

Indagine sui servizi sociali erogati dal *non profit* 2017

Nota metodologica

1. Premessa

La presente nota illustra la strategia campionaria adottata per lo svolgimento delle indagini sui servizi sociali erogati dal *non profit* condotta dall'Inapp nel corso del 2017. Nella nota sono descritti anche gli aspetti operativi che hanno caratterizzato il processo di produzione dell'informazione statistica e le procedure adottate nella realizzazione della fase di campo. In particolare, nei paragrafi seguenti, dopo una breve descrizione delle popolazioni di riferimento dell'indagine, viene descritto il piano di campionamento adottato, sono illustrati i principali aspetti che hanno caratterizzato la fase di implementazione dell'indagine e sono esplicitate le procedure adottate in fase di stima.

2. Popolazione di riferimento

La popolazione di riferimento dell'indagine è costituita dalle istituzioni *non profit* attive e appartenenti ai settori di attività economica "Sanità" e "Servizi Sociali", ovvero i codici 86, 87 e 88 della classificazione ATECO2007 che corrispondono ai codici dall'8 al 14 della classificazione ICNPO a 29 classi.

Sfruttando le informazioni ricavate dal censimento sulle istituzioni *non profit* realizzato dall'Istat nel 2011, si ha che tale popolazione ammonta a 35.830 istituzioni *non profit*.

Nella seguente tabella è riportata la distribuzione della popolazione di riferimento secondo le principali caratteristiche delle istituzioni *non profit*; tali caratteristiche sono utilizzate in fase di disegno quali domini pianificati e variabili di stratificazione.

Tabella 1 - Istituzioni non profit attive ed appartenenti ai settori di attività economica "Sanità" e "Servizi Sociali" secondo la collocazione territoriale, la forma giuridica, il settore e la presenza di lavoratori retribuiti. Anno 2011

		Valori assoluti	Percentuale
<i>Regione</i>	Piemonte	3086	8.6
	Valle d'Aosta	116	0.3
	Lombardia	6484	18.1
	Trentino-Alto Adige	1285	3.6
	Veneto	3058	8.5
	Friuli-Venezia Giulia	831	2.3
	Liguria	1135	3.2
	Emilia-Romagna	2663	7.4
	Toscana	2653	7.4
	Umbria	558	1.6
	Marche	955	2.7
	Lazio	2784	7.8
	Abruzzo	682	1.9
	Molise	238	0.7
	Campania	1963	5.5
	Puglia	1856	5.2
	Basilicata	429	1.2
	Calabria	1033	2.9
	Sicilia	2732	7.6
Sardegna	1289	3.6	
<i>Area geografica</i>	Nord-ovest	10821	30.2
	Nord-est	7837	21.9
	Centro	6950	19.4
	Mezzogiorno	10222	28.5
<i>Forma Giuridica</i>	Associazione riconosciuta	10134	28.3
	Associazione non riconosciuta	16544	46.2
	Cooperativa sociale	5606	15.7
	Altro	3546	9.9
<i>Settore di attività prevalente ICNPO</i>	Sanità	10934	30.5
	Assistenza sociale e protezione civile	24896	69.5
<i>Presenza di lavoratori retribuiti</i>	Sì	13317	37.2
	No	22513	62.8

La lista delle istituzioni *non profit* da intervistare, corredata delle informazioni necessarie alla realizzazione del disegno campionario e alla fase di contatto, è stata derivata dalla banca dati fornita dall'Istat contenente gli esiti del censimento delle istituzioni *non profit* del 2011.

