

Nota metodologica

1) PREVISIONI DELLA POPOLAZIONE RESIDENTE PER SESSO, ETÀ E REGIONE

Le previsioni demografiche regionali dell'Istat sono costruite con l'obiettivo di rappresentare il possibile andamento futuro della popolazione, in termini di numerosità totale e struttura per età e sesso. Le informazioni prodotte rappresentano uno strumento importante a supporto delle decisioni nelle politiche, come quelle relative ai sistemi pensionistici, sanitari, scolastici e abitativi. Le previsioni sono aggiornate periodicamente riformulando le ipotesi evolutive sottostanti la fecondità, la sopravvivenza, i movimenti migratori internazionali e quelli interni.

Le previsioni in base 1° gennaio 2025 sostituiscono quelle in base 2024 pubblicate dall'Istat il 28 luglio 2025. Titolare e responsabile della produzione e della diffusione delle previsioni è l'Istat, come documentato nel Programma statistico nazionale. L'impianto metodologico alla base delle correnti previsioni è il medesimo di quello implementato negli anni precedenti. Con questo nuovo rilascio si apre un nuovo ciclo triennale di produzione, il quarto della serie dopo quelli 2016-2018, 2019-2021¹ e 2022-2024, che vedrà in successione il rilascio entro due anni delle previsioni in base 2026 e 2027. Ad ogni apertura di ciclo viene implementato il modello previsivo generale, il quale viene modificato nei due anni successivi solo per introdurre correttivi di breve termine oltre che per aggiornare la popolazione base. La metodologia nazionale venne definita, tra il 2009 e il 2015, da un gruppo di lavoro che vide la collaborazione di ricercatori dell'Istat e dell'Università Luigi Bocconi di Milano.

L'approccio metodologico, attorno al quale ruota tutto il modello previsivo, è di tipo semi-probabilistico. La caratteristica fondamentale delle previsioni probabilistiche è quella di considerare l'incertezza associata ai valori previsti, determinando gli intervalli di confidenza delle variabili demografiche e dando la possibilità all'utente di poter scegliere autonomamente il grado di fiducia da assegnare ai risultati.

Rispetto all'approccio "deterministico", più largamente utilizzato su scala internazionale e in passato adottato anche dall'Istat (fino alle previsioni in base 2011), si tratta di un avanzamento metodologico significativo. Infatti, nel modello deterministico l'utente finale non dispone di misure di probabilità associate ai risultati. Cosicché, ulteriore vantaggio del metodo probabilistico è costituito dal fatto che l'utente può cessare di confidare acriticamente sul lavoro dei *projection makers*, che con le varianti "basso/alto" tipiche dell'approccio a scenari deterministici definiscono a priori i confini alternativi alla variante ritenuta "più probabile", generalmente identificata come "scenario centrale".

La quantificazione dell'incertezza non rappresenta peraltro l'unico vantaggio del modello probabilistico. Se ne individua anche un altro che si deve alla più efficace rappresentazione della potenziale evoluzione di una popolazione. Nel modello probabilistico, infatti, gli scenari definibili sono infiniti sul piano teorico (per quanto nella realtà, come si vedrà più avanti, se ne selezionano sempre un numero finito), per cui è possibile che ipotesi di bassa sopravvivenza si mescolino con ipotesi di alta fecondità o medio livello delle migrazioni, o il contrario. Invece, le ipotesi degli scenari alto/basso dell'approccio deterministico sono definite perseguendo una logica *output oriented*: lo scenario alto contempla ipotesi di massimo incremento della sopravvivenza, della fecondità e delle migrazioni, mentre, all'opposto, lo scenario basso contempla solo ipotesi di minimo. La costruzione di tali scenari contrapposti coglie, in effetti, l'obiettivo di determinare un futuro campo di variazione per la popolazione e le sue componenti strutturali, ma si fonda su ipotesi concomitanti che hanno una scarsa possibilità di verificarsi.

Le sezioni successive della nota, oltre a contenere informazioni di carattere generale, illustrano in maniera sintetica i passaggi che hanno reso possibile la costruzione delle previsioni demografiche. Tali sezioni includono informazioni sui seguenti aspetti:

- popolazione base
- tecnica di proiezione
- periodo di previsione
- panel di esperti
- questionario degli esperti e modello probabilistico
- il questionario elettronico
- relazione tra le previsioni nazionali e regionali
- dati di base
- componente correttiva di *nowcasting*
- intervalli di confidenza e scenario mediano
- previsioni regionali di fecondità
- previsioni regionali di mortalità

¹ Le previsioni in base 2019, teoricamente appartenenti al secondo ciclo di produzione, saltarono per l'indisponibilità nei tempi necessari della popolazione base, per la prima volta nella storia derivata dal nuovo Censimento permanente della popolazione, nonché per la necessità di produrre una ricostruzione storica della popolazione per il periodo 2002-2018 e di ricalcolare tutti gli indicatori demografici di riferimento per le previsioni (in particolare, probabilità di morte e tassi specifici di fecondità).

- previsioni regionali dei movimenti migratori internazionali
- previsioni regionali delle migrazioni interne
- confronto con le precedenti previsioni
- confronto con le previsioni prodotte da Eurostat e Nazioni Unite
- procedura analitica di calcolo

Popolazione base

La popolazione base è quella per sesso, singola classe di età e regione al 1° gennaio 2025, così come identificata dal Censimento permanente della popolazione e delle abitazioni. La popolazione include tutte le persone usualmente residenti in Italia, di qualunque cittadinanza, mentre non include né cittadini italiani residenti all'estero, né cittadini illegalmente o irregolarmente presenti sul territorio nazionale.

Tecnica di proiezione

Le previsioni sono condotte con tecnica iterativa tra il 1° gennaio e il 31 dicembre di ogni anno, utilizzando il cosiddetto metodo per “coorti-componenti”. In corrispondenza di ciascuna classe di età alla popolazione iniziale vengono sommate le immigrazioni (dall'estero o da altre regioni) mentre vengono sottratti decessi ed emigrazioni (per l'estero o per altre regioni), ottenendo così la popolazione in vita alla fine dell'anno. A ciò vanno aggiunti i nati nel corso dell'anno che, al netto dei decessi e dei movimenti migratori che li riguardano, risultano ancora in vita al 31 dicembre.

Per la popolazione (*stock*) l'età è definita in anni compiuti al 1° gennaio (da 0 a 110 anni e più). Lo stesso dicasi per i dati di flusso come quelli sui nati, sui morti e sui movimenti migratori. Ciò permette di identificare, sempre e comunque, gli eventi demografici per anno di nascita dei soggetti coinvolti, assicurando la coerenza richiesta all'interno dell'equazione della popolazione.

Si assume che gli eventi demografici possano occorrere linearmente in qualunque momento dell'anno. Tra l'evento di morte e quello di migrazione (per l'interno o con l'estero) si assume che sussista incompatibilità, ossia che non possano coinvolgere un medesimo individuo nello stesso anno.

I decessi si determinano moltiplicando la popolazione residente per classi di età al 1° gennaio per le rispettive probabilità (prospettive) di morte, quelle che cioè interessano soggetti appartenenti alla stessa coorte di nascita.

Le nascite di un dato anno si ottengono in tre passaggi. Nel primo si moltiplica il contingente medio di donne relativo a ogni età feconda (ottenuto come media delle popolazioni di tal età all'inizio e alla fine dell'anno) per il rispettivo tasso di fecondità. Nel secondo si effettua la somma dei nati per età della madre, ottenendo il totale dei nati nell'anno. Nel terzo si scompongono i nati per sesso utilizzando il rapporto fisso di 106 nati di sesso maschile ogni 100 nati di sesso femminile.

Le previsioni hanno un profilo territoriale e sono costruite nella logica del modello multi-regionale, modello che, con particolare riguardo ai flussi migratori interni, tratta simultaneamente e coerentemente le distinte unità territoriali di riferimento. Il modello previsivo sulle migrazioni interne parte dalla costruzione di una matrice multi-regionale di probabilità migratorie per regione di origine, regione di destinazione, sesso, ed età. Tale matrice, applicata alla popolazione a rischio di migrare, identifica in ogni anno di previsione una serie coerente di immigrati ed emigrati.

Periodo di previsione

Le previsioni coprono il periodo tra il 1° gennaio 2025 e il 1° gennaio 2080. Scopo principale è fornire indicazioni sul futuro sviluppo della popolazione nel breve termine (2030), quindi quello di fornirne nel medio (2050) e lungo termine (2080), precisando che quest'ultime vanno utilizzate con cautela dal momento che i risultati diventano tanto più incerti quanto più ci si allontana dall'anno base. Tale rischio è tanto più concreto quanto più si cala l'attenzione sulle unità territoriali più piccole, come nel caso di alcune regioni italiane.

Panel di esperti

Un panel di esperti nazionali ha supportato l'Istat nella definizione delle ipotesi demografiche di lungo periodo relative all'Italia. Le ipotesi sub-nazionali, invece, sono state curate dall'Istat sulla base di un'apposita metodologia “ponte” tra le ipotesi nazionali e quelle territoriali. Gli esperti che hanno risposto al questionario (con tecnica CAWI), fornendo informazioni utili e complete a definire le ipotesi, sono stati 108. Essi sono stati reclutati volontariamente dall'Istat nel panorama degli studiosi in campo demografico-sociale e in quello dei policy-makers che per competenza professionale seguono con interesse gli sviluppi demografici del Paese. In particolare, si tratta di 60 donne e 48 uomini, principalmente occupati nelle Università (53) o in Ente di Ricerca pubblico (39). L'età media dei rispondenti è pari a 53 anni mentre la loro esperienza lavorativa è pari in media a 25 anni. Oltre a fornire supporto alla costruzione delle ipotesi demografiche di lungo termine, gli esperti hanno risposto nel medesimo questionario anche a domande sul futuro delle famiglie residenti in Italia e a quello inerente le forze lavoro, alimentando così anche tali previsioni derivate. L'indagine ha avuto luogo nel primo trimestre 2026, le sue risultanze di lungo termine sono prese a riferimento per tutto il ciclo triennale di lavorazione (ossia per queste in base '25 e per le previsioni in base '26 e '27 che saranno diffuse nei prossimi anni).

Questionario degli esperti e modello probabilistico

Il metodo probabilistico adottato poggia sulle opinioni degli esperti (*expert-based model*) per definire l'evoluzione futura di alcuni indicatori demografici di intensità. Il modello rientra nella classe più ampia dei modelli random scenario e viene utilizzato per la definizione degli scenari probabilistici a livello nazionale. Il metodo trova applicazione chiedendo agli esperti di elicitare dei valori attesi, da cui in seguito sia possibile ottenerne l'evoluzione futura su base stocastica. Gli esperti sono chiamati a fornire dei valori a un dato anno "t" riguardo a una serie di indicatori demografici di sintesi, condizionatamente ai valori assunti dagli stessi indicatori in istanti di tempo precedenti l'anno "t" (Billari, Graziani e Melilli, 2012).

Il metodo ha il vantaggio di essere semplice e flessibile. Nel questionario, infatti, le componenti demografiche richieste sono: il numero medio di figli per donna; la speranza di vita alla nascita distinta per sesso; le immigrazioni e le emigrazioni con l'estero. Le altre informazioni necessarie alla produzione delle previsioni, come quelle riguardo alla distribuzione per età (cadenza) degli eventi demografici, sono tenute appositamente fuori e lavorate successivamente al fine di rendere parsimonioso il questionario e il modello previsivo medesimo.

PROSPETTO A1. VALORI MEDI, VARIANZE E CORRELAZIONI OTTENUTE SOTTO LE IPOTESI DEGLI ESPERTI PER TIPO DI INDICATORE. Anni 2024, 2050 e 2080

Indicatore	Numero medio di figli per donna	Speranza di vita alla nascita – Uomini	Speranza di vita alla nascita – Donne	Immigrazioni dall'estero (migliaia)	Emigrazioni per l'estero (migliaia)
Anno 2024					
Valore osservato	1,18	81,5	85,6	451	189
Anno 2050					
Ipotesi media	1,25	84,2	87,6	373	169
Ipotesi alta	1,39	85,3	88,7	448	207
Varianza	0,012	0,843	0,743	3.382	902
Anno 2080					
Ipotesi media condizionata all'ipotesi media 2050	1,31	86,3	89,6	402	181
Ipotesi media condizionata all'ipotesi alta 2050	1,49	87,8	91,0	495	230
Ipotesi alta condizionata all'ipotesi media 2050	1,52	88,2	91,3	518	242
Varianza	0,046	3,312	2,971	13.482	3.686
Correlazione 2050-2080					
Coefficiente di correlazione	0,65	0,62	0,64	0,62	0,62

Per ogni indicatore demografico si prendono in considerazione due istanti temporali: un anno intermedio "t1" e un anno "t2" corrispondente all'ultimo anno di previsione. Nel questionario sottoposto agli esperti si considera "t0 = 2024", "t1 = 2050", "t2 = 2080", generando in questo modo due sotto-intervalli, 2024-2050 e 2050-2080. Indicare il valore della speranza di vita alla nascita nell'anno 2080, dato il valore previsto nel 2050, è un pratico esempio di come funziona il meccanismo.

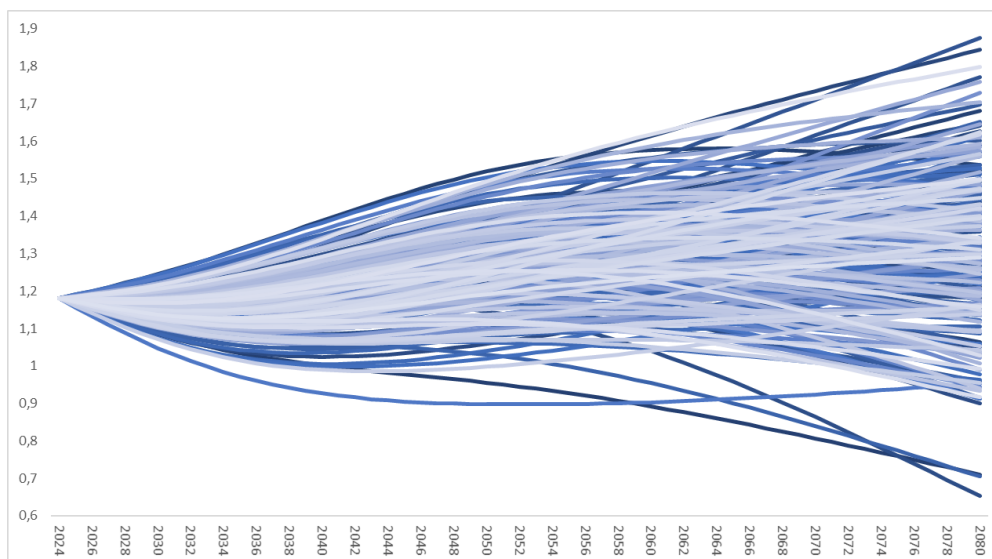
Gli indicatori demografici sono assunti, per semplicità di ipotesi, tra loro indipendenti (ad esempio, la fecondità non è influenzata dal livello delle migrazioni e viceversa), benché il modello permetta nella sua versione generalizzata la possibilità di far interagire tra loro le componenti demografiche in gioco. Si assume, inoltre, che la coppia di elicitazioni al 2050 e al 2080 di un dato indicatore possieda una distribuzione normale bivariata.

