
Il disegno campionario e l'estrazione delle famiglie

La rilevazione continua è stata progettata e realizzata con l'obiettivo di ottenere un elevato standard qualitativo delle informazioni raccolte, sin dal momento della rilevazione. Ciò incide in modo determinante sulle scelte metodologiche effettuate per la progettazione e realizzazione del disegno campionario, per la definizione delle strategie per il trattamento dei dati e per la definizione delle procedure di stima.

1. Il disegno campionario

La popolazione di riferimento è costituita da tutti **componenti delle famiglie residenti** in Italia, anche se temporaneamente all'estero. Sono escluse le famiglie che vivono abitualmente all'estero e i membri permanenti delle convivenze (istituti religiosi, caserme ecc.).

Il campionamento è a **due stadi**, le unità di primo stadio (UPS) sono i comuni mentre quelle di secondo stadio (USS) sono le famiglie anagrafiche.

Le UPS sono stratificate all'interno di ciascuna provincia sulla base della dimensione demografica dei comuni. Tale stratificazione conduce all'identificazione di due tipologie di comuni: gli autorappresentativi (AR), che fanno tutti parte del campione; i non autorappresentativi (NAR), selezionati in base alla dimensione demografica. I comuni AR con maggiore dimensione demografica sono coinvolti nella rilevazione tutte le settimane; gli altri comuni AR una settimana al mese. Ognuno dei comuni NAR viene coinvolto nella rilevazione una settimana al mese secondo uno schema di associazione casuale delle settimane ai comuni del campione. Tale disegno campionario rimane fisso nel tempo.

Per ciascun comune viene estratto dalla lista anagrafica un campione casuale semplice di famiglie. Tale campione è composto da gruppi di quattro famiglie, cosiddette quartine, in numero pari a quello delle famiglie del campione previste per il comune stesso. In tal modo si assegnano ad ogni famiglia base da intervistare (la prima del gruppo estratto) altre tre famiglie da utilizzare in luogo di quella base qualora questa non voglia o non possa partecipare all'indagine.

Complessivamente vengono estratte circa **76.800 famiglie** a trimestre. In un anno vengono dunque intervistate circa 300.000 famiglie.

Ad esclusione della distribuzione settimanale delle famiglie del campione, le principali caratteristiche del disegno campionario sono rimaste invariate rispetto alla precedente rilevazione trimestrale sulle forze di lavoro. Anche la nuova indagine prevede uno schema di rotazione delle famiglie campione del tipo 2, 2, 2 (ogni famiglia estratta nel campione partecipa all'indagine per due trimestri consecutivi, non viene intervistata nei due trimestri successivi e partecipa nuovamente all'indagine per altri due trimestri). Da ciò consegue, in primo luogo, che il 50% delle famiglie intervistate a 3 e a 12 mesi di distanza sono le stesse, mentre a 9 e 15 mesi la sovrapposizione del numero delle famiglie è del 25%; e, in secondo luogo, che ogni trimestre un quarto del campione delle famiglie si trovi rispettivamente alla prima, seconda, terza o quarta intervista.

Come la precedente indagine, la rilevazione continua è in grado di garantire stime a livello regionale con cadenza trimestrale e a livello provinciale in media d'anno.

2. La fase di estrazione delle famiglie

L'estrazione delle unità di secondo stadio effettuata dai **comuni del campione** deve seguire alcune rigide regole metodologiche: dall'utilizzo di un specifico passo di estrazione ai criteri di ordinamento delle liste anagrafiche. Una particolare attenzione va quindi rivolta ai problemi che emergono in questa fase dell'indagine al fine di evitare gli effetti distorsivi legati a procedure di estrazione metodologicamente non corrette.

Il sistema di estrazione delle famiglie chiama in causa gli attori principali dell'organizzazione della rilevazione: il centro di coordinamento, gli uffici territoriali e, ovviamente, i comuni che forniscono le liste dei nominativi. Tale sistema si basa su un software prodotto dall'Istat e sulle potenzialità delle trasmissioni telematiche.

Il processo di acquisizione dei nominativi delle famiglie del campione è stato progettato al fine di sfruttare le potenzialità dei comuni che hanno l'anagrafe informatizzata; di aiutare i comuni senza anagrafe informatizzata a registrare i nominativi del campione in modo uniforme; di ridurre i tempi di acquisizione del campione; di monitorare l'intero processo.

