

Nota metodologica¹

Nella presente nota si intende dettagliare le caratteristiche metodologiche e gli aspetti operativi dell'intero progetto che ha trovato la sua realizzazione nella conduzione di un'indagine campionaria.

Popolazione di riferimento

La popolazione di interesse dell'indagine è costituita dagli individui che nel 2006 hanno conseguito un dottorato di ricerca in un ateneo italiano, anche se di cittadinanza non italiana e che al momento del conseguimento del titolo avevano un'età compresa tra i 25 e 49 anni. La popolazione di riferimento è ricavata dagli archivi dell'ISTAT e in particolare dal Censimento dei dottori di ricerca (Anno 2006)².

Tale popolazione è pari a 9.727 individui, di cui il 51,1% sono donne e la cui classe di età modale al 2006 comprende gli individui di 30-34 anni (il 27,2% aveva meno di 30 anni, il 53,7% aveva tra i 30 e i 34 anni e il restante 17,1% aveva 35 anni o più).

Disegno campionario

La fase di progettazione del disegno campionario ha considerato le fonti di variabilità che verosimilmente determinano l'eterogeneità dei fenomeni, in modo da massimizzare l'efficienza delle stime prodotte. In base a tale assunto il disegno ha previsto un piano di campionamento stratificato con estrazione casuale senza reimmissione dalla lista anagrafica dei dottori di ricerca fornita dall'ISTAT. Il disegno ha assicurato inoltre la pianificazione ex-ante dei domini di analisi, identificati negli strati elementari definendo la numerosità campionaria negli strati in modo da garantire un livello predeterminato di attendibilità delle stime. La pianificazione ex-ante dei domini di studio ha consentito di rappresentare in maniera appropriata anche i segmenti poco numerosi della popolazione.

Sul piano metodologico tale attività si è servita delle opportune tecniche di allocazione negli strati di un campione di numerosità fissata, pari a 7.000 unità³.

¹ Marco Centra e Valentina Gualtieri

² Il Censimento dei dottori di ricerca è un'indagine ISTAT compresa nel Piano Statistico Nazionale (IST- 01585). L'anagrafe delle persone che hanno conseguito il dottorato di ricerca nel 2006 è stata fornita dall'ISTAT all'ISFOL a seguito di una convenzione stipulata tra i due enti di ricerca.

³ La numerosità effettiva del campione è risultata inferiore alla numerosità teorica dopo i controlli di coerenza e di qualità delle risposte ed è pari a 4.879 interviste.

La procedura allocativa adottata è di tipo complesso e combina l'allocazione proporzionale e quella con uguale errore standard nei domini di studio, fissato un determinato livello di attendibilità. In tal modo l'allocazione del campione ha consentito di produrre stime in diversi domini con errore campionario non dissimile, con il vantaggio di permettere confronti attendibili. Allo stesso tempo la procedura di allocazione ha consentito di controllare l'effetto del disegno del campione non proporzionale⁴,

La stratificazione ha considerato variabili territoriali (area geografica di residenza dell'individuo nel 2006), demografiche (sesso ed età) e caratterizzanti il dottorato di ricerca (area disciplinare del dottorato di ricerca) nell'ipotesi che tali elementi consentissero di recuperare parte della variabilità dei caratteri oggetto di interesse.

Di seguito sono elencate le variabili e le relative modalità utilizzate nella stratificazione.

1) Area geografica di residenza del dottore nel 2006

- Nord-ovest
- Nord-est
- Centro
- Sud
- Isole
- Mancante

2) Sesso

- Maschio
- Femmina

3) Classe di età in anni compiuti nel 2006

- 25-29 anni
- 30-34 anni
- 35 anni o più
- Mancante

4) Gruppo disciplinare del dottorato di ricerca

- Scienze matematiche e informatiche
- Scienze fisiche
- Scienze chimiche
- Scienze della terra
- Scienze biologiche
- Scienze mediche
- Scienze agrarie e veterinarie
- Ingegneria civile e Architettura
- Ingegneria industriale e dell'informazione
- Scienze dell'antichità, filologico-letterarie e storico-artistiche
- Scienze storiche, filosofiche, pedagogiche e psicologiche
- Scienze giuridiche
- Scienze economiche e statistiche
- Scienze politiche e sociali

⁴ La fase di allocazione si servirà della procedura contenuta in Centra, Falorsi (2007), pp. 24-32.

