

Le determinanti della buona riuscita di un'indagine via web

C. Cimini^{*}, G. Gasperoni^{**}, C. Girotti^{*}

^{*}Consorzio Interuniversitario ALMALAUREA, ^{**}Università di Bologna
chiara.cimini@almalaurea.it; giancarlo.gasperoni@unibo.it;
claudia.girotti@almalaurea.it

1. Introduzione

Da oltre dieci anni il Consorzio Interuniversitario AlmaLaurea¹ svolge annualmente un'indagine volta a monitorare gli esiti occupazionali dei laureati dopo uno, tre e cinque anni dal conseguimento del titolo universitario. La rilevazione garantisce l'analisi delle più recenti tendenze del mercato del lavoro, verificando le esperienze formative e lavorative compiute dai laureati nel primo quinquennio successivo al conseguimento del titolo.

L'indagine, che coinvolgeva dapprima i soli laureati del periodo maggio-agosto, è stata ampliata ai laureati dell'intero anno solare per due ordini di motivi:

1. per fornire agli atenei strumenti per pianificare nuovi corsi avendo a disposizione una documentazione articolata fino a livello del singolo corso di laurea;
2. avendo riscontrato che, diversamente da quanto osservato nel sistema pre-riforma, i laureati post-riforma del periodo maggio-agosto non sono rappresentativi dei laureati dell'intero anno solare.

Ovviamente l'estensione della rilevazione all'intero anno solare ha comportato una forte crescita del numero dei laureati da intervistare e ha fatto emergere la necessità di ridurre i tempi e i costi delle indagini.

A partire dalla rilevazione 2008, quindi, lo schema di indagine è stato rivisto e la tradizionale metodologia CATI (*Computer-Assisted Telephone Interview*) è stata affiancata dalla più moderna tecnica di rilevazione CAWI (*Computer-Assisted Web Interview*). Ciò è stato possibile fondamentalmente grazie a due fattori: da un lato, l'ampia e crescente disponibilità di indirizzi e-mail, generalmente aggiornati perché inseriti direttamente dai laureati nel proprio curriculum vitae consultabile dalle aziende; dall'altro, la maturata esperienza nel campo delle indagini on-line (oltre 20 indagini svolte

¹ ALMALAUREA è nata nel 1994 per iniziativa dell'Osservatorio Statistico dell'Università di Bologna e si è costituita in Consorzio Interuniversitario nel 2000. Raggruppa attualmente 53 Atenei che rappresentano circa il 70% dei laureati italiani. ALMALAUREA è nata al servizio dei laureati, delle Università e delle imprese; con questo obiettivo rende disponibile una documentazione tempestiva sul capitale umano formato ai più alti livelli e realizza la banca dati on-line dei laureati, anche con pluriennale esperienza, per agevolare l'incontro tra domanda e offerta di lavoro nel mercato nazionale, nonché per favorire la mobilità transnazionale. Per approfondimenti sulla metodologia di indagine: www.almalaurea.it.

da ALMALAUREA negli ultimi 5 anni)². Nello specifico, a partire dalla rilevazione 2008, lo schema di indagine prevede che tutti i laureati di primo e secondo livello³ in possesso di indirizzo di posta elettronica siano contattati via e-mail ed invitati a compilare un questionario on-line⁴. Successivamente, tutti coloro che non rispondono al questionario via web (nonché, ovviamente, i laureati senza indirizzo di posta elettronica) vengono contattati per una intervista telefonica. La combinazione dei due strumenti di rilevazione ha permesso ad ALMALAUREA di raggiungere gli abituali tassi di risposta ottenuti nelle tradizionali indagini sulla condizione occupazionale (ad esempio per i laureati 2007 sia pre che post-riforma indagati ad un anno dal conseguimento del titolo ha superato l'85%).

L'implementazione di questa nuova metodologia ha comportato la necessità di approfondire tale tematica, cercando di individuare i fattori che determinano una buona riuscita dell'indagine web e, di conseguenza, adattare gli schemi di rilevazione futuri ai collettivi di volta in volta indagati.

2. Caratteristiche strutturali delle indagini e collettivo di riferimento

La necessità di identificare le variabili che incidono sulla buona riuscita di un'indagine CAWI ha comportato l'esigenza di concentrarsi su un collettivo omogeneo, escludendo dall'analisi alcune indagini su collettivi o temi specifici (ad es. sul dialogo interculturale) che avrebbero potuto determinare una distorsione dei risultati. Si è dunque scelto di prendere in esame le sole indagini riguardanti gli esiti formativi e occupazionali dei laureati post-riforma ad un anno dalla laurea.

Si tratta di quasi 254 mila laureati indagati in sette differenti rilevazioni (Tab.1), di cui oltre 200 mila in possesso di indirizzo e-mail valido (79,6%)⁵. Il tasso di possesso e-mail, dato dal rapporto tra il numero dei laureati in possesso di indirizzo di posta elettronica valido e il complesso dei laureati indagati, varia a seconda del collettivo esaminato: è del 74% tra i laureati di primo livello del 2004, sale all'83% tra i laureati di secondo livello del periodo luglio-dicembre del 2007. Tale differenza è probabilmente imputabile alla

² Le prime sperimentazioni sono state compiute per approfondire in tempi ridotti e a costi contenuti specifici fenomeni di interesse, comunque generalmente inerenti gli esiti occupazionali dei laureati. Ne sono un esempio l'indagine sugli esiti occupazionali dei laureati under 23 del 2003 (i primi laureati di primo livello), l'indagine sui laureati residenti a Imola o l'indagine sul dialogo interculturale che ha avuto come obiettivo quello di valutare le opportunità che i laureati hanno di entrare in contatto con persone e culture diverse da quelle italiane, sia sul luogo di lavoro, sia nella vita personale.

³ Tra i laureati di secondo livello sono compresi i laureati specialistici e i laureati specialistici a ciclo unico.

