

Nota metodologica

RICOSTRUZIONE DELLA POPOLAZIONE RESIDENTE

Per sesso, età e Comune
Anni 2002-2018

Da un punto di vista tecnico, una ricostruzione elimina per definizione quella discontinuità che viene a crearsi nella serie storica della popolazione ogni qual volta un nuovo Censimento determina la popolazione legale del Paese. Con riferimento al 31 dicembre 2018, punto di arrivo della presente ricostruzione, è stata riscontrata una differenza pari a 617 mila unità tra la popolazione aggiornata a calcolo¹ dall'Istat e la popolazione censita: precisamente 60 milioni 433 mila la prima contro 59 milioni 817 mila della seconda, per una differenza relativa pari all'1%. Alla base di tale differenza hanno concorso errori di copertura censuaria ed errori accumulati negli schedari anagrafici dei Comuni negli anni precedenti al Censimento.

Il legame tra dimensione demografica di un Comune e scostamento tra popolazione aggiornata a calcolo e censita è direttamente proporzionale (Prospetto A1). Nei comuni fino a 15 mila abitanti lo scostamento è sempre inferiore all'1%, con un minimo pari a 0,4% nei Comuni fino a 1000 residenti e via via a salire fino a un massimo di 0,9% in quelli tra 7mila e 15mila. Oltre tale soglia di residenti gli scostamenti crescono ulteriormente e il massimo si ha nei Comuni tra 30mila e 100mila abitanti con una differenza pari all'1,5%. Nei centri urbani oltre i 100mila residenti lo scostamento relativo è solo pari allo 0,9%, ciononostante tali Comuni raccolgono una differenza di oltre 123mila unità in senso assoluto e il 20% del suo totale.

PROSPETTO A1. CONFRONTO TRA POPOLAZIONE CENSITA E POPOLAZIONE AGGIORNATA A CALCOLO AL 31 DICEMBRE 2018, PER DIMENSIONE DEMOGRAFICA DEI COMUNI

Dimensione demografica (residenti)	Numero di comuni	Popolazione aggiornata	Popolazione censita	Differenza assoluta	Differenza percentuale	Distribuzione percentuale (Italia=100)
Fino a 999	1.960	1.068.849	1.064.920	3.929	0,4	0,6
Da 1.000 a 2.999	2.464	4.555.970	4.530.029	25.941	0,6	4,2
Da 3.000 a 6.999	1.715	8.010.929	7.954.681	56.248	0,7	9,1
Da 7.000 a 14.999	1.039	10.627.945	10.529.004	98.941	0,9	16,0
Da 15.000 a 29.999	428	8.875.666	8.762.088	113.578	1,3	18,4
Da 30.000 a 49.999	164	6.373.289	6.279.875	93.414	1,5	15,1
Da 50.000 a 99.999	99	6.761.292	6.660.141	101.151	1,5	16,4
Da 100.000 in su	45	14.159.420	14.035.935	123.485	0,9	20,0
ITALIA	7.914	60.433.360	59.816.673	616.687	1,0	100,0

Oltre alla dimensione demografica dei Comuni esistono altre grandezze cui guardare onde interpretare il senso della differenza accusata tra le due popolazioni, aggiornata e censita. Una di queste è, ad esempio, la dimensione geografica, dove lo scostamento si presenta più marcato tra le regioni del Centro-sud. Differenze rilevanti in termini relativi si riscontrano soprattutto per Lazio e Sicilia (1,9%), quindi in Calabria (1,7%), Puglia (1,4%) e Campania (1,3%).

¹ Per popolazione aggiornata a calcolo s'intende quella rilevata aggiornando il dato di popolazione rilevato all'ultimo censimento con le poste demografiche correnti (nascite, decessi, iscrizioni e cancellazioni per trasferimento di residenza) trasmesse all'Istat da ciascun Comune. In questa sede il confronto è quindi tra la popolazione censita nel 2011 aggiornata con le poste del periodo 2011-2018 e la popolazione censita a fine 2018.

Tra le regioni del Nord, soltanto la Liguria presenta uno scarto superiore al dato medio nazionale (1,2% contro 1%). In termini assoluti lo scostamento è concentrato in massima parte nelle principali regioni italiane: Lombardia, Lazio, Campania e Sicilia nell'insieme raccolgono il 58% delle complessive 617 mila unità di differenza. Anche la struttura per età e sesso rappresenta una chiave interpretativa fondamentale per comprendere gli scostamenti in essere alla data del censimento, momento in cui è massimo l'effetto della revisione statistica prodotta. Calcolando le differenze per singola classe di età si rileva un profilo molto simile a quello sottostante le migrazioni: più elevato nelle classi di età intermedie, più contenuto in quelle senili, su valori intermedi nelle classi di età giovanili. In pratica, le classi di età interessate da una più profonda revisione statistica sono quelle in età attiva (15-64 anni). Infine, una componente esplicativa parimenti importante è quella relativa alla cittadinanza della popolazione. Gli stranieri, in possesso di una struttura per età più giovane e con meno legami radicati nel luogo di insediamento rispetto ai cittadini italiani, detengono una più alta propensione alla mobilità, alla ricerca di opportunità lavorative e condizioni di vita migliori. Ciò comporta una loro connaturata complessità a risultare censiti come residenti nel Comune di dimora abituale. Tant'è che, considerando lo scarto tra popolazione aggiornata e censita al 31 dicembre 2018 tra i soli cittadini stranieri, si rileva un livello pari a 263 mila unità di differenza, pari al 43% della differenza accusata a livello totale (617mila), nonostante che la popolazione straniera non superi i complessivamente i 5 milioni di residenti.

Quadro di sintesi della metodologia adottata

Alla luce delle considerazioni di cui al precedente paragrafo, riguardanti l'esistenza di uno scarto tra popolazione aggiornata e popolazione censita e quali ne siano i fattori scatenanti, ricostruire la serie storica della popolazione tra due censimenti porta a includere necessariamente nell'analisi il territorio, l'età e la cittadinanza dei residenti. Ciò porta a compiere una serie di scelte operative.

L'unità territoriale di base della ricostruzione è il singolo Comune. I risultati a livello sovra comunale, dal provinciale al nazionale, si determinano per aggregazione. La ricostruzione interessa, dunque, 7.914 Comuni, omogeneamente classificati nel corso del periodo in esame, utilizzando la classificazione Istat vigente nel 2019. Ciò comporta che i Comuni che nel corso del periodo ricostruito siano stati oggetto di fusione amministrativa sono considerati unitariamente. Analogamente, Comuni che abbiano modificato Provincia di appartenenza sono considerati come se tale variazione fosse a regime dall'inizio del periodo di ricostruzione. Sono soltanto due, al contrario, i casi di Comuni sorti per scorporo da preesistenti Comuni: Baranzate (MI), sorto a fine 2003 da Bollate; Mappano (TO), sorto a fine 2016 da parti di Borgaro Torinese, Caselle Torinese, Settimo Torinese e Leini.

Le generazioni interessate alla ricostruzione sono quelle nate tra il 1901 e il 2018. La metodologia è, infatti, a carattere "longitudinale", ossia punta a ricostruire l'andamento demografico intercensuario di ciascuna generazione singolarmente, con la ricostruzione dei totali ottenuta in ogni anno per somma delle singole esperienze generazionali. Questo perché le cento e oltre generazioni di una popolazione sono diversamente esposte ai rischi di subire gli eventi demografici (morte o migrazione) e, conseguentemente, diversamente esposte alla possibilità di essere sotto o sopra enumerate tanto nei confronti dei registri anagrafici (in termini di iscrizioni e cancellazioni anagrafiche) quanto nei confronti del censimento (in termini di copertura). Sotto quest'ultimo punto di vista, considerando la relativa diversità dei rispettivi comportamenti demografici, è parimenti importante la distinzione tra residenti di cittadinanza italiana e straniera.

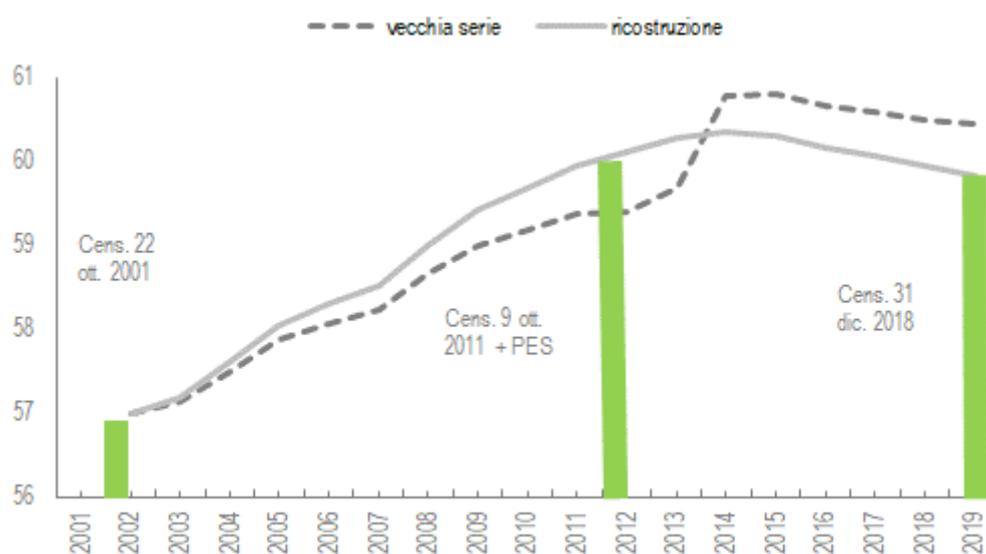
Il risultato finale del processo è che, correggendo alcune poste demografiche intercensuarie, lo scostamento rilevato a fine 2018 tra popolazione aggiornata e censimento viene annullato, distintamente per ciascuna delle 7.914 entità territoriali, nonché per genere, cittadinanza, e singola classe di età. Le poste demografiche oggetto di revisione sono le migrazioni, sia con l'interno sia con l'estero. Non sono invece soggette a revisione le componenti della dinamica naturale (nascite e decessi) e le acquisizioni della cittadinanza italiana.

