

Nota metodologica

Introduzione

Tra febbraio e luglio 2014 l'Istituto nazionale di statistica ha svolto la seconda edizione dell'indagine sull'inserimento professionale dei dottori di ricerca, intervistando quanti avevano conseguito il dottorato nel 2008 e nel 2010, con l'obiettivo di rilevarne la condizione occupazionale a sei e quattro anni di distanza dal conseguimento del titolo.

La rilevazione fa parte del sistema integrato di indagini sulla transizione istruzione-lavoro: il sistema informativo si compone anche dell'indagine sui percorsi di studio e di lavoro dei diplomati di scuola secondaria di secondo grado e dell'indagine sull'inserimento professionale dei laureati.

Il sistema rappresenta, da una parte, una rilevante fonte di dati per l'analisi comparativa della resa dei diversi titoli di studio sul mercato del lavoro, dall'altra, uno strumento utile per valutare l'efficacia del sistema di istruzione superiore nel suo complesso. Le indagini consentono, tra l'altro, di indagare sull'influenza dell'origine sociale, sia sul processo di selezione scolastica e universitaria, sia sul processo di transizione al lavoro.

La popolazione di riferimento

La rilevazione relativa ai dottori di ricerca ha riguardato due coorti, ovvero quanti hanno conseguito il titolo di dottore di ricerca negli anni solari 2008 e 2010. L'indagine, pertanto, ha rilevato la condizione occupazionale a sei e quattro anni dal titolo e, diversamente dalle altre rilevazioni del sistema, che sono campionarie, ha riguardato tutti i dottori di ricerca delle due leve.

L'indagine si è svolta in due fasi: nella prima sono stati coinvolti gli atenei, mentre la seconda ha interessato direttamente i dottori di ricerca.

La rilevazione presso gli atenei ha avuto la finalità di predisporre gli elenchi dei nominativi, dei recapiti telefonici e abitativi e degli indirizzi email dei dottori da contattare; le Università hanno inviato le liste in modalità telematica, trasmettendole al sito web dell'Istat dedicato all'indagine. I recapiti abitativi comunicati dagli atenei sono stati aggiornati tramite link nominativo alle Liste Anagrafiche Comunali; a loro volta i recapiti telefonici sono stati aggiornati tramite link alle liste Consodate, in vista dell'attività di sollecito telefonico.

L'universo di riferimento dell'indagine è costituito complessivamente da 22.469 dottori di ricerca: 11.229 del 2008 e 11.240 del 2010.

La rilevazione e il trattamento dei dati

La rilevazione è stata svolta utilizzando la tecnica CAWI (*Computer Assisted Web Interviewing*): si tratta cioè di un questionario online che il dottore di ricerca compila in modo autonomo, accedendo con codici personali e segreti.

Le interviste sono state svolte nel periodo compreso tra febbraio e luglio 2014. A tutti i dottori di ricerca coinvolti nell'indagine è stata inviata, prima dell'intervista, una lettera cartacea a firma del Presidente dell'Istat, contenente una breve presentazione dell'indagine e i codici personali di accesso.

Il questionario online è stato realizzato mediante un software sviluppato dall'Istat (GINO++). All'interno del questionario elettronico sono stati introdotti dei controlli interni di coerenza, sia tra le risposte fornite dagli intervistati, sia tra le risposte e le informazioni fornite dagli atenei.

Il questionario di rilevazione si articola in cinque sezioni. La prima sezione è dedicata al curriculum degli studi e alle attività di qualificazione, la seconda al lavoro, la terza alla ricerca del lavoro, la quarta alla mobilità territoriale (con particolare riferimento a quella verso altri Paesi) e nella quinta si raccolgono notizie relative alla famiglia.

Nella prima sezione sono richieste informazioni sui percorsi di studio intrapresi, ponendo particolare attenzione all'esperienza del corso di dottorato.