Sotto tali condizioni, una volta ottenuti i valori medi di risposta forniti dagli esperti è possibile stimare la varianza associata in ciascuno dei due istanti temporali futuri nonché la correlazione tra il primo e il secondo (Prospetto A1). Sulla base delle corrispondenti distribuzioni normali bivariate sono state quindi effettuate 3mila simulazioni col metodo Markov Chain Monte Carlo.

L'ultimo passo di stima è finalizzato al calcolo dei valori di ciascun parametro negli anni intermedi i due intervalli 2024-2050 e 2050-2080. Tale attività è realizzata, per ciascuna delle 3mila simulazioni, mediante interpolazione passante per i punti noti in corrispondenza degli anni 2024, 2050 e 2080. Le funzioni utilizzate per l'interpolazione sono di tipo quadratico per la mortalità e le migrazioni internazionali mentre si fa ricorso ad una funzione *spline* per ricavare i valori della fecondità. Si perviene quindi alla definizione di 3mila curve stocastiche per ogni indicatore demografico nazionale. A titolo di esempio, la Figura A1 descrive un fascio di 250 curve, ricavate dalla procedura sopra descritta e casualmente selezionate su 3mila, relative al numero medio di figli per donna.

La scelta di considerare un numero di 3mila simulazioni è frutto di un compromesso tra due necessità, entrambe strategiche: quella di rappresentare fedelmente l'aleatorietà degli eventi demografici e quella di ottimizzare i tempi macchina di elaborazione delle previsioni. Questi ultimi, nonostante la disponibilità di strumenti hardware/software sempre più potenti e sofisticati, costituiscono a loro volta un aspetto tecnico tutt'altro che secondario considerando la notevole mole di dati trattati.

FIGURA A1. EVOLUZIONE DEL NUMERO MEDIO DI FIGLI PER DONNA SULLA BASE DI 3.000 SIMULAZIONI STOCASTICHE OTTENUTE A PARTIRE DALLE VALUTAZIONI DEGLI ESPERTI. Anni 2024-2080



Il questionario elettronico Limesurvey

Per la realizzazione del questionario elettronico per la raccolta dei dati dagli esperti è stato utilizzato lo strumento *Limesurvey*, *software open-source*, distribuito sotto licenza *GNU General Public License (GPL)* e realizzato su piattaforma LAMP, per la creazione e la gestione di sondaggi online e questionari.

Esso permette, tramite un'interfaccia utente intuitiva, lo sviluppo rapido di questionari *web* e di gestire tutte le fasi successive della raccolta dati, dalla creazione della lista dei rispondenti e delle relative modalità di contatto, del monitoraggio della rilevazione, fino all'esportazione delle risposte. Sono consentite varie tipologie di domande, supporta sondaggi multilingue, ed è pienamente personalizzabile graficamente tramite *template* con *layout responsive*, cioè il cui contenuto si adatta alle dimensioni del *browser* del dispositivo utilizzato.

L'Istat ha iniziato da più di un decennio ad utilizzare il *software*, ospitandolo e aggiornandolo periodicamente sui propri server (esposti e interni), per diverse tipologie di processi di raccolta diretta dei dati in ambiente *web*.

Il questionario in oggetto, composto di una sezione anagrafica e di 7 sezioni tematiche (di cui 5 riservate alle previsioni demografiche, 1 alle previsioni delle famiglie e 1 alle previsioni della forza lavoro), è stato implementato rispettando fedelmente i controlli di coerenza e validazione tra i valori dei quesiti all'interno della stessa sezione e tra quelli appartenenti a sezioni differenti. A tal fine, è stato necessario riprogrammare (in javascript) l'interazione con l'utente per obbligarlo a inserire nei vari quesiti, di volta in volta, un opportuno numero di cifre decimali, personalizzare la tabella finale sulle posizioni familiari, nonché configurare alcuni aspetti grafici generali.

La lista degli esperti è stata precaricata sul sistema e a ognuno di essi è stato assegnato casualmente un codice univoco di partecipazione (*token*), tramite il quale viene composto il *link* (URL) per la compilazione del questionario. Ogni partecipante ha ricevuto nella propria casella di posta un'email di invito con tale *link* assieme a una breve informativa; lo stesso aveva la possibilità di accedere al questionario da qualsiasi *browser* e da qualsiasi dispositivo (*smartphone* compresi) e di compilare il questionario anche in diverse sessioni.

Il sistema, infine, ha permesso tramite il *backend* amministrativo di *Limesurvey* di sollecitare a più riprese i rispondenti che non avevano acceduto o completato il questionario.

Relazione tra le previsioni nazionali e regionali

Il modello probabilistico fornisce un set di 3mila simulazioni nazionali per ciascun indicatore demografico di sintesi. Poiché l'obiettivo delle previsioni Istat è anche quello di fornire indicazioni a livello territoriale, è stata implementata una procedura "ponte" tra la definizione degli *input* nazionali e quelli regionali. L'approccio è di natura *top-down* dal lato della costruzione delle ipotesi mentre, come si vedrà più avanti, è di natura *bottom-up* dal lato della produzione degli *output* finali.

L'azione principale è quella di derivare 3mila scenari regionali stocastici dai 3mila scenari nazionali frutto della consultazione tra gli esperti. La prima operazione in tal senso è quella di elaborare una previsione deterministica intermedia, applicando il modello multi-regionale a coorti componenti, includendovi anche la modellazione delle migrazioni interregionali. Da tale previsione, ottenuta estrapolando le tendenze regionali ritenute più probabili per

ciascuna componente (cfr. paragrafi successivi), vengono ricavati gli stessi indicatori di sintesi oggetto del modello stocastico precedentemente descritto, ossia numero medio di figli per donna, speranza di vita alla nascita e movimenti migratori con l'estero. Questa prima previsione intermedia, unica e deterministica, somiglia in sostanza a quella che in un approccio di tipo deterministico verrebbe etichettata col termine di "scenario centrale".

A questo punto, la transizione dal modello regionale deterministico al modello regionale stocastico si realizza moltiplicando, e reiterando 3mila volte il procedimento, la previsione dell'indicatore di sintesi deterministico regionale per il rapporto intercorrente tra la previsione nazionale stocastica e quella deterministica. In formula, indicando con "n" la generica simulazione ($n=1, \dots, 3000$), con "j" il codice territoriale regionale, con DR la previsione regionale deterministica, con SR quella stocastica, con DN e SN, rispettivamente, la previsione nazionale deterministica e stocastica, si ha:

$$SR_{t,n}^j = DR_t^j \times \frac{SN_{t,n}}{DN_t}$$

agganciando così, a ogni simulazione, il vettore dei valori regionali al valore nazionale stocastico di riferimento.

Si noti che per quanto concerne gli indicatori sintetici delle immigrazioni e delle emigrazioni con l'estero si ha che:

$$DN_t = \sum_j DR_t^j$$

$$SN_{t,n} = \sum_j SR_{t,n}^j$$

Ottenuti gli indicatori sintetici stocastici a livello regionale si passa alla costruzione degli *input* necessari all'applicazione del metodo per coorti-componenti, ossia le probabilità prospettive di morte per sesso ed età, i tassi specifici di fecondità per età della madre e la distribuzione degli immigrati/emigrati con l'estero per sesso ed età. La procedura associa quindi a ogni indicatore di sintesi la relativa cadenza del fenomeno (sesso ed età). Quest'ultima, non trattata in via stocastica, è quella che deriva dal modello deterministico regionale e, di simulazione in simulazione, riadattata allo specifico indicatore sintetico stocastico.

L'accoppiamento dei 3mila vettori di probabilità di morte (ciascun vettore sviluppa un numero di elementi pari a "numero di regioni X classi di età X sesso X anni di previsione") con i 3mila vettori di fecondità, e gli altrettanti sulle immigrazioni e le emigrazioni con l'estero e, infine, con le 3mila matrici O/D di probabilità di migrazione interna, è di tipo casuale.

Dopo aver introdotto una componente correttiva di *nowcasting* (cfr. paragrafo successivo) relativa ai primissimi anni di previsione, il modello coorti componenti viene fatto a quel punto girare 3mila volte, ottenendo così gli *output* richiesti: popolazione per sesso ed età, flussi demografici per sesso ed età, più la serie di indicatori demografici di supporto all'analisi che va dai tassi generici (di natalità, di mortalità ecc..) agli indicatori strutturali (età media, indice di dipendenza, di vecchiaia ecc..).

I risultati a livello nazionale (nonché di ripartizione geografica) nell'ambito di ciascuna simulazione regionale si ottengono per somma (approccio *bottom-up*). Pertanto, l'ammontare della popolazione prevista, dei decessi, e delle migrazioni, classificate per sesso ed età, e delle nascite per età della madre che si determinano a livello nazionale (o di ripartizione geografica) sono la sommatoria delle traiettorie previsive regionali. Anche gli assunti livelli nazionali (e di ripartizione) relativi agli indicatori di sintesi posti in diffusione, riguardo ad esempio la speranza di vita o il numero medio di figli per donna, sono ricalcolati ex-post sulla base di tali riepiloghi regionali.

È opportuno segnalare che la stocasticità introdotta a livello regionale, mutuata *top-down* da quella nazionale e limitata ai soli indicatori di sintesi, potrebbe risultare non sempre sufficiente a riprodurre l'aleatorietà dei vari eventi demografici. Ciò è particolarmente vero nelle piccole realtà territoriali dove l'incertezza tende a essere relativamente maggiore. Per questa ragione è più opportuno parlare di approccio semi-stocastico quando ci si riferisce alle previsioni regionali, per quanto la numerosità delle simulazioni condotte offra comunque ampia garanzia di rappresentatività della variabilità su scala regionale.

Una seconda osservazione riguarda il fatto che nel modello Istat è esclusa una trattazione statistica generalizzata della covarianza tra le Regioni (ad esempio: la previsione di incremento/decremento della fecondità in una data regione quanto condiziona o quanto è a sua volta condizionata dalla previsione di incremento/decremento in un'altra). A questa strada, esclusa anche per ragioni di parsimonia del modello statistico, se ne è preferita un'altra, quella della convergenza territoriale. Infatti, il modello regionale deterministico iniziale, successivamente trasformato in modello stocastico attraverso la procedura sopra descritta, è costruito su ipotesi di convergenza a lunghissimo termine (2125, ben oltre l'ultimo anno delle previsioni) tra le regioni per ciascuna componente demografica fondamentale. Ciò comporta che i 3mila scenari stocastici regionali raffigurano 3mila diverse ipotesi di convergenza dei comportamenti demografici sul territorio.

L'ipotesi principale alla base della convergenza è che le differenze socio-economiche e culturali attualmente in essere tra le regioni costituenti il territorio italiano siano destinate a scomparire nel lungo termine. Perciò, il loro progressivo annullamento comporterebbe anche un generalizzato riavvicinamento dei comportamenti demografici. L'idea di convergenza non è nuova in demografia ed esistono molti esempi di previsioni demografiche che la seguono (Eurostat e ONU, in particolare), tra cui anche quelle passate dell'Istat. Nelle previsioni Istat la convergenza è intesa come spostamento progressivo di un dato comportamento demografico verso un punto molto distante nel futuro che rappresenta l'istante di piena convergenza regionale (nel senso che a quel punto i valori risulterebbero identici per le diverse regioni), ma che in realtà è lungi dall'essere raggiunto all'interno dell'orizzonte previsivo considerato (2025-2080). Di fatto, risulta corretto in tale circostanza parlare più di modello di semi-convergenza che di modello di convergenza piena.

Dati di base

Le ipotesi definite a livello regionale nel modello deterministico preliminare, prima del passaggio al modello stocastico, sono state ottenute estrapolando i futuri trend dall'analisi delle serie storiche osservate. In particolare tali ipotesi sono state definite ricorrendo alle seguenti serie di dati:

- per la fecondità, i tassi specifici per età della madre del periodo 1977-2024;
- per la mortalità, le probabilità di morte per sesso ed età del periodo 1974-2024;
- per le migrazioni interne e internazionali, i trasferimenti di residenza per sesso ed età degli anni 2019 e 2021-2024 (escludendo quindi il 2020, anno della pandemia, per non influenzare le previsioni con un'annualità anomala ed episodica, durante la quale i trasferimenti di residenza sono stati frenati dalle misure di contenimento del Covid-19).

Componente correttiva di *nowcasting*

Prima di essere lanciate a pieno regime lungo l'orizzonte previsivo col metodo coorti-componenti, le previsioni incorporano un fattore correttivo di *nowcasting*². Con tale operazione si intende fare in modo che la previsione relativa ai primissimi anni risulti quanto più in linea alla tendenza emersa nell'ultimo periodo o nell'ultimo anno storico, pur preservando quella che è la potenziale variabilità del fenomeno studiato. Questo genere di operazione, che punta a contenere il cosiddetto effetto *jump-off*, si rivela particolarmente idonea in anni contraddistinti da improvvisi, e in quanto tali imprevedibili, cambiamenti della congiuntura demografica. È questo il caso, come avvenuto nel 2020 e in parte minore nel 2021, degli effetti provocati dalla pandemia Covid-19 su tutte le componenti del ricambio demografico. Non soltanto, per quanto prioritariamente, sulla mortalità, ma anche sulla natalità e sui trasferimenti di residenza interni e internazionali.

Dal momento che la popolazione base delle previsioni è quella rilevata al 1° gennaio 2025, si sono rese necessarie operazioni di correzione a breve termine degli *input* previsivi che hanno interessato i primi anni di previsione. La correzione, in particolare, sfrutta le informazioni del bilancio demografico provvisorio - Anno 2025, che l'Istat ha rilasciato nel marzo 2026³. Con ciò si vuole non solo tenere conto degli eventi demografici che hanno influenzato il 2025, ma anche quelli successivi entro i quali si presuppone che i suoi effetti possano esaurirsi e a far sì di ricordare la serie prevista degli *input* di breve termine con quelli di medio e lungo⁴.