Sul piano storico la ricostruzione copre il periodo dal 22 ottobre 2001 al 31 dicembre 2018. Qui, per facilità espositiva, si indica quale periodo il 2002-2018 quando ci si riferisce a dati di flusso. Quando invece ci si riferisce a dati di stock allora l'indicazione riportata è 1° gennaio 2002-2019, precisando che un dato di stock al 1° gennaio dell'anno t coincide in tutto e per tutto col dato al 31 dicembre dell'anno t-1. Gli elementi portanti della ricostruzione sono dunque i tre censimenti che hanno contraddistinto tale periodo, quelli del 2001 e del 2018, quali punti di partenza e di arrivo, e quello del 2011 quale punto intermedio. Di fatto, sul piano operativo le ricostruzioni sono due. La prima ricongiunge il censimento 2001 al censimento 2011, la seconda il censimento 2011 al censimento 2018.

Un'operazione preliminare alla ricostruzione medesima ha riguardato il censimento 2011. In effetti, in occasione di tale Censimento (9 ottobre 2011) si rilevò un notevole scarto tra la popolazione legale e quella che l'Istat aveva fino a quel momento potuto aggiornare a calcolo con i flussi anagrafici partendo dal Censimento 2001, ben 1 milione 352 mila individui in più a favore di questa seconda. Lo scarto così rilevante, dovuto ai noti errori di sovra e sottocopertura, comportò, secondo quanto previsto dall'art. 46 del Regolamento anagrafico, una folta fase di revisione e aggiornamento delle Anagrafi da parte dei Comuni nel periodo post-censuario. Con riferimento a quest'ultimo punto, la velocità con cui molti Comuni provvidero a rettificare le anagrafi comportò un notevole recupero di persone re-iscritte (in quanto non censite ma effettivamente residenti nel Comune alla data del Censimento) nel 2012 e soprattutto nel 2013, e quindi un repentino rialzo della popolazione residente nel periodo post-censuario dovuto non a fenomeni reali ma a procedure amministrative di rettifica delle posizioni anagrafiche. Ciò è particolarmente evidente guardando all'evoluzione della vecchia serie di popolazione tra il 2013 ed il 2014, dove in un solo anno si certificò un incremento (a calcolo) di circa 1 milione 100 mila individui su scala nazionale (Figura A1).

FIGURA A1. VECCHIA E NUOVA SERIE DELLA POPOLAZIONE RESIDENTE, ITALIA

Anni 2002-2019, dati al 1° gennaio, milioni (a)



(a) Nella vecchia serie le fonti sono la "Ricostruzione della popolazione residente 2001-2011" e la "Rilevazione su movimento e calcolo della popolazione residente" per gli anni 2012-2019.

La nuova ricostruzione punta invece a non distorcere le stime di popolazione con perturbazioni dovute a tali fenomeni. Da un punto di vista metodologico si è fatto ricorso ai risultati della *Post Enumeration Survey* (PES)². Di tale indagine, condotta su un campione di famiglie, il principale risultato fu una stima della sottocopertura netta del Censimento 2011 pari a 643 mila unità, pari all'1,1% della popolazione legale (al netto delle convivenze). I risultati della PES, quindi, sono in questa sede considerati come correttori statistici della popolazione legale 2011. In base a essi viene effettuata una correzione statistica³ della popolazione residente in famiglia al 9 ottobre 2011, che risulta pertanto pari alla somma della popolazione legale più la sottocopertura stimata dalla PES⁴.

Sul piano operativo la ricostruzione si articola in una serie di passi metodologici. Il primo di essi consiste nel produrre due aggiornamenti micro-fondati della popolazione per sesso, cittadinanza, anno di nascita e Comune di residenza alle date dei censimenti 2011 e 2018.

² La PES è normata dalla Commissione Europea (*Commission Regulation* n. 1151/2010).

³ La correzione ai dati censuari 2011 ha esclusive finalità statistiche e non modifica sul piano amministrativo e legale ciò che il Censimento sul campo ha rilevato.

⁴ La metodologia con la quale è stata modificata la consistenza del censimento 2011 con le risultanze PES è illustrata in appendice.

Il primo di tali conteggi usa quale stock di partenza la popolazione censita nel 2001 e via via effettua il conteggio della popolazione negli anni 2001-2011 considerando le poste demografiche di flusso in entrata e in uscita. Il secondo di tali conteggi adotta il medesimo procedimento ma parte dal censimento 2011 con lo scopo di conteggiare la popolazione a fine 2018.

Il secondo passo metodologico consiste nel confrontare la struttura comunale delle due popolazioni micro-fondate, in base alle caratteristiche del genere, della cittadinanza e della generazione, con quelle rilevate ai censimenti 2011 e 2018. Dal confronto tra i vettori di popolazione ai due riferimenti temporali emergono dei residui che rappresentano, al netto dei dati di flusso intermedi, il saldo demografico intercensuario cumulato di ciascuna generazione.

Il terzo e ultimo passo consiste nel ripartire nei due periodi intercensuari (2001-2011 e 2011-2018) i saldi residui suddetti, grazie ad un apposito modello ridistribuivo, dandone anche una spiegazione interpretativa. Ossia, quello che è un valore residuo all'apparenza neutro, frutto della differenza tra popolazione micro-fondata e censita, viene ricostruito sotto forma di immigrazioni o emigrazioni aggiuntive (con l'estero o con l'interno) che vanno ad integrare il bilancio demografico annuale di ciascun Comune.

Alla fine della procedura si ottengono facilmente quelli che rappresentano i principali risultati dell'esercizio e che, fondamentalmente, sono di due tipi: da un lato si riferiscono alla popolazione distinta per Comune, sesso, cittadinanza (nelle modalità italiana/straniera) e singolo anno di età al 1° gennaio 2002-2019; dall'altro si riferiscono ai bilanci demografici comunali ricostruiti in ciascuna annualità del periodo 2002-2018. Nei bilanci sono presenti misure di stock, relative al volume della popolazione all'inizio e alla fine di ciascun intervallo, e misure del movimento demografico, ossia nascite, decessi, immigrazioni ed emigrazioni interne, immigrazioni ed emigrazioni con l'estero, acquisizioni della cittadinanza italiana.

I paragrafi successivi illustrano in dettaglio quanto è stato fin qui sommariamente descritto.

Primo passo: aggiornamento micro-fondato della popolazione comunale tra due censimenti

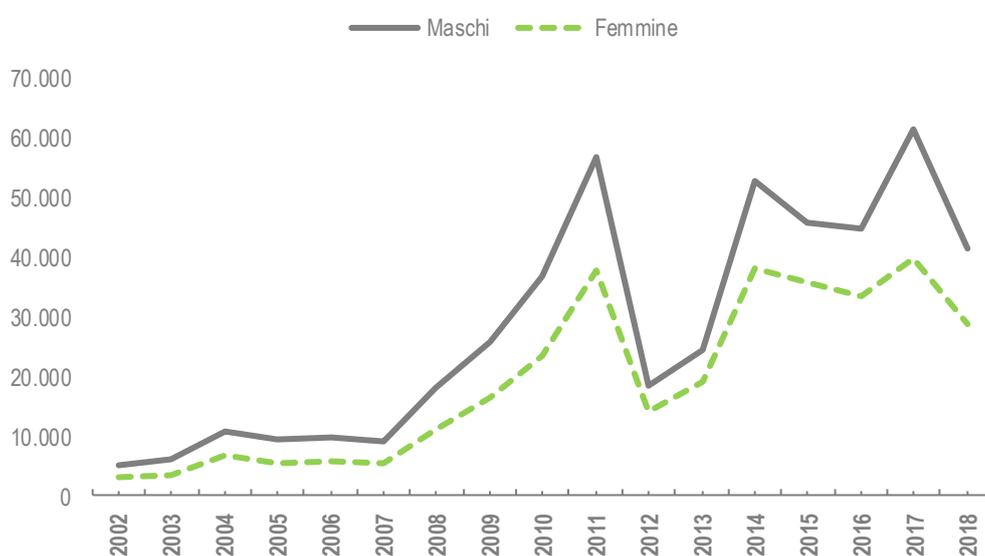
Ricostruire tra due censimenti equivale a tracciare il percorso demografico ottimale che ha caratterizzato una popolazione attraverso ciascuno dei suoi punti intermedi. Ai due censimenti sono note le caratteristiche strutturali della popolazione, ossia il suo ammontare distinto per sesso, anno di nascita, cittadinanza (italiana/straniera) e Comune di residenza. I punti intermedi del percorso si fanno corrispondere usualmente alla data del 31 dicembre dell'anno t (oppure, ma concettualmente è lo stesso, al 1° gennaio dell'anno $t+1$). Tra un punto intermedio e il successivo sono note le statistiche demografiche di flusso distribuite secondo le medesime caratteristiche strutturali.

Come si è visto nel paragrafo introduttivo, se si partisse dalla popolazione per sesso ed anno di nascita del censimento 2011, e aggiungendo a questa i flussi demografici che hanno avuto luogo durante il periodo intercensuario, distinti secondo le medesime caratteristiche, si ottiene una popolazione a calcolo finale che, tanto sotto il profilo dell'ammontare complessivo quanto sotto quello strutturale, è lontana da quella rilevata al censimento 2018. Errori di copertura censuaria ed errori di registro anagrafico concorrono a far sì che questo avvenga.

