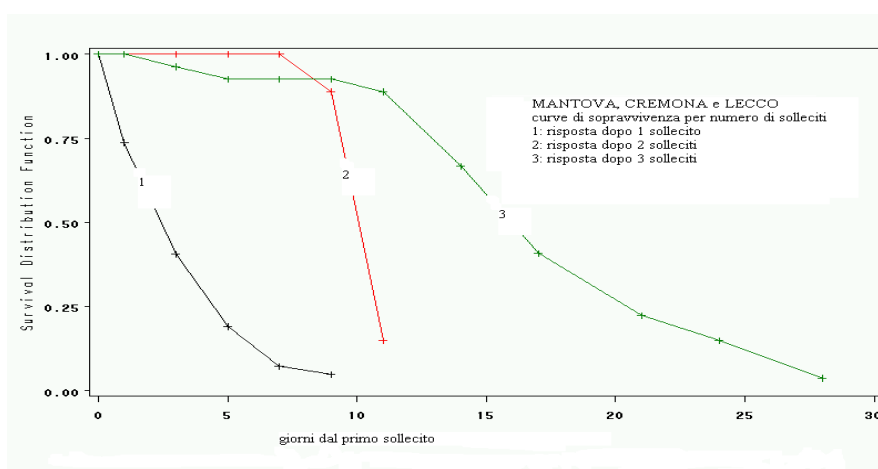


Fig. 4 - Gruppo comprendente le province di Cremona, Mantova e Lecco: curve di sopravvivenza

4.3 Le funzioni di rischio come funzioni del tempo

Tornando a considerare l'intero insieme dei dati raccolti osserviamo che le funzioni di rischio viste in funzione del tempo non sono le stesse per tutti i tipi di evento. La Figura 5 mostra il comportamento di ciascuna funzione di sopravvivenza dopo una doppia trasformazione logaritmica $\log(-\log(S_{ij}(t)))$. Il tempo è misurato in ore dalla spedizione (ricevimento) del messaggio di invito a partecipare all'indagine. Se le funzioni di rischio fossero proporzionali, le funzioni di sopravvivenza log-log dovrebbero essere parallele¹⁵.

Le curve di tipo 1 e di tipo 2 sono virtualmente simili. La curva per le uscite di tipo 3 è più bassa delle altre due curve e si muove accanto alle altre due nell'ultima parte del periodo di rilevazione, testimoniando la presenza di non proporzionalità. L'ipotesi di proporzionalità dei rischi può essere verificata adattando un modello logit multinomiale alla relazione tra il tipo di hazard ed il tempo. I risultati della Tabella 1 mostrano che l'effetto del tempo è altamente significativo, indicando di rifiutare l'ipotesi di proporzionalità.

¹⁴ I test di eguaglianza delle funzioni di sopravvivenza nei tre gruppi sono significativi al livello $p < 0.0001$ in ciascuna delle province o gruppi di province considerati.

¹⁵ Lo *shift* temporale tra le tre curve è da attribuire al fatto che i messaggi di sollecito non sono stati spediti alla stessa ora nelle date fissate.

Estratto da : S. Biffignandi, G. Gozzi (a cura di), Qualita' e Informazione Statistico Economica Territoriale: Aspetti del Processo di Formazione dei Dati e delle Metodologie di Analisi, F. Angeli. Milano, 2003.

In Tabella 2, inoltre, si possono esaminare i valori assunti dai parametri del modello multinomiale. Si nota come il parametro 3, coefficiente del confronto tra eventi di tipo 1 (risposte dopo un sollecito) e di tipo 3 (risposte dopo 3 solleciti): il rischio di eventi di tipo 1 decresce molto più rapidamente del rischio di eventi di tipo 3. La stessa conclusione si raggiunge dall'esame del parametro 4 per il confronto tra eventi di tipo 2 e di tipo 3. I rischi di risposta dopo 1 o 2 solleciti decrescono molto più rapidamente con il tempo dei rischi di risposta dopo 3 solleciti. Si può concludere che il rischio di rispondere dopo 3 solleciti è più basso degli altri due (anche se il comportamento non è omogeneo in tutto il periodo, vedi Figura 1), e che esso decresce meno rapidamente nel tempo.