La seconda sezione è dedicata al lavoro e, nel caso in cui il dottore dichiara di svolgere una attività lavorativa (comprese attività formative retribuite, come borse di studio e assegni di ricerca), ne approfondisce i seguenti aspetti: tipo di occupazione, settore di attività economica, posizione nella professione, tipo di contratto, professione (codificata secondo la CP2011), orario di lavoro, reddito, luogo di lavoro, eventuali attività di ricerca e sviluppo, grado di soddisfazione rispetto al lavoro svolto. Una sottosezione rileva la produzione scientifica del dottore di ricerca dal dottorato ad oggi, indipendente dall'eventuale lavoro svolto: articoli, volumi monografici, partecipazione a progetti di ricerca ecc. Un'ulteriore sottosezione mira a descrivere la prima esperienza lavorativa intrapresa subito dopo la conclusione del dottorato, che per molti coincide con l'ingresso nel modo del lavoro.

La terza sezione, dedicata alla ricerca di lavoro, è rivolta a tutti i dottori, occupati o meno. La sezione relativa alla mobilità mira a tracciare i percorsi di mobilità dai tempi dell'università al momento dell'intervista, con particolare attenzione agli spostamenti per studio e lavoro. L'ultima sezione, infine, raccoglie informazioni sulla famiglia attuale e su quella di origine, permettendo l'analisi della mobilità sociale.

L'Istat si è avvalsa di una ditta esterna per fornire un servizio di Contact Center ai dottori di ricerca coinvolti nell'indagine. Gli operatori, selezionati tra persone di esperienza e adeguata specializzazione, hanno gestito i contatti in entrata e in uscita con i rispondenti, tramite una casella di posta elettronica dedicata e un numero telefonico gratuito. La ditta si è fatta carico anche di effettuare periodici e cadenzati contatti email e telefonici di sollecito, volti ad ottenere la massima partecipazione e, ove il questionario fosse stato lasciato in sospeso, il completamento dell'intervista.

Al termine della fase di rilevazione una piccola quota di interviste è stata scartata (237 questionari), in quanto incomplete e ritenute non sufficientemente informative. I questionari compilati in modo utile sono stati sottoposti ad un piano di controllo e correzione, consistente in correzioni deterministiche e probabilistiche per il trattamento delle inconsistenze e delle mancate risposte parziali, secondo procedure e accurati standard qualitativi adottati in Istat.

I tassi di risposta

La partecipazione dei dottori alla rilevazione è risultata molto elevata: sono state realizzate 16.322 interviste valide su una popolazione di 22.459 dottori, corrispondente ad un tasso di risposta pari al 72,64%. Nel dettaglio, sono state raccolte 7888 interviste sui 11.229 dottori del 2008 e 8.434 sui 11.240 dottori 2010.

La risposta è stata più elevata nella coorte più di recente (75,04 % dei dottori 2010, rispetto al 70,25 % della coorte 2008), per la naturale mobilità territoriale di questa fascia di popolazione.

A seguire, sono riportati i tassi di risposta secondo le principali variabili che sono utilizzate per calcolare i pesi di correzione per mancata risposta totale, che sono:

- ateneo, con aggregazione di quelli con meno di 10 unità (tabella 1);
- area disciplinare e sesso (tabella 2).

TABELLA 1. TASSI DI RISPOSTA PER ATENEIO E ANNO DI CONSEGUIMENTO DEL TITOLO.