⁴ A differenza delle indagini CATI svolte in collaborazione con una società esterna, le indagini di tipo CAWI sono condotte internamente da ALMALAUREA in tutte le sue fasi: dalla costruzione del questionario alla fase di test dello stesso, dalla spedizione della e-mail di invito alla gestione dei solleciti, alla raccolta dei dati.

⁵ Il restante 20,4% comprende non solo i laureati non in possesso di indirizzo e-mail ma anche coloro che, pur avendo un indirizzo, non hanno ricevuto l'e-mail di invito a causa di "rimbalzi permanenti" (ad es. e-mail errata) o non hanno potuto partecipare all'indagine a causa di "rimbalzi temporanei" (ad es. mail box piena, risposta automatica di assenza o di modifica dell'indirizzo e-mail). Nel complesso quindi i laureati non contattati tramite e-mail non hanno compilato, alla vigilia della laurea, il questionario ALMALAUREA, risiedono al Centro o al Sud ed appartengono al gruppo medico; hanno inoltre una scarsa conoscenza degli strumenti informatici.

maggiore diffusione della posta elettronica tra i laureati in anni recenti e al minore numero di rimbaldi⁶ registrati negli ultimi anni, cioè alla migliore qualità dell'informazione.

Il collettivo dei laureati in possesso di indirizzo e-mail è composto sia da laureati di primo livello del 2004 (17%), del 2005 (24% in totale, il 9% del periodo gennaio-aprile e il 15% del periodo settembre-dicembre), del 2007 (in generale 43%, il 17% del periodo gennaio-giugno ed il 26% del periodo luglio-dicembre) sia da laureati di secondo livello del 2007 (16% in totale, di cui il 6% del periodo gennaio-giugno e il 10% del periodo luglio-dicembre)⁷.

Tab. 1 - Composizione del collettivo oggetto di analisi

Codice indagine	Tipo di laureati e periodo di conseguimento della laurea	Totale laureati		Laureati in possesso di e-mail	
		Valori assoluti	Tasso di possesso e-mail	Valori assoluti	Composizione per indagine (% colonna)
07	L 2004 ad UN anno	47.380	74,0	35.043	17,4
08	L 2005 ad UN anno genn.-apr.	23.697	74,7	17.696	8,8
10	L 2005 ad UN anno sett.-dic.	39.054	77,9	30.415	15,1
16	L 2007 ad UN anno genn.-giugno	41.591	81,1	33.711	16,7
18	LS 2007 ad UN anno genn.-giugno	13.956	83,9	11.706	5,8
22	L 2007 ad UN anno luglio-dic.	63.848	83,4	53.258	26,4
23	LS 2007 ad UN anno luglio-dic.	24.114	83,2	20.062	9,9
TOTALE		253.640	79,6	201.891	100,0

Nota: L = laureate di primo livello; LS = laureate di secondo livello

Fonte: ALMALAUREA, anni vari.

Le indagini prese in considerazione, benché molto simili per gli argomenti trattati, si differenziano invece per altre caratteristiche strutturali come ad esempio l'ora e il giorno di avvio, la durata dell'indagine, il numero di domande presenti nel questionario⁸, il numero di solleciti effettuati⁹ e il giorno in cui sono stati inviati (Tab. 2).

In particolare, tutte le indagini sui laureati di primo e secondo livello del 2007 sono state avviate di lunedì, le due indagini sui laureati del 2005 di martedì e solo l'indagine sui laureati del 2004 di giovedì. Inoltre quattro indagini sono partite all'ora di pranzo (tra

⁶ I rimbaldi sono in media del 4% sull'intero collettivo dei laureati indagati ma oscillano tra il 5% sui laureati di primo livello del 2004 e del periodo settembre-dicembre del 2005 e l'1% sui laureati del primo periodo del 2007.

⁷ Per far sì che i laureati fossero intervistati alla stessa distanza temporale dalla laurea, condizione necessaria per garantire la comparabilità dei dati riferiti ad un intero anno solare, tutte le indagini, ad esclusione di quella sui laureati del 2004, sono state suddivise in due o tre periodi di rilevazione: per il 2005 sono stati intervistati prima i laureati di gennaio-aprile, poi quelli di maggio-agosto (solo CATI) ed infine quelli di settembre-dicembre; per il 2007 invece la prima *tranche* ha riguardato i laureati del periodo gennaio-giugno e la seconda quelli di luglio-dicembre.

⁸ L'informazione relativa al numero di domande distingue tra: domande visibili (cioè domande visualizzate nel momento in cui si accede al questionario on-line), domande minime (cioè quelle corrispondenti al percorso più breve, generalmente associato a laureati che non lavorano) e domande massime previste dal questionario (cioè quelle corrispondenti al percorso più lungo, rivolte quindi a laureati occupati).

⁹ Come noto, il sollecito corrisponde ad un'ulteriore e-mail inviata dopo circa una settimana dalla precedente a tutti coloro che non hanno ancora risposto al questionario on-line. Nelle indagini qui considerate si è utilizzato almeno un sollecito.

le 12.30 e le 14.00) mentre tre sono partite nella prima mattinata (indagine sui laureati 2004, sui laureati 2005 del periodo gennaio-aprile, sui laureati di secondo livello del periodo luglio-dicembre del 2007).