Le prime scelte operative sono dunque rivolte ad attenuare tale questione, riguardando una serie di interventi che nel loro insieme configurano una valutazione di completezza ed attendibilità delle statistiche di flusso e stock disponibili. In primo luogo, in merito alla struttura della popolazione per sesso ed anno di nascita rilevate ai censimenti. In questa sede vengono assunte per date, senza alcun intervento, le evidenze dei censimenti 2001 e 2018, ovvero i punti di partenza e arrivo del processo di ricostruzione. Invece, come già anticipato nel paragrafo precedente, il censimento 2011 viene corretto per una quota aggiuntiva pari alla sottocopertura netta evidenziata dalla PES, che viene ridistribuita su base comunale (si veda appendice).

Per quanto attiene i flussi demografici da considerare nei punti intermedi, nel processo ricostruttivo sono presi in considerazione: i nati, i decessi; le iscrizioni e cancellazioni anagrafiche per trasferimento di residenza con l'interno e con l'estero, le acquisizioni della cittadinanza italiana⁵. Sono dunque escluse tutte le altre voci che in anni correnti influenzano anch'esse il conteggio della popolazione e, in particolare, le iscrizioni e cancellazioni anagrafiche per altri motivi, ad eccezione di una particolare tipologia, ossia le cancellazioni d'ufficio per irreperibilità riguardanti cittadini stranieri. Per quanto la loro evoluzione del tempo si presenti in alcuni momenti occasionale, poiché dipendono dal periodo in cui una Anagrafe stabilisce di effettuare tutta una serie di controlli circa la effettiva residenza di propri cittadini, quelli inerenti i cittadini stranieri risultano oltremodo interessanti in quanto attenuano il problema della mancate cancellazioni per l'estero di tipo ordinario⁶. Peraltro, fatta eccezione nel biennio 2012-2013 dove si assiste a una loro contrazione ma che fisiologicamente è in linea col momento post-censuario, la serie storica presenta un'evoluzione crescente che è strettamente correlata all'aumento della presenza straniera in Italia (Figura A2).

FIGURA A2. CANCELLAZIONI PER IRREPERIBILITA' DI CITTADINI STRANIERI, ITALIA. Anni 2002-2018



Con riferimento alla ricostruzione prodotta per il periodo 2011-2018 siano, rispettivamente, $P_x^{9.10.11}$ e P_x^{18} le popolazioni censite ai censimenti del 9 ottobre 2011 e del 31 dicembre 2018 per la generica coorte di nascita x , prescindendo nella notazione dal Comune di residenza, dal genere e dalla cittadinanza⁷.

⁵ Vale la pena ricordare brevemente le fonti di tali informazioni, tutte Istat. Le statistiche sulle nascite provengono dalla rilevazione "Iscritti in anagrafe per nascita". Le statistiche sui decessi sono state recuperate dalla rilevazione sulle "Cause di morte" fino al 2010 e dalla rilevazione sui "Cancellati dall'anagrafe per decesso" dal 2011 in poi. Le statistiche relative ai trasferimenti di residenza sono desunte dalla rilevazione "Movimento migratorio della popolazione residente", le acquisizioni di cittadinanza provengono infine dalla rilevazione "Movimento e calcolo della popolazione residente straniera".

⁶ A differenza di un cittadino italiano, che nel momento di emigrare all'estero per un periodo prolungato ha interesse nel dichiarare il suo trasferimento iscrivendosi all'AIRE, un cittadino straniero non vive la medesima condizione. Non ha vincoli amministrativi, se non specificamente richiesti dal Paese di nuova accoglienza, e può trarre vantaggio dal mantenere una residenza fittizia in Italia. Da qui scaturisce, prevalentemente, il motivo per cui le cancellazioni per l'estero risultano ampiamente sotto enumerate in anni correnti.

⁷ Poiché la procedura di ricostruzione è metodologicamente identica, che si operi sul 2001-2011 o sul 2011-2018, da questo punto in poi la formulazione fa riferimento soltanto al secondo di tali intervalli.

Attraverso un procedimento iterativo si va a determinare alla fine di ciascun intervallo annuale una popolazione a calcolo sulla base dei microdati di flusso sopra specificati. In particolare, per una generica coorte presente a entrambi i censimenti (Figura A3) si deriva il suo ammontare calcolato al 31 dicembre 2011 mediante l'espressione:

$$C_x^{11} = P_x^{9.10.11} - m_x^{9.10.11-31.12.11} + ii_x^{9.10.11-31.12.11} - ci_x^{9.10.11-31.12.11} + ie_x^{9.10.11-31.12.11} - ce_x^{9.10.11-31.12.11} \pm ac_x^{9.10.11-31.12.11} - cr_x^{9.10.11-31.12.11}$$

ove a destra dell'uguale sono riportati, dopo la popolazione censita al 2011, i flussi demografici in entrata e in uscita del periodo 9 ottobre-31 dicembre 2011, tra cui, nello specifico: i decessi (m); gli iscritti da altri Comuni (ii); i cancellati per altri Comuni (ci); gli iscritti dall'estero (ie); i cancellati per l'estero (ce); le acquisizioni della cittadinanza italiana (ac); le cancellazioni per irreperibilità (cr).

Si osservi che il simbolo \pm davanti alle acquisizioni di cittadinanza sta a significare che si tratta di una componente positiva per i cittadini italiani e negativa per gli stranieri. Infine, come già anticipato, la componente *cr* riguarda i soli cittadini stranieri, per cui essa risulta nulla per gli italiani.

Con notazione analoga, la popolazione calcolata al 31 dicembre degli anni 2012-2018 si determina genericamente attraverso la seguente formulazione:

$$C_x^t = C_x^{t-1} - m_x^t + ii_x^t - ci_x^t + ie_x^t - ce_x^t \pm ac_x^t - cr_x^t$$

In definitiva per una data generazione "x" si è determinato il vettore:

$$C = \{C_x^{11}, C_x^{12}, C_x^{13}, C_x^{14}, C_x^{15}, C_x^{16}, C_x^{17}, C_x^{18}\}$$

esprimente l'evoluzione della generazione al 31 dicembre 2011-2018.

Per le generazioni intercensuarie si usa la stessa procedura ma ovviamente avviando l'aggiornamento della popolazione a partire dall'ammontare dei nati. In generale, per una generazione intercensuaria vale la seguente relazione iniziale:

$$C_x^t = n_x^t - m_x^t + ii_x^t - ci_x^t + ie_x^t - ce_x^t \pm ac_x^t - cr_x^t$$

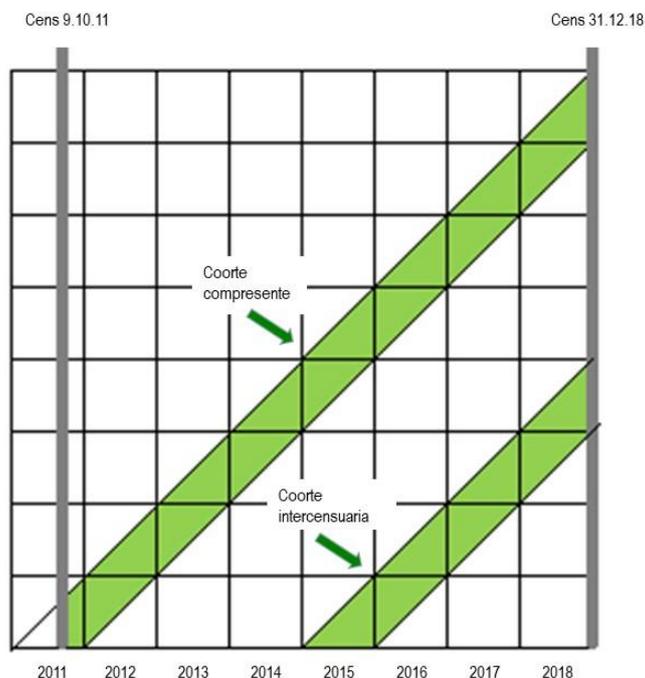
dove la notazione *n* rappresenta le nascite rilevate nell'anno.

Caso particolare è quello che riguarda la generazione 2011, la cui popolazione calcolata al 31 dicembre 2011 corrisponde a:

$$C_{11}^{11} = P_{11}^{9.10.11} + n_{11}^{9.10.11-31.12.11} - m_{11}^{9.10.11-31.12.11} + ii_{11}^{9.10.11-31.12.11} - ci_{11}^{9.10.11-31.12.11} + ie_{11}^{9.10.11-31.12.11} - ce_{11}^{9.10.11-31.12.11} \pm ac_{11}^{9.10.11-31.12.11} - cr_{11}^{9.10.11-31.12.11}$$

nella quale si tiene conto tra le poste demografiche anche dei nati nel periodo 9 ottobre-31 dicembre 2011.

FIGURA A3. RAPPRESENTAZIONE IN SCHEMA DI LEXIS DI UNA COORTE COMPRESENTE E DI UNA COORTE INTERCENSUARIA AI CENSIMENTI 2011 E 2018



Secondo passo: confronto tra popolazione calcolata e popolazione censita al 31 dicembre 2018

Dopo aver ottenuto a calcolo la popolazione di ciascuna generazione fino al 31 dicembre 2018, si può passare al confronto con la popolazione censita. Per ciascuna generazione si ha la differenza:

$$\Omega_x^{P-C} = P_x^{18} - C_x^{18}$$

ovvero una quantità residuale cumulata sulla quale, in fin dei conti, si concentra tutta la questione della ricostruzione intercensuaria. Infatti, il principio chiave alla base di ogni ricostruzione consiste nel ridistribuire tale saldo residuale di ogni generazione in ciascuno degli n intervalli intercensuari, pervenendo in ultima analisi alla determinazione della popolazione ricostruita al 31 dicembre di ogni anno. Le soluzioni per pervenire a tale risultato sono molteplici. Dalla più semplice, che consiste nel ridistribuire linearmente nel tempo i saldi tra popolazione censita e calcolata, a soluzioni più sofisticate come ad esempio quella di considerare parametri demografici per tenere conto del diverso profilo migratorio o del diverso livello di sopravvivenza tra le generazioni.