In termini numerici la procedura allocativa è stata effettuata su 518 strati.

La scelta del contatto telefonico o via web come tecnica di rilevazione da utilizzare per raccogliere le informazioni sul campo ha creato un parziale disallineamento tra la popolazione obiettivo teorica e la popolazione effettiva da cui è tratto il campione di rispondenti. Nonostante, infatti, sia stata effettuata una laboriosa fase di recupero dei contatti telefonici e degli indirizzi e-mail di tutti i dottori appartenenti alla popolazione di riferimento, molti dei quali già presenti sugli archivi forniti dall'ISTAT, alcuni nominativi sono stati scartati dalla popolazione fattuale. L'entità di questo scarto, da cui deriva un errore non campionario, è trascurabile per le stime di totali e di indicatori a livello di aggregazione utilizzato per le analisi condotte successivamente sui dati. Ciò nonostante la particolare metodologia utilizzata nella fase di stima basata su tecniche di calibrazione, come esplicitato più avanti, ha consentito di recuperare l'eventuale parte della distorsione indotta da specifici effetti di selezione del campione dei rispondenti.

Conduzione dell'indagine

La rilevazione è stata condotta con tecnica CATI e si avvalsa di intervista web solo nel caso in cui è stata constatata un'oggettiva impossibilità a svolgere l'intervista telefonica⁵. Le modalità di rilevazione per questa particolare popolazione di cui non si conosceva a priori il luogo di residenza nel momento dell'intervista, sono state definite con l'obiettivo di minimizzare il numero di mancate risposte totali al fine di evitare fenomeni di distorsione delle stime dovute ad autoselezione del campione dei rispondenti.

A tal fine il contatto iniziale con l'individuo da intervistare è stata curata in modo particolarmente scrupoloso e sono state messe in atto differenti procedure volte a evitare la "caduta" di nominativi.

Il primo *step* della fase di contatto ha previsto l'invio di una lettera ufficiale da parte dell'ISFOL a firma del Presidente, che aveva il principale obiettivo di sensibilizzare la popolazione oggetto d'indagine. Nella lettera erano inoltre illustrati i contenuti della rilevazione e veniva esplicitata l'importanza del processo conoscitivo che l'indagine permette di realizzare.

Successivamente all'invio della lettera, e contestualmente all'avvio della rilevazione è stata attuata una scrupolosa fase di "inseguimento" degli individui. In altri termini si è cercato, tramite l'aiuto dei familiari dell'intervistato, di inseguire e intervistare l'individuo anche e soprattutto nel caso di un suo eventuale trasferimento con annesso cambio di residenza.

Come ausilio a tale fase è stata predisposta una scheda contatto (vedi allegato "Scheda contatti") tramite la quale sono state registrate tutte le informazioni inerenti l'individuo e i suoi eventuali spostamenti nel territorio italiano ed estero. Le informazioni reperite nella scheda sono state utilizzate in fase di stima per correggere la distorsione indotta da mancata risposta totale.

Durante la fase di campo, a causa dell'impossibilità di intervistare tutti gli individui facenti parte del campione estratto si è fatto ricorso a un campione suppletivo vincolato dal sistema

⁵ In dettaglio sono state effettuate 4.103 interviste CATI e 776 interviste CAWI.

di regole per definire la caduta di un contatto (rifiuto esplicito, contatto irreperibile, ecc.). Per ciascuna unità del campione base e del campione suppletivo è stata tenuta traccia dell'esito del contatto, in modo da controllare il processo di autoselezione del campione e ridurre la distorsione delle stime prodotte.

La verifica degli strumenti di rilevazione (tecnica di comunicazione, sistema CATI, questionario) è stata effettuata tramite un'apposita indagine pilota condotta un mese prima l'avvio della fase estensiva su un piccolo campione selezionato di dottori di ricerca non appartenenti alla popolazione d'indagine (sono stati utilizzati alcuni nominativi di dottori di ricerca del 2004).

