

Nota tecnica sulla produzione dei dati del Censimento Permanente: la stima della popolazione residente per sesso, età cittadinanza, grado di istruzione e condizione professionale per gli anni 2018 e 2019

Dalla rilevazione “porta a porta” al Censimento permanente

Da ottobre 2018 l'Istat ha avviato il Censimento permanente della popolazione e delle abitazioni, che sostituisce il Censimento della Popolazione decennale, basato sull'integrazione tra le informazioni reperibili dalle fonti amministrative e quelle acquisite con indagini campionarie a rotazione.

L'uso combinato di dati da registri e indagini garantisce la tradizionale offerta informativa del Censimento sulle principali caratteristiche demografiche e socio-economiche della popolazione residente in Italia e, al tempo stesso, consente da un lato di contenere i costi e il disturbo statistico sulle famiglie, dall'altro di non sovraccaricare la rete di rilevazione dei Comuni.

Obiettivo specifico del Censimento permanente resta dunque la produzione dei dati relativi al conteggio a livello comunale della popolazione residente e alla sua distribuzione per sesso, età, cittadinanza, grado di istruzione e condizione professionale, unitamente ai alla produzione dei dati sulle famiglie e sulle abitazioni.

Il passaggio dal Censimento tradizionale, basato sulla rilevazione “porta a porta”, esaustiva e simultanea, condotta su tutto il territorio nazionale, al Censimento permanente è stato graduale e in futuro sarà caratterizzato progressivamente da un uso sempre maggiore dei dati amministrativi. Fondamentale in tal senso è stato il Censimento del 2011 che, pur nell'ambito di una metodologia ancora sostanzialmente convenzionale (rilevazione sul campo esaustiva e simultanea), si è basato fortemente sull'uso delle fonti amministrative e sulla rilevazione campionaria di un sottoinsieme di informazioni socio economiche. In particolare, nel 2011 le fonti amministrative sono state utilizzate per guidare la rilevazione: le Liste Anagrafiche Comunali per la spedizione postale dei questionari alle famiglie iscritte in anagrafe, e la Lista Integrativa da Fonti Ausiliarie, per rilevare la popolazione abitualmente dimorante non ancora iscritta in anagrafe.

Il Censimento permanente si basa però su un diverso rapporto tra rilevazione sul campo e fonti amministrative. Se nel Censimento del 2011 la rilevazione costituiva la base della produzione dei dati censuari, nel Censimento permanente della popolazione il perno è costituito dal Registro di Base degli Individui (d'ora in avanti RBI) che, insieme ai registri tematici (quali quelli sull'occupazione e sull'istruzione), viene sottoposto al vaglio delle indagini campionarie annuali con il fine di correggere e integrare le informazioni in essi contenute. Questo passaggio è stato reso possibile dalla regolare acquisizione di fonti amministrative, e dal loro trattamento e utilizzo a fini statistici, operati nell'ambito del processo di modernizzazione dell'Istituto e nel contesto del ruolo via via maggiore assunto dalle fonti amministrative nell'ambito della produzione statistica ufficiale a livello internazionale.

Il RBI è un ambiente informativo interno all'Istituto a supporto dei processi produttivi statistici; in particolare, è l'infrastruttura di base per la produzione delle statistiche ufficiali riferite alla popolazione e il riferimento per l'estrazione dei campioni delle indagini previste per il Censimento permanente (cfr. legge censuaria e Piano Generale di Censimento) e di tutte le indagini campionarie sulle famiglie.

In accordo con l'impianto metodologico del Censimento permanente, e nel rispetto delle norme sulla protezione dei dati personali, il RBI contiene dati anonimi “statistici”, ossia derivanti da un processo di elaborazione e validazione statistica a partire da fonti amministrative e di indagine, e riferiti a un numero limitato di variabili funzionali alla rappresentazione delle principali caratteristiche strutturali della popolazione e delle famiglie.

Il RBI è “consolidato” con cadenza annuale con riferimento al 31 dicembre. Tale consolidamento è realizzato mediante l'integrazione dei flussi individuali anagrafici della dinamica demografica (nascite, decessi, trasferimenti di residenza da e per altro comune o da e per l'estero). L'applicazione del modello di contabilità demografica MIDEA (Micro-DEmographic Accounting) consente di sfruttare le potenzialità della base dati micro (flussi+stock) per produrre indicatori più accurati e innovativi sulla dinamica demografica, tenendo conto della sequenza degli eventi demografici sperimentati dagli individui.