3. Piano di campionamento

Al fine di massimizzare l'efficienza delle stime prodotte, nella fase di pianificazione del disegno campionario sono state considerate le possibili fonti di variabilità che possono determinare l'eterogeneità dei fenomeni oggetto di studio. Il disegno campionario scelto ha previsto l'estrazione di un campione selezionato attraverso un campionamento stratificato; è stato adottato un piano di campionamenti di tipo probabilistico con estrazione casuale senza reimmissione delle unità campionarie dalla lista delle istituzioni *non profit* derivante dal censimento condotto dall'Istat nel 2011. Grazie alle informazioni già rilevate durante il censimento è stato inoltre possibile adottare un piano di campionamento con pianificazione *ex ante* dei domini di studio (Bethel, 1989). In altri termini, utilizzando un'allocazione multidominio con domini pianificati, il numero di unità da campionare in ciascuno strato è stato calcolato in modo da assicurare un livello predeterminato di attendibilità delle stime riferite a particolari domini di interesse, ottenuti come aggregazione di strati elementari (Choudhry et alii, 2012). Una volta definito lo schema di stratificazione, l'allocazione multidominio con domini pianificati è una soluzione alternativa all'allocazione proporzionale, vale a dire all'allocazione che prevede per ciascuno strato una quota del campione pari alla medesima quota della popolazione nello stesso strato. L'allocazione proporzionale, pur presentando un vantaggio rispetto alle altre tipologie di allocazione poiché garantisce un livello di efficienza complessiva del disegno pari al disegno casuale semplice, può dimostrarsi, tuttavia, poco adeguata quando si è interessati a produrre stime riferite a domini poco numerosi nella popolazione, dove il livello di precisione delle stime risulta spesso insufficiente a causa della bassa quota di campione attribuita. Una scelta diversa prevede di allocare il campione negli strati in funzione del livello minimo di precisione fissata su particolari domini di interesse.

Nello schema seguente è riportato l'elenco delle variabili di stratificazione e dei domini pianificati, nonché i vincoli posti nella procedura allocativa.

Schema 1- Principali caratteristiche dei piani di campionamento

Variabili di Stratificazione	Domini Pianificati	Vincoli procedura allocativa
<ul style="list-style-type: none"> • Regione (20 modalità) • Settore di attività (Sanità, Assistenza sociale e protezione civile) • Forma giuridica (Associazione riconosciuta, Associazione non riconosciuta, Cooperativa sociale, Altro) • Presenza di lavoratori retribuiti (Sì, No) 	<ul style="list-style-type: none"> • Concatenazione dell'area geografica con il settore di attività • Concatenazione della forma giuridica con il settore di attività • Concatenazione della presenza di lavoratori retribuiti con il settore di attività • Regione 	<p>La procedura allocativa ha previsto per stime de 10% un CV massimo:</p> <ul style="list-style-type: none"> • del 15% sul dominio regionale • del 12% sul dominio concatenato forma giuridica/settore di attività • del 10% sui restanti domini concatenati.

Da tale allocazione è derivata una dimensione campionaria totale pari a 7.250 unità. La probabilità di inclusione delle unità nel campione è stata fissata uguale all'interno degli strati (estrazione con probabilità uguale).

4. Aspetti operativi e procedure adottate per la fase di campo

Per limitare i costi del progetto e il disturbo statistico sulle istituzioni indagate, soprattutto in ragione della lunghezza e della complessità del questionario somministrato, l'indagine è stata condotta *in house* con tecnica CAWI (Computer-assisted web interviewing).

L'applicativo informatico utilizzato è la *Lime Survey*^A, un *tool* completamente *open source e freeware*. *Lime Survey* è uno strumento molto flessibile che permette di suddividere le sezioni del questionario in gruppi di domande distinti, di impostare dei filtri costruiti su una logica condizionale parametrabile a seconda delle risposte date in precedenza nel questionario. Altro vantaggio di *Lime Survey*, oltre alla possibilità di svolgere la rilevazione senza il ricorso ad esternalizzazioni è che permette, in un unico applicativo, di gestire in maniera verticale tutte le fasi dell'indagine a partire dal caricamento delle liste dei partecipanti, passando per la gestione degli identificativi e dei moduli mail di invito all'indagine e del sollecito, per finire a una sezione *built in* di statistiche estraibili dai risultati (non manca la possibilità di esportare integralmente i risultati dell'indagine nei più comuni formati).

Al fine di raggiungere il più possibile la numerosità campionaria definita in fase di allocazione del campione, limitando al contempo la distorsione da errore non campionario potenzialmente causata dal un numero eccessivo di sostituzioni, al campione base sono stati affiancati due campioni suppletivi estratti con pari disegno campionario e numerosità del campione base (tabella 2).

Tabella 2. -Tipo di campioni e rispondenti. Valori assoluti e %

Tipo di campione	n. rispondenti	% di risposta
Campione base	2.496	34
I suppletivo	2.295	33
II suppletivo	2.136	35

Fonte: dati Inapp 2018

La fase di campo è stata gestita attraverso un primo contatto nel quale è stata inviata una lettera di presentazione del progetto e di invito a partecipare alla rilevazione. Successivamente si è proceduto all'invio del link per la compilazione del questionario *on-line* e ad una fase di monitoraggio dell'andamento dell'indagine accompagnata da momenti di sollecito alla compilazione.