Tab. 2 – Caratteristiche strutturali delle indagini oggetto di analisi

Variabili strutturali	Codice indagine						
	07	08	10	16	18	22	23
Giorno di avvio dell'indagine	giovedì	martedì	martedì	lunedì	lunedì	lunedì	lunedì
Ora di avvio dell'indagine	9.00-11.30	9.00-11.30	12.30-14.00	12.30-14.00	12.30-14.00	12.30-14.00	9.00-11.30
Durata dell'indagine in giorni	27	21	21	30	29	23	22
Numero di solleciti effettuati	2	1	1	2	2	2	2
Giorno di invio del 1° sollecito	giovedì	martedì	mercoledì	martedì	martedì	martedì	martedì
Giorno di invio del 2° sollecito	giovedì	-	-	lunedì	lunedì	mercoledì	mercoledì
Numero di domande visibili	6,5	6,5	7	8	8	8	8
Numero minimo di domande	7,5	9,5	9	11	10	11	10
Numero massimo di domande	50,5	18,5	18	23	19	23	19

* Nel caso di batterie di domande il valore attribuito a ciascuna variabile è di 0,5

Le varie indagini hanno inoltre avuto durata diversa compresa tra i 21 giorni delle due indagini sui laureati del 2005 e i 30 giorni dell'indagine sui laureati di primo livello del periodo gennaio-giugno del 2007.

Rilevante è inoltre sottolineare che, nello strutturare lo schema di rilevazione delle prime indagini, si è tenuto conto della necessità di raccogliere il maggior numero di risposte senza sollecitare eccessivamente i laureati. Di conseguenza, la durata dell'indagine, il numero di solleciti da effettuare e il giorno in cui inviarli sono stati definiti in corso di rilevazione, sulla base dell'andamento dei tassi di risposta giornalieri. Al contrario, l'introduzione in anni più recenti della doppia metodologia (CAWI e CATI) ha comportato la necessità di stabilire a priori lo schema di rilevazione dell'indagine web: in particolare, devono essere definite sia la durata, in quanto il termine della CAWI coincide con l'avvio della rilevazione telefonica, sia il numero dei solleciti da effettuare (generalmente due per massimizzare, nel tempo a disposizione, il numero di risposte ottenute, minimizzando così il numero di laureati da intervistare telefonicamente).

3. Analisi dei tassi di risposta

La partecipazione alle indagini web qui analizzate è stata elevata se si considera che nel complesso ha risposto al questionario on-line il 43% dei laureati coinvolti¹⁰. Tale risultato è molto buono per lo strumento di rilevazione utilizzato, anche se decisamente

¹⁰ Il tasso di risposta è calcolato su coloro che sono in possesso di indirizzo e-mail valido; è depurato cioè da coloro per i quali non è noto l'indirizzo e-mail e dai cosiddetti rimbalzi (cfr. note 6 e 7).

più basso rispetto a quello generalmente osservato nelle indagini telefoniche ALMALAUREA: ad esempio, l'indagine CATI sulla condizione occupazionale dei laureati pre-riforma svolta da ALMALAUREA nel 2008 ha ottenuto un tasso di risposta che oscilla tra il 75% (per i laureati indagati a cinque anni dalla laurea) e il 90% (per i laureati ad un anno).

Il tasso di risposta al termine della rilevazione varia notevolmente a seconda dell'indagine web (Fig. 1): dal 38% per i laureati di secondo livello del periodo luglio-dicembre 2007 al 52% per i laureati di primo livello dell'anno solare 2004.

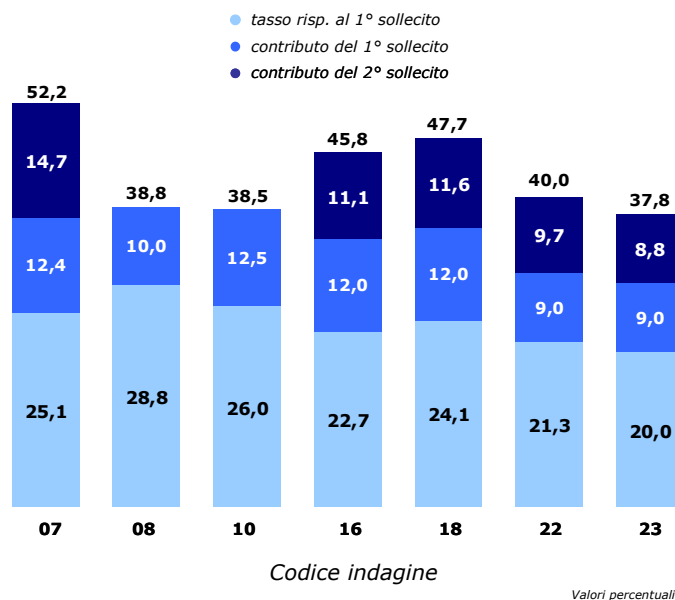


Fig. 1 - Tassi di risposta finali e contributo dei vari solleciti per singola indagine

Fonte: ALMALAUREA, anni vari.

La partecipazione all'indagine è determinata, come si vedrà meglio in seguito, da vari fattori, legati alle caratteristiche della popolazione e della rilevazione; tra questi, rilevante è il contributo dato dai solleciti: in corrispondenza di questi si registra un aumento del numero di questionari raccolti che varia tra i 9 e i 12,5 punti percentuali per il primo sollecito e tra i 9 e i 15 punti per il secondo¹¹.

Escludendo l'indagine 23 (laureati di secondo livello del periodo luglio-dicembre 2007), i minori tassi di risposta si sono registrati tra i laureati di primo livello del 2005 (indagini 8 e 10). Ciò tuttavia non sembra dipendere da uno scarso interesse per l'indagine: questi laureati hanno fatto registrare infatti i maggiori tassi di risposta iniziali¹² (rispettivamente, 29 e 26%) e sono stati sollecitati solo una volta¹³. Si può ritenere,

¹¹ Il contributo del primo sollecito è calcolato come rapporto tra il numero di questionari compilati dal giorno di invio del primo sollecito al giorno prima dell'invio del secondo sollecito (o al termine dell'indagine nel caso di un solo sollecito) e il totale dei laureati indagati. Analogamente, il contributo del secondo sollecito è calcolato come rapporto tra il numero di questionari compilati dal giorno di invio del secondo sollecito al giorno di chiusura della rilevazione e il totale dei laureati indagati.

¹² Il tasso di risposta iniziale corrisponde alla percentuale di risposte ottenute prima dell'invio del primo sollecito.