I principali output del sistema MIDEA applicati al RBI “consolidato” sono il bilancio demografico della popolazione residente comunale e la struttura della popolazione residente comunale per sesso, età, stato civile e cittadinanza. La produzione tempestiva di questi output anticipa i risultati definitivi annuali del Censimento permanente con l’obiettivo di assicurare la continuità della produzione statistica ufficiale, soddisfacendo al tempo stesso le tempistiche stringenti richieste tanto dagli utilizzatori istituzionali dei dati di popolazione, quanto dal regolamento europeo sulle statistiche demografiche.

Questi stessi output vengono rielaborati e diffusi successivamente come definitivi a seguito della correzione del RBI mediante l’integrazione dei pesi prodotti dal Censimento permanente. RBI “corretto” assicura pertanto l’allineamento dei dati censuari con quelli delle statistiche demografiche.

L’integrazione del Censimento con il Sistema dei Registri è finalizzata da un lato a correggere gli errori di sovra e sotto-copertura del RBI (individuando le persone presenti nel Registro come residenti ma non trovate sul territorio e quelle che invece sono state trovate sul territorio come abitualmente dimoranti ma non risultano nel Registro) e, dall’altro, a raccogliere le informazioni che attualmente non sono disponibili nei dati di fonte amministrativa.

La grande innovazione che il Censimento permanente introduce sullo scenario statistico è, appunto, la continuità della raccolta e della diffusione dei dati necessari al conteggio di popolazione e delle informazioni socio-demografiche, con un progressivo aggiornamento del Registro di base degli individui che permetterà la diffusione dei dati comunali con periodicità annuale (e con periodicità più ampia anche a livello sub-comunale).

Va sottolineato inoltre che il nuovo impianto censuario introduce, nell’ambito delle statistiche sulla popolazione, una completa integrazione con gli output delle statistiche demografiche, contribuendo in tal modo a una maggiore stabilità e coerenza di tutto il sistema dei dati che riguardano la popolazione.

Va ricordato, a tal proposito, che con il Censimento tradizionale veniva determinata una popolazione differente dalla popolazione anagrafica e dalla popolazione statistica annualmente calcolata dall’Istat. Su quella base si determinava un’interruzione delle serie annuali ed era necessario, al termine del Censimento, procedere alla ricostruzione delle serie storiche precedenti l’anno di Censimento e a una revisione anagrafica di notevole entità. L’intero sistema statistico che utilizza dati di popolazione (indicatori demografici, indicatori sociali, forze di lavoro, contabilità nazionale) doveva infatti essere rivisto alla luce dei risultati censuari.

Grazie all’introduzione del Censimento permanente, una volta effettuata l’ultima ricostruzione delle serie storiche di popolazione (cioè dall’edizione censuaria del 2011 alla nuova edizione del 2018), non sarà più necessario procedere all’allineamento del calcolo della popolazione effettuato sulla base dell’aggiornamento dei flussi demografici con i risultati censuari. In tal senso, il Censimento (con cadenza annuale) e le statistiche demografiche (che producono aggiornamenti mensili) convergono annualmente sullo stesso conteggio di popolazione.

La convergenza tra Censimento e statistiche demografiche risulta ancora più significativa se si considera la funzione dell’Anagrafe Nazionale della Popolazione Residente (ANPR), che sta centralizzando tutte le informazioni anagrafiche in un unico sistema (che riguarda al momento circa l’85% dei Comuni e l’82% della popolazione). La compresenza del Censimento permanente della popolazione e dell’ANPR, già in connessione sulla base di un Accordo di Servizio tra il Ministero dell’Interno e l’Istat per la gestione continua degli aggiornamenti delle statistiche demografiche, è il fattore innovativo e strategico che modifica strutturalmente il sistema delle statistiche della popolazione.

Da ultimo, ma certo non per ordine di importanza, va ricordato il fatto che le innovazioni introdotte con il Censimento permanente rappresentano una grande opportunità per lo sviluppo e il consolidamento della funzione statistica nell’ambito del Sistema statistico nazionale. La realizzazione dei censimenti si basa, infatti, su una complessa rete istituzionale che vede la partecipazione di soggetti istituzionali coinvolti a vario titolo sul piano dell’organizzazione e della realizzazione delle operazioni censuarie (il Ministero dell’Interno, le Regioni, le Province e, in particolare, i Comuni, che hanno sempre gestito tutte le operazioni di raccolta dei dati in stretto raccordo con l’Istat). In tal senso, il nuovo Censimento permanente comporta anche la continuità e il progressivo consolidamento di questa rete istituzionale, chiamata ogni anno a condividere (con le componenti rappresentative dei diversi soggetti istituzionali) e a supportare sul territorio le operazioni censuarie.