Tab. 1 - Modello multinomiale: analisi della varianza

Fonte	GF	Chi-Quadro	Pr> ChiQuadro
Intercetta	2	110.72	<.0001
Tempo	2	104.75	<.0001
Likelihood Ratio	610	386.65	1.0000

Tab.2 - Modello multinomiale: analisi delle stime

Effetto	Parametro	Stima	Errore Standard	Chi-Quadro	Pr>ChiQuadro
Intercetta	1	17.0835	1.6345	109.25	<.0001
	2	5.5694	1.0531	27.97	<.0001
Tempo	3	-0.0314	0.00307	104.75	<.0001
	4	-0.00860	0.00165	27.14	<.0001

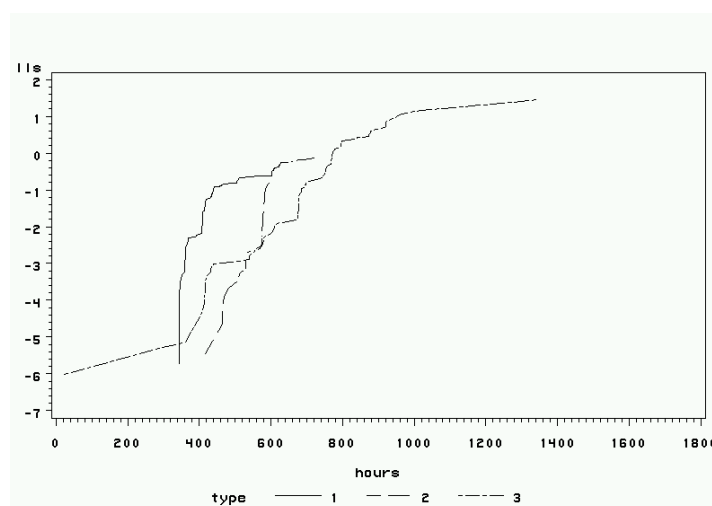


Fig. 5 - Funzioni di sopravvivenza Log-log per i tre tipi di auto-intervista

4.4 Effetto delle covariate sulle funzioni di rischio

L'effetto delle covariate su ciascun tipo di hazard può essere verificato adattando un modello di Cox (equazione (2)) a ciascun tipo di auto-intervista. L'obiettivo è verificare se l'effetto delle covariate è lo stesso sui diversi tipi di uscita per auto-intervista (risposta)¹⁶.

Il modello include le seguenti variabili:

- TEMPO [ore]: tempo trascorso prima dell'auto-intervista, misurato in ore, minuti e secondi dall'invio (ricevimento) dell'invito a partecipare all'indagine;
- SOLLECITI [s]: s=0, nessun messaggio di sollecito, s=1, auto-intervista dopo il 1^o messaggio di sollecito, s=2, auto-intervista dopo il 2^o messaggio di sollecito, auto-intervista dopo il 3^o messaggio di sollecito;
- ETÀ [eta]: età dell'impresa in anni al momento dell'indagine.
- DIMENSIONE DELL'IMPRESA [small]: dimensione dell'impresa in termini di addetti: small=1, numero di addetto superiore o uguale a 20, small=0 altrimenti;
- FORMA GIURIDICA DELL'IMPRESA [di]: di= 1 se l'impresa è una ditta individuale, di=0 altrimenti;
- FREQUENZA DI CONNESSIONE AD INTERNET [fuicf]: frequenza di uso di Internet per comunicazioni con clienti e fornitori, fuicf=1, connessione quotidiana ad Internet, fuicf=0 altrimenti;
- PRATICA DEL COMMERCIO ELETTRONICO [q6ind]: q6ind=1 se l'impresa pratica commercio elettronico, q6ind=0 altrimenti.