Dal punto di vista computazionale la rivisitazione delle ipotesi di breve termine si realizza applicando dei fattori di correzione. Sia, ad esempio, E_b^j il numero di eventi demografici previsti nel primo anno sulla base dello scenario mediano nella regione j . E sia, invece, $\hat{E}_b^j$ il valore osservato di tali eventi oppure, in assenza del valore realmente osservato, la migliore stima che si possa ottenere (ad esempio, utilizzando procedure di *nowcasting* o analoghi modelli statistici). Il rapporto:

$$r_b^j = \hat{E}_b^j / E_b^j$$

rappresenta il fattore di correzione da applicare alle misure statistiche che danno luogo agli eventi di tipo "E" nell'anno "b" per la regione j . Se ad esempio tali eventi fossero il numero totale di nascite allora la quantità:

$$\hat{f}_{b,x}^{n,j} = r_b^j \cdot f_{b,x}^{n,j} \text{ con } x=14, \dots, 50 \text{ e } n=1, \dots, 3000$$

rappresenta la serie dei tassi specifici di fecondità per età della madre (n -esima simulazione) corretta per l'anno "b". Analoghe considerazioni valgono per la determinazione dei coefficienti di correzione inerenti la mortalità e i movimenti migratori. Per quanto concerne il 2025 i fattori di correzione sono stati costruiti rapportando i dati del

² Con il termine *nowcasting* (letteralmente "prevedere adesso"), contrapposto al più noto *forecasting* ("prevedere il futuro"), si intende la capacità di prevedere cosa sia accaduto nel recentissimo passato, anche in assenza di informazioni statistiche complete, e cosa possa accadere nell'immediato futuro. Si differenzia dall'approccio previsivo tradizionale perché si concentra sul breve termine (in genere non oltre cinque anni), piuttosto che su proiezioni a lungo termine. Poiché si basa sugli ultimi dati disponibili, anche se parziali (ad esempio, è possibile che per l'anno precedente l'anno base delle previsioni sia disponibile il totale dei decessi ma non ancora la loro articolazione per singola età) è in grado di fornire informazioni utili a comprendere quando lo scenario previsivo di fondo si discosta oltre una certa soglia dall'ultima realtà rilevata, dando così la possibilità di correggerne il tiro.

³ Cfr.: Istat, Indicatori demografici – anno 2025, <https://www.istat.it/comunicato-stampa/indicatori-demografici-anno-2025/>.

⁴ Tenendo peraltro presente il meccanismo iterativo di calcolo offerto dal metodo coorti componenti, ovvero un meccanismo di continua interazione stock-flussi nel tempo, la correzione imposta nei primi anni ha effetti strutturali anche sui risultati di tutti quelli successivi, fino al 2080 in questo caso.

bilancio demografico provvisorio di ciascuna regione, rilasciati nel marzo 2026 dall'Istat, alle previsioni preliminarmente prodotte per tale anno.

Per gli anni successivi al 2025 i fattori di correzione vengono applicati per un periodo limitato dell'orizzonte previsivo, elaborando pesi che tendono progressivamente a uno. In particolare, il numero di anni per cui il fattore di correzione si applica alla serie di interesse si desume da:

$$Y^j = \text{abs}(1 - r_b^j) \cdot \epsilon$$

con ϵ quantità arbitraria, scelta opportunamente per fare in modo di assicurare che il numero di anni per garantire il rientro dalle previsioni di breve a quelle di medio-lungo termine non superi i cinque anni in media regionale e non superi i dieci per una singola regione. A questo punto, i livelli dei fattori di correzione per gli anni successivi a "b", per un totale di "Y" anni, sono dati da:

$$r_t^j = \frac{r_b^j \cdot (b + Y^j - t) + (t - b)}{Y^j} \quad \text{con } t = b, b + 1, \dots, b + Y^j - 1$$

Anche le migrazioni interne incorporano un fattore correttivo di *nowcasting* ma con una procedura più articolata. In tal caso, infatti, si tratta di correggere una serie di matrici O/D di probabilità di migrazione per sesso, età, regione di origine e regione di destinazione, materialmente costruite sul quinquennio storico 2019 e 2021-2024 (2020 escluso, per non incorporare l'effetto Covid-19 sulla mobilità residenziale di tale annualità) alla luce delle analoghe informazioni riassuntive relative al 2025, provvisoriamente rilasciate dall'Istat nel marzo 2026. In estrema sintesi, l'utilizzo di cinque anni di migrazioni interne consente la costruzione di matrici O/D solide e rappresentative a livello di singola classe di età, anche in corrispondenza di quelle età o di quella realtà territoriali dove gli eventi migratori risultano più sporadici. Tuttavia, l'intensità complessivamente espressa da tali matrici (3mila in totale, tra loro variamente articolate sulla base della variabilità espressa nel quinquennio di loro costruzione) risente inevitabilmente dei livelli migratori medi di tale periodo, rappresentati dallo scenario mediano. La procedura di *nowcasting*, quindi, in questo caso consente di calibrare le matrici O/D finali sui livelli di intensità medi espressi nell'anno base (perché già rilevati o perché stimati).

Rimandando ai successivi paragrafi per quanto attiene la costruzione delle matrici O/D di probabilità costruite sul quinquennio storico sopra citato, la loro correzione è stata impostata come segue.

Prescindendo nella notazione dal sesso, rispetto al quale la procedura correttiva è identica, siano, rispettivamente, IMM_{25}^i ed EMI_{25}^i il totale di immigrati da altre regioni e il totale degli emigrati per altre regioni previsti nella regione "i" nel primo anno sulla base dello scenario mediano. E siano, invece, $\widehat{IMM}_{25}^i$ e $\widehat{EMI}_{25}^i$ gli analoghi valori realmente osservati in tale anno oppure, in assenza di essi, la loro migliore stima. Sia, infine, $m_{x,med}^{i,j}$ la probabilità prospettiva di migrare per un individuo di età "x" tra la regione "i" e la regione "j" nello scenario mediano. La quantità:

$$r^{ij} = \frac{\widehat{EMI}_{25}^i}{EMI_{25}^i} \times \frac{\widehat{IMM}_{25}^j}{IMM_{25}^j} \quad \forall (i,j) \text{ con } i \neq j$$

definisce l'elemento base di una matrice di 21x21 fattori correttivi (escludendo la diagonale principale che presenta valori unitari), che permette di parametrizzare l'intensità della matrice O/D delle probabilità migratorie per età dello scenario mediano, rendendola tale da riprodurre (approssimativamente) i valori osservati (o stimati) nell'anno base. In particolare, si avrà:

$$\widehat{EMI}_{25}^i = \sum_{j,x} \left({}_{25}P_x^i \cdot {}_{25}\lambda_x^i \cdot m_{x,med}^{ij} \cdot r^{ij} \right) \quad (5)$$

$$\widehat{IMM}_{25}^i = \sum_{j,x} \left({}_{25}P_x^j \cdot {}_{25}\lambda_x^j \cdot m_{x,med}^{ji} \cdot r^{ji} \right)$$

relazioni nelle quali l'espressione del tipo $P_x \cdot \lambda_x$ dentro la parentesi identifica gli esposti a rischio di migrare al netto della mortalità (si veda paragrafo finale sulla procedura di calcolo per ulteriori dettagli).

Ottenuto il fattore correttivo in questione, questo può essere applicato per correggere le 3mila matrici O/D prodotte

$$\widehat{m}_{x,s}^{ij} = r^{ij} \times m_{x,s}^{ij} \quad \forall x, s \text{ e } (i,j) \text{ con } i \neq j \text{ e } s = 1, \dots, 3000$$

Intervalli di confidenza e scenario mediano

Una volta lanciata la procedura di calcolo inerente 3mila simulazioni regionali, il margine di incertezza viene calcolato per tutti i possibili livelli informativi, dalla popolazione prevista ai dati di flusso, sulla base delle componenti strutturali del sesso e dell'età. Tali margini di incertezza dipendono a loro volta dall'incertezza insita nei futuri livelli di mortalità, fecondità e migratorietà che sono anch'essi resi disponibili. La diffusione dei risultati contempla il rilascio dei soli intervalli di confidenza del 90%, dell'80% e del 50% ma è possibile definire intervalli su qualunque scala di interesse. L'intervallo di confidenza restituisce l'informazione su quanto sia probabile che

un determinato indicatore demografico ricada entro prefissati limiti. Sotto questo punto di vista è opportuno ricordare che tale probabilità rappresenta essa stessa una previsione, in quanto fondata su ipotesi la cui validità è incerta. In nessun caso, inoltre, gli estremi dell'intervallo di confidenza vanno interpretati come estremi limiti, superiori o inferiori, del futuro comportamento demografico.

La costruzione di un intervallo di confidenza è qui fondata sulla determinazione dei percentili nella distribuzione delle 3mila simulazioni. Ad esempio, l'intervallo di confidenza al 90% per un dato indicatore è determinato considerando i valori della distribuzione che ricadono tra il 5° e il 95° percentile. Si ricorda, inoltre, che l'incertezza si riferisce sempre e comunque al dominio dello specifico parametro stimato. I limiti dell'intervallo di confidenza per un dato livello gerarchico sono stimati in proprio, e non costruiti per sommatoria di limiti ottenuti a un livello di disaggregazione gerarchicamente inferiore. Il criterio trova applicazione anche in contesti gerarchici non territoriali; ad esempio, nella composizione per età della popolazione o in quella per sesso⁵.

Con lo scopo di definire una previsione “puntuale” che possa essere adottata quale riferimento più probabile dell'evoluzione demografica futura è stato definito lo “scenario mediano”. Tale scenario corrisponde a una 3001-esima simulazione, ottenuta per costruzione, ma che di fatto non è stata rilevata nel campo di osservazione delle 3mila simulazioni. Il set di ipotesi viene identificato prendendo a riferimento il valore mediano tra tutte le simulazioni a livello delle singole componenti demografiche (fecondità, mortalità, migrazioni) nell'ambito delle possibili combinazioni delle covariate età, regione e anno di previsione. Ad esempio, il tasso specifico di fecondità all'età di 32 anni della regione Toscana, nell'anno 2065, dello scenario mediano è identificato come il valore mediano con tali caratteristiche individuato tra tutte le simulazioni. Il medesimo tasso specifico ma all'età successiva, o nell'anno successivo, è individuato con lo stesso procedimento ma esso scaturisce, verosimilmente, da una simulazione diversa. Per l'identificazione dello scenario mediano sulla mortalità e sulle migrazioni la procedura è identica ma con l'ulteriore covariata del sesso. Per le migrazioni interne, inoltre, le covariate territoriali riguardano la regione di origine e di destinazione.

Lo scenario è pertanto “mediano” dal lato degli *input* fondamentali. Dal punto di vista degli *output* (popolazione e flussi previsti) che tale scenario genera una volta lanciata la procedura per coorti-componenti, per le proprietà tipiche della mediana esso restituisce valori molto prossimi a quelli mediani.

Previsioni regionali di fecondità

Per la fecondità regionale le previsioni hanno riguardato i classici parametri di intensità e cadenza, ossia il numero medio di figli per donna e la distribuzione dei tassi specifici di fecondità per età della madre.

Il numero medio di figli per donna è stato rappresentato ricorrendo a modelli di tipo ARIMA(n,p,k) ricercando, distintamente per singola regione, quello più idoneo a prevedere l'intensità futura del comportamento riproduttivo, sulla base della serie storica 1977-2024. Il modello prevalentemente utilizzato è di tipo ARIMA(2,0,0) con intercetta.

Il profilo per età della fecondità è stato modellato ricorrendo a un sistema di funzioni *quadratic splines* (Schmertmann, 2003). Tale modello descrive funzionalmente la curva dei tassi specifici di fecondità standardizzati in funzione di tre parametri: l'età di inizio dell'età fertile α ; l'età P in cui la fecondità raggiunge il suo livello massimo; l'età H, successiva a P, nella quale la fecondità si dimezza rispetto al livello massimo. Per tasso specifico di fecondità standardizzato si intende il tasso specifico di fecondità normalizzato all'unità in corrispondenza del valore massimo individuato all'interno della sua distribuzione per età.

Il modello di *quadratic splines* adatta cinque polinomi di secondo grado alle curve di fecondità. La funzione finale risulta continua con la derivata prima anch'essa continua. Inoltre, grazie ad opportune restrizioni matematiche essa è univocamente determinata dai tre parametri $[\alpha, P, H]$ sopra menzionati.

In pratica, la previsione del tasso specifico di fecondità si trasforma nella previsione dei tre parametri (attraverso modelli ARIMA) che lo esprimono funzionalmente, una volta stimata la serie nel periodo 1977-2024. Per fare questo è stata adottata un'ipotesi di convergenza tra le regioni italiane, assumendo che le differenze territoriali in termini di comportamento riproduttivo tendano a diminuire nel lungo periodo. Dal punto di vista operativo la piena convergenza è stata fissata nel 2125. In particolare il vincolo di convergenza prevede che, dal 2025 al 2125, i parametri del vettore regionale $[\alpha, P, H]$ convergano linearmente ai valori di un ipotetico vettore nazionale, appositamente disegnato per l'operazione.

Previsioni regionali di mortalità

Le previsioni di mortalità sono state prodotte ricorrendo al modello di Lee-Carter (1992) nella variante proposta da Lee-Miller (2001), modello nel quale la procedura di *adjustment* riconduce le probabilità teoriche di morte a riprodurre precisamente il livello osservato della speranza di vita alla nascita, anziché il totale dei decessi osservati come nella versione originale. Inoltre, qui il modello viene applicato alla distribuzione per età delle probabilità di morte anziché a quella dei tassi specifici di mortalità della formulazione originale.

⁵ Ad esempio, l'intervallo di confidenza della popolazione totale è definito ricorrendo ai percentili della sua distribuzione e non alla somma dei limiti degli intervalli rilevati nelle singole classi di età; lo stesso dicasi per l'Italia, il cui intervallo è sempre costruito sulla sua distribuzione dei percentili e non sulla sommatoria dei limiti degli intervalli rilevati nelle singole regioni.

Il modello approssima la forma logaritmica delle probabilità di morte utilizzando tre parametri sintetici, di cui uno legato al *trend* $k(t)$ e due legati alla distribuzione per età $[a(x), b(x)]$.

Come per la fecondità, anche per la mortalità la costruzione del modello origina dalla definizione di uno scenario di riferimento provvisorio a livello nazionale. La previsione si determina proiettando nel futuro il solo parametro nazionale di trend $k(t)$, la cui serie è individuata sul periodo 1974-2024, mentre i parametri $a(x)$ e $b(x)$ rimangono in questa fase invariati nel tempo. In particolare, per via della sua sostanziale linearità il parametro $k(t)$ è stato proiettato al 2080 con la tecnica del *random walk with drift*.

