



# Istituto Nazionale di Statistica

## CENSIMENTO PERMANENTE DELLA POPOLAZIONE E DELLE ABITAZIONI ALLEGATO A

### Utilizzo delle fonti amministrative a fini censuari: popolamento e correzione del Registro Base degli Individui (RBI)

#### 1. Le fonti amministrative per il popolamento del Registro. Criteri per la selezione degli individui eleggibili e la valorizzazione delle variabili core

La struttura di base del RBI è data dall'insieme degli individui identificati dal Sistema Integrato di Microdati dell'Istat<sup>1</sup> (SIM) che risultino *eleggibili*, cioè tutti i soggetti (inclusi gli stranieri e i non residenti in Italia):

- (i) che compaiono in almeno una delle fonti<sup>2</sup> amministrative che SIM tratta;
- (ii) per i quali sono "noti" (ad un prefissato livello di attendibilità basato su analisi statistiche di qualità<sup>3</sup>) i valori relativi ad un insieme ristretto di variabili *invarianti nel tempo*: 'genere', 'data di nascita' e 'luogo di nascita'. Tali variabili sono denominate "*variabili di eleggibilità*".

L'insieme degli individui eleggibili rappresenta l'insieme più completo possibile su cui l'Istat opera la selezione per l'identificazione delle popolazioni statistiche di interesse per le proprie indagini; in particolare, a fini censuari, la popolazione di riferimento è quella abitualmente dimorante.

Per costruzione, il RBI non potrà contenere individui non presenti in SIM. A titolo di esempio, non saranno presenti in RBI:

1. cittadini non UE, clandestini o irregolari, che hanno vissuto per un certo periodo sul territorio italiano senza permesso o che sono passati sul territorio italiano in maniera clandestina;
2. cittadini UE che vivono o sono passati sul territorio italiano senza entrare in contatto con alcuna delle fonti che alimentano SIM.

<sup>1</sup> Il lavoro è inserito nel Programma Statistico Nazionale con il codice IST-02270.

<sup>2</sup> Ai fini del popolamento di RBI sono trattate esclusivamente le fonti amministrative pseudonimizzate in SIM specificate nel PGC.

<sup>3</sup> Analisi basate sui valori delle variabili di eleggibilità del RBI come sesso, data di nascita e luogo di nascita: % di informazioni mancanti, % di informazioni errate (fuori dominio), % di informazioni incompatibili (es. 31 febbraio nella data di nascita, ecc.).



# Istituto Nazionale di Statistica

La conoscenza delle popolazioni di cui al punto 1 esula dal campo di osservazione del censimento permanente della popolazione e, pertanto, non costituisce oggetto di trattamento statistico. Al contrario, per gli individui di cui al punto 2, l'interesse censuario riguarda solo coloro che sarebbero candidati ad essere conteggiati tra la popolazione abitualmente dimorante. L'ammontare di questa sotto-popolazione, non identificabile a livello micro, deve essere necessariamente stimato con opportuni metodi statistici (stima della sotto copertura del RBI).

Contribuiscono alla struttura del RBI anche le variabili utili al censimento quali: residenza anagrafica, famiglia o convivenza anagrafica, cittadinanza e titolo di studio.

Alcune di queste variabili provengono da fonti che appare lecito considerare particolarmente affidabili rispetto ad una specifica popolazione di interesse. Un tipico esempio è fornito dalla variabile 'residenza anagrafica', come desunta dalle fonti anagrafiche a valle del processo di validazione statistica. In questi casi la valorizzazione della variabile si ottiene mutuandola dalle Liste Anagrafiche Comunali (LAC) "validate"<sup>4</sup>.

Più in generale, le modalità delle variabili da riferire a ciascun individuo sono selezionate secondo una gerarchia di fonte:

- se l'individuo è presente in LAC, si considerano i valori di LAC;
- se l'individuo non è presente in LAC, ma è presente in Anagrafe Tributaria (AT), si considerano i valori da AT;
- se l'individuo non è presente né in LAC né in AT ma è definito "cluster 1" (ossia l'individuo appare in una sola fonte nel SIM), si prendono i valori presenti nella fonte;
- tutte le altre casistiche generano dei "residui" che dovranno essere trattati statisticamente.

In accordo con il principio della qualità delle informazioni personali (intesa come correttezza rispetto al dato reale), uno dei criteri più semplici da adottare per gli individui residuali di cluster 2 o + (ossia gli individui che appaiono in 2 o più fonti non anagrafiche e non AT) è quello di selezionare il valore modale della variabile, ossia secondo il criterio della prevalenza.