¹³ Per queste indagini si è fatto ricorso ad un solo sollecito visti gli ottimi risultati ottenuti nella prima settimana e con il (primo) sollecito. Come detto in precedenza, infatti, nelle prime indagini web svolte da ALMALAUREA, lo schema di rilevazione non era definito a priori; alcune scelte, relative ad esempio al numero di solleciti da

quindi, che scegliendo di effettuare un secondo sollecito, il tasso di risposta finale sarebbe stato addirittura superiore a quello medio raggiunto dalle rilevazioni con due solleciti.

L'analisi dei tassi di risposta giornalieri evidenzia un andamento simile delle curve che, al di là dei valori registrati nelle singole indagini, dipendono dal numero di solleciti e dal momento di avvio dell'indagine (mattina o pomeriggio). Per motivi di sintesi e per una migliore leggibilità, nella figura 2 sono rappresentate le curve dei tassi di risposta giornalieri di due sole indagini (relative ai laureati di primo livello, rispettivamente, del periodo gennaio-aprile 2005 e gennaio-giugno 2007).

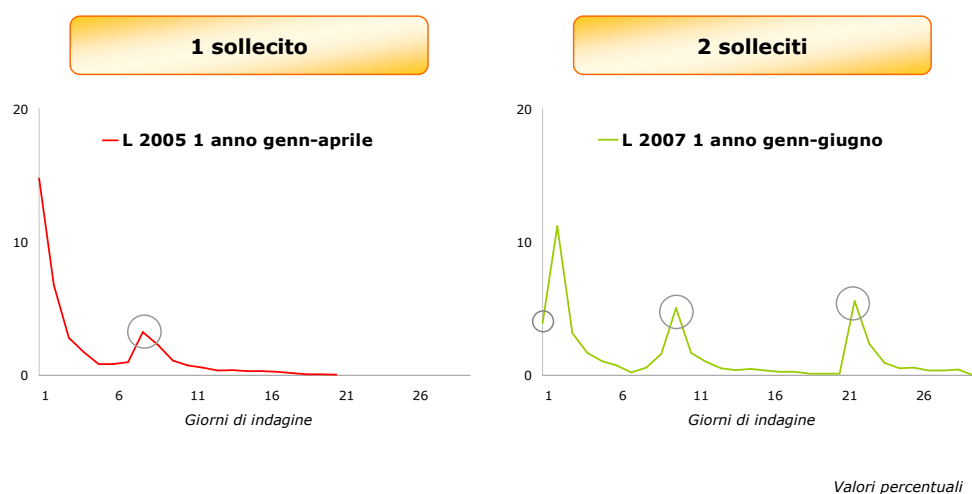


Fig. 2 - Andamento dei tassi di risposta giornalieri per due indagini tipo

Nota: nella figura è evidenziato sia l'avvio dell'indagine sia i picchi delle risposte in corrispondenza dei solleciti

Fonte: ALMALAUREA, anni vari.

Tali curve possono essere considerate rappresentative dell'andamento dei tassi di risposta giornalieri di tutte le indagini web analizzate: in particolare, la prima di un'indagine avviata di mattina e che ha utilizzato un solo sollecito; la seconda di un'indagine avviata di pomeriggio e che ha utilizzato due solleciti. Dal grafico si osserva chiaramente come la massima partecipazione sia limitata ai primissimi giorni di indagine, con un crollo consistente delle risposte nei giorni successivi: in particolare, il numero massimo di questionari compilati si osserva il primo giorno di indagine se questa è avviata di mattina, o il secondo giorno se è avviata di pomeriggio. Il sollecito ha un effetto positivo, determinando un picco del numero di risposte in corrispondenza del suo invio; il numero di picchi osservati è ovviamente funzione del numero di solleciti effettuati. Tuttavia, anche in questo caso l'effetto è limitato ai giorni immediatamente successivi ed è seguito da un'ulteriore caduta del numero di risposte.

effettuare e al giorno in cui inviarli, sono state compiute in corso di rilevazione, in particolare in base all'andamento dei tassi di risposta giornalieri.

4. Una misura della propensione a rispondere e della velocità di risposta ad un'indagine web

La banca-dati ALMALAUREA dispone di informazioni relative alle caratteristiche socio-demografiche dei laureati (genere, età alla laurea), percorso universitario compiuto (area geografica dell'ateneo di conseguimento della laurea, gruppo disciplinare, voto di laurea, regolarità negli studi¹⁴, intenzione di proseguire gli studi, ipotesi di reinscrizione all'università), ulteriori conoscenze acquisite come ad esempio la conoscenza degli strumenti informatici.

Queste informazioni derivano da due fonti: da un lato le Università che forniscono ad ALMALAUREA tutte le informazioni amministrative riguardanti i propri laureati, dall'altro gli stessi laureati che, compilando un questionario alla vigilia della laurea¹⁵ e pubblicando il proprio curriculum sul sito di ALMALAUREA, mettono a disposizione delle aziende, alla ricerca di personale, informazioni riguardanti la propria carriera universitaria conclusa, le proprie esperienze universitarie e lavorative pregresse, le proprie competenze, le proprie aspirazioni etc. Quest'ampia disponibilità di informazioni sui laureati unita ai dati riguardanti le caratteristiche strutturali dell'indagine (giorno e ora di avvio dell'indagine, numero dei solleciti etc.) ha reso possibile l'individuazione di un profilo degli intervistati e dei non intervistati delle indagini CAWI prese in esame attraverso l'utilizzo della procedura statistica DEMOD. Tale metodo permette infatti di identificare le variabili che, indipendentemente dalla loro natura, caratterizzano un determinato gruppo di soggetti. Attraverso opportuni test probabilistici (di fatto dei chi-quadrato), infatti, si individuano le modalità o le medie delle variabili (a seconda della tipologia) che risultano significativamente diverse rispetto al complesso della popolazione in esame; queste possono essere tra l'altro ordinate in modo decrescente a seconda della propria capacità di connotare il gruppo.