Le ipotesi a livello regionale vengono fatte discendere dal provvisorio scenario di riferimento nazionale, stimando in primo luogo i valori regionali dei tre parametri nel 1974-2024 con la stessa metodologia e, successivamente, facendo convergere ogni parametro regionale al corrispondente parametro nazionale al 2125. Pertanto, come conseguenza del processo di convergenza e diversamente dall'impostazione classica del modello di Lee-Carter, qui si fanno variare nel tempo anche i parametri regionali $a(x)$ e $b(x)$.

Previsioni regionali dei movimenti migratori internazionali

Al fine di catturare le tendenze più recenti, le previsioni dei flussi migratori con l'estero concentrano l'analisi solo sugli ultimi cinque anni validi, vale a dire sul 2019 e sul 2021-2024. Il 2020 è stato volutamente censurato per non incorporare nelle previsioni gli effetti delle misure di *lockdown*. Il modello è piuttosto semplice nella sua formulazione, superando la complessità di ragionare sull'analisi di lunghe serie storiche per prevedere i flussi migratori internazionali. Senza dimenticare che a questo livello delle operazioni si tratta di strutturare un modello deterministico intermedio, i cui valori sono successivamente calibrati sulle intensità prodotte dal modello stocastico *expert-based*.

Nel primo anno di proiezione i valori totali delle immigrazioni e delle emigrazioni con l'estero sono posti pari al valore medio osservato nel corso degli ultimi cinque anni validi. In conformità con il quadro generale di convergenza del modello deterministico, si suppone quindi che in ciascuna regione ingressi e uscite convergano linearmente nel lungo periodo (2125) al medesimo livello, ossia alla semisomma iniziale dei due valori.

Una volta determinati i totali dei flussi in ingresso e in uscita fino al 2080, le distribuzioni per sesso ed età associate vengono derivate applicando il modello Castro-Rogers (Rogers et al., 1978) alla serie 2019 e 2021-2024. Con tale modello si dimostra che il caratteristico profilo per età delle migrazioni (qualunque esse siano, in ingresso o in uscita, con l'estero o con l'interno) può essere descritto, indipendentemente dall'intensità del fenomeno, da una funzione matematica composta di quattro componenti additive e fino a 11 parametri predittivi. Tali parametri, la cui stima nel periodo osservato viene prodotta grazie a una procedura generalizzata per modelli non lineari (categoria nella quale la funzione Castro-Rogers ricade), vengono tenuti costanti nel periodo di previsione. Il risultato conclusivo è dunque che l'intensità globale dei flussi migratori con l'estero possa variare nel tempo ma sulla base di una composizione per età fissa.

Previsioni regionali delle migrazioni interne

Le migrazioni interregionali sono trattate secondo un approccio multidimensionale, che permette di considerare simultaneamente le aree di origine e destinazione dei flussi migratori, e di definire gli ingressi in una determinata area come somma delle uscite con quella destinazione da tutte le altre aree del sistema. Il sistema è per costruzione coerente per tutti gli anni di previsione in quanto la riga e la colonna marginali della matrice O/D, corrispondenti rispettivamente ai flussi in entrata e in uscita in/da ciascuna regione, danno la stessa somma, corrispondente all'ammontare complessivo dei movimenti interni al territorio nazionale.

Come per le migrazioni internazionali, le previsioni dei flussi migratori interni concentrano l'analisi solo sugli ultimi cinque anni validi, vale a dire sul 2019 e sul 2021-2024, con il 2020 volutamente censurato.

Le probabilità di migrazione sono calcolate tenendo contemporaneamente conto dei livelli complessivi di mobilità interregionale espressi dalle singole regioni e delle strutture per età riferite a macroaree omogenee in quanto a comportamento migratorio. La metodologia di stima si basa sull'utilizzo di un duplice input iniziale; da un lato è predisposta una matrice origine/destinazione 21x21 contenente i trasferimenti interregionali, senza distinzione di età, dall'altro sono approntati dei vettori di probabilità di migrazione che, distintamente per singolo anno di età, esprimono la propensione a trasferirsi da una regione di una macroarea geografica verso un'altra regione della stessa macroarea o verso una regione di una macroarea differente.

Le macroaree geografiche – individuate grazie a una *cluster analysis* secondo il metodo aggregativo di Ward⁶ – raggruppano regioni che nel quinquennio storico in esame manifestano un comportamento migratorio omogeneo, consentendo il loro trattamento simultaneo nella fase di calcolo delle probabilità di migrazione per età. Le macroaree individuate sono 5 e precisamente:

- 1) Piemonte, Lombardia, Veneto, Friuli-VG, Emilia-Romagna, Toscana;
- 2) Umbria, Marche, Abruzzo, Sardegna;
- 3) Valle d'Aosta, Liguria, Prov. aut. di Bolzano e Trento;
- 4) Lazio;

⁶ Il metodo di Ward è un criterio iterativo applicato nell'analisi gerarchica dei cluster basato sul concetto di varianza minima. La funzione obiettivo da minimizzare, infatti, ad ogni passaggio in cui le unità statistiche da accoppiare vengono via via individuate e progressivamente raggruppate è la somma dei quadrati degli errori.

5) Campania, Puglia, Basilicata, Calabria, Sicilia.

La scelta di aggregare le regioni in macroaree di maggiori dimensioni, con il solo Lazio non agganciato a nessun'altra regione, consente di risolvere il problema delle bassissime frequenze dei trasferimenti interregionali in corrispondenza di alcune classi di età, che rendono difficoltoso il calcolo delle probabilità di migrazione.

Gli elementi dei vettori delle probabilità sono calcolati ponendo al numeratore il numero di trasferimenti di residenza interzonali classificati per regione di origine, regione di destinazione, sesso e singolo anno di età; al denominatore sono invece posti gli esposti al rischio di migrare cioè la popolazione per sesso età e regione, depurata dall'effetto della mortalità. Prescindendo nella notazione dal sesso, le probabilità di migrazione interzonali sono dunque calcolate in base alla formula:

$$t\mu_x^{hk} = \frac{\sum_{i,j} tM_x^{ij}}{\sum_i tP_x^i \left(1 - \frac{tq_x^i}{2}\right)} \quad \text{con} \quad \begin{cases} i \in h; i = 1, \dots, n(h) \\ j \in k; j = 1, \dots, n(k) \\ i \neq j; t = 2019, 2021, \dots, 2024 \end{cases}$$

dove al numeratore sono rappresentati i trasferimenti dalle regioni della macroarea h a quelle della macroarea k in età x, mentre al denominatore tP_x^i e tq_x^i rappresentano, rispettivamente, la popolazione di età x e la probabilità di morte all'età x della regione i (appartenente ad h). Si adotta l'ipotesi, infatti, che la migrazione e la morte siano distribuiti uniformemente nel corso dell'anno, per cui nella metà dei casi la morte interviene prima della migrazione, mentre nella restante metà la morte segue la migrazione.

I vettori di probabilità così ottenuti, a livello di ciascuna annualità, sono successivamente perequati utilizzando la funzione Castro-Rogers. Compiuta la costruzione delle probabilità migratorie interzonali, il passo successivo consiste nel calibrare tale set informativo sui livelli migratori specifici di ciascuna regione, affinché le probabilità migratorie finali possano esprimere un'intensità pari a quella che ciascuna regione ha potuto esprimere in ognuna delle 5 annualità considerate. La calibrazione delle probabilità viene effettuata ricorrendo a una matrice di coefficienti correttivi, i cui elementi, per ciascuna coppia (i,j) di regioni, sono pari a:

$$tC^{ij} = \frac{tM^{ij}}{\sum_x tP_x^i \cdot t\mu_x^{hk}} \quad \text{con } i \in h; j \in k; i \neq j; t = 2019, 2021, \dots, 2024$$

espressione che individua il rapporto esistente tra il reale flusso dei trasferimenti tra i e j nell'anno t e quelli attesi, secondo la condizione che la popolazione in i sia sottoposta alle probabilità migratorie tra le macroaree h e k, cui i e j afferiscono. A questo punto le probabilità migratorie finali si ottengono tramite la relazione:

$$t\mu_x^{ij} = tC^{ij} \cdot t\mu_x^{hk} \quad \text{con } i \in h; j \in k; i \neq j; t = 2019, 2021, \dots, 2024$$

tale probabilità specifica per età (80)⁷, regione di origine (21) e di destinazione (21) costituisce la componente elementare della matrice O/D composta di 80x21x21=35.280 celle per ciascuna annualità del quinquennio considerato.

Della variabile $t\mu_x^{ij}$ si assume che questa rappresenti una variabile casuale di tipo normale con media pari al valore medio del quinquennio e varianza pari alla varianza rilevata nel quinquennio:

$$\mu_x^{i,j} = E(t\mu_x^{ij})$$

$$\sigma_x^{i,j} = E(t\mu_x^{ij} - \mu_x^{i,j})^2$$

Con una tecnica random di tipo Montecarlo, dalle suddette variabili casuali di media e varianza vengono estratti 3mila valori per ciascuno degli elementi della matrice di O/D, dando così luogo alla generazione casuale di 3mila matrici tra loro diverse. La matrice O/D relativa allo scenario stocastico mediano viene identificata prendendo a riferimento il valore mediano tra tutte le simulazioni nell'ambito delle possibili combinazioni delle covariate, età, regione di origine e regione di destinazione. Tale matrice mediana è quella anche usata con lo scopo preliminare di produrre la previsione deterministica della popolazione, antecedente la transizione al modello stocastico vero e proprio (cfr. precedente paragrafo sulla relazione tra previsioni nazionali e regionali).

Si noti che nell'ambito di ciascuna simulazione (compresa quella relativa allo scenario mediano) la matrice O/D è supposta invariante nel tempo. Il che non comporta che i risultati prodotti dalle diverse simulazioni siano tutti uguali e costanti. Non solo le singole matrici O/D sono diverse l'una dall'altra, esse sono combinate in modalità del tutto casuale nei diversi scenari, e i flussi migratori interni che ne derivano evolvono nel tempo in virtù delle variazioni che interessano livello e struttura per età della popolazione esposta al rischio di migrare.

⁷ In quanto 80 sono le classi di età considerate sul quale è applicato il metodo, dall'età compiuta 0 all'età compiuta 79 (età espressa con riferimento l'inizio dell'anno). Per i nati nell'anno, mediamente esposti a rischio di migrare per sei mesi, si impone che la loro probabilità migratoria sia pari al 50% di quella che si ricava a 0 anni compiuti. Per le età 80-94 anni, invece, i valori assunti sono interpolati linearmente tra il valore previsto all'età di 79 anni e il valore all'età di 95 anni, supposto essere pari a 0. Ne consegue che dopo i 95 anni e fino a 120 anni e oltre, tutte le probabilità migratorie sono supposte nulle. Nel complesso, tenuto conto di ciò, la matrice O/D finale esprime 122x21x21 elementi, nella quale fino a 79 anni sono presenti valori del modello di previsione, da 80 a 94 anni valori interpolati linearmente e, infine, da 95 anni in su e sulla diagonale principale (laddove i=j) soltanto valori nulli.

Confronto con le precedenti previsioni Istat

Una valutazione del cambiamento occorso tra i due ultimi round previsivi può essere effettuata confrontando gli scenari mediani delle previsioni in base '24 e '25.

In primo luogo, va rilevata una differenza quasi insignificante tra la popolazione totale base '25 (58 milioni 943mila) e quella che era stata stimata nello scenario mediano alla stessa data dalle previsioni in base '24 (58 milioni 934mila), frutto perlopiù degli aggiustamenti statistici ex-post intervenuti nel corso della validazione dei dati censuari in termini di sotto e sovra copertura.

Dal lato dei flussi previsti nel tratto di comune proiezione (2025-2080) si coglie un numero di nascite molto più basso nell'esercizio base '25 (18,1 contro 20,1 milioni). Riguardo ai decessi si ha invece un numero di decessi più alto (43,5 milioni rispetto a 43,0 milioni). Le previsioni sulle poste migratorie con l'estero sono invece molto diverse: il saldo migratorio netto globale è pari a 12,1 milioni di individui nelle previsioni in base '25, molto più alto dei 9,6 milioni previsto con l'esercizio base '24.

PROSPETTO A2. CONFRONTO TRA LE IPOTESI SUI PRINCIPALI INDICATORI SINTETICI DELLO SCENARIO MEDIANO IN BASE 2024 E DELLO SCENARIO MEDIANO IN BASE 2025. Anni 2025, 2030, 2050 e 2080

Scenario mediano	Numero medio di figli per donna	Speranza di vita alla nascita – Uomini	Speranza di vita alla nascita – Donne	Immigrazioni dall'estero (migliaia)	Emigrazioni per l'estero (migliaia)
Anno 2025					
Base 2024	1,20	81,8	85,7	435	185
Base 2025	1,14	82,0	85,8	440	144
Anno 2030					
Base 2024	1,26	82,0	85,9	357	161
Base 2025	1,16	82,3	86,1	423	176
Anno 2050					
Base 2024	1,38	84,3	87,8	301	136
Base 2025	1,25	84,1	87,6	372	169
Anno 2080					
Base 2024	1,46	86,1	89,7	305	143
Base 2025	1,30	86,2	89,6	404	183

Gli effetti combinati delle differenti entità dei flussi dal 2025 al 2080 determinano una differenza nulla tra le popolazioni finali dei due distinti esercizi previsivi (circa 40mila unità in più al 1° gennaio 2080 per lo scenario mediano base '24). Sotto questo punto di vista il prospetto A2 mette in risalto come il processo di ridefinizione delle ipotesi per tutte le componenti demografiche abbia interessato profondamente la fecondità e le migrazioni internazionali, sia nel breve che nel medio e lungo periodo. Nel nuovo rilascio si avranno meno nati e più emigrati ma anche molti più immigrati, cosicché la popolazione prevista al 2080 è pari a 45,8 milioni sia nell'esercizio base '24 che nell'esercizio base '25.

Confronto con le previsioni realizzate da Eurostat e Nazioni Unite

Per confrontare le previsioni prodotte dall'Istat con quelle di altri Enti ha senso prendere a riferimento le previsioni rilasciate dall'Eurostat e dalla *United Nations Population Division* (UNPD). Da anni, l'istituto statistico dell'Unione europea assolve il compito di produrre previsioni demografiche con cadenza regolare per tutti i Paesi membri. Le ultime rilasciate sono in base '25, il cui principale scenario di riferimento è il cosiddetto *baseline scenario*. Anche la UNPD, a sua volta, produce previsioni demografiche con regolarità attraverso i *World Population Prospects*, all'interno dei quali sono contemplati tutti i Paesi del globo. Le sue ultime previsioni disponibili sono in base '24 e il principale scenario di riferimento è la *medium variant*.