La fase di somministrazione del questionario ha previsto inoltre una serie di procedure e strumenti finalizzati a garantire la qualità dei dati raccolti, tra i quali la selezione e formazione degli intervistatori e il controllo in itinere della fase di campo tramite l'adozione di predefiniti indicatori di qualità.

Questionario

Attraverso il questionario di rilevazione sono state raccolte informazioni di carattere socio anagrafico, relative ai percorsi di istruzione, di mobilità geografica, nonché indicazioni sulla situazione occupazionale e sugli aspetti legati alle motivazioni e alla soddisfazione.

Nello specifico, quindi, il questionario si compone di differenti sezioni (Schema 1).

La prima sezione è dedicata alla STORIA ACCADEMICA dell'intervistato con riferimento alle caratteristiche del percorso di studi e, in particolare, alle esperienze all'estero nel periodo di formazione.

Nella seconda parte si rilevano informazioni sulla attuale SITUAZIONE OCCUPAZIONALE, indagando sulle modalità di ricerca di lavoro, sulla tipologia di impiego e sui contenuti e compiti del lavoro, dando particolare rilievo a quesiti sull'attinenza del lavoro con il dottorato di ricerca conseguito.

La terza sezione, considerata centrale nella rilevazione, si concentra sulla MOBILITÀ GEOGRAFICA dei dottori di ricerca ed è orientata a comprendere le motivazioni e i fattori sottostanti la scelta di mobilità o di non mobilità, ma anche le aspettative per il futuro e le intenzioni sulle opzioni circa la mobilità futura. Per mobilità territoriale non si intendono solo ed esclusivamente gli spostamenti internazionali ma anche i cambiamenti di domicilio all'interno dell'Italia, ovvero da una regione a un'altra.

La quarta sezione serve a ricostruire la SITUAZIONE OCCUPAZIONALE PASSATA, riferendosi al periodo successivo al conseguimento del dottorato di ricerca. In questa parte dell'intervista si vuole dar conto di tutte le esperienze lavorative che hanno comportato uno spostamento territoriale (nazionale o regionale) di una durata superiore a sei mesi.

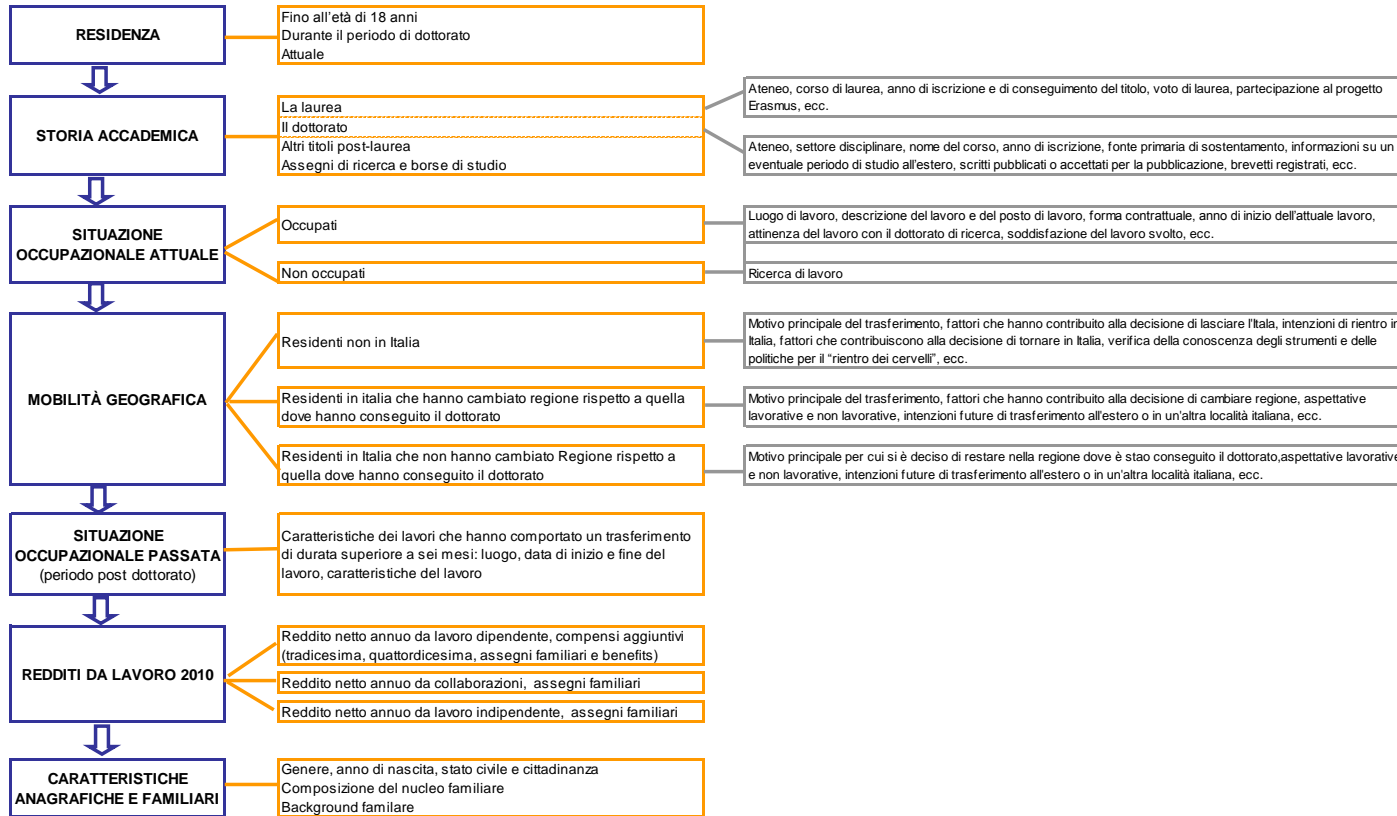
La penultima sezione del lavoro è dedicata alle RETRIBUZIONI NEL 2010 e permetterà di verificare i differenziali retributivi rispetto alle diverse localizzazioni territoriali e al livello e al profilo professionale.