La classificazione delle imprese rispondenti per numero di solleciti e variabili esplicative introdotte nel modello dà origine alle seguenti distribuzioni di frequenza:

Variabili indipendenti e numero di solleciti

Solleciti	Q6ind		Fuicf		small	
	0	1	0	1	0	1
0 (120)	36	84	53	67	59	61
1 (124)	37	87	46	78	74	50
2 (95)	37	58	41	54	61	34
3 (88)	38	50	34	54	43	45
Totale (n=427) ¹⁷	148	279	174	253	237	190

¹⁶ L'adattamento dei modelli è stato effettuato testando l'effetto delle covariate sulla verosimiglianza con il test del rapporto di verosimiglianza (forward selection procedure). I vari modelli sono stati confrontati tramite test standard (test del rapporto di verosimiglianza per modelli annidati e Akaike Information Criterion per i modelli non annidati).

¹⁷ I modelli sono stati stimati sui 427 rispondenti per i quali nessuna delle variabili indipendenti presentava valori mancanti.

Estratto da : S. Biffignandi, G. Gozzi (a cura di), Qualita' e Informazione Statistico Economica Territoriale: Aspetti del Processo di Formazione dei Dati e delle Metodologie di Analisi, F. Angeli. Milano, 2003.

Variabili indipendenti e numero di solleciti

Solleciti	eta			di	
	1-10	11-20	21+	0	1
0 (120)	29	33	58	104	16
1 (124)	28	39	57	110	14
2 (95)	19	30	46	88	7
3 (88)	23	28	37	80	8
Totale (n=427)	99	130	198	382	45

Sono stati adattati i seguenti modelli:

- modello base: il modello tratta tutti i tipi di evento nello stesso modo. Tale modello è incluso per motivi di confronto, qui la censura è sulle imprese che hanno completato l'auto-intervista dopo almeno un sollecito;
- modello 1: questo è il modello per le uscite di tipo 1 (auto-interviste dopo il 1° sollecito), le uscite di tipo 2 e di tipo 3 sono trattate come dati censurati;
- modello 2: questo è il modello per le uscite di tipo 2 (auto-interviste dopo il 2° sollecito), le uscite di tipo 1 e di tipo 3 sono trattate come dati censurati;
- modello 3: questo è il modello per le uscite di tipo 3 (auto-interviste dopo il 3° sollecito), le uscite di tipo 2 e di tipo 1 sono trattate come dati censurati.

La Tabella 3 mostra i risultati ottenuti quando non si distingue tra i diversi tipi di uscite per l'intero insieme dei dati raccolti (modello base).

L'effetto più forte è quello in corrispondenza alla frequenza di uso di Internet (fuicf): il sollecito sembra essere più efficace per le aziende che controllano frequentemente le loro caselle di posta elettronica. Gli utenti assidui di Internet rischiano di uscire per auto-intervista 1.2 volte di più degli altri per effetto di almeno un messaggio di sollecito.

Tab. 3- Analisi delle uscite dopo almeno un sollecito

Testing Global Null Hypothesis: BETA=0						
Test	Chi-Square	DF	Pr > ChiSq			
Likelihood Ratio	7.6784	3	0.0531			
Score	7.5728	3	0.0557			
Wald	7.5376	3	0.0566			
Analysis of Maximum Likelihood Estimates						
Variable	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio
SMALL	1	-0.20013	0.12179	2.7001	0.1003	0.819
fuicf	1	0.22183	0.12493	3.1527	0.0758	1.248

Estratto da : S. Biffignandi, G. Gozzi (a cura di), Qualita' e Informazione Statistico Economica Territoriale: Aspetti del Processo di Formazione dei Dati e delle Metodologie di Analisi, F. Angeli. Milano, 2003.

q6ind	1	-0.14167	0.12290	1.3286	0.2490	0.868
-------	---	----------	---------	--------	--------	-------

L'uscita dopo 2 solleciti mostra un pattern diverso (modello 2 - Tabella 4). La frequenza di uso di Internet per le comunicazioni con clienti e fornitori (fuicf) non sembra importante, e nemmeno la pratica del commercio elettronico (q6ind) ha un effetto di rilievo. Notiamo un effetto della dimensione dell'azienda (small): aziende di piccola dimensione subiscono una diminuzione di .52 nel rischio di uscire dopo 2 solleciti.