Occorre evidenziare in premessa che, nonostante la comparabilità sul piano della tecnica proiettiva, gli esercizi prodotti dai due organismi internazionali presentano alcune differenze metodologiche rispetto a quello italiano. Tra queste, il fatto che i due modelli internazionali qui presi in esame sono di tipo uni-nazionale, ossia proiettano la popolazione residente in Italia nel suo insieme non tenendo conto dello sviluppo delle regioni.

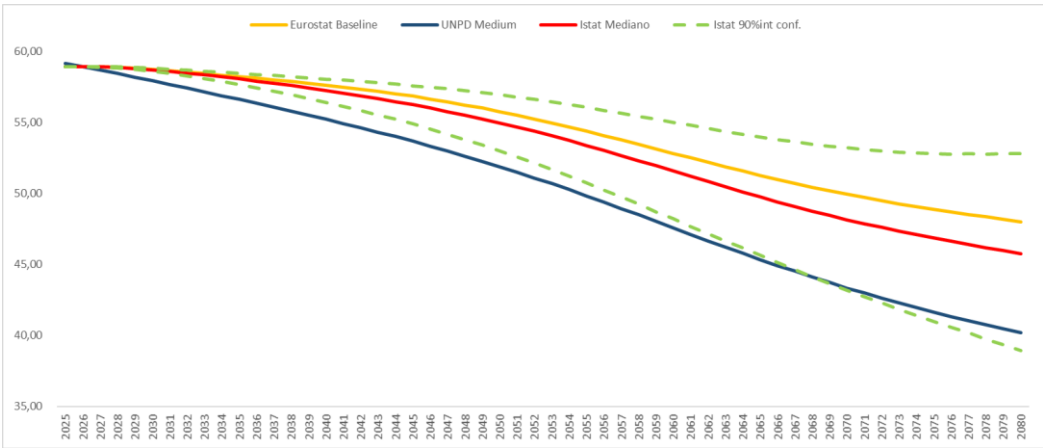
Il prospetto A3 presenta le principali ipotesi di scenario messe a confronto. Per quanto attiene i flussi migratori il confronto è limitato al saldo migratorio con l'estero in quanto sia Eurostat sia UNPD costruiscono le ipotesi direttamente su tale indicatore (senza distinzione tra immigrati ed emigrati).

Le previsioni UNPD presentano ipotesi assai contenute sul piano dei flussi migratori netti con l'estero, non solo nella fase di proiezione iniziale ma su tutto l'orizzonte previsivo. Lo scenario Eurostat sulle migrazioni internazionali prevede invece una visione molto più ottimistica e vicina all'ipotesi mediana Istat.

PROSPETTO A3. CONFRONTO TRA LE IPOTESI DEI PRINCIPALI INDICATORI SINTETICI CURATE DA ISTAT, EUROSTAT E UNPD. Anni 2024, 2030, 2050 e 2080

Scenario	Numero medio di figli per donna	Speranza di vita alla nascita – Uomini	Speranza di vita alla nascita – Donne	Saldo migratorio estero (migliaia)
Anno 2025				
Istat Mediano	1,14	82,0	85,8	296
Eurostat Baseline	1,14	81,7	85,8	296
UNPD Medium	1,21	81,9	86,0	75
Anno 2030				
Istat Mediano	1,16	82,3	86,1	247
Eurostat Baseline	1,18	82,2	86,2	244
UNPD Medium	1,25	82,8	86,6	44
Anno 2050				
Istat Mediano	1,25	84,1	87,6	203
Eurostat Baseline	1,28	84,2	88,1	210
UNPD Medium	1,35	85,4	89,0	43
Anno 2080				
Istat Mediano	1,30	86,2	89,6	221
Eurostat Baseline	1,39	87,4	90,9	209
UNPD Medium	1,43	88,8	92,5	39

FIGURA A2. POPOLAZIONE TOTALE PREVISTA SECONDO GLI SCENARI ISTAT, EUROSTAT E UNPD. Anni 2024-2080, milioni di residenti



Il confronto sulle ipotesi di fecondità è invece più articolato. Nel medio-lungo termine Eurostat e UNPD producono ipotesi più favorevoli. Nel periodo di avvio delle previsioni, le ipotesi Eurostat ed Istat sono simili, mentre UNPD produce valori subito più alti a causa di una base dati meno aggiornata che non tiene conto dell'ulteriore diminuzione della fecondità avvenuta nel biennio 2024-2025. Più distanti risultano le ipotesi sulla sopravvivenza, laddove Eurostat e soprattutto UNPD evidenziano aspettative molto favorevoli circa l'allungamento della vita media, che nel modello Istat si intravedono solo in parte.

Lo sviluppo delle diverse ipotesi demografiche dà quindi luogo a delle differenze in termini di risultati attesi che, per quanto riguarda l'evoluzione della popolazione totale, è possibile apprezzare nella figura A2. Le previsioni UNPD danno un'evoluzione della popolazione pessimista che approssima il limite inferiore dell'intervallo di confidenza delle previsioni Istat. Lo scenario Eurostat risulta leggermente più ottimista, basandosi su ipotesi riguardanti fecondità e sopravvivenza più sostenute soprattutto nel lungo periodo. Ciononostante, la traiettoria evolutiva della popolazione risulta coerente tra i tre scenari. Tutti, infatti, ne prevedono un progressivo declino che tende ad accentuarsi nel medio-lungo termine.

Procedura analitica di calcolo

Una volta predisposti gli input relativi a tutte le componenti demografiche si può attivare la procedura di calcolo della popolazione per ciascun anno di previsione. Il metodo analitico di riferimento che ricalca il concetto di equazione della popolazione è universalmente noto come modello delle componenti di coorte. Tale modello, simula l'evoluzione in avanti della popolazione che, tenendo conto del naturale processo di avanzamento dell'età, la modifica anno dopo anno sulla base del saldo naturale (differenza tra nascite e decessi) e del saldo migratorio (differenza tra movimenti migratori in entrata e movimenti migratori in uscita). La procedura consente di ottenere un generico contingente di popolazione ${}_{t+1}P_{x+1}$ a partire dal contingente iniziale ${}_tP_x$, dove x rappresenta l'età e t l'anno di calendario. Per età x si intende quella in anni compiuti al 1° gennaio dell'anno, sia con riferimento al dato di stock (popolazione), sia con riferimento ai flussi che si concretizzano lungo tutto l'anno (nascite, decessi, migrazioni), onde favorire coerenza nei calcoli rispetto alle generazioni. Nella notazione si prescinde dalla variante previsiva dal momento che la procedura di calcolo è sempre la stessa per ciascuna delle simulazioni.

Il passo preliminare consiste nella determinazione del numero dei nati. Se definiamo con

${}_tF_x^i$ la popolazione femminile di età x della regione i-ma al 1° gennaio dell'anno t

${}_tD_x^i$ le donne decedute all'età x della regione i-ma nell'anno t

${}_t\lambda_x^i$ la probabilità di sopravvivere fino al 1° gennaio t+1 per il contingente di donne ${}_tF_x^i$

${}_tE_x^i$ le donne emigrate di età x dalla regione i-ma nell'anno t, con destinazione altre regioni

${}_tEW_x^i$ le donne emigrate di età x dalla regione i-ma nell'anno t, con destinazione l'estero

${}_tI_x^i$ le donne immigrate di età x nella regione i-ma nell'anno t, provenienti da altre regioni

${}_tIW_x^i$ le donne immigrate di età x nella regione i-ma nell'anno t, provenienti dall'estero

${}_tm_x^{ij}$ la probabilità per una donna di età x di emigrare dalla regione i-ma alla regione j-ma nell'anno t

avremo per $x=14, 15, \dots, 49$

$${}_{t+1}F_{x+1}^i = {}_tF_x^i - {}_tD_x^i - {}_tEW_x^i + {}_tIW_x^i - {}_tE_x^i + {}_tI_x^i \quad (1)$$

dove

$${}_tD_x^i = {}_tF_x^i \cdot (1 - {}_t\lambda_x^i)$$

$${}_tE_x^i = \sum_j {}_tF_x^i \cdot {}_t\lambda_x^i \cdot {}_tm_x^{ij} \quad (2)$$

$${}_tI_x^i = \sum_j {}_tF_x^j \cdot {}_t\lambda_x^j \cdot {}_tm_x^{ji} \quad (3)$$

Si osservi il modo in cui, nella (2) e nella (3), agiscono le probabilità di sopravvivenza della popolazione femminile di età x, introdotte per tenere conto dell'effetto della mortalità sui contingenti esposti a rischio di migrazione. Tale formulazione significa, in pratica, ipotizzare l'incompatibilità dei due eventi morte e migrazione nello stesso intervallo di tempo, in quanto la probabilità di migrare viene applicata solo alle quantità ${}_tF_x^i \cdot {}_t\lambda_x^i$ e ${}_tF_x^j \cdot {}_t\lambda_x^j$, che rappresentano l'ammontare di sopravvissuti alla fine del periodo stesso⁸. Si noti, inoltre, che nella (2) l'indice di sommatoria j ($j \neq i$) è esteso a tutte le possibili destinazioni, ovvero 19 regioni italiane più 2 province autonome, per un complesso di 21 destinazioni. Nella (3) l'indice di sommatoria j ($j \neq i$) riguarda invece le aree di origine, per un complesso di 21 origini.

⁸ Così operando, se chiamiamo in generale q la probabilità di morire e m la probabilità di emigrare, si commette un errore pari a circa $qm/2$. Infatti, il modello adottato si basa sull'indipendenza delle probabilità di morte (dall'influenza delle migrazioni) e delle probabilità di migrazione (dall'influenza della mortalità). In particolare, quest'ultime sono state modificate in modo tale da eliminare l'influenza dell'evento perturbatore mortalità, che fa sentire i suoi effetti soprattutto alle età anziane. Sotto tali condizioni possiamo definire le seguenti probabilità:

(1-q) (1-m) sopravvivenza e permanenza

(1-q) m sopravvivenza e migrazione

q (1-m) morte (per i permanenti)

q m morte (per i migranti)

Correttamente, quindi, soltanto q (1-m)+qm/2 decessi andrebbero assegnati alla regione di origine, mentre qm/2 di essi andrebbero ripartiti tra le regioni di destinazione. Assegnare tutti i decessi alla regione di origine, introducendo così l'ipotesi di incompatibilità tra i due eventi considerati, significa semplificare notevolmente i calcoli, commettendo un errore per eccesso praticamente irrilevante che, oltre tutto, è compensato dall'errore per difetto che si commette non computando in entrata le frazioni di qm/2 originate da altre regioni e destinate a quella considerata.

La (1) rappresenta la popolazione femminile in età feconda al 1° gennaio $t+1$. In base ad essa, e ai quozienti di fecondità per classi annuali di età è possibile determinare le nascite previste:

$${}_tN^i = \sum_x 0,5 \cdot ({}_tF_x^i + {}_{t+1}F_{x+1}^i) \cdot {}_tf_x^i \quad (4)$$

La (4) fornisce il numero di nati vivi nell'anno t . Tale ammontare viene quindi ripartito per sesso sulla base della composizione di 106 maschi ogni 100 femmine. Le due quantità ${}_tNM^i$ e ${}_tNF^i$, rispettivamente nati di sesso maschile e femminile, vengono quindi inserite al primo posto dei vettori della popolazione maschile e femminile⁹. Dopo la determinazione dei nati può partire il calcolo iterativo per ottenere la popolazione al 1° gennaio dell'anno $t+1$; visto che la procedura è identica per i due sessi, la popolazione viene indicata in notazione con il solo simbolo P . Per ottenere il vettore ${}_{t+1}P_{x+1}$ a partire dal vettore completo ${}_tP_x$ si sfrutta il procedimento analogo a quello utilizzato per la determinazione della (1).

Utilizzando la simbologia ormai nota, avremo per $x = 0, 1, \dots, 94$:

$${}_{t+1}P_{x+1}^i = {}_tP_x^i - {}_tD_x^i - {}_tEW_x^i + {}_tIW_x^i - {}_tE_x^i + {}_tI_x^i$$

mentre per la popolazione di 0 anni compiuti al 1° gennaio $t+1$ si ha:

$${}_{t+1}P_0^i = {}_tN^i - {}_tD_n^i - {}_tEW_n^i + {}_tIW_n^i - {}_tE_n^i + {}_tI_n^i$$

espressione nella quale il primo termine a destra dell'uguale identifica i nati mentre quelli successivi i possibili eventi (di morte o migrazione) che li possono interessare tra il momento della nascita e la fine dell'anno. Dai 95 anni di età in poi i movimenti migratori sono supposti nulli, pertanto, l'equazione della popolazione si limita a considerare la sola eliminazione per decesso:

$${}_{t+1}P_{x+1}^i = {}_tP_x^i - {}_tD_x^i$$

Per la classe aperta 120 anni e oltre si determina prima la quantità:

$${}_{t+1}P_{121+}^i = {}_tP_{120+}^i - {}_tD_{120+}^i$$

Quindi si riconsolida l'ultimo elemento del vettore di popolazione mediante l'espressione¹⁰:

$${}_{t+1}P_{120+}^i = {}_{t+1}P_{120}^i + {}_{t+1}P_{121+}^i$$

Si ha, infine:

$${}_tD_x^i = {}_tP_x^i \cdot (1 - {}_t\lambda_x^i)$$

$${}_tE_x^i = \sum_j {}_tP_x^i \cdot {}_t\lambda_x^i \cdot {}_tm_x^{ij} \quad (5)$$

$${}_tI_x^i = \sum_j {}_tP_x^j \cdot {}_t\lambda_x^j \cdot {}_tm_x^{ji} \quad (6)$$

dove, per quanto riguarda le espressioni (5) e (6), va osservato che le probabilità di migrazione dei nati (per $x=n$) tengono conto del fatto che costoro sono esposti al rischio di emigrare per mezzo anno in media. Si noti, infine, che la (5) e la (6) sono strutturalmente identiche (cambia solo la popolazione esposta a rischio) alla (2) e alla (3); anche in questo caso viene ipotizzata l'incompatibilità dei due eventi morte e migrazione, e valgono le considerazioni svolte nella nota 8.

⁹ Ciascun vettore contiene di conseguenza 122 elementi, dei quali il primo identifica i nati vivi, l'ultimo la classe di età aperta (120 e più) e i restanti le singole età da 0 a 119 anni.

¹⁰ In diffusione sono rilasciati i dati fino alla classe aperta 110 anni e oltre.