I risultati di questa analisi, presentati di seguito, devono essere letti quindi in termini probabilistici come la maggiore o minore probabilità di avere certe caratteristiche rispetto alla media della popolazione oggetto di analisi.

Intervistati. I partecipanti ad un'indagine CAWI hanno conseguito con maggiore probabilità lauree nei gruppi ingegneria, politico-sociale, economico-statistico in atenei del Centro Italia (e risiedono al Centro) e hanno avuto buone performance di riuscita negli studi: hanno un voto di diploma e di laurea superiore alla media, un'età alla laurea

¹⁴ Un laureato è definito regolare o in corso se ha conseguito il titolo universitario nei tempi previsti dall'ordinamento del suo corso (tre anni per i laureati di primo livello, due anni per i laureati specialistici, cinque o sei anni per i laureati specialistici a ciclo unico); in caso contrario si parla di laureati irregolari o fuori corso. A titolo di esempio la quota di laureati fuori corso in Italia è del 58% tra i laureati di primo livello del 2007 e del 41% tra i laureati di secondo livello del 2007.

¹⁵ Si tratta di un questionario somministrato ai laureandi prima del conseguimento del titolo in cui si raccolgono informazioni relative alle esperienze universitarie pregresse, alle attività lavorative svolte durante il corso di studi, alla soddisfazione per l'Università e le sue strutture etc.

inferiore alla media, sono per lo più laureati puri¹⁶, di primo livello, hanno conseguito il titolo in corso o al più un anno fuori corso e hanno frequentato regolarmente quasi tutte le lezioni. Il 70% ha genitori con almeno un diploma di scuola secondaria superiore (il 25% almeno un genitore con laurea, il 45% un diploma di scuola secondaria), il 32% proviene dalla classe media impiegatizia¹⁷ e con maggior probabilità è uno studente-lavoratore¹⁸ (65%) o non ha avuto alcuna esperienza di lavoro (26%). Non hanno, inoltre, effettuato periodi di studio all'estero, né hanno precedenti esperienze universitarie; hanno intenzione di proseguire gli studi dopo la laurea e, se potessero tornare indietro, si riscriverebbero all'Università, nell'80% dei casi allo stesso corso. Inoltre, coloro che partecipano ad un'indagine CAWI hanno con maggiore probabilità compilato il questionario prima della laurea, hanno una buona conoscenza della navigazione in rete, dei programmi Word Processor, dei fogli elettronici e dei sistemi operativi e, più in generale, come era ovvio attendersi, il 72% conosce più di tre strumenti informatici.

Per quanto riguarda le variabili strutturali, gli intervistati rispondono con maggiore probabilità se la distanza tra l'avvio della riforma universitaria e l'inizio dell'indagine¹⁹ è di 1 anno, se l'indagine è stata avviata in mattinata, prevalentemente di giovedì, se la durata dell'indagine e il numero dei solleciti sono superiori alla media, se il numero delle domande visibili e i giorni trascorsi dalla laurea all'avvio dell'indagine²⁰ sono inferiori alla media.

Non intervistati. I non intervistati sono con maggiore probabilità laureati in atenei del Nord Italia (59%) e residenti al Nord (52%), hanno un'età alla laurea più elevata della media così come la durata degli studi, mentre il voto di laurea e di diploma è inferiore alla media. Sono con maggiore probabilità laureati ibridi²¹, alla laurea non hanno manifestato l'intenzione di proseguire gli studi, provengono prevalentemente da famiglie economicamente più avvantaggiate ed hanno inoltre una scarsa conoscenza degli strumenti informatici (realizzazione siti web, sistemi operativi, multimedia...).

Inoltre l'indagine a cui avrebbero dovuto partecipare è stata avviata con maggiore probabilità nel primo pomeriggio, di lunedì o martedì; il tempo intercorso dall'avvio della

¹⁶ Sono considerati laureati puri coloro che hanno compiuto il loro percorso di studi per intero ed esclusivamente nel nuovo ordinamento, mentre sono considerati ibridi coloro che hanno ottenuto il titolo concludendo un percorso di studi iniziato nel vecchio ordinamento.

¹⁷ Tale schema è stato proposto da A. Cobalti e A. Schizzerotto, 1994, in "La mobilità sociale in Italia", Il Mulino, Bologna

¹⁸ I lavoratori-studenti sono i laureati che hanno dichiarato di avere svolto attività lavorative continuative a tempo pieno per almeno la metà della durata degli studi sia nel periodo delle lezioni universitarie sia al di fuori delle lezioni. Gli studenti-lavoratori sono tutti gli altri laureati che hanno compiuto esperienze di lavoro nel corso degli studi universitari.

¹⁹ Dal momento che nelle prime indagini sui laureati post-riforma si poneva l'accento sul fatto che la partecipazione era importante per valutare i primi effetti della riforma universitaria, è stata costruita questa variabile per cercare di misurare l'effetto di questa diversa comunicazione fatta ai laureati sulla partecipazione all'indagine.

²⁰ La distanza dell'intervista dalla laurea è calcolata come differenza tra la data di avvio dell'indagine e la data di laurea.

²¹ Cfr. nota 17.

riforma universitaria all'inizio dell'indagine è stato di 4 anni (60%) o di 2 anni (26%) e il tempo tra la laurea e l'avvio dell'indagine è superiore alla media. Il numero di domande visibili è superiore alla media mentre la durata dell'indagine e il numero dei solleciti è inferiore alla media.

La probabilità di rispondere ad un'indagine web sembra dunque dipendere da un certo insieme di variabili strutturali ed individuali che, verosimilmente, determina anche diversi tempi di risposta. Per studiare in maniera più approfondita il fenomeno si è utilizzato il modello di regressione di Cox (anche conosciuto come *proportional hazard regression model*) che consente di stimare le curve dei tempi di risposta tenendo sotto controllo l'effetto esercitato da un insieme di variabili esplicative. Attraverso il modello si è infatti appurato che diversi sono gli elementi che influenzano la velocità di risposta ad un'indagine web: in particolare, il percorso di studio compiuto, le conoscenze informatiche, il voto di laurea, l'intenzione di proseguire gli studi e l'area geografica dell'ateneo di conseguimento del titolo.