Il metodo redistributivo prescelto per la ricostruzione 2002-2018 è analogo a quello delle due ricostruzioni precedenti⁸, e sfrutta le potenzialità che derivano dall'avere a disposizione le evidenze prodotte dall'indagine corrente sulla struttura per età della popolazione: la rilevazione sulla popolazione residente comunale per sesso, anno di nascita e stato civile (Posas). Fino a fine 2018, istante da cui prende avvio il Censimento permanente della popolazione con produzione annuale di risultati, la rilevazione Posas è stata la fonte ufficiale in anni post-censuari (e fino a sua eventuale sostituzione con una ricostruzione, come nel presente studio) della struttura per età della

⁸ Le passate ricostruzioni Istat hanno utilizzato il principio di ridistribuire linearmente nel tempo i saldi generazionali residui, fino alla ricostruzione 1972-81. La ricostruzione 1982-91, invece, si basava su un principio redistributivo che teneva conto del diverso profilo migratorio tra le generazioni. Nella ricostruzione 1992-2001 si è invece passati a sfruttare le evidenze fornite dalla rilevazione corrente sulla struttura per età della popolazione. Approccio, questo, confermato nella ricostruzione 2001-2011, qui ormai superata dalla presente ricostruzione che si estende fino al 2018.

popolazione⁹. I suoi totali corrispondono all'aggiornamento a calcolo della popolazione residente, quelli che per intendersi hanno prodotto al 31 dicembre 2018 una differenza di 617mila unità su scala nazionale rispetto alla popolazione censita. Nonostante tale divario, la fonte Posas si rileva preziosa in quanto le sue variazioni annuali a livello di singola coorte si prestano, come si vedrà, a redistribuire negli anni il saldo residuale Ω_x^{p-c} .

Analiticamente, sulla base della rilevazione Posas l'evoluzione storica di una generazione tra la fine del 2011 e la fine del 2018 può essere rappresentata dal vettore:

$$R = \{R_x^{11}, R_x^{12}, R_x^{13}, R_x^{14}, R_x^{15}, R_x^{16}, R_x^{17}, R_x^{18}\}$$

È utile, a questo punto, esaminare la singola componente del vettore R, a confronto con quella del vettore C. La variazione annuale tra due popolazioni calcolate successive è pari a:

$$\Delta_{x,C}^t = C_x^t - C_x^{t-1} = -m_x^t + ii_x^t - ci_x^t + ie_x^t - ce_x^t \pm ac_x^t - cr_x^t$$

d'altra parte la variazione annuale tra due popolazioni Posas successive è invece pari a

$$\Delta_{x,R}^t = R_x^t - R_x^{t-1} = -m_x^t + ii_x^t - ci_x^t + ie_x^t - ce_x^t \pm ac_x^t - cr_x^t + am_x^t = \Delta_{x,C}^t + am_x^t$$

dove la quantità am_x^t rappresenta il saldo delle iscrizioni e cancellazioni anagrafiche per altri motivi non incluse nelle componenti di flusso che hanno determinato la popolazione calcolata¹⁰. Si può a questo punto facilmente verificare che la differenza tra popolazione Posas e popolazione calcolata è pari a

$$R_x^t - C_x^t = \sum_{k=11}^t am_x^k$$

dove, la quantità a destra dell'uguale è la variazione cumulata della rilevazione Posas tra il primo e il t-esimo intervallo intercensuario, al netto delle variazioni subite nella popolazione calcolata. Pertanto, a meno di tale fattore, gli elementi dei vettori C e R risultano, come è nelle aspettative, positivamente correlati. Questo risultato, tutto sommato semplice, si rivela particolarmente utile in quanto permetterà, come illustrato nel successivo paragrafo, di considerare le variazioni annuali della rilevazione Posas come proxy di quelle relative alla popolazione calcolata e, in conclusione, di redistribuire negli anni il saldo emerso dal confronto tra quest'ultima e la popolazione censita a fine 2018, ossia la quantità Ω_x^{p-c} .

Terzo passo: redistribuzione del saldo residuo e ricostruzione della popolazione per generazione

L'ultimo passo del procedimento di ricostruzione consiste nel distribuire le differenze rilevate tra i contingenti di popolazione, censiti e calcolati al 31 dicembre 2018, nell'arco del periodo intercensuario. Per comprendere il senso delle scelte operate, si vedrà come si è pervenuti alla ricostruzione di ciascuna generazione avvalendosi, passo dopo passo, dell'ausilio di un esempio pratico condotto sulla generazione 1997 del Comune di Roma. Per tale generazione il quadro delle informazioni sin qui descritte è quello riportato nel Prospetto A2.

⁹ Unitamente alla rilevazione Posas qui è considerata anche la rilevazione Strasa, quella cioè inerente la distribuzione della popolazione residente comunale per sesso ed anno di nascita di cittadinanza straniera. Per differenza tra le due rilevazioni si ottiene la distribuzione omologa dei residenti di cittadinanza italiana.

¹⁰ Oltre a questo, in tale quota residuale possono avere un peso anche altri fattori. In primo luogo quelli sottostanti le procedure di controllo e correzione dei dati in fase di validazione dei risultati delle indagini correnti annuali sulla popolazione residente (di stock e flusso). Un altro fattore è che in anni correnti esiste uno sfasamento nelle procedure amministrative inerenti i trasferimenti di residenza interni al territorio nazionale, che fanno sì che il numero delle iscrizioni non equivalga mai il numero delle cancellazioni come atteso. Ciò si deve alla diversa tempistica degli adempimenti statistici tra comune d'origine e comune di destinazione.

PROSPETTO A2. POPOLAZIONE DELLA COORTE 1997 RESIDENTE NEL COMUNE DI ROMA PER SESSO E CITTADINANZA. Anni 2011-2018

FONTE / ANNI	Italiani		Stranieri		Italiani		Stranieri	
	M	F	M	F	M	F	M	F
Censimento 2011	10.637	10.044	1.093	993				
Censimento 2018	11.353	10.835	1.804	1.162				
	POPOLAZIONE CALCOLATA				RILEVAZIONE POSAS			
2011	10.623	10.029	1.128	1.018	10.655	10.073	1.133	1.026
2012	10.630	10.059	1.286	1.135	10.712	10.227	1.366	1.271
2013	10.652	10.081	1.413	1.214	11.056	10.461	1.304	1.239
2014	10.696	10.131	1.541	1.261	11.115	10.531	1.411	1.231
2015	10.752	10.207	1.669	1.306	11.295	10.721	1.362	1.105
2016	10.793	10.255	1.930	1.348	11.422	10.826	1.509	1.078
2017	10.779	10.284	2.223	1.452	11.386	10.839	1.796	1.188
2018	10.759	10.296	2.445	1.593	11.347	10.821	1.991	1.277
Differenza 2018 censita – calcolata	594	539	-641	-431				

PROSPETTO A3. SALDI ANNUALI DELLA COORTE 1997 RESIDENTE NEL COMUNE DI ROMA PER SESSO E CITTADINANZA. Anni 2011-2018

ANNI	Italiani		Stranieri		Italiani		Stranieri	
	M	F	M	F	M	F	M	F
	POPOLAZIONE CALCOLATA				RILEVAZIONE POSAS			
2011 (a)	-14	-15	35	25	18	29	40	33
2012	7	30	158	117	57	154	233	245
2013	22	22	127	79	344	234	-62	-32
2014	44	50	128	47	59	70	107	-8
2015	56	76	128	45	180	190	-49	-126
2016	41	48	261	42	127	105	147	-27
2017	-14	29	293	104	-36	13	287	110
2018	-20	12	222	141	-39	-18	195	89

(a) Periodo 9 ottobre – 31 dicembre.

Nell'esempio, si osserva per i cittadini italiani (maschi o femmine) un residuo positivo tra popolazione censita e calcolata (rispettivamente pari a +594 e +539), mentre per gli stranieri se ne osserva uno negativo (-641 e -431). Si tratterà, quindi, nel primo caso di distribuire saldi positivi aggiuntivi nel periodo intercensuario a quella che è risultata la dinamica della popolazione calcolata, negativi nell'altro. Il Prospetto A3 illustra le variazioni annuali per la rilevazione Posas e per la popolazione calcolata.

Il saldo tra popolazione censita e popolazione calcolata (Ω_x^{P-C}), ossia la quota da distribuire nei singoli intervalli intercensuari per definire la ricostruzione per generazione, può assumere valori positivi, negativi o nulli. In quest'ultimo caso la popolazione ricostruita coincide, di fatto, con la popolazione calcolata e pertanto non sono necessarie ulteriori operazioni. Nel caso in cui il saldo tra popolazione censita e popolazione calcolata risulti diverso da zero si accorda l'ipotesi che i saldi annuali aggiuntivi siano sempre del medesimo segno. Pertanto, essi saranno sempre positivi se il saldo cumulato tra popolazione censita e calcolata è anch'esso tale (caso relativo agli italiani nell'esempio del Prospetto A2), viceversa assumeranno sempre segno negativo (come nell'esempio riferito agli stranieri).