2) PREVISIONI REGIONALI DELLE FAMIGLIE

Le previsioni delle famiglie illustrano l'andamento futuro del numero e della tipologia di famiglie che risiederanno in Italia dal 2025 al 2050. Si tratta di proiezioni derivanti dall'applicazione di un metodo statico, basato sui *Propensity rates*, applicati alla popolazione prevista. L'obiettivo è quello di offrire, insieme alle previsioni della popolazione, un sistema integrato di informazioni utili a numerosi utenti, pubblici e privati, che si occupano di beni e servizi destinati alle famiglie piuttosto che ai singoli individui. Data l'importanza del ruolo della famiglia, sia a livello protettivo sia per le scelte e i percorsi individuali, la domanda di informazioni su questo collettivo nasce dalle necessità di pianificazione in vari ambiti. Si pensi innanzitutto alle decisioni da prendere nelle politiche di natura economica e sociale, come ad esempio quelle relative ai sistemi abitativi, sociali e assistenziali nei riguardi di giovani e anziani, ma anche alla produzione di beni di consumo durevoli per le famiglie e ai consumi di energia. Come per le previsioni demografiche, anche le previsioni delle famiglie poggiano su un sistema *expert-based* per definire le ipotesi di lungo termine. In particolare, lo stesso panel di esperti nazionali che ha concorso alla definizione delle ipotesi demografiche ha anche supportato l'Istat nella definizione delle ipotesi riguardanti le famiglie relative all'Italia¹¹. Le ipotesi sub-nazionali, invece, sono state curate dall'Istat sulla base di un'apposita metodologia "ponte" tra le ipotesi nazionali e quelle territoriali, come si spiega più avanti.

Dettaglio territoriale e periodo di previsione

Le previsioni delle famiglie hanno base 1.1.2025 e sono diffuse sia per il totale Italia sia con un dettaglio territoriale che scende fino al livello regionale, incluse le province autonome di Bolzano e Trento. Il periodo di previsione si estende fino al 2050.

Dati e popolazione base

I dati utilizzati sono:

- 1) per la popolazione base, le Previsioni probabilistiche della popolazione residente per sesso, età e regione – base 1.1.2025 dello scenario mediano;
- 2) per stimare la popolazione che vive in convivenza istituzionale per sesso e singolo anno di età, il Registro base della popolazione – 1° gennaio 2024 e 2025;
- 3) per quanto concerne le strutture familiari, l'Indagine Multiscopo "Aspetti della vita quotidiana", che offre la possibilità di utilizzare una lunga serie storica (dal 2002 al 2025).

Per il concetto di famiglia si fa riferimento alla definizione di "famiglia di fatto", che fino al 2022 identificava come famiglia *"l'insieme di persone legate da vincoli di matrimonio, parentela, affinità, adozione, tutela, o da vincoli affettivi, coabitanti e aventi dimora abituale nello stesso Comune"*. Nel 2023 la definizione di famiglia cambia leggermente con l'inserimento del concetto di condivisione delle risorse economiche: *"nella famiglia vanno incluse tutte le persone legate da vincoli di parentela o meno, dimoranti abitualmente nella stessa abitazione, che condividono il reddito (contribuendovi e/o beneficiandone) e/o le spese familiari"*¹².

Inoltre, dall'edizione 2023 l'indagine AVQ ha introdotto una nuova procedura per il riporto all'universo del numero totale di famiglie. Al fine di rendere coerenti, infatti, i risultati di indagine con le evidenze annualmente scaturite dal Censimento permanente della popolazione e delle abitazioni, i vincoli di calibrazione per la costruzione dei coefficienti di riporto sono definiti nell'ambito del sistema di "Stime anticipatorie degli indicatori demografici e sociali". In particolare, attraverso tale sistema, che fornisce la base per il riporto all'universo di tutte le indagini campionarie dell'Istat, viene costruita in veste anticipatoria una stima della distribuzione delle famiglie per numero di componenti che il Censimento permanente rilascia in forma definitiva solo l'anno successivo a quello cui si riferiscono i dati di indagine¹³.

La metodologia alla base delle previsioni

Il modello utilizzato per la previsione delle famiglie si basa su un adattamento al contesto italiano del metodo noto come *"Propensity model"*. Si tratta di un metodo macro statico¹⁴ che va oltre il classico *Headship rate model*, superando il concetto di 'capo famiglia' e fornendo un set di informazioni molto più dettagliato. Come prodotto finale si ottengono le previsioni del numero di famiglie future, la loro dimensione media e composizione, nonché la popolazione prevista nei diversi ruoli in famiglia. Alla base del metodo vi sono i *Propensity rates*, o tassi di propensione familiare, costruiti come la proporzione di persone di età x nella posizione familiare i al tempo t :

¹¹ Il questionario sulle famiglie contempla quesiti sul 2050, nei quali gli esperti hanno risposto sui cambiamenti da loro attesi delle principali strutture familiari (persone sole, coppie con o senza figli, famiglie monogenitore).

¹² Il cambiamento della definizione di famiglia di AVQ nel 2023 è una conseguenza della modifica del regolamento Eurostat sulle statistiche sociali, in cui viene adottata questa definizione.

¹³ Tale innovazione di processo poggia su una metodologia che usufruisce della serie storica corrente circa la distribuzione delle famiglie per numero di componenti di fonte censuaria e, attraverso un opportuno modello estrapolativo di nowcasting, è in grado di stimare numero di famiglie e sua articolazione per l'anno per cui si rendono necessarie le informazioni.

¹⁴ Un modello macro prevede l'evoluzione delle famiglie attraverso grandezze aggregate (tassi, proporzioni), mentre un modello micro simula l'evoluzione delle famiglie a partire dalle traiettorie dei singoli individui. A sua volta, un modello statico prevede le famiglie a istanti temporali distinti (es: all'inizio di ogni anno) senza rappresentare il processo di transizione tra un istante e il successivo. Al contrario, in un modello dinamico si prevede esplicitamente il processo di cambiamento nel tempo, modellando le transizioni tra istanti diversi con le traiettorie assunte dai singoli individui o da loro gruppi.

$$\text{Propensity Rate}_{x,i,t} = \frac{P_{x,i,t}}{P_{x,t}}$$

Ad esempio, la propensione per una persona con 30 anni di età a vivere in coppia con un partner sarà data dal numero di 30enni che vivono in coppia rapportati a tutta la popolazione dei 30enni.

I vantaggi del metodo sono molteplici: si lega facilmente alle previsioni della popolazione; non è necessario analizzare le transizioni tra le potenziali posizioni familiari, tipiche di un modello dinamico; è semplice da applicare e fornisce buoni dettagli nei risultati. In quanto metodo statico, non coglie in modalità istantanea i processi di formazione e scioglimento delle famiglie. Così l'applicazione dei tassi di propensione alla popolazione residente di riferimento può in alcuni casi determinare delle incoerenze, ad esempio tra sessi o per le posizioni familiari all'interno delle età, problematica che è quindi necessario aggiustare ex-post.

Il metodo si articola in 5 passi:

- Passo 1. Stimare la popolazione base prevista che vive in famiglia
- Passo 2. Calcolare le propensioni familiari (*Propensity rates*)
- Passo 3. Ipotizzare i trend futuri delle propensioni familiari
- Passo 4. Derivare la popolazione prevista nelle diverse posizioni familiari
- Passo 5. Calcolare il numero, la tipologia e la dimensione delle famiglie previste.

Di seguito si illustrano i vari passi in dettaglio.

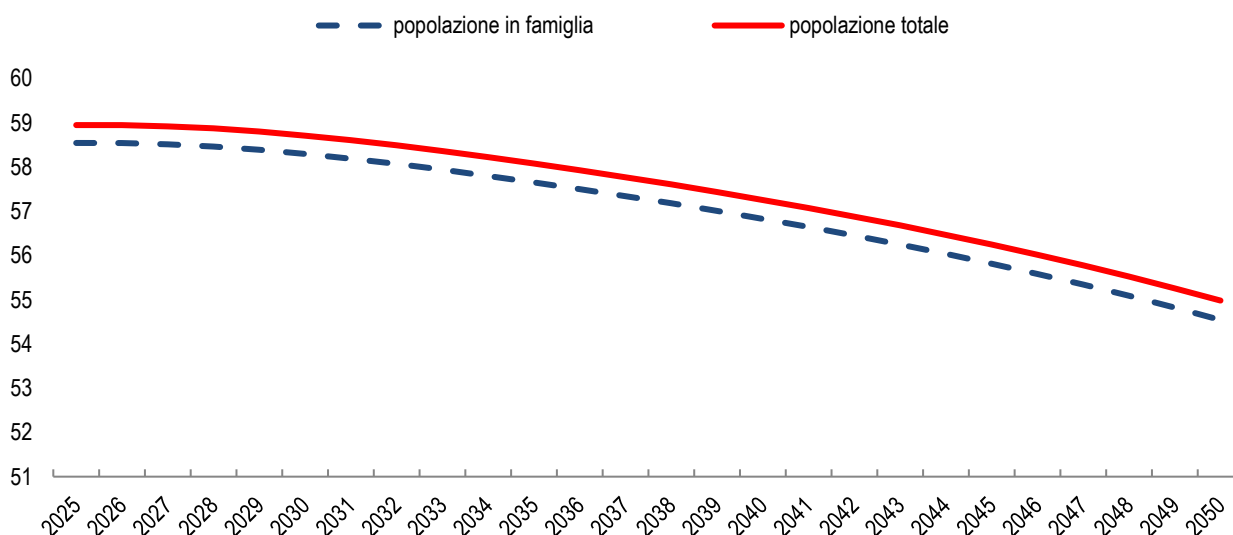
Passo 1. Stimare la popolazione base e prevista che vive in famiglia

Poiché la popolazione base di riferimento è rappresentata dalla popolazione residente per sesso età e regione al 1° gennaio 2025 dell'ultimo Censimento permanente effettuato dall'Istat, è necessario effettuare preliminarmente una stima della popolazione che vive in famiglia, escludendo gli individui che risiedono in convivenza istituzionale (ospedali, caserme, carceri, case di riposo, strutture religiose, ecc.). Tale operazione va quindi ripetuta per tutti gli anni di previsione, defalcando dalle previsioni regionali (il cui riferimento è anche in questo caso la popolazione complessivamente residente) la quota parte di popolazione istituzionalizzata.

Ricorrendo ai dati del Registro base della popolazione all'1.1.2024 e all'1.1.2025, sono state calcolate le incidenze percentuali (medie) della popolazione che vive in istituzione per sesso, classe quinquennale di età e regione. Data la sostanziale stabilità nel tempo di questa popolazione, si è ipotizzato un trend costante per tutto il periodo di previsione.

Applicando alla popolazione totale, dal 2025 al 2050, il complemento a 100 di queste incidenze, si è ottenuta la popolazione che vive in famiglia per regione, sesso e classe di età (Figura A3).

FIGURA A3. POPOLAZIONE RESIDENTE TOTALE E IN FAMIGLIA. Anni 2025-2050, scenario mediano, in milioni



Passo 2. Calcolare le propensioni familiari (*Propensity rates*)

Il secondo passo del modello consiste nel calcolare le propensioni a vivere in una determinata posizione familiare per sesso e classe di età quinquennali. Si considerano le seguenti posizioni familiari:

1. Persona sola
2. Persona in coppia senza figli
3. Persona in coppia con almeno un figlio di età fino a 19 anni
4. Persona in coppia con tutti i figli di 20 anni di età e più

5. Genitore solo con almeno un figlio di età fino a 19 anni
6. Genitore solo con tutti i figli di 20 anni di età e più
7. Figlio (che vive con almeno un genitore)
8. Altra persona che vive in un nucleo familiare
9. Persona in famiglia multipersonale (ad esempio due fratelli che convivono o un figlio divorziato che è tornato a casa da un genitore)
10. Persona in famiglia con due o più nuclei (ad es. due persone in coppia che vivono con la figlia e il genero).

Le posizioni da 2 a 8 si riferiscono agli individui in famiglie composte da un solo nucleo familiare. In generale il “figlio” segue la definizione adottata in tutte le indagini Multiscopo che li considera tali solo se celibi/nubili. Le persone che vivono in famiglie con due o più nuclei sono state considerate in una categoria a sé, pur costituendo una quota esigua del totale delle famiglie (circa l'1,5%).

Come si è accennato, le propensioni (o *Propensity rates*) sono costruite come la proporzione di persone di età x nella categoria i . In questo contesto, la variabile età è stata considerata in classi quinquennali e i tassi sono stati disaggregati anche per sesso, variabile che risulta molto discriminante nei comportamenti familiari. Di seguito tali tassi sono denominati come *Living Arrangement Propensities* (LAP).

$$\text{Propensity Rate}_{x,i,s,t} = \frac{P_{x,i,s,t}}{P_{x,s,t}} = \text{LAP}_{x,i,s,t}$$

dove x = classe quinquennale di età 0-4, 5-9, ..., 80-84, 85+, i = posizione familiare, s = sesso, t = tempo.

Le LAP sono calcolate ricorrendo ai dati dell'indagine campionaria Istat Aspetti della vita Quotidiana, su tutta la serie storica 2002-2025. Poiché il dettaglio regionale per sesso e classi di età porta ad avere una esiguità di dati nelle regioni piccole, si è deciso di effettuare un raggruppamento per “macro-regioni”.

Attraverso analisi statistiche multivariate, condotte sulla base di fattori di contesto sociodemografico¹⁵, si è arrivati a definire i seguenti 5 gruppi di regioni:

- Gruppo 1 - Nord-ovest (Piemonte, Valle d'Aosta, Lombardia, Liguria)
- Gruppo 2 - Est Adriatico (Veneto, Emilia-Romagna, Trentino-Alto Adige, Friuli-Venezia Giulia, Marche)
- Gruppo 3 - Tirrenico (Toscana, Lazio)
- Gruppo 4 - Sud (Campania, Puglia, Calabria, Sicilia)
- Gruppo 5 - Centrale (Umbria, Sardegna, Abruzzo, Molise, Basilicata).

Passo 3. Ipotizzare i trend futuri delle propensioni familiari

La previsione dell'evoluzione futura delle propensioni familiari si è basata sull'introduzione di un indicatore sintetico che riassume i comportamenti familiari della popolazione nel tempo. Tale indicatore, definito sotto il nome di *Tasso di Propensione Totale per posizione familiare* (TPT), è dato dalla somma per età delle LAP ponderate con gli anni vissuti alle varie età.

$$\text{TPT}_{i,s,t} = \sum_{x=0-4}^{85+} \text{LAP}_{x,i,s,t} * L_{x,s,t} = \sum_{x=0-4}^{85+} \frac{P_{x,i,s,t}}{P_{x,s,t}} * 100 * L_{x,s,t}$$

dove i = posizione familiare, s = sesso, x = classe quinquennale di età, t = tempo.