I comuni ricevono un'informativa che riporta: le regole da seguire per effettuare un'estrazione metodologicamente corretta; le modalità di trasmissione dei nominativi (in base al grado di informatizzazione del comune); i tempi da rispettare. Alla comunicazione è allegato un particolare software applicativo prodotto dall'Istituto, denominato Gefi (Gestione elenchi famiglie per le indagini), per la registrazione controllata dei nominativi e per la produzione del file da inviare all'Istat (i comuni che non possiedono personal computer effettuano l'estrazione manualmente e inviano le liste su supporto cartaceo agli uffici territoriali).

Le informazioni sono sottoposte ad un doppio sistema di controllo, uno centralizzato presso l'Istat (Servizio formazione e lavoro) per l'individuazione di eventuali informazioni errate o mancanti, e l'altro presso gli uffici territoriali per l'integrazione delle informazioni o il rifiuto del file trasmesso dal comune). Questi ultimi hanno il fondamentale compito di gestire i rapporti con i comuni nella fase preliminare, nel corso dell'estrazione e dopo l'effettuazione dei controlli.

Il Servizio formazione e lavoro, oltre a supportare il lavoro degli uffici territoriali e a verificare la correttezza delle informazioni raccolte, governa il complesso algoritmo di assegnazione dei nominativi ai rilevatori. In base a tale algoritmo, costruito nel rispetto delle caratteristiche del disegno campionario, i nominativi raccolti vengono associati ad ogni singolo rilevatore assegnando anche la specifica settimana di calendario in cui intervistare la famiglia.

La tavola seguente riporta i risultati più significativi dell'estrazione dei nominativi delle famiglie del campione avvenuta nel corso 2003.

Comuni del campione, informatizzazione dell'anagrafe, modalità di estrazione e criteri di ordinamento delle famiglie per regione. Anno 2003, valori assoluti e composizioni per 100 comuni della stessa regione

Regioni	N. comuni	Anagrafe informatizzata	Estrazione informatizzata (b)	Criteri di ordinamento delle famiglie			
				Alfabetico di via	Cognome intestat.	Data di iscrizione in anagrafe	Altro
Piemonte/Valle d'Aosta(a)	143	96,5	68,1	86,7	5,6	3,5	4,2
Lombardia	173	98,3	71,8	86,7	1,7	2,9	8,7
Provincia di Bolzano	25	100,0	96,0	84,0	8,0	4,0	4,0
Provincia di Trento	48	100,0	93,8	85,4	2,1	8,3	4,2
Veneto	72	97,2	51,4	87,5	2,8	-	9,7
Friuli-Venezia Giulia	37	100,0	97,3	100,0	-	-	0,0
Liguria	25	100,0	88,0	80,0	8,0	8,0	4,0
Emilia-Romagna	69	100,0	79,7	85,5	2,9	5,8	5,8
Toscana	77	100,0	81,8	88,3	1,3	3,9	6,5
Umbria	20	100,0	75,0	100,0	-	-	0,0
Marche	32	96,9	61,3	81,2	3,1	9,4	6,3
Lazio	53	98,1	76,9	79,3	9,4	1,9	9,4
Abruzzo	32	100,0	84,4	87,5	-	9,4	3,1
Molise	28	92,9	46,2	67,9	21,4	7,1	3,6
Campania	89	91,0	75,3	73,1	10,1	10,1	6,7
Puglia	72	98,6	67,6	79,2	6,9	4,2	9,7
Basilicata	36	88,9	62,5	80,5	11,1	2,8	5,6
Calabria	76	93,4	57,7	81,6	7,9	1,3	9,2
Sicilia	103	95,1	70,4	73,8	5,8	9,7	10,7
Sardegna	36	100,0	66,7	94,4	-	-	5,6
Italia	1.246	97,0	72,2	83,5	5,1	4,6	6,8

(a) I dati relativi alle due regioni sono unificati in quanto l'ufficio regionale Istat di Torino gestisce entrambe le regioni;
(b) per 100 comuni con anagrafe informatizzata.