

**NOTA
METODOLOGICA**

*IV Indagine sui servizi sociali erogati dal
non profit*

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 2021. 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 "Istruzione e ricerca – ateco 72 e 85", "Sanità – ateco 86 e parte di 87", "Assistenza Sociale – ateco 87 e 88" e limitatamente alle cooperative sociali "Sviluppo economico e coesione sociale – ateco vari" i codici 2, 3, 4 e parte del 6 della classificazione ICNPO a X classi.

La popolazione, secondo il Registro Statistico delle Istituzioni non Profit 2017 dell'Istat ammonta 63.898 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 "Istruzione" "Sanità", "Assistenza Sociale" e coop. Sociali del settore "sviluppo economico e coesione sociale" secondo la collocazione territoriale, la forma giuridica, il settore e la presenza di lavoratori retribuiti. Anno 2017

		<i>Valori assoluti</i>	<i>Percentuale</i>
<i>Regione</i>	Piemonte	5.259	8,2
	Valle d'Aosta	192	0,3
	Lombardia	11.309	17,7
	Trentino-Alto Adige	1.831	2,9
	Veneto	5.329	8,3
	Friuli-Venezia Giulia	1.347	2,1
	Liguria	1.762	2,8
	Emilia-Romagna	4.291	6,7
	Toscana	4.501	7,0
	Umbria	997	1,6
	Marche	1.767	2,8
	Lazio	6.391	10,0
	Abruzzo	1.280	2,0
	Molise	466	0,7
	Campania	4.524	7,1
	Puglia	3.165	5,0
	Basilicata	839	1,3
	Calabria	1.923	3,0
	Sicilia	4.712	7,4
	<i>Area geografica</i>	Sardegna	2.013
Nord-ovest		18.522	29,0
Nord-est		12.798	20,0
Centro		13.656	21,4
<i>Forma Giuridica</i>	Mezzogiorno	18.922	29,6
	Associazione	39.569	61,9
	Cooperative sociale	15.166	23,7
	Fondazione	4.233	6,6
	Altro	4.930	7,7
<i>Settore di attività prevalente ICNPO</i>	Istruzione e ricerca	13.915	21,8
	Sanità	12.235	19,1
	Assistenza sociale	32.245	50,5
	Svil. Econ. e coesione soc (solo coop. Sociali)	5.503	8,6
<i>Presenza di lavoratori retribuiti</i>	Si	29.389	46,0
	No	34.509	54,0
	Totale	63.898	100,0

Fonte: elaborazioni su dati Istat – Registro Statistico delle Istituzioni non Profit 2017

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 dati del Registro Statistico delle Istituzioni non Profit 2017.

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 Registro Statistico delle Istituzioni non profit dell'Istat, 2017. 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 al, 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

<ul style="list-style-type: none"> • Regione (20 modalità) • Settore di attività (Istruzione, Sanità, Assistenza sociale) • Forma giuridica (Associazione, Cooperativa sociale, Fondazione, Altro) • Presenza di lavoratori retribuiti (Si, No) 	
Domini Pianificati	Vincoli procedura allocative (per stime del 10%)
Concatenazione dell'area geografica con il settore di attività	c.v. massimo 15%
Concatenazione della forma giuridica con il settore di attività	c.v. massimo 10%
Concatenazione della presenza di lavoratori retribuiti con il settore di attività	c.v. massimo 10%
Regione	c.v. massimo 20%

Da tale allocazione è derivata una dimensione campionaria totale pari a 7.364 unità. La probabilità di inclusione delle unità nel campione è stata fissata uguale all'interno degli strati (estrazione con probabilità uguale).

3.1 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 *Lime Survey*¹, 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 numero eccessivo di sostituzioni, al campione base sono stati affiancati 4 campioni suppletivi estratti con pari disegno campionario e numerosità del campione base.

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 7 Gennaio 20127 per concludersi il 31 Maggio 2021. **I dati si riferiscono alla situazione esistente nelle organizzazioni al 31 dicembre del 2020.**

4. 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.