Ateneio	Dottori 2008	Dottori 2010	Ateneio	Dottori 2008	Dottori 2010
Torino - Università degli studi	74,8	76,0	Pisa - Università degli studi	68,5	73,6
Torino - Politecnico	79,2	83,7	Pisa - Scuola Normale Superiore	74,5	64,3
Novara - Università degli studi del Piemonte orientale A. Avogadro	71,4	71,4	Pisa - Scuola Superiore di Studi Universitari e Perfezionamento S. Anna	55,3	58,2
Alessandria - Università degli studi del Piemonte orientale A. Avogadro	81,8	81,8	Siena - Università degli studi	73,5	77,7
Genova - Università degli studi	72,6	72,8	Perugia - Università degli studi	78,9	84,3
Varese - Università Insubria	62,0	83,1	Viterbo - Università della Tuscia	93,7	84,6
Milano - Università degli studi	67,9	72,4	Roma - Università degli studi La Sapienza	65,0	76,3
Milano - Politecnico	69,2	71,7	Roma - Università degli studi di Tor Vergata	74,4	71,7
Milano - Università commerciale Luigi Bocconi	63,6	68,6	Roma - Libera Università Maria SS. Assunta (LUMSA)	(a)	81,8
Milano - Università Cattolica del Sacro Cuore	71,9	70,9	Roma - Libera Università internazionale degli studi sociali Guido Carli (LUISS)	68,4	63,4
Piacenza - Sede Università Cattolica del Sacro Cuore di Milano	100,0	70,6	Roma - Università degli studi del Foro Italico	(a)	83,3
Roma - Sede Università Cattolica del Sacro Cuore di Milano	62,7	70,4	Roma - Università degli studi Roma Tre	77,7	83,9
Milano - Libera Università di Lingue e Comunicazione (IULM)	85,0	80,0	Cassino - Università degli studi	66,7	75,4
Milano - Università Vita-Salute San Raffaele	44,4	77,8	Benevento - Università degli studi del Sannio	80,0	72,4
Milano-Bicocca - Università degli studi	65,4	80,3	Napoli - Università degli studi Federico II	68,5	68,4
Bergamo - Università degli studi	75,0	80,4	Napoli - Università degli studi Parthenope	70,6	82,8
Brescia - Università degli studi	78,7	86,0	Napoli - Università degli studi Orientale	52,4	65,5
Pavia - Università degli studi	69,8	75,2	Napoli - Istituto Universitario Suor Orsola Benincasa	79,3	84,2
Pavia - Istituto Universitario di Studi Superiori	(a)	64,7	Napoli - Seconda Università degli studi	51,8	56,2
Bolzano - Libera Università	41,7	68,8	Salerno - Università degli studi	62,0	72,3
Trento - Università degli studi	69,9	68,5	L'Aquila - Università degli studi	76,9	82,5
Verona - Università degli studi	78,6	77,3	Teramo - Università degli studi	64,3	76,6
Venezia - Università degli studi Ca Foscari	72,4	82,0	Pescara - Sede Università D'Annunzio di Chieti	52,6	60,0
Venezia - Istituto Universitario di Architettura Iuav	68,2	74,3	Chieti - Università degli studi Gabriele D'Annunzio	56,8	69,9
Padova - Università degli studi	74,0	77,7	Campobasso - Università degli studi del Molise	81,6	83,8
Udine - Università degli studi	80,2	80,8	Foggia - Università degli studi	79,1	82,5
Trieste - Università degli studi	74,6	81,8	Bari - Università degli studi	55,9	62,0
Trieste - Scuola Internazionale Superiore di Studi Avanzati	68,9	75,9	Bari - Politecnico	75,6	89,6
Parma - Università degli studi	67,9	71,2	Lecce - Università del Salento	80,4	77,6
Modena - Sede Università di Modena e Reggio Emilia	61,2	66,0	Potenza - Università degli studi della Basilicata	86,8	74,6
Bologna - Università degli studi	76,7	79,8	Arcavacata di Rende - Università degli studi della Calabria	67,5	68,3
Ravenna - Sede Università degli studi di Bologna	(a)	85,7	Catanzaro - Università degli studi Magna Grecia	63,0	66,7
Ferrara - Università degli studi	70,2	80,5	Reggio Calabria - Università degli studi Mediterranea	81,1	85,5
Urbino - Università degli studi	75,4	83,0	Palermo - Università degli studi	65,3	78,4
Ancona - Università Politecnica delle Marche	70,9	88,3	Messina - Università degli studi	72,8	74,1
Macerata - Università degli studi	74,1	77,1	Catania - Università degli studi	62,7	65,1
Camerino - Università degli studi	63,9	64,0	Sassari - Università degli studi	66,4	71,6
Lucca - Scuola IMT Alti Studi	91,7	69,2	Cagliari - Università degli studi	80,1	84,2
Firenze - Università degli studi	69,0	80,9	Altri Atenei (b)	75,4	87,8
			Totale	70,2	75,0

(a) Sono inclusi in "Altri Atenei".

(b) I seguenti Atenei non compaiono nella lista ma sono inclusi in "Altri Atenei", in quanto in entrambi gli anni hanno un numero di dottori di ricerca inferiore alle dieci unità: Vercelli - Università degli studi del Piemonte orientale A. Avogadro; Castellanza - Università Carlo Cattaneo (LIUC); Forlì - Sede Università degli studi di Bologna; Ascoli Piceno - Sede Università degli studi di Camerino; Firenze - Istituto Italiano di Scienze Umane; Siena - Università per stranieri; Roma - Università Campus Bio-Medico; Roma - Libera Università degli Studi Per Innovazione e le Organizzazioni (ex Libera Università San Pio V di Roma); Roma - Università telematica Guglielmo Marconi; Isernia - Sede Università del Molise.