L'ultima parte del questionario intende raccogliere informazioni sulle CARATTERISTICHE ANAGRAFICHE E FAMILIARI dell'intervistato e in particolare cattura informazioni sull'origine sociale del dottore (misurata tramite il titolo di studio, la condizione occupazionale e la

professione dei genitori) e sulla famiglia attuale (numero di componenti, tipologia familiare e informazioni sul partner).

In allegato alla presente nota è comunque possibile visionare l'intero questionario utilizzato per la rilevazione.

Schema 1: La struttura del questionario dell' "Indagine ISFOL sulla mobilità geografica dei dottori di ricerca"



Fase di stima

La fase di stima è un elemento particolarmente delicato nell'intero processo conoscitivo basato su procedure inferenziali. In questa fase è possibile correggere eventuali fattori di distorsione indotti da selezione del campione e da mancata risposta totale. Alcune tecniche sono particolarmente utili nel garantire un'elevata efficienza delle stime e, in ultima analisi, una rappresentazione accurata dei fenomeni di interesse.

La tecnica utilizzata per la costruzione dello stimatore dell'indagine sulla mobilità geografica dei dottori di ricerca si fonda sull'approccio predittivo e prevede la messa a punto di uno stimatore basato sulla teoria degli stimatori parzialmente assistiti da modello, sulla base degli stimatori di regressione. È stato seguito perciò un approccio riconducibile a modelli di superpopolazione (Dorfman et al., 2000), che contestualmente permette la messa a punto di stimatori calibrati (Deville e Särndal, 1992) basati su una serie di informazioni ausiliarie disponibili sulla popolazione di riferimento dell'indagine. Oltre a sfruttare le informazioni delle variabili ausiliarie riducendo la varianza campionaria, tale classe di stimatori gode di una serie di proprietà tra le quali quella della calibrazione, secondo la quale le stime dei totali delle variabili ausiliarie utilizzate come regressori, corrispondono ai totali noti sulla popolazione. In tal modo è possibile calibrare la popolazione stimata rispetto ai totali noti ricavati dalla popolazione di riferimento, disaggregati secondo specifiche caratteristiche.