Le indagini per il Censimento permanente

Obiettivo del Censimento permanente della popolazione è correggere gli errori di copertura di RBI e produrre i dati necessari a integrare nei registri tematici le informazioni mancanti o di qualità non sufficiente, in modo da assicurare la completezza e la qualità richiesta e garantita dai dati censuari. Il disegno di rilevazione per il Censimento Permanente si articola in due distinte indagini campionarie, una Areale e l'altra da Lista.

Entrambe le indagini sono progettate per rilevare sul campo le informazioni necessarie a migliorare la qualità di RBI, che può essere affetto da errori di sovra-copertura (inclusione nel Registro di individui che non hanno più dimora abituale nel comune) e di sotto-copertura (mancata inclusione nel Registro di individui che hanno dimora abituale nel comune). Inoltre, entrambe le indagini rilevano le informazioni necessarie a integrare i registri tematici (quali quello dell'istruzione e dell'occupazione), che non contengono tutte le informazioni necessarie per produrre i dati censuari.

Nel corso del primo ciclo del Censimento permanente (2018-2021) il disegno ha previsto la realizzazione di entrambe le indagini tutti gli anni nei comuni autorappresentativi e un anno su quattro, a rotazione, nei comuni cosiddetti non autorappresentativi¹.

Annualmente i comuni coinvolti nelle rilevazioni sono circa 2.850 (di cui 1.150 autorappresentativi e 1.700 non autorappresentativi), per un totale di circa 1 milione e mezzo di famiglie; di queste, 950.000 sono incluse nella rilevazione da Lista e 450.000 in quella Areale. Al termine del ciclo, tutti i comuni italiani saranno stati sondati almeno una volta.

L'Indagine Areale è basata su un campione di indirizzi e/o sezioni di Censimento estratto dal Registro base dei luoghi (cfr. RSBL, nel prospetto 1). Nella fase preliminare dell'indagine, i rilevatori procedono alla verifica degli indirizzi campionati e, successivamente, alla rilevazione porta a porta "alla cieca", ovvero senza l'ausilio di liste precompilate, per rilevare in maniera esaustiva tutte le famiglie dimoranti abitualmente presso gli indirizzi confermati nella fase preliminare. Tra le modalità di raccolta dei dati è prevista, oltre all'intervista tramite rilevatore (CAPI²), anche l'autocompilazione assistita a domicilio o presso i Centri Comunali di Rilevazione. L'indagine areale consente di determinare, ai fini del conteggio di popolazione, gli individui dimoranti abitualmente sul territorio e non iscritti come residenti in RBI, cioè l'errore di sotto-copertura.

L'Indagine da Lista è basata su un campione di famiglie estratto da RBI. Nella fase di restituzione spontanea le famiglie procedono alla compilazione autonoma del questionario di Censimento via internet (CAWI³), contattando telefonicamente il comune (CATI⁴) oppure recandosi presso i Centri Comunali di Rilevazione. Nella fase successiva, i rilevatori vengono inviati sul campo per il recupero delle mancate risposte, procedendo ad intervistare le famiglie con tecnica CAPI o a registrare nel Sistema di Gestione della Rilevazione un esito corrispondente alla situazione accertata sul campo.

Ai fini del conteggio 2018 e 2019, l'Indagine da Lista è stata usata per determinare la misura dell'errore di sovra-copertura di RBI, congiuntamente ai "segnali di vita" amministrativi forniti dall'Archivio Integrato dei Dimoranti Abitualmente in Italia (AIDA).

Inoltre, ai fini della produzione degli ipercubi su istruzione, paese estero di cittadinanza e condizione professionale, le informazioni rilevate in entrambe le indagini sono state utilizzate congiuntamente ai dati del RBI e dei registri tematici su occupazione e istruzione attraverso l'impiego di modelli statistici.

¹ I comuni AR sono i comuni con popolazione maggiore di 17.800 abitanti e i comuni più piccoli che non ruotavano nel campione della Rilevazione Istat sulle Forze di lavoro di giugno 2017. Fanno eccezione i comuni non autorappresentativi con popolazione inferiore a 1.000 abitanti, nei quali viene condotta la sola rilevazione Areale. Inoltre, tra questi, i comuni con meno di 300 abitanti effettuano la rilevazione areale su tutto il territorio comunale.

² Computer Assisted Personal Interviewing (CAPI).

³ Computer Assisted Web Interviewing (CAWI).

⁴ Computer Assisted Telephone Interviewing (CATI).