TABELLA 2. TASSI DI RISPOSTA PER AREA DISCIPLINARE, SESSO E ANNO DI CONSEGUIMENTO DEL TITOLO

Area disciplinare	Dottori 2008			Dottori 2010		
	Uomo	Donna	Totale	Uomo	Donna	Totale
Scienze matematiche e informatiche	70,6	84,5	74,8	79,9	83,9	81,3
Scienze fisiche	72,5	63,8	69,7	76,9	77,4	77,1
Scienze chimiche	73,3	71,7	72,4	71,3	76,9	74,2
Scienze della terra	77,6	77,4	77,5	81,6	82,9	82,2
Scienze biologiche	69,5	71,5	70,9	74,6	77,7	76,6
Scienze mediche	61,5	67,0	65,1	66,8	73,0	70,7
Scienze agrarie e veterinarie	76,5	74,3	75,4	74,3	72,8	73,5
Ingegneria civile e Architettura	68,4	74,8	71,3	79,0	73,5	76,5
Ingegneria industriale e dell'informazione	73,5	69,7	72,5	77,6	76,5	77,3
Scienze dell'antichità, filologico-letterarie e storico-artistiche	72,5	70,8	71,4	81,0	80,6	80,8
Scienze storiche, filosofiche, pedagogiche e psicologiche	72,3	71,4	71,8	76,0	74,4	75,0
Scienze giuridiche	64,5	65,3	65,0	69,2	67,5	68,3
Scienze economiche e statistiche	70,6	72,1	71,4	76,1	73,1	74,6
Scienze politiche e sociali	70,6	66,8	68,3	76,4	76,7	76,6
Totale	70,3	70,2	70,2	75,0	75,1	75,0

La correzione per mancata risposta totale

L'indagine è di tipo totale, in quanto ha l'obiettivo di rilevare tutte le unità della popolazione oggetto di indagine; tuttavia, nel corso della fase di raccolta delle informazioni, come accade per tutte le indagini statistiche, alcune delle unità si trovano nell'impossibilità di partecipare. Di conseguenza, i dati raccolti sono affetti da un tipo di errore non campionario da mancata risposta, di cui si deve tenere conto nella fase di calcolo delle statistiche relative alle variabili di interesse. Infatti, se i non rispondenti differiscono sistematicamente dai rispondenti relativamente alle variabili di interesse per le indagini, le stime sono affette da distorsione. Poiché le variabili di interesse sui non rispondenti non sono rilevate, non è possibile valutare l'entità di questa distorsione. Tuttavia, in alcuni casi è possibile ridurre la distorsione da mancata risposta: una situazione in cui questo è possibile è quella in cui sono noti i valori di una o più variabili ausiliarie, sia sui rispondenti che sui non rispondenti. Sulla base di queste variabili è possibile suddividere la popolazione in classi e costruire una stima alternativa di una generica variabile di interesse Y .

Formalizzando, il totale della generica variabile Y , calcolato sull'intera popolazione (U), è:

$$Y = \sum_{k \in U} y_k \quad (1)$$

Poiché l'insieme dei rispondenti (S) non coincide con l'intera popolazione, si può ottenere uno stimatore corretto della generica variabile oggetto di indagine Y tramite la seguente formula:

$$\hat{Y} = \sum_{j \in S} y_j \frac{1}{\pi_j} \quad (2)$$

dove π_j è il tasso di risposta della specifica cella a cui j appartiene. Utilizzando tale metodo di stima si effettua di fatto una post-stratificazione delle unità rispondenti rispetto alle variabili note a priori su rispondenti e non rispondenti. Si assegna, in pratica, a ciascun rispondente un coefficiente di riporto all'universo (o peso finale), permettendo ai rispondenti di rappresentarsi non solo se stessi, ma anche i non rispondenti che hanno le loro stesse caratteristiche in termini di variabili ausiliarie.