Per le uscite dopo 3 solleciti (modello 3 - Tabella 5) la pratica del commercio elettronico (q6ind) è l'unica variabile significativa: le imprese che praticano commercio elettronico hanno un rischio più basso del 66% di uscire per auto-intervista dopo 3 solleciti. Il quadro cambia quando si mette a fuoco il rischio di uscire per auto-intervista dopo un solo sollecito: il modello precedente non risulta più significativo (modello 1 - Tabella 6).

Tab. 4 - Modello 2: analisi delle uscite dopo 2 solleciti

Testing Global Null Hypothesis: BETA=0						
Test		Chi-Square	DF	Pr > ChiSq		
Likelihood Ratio		7.5699	3	0.0558		
Score		7.4423	3	0.0591		
Wald		7.2684	3	0.0638		
Analysis of Maximum Likelihood Estimates						
Variable	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio
SMALL	1	-0.54451	0.21893	6.1857	0.0129	0.580
fuicf	1	0.08926	0.21794	0.1677	0.6821	1.093
q6ind	1	-0.16730	0.21553	0.6025	0.4376	0.846

Tab. 5 - Modello 3: analisi delle uscite dopo 3 solleciti

Testing Global Null Hypothesis: BETA=0						
Test		Chi-Square	DF	Pr > ChiSq		
Likelihood Ratio		14.6499	3	0.0021		
Score		15.6206	3	0.0014		
Wald		14.9201	3	0.0019		
Analysis of Maximum Likelihood Estimates						
Variable	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio
SMALL	1	0.11694	0.23369	0.2504	0.6168	1.124
fuicf	1	0.39719	0.25775	2.3746	0.1233	1.488
q6ind	1	-0.89506	0.23952	13.9645	0.0002	0.409

Estratto da : S. Biffignandi, G. Gozzi (a cura di), Qualita' e Informazione Statistico Economica Territoriale: Aspetti del Processo di Formazione dei Dati e delle Metodologie di Analisi, F. Angeli. Milano, 2003.

Tab. 6 - Modello 1: analisi delle uscite dopo 1 sollecito

Testing Global Null Hypothesis: BETA=0						
Test		Chi-Square	DF	Pr >	ChiSq	
Likelihood Ratio		5.8430	3	0.1195		
Score		5.7221	3	0.1259		
Wald		5.6728	3	0.1287		

Analysis of Maximum Likelihood Estimates						
Variable	DF	Parameter Estimate	Standard Error	Chi-Square	Pr >	Hazard Ratio
SMALL	1	-0.10567	0.18992	0.3096	0.5779	0.900
fuicf	1	0.27305	0.19134	2.0363	0.1536	1.314
q6ind	1	0.33820	0.19838	2.9064	0.0882	1.402

L'analisi di sopravvivenza condotta con i modelli a rischi concorrenti è stata effettuata anche separatamente per province. In particolare, si è operato come in precedenza: le province di Bergamo, Varese, Brescia sono state trattate ciascuna individualmente, mentre le province di Mantova, Cremona e Lecco sono state considerate come un unico strato, omogeneo per le caratteristiche dimensionali e di struttura delle imprese presenti sul territorio. In questo lavoro sono riportati solo i modelli risultati significativi in ogni provincia o gruppo di province. Per la provincia di Bergamo non si sono avuti risultati significativi. In tutte le altre province l'attenzione si è focalizzata sul Modello 3 che analizza il rischio concorrente di uscire dopo il numero massimo di 3 solleciti consentito nell'indagine.

Per ciascuna provincia o gruppo di province sono state inserite nel modello le seguenti variabili e tutte le loro interazioni.