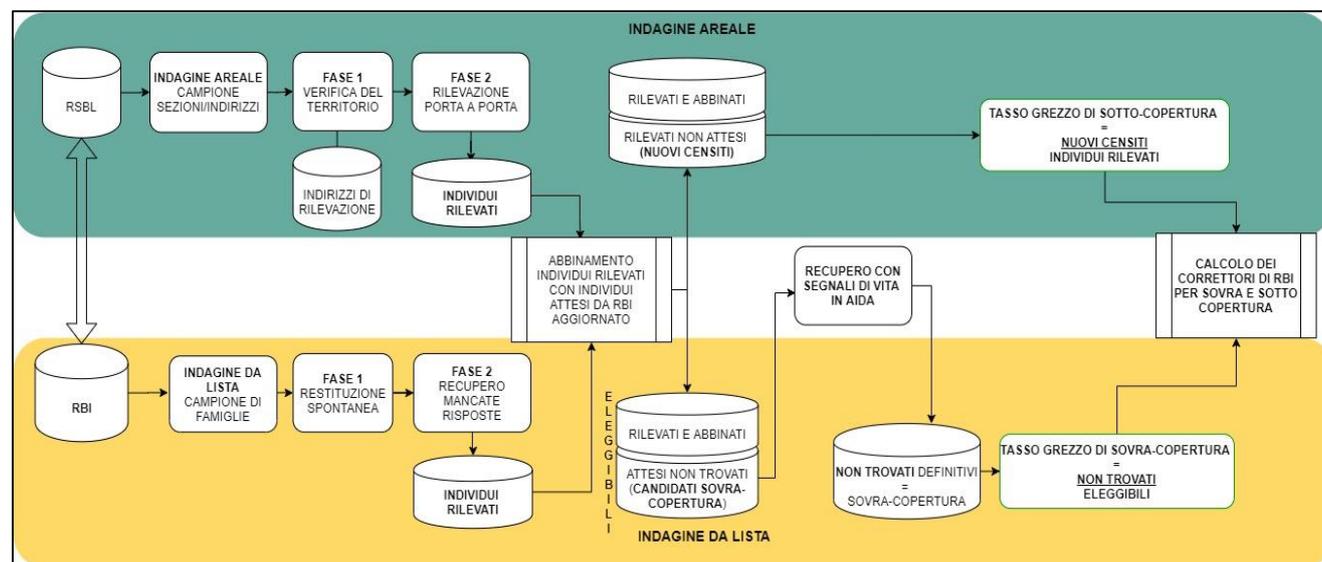
Il conteggio 2018-2019

Il Registro di Base degli Individui e le informazioni per sesso, età e cittadinanza che esso fornisce al 31 dicembre di ogni anno rappresentano la fonte principale di input per il Censimento permanente della popolazione. Il conteggio della popolazione censita alla fine del 2018 e del 2019 è ottenuto attraverso l'applicazione di coefficienti di correzione ai dati individuali contenuti in RBI, al fine di eliminare gli errori di sotto-copertura e sovra copertura.

In particolare, l'indagine Areale ha consentito di calcolare il coefficiente di correzione dell'errore di sotto-copertura attraverso l'individuazione di persone che, all'esito dell'indagine, hanno la dimora abituale nel comune ma non sono incluse come residenti in RBI, mentre l'indagine da Lista ha consentito di determinare la sovra-copertura di RBI, vale a dire l'aggregato di individui che nel Registro risultano residenti ma che non sono stati trovati sul territorio dai rilevatori anche dopo tentativi ripetuti di contatto (cfr. Prospetto 1). Inoltre, per il calcolo della sovra-copertura, prima di effettuare il calcolo del coefficiente finale di correzione, si è proceduto al recupero dei non rispondenti sulla base degli archivi amministrativi di fonte non anagrafica (AIDA), nel caso in cui per questi individui si registrassero "segnali di vita" forti nel comune di residenza. In particolare, sono stati considerati presenti sul territorio comunale quegli individui che, seppur non rilevati nell'indagine da Lista, avevano nello stesso comune (cioè nel comune in cui risultavano residenti in RBI) segnali di lavoro validi per almeno 8 mesi nell'anno solare, erano iscritti alla scuola primaria o secondaria, figuravano nel casellario dei pensionati, o erano familiari a carico dei coniugi o dei genitori dichiaranti il reddito relativo all'anno precedente la data di riferimento delle indagini censuarie.

In sintesi, al termine del processo, a tutti gli individui che in RBI risultano residenti nel comune viene agganciato un peso, che consente di correggere il conteggio di popolazione sulla base degli errori di sovra e sotto-copertura del Registro stimati per quel particolare comune.