Gli $L_{x,s,t}$, che rappresentano gli anni vissuti per classe di età x e sesso s nell'anno t , sono desunti dalle tavole di mortalità proiettate dello scenario mediano.

Sotto l'ipotesi di indipendenza tra mortalità e posizione familiare, il TPT per una data posizione familiare rappresenta il numero di anni che mediamente si aspetta di vivere in quella posizione una generazione di individui, assumendo nel corso della vita i comportamenti familiari e le condizioni di mortalità osservate in un determinato anno di calendario. Si tratta, quindi, di una vera e propria aspettativa di vita in quello stato familiare, traslata dalla dimensione osservazionale trasversale a quella longitudinale. In altri termini, assume lo stesso significato che hanno indicatori trasversali più conosciuti, come ad esempio il numero medio di figli per donna, la speranza di vita alla nascita o il tasso di nuzialità totale. Per l'implementazione nella formula di calcolo della funzione degli anni vissuti il TPT ricorda assai da vicino il tasso netto di riproduttività.

Se nel 2005 un uomo contava di vivere come persona sola mediamente 6 anni (su una speranza di vita totale di 78,1), nel 2025 il tempo atteso in questo stato sale a 12,4 anni (su un totale di 82). A seguito del calo della natalità, invece, nel 2005 le donne si aspettavano di vivere come persona in coppia con almeno un figlio fino a 19 anni di età figli per 13,8 anni (su un totale di 83,5), ma nel 2025 questo tempo atteso è sceso a 13,1 anni (su una

¹⁵ Allo scopo di individuare gruppi omogeni di regioni in base alle strutture familiari e alla loro evoluzione nel tempo, è stata effettuata un'analisi in componenti principali dinamica tramite la metodologia STATIS. L'analisi ha preso in esame le principali variabili socio-demografiche a livello regionale negli anni 2002-2019, tra cui: tassi di fecondità, età medie al parto, numero medio di componenti familiari, tassi di separazione e divorzio, tassi di occupazione femminile, tassi di migrazione interna ed estera, incidenza di alcune tipologie familiari (persone sole, coppie con e senza figli, genitori soli, ecc.). La procedura è stata ottimizzata eliminando le variabili con bassa variabilità latente spiegata dall'asse.

aspettativa di vita totale nel frattempo salita a 85,8 anni). Per fare un ultimo esempio, il tempo nello stato di “figlio” è aumentato da 30,7 a 30,8 anni per i maschi e da 28,0 a 28,5 per le femmine, a causa della prolungata permanenza dei giovani all’interno della famiglia di origine.

Per ipotizzare i trend futuri delle propensioni, si è proceduto alla previsione dell’“intensità totale” del tempo trascorso in ogni posizione familiare, per poi stimarne la “cadenza” o distribuzione nelle varie età ($LAP_{x,i,s,t}$) in ogni anno di previsione. Prevedere in una prima fase l’intensità totale ha consentito, da un lato, di tradurre più facilmente le ipotesi previsive sui comportamenti familiari, dall’altro, di mantenere insieme gli andamenti delle varie posizioni familiari che, se previste separatamente per fasce di età, risulterebbero di più difficile controllo con il rischio di ottenere risultati inattendibili (ad esempio, un tasso per la posizione familiare di figlio più alto alle età intermedie che in quelle giovanili).

L’obiettivo finale del passo 3, ossia la definizione delle LAP previste dal 2025 al 2050 per regione, è stato conseguito effettuando prima la previsione nei 5 gruppi territoriali stabiliti (passo 3.1), per poi arrivare da questi al dettaglio regionale (passo 3.2).

PROSPETTO A4. TASSI DI PROPENSIONE TOTALE PER POSIZIONE FAMILIARE E SESSO. Anni 2005, 2015 e 2025.

Posizione familiare	MASCHI			FEMMINE		
	2005	2015	2025	2005	2015	2025
Persona sola	6,0	8,4	12,4	10,9	12,2	14,0
Persona in coppia senza figli	13,5	14,1	14,6	12,5	13,1	13,6
In coppia con almeno un figlio fino a 19 anni	14,0	12,9	12,1	13,8	13,3	13,1
In coppia con soli figli di 20 anni e più	8,7	7,5	6,0	8,3	6,9	5,6
Genitore solo con almeno un figlio fino a 20 anni	0,3	0,3	0,4	1,3	1,9	2,5
Genitore solo con soli figli di 20 anni e più	0,7	0,8	0,9	3,0	3,0	3,1
Figlio/a	30,7	30,9	30,8	28,0	28,2	28,5
Altra persona in famiglia con 1 nucleo	0,8	1,1	0,9	1,9	1,5	1,2
Persona in famiglia multipersonale	1,1	1,4	1,9	1,6	1,6	2,0
Persona in famiglia con 2+ nuclei	2,2	2,7	2,1	2,2	2,9	2,3
Totale	78,1	80,1	82,0	83,5	84,5	85,8

Passo 3.1 Previsione delle LAP nei 5 gruppi territoriali

L’intensità totale di ogni posizione familiare e sesso ($TPT_{i,s,t}$) è stata prevista attraverso la combinazione di due informazioni:

- l’estrapolazione dei trend nel periodo 2002-2024, utilizzando modelli di analisi delle serie storiche. Per ogni posizione familiare e sesso sono stati applicati modelli di tipo ARIMA, *Random walk with drift* o Linear Trend (Prospetto A5);
- le opinioni degli esperti sui futuri livelli al 2050 a livello nazionale relativamente ai TPT delle principali posizioni familiari per sesso.

PROSPETTO A5. MODELLI PREVISIVI DEI TASSI DI PROPENSIONE TOTALE PER POSIZIONE FAMILIARE E SESSO (Modello prevalente tra i 5 gruppi territoriali) *

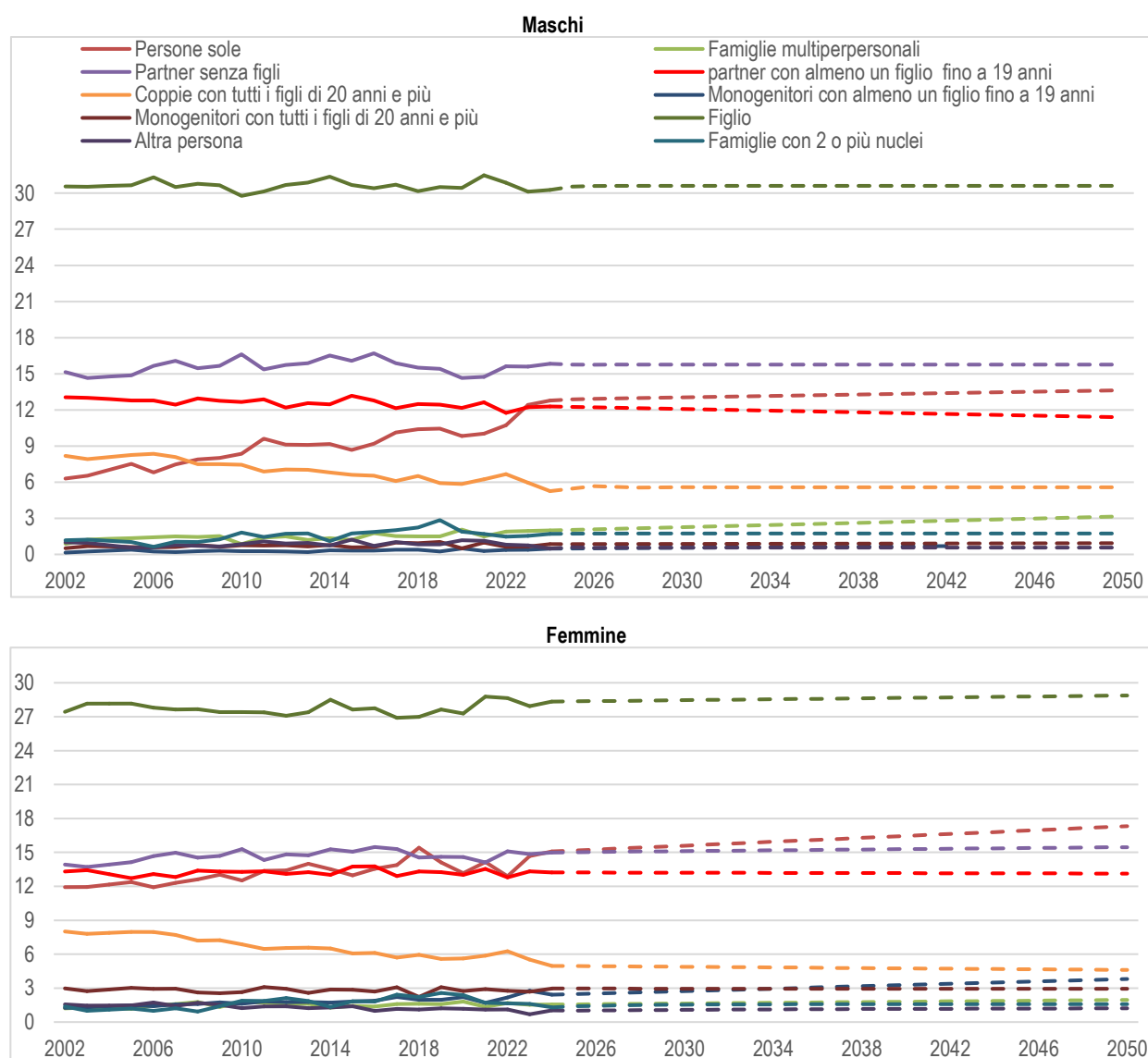
Posizione familiare	MASCHI	FEMMINE
Persona sola	RWD ARIMA(1,1,0)	RWD ARIMA(0,1,0)
Persona in famiglia multipersonale	RWD	Linear Trend
Persona in coppia senza figli	ARIMA(1,1,0)	Linear Trend
Persona in coppia con almeno un figlio fino a 19 anni	RWD	ARIMA(2,2,0)
Persona in coppia con tutti i figli di 20 anni e più	RWD ARIMA(0,1,0)	RWD ARIMA(0,1,0)
Genitore solo con almeno un figlio fino 19 anni	RWD ARIMA(0,1,0)	RWD
Genitore solo con tutti i figli di 20 anni e più	RWD ARIMA(0,1,0)	ARIMA(1,0,0)
Figlio/a	ARIMA(1,0,0)	RWD
Altra persona	ARIMA(2,1,0)	ARIMA(1,0,1)
Persona in famiglia con 2+ nuclei	RWD	RWD ARIMA(0,2,0)

*RWD=Random Walk with Drift model; ARIMA=AutoRegressive Integrated Moving Average model.

Per quanto riguarda l'estrapolazione dei trend, un esempio di risultato dei suddetti modelli è quello rappresentato in Figura A4, dove viene raffigurato il comportamento del gruppo territoriale del Nord-ovest per le principali posizioni familiari. Più in generale, con riguardo a ciascuna zona del Paese, le variazioni del tempo trascorso nei vari ruoli familiari si traducono in:

- un aumento delle persone sole,
- una diminuzione dei partner con figli,
- un lieve aumento dei partner senza figli,
- un leggero aumento delle persone nel ruolo di figlio,
- un leggero aumento dei genitori soli, soprattutto se padri

FIGURA A4. TASSI DI PROPENSIONE TOTALE PER POSIZIONE FAMILIARE. NORD-OVEST. Anni 2002-2050



Nel Prospetto A6 vengono invece riportati i valori medi al 2050 espressi dagli esperti. Ad ogni esperto è stata richiesta un'opinione in modo tale che la somma dei TPT (per sesso) da loro indicati al 2050 risultasse uguale al valore della speranza di vita da loro prevista al 2050 (come loro ipotesi centrale). Gli esperti che hanno fornito informazioni complete sono stati 89. Con un'interpolazione di tipo lineare sono stati successivamente calcolati i valori dei TPT *expert-based* (a livello Italia e per sesso) per tutto il periodo 2025-2050.

I valori finali dei tassi di propensione totale per ciascun gruppo di regioni sono stati ricavati utilizzando la seguente formula:

$$TPT_{s,i,t,g} = \frac{STPT_{s,i,t,g}}{MSTPT_{s,i,t}} \cdot ETPT_{s,i,t} \quad t = 2025, \dots, 2050$$

Dove $STPT_{s,i,t,g}$ è il valore previsto per il gruppo g attraverso il modello in serie storica in corrispondenza del sesso s , della posizione familiare i e dell'anno t ; $MSTPT_{s,i,t}$ è la media ponderata dei gruppi dei valori previsti con i modelli in serie storica in corrispondenza del sesso s , della posizione familiare i e dell'anno t e dove la numerosità della popolazione dei gruppi di regioni è utilizzata come peso nella media; $ETPT_{s,i,t}$ è il valore previsto con il questionario degli esperti in corrispondenza del sesso s , della posizione familiare i e dell'anno t . In questo modo i valori iniziali (previsti con i modelli in serie storica) dei TPT previsti nel singolo gruppo vengono riproporzionati

in base ad un coefficiente che è pari al rapporto tra il valore fornito dall'esperto e quello proveniente dai modelli in serie storica a livello nazionale. Il risultato finale di tale operazione avvicina i valori finali alle opinioni degli esperti¹⁶.

Ai valori così ottenuti, aggiornati con il valore del TPT del 2025, è stato applicato un fattore correttivo di breve periodo, in modo che la previsione relativa ai primissimi anni risulti quanto più in linea alle tendenze degli ultimi anni osservati, utilizzando la metodologia spiegata nel paragrafo "Componente correttiva di nowcasting" delle previsioni della popolazione.

PROSPETTO A6. VALORI MEDI DELLE OPINIONI DEGLI ESPERTI SUI TASSI DI PROPENSIONE TOTALE PER POSIZIONE FAMILIARE (TPT) E SESSO. Anno 2050

Posizione familiare	MASCHI	FEMMINE
Persona sola	11,8	14,3
Persona in coppia senza figli	14,1	13,4
In coppia con figli	18,8	19,2
Genitore solo con figli	2,3	5,8
Figlio/a	32,1	29,5
Altra posizione	5,1	5,4
Totale	84,2	87,6

La stima della cadenza prevista, ossia delle $LAP_{x,i,s,t}$ dal 2025 al 2050, è stata ottenuta utilizzando i TPT previsti, gli anni vissuti previsti e le distribuzioni osservate dai dati dell'indagine AVQ nel triennio 2023-25. A tal fine, le distribuzioni medie 2023-25 di ogni ruolo in famiglia sono state ponderate con due coefficienti: uno che tenesse conto della previsione del TPT nell'anno t rispetto a quello del triennio 2023-25:

$$WP_{s,i,t} = \frac{TPT_{s,i,t}}{TPT_{s,i,2023-25}} \quad t = 2025, \dots, 2050$$

e un coefficiente che esprimesse i cambiamenti della mortalità nel tempo:

$$WL_{x,s,t} = \frac{L_{x,s,2023-25}}{L_{x,s,t}} \quad t = 2025, \dots, 2050$$

Pertanto le propensioni familiari per tutto l'orizzonte previsivo sono state calcolate con la formula:

$$LAP_{x,s,i,t} = LAP_{x,s,i,2023-25} * WP_{s,i,t} * WL_{x,s,t} \quad t = 2025, \dots, 2050$$

dove: x=classi di età 0-4, ..., 85+, s=sexso, i=ruolo in famiglia.