- ETA: indicatore dell'età dell'impresa quale compare dal Registro Ditte: 1 se età maggiore dell'età media nella provincia, 0 altrimenti.
- DI: indicatore di forma giuridica: 1 se ditta individuale, 0 altrimenti.
- SMALL: indicatore di classe dimensionale: 1 se numero di addetto superiore o uguale a 20, small=0 altrimenti.
- FUICF: Frequenza di connessione ad Internet per comunicazioni con clienti e fornitori, fuicf=1, connessione quotidiana ad Internet, fuicf=0 altrimenti.
- Q6IND: Pratica del commercio elettronico: 1 se l'impresa pratica commercio elettronico, 0 altrimenti.
- FUIEI: Frequenza di connessione ad Internet per documentazione, 1, connessione quotidiana ad Internet, fuicf=0 altrimenti.
- FUICO: Frequenza di connessione ad Internet per motivi commerciali, 1, connessione quotidiana ad Internet, 0 altrimenti.
- SERA: indicatore del momento della auto-intervista: 1 dopo le 19, 0 prima delle 19.

Estratto da : S. Biffignandi, G. Gozzi (a cura di), Qualità e Informazione Statistico Economica Territoriale: Aspetti del Processo di Formazione dei Dati e delle Metodologie di Analisi, F. Angeli. Milano, 2003.

- WEB: indicatore della presenza di pagina Web dell'impresa: 1se sì, 0 altrimenti.

Nelle tabelle dei risultati sono riportate solo le interazioni e le variabili significative. I risultati dell'adattamento del modello base alle imprese attive della provincia di Brescia sono riportati nella Tabella 7. Il rischio di uscire dopo almeno un sollecito appare legato significativamente solo a due variabili: il momento del collegamento (SERA=1) e l'indicatore congiunto della forma giuridica e dell'età dell'impresa (ETABSDI=1). Il rischio concorrente di aver bisogno di almeno un sollecito per partecipare all'indagine è più alto per le imprese con forma giuridica di ditta individuale e con età superiore all'età media delle imprese della provincia (l'età è calcolata in anni dalla data di inizio attività certificata nel Registro Ditte CCIAA: età media=17.6, s.q.m.=8.98, in tutto 170 imprese attive dal Registro). Dai dati raccolti sembra che chi si collega di sera (dopo le 19) abbia un bisogno minore di solleciti (hazard ratio <1): forse il risultato segnala che compilazioni fuori dall'orario di lavoro e che quindi non intralciano la normale attività dell'impresa sono preferite dagli imprenditori.

Tab. 7 - Brescia: analisi delle uscite dopo almeno 1 sollecito

Testing Global Null Hypothesis: BETA=0						
Test		Chi-Square	DF	Pr >	ChiSq	
Likelihood Ratio		17.2971	9		0.0443	
Score		19.7659	9		0.0194	
Wald		17.6508	9		0.0394	

Analysis of Maximum Likelihood Estimates						
Variable	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio
etabs	1	-0.36055	0.42420	0.7224	0.3953	0.697
DI	1	-0.32750	0.54326	0.3634	0.5466	0.721
SMALL	1	-0.19383	0.22542	0.7394	0.3899	0.824
FUICF	1	0.14780	0.30107	0.2410	0.6235	1.159
Q6IND	1	-0.21067	0.30940	0.4636	0.4959	0.810
sera	1	-1.51231	0.73586	4.2237	0.0399	0.220
etabsdi	1	2.19167	0.85199	6.6173	0.0101	8.950
etabsfui	1	0.55837	0.41250	1.8322	0.1759	1.748
etabsq6	1	0.10360	0.42301	0.0600	0.8065	1.109

I risultati dell'analisi delle uscite di tipo 3 (auto-intervista dopo 3 solleciti) sono contenuti in Tabella 8. Abbiamo già visto che il rischio di essere sollecitate almeno una volta è maggiore per le ditte individuali attive da più di 17 anni (cfr. Tabella 7): i risultati dell'adattamento del Modello 3 confermano il maggior bisogno di richiamare l'attenzione di queste aziende alla partecipazione all'indagine (hazard ratio >1 per la variabile che indica ditte individuali "anziane"). La pratica del commercio elettronico pare rendere meno urgente il bisogno di molti solleciti e curiosamente le aziende attive da più di 17 anni anche se si dichiarano utenti Internet corrono un rischio di 1.7 volte più elevato delle altre di rispondere dopo 3 solleciti.