PROSPETTO 1. SCHEMA SINTETICO DEL PROCESSO DI DETERMINAZIONE DEGLI ERRORI DI COPERTURA DI RBI



Fonte: Istat, 2020

La procedura di calcolo dei correttori di sotto e sovra copertura

Al fini del calcolo dei correttori, ad ogni individuo rilevato alle indagini del Censimento permanente sono associate le caratteristiche demografiche di genere, età e cittadinanza e il comune di residenza. Quindi sarebbe possibile calcolare tassi grezzi di sotto e sovra copertura per tutti i profili individuati dall'incrocio di queste variabili ma, date le esigue numerosità campionarie riscontrate per alcuni profili in comuni medio-piccoli, e per poter garantire la robustezza e l'accuratezza delle stime finali prodotte, il profilo considerato per il calcolo dei tassi di copertura è costituito dalle variabili "comune di residenza" e "cittadinanza" (italiana e straniera).



La procedura di stima dei correttori di sotto e sovra-copertura del registro RBI può essere schematizzata nei passi che seguono (Cfr. prospetto 1):

1. calcolo del tasso “grezzo” di sotto copertura, che è dato dal rapporto tra il numero di individui nuovi censiti (rilevati all’indagine e non residenti in RBI) e il numero totale di individui rilevati all’indagine Areale

$$p_{sotto} = \frac{\text{Nuovi censiti}}{\text{Rilevati}}$$

2. calcolo del tasso “grezzo” di sovra copertura, che è dato dal rapporto tra gli individui residenti di RBI non trovati al Censimento, e al denominatore gli stessi individui non trovati al Censimento più gli individui residenti di RBI che hanno risposto all’indagine da Lista

$$\frac{\text{Non trovati}}{\text{Non trovati} + \text{Rilevati residenti in RBI}}$$

Occorre ricordare, però, che per quanto riguarda gli individui “non trovati” all’indagine da Lista, come già detto, si è proceduto al recupero degli individui con “segnali di vita” amministrativi nello stesso comune in cui risultano residenti in RBI. Più precisamente, gli individui con segnali di presenza sul territorio comunale di forte intensità, cioè con un segnale continuativo di 8-12 mesi nel periodo precedente la data di riferimento del Censimento, hanno concorso a “correggere” la potenziale sovra-copertura del Registro stimata dalla sola indagine.

Dopo aver calcolato i due tassi come descritto, il rapporto tra le due componenti definisce il “correttore grezzo” di sovra e sotto-copertura:

$$corr = \frac{1 - p_{sovra}}{1 - p_{sotto}}$$

In particolare può verificarsi quanto segue:

- a) il correttore è pari ad 1 quando, il registro RBI, per quel comune, non è affetto da errori di sovra e sotto-copertura oppure i due fenomeni hanno la stessa intensità e quindi si compensano; in questo caso tutti gli individui residenti nel comune, presenti nel registro, rappresentano loro stessi e possono essere contati con peso pari ad 1 per fornire il dato comunale corretto;
- b) il correttore è superiore ad 1, quando in RBI il fenomeno della sotto-copertura è superiore al fenomeno della sovra copertura, quindi il peso da applicare agli individui del comune deve essere incrementato di una certa quantità per fornire il dato comunale corretto;
- c) il correttore è inferiore ad 1, quando in RBI il fenomeno della sovra-copertura è superiore al fenomeno della sotto-copertura, quindi il peso da applicare agli individui del comune deve essere ridotto di una certa quantità per fornire il dato comunale corretto.

Come già evidenziato, per garantire la robustezza e l’accuratezza delle stime prodotte, i correttori non sono calcolati a livello individuale ma per profilo di comune e cittadinanza (italiano/straniero).

Dopo il calcolo del correttore “grezzo” di sovra e sotto-copertura, si procede al calcolo delle stime dirette per i comuni campione. Successivamente, al fine di ottenere stime dei correttori di sovra e sotto-copertura per i comuni non inclusi nel campione annuale e stime più efficienti per i comuni campione, è necessario procedere al calcolo delle stime indirette.



Il calcolo delle stime dirette e indirette

Il passo successivo al calcolo dei tassi grezzi di sovra e sotto-copertura è consistito nell'individuazione dei comuni da escludere dal processo di stima diretta del conteggio di popolazione (cosiddetti *outlier*), sulla base di indicatori di qualità dell'indagine o del RSBL (utilizzato per l'estrazione del campione areale).

Dopo aver effettuato le stime dirette per i comuni partecipanti alle indagini (a esclusione dei comuni *outlier*) si è proceduto al calcolo delle stime indirette. Infatti, circa 5.000 comuni non sono inclusi nel campione annuale e di conseguenza per essi non è possibile calcolare stime dirette del tasso di sovra e sotto-copertura. Inoltre, le stime dirette, pur essendo in generale caratterizzate da una bassa componente distorsiva, possono avere in alcuni domini di stima una variabilità elevata. Questo può verificarsi perché sono prodotte per sottopopolazioni di piccole dimensioni, ad esempio gli stranieri, per le quali la dimensione campionaria è insufficiente per produrre stime di buona qualità. Questa criticità è acuita dal fatto che per domini molto specifici, quali quelli per classi di età, la sovra-copertura e la sotto-copertura di RBI risultano particolarmente difficili da stimare con buona precisione.