Come si può osservare nel Prospetto A3, anche il saldo residuo Posas può alternare negli anni valori positivi a valori negativi. Un'operazione preliminare è stata quindi quella di ottenere una distribuzione modificata di tali saldi annuali,

nella quale tutti i valori presentino lo stesso segno del saldo cumulato tra popolazione censita e calcolata. Questa operazione, di fatto corrispondente a una “traslazione”, è un espediente strumentale cui si ricorre tramite una delle seguenti espressioni:

$$\Delta_{x,N}^t = \Delta_{x,R}^t + (\text{Max } \Delta_{x,R}^t - \text{Min } \Delta_{x,R}^t) \quad \text{se } \Omega_x^{P-C} > 0 \text{ con } \Delta_{x,R}^t \text{ positivi e negativi}$$

$$\Delta_{x,N}^t = \Delta_{x,R}^t \quad \text{se } \Omega_x^{P-C} > 0 \text{ con } \Delta_{x,R}^t \text{ positivi o nulli}$$

$$\Delta_{x,N}^t = \Delta_{x,R}^t - (\text{Min } \Delta_{x,R}^t + \text{Media } \Delta_{x,R}^t) \quad \text{se } \Omega_x^{P-C} > 0 \text{ con } \Delta_{x,R}^t \text{ negativi o nulli}$$

$$\Delta_{x,N}^t = \Delta_{x,R}^t - (\text{Max } \Delta_{x,R}^t - \text{Min } \Delta_{x,R}^t) \quad \text{se } \Omega_x^{P-C} < 0 \text{ con } \Delta_{x,R}^t \text{ positivi e negativi}$$

$$\Delta_{x,N}^t = \Delta_{x,R}^t \quad \text{se } \Omega_x^{P-C} < 0 \text{ con } \Delta_{x,R}^t \text{ negativi o nulli}$$

$$\Delta_{x,N}^t = \Delta_{x,R}^t - (\text{Max } \Delta_{x,R}^t + \text{Media } \Delta_{x,R}^t) \quad \text{se } \Omega_x^{P-C} < 0 \text{ con } \Delta_{x,R}^t \text{ positivi o nulli}$$

La Figura A4 presenta i risultati dell’operazione di traslazione applicata alla generazione 1997 del Comune di Roma, con la quale si ottiene per ciascun sesso e cittadinanza una nuova serie di saldi residui di segno identico a quello del saldo cumulato tra popolazione censita e calcolata (Ω_x^{P-C}). Nella medesima figura è rappresentato anche il saldo residuale ricostruito che, a sua volta, si ottiene in base all’espressione:

$$\Delta_{x,P}^t = \frac{\Delta_{x,N}^t \times g^t}{\sum_{k=11}^{18} (\Delta_{x,N}^k \times g^k)} \times \Omega_x^{P-C}$$

dove, di fatto, i saldi traslati sono riproporzionati sulla quantità Ω_x^{P-C} sottoponendoli preliminarmente a media ponderata con i giorni presenti nei vari anni (g^t), per tenere conto della loro diversa numerosità¹¹.

In questo modo si ha quindi che

$$\Omega_x^{P-C} = \sum_{k=11}^{18} \Delta_{x,P}^k$$

ossia, la differenza emersa tra popolazione censita e popolazione calcolata al censimento 2018 viene suddivisa in valori distinti, ognuno dei quali corrispondente a ciascuno degli intervalli intercensuari.

In Figura A4 si rileva che rispetto alle variazioni annuali Posas quelle ricostruite presentano una serie di caratteristiche interessanti. Si osserva, prima di tutto, che i valori ricostruiti sono sempre dello stesso segno (per costruzione), inoltre presentano un andamento nel tempo più regolare e si attestano mediamente su valori più bassi. La nuova serie mantiene, infine, una discreta somiglianza, a meno di fattori di scala variabili nel tempo, con quella da cui si è partiti.

Dopo la determinazione del saldo residuo ricostruito è possibile ricavare a calcolo la popolazione ricostruita al 31 dicembre di ogni anno. Quest’ultima è infatti data da:

$$P_x^{11} = P_x^{9.10.11} + \Delta_{x,C}^{9.10.11-31.12.11} + \Delta_{x,P}^{9.10.11-31.12.11}$$

con riferimento al 31 dicembre 2011, e in generale da

$$P_x^t = P_x^{t-1} + \Delta_{x,C}^t + \Delta_{x,P}^t$$

¹¹ Sono 84 quelli considerati nel 2011, quindi 366 nei bisestili e 365 nei rimanenti.

con riferimento al 31 dicembre 2012-2018, espressione che si riferisce ad una generazione comprese ai due censimenti. Il primo addendo a destra dell'uguale rappresenta la popolazione al 31 dicembre dell'anno precedente, il secondo rappresenta la componente incrementale ottenuta a calcolo, il terzo rappresenta la componente residua finora trattata. Formulazioni analoghe si hanno, con le opportune modifiche di notazione, nel caso di generazione intercensuaria.

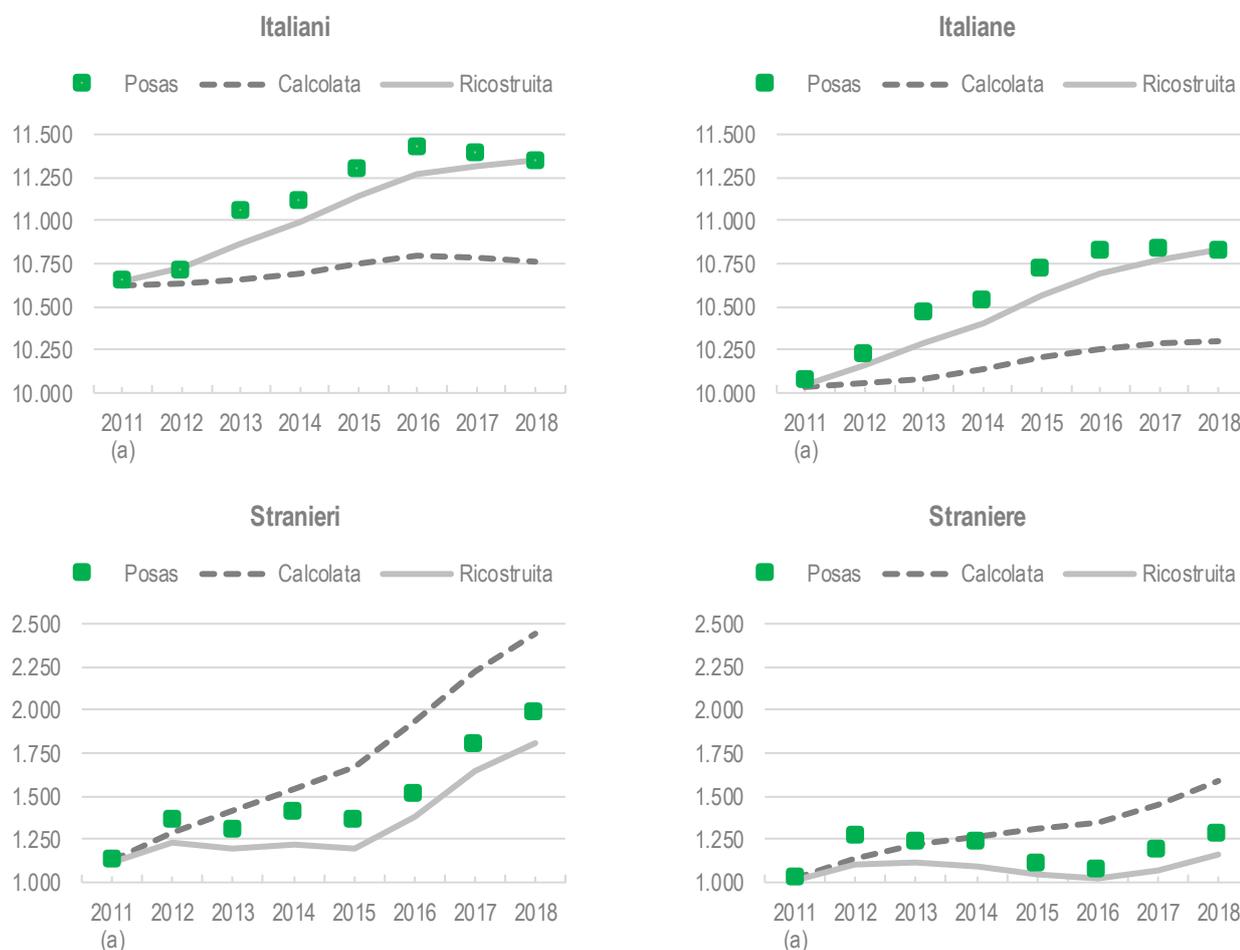
FIGURA A4. TRASLAZIONE DEL SALDO ANNUALE POSAS E SALDO RICOSTRUITO PER LA COORTE 1997 RESIDENTE NEL COMUNE DI ROMA. Anni 2011-2018



(a) Periodo 9 ottobre – 31 dicembre.

Tornando ora alla generazione 1997 residente nel Comune di Roma, fin qui presa ad esempio dimostrativo, la successiva Figura A5 ne evidenzia i risultati principali, rappresentando nell'ordine la popolazione rilevata in anni correnti (Posas), la popolazione ottenuta a calcolo e quella ricostruita.