Il modello di regressione di Cox segue l'approccio dei modelli sull'analisi dei dati di sopravvivenza, attraverso i quali si cerca di esplorare simultaneamente l'effetto di alcune variabili sul tempo di sopravvivenza atteso e di confrontare distribuzioni di sopravvivenza fra diverse sub-popolazioni.

Nella nostra analisi il modello di regressione di Cox è stato utilizzato per misurare la funzione di sopravvivenza nella condizione di non intervistato che caratterizza ciascun individuo i , $S_i(t)$, dato un certo insieme di variabili esplicative.

Formalmente il modello è espresso attraverso la funzione:

$$S_i(t) = S_0(t) \exp\left(\sum_{j=1}^J x_{ij} \beta_j\right)$$

dove $S_i(t)$ indica la funzione di sopravvivenza che caratterizza l'individuo i , ovvero la probabilità di non essere ancora stato intervistato dopo t mesi dall'avvio dell'indagine; x_{ij} indica il valore assunto dalla variabile X_j in corrispondenza dell'individuo i ; β_j indica il parametro che esprime l'effetto esercitato dalla variabile X_j sulla funzione di sopravvivenza; e $S_0(t)$ indica la funzione di sopravvivenza di base, cioè quella relativa all'individuo in corrispondenza del quale $x_{ij}=0$ per ogni $j=1, \dots, J$.

Il modello è stato applicato agli oltre 200 mila laureati contattati tramite e-mail ed invitati a compilare il questionario ospitato sul sito di ALMALAUREA²². Le variabili potenzialmente rilevanti nel determinare differenti tempi di risposta, e quindi considerate nel modello, sono di due tipi: strutturali, che, come già anticipato, sono legate alle caratteristiche della rilevazione, e individuali, ossia legate alle caratteristiche della

²² Come esplicitato nel precedente § 2, le analisi sono state effettuate sul collettivo di laureati con e-mail valida. Sono stati cioè esclusi dall'analisi i laureati privi di indirizzo e-mail e i rimbaldi.

popolazione indagata. Le variabili strutturali considerate sono: giorno della settimana e ora di avvio dell'indagine, numero di domande visibili nel momento in cui si accede al questionario on-line, numero di domande minime e massime previste dal questionario²³, distanza dell'intervista dalla laurea e numero di anni dall'avvio della riforma; non sono state considerate le informazioni relative al numero di solleciti effettuati e al giorno della settimana in cui sono stati inviati in quanto, come precedentemente detto, per alcune indagini non si è trattato di variabili ante-rilevazione ma di elementi valutati successivamente, soprattutto sulla base dell'andamento dei tassi di risposta giornalieri. Le variabili individuali inserite nel modello sono: genere, età alla laurea, area geografica dell'ateneo di conseguimento del titolo²⁴, gruppo disciplinare, voto di laurea, regolarità negli studi, intenzione di proseguire gli studi, ipotesi di reinscrizione all'università²⁵ e numero di strumenti informatici conosciuti. Di queste, solo il giorno della settimana di avvio dell'indagine e il numero minimo di domande previste dal questionario non sono risultate significative, e non sono pertanto entrate nel modello.

Nella seguente tabella sono riportate le stime dei parametri del modello di Cox: i parametri B stimati dal modello rappresentano l'effetto esercitato dalla singola modalità considerata rispetto alla classe di riferimento (indicato in grassetto nella tabella) della variabile esplicativa.

Se il parametro B stimato assume un valore positivo significa che in corrispondenza di quella modalità diminuisce, rispetto alla classe di riferimento, la probabilità di rimanere nella condizione di non intervistato; al contrario, se il parametro assume un valore negativo, allora aumenta la probabilità di rimanere nella condizione di non intervistato. L'EXP (B) è il tasso con cui varia l'ODDS RATIO (dato dal rapporto tra la probabilità che si verifichi l'evento "intervista" e la probabilità che si verifichi l'evento opposto) al variare della variabile indipendente o delle modalità della variabile indipendente. La probabilità che si verifichi l'evento intervista aumenta se l'EXP (B) è maggiore di 1, diminuisce se è minore di 1, rimane immutata se l'EXP assume valore unitario.

In genere si è soliti valutare le sole modalità in corrispondenza delle quali il livello di significatività (riportato nella colonna Sig. della tabella 3) è inferiore a 0,05. Nel caso in esame tutte le modalità delle variabili risultate significative corrispondono a livelli di significatività inferiori a 0,035.

²³ Cfr. nota 9.

²⁴ Si è scelto di considerare l'area geografica dell'ateneo di conseguimento della laurea, e non l'area geografica di residenza, per il suo maggior potere esplicativo.

²⁵ L'ipotesi di reinscrizione è stata inserita nel modello perché può essere interpretata come proxy dell'affezione dei laureati verso l'Università in cui hanno conseguito il titolo.