---

<sup>4</sup> Il trattamento ai fini di validazione statistica prevede l'eliminazione degli eventuali doppi, dei deceduti o trasferiti all'estero, in coerenza con quanto risulta dalle fonti anagrafiche sui flussi della dinamica demografica, e il trattamento delle mancate risposte, dei fuori dominio ecc., effettuato con procedure statistiche consolidate e riconosciute a livello internazionale (Regolamento europeo sulla qualità della statistica ufficiale, Regolamento europeo sulle statistiche demografiche, solo per citarne i più importanti).



# Istituto Nazionale di Statistica

Una variabile fondamentale per il censimento permanente e le statistiche demografiche è la “dimora abituale”. In RBI la dimora abituale è validata sulla base delle risultanze del sistema MIDEA-ANVIS<sup>5</sup> che tratta i flussi anagrafici in modo da assicurare la coerenza tra due stock successivi di popolazione e la dinamica demografica.

Gli individui presenti nel RBI a cui è attribuito il flag “abituale dimorante” sulla base delle risultanze di MIDEA-ANVIS costituiscono la lista da cui estrarre le famiglie per l’indagine diretta del censimento permanente.

## **2. La correzione del conteggio nel RBI con il censimento**

La stima della dimora abituale tramite il sistema MIDEA-ANVIS, essendo basata sulle fonti anagrafiche, risente in parte degli errori di copertura delle stesse. Si pone quindi la necessità di correggere questo errore ricorrendo ad una modifica che consenta di allineare il Registro alle risultanze censuarie.

In accordo con la legge censuaria, con il PGC e con i trattamenti descritti nel Programma Statistico Nazionale, è previsto l’utilizzo delle fonti amministrative non anagrafiche individuate (cfr. par. 2.5 PGC) al fine di migliorare la qualità statistica del conteggio di popolazione nel Registro.

Le stime della sotto sovra copertura di RBI, ottenute dai segnali desumibili dalle fonti amministrative e dalle indagini dirette, vengono utilizzate per la costruzione di correttori per *un determinato livello di aggregazione della popolazione*.

Il procedimento che porta al calcolo dei correttori prevede diverse fasi. Si parte dai segnali di vita desumibili dalle fonti amministrative e dai dati delle indagini dirette, dai quali si desumono le informazioni per la determinazione dei correttori del RBI per sotto e sovra copertura. A tal fine è prevista una fase di linkage/accoppiamento che richiederà l’associazione dei dati raccolti con quelli degli archivi indicati nel PGC.

L’utilizzo dei dati dell’indagine Areale (A) effettuata solo in alcuni comuni, inoltre, rappresenta, per ciascun incrocio delle variabili sesso, classe di età, cittadinanza (italiana/straniera) e classe comunale, un ulteriore strumento di verifica (oltre ai segnali desumibili dalle fonti amministrative) che, utilizzato per tutte le unità del Registro appartenenti a tali comuni, può contribuire al calcolo del totale della popolazione corretto per sovra e sotto copertura.

---

<sup>5</sup> Il lavoro MIDEA (Micro Demographic Accounting) ANVIS (Anagrafe Virtuale Statistica) dell’Istat è inserito nel Programma Statistico Nazionale con il codice IST-02703.



## Istituto Nazionale di Statistica

I correttori sono applicati agli individui che nel RBI sono classificati come abitualmente dimoranti. Nei comuni dove viene effettuata anche l'Indagine A, tutti gli individui appartenenti ad uno stesso incrocio delle variabili sesso, classe di età, cittadinanza (italiana/straniera) e classe comunale hanno lo stesso correttore. Gli incroci hanno numerosità che non consentono di identificare gli individui che li compongono.

L'impiego dei segnali di vita (desunti dalle fonti amministrative o dall'indagine da Lista) e degli eventuali correttori desunti dall'indagine A consente di utilizzare il RBI per produrre stime della numerosità e della struttura della popolazione coerenti con i risultati del censimento permanente e corrette "statisticamente" rispetto agli errori di sovra o sotto copertura del Registro.

Questi correttori sono prodotti in modo tale che la loro *somma* nel RBI restituisca l'ammontare *stimato* della popolazione di interesse (nell'esempio: la popolazione abitualmente dimorante in Italia) *al livello di aggregazione specificato*.

Una volta corretto per sovra e sotto copertura il RBI viene utilizzato per la produzione ufficiale della popolazione abitualmente dimorante (popolazione legale) e come base a cui riportare i risultati relativi alle variabili socio-economiche delle indagini censuarie.