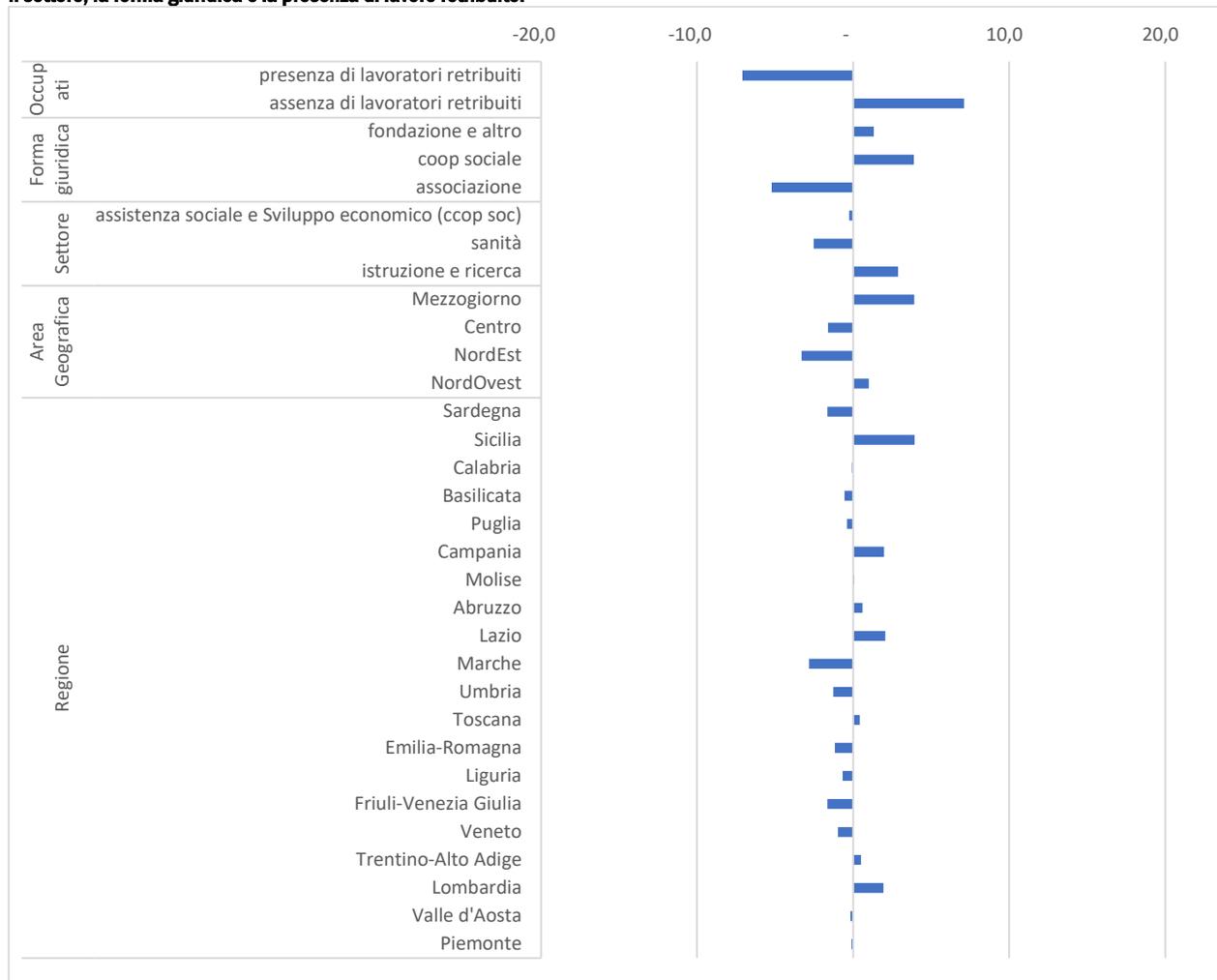
Tabella 3. - Campione effettivo secondo la collocazione territoriale, la forma giuridica, il settore e la presenza di lavoratori retribuiti

		Valori assoluti	Percentuale
<i>Regione</i>	Piemonte	388	8,3
	Valle d'Aosta	22	0,5
	Lombardia	726	15,5
	Trentino-Alto Adige	112	2,4
	Veneto	437	9,3
	Friuli-Venezia Giulia	179	3,8
	Liguria	162	3,5
	Emilia-Romagna	369	7,9
	Toscana	310	6,6
	Umbria	131	2,8
	Marche	262	5,6
	Lazio	374	8,0
	Abruzzo	65	1,4
	Molise	32	0,7
	Campania	248	5,3
	Puglia	254	5,4
	Basilicata	91	1,9
	Calabria	145	3,1
	Sicilia	165	3,5
	Sardegna	222	4,7

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.

Area geografica	Nord-ovest	1.298	27,7
	Nord-est	1.098	23,4
	Centro	1.078	23,0
	Mezzogiorno	1.220	26,0
Forma Giuridica	Associazione	3.158	67,3
	Cooperativa sociale	1.098	19,7
	Fondazione e altro	612	13,0
Settore di attività prevalente ICNPO	istruzione e ricerca	926	19,7
	Sanità	1.098	23,0
	Assistenza sociale e Sviluppo economico	2.687	57,3
Presenza di lavoratori retribuiti	No	2.164	46,1
	Sì	2.530	53,9
	Totale	4.694	100,0

Figura 1. - Campione effettivo Vs Popolazione di riferimento: differenza nelle composizioni percentuali secondo la regione, la ripartizione geografica, il settore, la forma giuridica e la presenza di lavoro retribuito.



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.

5. 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.

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 Registro Statistico delle Istituzioni non profit 2017 dell'Istat. 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 ausiliarie (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. - Informazioni 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

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.

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: $w_i = d_i * c_i$.

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
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 Registro Statistico delle Istituzioni non profit 2017 dell'Istat.