FIGURA A5. POPOLAZIONE POSAS, CALCOLATA E RICOSTRUITA PER LA COORTE 1997 RESIDENTE NEL COMUNE DI ROMA. Anni 2011-2018



Quarto passo: esplicitazione del bilancio demografico intercensuario ricostruito

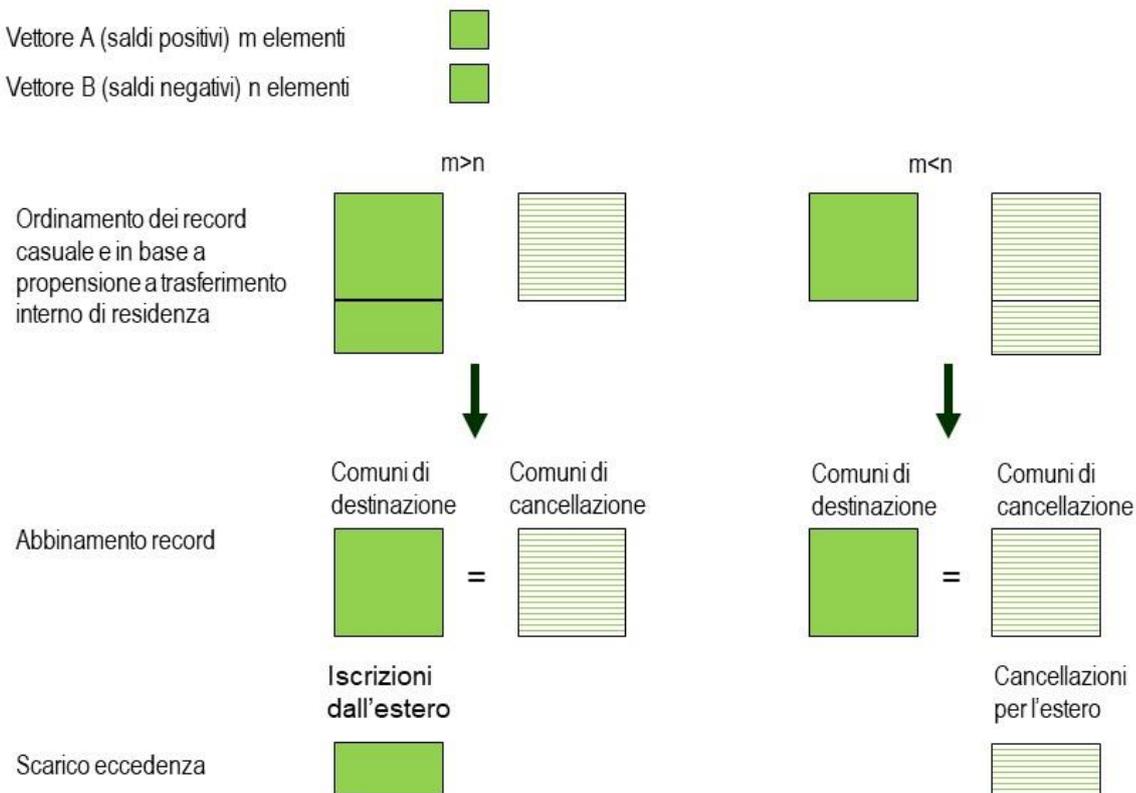
L'ultima parte del processo di ricostruzione è riservata all'interpretazione demografica del saldo additivo $\Delta_{x,p}^t$. Come si è visto, nei diversi intervalli intercensuari tale componente assume per costruzione sempre lo stesso segno, positivo o negativo. Da questo punto di vista, quindi, essa può interpretarsi come quota addizionale di iscrizioni o cancellazioni "statistiche" per trasferimento di residenza, in Italia o con l'estero, non contemplate tra quelle conteggiate nella popolazione calcolata e ufficialmente rilevate in anni correnti. Il meccanismo che permette di trasformare i saldi additivi $\Delta_{x,p}^t$ in potenziali iscrizioni/cancellazioni ulteriori opera (nell'ambito di ciascuna coorte, sesso e cittadinanza) sulla scala di tutti i 7.914 Comuni oggetto di ricostruzione contemporaneamente. In particolare, i Comuni vengono suddivisi in due distinti vettori, da un lato quelli con valori positivi dei saldi, dall'altro quelli con valori negativi.

Nell'ambito di ciascun vettore i Comuni entrano con un numero di osservazioni (record) pari all'entità del saldo additivo. Tali record vengono, quindi, ordinati prima casualmente e, successivamente, in base alla propensione storica posseduta da ciascun Comune nel registrare un trasferimento di residenza interno rispetto a un trasferimento di residenza con l'estero¹².

¹² Tale propensione media è stata calcolata sui dati reali relativi al periodo 2012-2018.

Ne risulteranno due vettori nei quali le posizioni in alto sono occupate da record casuali di Comuni in cui è più alta la propensione a un trasferimento interno mentre nelle posizioni in basso andranno a collocarsi record casuali di Comuni in cui tale propensione è inferiore. Si supponga adesso che il vettore di sinistra, quello dei Comuni con saldi additivi positivi abbia m elementi, mentre quello di destra per Comuni con saldi additivi negativi ne abbia n . A questo punto viene fatto un affiancamento record per record (merge) tra i due vettori, derivando così un numero di trasferimenti interni, con Comuni di origine e destinazione noti, pari alla lunghezza minima di record tra i due vettori (minimo tra m e n). I record in eccesso in coda, che non risultano in tale modo accoppiati, pari alla differenza tra i due vettori in termini di record posseduti ($m-n$ se $m > n$, $n-m$ se $n > m$) vengono assegnati a trasferimenti con l'estero. A seconda dei casi saranno assegnate $m-n$ iscrizioni dall'estero o $n-m$ cancellazioni per l'estero (Figura A6).

FIGURA A6. RAPPRESENTAZIONE SCHEMATICA DEL MECCANISMO DI TRASFORMAZIONE DEI SALDI RESIDUI RICOSTRUITI IN TRASFERIMENTI DI RESIDENZA INTERNI O CON L'ESTERO



Al termine della procedura si ottengono quote aggiuntive di flusso che rendono possibile esplicitare il bilancio demografico in ciascuna delle sue componenti. Cosicché, ad esempio, alle cancellazioni per altri Comuni ci_x^t rilevate negli anni correnti, già utilizzate nell'equazione della popolazione calcolata, andrà ad aggiungersi quando presente un'ulteriore quota di iscrizioni pari a $ci_{x.P}^t$. Nel bilancio demografico ricostruito il totale delle iscrizioni da altri Comuni sarà pari perciò a $ci_x^t + ci_{x.P}^t$. Analogamente risulterà per le iscrizioni da altri Comuni e dall'estero, dove i valori rappresentati saranno pari, rispettivamente, a $ii_x^t + ii_{x.P}^t$ e $ie_x^t + ie_{x.P}^t$. Infine, ricordando che le cancellazioni per irreperibilità (degli stranieri) sono equiparate a trasferimenti all'estero, le cancellazioni per trasferimento di residenza all'estero di bilancio complessive sono poste pari a $ce_x^t + ce_{x.P}^t + cr_x^t$.

L'effetto conclusivo del procedimento di esplicitazione delle quantità residuali $\Delta_{x,p}^t$ in poste del bilancio demografico è rappresentato, a livello nazionale indistintamente dalla coorte, dal genere e dalla cittadinanza nel successivo Prospetto A4. Tutte le poste risultano rivalutate al rialzo al termine del processo di ricostruzione. In particolar modo le cancellazioni per l'estero, rispetto alle quali si evidenziano nei vari anni incrementi da un minimo del 53% a un massimo del 202%. Un risultato, questo, che non sorprende poiché la questione del divario che si riscontra ogni qual volta in occasione di una tornata censuaria, quella cioè tra popolazione aggiornata correntemente a calcolo e popolazione censita (617 mila nel 2018), è una questione prevalentemente di mancate cancellazioni per l'estero negli anni precedenti al Censimento. Ma non solo ciò, poiché in occasione di un Censimento le tematiche della sottocopertura e della sovracopertura viaggiano in parallelo. Infatti, sono presenti incrementi, pur se di intensità minore, anche tra le iscrizioni dall'estero, che variano da un minimo dell'8% a un massimo del 38%, e tra i trasferimenti interni al territorio nazionale, che a loro volta variano da un minimo del 4% a un massimo del 17%.

PROSPETTO A4. TRASFERIMENTI DI RESIDENZA DA RILEVAZIONI CORRENTI E RICOSTRUITI PER TIPOLOGIA DI TRASFERIMENTO, ITALIA, Anni 2002-2018

Anni	Trasferimenti di residenza interni			Iscrizioni dall'estero			Cancellazioni per l'estero		
	Da rilevazioni correnti	Ricostruiti	Differenza (%)	Da rilevazioni correnti	Ricostruite	Differenza (%)	Da rilevazioni correnti	Ricostruite	Differenza (%)
2002	1.308.931	1.458.973	11,5	249.969	343.655	37,5	55.029	131.352	138,7
2003	1.314.393	1.468.772	11,7	495.038	595.705	20,3	66.426	127.688	92,2
2004	1.412.538	1.585.953	12,3	467.778	558.891	19,5	69.892	142.454	103,8
2005	1.443.247	1.608.581	11,5	340.184	415.245	22,1	68.795	157.335	128,7
2006	1.479.144	1.640.217	10,9	312.426	389.083	24,5	78.446	169.472	116,0
2007	1.463.215	1.632.812	11,6	576.176	650.089	12,8	68.333	152.177	122,7
2008	1.481.667	1.653.601	11,6	549.471	615.881	12,1	84.526	188.591	123,1
2009	1.385.691	1.552.324	12,0	453.975	513.650	13,1	82.926	221.120	166,6
2010	1.391.936	1.558.686	12,0	469.287	528.623	12,6	81.008	244.898	202,3
2011	1.438.520	1.501.523	4,4	403.038	449.515	11,5	88.013	246.010	179,5
2012	1.580.930	1.728.749	9,4	350.772	412.996	17,7	106.216	162.175	52,7
2013	1.386.617	1.617.229	16,6	307.454	362.871	18,0	125.735	207.827	65,3
2014	1.325.582	1.479.721	11,6	277.631	312.367	12,5	136.328	267.019	95,9
2015	1.317.696	1.475.842	12,0	280.078	306.101	9,3	146.955	276.095	87,9
2016	1.337.777	1.496.027	11,8	300.823	326.760	8,6	157.065	281.915	79,5
2017	1.342.659	1.499.637	11,7	343.440	370.804	8,0	155.110	308.859	99,1
2018	1.366.026	1.531.483	12,1	332.324	364.109	9,6	156.960	291.819	85,9

APPENDICE

Aggiustamento della popolazione legale al 9 ottobre 2011, per Comune, sesso, anno di nascita e cittadinanza sulla base dei correttori della *Post Enumeration Survey* (PES)

Come argomentato nei paragrafi precedenti, le evidenze del Censimento 2011 sono state oggetto di rivisitazione per via di un problema relativo alla sua eccessiva sottocopertura, che avrebbe dato luogo a una serie storica intercensuaria della popolazione fortemente perturbata. Da un punto di vista metodologico si è fatto ricorso, quindi, ai risultati della *Post Enumeration Survey* (PES), indagine campionaria che ha permesso di correggere al rialzo la popolazione legale di un'entità, su scala nazionale, pari alla stima della sua sottocopertura netta (643 mila unità, pari all'1,1% della popolazione legale, al netto delle convivenze). I passi successivi illustrano come a partire dai dati PES sia stato possibile correggere la popolazione legale, fino a livello di singolo Comune, sesso, cittadinanza (italiana/straniera) e anno di nascita.