Tab. 3 - Modello di Cox per la stima dei tempi di risposta ad un'indagine web: stima dei parametri e bontà di adattamento del modello

	B	Exp(B)	Sig.
Ora di avvio dell'indagine (12.30-14.00=0)	-0,144	0,866	0,000
Numero massimo di domande	-0,016	0,984	0,000
Numero di domande visibili	-0,416	0,660	0,000
Distanza (in semestri) dell'intervista da laurea	-0,044	0,957	0,000
Anni dall'avvio della riforma (1 anno=0)			0,000
2 anni dalla riforma	-0,673	0,510	0,000
4 anni dalla riforma	-0,240	0,787	0,013
Genere (Uomini=0)	0,063	1,065	0,000
Area geografica dell'ateneo (Nord=0)			0,000
Centro	0,103	1,108	0,000
Sud e Isole	0,032	1,033	0,000
Gruppo disciplinare (Scientifico=0)			0,000
Chimico-farmaceutico	-0,248	0,780	0,000
Geo-biologico	-0,116	0,890	0,000
Medico	-0,736	0,479	0,000
Ingegneria	-0,087	0,917	0,000
Architettura	-0,441	0,643	0,000
Agrario	-0,209	0,811	0,000
Economico-statistico	-0,256	0,774	0,000
Politico-sociale	-0,289	0,749	0,000
Giuridico	-0,437	0,646	0,000
Letterario	-0,342	0,711	0,000
Linguistico	-0,244	0,784	0,000
Insegnamento	-0,444	0,642	0,000
Psicologico	-0,217	0,805	0,000
Educazione fisica	-0,451	0,637	0,000
Difesa e sicurezza	-1,385	0,250	0,000
Voto di laurea	0,008	1,008	0,000
Età alla laurea	-0,008	0,992	0,000
Regolarità negli studi (In corso=0)			0,002
I anno f.c.	-0,018	0,982	0,034
II anni f.c. e oltre	-0,035	0,966	0,001
Si reiscriverebbe (Sì, allo stesso corso=0)			0,000
Sì, ad un altro corso	0,021	1,022	0,021
No, non mi iscriverei più	-0,070	0,933	0,024
Non disponibile	-0,171	0,843	0,000
Intenzione di proseguire gli studi (Sì=0)			0,000
No	-0,108	0,898	0,000
Non disponibile	-0,239	0,788	0,000
N. strumenti informatici conosciuti (Almeno tre strumenti=0)			0,000
Nessuno strumento	-0,353	0,703	0,000
Al più due strumenti	-0,115	0,891	0,000
Non disponibile	-0,168	0,845	0,021

Chi-quadrato del modello: 6420,067

Gradi di libertà del modello: 36

Significatività: 0,000

Fonte: ALMALAUREA, anni vari.

La probabilità di rimanere nella condizione di non intervistato può essere letta anche in termini di rischio, ossia rischio di subire l'evento intervista. Dunque, la maggiore o minore probabilità di rimanere nella condizione di non intervistato corrisponde ad una minore o maggiore probabilità di subire l'evento intervista²⁶ che, tenendo conto del tempo, può essere letta come velocità di risposta.

Per una lettura più efficace dei risultati sono riportati i grafici relativi alle due variabili maggiormente esplicative: in particolare, vengono poste a confronto le curve di rischio di subire l'evento "intervista" al tempo t per diversi gruppi disciplinari²⁷ (Fig. 3) e per numero di strumenti informatici conosciuti (Fig. 4).

A parità di condizioni, i laureati dei gruppi scientifico e ingegneria presentano le minori probabilità al tempo t di rimanere nella condizione di non intervistato e, conseguentemente, la maggiore velocità di risposta. Al contrario, i laureati dei gruppi difesa e sicurezza e medico sono caratterizzati da tempi di risposta decisamente più lunghi. Per esempio, al decimo giorno di indagine, per i laureati di questi ultimi due gruppi si osserva un tasso di risposta inferiore al 22%, mentre per i colleghi dei gruppi scientifico e ingegneria superiore al 35%, quota che non viene raggiunta dai primi nemmeno al termine dell'intero periodo di rilevazione: si stima infatti che il tasso di risposta al trentesimo giorno sia del 32% per i laureati del gruppo medico e solo del 18% per quelli del gruppo difesa e sicurezza.

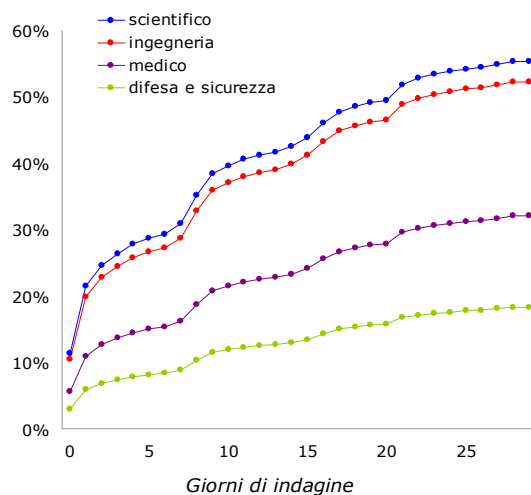


Fig. 3 - Tempi di risposta ad un'indagine web per gruppo disciplinare

Nota: Nella figura sono riportati i gruppi disciplinari più rappresentativi

Fonte: ALMALAUREA, anni vari.

²⁶ Il rischio o probabilità di risposta è, come ovvio, data dal complemento a 1 della probabilità di rimanere nella condizione di non intervistato.

²⁷ Per una migliore lettura, sono state rappresentate le curve dei tempi di risposta solo di alcuni gruppi disciplinari, quelli che presentano i minimi (curve più alte) e massimi (curve più basse) tempi di risposta. Le curve dei restanti gruppi si collocano in una posizione intermedia rispetto alle curve visualizzate.

Inoltre, come ci si poteva attendere, all'aumentare del numero di strumenti informatici conosciuti aumenta la probabilità di risposta al tempo t. Dalla figura 4 è infatti evidente la maggiore velocità di risposta per i laureati che conoscono almeno tre strumenti informatici; più lunghi, invece, i tempi di risposta per coloro che non conoscono alcuno strumento informatico.