Non avendo prodotto ipotesi di variazione annuale circa la distribuzione per età delle LAP, si è implicitamente ipotizzato che i comportamenti in termini di scelte familiari mantengano in futuro una distribuzione per classi di età proporzionale a quella rilevata nell'indagine AVQ nel triennio 2023-2025. A seguito della procedura di stima, la somma delle LAP per posizione familiare in ogni classe di età approssima ma non sempre eguaglia il valore di 100, pertanto si rendono necessari degli aggiustamenti ex-post.

Passo 3.2. Previsione delle LAP nelle regioni

Per portare le previsioni delle famiglie a livello regionale è necessario fare in modo che ogni regione recuperi la propria specificità socio-demografica all'interno del gruppo previsivo di cui fa parte. A tale scopo si è definito un fattore di correzione regionale da applicare alle LAP previste dei diversi gruppi territoriali in modo da ottenere quelle specifiche di ciascuna regione:

$$FC_{r,i} = \frac{TPT_{2023-25,i,r}}{TPT_{2023-25,i,G}}$$

dove i=ruolo familiare, r=regione, G=gruppo cui appartiene la regione r.

Le LAP previste per i gruppi territoriali vengono quindi moltiplicate per il fattore di correzione regionale così calcolato, determinando la serie delle LAP regionali dal 2025 al 2050. Ad esempio, per la posizione familiare persona sola maschio, il TPT rilevato in Piemonte è pari a 11,9 mentre nel gruppo 1 è pari a 11,1. Il fattore di correzione equivale dunque in questo caso a 1,07. Questo vuol dire che, poiché il Piemonte ha un TPT più alto di quello del gruppo cui appartiene, occorre fare un aggiustamento moltiplicando tutte le LAP alle diverse età e ai vari anni di previsione per 1,07, aumentandone leggermente il livello.

¹⁶ Poiché le posizioni familiari inserite nel questionario per esperti sono più aggregate rispetto a quelle tradizionalmente usate nei modelli in serie storica (per es. i partner in coppia con figli non sono disaggregati per età dei figli), queste ultime vengono prima riunite e poi di nuovo disaggregate in base alle proporzioni interne iniziali.

Passo 4. Derivare la popolazione prevista nelle diverse posizioni familiari

In questo passo, le propensioni regionali vengono applicate alla popolazione prevista che vive in famiglia ottenuta al Passo 1. Si ricava così la popolazione prevista nei diversi stati familiari per sesso, classe di età e regione dal 2025 al 2050.

Passo 5. Calcolare il numero, la tipologia e la dimensione delle famiglie previste

Il numero previsto di famiglie si ottiene direttamente dalla popolazione distinta per posizione familiare, sesso ed età, in quanto:

- ogni “persona sola” rappresenta 1 famiglia (coefficiente=1);
- le persone in coppia costituiscono lo 0,5 di una famiglia (coefficiente=0,5);
- ogni “genitore solo” rappresenta 1 famiglia (coefficiente=1);
- le famiglie “multipersonali” si ottengono dividendo il numero di persone che vivono in famiglie multipersonali per la dimensione media di questo tipo di famiglie, sostanzialmente stabile nel tempo e pari a circa 2,1 componenti (coefficiente=2,1);
- le famiglie “con due o più nuclei” si ricavano dividendo il numero di persone che vivono in famiglie con due o più nuclei per la dimensione media di questo tipo di famiglie, che assume valori di coefficiente tra 5,0 (nel gruppo 3) e 5,4 (nel gruppo 4) a seconda del gruppo territoriale di riferimento.

Applicando i coefficienti alla popolazione di sesso *s* ed età *x* si ottiene come prodotto finale il numero di famiglie distinto per le tipologie familiari di interesse.

Il numero medio di componenti familiari viene quindi calcolato rapportando la popolazione che vive in famiglia per il numero di famiglie. Può essere disaggregato per le famiglie totali e per quelle con almeno un nucleo familiare (escludendo le persone sole e le famiglie multipersonali).

Ai fini della diffusione, le famiglie multipersonali e quelle con due o più nuclei sono considerate insieme nella modalità “altro tipo di famiglia”.

3) DIFFUSIONE DEI DATI E TERMINI DI UTILIZZO CIRCA LE PREVISIONI DEMOGRAFICHE E DELLE FAMIGLIE

Il quadro dettagliato delle ipotesi sottostanti le previsioni e dei principali risultati è consultabile sia sul sito generalista di Istituto esploradati.istat.it (raggiungibili lungo il percorso Home>IstatData>Categorie>Popolazione e famiglie>Previsioni demografiche) sia sul sito tematico demo.istat.it (Home>Scenari demografici).

La diffusione dei dati inerenti le **previsioni della popolazione** è articolata in tre sezioni comprendenti tabelle elaborabili sulla struttura per sesso e singola classe d'età della popolazione, sulle componenti del bilancio demografico e sui principali indicatori demografici. Ciascuna tabella riporta i valori dello scenario mediano e i limiti, inferiore e superiore, degli intervalli di confidenza al 90%, all'80% e al 50%.

Tra le componenti del bilancio sono inclusi:

- popolazione a inizio e fine anno, saldo totale
- nascite e decessi, saldo naturale
- iscritti e cancellati con l'estero, saldo migratorio estero
- iscritti e cancellati con l'interno, saldo migratorio interno.

I dati sopra descritti e quelli relativi alla distribuzione per età della popolazione sono arrotondati all'unità.

Per quanto riguarda gli indicatori demografici le tabelle comprendono:

- tassi di natalità, mortalità e crescita naturale
- tasso immigratorio dall'estero, emigratorio per l'estero e tasso migratorio netto con l'estero
- tasso immigratorio dall'interno, emigratorio per l'interno e tasso migratorio netto con l'interno
- tasso migratorio netto totale e tasso di crescita totale
- età media della popolazione
- % di popolazione 0-14 anni, 15-64 anni, 65 anni e più, 85 anni e più
- indici di dipendenza strutturale, di dipendenza degli anziani e di vecchiaia
- numero medio di figli per donna
- speranza di vita alla nascita e a 65 anni di età per sesso.

La diffusione dei dati inerenti le **previsioni delle famiglie** è articolata in tre sezioni comprendenti tabelle elaborabili sulla struttura per sesso, classe quinquennale d'età e posizione familiare della popolazione, sulla distribuzione delle famiglie per tipologia e sul numero medio di componenti familiari. Tutti i risultati si riferiscono allo scenario mediano.

La riproduzione delle informazioni contenute nella presente nota e nelle banche dati esploradati.istat.it e demo.istat.it è lasciata libera, a condizione che venga citata la fonte Istat.

L'Istat produce periodicamente le previsioni demografiche nel quadro della linea di attività "Sistema di stime e previsioni della popolazione", in conformità a quanto indicato nel Programma Statistico Nazionale, progetto "Previsioni demografiche" (cod. PSN IST-01448).

4) RIFERIMENTI BIBLIOGRAFICI

1. Australian Bureau of Statistics – ABS. 2024. Household and family projections. Reference period 2021 – 2046. <https://www.abs.gov.au/statistics/people/population/household-and-family-projections-australia/latest-release>
2. Bell M., Cooper J., & Les M. (1995). *Household and family forecasting models: A review*. Commonwealth Department of Housing and Regional Development.
3. Billari, F.C., Corsetti G., Graziani R., Marsili M. e Melilli E. (2014), *Towards stochastic forecasts of the Italian population: an experiment with conditional expert elicitations*. Proceedings of the 6th Eurostat/UNECE Work Session on Demographic Projections, pagg. 326-338, Istat, 2014.
4. Billari, F.C., Corsetti G., Graziani R., Marsili M. e Melilli E. (2014), *A stochastic multi-regional model for Italian population projections*. Budapest, 25-28 giugno 2014, *European Population Conference*. <http://epc2014.princeton.edu/papers/140361>.
5. Corsetti G., Marsili M. (2012), *A stochastic population projection from the perspective of a national statistical office*. *European Population Conference*. Stoccolma, 13-16 giugno 2012, EAPS. <http://epc2012.princeton.edu/papers/120635>.
6. Corsetti G., Marsili M. (2013), *Previsioni stocastiche della popolazione nell'ottica di un Istituto nazionale di statistica*. *Rivista di statistica ufficiale*, n. 2-3, p. 5-29, Istat.
7. Eurostat (2026), *Population projections in the EU (Europop2025)*, *Statistics Explained*, April 2026.
8. Istat (1989), *Previsioni della popolazione residente per sesso, età e regione – Base 1.1.1988*, *Note e Relazioni*, n.4.
9. Istat (1989), *Previsioni del numero di famiglie italiane dal 1995 al 2020*. Notiziario. Serie 4. Foglio 41 – Anno X No. 19. Dicembre 1989.
10. Istat (1997), *Previsioni della popolazione residente per sesso, età e regione – Base 1.1.1996*, *Informazioni*, n. 34.
11. Istat (2001), *Previsioni della popolazione residente base 1° gennaio 2000*, *Statistiche in breve*, www.istat.it.
12. Istat (2003), *Previsioni della popolazione residente per sesso, età e regione, base 1.1.2001*, *Informazioni* n.13.
13. Istat (2006), *Previsioni demografiche nazionali 1° gennaio 2005-1° gennaio 2050*, www.istat.it, *Nota informativa*, 22 marzo 2006.
14. Istat (2008), *Previsioni demografiche 1° gennaio 2007-1° gennaio 2051*, www.istat.it, *Nota Informativa*, 19 giugno 2008.
15. Istat (2011), *Il futuro demografico del paese - Previsioni regionali della popolazione residente al 2065*, *Statistiche Report*, www.istat.it, 28 dicembre 2011.
16. Istat (2017), *Il futuro demografico del Paese – Previsioni regionali della popolazione residente al 2065*, *Statistiche Report*, www.istat.it, aprile 2017.
17. Istat (2018), *Il futuro demografico del Paese – Previsioni regionali della popolazione residente al 2065 – base 1.1.2017*, *Statistiche Report*, www.istat.it, maggio 2018.
18. Istat (2019), *Il futuro demografico del paese - Previsioni regionali della popolazione residente al 2065 – base 1.1.2018*, *Nota metodologica*, www.istat.it, ottobre 2019.
19. Istat (2021), *Previsioni della popolazione residente e delle famiglie – base 1.1.2020*, *Statistiche Report*, www.istat.it, novembre 2021.
20. Istat (2022), *Previsioni della popolazione residente e delle famiglie – base 1.1.2021*, *Statistiche Report*, www.istat.it, settembre 2022.
21. Istat (2023), *Previsioni della popolazione residente e delle famiglie – base 1.1.2022*, *Statistiche Report*, www.istat.it, settembre 2023.
22. Istat (2024), *Previsioni della popolazione residente e delle famiglie – base 1.1.2023*, *Statistiche Report*, www.istat.it, luglio 2024.
23. Istat (2025), *Previsioni della popolazione residente e delle famiglie – base 1.1.2024*, *Statistiche Report*, www.istat.it, luglio 2025.
24. Istat (2026), *Indicatori demografici – Anno 2025*, *Statistiche Report*, www.istat.it, marzo 2026. <https://www.istat.it/comunicato-stampa/indicatori-demografici-anno-2025/>.
25. Keilman, N. (2019). Family Projection Methods: A Review. In: Schoen, R. (eds) *Analytical Family Demography. The Springer Series on Demographic Methods and Population Analysis*, vol 47. Springer, Cham. https://doi.org/10.1007/978-3-319-93227-9_12.
26. Lo Conte, M., Marsili, M., Corsetti, G., & Meli, E. (2023). *Expected developments in the number and type of Italian households until 2040*. *Rivista di Statistica Ufficiale*, 5. <https://www.istat.it/wp-content/uploads/2023/12/Rivista-statistica-ufficiale-1-2-3-2023.pdf#page=5>.
27. Lee R. D., Carter L. R. (1992). *Modeling and Forecasting U. S. Mortality*, *Journal of the American Statistical Association*, Vol. 87, No. 419 (Sep., 1992), pp. 659-671.

28. Lee R. D., Miller T. (2001). *Evaluating the performance of the Lee-Carter method for forecasting mortality*, *Demography*, 38 (4): 537–549.
29. Lo Conte, M., Corsetti, G., De Rose, A., Marsili, M., Meli, E. (2023). *The future of the Italian family: Evidence from a household projection model*. In *The Demography of Transforming Families* (pp. 93-118). Cham: Springer International Publishing.
30. Marsili M. (2007), *Demographic projections: the impact of net international migration on population ageing in Italy*, Atti del Convegno Intermedio della SIS 2007 “Rischio e Previsione”, Università Ca’ Foscari, Venezia, 6-8 giugno.
31. Marsili M. (2020) *Scenari demografici, previsioni per l’uso*, Atti della 13° Conferenza nazionale di statistica, Dall’incertezza alla decisione consapevole: un percorso da fare insieme, Roma, 4-6 luglio 2018, pagg. 246-252, Istat, 2020.
32. McDonald, P., R. Kippen, and J. Temple (2006). *Net transition probabilities: an approach to subnational level projections of households and housing demand based on census data*. *Population, Space and Place*, Volume 12, N. 6: 479-495.
33. Rogers A., Raquillet R., Castro L. J. (1978). *Model Migration Schedules and Their Applications*, *Environment and Planning A*, vol. 10(5), pages 475-502, May.
34. Schmertmann C. (2003). *A system of model fertility schedules with graphically intuitive parameters*, *Demographic Research*, Volume 9 - Article 5 | Pages 81–110.
35. Terra Abrami V. (1998), *Le previsioni demografiche*, Il Mulino, Bologna.
36. UNECE (2018), *Recommendations on Communicating Population Projections*, United nations economic commission for Europe, United Nations, New York and Geneva, agosto 2018.
37. United Nations, Department of Economic and Social Affairs, Population Division (2024). *World Population Prospects 2024: Methodology of the United Nations population estimates and projections* (UN DESA/POP/2024/DC/NO. 10).