Il primo obiettivo è stato quello di determinare l'ammontare di popolazione recuperabile a livello regionale.

Il Prospetto A5 illustra il confronto regionale tra Censimento e indagine PES rispetto alla popolazione residente in famiglia. I tassi di sottocopertura (al lordo della sovracopertura) per ciascuna regione ($r=1, \dots, 20$) sono definiti come segue:

$$\text{undercov}^r = \frac{P_{pes}^r - P_{cen}^r}{P_{pes}^r} \times 100$$

espressione che al numeratore contiene la differenza tra le popolazioni, rispettivamente, di fonte PES e censuaria. Tali tassi hanno dato origine ad una sottocopertura complessiva lorda di 871 mila individui residenti in famiglia a livello nazionale, per un tasso dell'1,45%.

La sovracopertura, ossia la rilevazione di unità non appartenenti al campo di osservazione del Censimento oppure la replicazione di unità già rilevate, è stimata su scala nazionale in 229 mila unità, per un tasso pari allo 0,38%. Il tasso di sovracopertura regionale (overcov^r) è definito dal rapporto tra le unità di sovracopertura e la popolazione lorda stimata dalla PES (moltiplicato 100).

PROSPETTO A5. POPOLAZIONE RESIDENTE IN FAMIGLIA E TASSI DI COPERTURA AL 9 OTTOBRE 2011

Regioni	Censimento	PES (stima al lordo sovracopertura)	Tasso di sottocopertura lordo	Sottocopertura	Tasso di sovracopertura	Sovracopertura	Tasso di sottocopertura netto
Piemonte	4.330.677	4.388.399	1,32	57.722	0,14	6.288	1,17
Valle d'Aosta	125.986	126.782	0,63	796	0,27	338	0,36
Lombardia	9.648.019	9.805.379	1,60	157.360	0,67	66.199	0,94
Trentino-Alto Adige	1.018.317	1.028.830	1,02	10.513	0,40	4.123	0,62
Veneto	4.817.382	4.872.438	1,13	55.056	0,46	22.457	0,67
Friuli-Venezia	1.208.413	1.214.516	0,50	6.103	0,19	2.255	0,32
Liguria	1.560.183	1.584.695	1,55	24.512	0,25	4.021	1,30
Emilia-Romagna	4.317.130	4.369.882	1,21	52.752	0,18	7.789	1,03
Toscana	3.655.690	3.722.904	1,81	67.214	0,11	4.318	1,69
Umbria	879.371	886.506	0,80	7.135	0,12	1.096	0,68
Marche	1.534.536	1.545.618	0,72	11.082	0,15	2.323	0,57
Lazio	5.474.328	5.590.105	2,07	115.777	0,47	26.824	1,60
Abruzzo	1.303.335	1.337.381	2,55	34.046	0,75	10.237	1,79
Molise	312.530	313.377	0,27	847	0,10	325	0,17
Campania	5.753.672	5.835.421	1,40	81.749	0,33	19.665	1,07
Puglia	4.042.867	4.122.902	1,94	80.035	0,67	28.147	1,27
Basilicata	576.420	578.951	0,44	2.531	0,18	1.042	0,26
Calabria	1.953.285	1.971.938	0,95	18.653	0,43	8.618	0,51
Sicilia	4.986.676	5.056.508	1,38	69.832	0,21	10.898	1,17
Sardegna	1.633.415	1.650.465	1,03	17.050	0,10	1.703	0,93
ITALIA	59.132.232	60.002.997	1,45	870.765	0,38	228.666	1,07

Il tasso di sottocopertura netto, che si ottiene sottraendo il tasso di sovracopertura al tasso di sottocopertura lordo, viene utilizzato per determinare l'ammontare di popolazione recuperabile a livello regionale:

$$Y^r = P_{pes}^r \times (\text{undercov}^r - \text{overcov}^r) / 100$$

Ottenuta tale quantità, si passa al secondo obiettivo, che consiste nel suddividere l'ammontare di popolazione regionale recuperabile tra residenti italiani e stranieri.

A tal fine si prendono in esame le informazioni PES inerenti la distribuzione dei residui, ovvero degli individui non abbinati al Censimento, per cittadinanza e regione (Prospetto A6). Definita come $q^{r,z}$ la quota di residui della regione r di cittadinanza z , la suddivisione per cittadinanza si ottiene moltiplicando tale quota per l'ammontare recuperabile di popolazione ottenuto al punto precedente:

$$Y^{r,z} = Y^r \times q^{r,z}$$

Sulla base di tale operazione emerge che a fronte di complessivi 642 mila residenti in famiglia recuperabili sul territorio nazionale¹³, 399 mila sono italiani mentre 243 mila sono stranieri.

PROSPETTO A6. DISTRIBUZIONE DEI RESIDUI PES (NON ABBINATI AL CENSIMENTO) PER CITTADINANZA E REGIONE

Regioni	residui campionari di italiani	residui campionari di stranieri	totale residui campionari	quota residui stranieri	Sottocopertura netta	Italiani recuperabili	Stranieri recuperabili
Piemonte	98	130	228	0,57	51.434	22.108	29.326
Valle d'Aosta	14	12	26	0,46	458	247	211
Lombardia	256	272	528	0,52	91.161	44.199	46.962
Trentino-Alto Adige	45	80	125	0,64	6.390	2.300	4.090
Veneto	77	96	173	0,55	32.599	14.509	18.090
Friuli-Venezia	20	33	53	0,62	3.848	1.452	2.396
Liguria	86	64	150	0,43	20.491	11.748	8.743
Emilia-Romagna	141	149	290	0,51	44.963	21.861	23.102
Toscana	170	118	288	0,41	62.896	37.126	25.770
Umbria	45	51	96	0,53	6.039	2.831	3.208
Marche	51	39	90	0,43	8.759	4.963	3.796
Lazio	394	190	584	0,33	88.953	60.013	28.940
Abruzzo	59	23	82	0,28	23.809	17.131	6.678
Molise	34	18	52	0,35	522	341	181
Campania	351	127	478	0,27	62.084	45.589	16.495
Puglia	165	31	196	0,16	51.888	43.681	8.207
Basilicata	12	17	29	0,59	1.489	616	873
Calabria	126	45	171	0,26	10.035	7.394	2.641
Sicilia	198	42	240	0,18	58.934	48.621	10.313
Sardegna	44	12	56	0,21	15.347	12.058	3.289
ITALIA	2.386	1.549	3.935	0,39	642.099	398.788	243.311

Nel passo successivo la quantità di popolazione recuperabile, distinta per regione e cittadinanza, viene ulteriormente articolata per singolo Comune. A tal fine si prendono in considerazione dati di fonte anagrafica (dalla rilevazione sul movimento e calcolo della popolazione residente) e, naturalmente, censuaria. L'attribuzione dei recuperi su base comunale ha approccio di tipo stocastico.

Alla data del Censimento 2011 sono note, per ciascun Comune, sia la popolazione censita sia la popolazione aggiornata a calcolo, quest'ultima ricavata per l'appunto "a calcolo" partendo dalla popolazione del Censimento 2001 e aggiungendo i flussi demografici intercorsi nel decennio intercensuario 2001-2011. A scopo illustrativo il successivo Prospetto A7 riporta un ipotetico esempio di una regione composta di soli 6 Comuni, in cui sono descritte le statistiche della popolazione calcolata e di quella censita.

¹³ Con questa operazione si porta la popolazione residente al 9 ottobre 2011 da 59.433.744 unità della Legale a 60.075.843 unità. Le unità in aggiunta riguardano solo residenti in famiglia, confermando i residenti in convivenza istituzionale di fonte censuaria.

PROSPETTO A7. POPOLAZIONE (CALCOLATA E CENSITA) PER CITTADINANZA NEI COMUNI DI UNA REGIONE FITTIZIA AL 9 OTTOBRE 2011

Comune	Italiani a calcolo	Stranieri a calcolo	Italiani censiti	Stranieri censiti	Differenza italiani	Differenza stranieri	Differenza italiani %	Differenza stranieri %
A	8.437	281	8.364	222	73	59	13,7	1,9
B	866	23	855	26	11	-3	2,1	0,0
C	13.015	347	12.772	250	243	97	45,6	3,1
D	5.827	120	5.812	100	15	20	2,8	0,6
E	2.048	38	2.042	35	6	3	1,1	0,1
F	186.370	18.937	186.185	15.938	185	2.999	34,7	94,4
Totale	216.563	19.746	216.030	16.571	533	3.175	100,0	100,0

L'idea di fondo alla base del meccanismo redistributivo è quello di considerare la distribuzione comunale dei residui per cittadinanza tra popolazione calcolata e popolazione legale (ultime due colonne di Prospetto A7) come proxy della sottocopertura con la quale redistribuire l'ammontare di popolazione recuperabile $Y^{r,z}$, ottenuta al passo precedente. In primo luogo va considerata la funzione di ripartizione delle frequenze relative dei residui. Ad esempio, dai dati del Prospetto A7, per i residenti italiani si ottiene la seguente distribuzione (ordinata per comodità in senso ascendente):