I metodi di stima indiretta sono quindi necessari per poter calcolare le stime di sovra e sotto-copertura e del correttore per i comuni non campione (e per i comuni per i quali non è stato possibile calcolare le stime dirette, perché individuati come *outlier*). Inoltre, l'utilizzo di metodi di stima indiretta migliora anche l'efficienza (in termini di errore associato alle stime) delle stime dirette relative ai comuni campione.

Il modello adottato per le stime indirette appartiene ai metodi di stima per piccole aree, che hanno come finalità quella di migliorare la precisione delle stime dirette, se queste non sono ritenute sufficientemente affidabili, e di fornire una stima anche per i domini non pianificati in fase di campionamento e per i quali, ovviamente, la stima diretta non è disponibile (i comuni non campione).

Le stime dirette di sovra e sotto-copertura per ogni profilo comunale sono state calcolate attraverso un processo di calibrazione dei dati campionari. In dettaglio, il processo di calibrazione vincola, per ogni comune, i pesi campionari dell'indagine ai totali noti di popolazione per sesso e cittadinanza desunti da RBI.

Il peso assegnato alla stima diretta nel calcolo della stima indiretta è funzione della sua variabilità. Tanto più le stime dirette sono precise e, quindi, affidabili, tanto maggiore è il peso assegnato alla stima diretta e minore il peso dato alla stima sintetica (cosa che in genere accade per i grandi comuni). Al contrario, tanto più le stime dirette sono poco precise e, quindi, ritenute meno affidabili per la produzione delle stime di sovra e sotto-copertura, tanto minore è il peso dato alla stima diretta e tanto maggiore il peso attribuito alla stima sintetica, circostanza che in genere si verifica nei comuni di più piccola dimensione demografica.

Il “correttore medio” del Censimento permanente della popolazione 2018-2019

L'obiettivo di produrre i conteggi di popolazione per domini molto specifici e la necessità di garantire una maggiore stabilità delle stime richiede di adottare, per la costruzione dei correttori di sovra e sotto-copertura, un correttore medio di popolazione che sintetizza i risultati ottenuti da due rilevazioni annuali consecutive.

Più precisamente, per ogni dominio di stima e separatamente per sovra e per sotto-copertura, si è provveduto a calcolare la media delle stime dei correttori tra gli anni 2018 e 2019, ponderate con le rispettive ampiezze demografiche. La stima del correttore medio 2018-2019 viene così ottenuta come rapporto tra le medie ponderate delle stime del correttore di sovra-copertura e del correttore di sotto-copertura.

A tutti gli individui di un determinato comune con la stessa cittadinanza (italiana o straniera), presenti in RBI come residenti in famiglia e con un'età inferiore a 98 anni, è stato quindi applicato lo stesso correttore sia per il 2018 che per il 2019. Fanno eccezione gli individui residenti in convivenza e gli individui residenti in famiglia con un'età pari o maggiore a 98 anni, per i quali il correttore è stato posto pari a 1⁵.

⁵ Gli individui residenti in convivenza non rientrano nel campo di osservazione delle rilevazioni 2018-2019, mentre per gli individui di 98 anni o più l'Istat ha una rilevazione e un monitoraggio dedicati, e quindi si ritiene che il loro conteggio non sia affetto da errori di copertura.



Le stime degli ipercubi su grado di istruzione, condizione professionale e paese estero di cittadinanza

Il cambio di paradigma censuario ha imposto una totale revisione, sia concettuale che metodologica, del processo di costruzione degli ipercubi censuari. La nuova strategia è finalizzata a massimizzare l'integrazione tra il contributo informativo delle indagini campionarie del Censimento permanente e i dati già inclusi nei registri statistici e negli archivi amministrativi, in modo da assicurare, oltre ad una solida base inferenziale, la completa coerenza del sistema delle statistiche censuarie.