La rilevazione dei dati è partita il 27 febbraio 2017 per concludersi il 7 luglio 2017. **I dati si riferiscono alla situazione esistente nelle organizzazioni al 31 dicembre del 2016.**

5. Esito dell'indagine: il campione effettivo

Il campione teorico di istituzioni cui sottoporre il questionario d'indagine ha subito, durante la fase di campo, delle cadute. Tali cadute sono avvenute sia in ragione della obsolescenza delle informazioni contenute nel sampling frame (che riportava informazioni relative al 2011), sia per rifiuto esplicito all'intervista. Inoltre, a esito della compilazione del questionario alcune interviste sono state scartate a causa dell'incompletezza delle informazioni rilasciate.

Il campione effettivo ammonta a 6.927 interviste.

La distribuzione del campione effettivo secondo le caratteristiche delle istituzioni utilizzate in fase di disegno campionario, è riportata nella tabella 2. La figura 2, invece, mostra la differenza nelle composizioni percentuali tra il campione effettivo e la popolazione di riferimento.

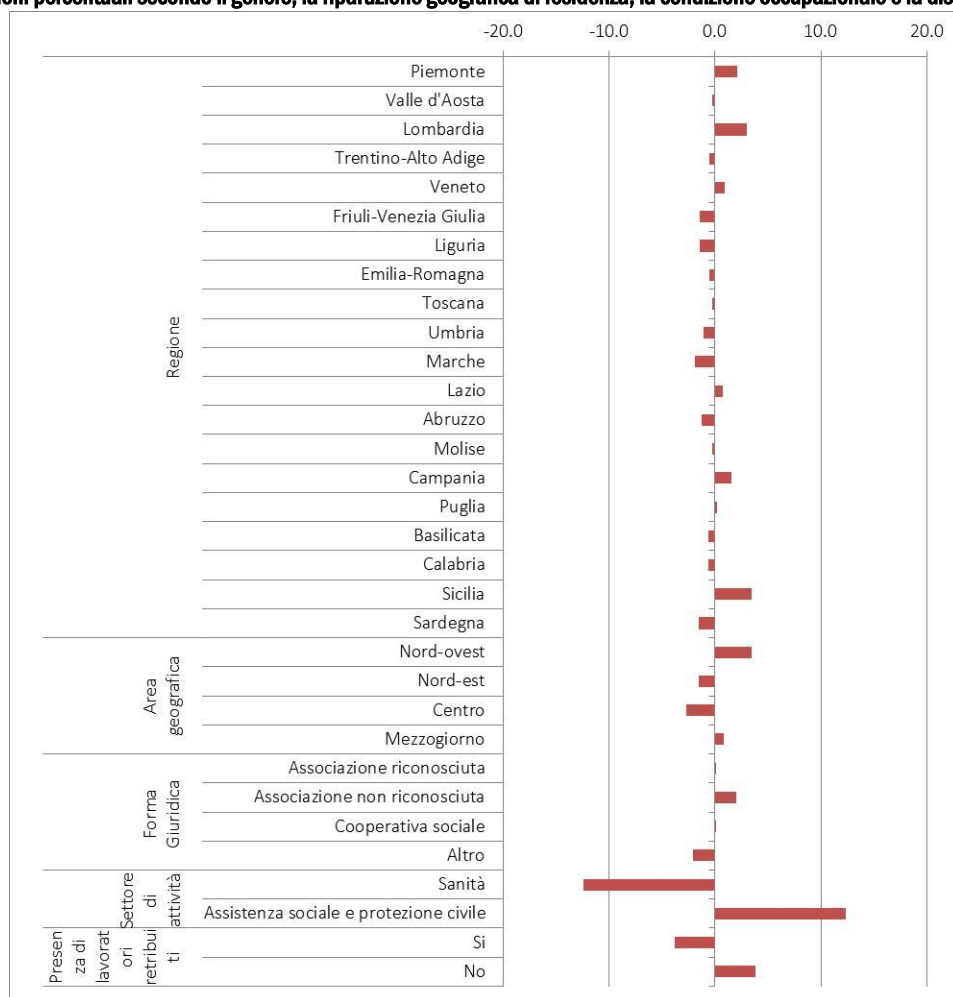
Tabella 3. - Campione effettivo secondo la collocazione territoriale, la forma giuridica, il settore e la presenza di lavoratori retribuiti