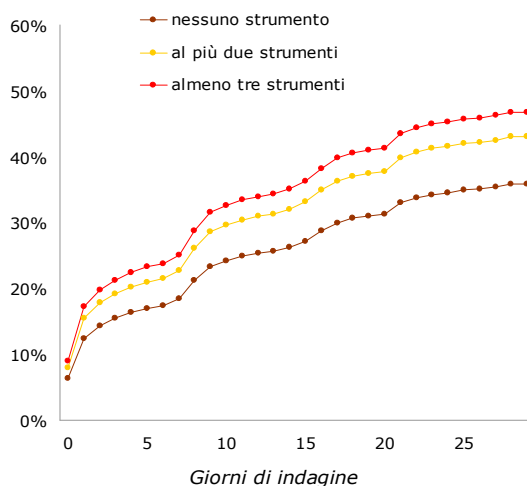


Fig. 4 - Tempi di risposta ad un'indagine web per numero di strumenti informatici conosciuti

Fonte: ALMALAUREA, anni vari.

Oltre al gruppo disciplinare e alle conoscenze informatiche, altre variabili individuali influenzano i tempi di risposta, che si riducono, in particolare, per i laureati con buone performance negli studi (elevato voto di laurea, ridotta età alla laurea, regolarità negli studi).

Inoltre tempi di risposta più brevi sono associati a laureati in atenei del Centro e, anche se in misura minore, del Sud (rispetto ai colleghi del Nord) e a coloro che alla vigilia della laurea hanno dichiarato di voler proseguire gli studi universitari.

Tra tutte le variabili strutturali, il numero di domande visibili è quella che esercita la maggiore influenza sulla velocità di risposta, a conferma quindi, come è noto, che l'impatto visivo è importante e che un numero eccessivo di domande può disincentivare i rispondenti a proseguire la compilazione del questionario: all'aumentare del numero di domande visibili, aumentano infatti i tempi di risposta.

Meno apprezzabili, ma comunque significativi gli effetti dell'ora di avvio dell'indagine e del numero massimo di domande previste dal questionario. Come visto nel precedente paragrafo 3, il maggior tasso di risposta giornaliero si registra nella prima giornata se l'indagine è avviata di mattina, o nella seconda se l'indagine è avviata di pomeriggio, proprio a causa del minor tempo a disposizione per compilare il questionario. Tuttavia, l'avvio in mattinata mostra una minore velocità di risposta rispetto all'avvio tra le 12.30 e

le 14.00, forse perché i laureati già impegnati in altre attività (ad es. lavorative o di studio) le sospendono verosimilmente nell'ora di pranzo: questo rappresenta dunque il momento migliore per avviare un'indagine web, quello in cui si ha una risposta più immediata e in cui i laureati sono maggiormente predisposti a rispondere.

Il numero massimo di domande previste dal questionario influisce negativamente sui tempi di risposta: all'aumentare del numero massimo di domande, infatti, diminuisce la velocità di risposta, in analogia a quanto osservato per le domande visibili, anche se l'effetto è inferiore.

5. Considerazioni conclusive

Il crescente ricorso da parte di ALMALAUREA alle indagini CAWI ha reso necessario approfondire la metodologia che ne sta alla base individuando quali sono le caratteristiche socio-demografiche e di curriculum degli intervistati e dei non intervistati e quali le caratteristiche strutturali dell'indagine che determinano una maggiore velocità di risposta.

Da queste analisi è emerso che:

1. i rispondenti sono un campione auto-selezionato caratterizzato da migliori performance nello studio e maggiori conoscenze informatiche;
2. i solleciti determinano un picco delle risposte al questionario web innalzando il tasso di risposta complessivo; tale effetto è però limitato ai giorni immediatamente successivi;
3. la velocità di rispondere ad un'indagine web dipende sia da fattori legati alle caratteristiche dei laureati (gruppo disciplinare di conseguimento della laurea, intenzione di proseguire gli studi etc.) sia da caratteristiche strutturali dell'indagine (giorno e ora di avvio, durata, etc...).

Tali evidenze potranno essere utili in futuro per adattare lo schema di rilevazione al collettivo da indagare, al fine di massimizzare l'efficienza e l'efficacia di questo strumento. In altre parole, alla luce di questi risultati per le future rilevazioni si valuterà l'opportunità o meno di effettuare indagini via web e si stabiliranno le caratteristiche dell'indagine stessa: numero di domande previste dal questionario, giorno e ora d'avvio dell'indagine, numero di solleciti, durata dell'indagine. Per esempio, per i laureati caratterizzati da brevi tempi di risposta (ad es. laureati dei gruppi scientifico e ingegneria o con buone conoscenze informatiche) si potrebbe pensare di raccogliere le informazioni solo attraverso il questionario on-line senza ricorrere alla successiva indagine telefonica; al contrario, per i laureati caratterizzati da tempi di risposta più lunghi (ad esempio appartenenti ai gruppi difesa e sicurezza e medico o con scarse conoscenze informatiche) potrebbe essere conveniente ricorrere direttamente ed esclusivamente all'indagine telefonica.

Bibliografia

Agresti A. (1990), *Categorical Data Analysis*, Wiley, New York.

Biffignandi S., Pratesi M.(2003), *Tempestività e qualità: aspetti concettuali e comportamenti di risposta nelle indagini via Internet. Un'applicazione ad un'indagine regionale sulle imprese*, 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.

Camillo F., Girotti C. (2007), *L'impatto dell'integrazione di tecniche multiple di rilevazione nell'indagine sulla condizione occupazionale ALMALAUREA: una misura di propensity score in spazi condizionati multivariati*, in: *IX Profilo dei laureati italiani. La riforma allo specchio*, Consorzio Interuniversitario ALMALAUREA (a cura del), Il Mulino, 289-309.

Cammelli A. (2009), *XI rapporto sulla condizione occupazionale dei laureati. Occupazione e occupabilità dei laureati a dieci anni dalla Dichiarazione di Bologna*, disponibile su www.almalaurea.it/universita/occupazione/occupazione07

Cobalti A., Schizzerotto A., 1994, "La mobilità sociale in Italia", Il mulino, Bologna

Hosmer D. W. and Lemeshow S. (1989), *Applied Logistic Regression*, John Wiley & Sons, New York.