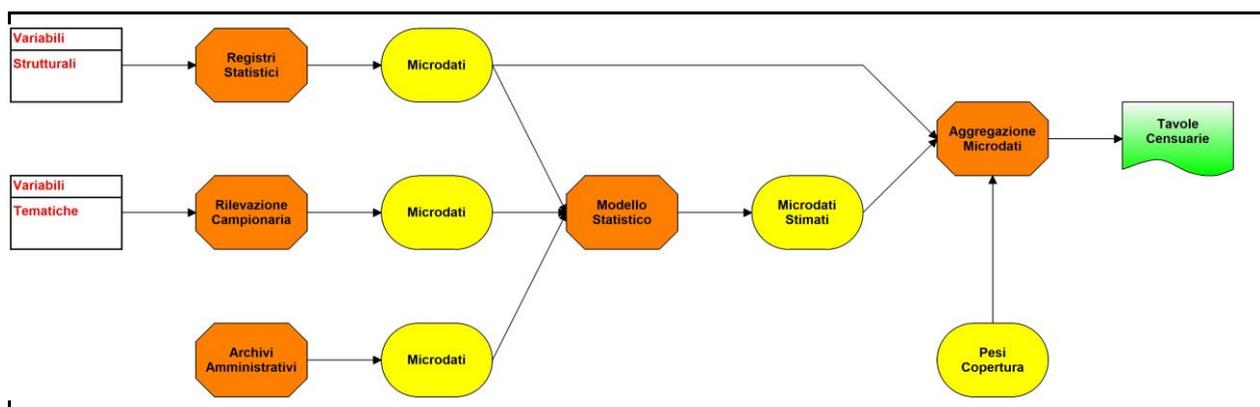
Ciò ha comportato l'abbandono del processo di stima adottato nel 2011 che, prevedendo per alcune variabili una rilevazione esaustiva e per altre campionaria, aveva implicato per le seconde il riporto all'universo delle informazioni campionarie. La scelta fatta per la definizione dell'architettura alla base della costruzione degli ipercubi censuari ricalca, in un certo qual modo, quanto avveniva nelle tornate censuarie precedenti al 2011, nelle quali, essendo prevista la raccolta delle caratteristiche socio-demografiche attraverso un'indagine esaustiva per tutte le variabili, gli ipercubi censuari venivano desunti mediante aggregazione dei dati individuali.

Precisamente, l'approccio adottato consiste nella creazione di una base di microdati individuali che utilizza:

- ✓ per le principali variabili demografiche, nello specifico sesso, età e cittadinanza, i dati individuali presenti nei registri statistici, che sono denominate *variabili strutturali*;
- ✓ per le restanti variabili, i valori stimati attraverso modelli statistici che usano congiuntamente i dati di indagine e le informazioni presenti nei registri e negli archivi amministrativi. Le variabili attualmente validate e diffuse che ricadono in questa categoria, denominate *variabili tematiche*, sono quelle relative al grado di istruzione, alla condizione professionale e non professionale.

Quindi, gli ipercubi censuari sono calcolati mediante la semplice aggregazione dei valori contenuti nella base di microdati, pesando ciascun record individuale con il corrispondente valore del correttore di sopra e sotto-copertura. Questa architettura è perfettamente consistente con l'informazione contenuta nel sistema dei registri e rispetta, inoltre, i requisiti di coerenza interna, nel senso che una variabile presente in più ipercubi non vede la sua distribuzione statistica cambiare a seconda dell'ipercubo da cui è desunta.

PROSPETTO 2. SCHEMA SINTETICO DELL'ARCHITETTURA PER LA PRODUZIONE DEGLI IPERCUBI DEL CENSIMENTO PERMANENTE DELLA POPOLAZIONE



Fonte: Istat, 2020.



Ad esempio, per gli ipercubi relativi al grado di istruzione, alla condizione professionale e all'area geografica di cittadinanza sono descritte, di seguito, le variabili che ne definiscono la struttura e le fonti dei dati che hanno concorso al loro calcolo.

Ipercubo su grado di istruzione

Il titolo di studio più elevato conseguito da un individuo è stato rilevato sulla popolazione di 9 anni e più. Il corrispondente ipercubo, schematizzato nel prospetto 3, è interamente desunto dalle seguenti fonti di dati:

- il Registro Base degli Individui per ciò che concerne la variabile sesso;
- il Registro Tematico del Titolo di Studio per quanto riguarda la variabile grado di istruzione.

Quest'ultimo Registro è il risultato dell'integrazione tra le seguenti fonti di dati:

- a) la Base Informativa su Istruzione e Titoli di Studio (nella quale sono registrati i dati sui percorsi di studio e sui titoli conseguiti nel territorio italiano dal 2011 in poi);
- b) il 15° Censimento generale della popolazione e delle abitazioni del 2011;
- c) l'indagine campionaria del Censimento permanente.