La questione centrale nell'utilizzo di questo stimatore post-stratificato è l'individuazione di un'opportuna suddivisione della popolazione in classi (definite *classi di aggiustamento per mancata risposta*¹) per le quali: (a) sia ragionevole supporre che rispondenti e non rispondenti appartenenti a una stessa classe abbiano un comportamento simile relativamente alle variabili di interesse, e (b) siano presenti un numero congruo di unità rispondenti.

Nell'indagine in esame le variabili prese in considerazione per la costruzione delle classi di aggiustamento sono quelle presenti nell'archivio della popolazione dei dottori dei due anni considerati, ossia: l'anno di dottorato, la cittadinanza (italiano, straniero), l'area disciplinare, il sesso e l'ateneo. Relativamente a queste variabili sono note, sull'intera popolazione, sia le distribuzioni marginali sia le distribuzioni congiunte. Da un'analisi sui dati è emerso che, a causa dell'elevato numero di classi così definite, il numero di rispondenti fosse spesso esiguo. Il coefficiente di riporto non è quindi stato calcolato come inverso del tasso di risposta, ma si è ritenuto opportuno ricorrere a una procedura alternativa la metodologia degli *stimatori di calibrazione*, metodologia usuale nel caso di indagini campionarie². Tale procedura consiste nella risoluzione di un problema di minimo vincolato, in cui la funzione da minimizzare è una funzione di distanza (opportunamente prescelta) tra i pesi base e i pesi finali e i vincoli sono definiti dalla condizione di uguaglianza tra le stime dei totali noti di popolazione ottenute mediante i pesi finali e i corrispondenti valori noti. In una indagine campionaria, i pesi base derivano dal disegno di campionamento e sono ottenuti come inverso della probabilità di inclusione delle unità nel campione. Nel caso in esame, poiché si tratta di un'indagine censuaria, l'inclusione è certa, e i pesi base sono quindi tutti pari ad uno; in questo contesto, quindi, i pesi finali coincidono con i correttori per mancata risposta.

La metodologia può essere espressa, in termini formali, come:

$$\begin{cases} \text{Min}\{\sum_{j \in S} \text{dist}(k_j, w_j) \\ \sum_{j \in S} x_j * w_j = t \end{cases} \quad (3)$$

dove k_j è il peso base, identicamente pari ad 1 per ogni j ; w_j è il peso finale dell'unità j ; t è il vettore dei totali noti e x_j è il vettore delle variabili ausiliarie osservate sulla j -esima unità dei rispondenti.

I totali noti considerati per le variabili ausiliarie sono dati dalle distribuzioni marginali delle seguenti variabili: (i) cittadinanza (italiano, straniero), (ii) area disciplinare distintamente per sesso, (iii) ateneo, con accorpamento degli atenei con meno di 10 unità (11 atenei nel 2008 e 10 atenei nel 2010, tra quanti nei rispettivi anni hanno avuto almeno un dottore di ricerca).

La funzione di distanza utilizzata è la funzione logaritmica troncata; l'adozione di una particolare procedura iterativa ha fatto sì che fosse raggiunta non solo una correttezza formale, ma anche la correttezza sostanziale, con nessun dottore di ricerca che avesse coefficiente inferiore all'unità³.

Contatti

Per informazioni riguardanti l'indagine è possibile contattare:

Francesca Gallo, e-mail: gallo@istat.it

Emilia Arcaleni, e-mail: arcaleni@istat.it

Il documento è stato curato da Francesca Gallo (paragrafi 1 e 2), Emilia Arcaleni (paragrafi 3 e 4) e Andrea Cutillo (paragrafo 5).

¹ Sarndal C.E., Swensson B., Wretman J. (1992) *Model assisted survey sampling*, Springer Verlag, New York, Chapter 15.

² La metodologia è illustrata da Deville, J.C. e Särndal, C.E. in *Calibration Estimation in Survey Sampling*, Journal of the American Statistical Association, Vol. 87, n.418, 1992.

³ La correttezza formale è assicurata dal rispetto dei totali noti, cosa che si potrebbe raggiungere anche nel caso in cui alcuni coefficienti di riporto fossero negativi (pesi finali negativi sono comunque esclusi grazie all'utilizzo della distanza logaritmica troncata). La correttezza sostanziale è invece raggiunta facendo sì che nessun dottore di ricerca rappresenti meno di se stesso, abbia cioè un peso inferiore all'unità.