Una stima del totale di una variabile Y è data dalla seguente espressione:

$$\tilde{Y} = \sum_{l \in S} y_l w_l$$

dove con riferimento al campione S si indica con y_l il valore della variabile Y osservato nella l -ma unità campionaria e con w_l il peso di riporto all'universo associato alla l -ma unità campionaria. La metodologia di calcolo dei pesi di riporto all'universo w_l secondo la tecnica della calibrazione prevede in primo luogo il calcolo del peso diretto d_l , sulla base del disegno campionario, definito come l'inverso della probabilità di inclusione della l -ma unità campionaria:

$$d_l = 1/\pi_l;$$

successivamente prevede il calcolo di un fattore correttivo ω ottenuto vincolando a totali noti le stime dirette.

$$\tilde{Y} = \sum_{l \in S} y_l w_l = \sum_{l \in S} y_l d_l \omega_l$$

Il coefficiente di riporto all'universo w_l si ottiene moltiplicando il peso diretto per il fattore correttivo: $w_l = d_l \omega_l$

I fattori correttivi ω_l si ottengono risolvendo il seguente problema di minimo vincolato, in cui si minimizza una funzione di distanza tra i pesi diretti d_l e i pesi finali w_l ; i vincoli sono definiti dalla condizione di uguaglianza tra le stime campionarie dei totali di alcune variabili ausiliarie e i totali noti degli stessi:

$$\begin{cases} \min \left\{ \sum_{l \in S} \text{dist}(d_l, w_l) \right\} \\ \sum_{l \in S} y_l w_l = \mathbf{t} \end{cases}$$

dove \mathbf{t} è il vettore dei totali noti e \mathbf{x}_l è il vettore delle variabili ausiliarie osservate nella l -ma unità campionaria (ISTAT, 2006). La soluzione del problema è data dalla seguente espressione che definisce il fattore correttivo ω_l per ciascuna unità del campione (Deville e Särndal, 1992).

$$\omega_l = 1 + \left(\mathbf{t} - \sum_{l \in S} \mathbf{x}_l w_l \right)' \left(\sum_{l \in S} \mathbf{x}_l \mathbf{x}_l' w_l \right)^{-1} \mathbf{x}_l$$

Gli aggregati di riferimento, utilizzati come totali noti dalla procedura di calibrazione, sono stati ricavati dal Censimento dei dottori di ricerca dell'ISTAT. Il ricorso all'approccio predittivo e l'utilizzo di uno stimatore calibrato hanno permesso inoltre di correggere eventuali distorsioni delle stime derivate da autoselezione del campione, ottimizzando così l'informazione ausiliaria sulla popolazione.

Le informazioni ausiliarie utilizzate nella costruzione dello stimatore calibrato sono riportate nello schema seguente:

Schema 2: Informazione ausiliarie per la procedura di calibrazione

Descrizione	Modalità
Area geografica di residenza del dottore nel 2006	Nord ovest
	Nord est
	Centro
	Sud
	Isole
Sesso	Uomo
	Donna
Età in classi in anni compiuti al 2006	25-29 anni
	30-34 anni
	35 anni o più
Gruppo disciplinare del dottorato di ricerca	Scienze matematiche e informatiche
	Scienze fisiche
	Scienze chimiche
	Scienze della terra
	Scienze biologiche
	Scienze mediche
	Scienze agrarie e veterinarie
	Ingegneria civile e Architettura
	Ingegneria industriale e dell'informazione
	Scienze dell'antichità, filologico-letterarie e storico-artistiche
	Scienze storiche, filosofiche, pedagogiche e psicologiche
	Scienze giuridiche
	Scienze economiche e statistiche
	Scienze politiche e sociali

Le variabili di calibrazione sono state utilizzate seconda la seguente nidificazione:

Schema 3 : Nidificazione per la procedura di calibrazione

		N. vincoli ⁶
Area geografica	Disciplina	85
Età in classi	Disciplina	42
Disciplina	Sesso	28
Sesso	Età in classi	6
Area geografica	Sesso	10
Area geografica	Età in classi	15

Lo stimatore così ottenuto, applicato come coefficiente moltiplicativo delle unità campionarie, ha permesso di produrre stime sulla popolazione di riferimento in modo che gli aggregati riferiti a ciascuna nidificazione riportata nel piano di calibrazione, coincidessero con i corrispondenti totali noti ricavati dai dati ISTAT del Censimento dei dottori di ricerca del 2006.

⁶ Dalla procedura di calibrazione sono tolti i vincoli ridottanti.