Comune	Frequenza relativa Italiani	Intervallo di ripartizione
E	1,1	0 – 1,1
B	2,1	1,1 – 3,2
D	2,8	3,2 – 6,0
A	13,7	6,0 – 19,7
F	34,7	19,7 – 54,4
C	45,6	54,4 – 100,0
Totale	100	

Il meccanismo di attribuzione dei recuperi regionali ai singoli Comuni è di tipo stocastico. Si genera, cioè, un numero casuale compreso tra 0 e 100 e si osserva in quale intervallo di ripartizione va a ricadere. Quindi si attribuisce un recupero al Comune nel cui dominio rientra il numero casuale estratto (ad esempio, se il numero generato casualmente fosse 45 si attribuirebbe un recupero al Comune F). Dopodiché, si ripete lo stesso meccanismo casuale per un numero di volte pari al totale dei recuperi regionali (ossia $Y^{r,z}$, con z=italiana in questo esempio) e si ottiene la distribuzione $Y^{c,z}$, ovvero la distribuzione territoriale su base comunale (c) dei recuperi da assegnare. A tale meccanismo, che vale in generale per tutti i Comuni, sono poste alcune limitazioni. Nei casi di Comuni con sovracopertura censuaria (cioè con popolazione censita superiore a quella calcolata, come nel caso degli stranieri del Comune B dell'esempio) non si procede ad alcun recupero. Il che comporta che alcuni Comuni possono non essere soggetti a recuperi (né per gli italiani né per gli stranieri) o possono esserlo soltanto in parte (solo per gli italiani o solo per gli stranieri). Inoltre, non partecipano al criterio di attribuzione dei residui i Comuni in cui una eventuale ripartizione su base proporzionale dei residui, anziché stocastica, non comporterebbe l'assegnazione di neanche un solo residuo aggiuntivo. Infine, solo per l'assegnazione di residui di stranieri, sono esclusi i Comuni in cui la popolazione straniera, di sesso maschile o femminile, non sia presente. Sulla base di questi criteri i Comuni che partecipano all'assegnazione dei residui, su base stocastica, scendono a 6512 per la popolazione italiana e a 6452 per quella straniera.

Il passo successivo consiste nel ridistribuire la popolazione comunale residua in base al genere e, dopo aver individuato la quantità $Y^{c,z}$, si punta a ricavare la quantità $Y^{c,z,s}$. Per raggiungere lo scopo si riprendono in esame i parametri regionali forniti dalla PES e, in particolare, le distribuzioni dei residui riportate nel successivo Prospetto A8. Separatamente per italiani e stranieri, da tali distribuzioni si ottengono le frequenze relative per sesso e, a partire da queste ultime, si creano gli intervalli di ripartizione¹⁴.

Il meccanismo di attribuzione dei recuperi in base al genere è simile a quello utilizzato al passo precedente, ossia ancora una volta di tipo stocastico. Si genera un numero casuale compreso tra 0 e 100 e si osserva in quale intervallo di ripartizione va a ricadere. Quindi si attribuisce un recupero al genere, maschile o femminile, nel cui dominio rientra il numero casuale estratto. A seguire, si ripete lo stesso meccanismo casuale per un numero di volte pari al totale dei recuperi comunali da ottenere (cioè $Y^{c,z}$) e si ricava la distribuzione richiesta $Y^{c,z,s}$, ossia la distribuzione per cittadinanza e sesso dei recuperi da attribuire a ciascun Comune.

PROSPETTO A8. DISTRIBUZIONE DEI RESIDUI PES (NON ABBINATI AL CENSIMENTO) PER CITTADINANZA, REGIONE E SESSO

Regioni	Italiani						Stranieri					
	Frequenza		Frequenza %		Intervallo di ripartizione		Frequenza		Frequenza %		Intervallo di ripartizione	
	M	F	M	F	M	F	M	F	M	F	M	F
Piemonte	56	42	57	43	0 - 57	57 - 100	66	64	51	49	0 - 51	51 - 100
Valle d'Aosta	8	6	57	43	0 - 57	57 - 100	4	8	33	67	0 - 33	33 - 100
Lombardia	132	124	52	48	0 - 52	52 - 100	148	124	54	46	0 - 54	54 - 100
Trentino-Alto Adige	23	22	51	49	0 - 51	51 - 100	40	40	50	50	0 - 50	50 - 100
Veneto	44	33	57	43	0 - 57	57 - 100	39	57	41	59	0 - 41	41 - 100
Friuli-Venezia Giulia	11	9	55	45	0 - 55	55 - 100	20	13	61	39	0 - 61	61 - 100
Liguria	42	44	49	51	0 - 49	49 - 100	40	24	63	37	0 - 63	63 - 100
Emilia-Romagna	84	57	60	40	0 - 60	60 - 100	73	76	49	51	0 - 49	49 - 100
Toscana	86	84	51	49	0 - 51	51 - 100	64	54	54	46	0 - 54	54 - 100
Umbria	19	26	42	58	0 - 42	42 - 100	26	25	51	49	0 - 51	51 - 100
Marche	24	27	47	53	0 - 47	47 - 100	19	20	49	51	0 - 49	49 - 100
Lazio	196	198	50	50	0 - 50	50 - 100	91	99	48	52	0 - 48	48 - 100
Abruzzo	32	27	54	46	0 - 54	54 - 100	9	14	39	61	0 - 39	39 - 100
Molise	22	12	65	35	0 - 65	65 - 100	10	8	56	44	0 - 56	56 - 100
Campania	167	184	48	52	0 - 48	48 - 100	67	60	53	47	0 - 53	53 - 100
Puglia	87	78	53	47	0 - 53	53 - 100	12	19	39	61	0 - 39	39 - 100
Basilicata	6	6	50	50	0 - 50	50 - 100	10	7	59	41	0 - 59	59 - 100
Calabria	70	56	56	44	0 - 56	56 - 100	21	24	47	53	0 - 47	47 - 100
Sicilia	109	89	55	45	0 - 55	55 - 100	21	21	50	50	0 - 50	50 - 100
Sardegna	27	17	61	39	0 - 61	61 - 100	5	7	42	58	0 - 42	42 - 100

Assegnate a ciascun Comune le quantità $Y^{c,z,s}$, l'ultimo passaggio consiste nel trasformare i dati per singolo anno di nascita (x), pervenendo a definire le quantità $Y^{c,z,s,x}$. Come nei due passi precedenti, relativi al Comune di residenza e al genere, si sfrutta un meccanismo di attribuzione casuale. L'elemento di riferimento da cui attingere le informazioni riguardanti la struttura per anno di nascita è la popolazione legale del Censimento, con tutte le caratteristiche strutturali sin qui considerate, ossia $P_{cen}^{c,z,s,x}$. Di tale vettore vengono prese in considerazione le distribuzioni di frequenze relative (in modo che faccia 100 il totale delle classi di nascita nell'ambito del singolo Comune, genere e

¹⁴ Si accorda, pertanto, l'ipotesi che la distribuzione dei residui per genere di fonte PES sia rappresentativa in ciascuna Regione dei suoi Comuni.

cittadinanza) e derivate le rispettive funzioni di ripartizione. Queste ultime, a loro volta fungeranno da urna casuale per l'estrazione di tanti numeri casuali, ognuno corrispondente a un anno di nascita, quanti sono necessari a coprire le quantità $Y^{c,z,s}$.

Riepilogando, nel primo passo della procedura si determina l'ammontare di popolazione da assegnare a ciascuna Regione, mentre nel secondo si suddivide ulteriormente questa quantità per Cittadinanza. Entrambi questi passi sfruttano le informazioni della PES. Nel terzo passo la popolazione regionale così recuperata viene distribuita per singolo Comune, ricorrendo agli scarti tra popolazione calcolata e popolazione censita. L'assegnazione dei singoli residui avviene, da questo punto in poi, con un meccanismo di assegnazione casuale. Il quarto passo distribuisce ulteriormente i residui comunali per Sesso, utilizzando dati regionali del campione PES. Il quinto passo perviene al risultato finale di ottenere la distribuzione dei residui per Comune, Cittadinanza, Sesso e Anno di nascita, utilizzando in questo caso la struttura del Censimento quale distribuzione di riferimento.

I primi due passi concorrono a definire la cornice di riferimento per l'aggiustamento della popolazione legale, fornendo risultati aggregati dei residui fissi, ovvero ottenuti in maniera deterministica. Dal terzo passo in poi, invece, si usa un meccanismo di assegnazione casuale che, nelle condizioni date, è reiterabile infinite volte. Sotto questo punto di vista, per pervenire all'assegnazione finale dei residui per Comune, Cittadinanza, Sesso e Anno di nascita, ovvero alla struttura definitiva del vettore $\hat{Y}^{c,z,s,x}$, si è deciso di ricorrere a una serie ampia di simulazioni, condotte sui passi 3-5, e di preferire, tra queste, quella che meglio rispondesse a un criterio di massima prossimità rispetto alla media delle n simulazioni¹⁵. In particolare, delle n simulazioni sono stati derivati per somma i risultati a livello nazionale mentre quale criterio di prossimità si è optato per "il minimo totale degli scarti rispetto alla media nazionale delle simulazioni". In notazione, la simulazione prescelta è stata quella che ha soddisfatto il seguente criterio:

$$\sum_{z,s,x} \left(Y_i^{z,s,x} - E(Y_i^{z,s,x}) \right)^2 = Min; \quad i = 1, \dots, n$$

in conclusione si perviene ad aggiustare la struttura per Comune, cittadinanza, sesso e anno di nascita della popolazione legale ponendo:

$$\hat{P}^{c,z,s,x} = P_{cen}^{c,z,s,x} + \hat{Y}^{c,z,s,x}$$

con $\hat{Y}^{c,z,s,x}$ pari al vettore con minimo livello nazionale di scarto rispetto alla distribuzione nazionale media delle simulazioni.

Per informazioni tecniche e metodologiche

Marco Marsili

marsili@istat.it

¹⁵ Si è adottato $n=30$.