		Valori assoluti	Percentuale
<i>Regione</i>	Piemonte	449	6.5
	Valle d' Aosta	39	0.6
	Lombardia	1043	15.1
	Trentino-Alto Adige	286	4.1
	Veneto	528	7.6
	Friuli-Venezia Giulia	259	3.7

¹ L'indagine è stata condotta avvalendosi di Lime Survey, uno strumento open source basato sull'acquisizione dei dati mediante l'accesso a pagine web. In sintesi, il questionario è stato informatizzato secondo quanto richiesto da Lime Survey prevedendo schermate successive. I rispondenti hanno potuto così avvalersi di una modalità di compilazione del questionario interamente on-line, beneficiando dell'opportunità di inserire, dalla propria postazione di lavoro (o mediante qualsiasi connessione internet), le proprie risposte e di poter essere guidati nel corso dell'intera compilazione. Per consentire la massima flessibilità di compilazione questa poteva essere interrotta e ripresa in qualsiasi momento senza perdere le informazioni già inserite. Infine, il gruppo di lavoro, si è dotato di un sistema di archiviazione delle risposte mediante trasferimento dei dati dalla rete ad un server ISFOL dedicato. Per saperne di più: www.limesurvey.org.

	Liguria	324	4.7
	Emilia-Romagna	549	7.9
	Toscana	533	7.7
	Umbria	188	2.7
	Marche	318	4.6
	Lazio	491	7.1
	Abruzzo	216	3.1
	Molise	65	0.9
	Campania	271	3.9
	Puglia	348	5.0
	Basilicata	129	1.9
	Calabria	244	3.5
	Sicilia	290	4.2
	Sardegna	357	5.2
<i>Area geografica</i>	Nord-ovest	1855	26.8
	Nord-est	1622	23.4
	Centro	1530	22.1
	Mezzogiorno	1920	27.7
<i>Forma Giuridica</i>	Associazione riconosciuta	1954	28.2
	Associazione non riconosciuta	3059	44.2
	Cooperativa sociale	1082	15.6
	Altro	832	12.0
<i>Settore di attività prevalente ICNPO</i>	Sanità	2971	42.9
	Assistenza sociale e protezione civile	3956	57.1
<i>Presenza di lavoratori retribuiti</i>	Si	2840	41.0
	No	4087	59.0
Totale		6927	100.0

Figura 1. - Campione effettivo Vs Popolazione di riferimento (persone con meno di 35 anni al 2016, con laurea triennale): differenza nelle composizioni percentuali secondo il genere, la ripartizione geografica di residenza, la condizione occupazionale e la disciplina universitaria



Al fine di correggere parte dell'autoselezione del campione effettivo rispetto a quello teorico e di poter sviluppare analisi di tipo inferenziali è stata implementata una procedura di stima che ha permesso di calcolare i pesi di riporto all'universo a partire dalla probabilità di inclusione attribuita a ciascuna unità del campione nella fase di disegno.

6. La fase di stima

Tramite la fase di stima è possibile correggere eventuali fattori di distorsione dovuto al disallineamento tra campione teorico e campione effettivo e più in generale al fenomeno della mancata risposta totale.

Il principio su cui è basato ogni metodo di stima campionaria è che le unità appartenenti al campione rappresentino anche le unità della popolazione che non sono incluse nel campione. Questo principio viene realizzato attribuendo ad ogni unità campionaria un peso che indica il numero di unità della popolazione rappresentate dall'unità medesima. Per fare ciò si individuano funzioni dei dati campionari tramite le quali si producono stime relative alla popolazione, ossia che permettono di fare inferenza.

Le principali indagini svolte nell'ambito della statistica ufficiale impiegano stimatori che utilizzano variabili ausiliarie (stimatori indiretti) altamente correlate con le variabili di interesse per migliorare l'accuratezza, la correttezza e la consistenza delle stime. Tra gli stimatori che sfruttano le variabili ausiliarie i più noti sono lo stimatore di regressione generalizzata² e lo stimatore calibrato (o di ponderazione vincolata). Tali stimatori, oltre a consentire la correzione degli errori non campionari, permettono anche di aumentare la precisione delle stime riduzione dell'errore campionario.

² Lo stimatore di regressione generalizzata è un caso particolare dello stimatore calibrato e coincide con quest'ultimo quando la distanza utilizzata è quella del χ^2 . Lo stimatore calibrato tende asintoticamente allo stimatore di regressione generalizzata.