Ipercubo su condizione professionale e non professionale

L'ipercubo sulla condizione professionale e non professionale è definito per la popolazione in età lavorativa (15 anni e oltre). Tale popolazione è ripartita in tre gruppi distinti: occupati, disoccupati, inattivi. Nell'applicazione di questi criteri viene seguito un principio gerarchico: prima si identificano gli occupati, successivamente, tra tutti i non occupati, le persone in cerca di occupazione e, infine, le persone inattive ovvero quelle non incluse tra gli occupati o i disoccupati. Gli occupati e le persone in cerca di occupazione costituiscono l'aggregato delle Forze di lavoro mentre le restanti categorie vanno a costituire l'aggregato delle Non forze di lavoro. Il prospetto 4 riporta le classificazioni delle variabili che definiscono l'ipercubo.

PROSPETTO 3. TAVOLA SINOTTICA DELL'IPERCUBO SUL GRADO DI ISTRUZIONE

Popolazione residente di 9 anni e più	Sesso (3) x Condizione (9)
Sesso (3)	Grado di istruzione (9)
Totale	Totale
Maschi	Analfabeti
Femmine	Alfabeti privi di titolo di studio
	Licenza di scuola elementare
	Licenza di scuola media inferiore o di avviamento professionale
	Diploma di scuola secondaria superiore (comprende la qualifica 2/3) anni, l'attestato IFP, la maturità di 4/5 anni e l'IFTS)
	Diploma di tecnico superiore ITS; Laurea o Diploma accademico AFAM di I livello; Diploma universitario (2-3 anni), Scuola diretta a fini speciali, altro diploma terziario non universitario
	Laurea magistrale/specialistica (biennale, a ciclo unico, diploma di laurea di 4-6 anni); Diploma accademico II livello (compresi i titoli del vecchio ordinamento - livello unico);
	Dottorato di ricerca/Diploma accademico di formazione alla ricerca

Fonte: Istat, 2020.



PROSPETTO 4. TAVOLA SINOTTICA DELL'IPERCUBO SU CONDIZIONE PROFESSIONALE E NON PROFESSIONALE

Popolazione residente di 15 anni e più		Sesso (3) x Condizione (9)	
Sesso (3)	Condizione professionale e non professionale (9)		
Totale	Totale		
Maschi	Forze di lavoro		
Femmine		Occupato/a	
		In cerca di occupazione	
	Non forze di lavoro		
		Percettore/rice di una o più pensioni per effetto di attività lavorativa precedente o di redditi da capitale	
		Studente/ssa	
	Casalingo/a		
	In altra condizione		

Fonte: Istat, 2020.

Il processo di stima dell'ipercubo prevede l'integrazione, attraverso un modello di stima, tra i dati campionari e le seguenti fonti di dati:

- ✓ il Registro Base degli Individui per ciò che concerne le variabili demografiche sesso, età e cittadinanza;
- ✓ il Registro Tematico del Titolo di Studio per quanto riguarda la variabile grado di istruzione;
- ✓ il Registro degli Occupati per le informazioni relative ai segnali amministrativi di lavoro, di pensione, di studio e le variabili legate al reddito e ai trasferimenti monetari per l'integrazione del reddito.

In particolare, il Registro degli Occupati integra in un'unica fonte di dati le informazioni provenienti dalle seguenti basi dati:

- ✓ il Registro Annuale sulle Retribuzioni e sui Costo del Lavoro a livello Individuale (in cui sono registrate le caratteristiche del rapporto di lavoro, le misure di input di lavoro, ossia ore retribuite, retribuzioni, oneri sociali, le caratteristiche del lavoratore e del datore di lavoro);
- ✓ il Registro Base degli Individui;
- ✓ le informazioni desunte dall'indagine campionaria della Rilevazione sulle Forze di Lavoro e quella del Censimento permanente.



Ipercubo su area geografica di cittadinanza

L'Ipercubo sull'area geografica di cittadinanza risponde all'esigenza di misurare la presenza e la cittadinanza degli stranieri abitualmente dimoranti sul territorio. Il prospetto 5 descrive in modo sintetico le caratteristiche dell'ipercubo rispetto alle variabili che lo definiscono: sesso e area geografica di cittadinanza. Entrambe le variabili sono completamente desunte dal Registro Base degli Individui.

PROSPETTO 5. TAVOLA SINOTTICA DELL'IPERCUBO SU AREA GEOGRAFICA DI CITTADINANZA

Popolazione straniera residente		Sesso (3) X Area geografica di cittadinanza (19)	
Sesso (3)	Area geografica di cittadinanza (19)		
Totale	Totale		
Maschi	Europa		
Femmine			Unione Europea 28
			Europa centro-orientale
			Altri paesi europei
	Africa		
			Africa settentrionale
			Africa occidentale
			Africa orientale
			Africa centro-meridionale
	Asia		
			Asia occidentale
			Asia centro-meridionale
			Asia orientale
	America		
			America settentrionale
			America centro-meridionale
Oceania			
Apolide			

Fonte: Istat, 2020.