Tab. 8 - Brescia: analisi delle uscite dopo 3 solleciti

Testing Global Null Hypothesis: BETA=0

Estratto da : S. Biffignandi, G. Gozzi (a cura di), Qualita' e Informazione Statistico Economica Territoriale: Aspetti del Processo di Formazione dei Dati e delle Metodologie di Analisi, F. Angeli. Milano, 2003.

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	14.5915	7	0.0416
Score	17.5516	7	0.0142
Wald	14.8359	7	0.0382

Analysis of Maximum Likelihood Estimates						
Variable	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio
etabs	1	-1.19360	0.65406	3.3302	0.0680	0.303
DI	1	-0.48309	0.72063	0.4494	0.5026	0.617
SMALL	1	0.14621	0.38042	0.1477	0.7007	1.157
Q6IND	1	-0.89259	0.50901	3.0751	0.0795	0.410
etabsdi	1	3.84289	1.44383	7.0841	0.0078	46.660
etabsfui	1	1.26571	0.59969	4.4547	0.0348	3.546
etabsq6	1	-0.20521	0.74038	0.0768	0.7816	0.814

Per la provincia di Varese i risultati dell'analisi delle uscite di tipo 3 (auto-intervista dopo 3 solleciti) sono contenuti in Tabella 9. La pratica generale del commercio elettronico pare rendere meno urgente il bisogno di molti solleciti: dai risultati appare che essa riduce il rischio di rispondere dopo 3 solleciti più del 50% in confronto alle altre aziende. Anche l'uso di Internet per ricerche e documentazioni (FUIEI) abbassa il rischio di rispondere dopo 3 solleciti. Non è significativo invece l'uso di Internet strettamente per motivi commerciali (FUICO). Invece, curiosamente, le aziende che secondo il Registro Ditte hanno una pagina Web (variabile WEB) corrono un rischio più elevato di rispondere dopo 3 solleciti. Quest'ultimo risultato, francamente contro intuitivo e ottenuto solo nella provincia di Varese, può essere dovuto al fatto che la notizia sulla presenza di una pagina Web deriva non dall'auto-intervista ma da una variabile del Registro Ditte rilevata su indicazione volontaria dell'impresa. L'attendibilità dell'indicazione e la sua rispondenza all'effettiva realtà vissuta dall'azienda sembrano quindi discutibili.

Per il gruppo di province Mantova, Cremona e Lecco i risultati dell'analisi delle uscite di tipo 3 (auto-intervista dopo 3 solleciti) sono contenuti in Tabella 10. In questo contesto le ditte individuali hanno un rischio di uscire dopo 3 solleciti ridotto quasi del 90% rispetto alle altre aziende: il messaggio di invito a partecipare all'indagine pare arrivare, nel caso delle ditte individuali, più facilmente in modo diretto a chi ha le competenze e l'autorità per rispondere subito al questionario. Anche in queste province l'uso frequente di Internet è significativamente legato al rischio di uscire dopo 3 solleciti invece che dopo 1 o 2 soltanto. La direzione dell'effetto sul rischio di uscire dopo 3 solleciti è però opposta rispetto alle altre province. L'uso di Internet per ricerche e documentazioni (FUIEI) alza di 6 volte il rischio di rispondere dopo 3 solleciti. Nello stesso senso, rischio 3 volte più elevato, è l'effetto dell'uso di Internet strettamente per motivi commerciali (FUICO). La giustificazione del risultato potrebbe essere trovata nella tentazione di ritardare la compilazione, ritenuta un compito tecnicamente facile, da parte di titolari che spesso hanno ricevuto l'invito a partecipare direttamente al proprio indirizzo personale di posta elettronica, usato spesso, nel caso di ditte individuali, anche per l'impresa.