Lo stimatore scelto per l'indagine in oggetto, è lo stimatore calibrato, messo a punto da Deville e Särndal (Deville e Särndal, 1992) e comunemente usato nelle indagini condotte da Istat e Inapp³.

Gli aggregati di riferimento, utilizzati come totali noti dalla procedura di calibrazione dell'indagine sui servizi sociali, sono ricavati sempre dal censimento delle istituzioni *non profit* condotto dall'Istat nel 2011. Nello specifico, tali aggregati prendono in considerazione le variabili utilizzate nella fase di disegno del campione e nella determinazione dei domini di studio, ma con una più fitta nidificazione delle stesse.

Nei due schemi seguenti sono presentati l'elenco delle variabili ausiliare (e delle relative modalità) adottate nelle procedure di calibrazione, nonché il piano di nidificazione di tali informazioni, ossia il modo in cui le variabili ausiliarie sono state concatenate.

Schema 2. - Informazione ausiliarie adottate nella procedura di calibrazione

Descrizione delle variabili	Modalità
Regione	20 regioni
Area geografica	Nord-ovest Nord-est Centro Mezzogiorno
Forma Giuridica	Associazione riconosciuta Associazione non riconosciuta Cooperativa sociale Altro
Settore di attività prevalente	Sanità Assistenza sociale e protezione civile
Presenza lavoratori retribuiti	Si No

Schema 3. - Nidificazione per la procedura di calibrazione

				N. vincoli
Area geografica	Forma Giuridica	Settore di attività prevalente	Presenza lavoratori retribuiti	64
Regione	Forma Giuridica			80

³ La tecnica di calibrazione si sviluppa in termini di "correzione": le variabili ausiliarie vengono calibrate con i totali noti al fine di correggere i pesi base assegnati alle unità dallo stimatore di Horvitz Thompson in base al disegno campionario prescelto. La tecnica di calibrazione si fonda sull'approccio predittivo (Dorfman et al., 2000) e prevede la messa a punto di uno stimatore basato sulla teoria degli stimatori parzialmente assistiti da modello sulla base degli stimatori di regressione.

Considerato un generico stimatore del totale: $\hat{y} = \sum_{i=1}^n y_i w_i$

La metodologia di calcolo dello stimatore calibrato prevede:

1) Il calcolo del peso base/diretto d_i secondo il disegno di campionamento, ottenuto come reciproco della probabilità di inclusione di ogni unità campionaria: $d_i = 1/\pi_i$

2) Il calcolo di un fattore correttivo c_i , che consente di soddisfare la condizione di uguaglianza tra i totali noti della popolazione e le corrispondenti stime campionarie dei totali delle variabili ausiliarie.

c_i è ottenuto risolvendo un problema di minimo vincolato in cui si minimizza una funzione di distanza tra il peso diretto e il peso finale e i vincoli sono definiti dalla condizione che le stime campionarie dei totali di popolazione sopra definiti coincidano con i valori noti degli stessi.

$$\begin{cases} \min \sum_{i \in S} \text{dist}(d_i, w_i) \\ \sum_{i \in S} x_i w_i = T \end{cases}$$

Dove T è il vettore dei totali noti.

La soluzione del problema è data dalla seguente espressione che definisce il fattore correttivo $c_i \forall i$

$$c_i = 1 + \left[T - \sum_{i \in S} x_i d_i \right] \left[\sum_{i \in S} x_i x_i' d_i \right]^{-1} x_i$$

3) Si calcolano i pesi finali: $w_i = d_i * c_i$.

Regione	Settore di attività prevalente			40
Regione	Presenza di lavoratori retribuiti			40

Il ricorso all'approccio predittivo e l'utilizzo di uno stimatore calibrato permette dunque di correggere eventuali distorsioni delle stime derivate da autoselezione dei campioni, nella misura in cui l'autoselezione è correlata con le variabili ausiliare utilizzate nella procedura. Lo stimatore calibrato così ottenuto consente di recuperare, oltre alla distorsione diretta, anche la quota di distorsione delle stime relative a fenomeni di interesse di cui non si conosce la distribuzione nella popolazione ma che si ipotizzano correlati alle variabili utilizzate nella definizione dei vincoli di calibrazione.

I pesi ottenuti dalla procedura di calibrazione, applicati come coefficienti moltiplicativi delle unità campionarie, consentono 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 dal censimento delle istituzioni *non profit* condotto dall'Istat nel 2011.