Estratto da : S. Biffignandi, G. Gozzi (a cura di), *Qualità e Informazione Statistico Economica Territoriale: Aspetti del Processo di Formazione dei Dati e delle Metodologie di Analisi*, F. Angeli. Milano, 2003.

Tab. 9 - Varese: analisi delle uscite dopo 3 solleciti

Testing Global Null Hypothesis: BETA=0						
Test		Chi-Square	DF	Pr > ChiSq		
Likelihood Ratio		20.5377	7	0.0045		
Score		20.7174	7	0.0042		
Wald		15.2239	7	0.0332		

Analysis of Maximum Likelihood Estimates						
Variable	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio
etava	1	1.17992	0.72068	2.6806	0.1016	3.254
DI	1	-0.97983	1.06348	0.8489	0.3569	0.375
SMALL	1	1.17724	0.76718	2.3547	0.1249	3.245
Q6IND	1	-1.78196	0.65065	7.5008	0.0062	0.168
WEB	1	2.84374	1.00867	7.9484	0.0048	17.180
FUIEI	1	-2.13212	0.86760	6.0393	0.0140	0.119
FUICO	1	1.36391	0.85596	2.5390	0.1111	3.911

Tab. 10- Gruppo Mantova, Cremona, Lecco: analisi delle uscite dopo 3 solleciti

Testing Global Null Hypothesis: BETA=0						
Test		Chi-Square	DF	Pr > ChiSq		
Likelihood Ratio		22.0379	7	0.0025		
Score		22.7462	7	0.0019		
Wald		17.3476	7	0.0153		

Analysis of Maximum Likelihood Estimates						
Variable	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio
etamc	1	0.55881	0.58676	0.9070	0.3409	1.749
DI	1	-2.65793	1.40488	3.5794	0.0585	0.070
SMALL	1	-0.52415	0.55834	0.8813	0.3478	0.592
Q6IND	1	-0.66029	0.47889	1.9011	0.1680	0.517
web	1	-0.24778	0.49074	0.2549	0.6136	0.781
FUIEI	1	1.82346	0.90658	4.0456	0.0443	6.193
FUICO	1	1.27696	0.60023	4.5260	0.0334	3.586

5. Considerazioni conclusive

Nelle indagini via Internet basate su una lista di indirizzi di posta elettronica il primo contatto ed i successivi messaggi di sollecito sono cruciali nel suscitare l'attenzione dei rispondenti e promuovere la loro partecipazione all'indagine. La tempestività è dovuta in larga parte all'efficacia degli stimoli ottenuti tramite le operazioni di sollecito. L'effetto dello stimolo si esaurisce nel breve periodo (circa 2 giorni dall'invio del messaggio di sollecito). L'efficacia del sollecito si perde dopo l'invio di diversi messaggi di sollecito (nel nostro caso 3). Addirittura dopo 4 stimoli si vede che il loro effetto diviene irrilevante. Il rischio di rispondere dopo uno, due o tre solleciti dipende

Estratto da : S. Biffignandi, G. Gozzi (a cura di), *Qualità e Informazione Statistica Economica Territoriale: Aspetti del Processo di Formazione dei Dati e delle Metodologie di Analisi*, F. Angeli. Milano, 2003.

dalle caratteristiche dell'impresa. Frequenza di uso di Internet, pratica del commercio elettronico e dimensioni dell'azienda in termini di addetti sono risultati significativamente legati alla reattività ai solleciti. Le imprese di piccole dimensioni che partecipano all'indagine – quelle per le quali la catena di re-indirizzamenti del messaggio di invito è breve – mostrano una reazione immediata agli stimoli. In effetti, è ragionevole pensare che l'accesso alla rete giochi un ruolo diverso in imprese di diversa dimensione e che esso sia più o meno direttamente gestito dallo stesso personale deputato alla compilazione del questionario.

Nelle singole province sono generalmente confermati i risultati ottenuti per l'intero insieme di imprese. A Mantova, Cremona e Lecco le ditte individuali rischiano meno delle altre di rispondere dopo tre solleciti. A Brescia, invece, osserviamo una reattività peggiore nelle ditte individuali attive da un numero di anni maggiore della media. La pratica del commercio elettronico abbassa il rischio di rispondere dopo 3 solleciti sia per le imprese di Brescia sia per quelle di Varese.

I risultati ottenuti, pur nei limiti dell'applicazione effettuata, permettono di concludere che i guadagni in tempestività attesi con indagini via Internet possano e debbano essere guidati bilanciando in modo opportuno i tempi di reazione agli stimoli (solleciti) e la periodicità dei solleciti stessi.

Riferimenti bibliografici

- Allison P.D. (1998), *Survival Analysis using the SAS System. A Practical Guide*, SAS Institute".
- Batagelj Z., Lozar K., Vehovar V. (1998a), "Who are non respondents in Web Surveys?", Paper presented at the 9th International Workshop on Household Survey Non-response, Bled, September 1998.
- Batagelj Z., Lozar K., Vehovar V. (1998b), "Technical and Methodological Issues in WWW Surveys", Meeting of the American Association for Public Opinion Research, St. Louis, May.
- Biffignandi S., Pratesi M. (2000a), Le indagini sulle imprese: aspetti metodologici e un'analisi degli intervistati, in "Tecnologie informatiche e fonti amministrative nella formazione dei dati economici", C. Filippucci ed., F. Angeli, Milano.
- Biffignandi S., Pratesi M. (2000b), "Modelling firm response and contact probabilities in Web Surveys", International Conference on Establishment Surveys, June 2000, Proceedings.
- Biffignandi S., Pratesi M. (2000c), "Modeling the respondents' profile in a Web survey on firms in Italy", *Metodoloski zvezki*, n.18 Ljubljana, Slovenia.
- Bosnjak M., Tuten T.L (2001), "Classifying response behaviour in Web-based Surveys", *Journal of Computer-Mediated Communication* 6 (3).
- Cox D. R. Oakes D. (1984), *Analysis of survival data*, Chapman & Hall, London.
- Clayton R.C., Werking G.S. (2000), "Business Surveys of the Future: the World Wide Web as a Data Collection Methodology", in Couper M. et al., *Computer Survey Information Collection*, J. Wiley, New York.
- Couper M.P., Blair J., Triplett T. (1999), "A Comparison of Mail and E-mail for a Survey of Employees", *Journal of Official Statistics*, vol. 15, n. 1.
- Couper M.P. (2000), "Web Surveys. A Review and Guide", *Public Opinion Quarterly*, 64, 4.
- Dillman D.A. (1998), "Mail and Other Self-Administered Surveys in the 21st Century: the Beginning of a New Era".
- Elvers E., Rosén B., *Encyclopedia of Statistical Sciences*, Quality Concepts for Official Statistics, Wiley & Sons, 1997.
- Metha R., Sividas E. (1995), "Comparing Response Rates and Response Content in Mail Versus Electronic Mail Surveys", *Journal of Marketing Research Society*, vol. 37.
- Schaefer D., Dillman D.A. (1998), "Development of a Standard E-mail Survey Methodology: Results of an Experiment", Paper presented at the American Association for Public Opinion Research.

Estratto da : S. Biffignandi, G. Gozzi (a cura di), *Qualità e Informazione Statistica Economica Territoriale: Aspetti del Processo di Formazione dei Dati e delle Metodologie di Analisi*, F. Angeli. Milano, 2003.

Smith C. (1997), "Casting the Net: Surveying an Internet Population", *Journal of Communication Mediated by Computers*.