

Nota metodologica

Le previsioni demografiche regionali dell'Istat sono costruite con l'obiettivo di rappresentare il possibile andamento futuro della popolazione, sia in termini di numerosità totale sia in termini di struttura per età e sesso. Le informazioni prodotte rappresentano uno strumento importante a supporto delle decisioni nelle politiche economiche e sociali, 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 2021 sostituiscono quelle in base 2020 pubblicate dall'Istat nel novembre 2021. 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 nel precedente ciclo triennale, che diede luogo al rilascio in successione delle previsioni in base 2016, 2017 e 2018. Tale metodologia venne definita, tra il 2009 ed 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 previsivo 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 selezioni 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. Tali sezioni includono informazioni sui seguenti aspetti:

- popolazione base
- tecnica di proiezione
- periodo di previsione
- panel di esperti
- questionario degli esperti e modello probabilistico
- 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

- 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
- diffusione dati e termini di utilizzo
- informazioni di contatto e richieste dati personalizzate.

Popolazione base

La popolazione base è quella articolata per sesso, singola classe di età e regione al 1° gennaio 2021, 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 che non risultino iscritti presso alcuna anagrafe.

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 2020 e il 1° gennaio 2070. Scopo principale è fornire indicazioni sul futuro sviluppo della popolazione nel breve termine (2030), quindi quello di fornirne nel medio (2050) e lungo termine (2070), 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 (2020). 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 relative all'Italia. Le ipotesi relative alle regioni, invece, sono state curate dall'Istat sulla base di un'apposita metodologia "ponte" tra le ipotesi nazionali e quelle regionali. Gli esperti che hanno risposto al questionario (con tecnica CAWI), fornendo informazioni utili e complete a definire le ipotesi, sono stati 86. Essi sono stati reclutati volontariamente tra i partecipanti alla 13^a edizione delle Giornate di Studio della Popolazione organizzata dall'Associazione Italiana per gli Studi di Popolazione (AISP), che ha avuto luogo a Milano tra il 24 e il 26 gennaio 2019 presso l'Università Bocconi. In particolare, si tratta di 50 donne e 36 uomini, principalmente occupati nelle Università (21 del Nord Italia, 11 rispettivamente del Centro e del Mezzogiorno e 10 afferenti Università estere) o in altro Ente di Ricerca pubblico (24). L'età media dei rispondenti è pari a 44 anni mentre la loro esperienza lavorativa è pari in media a 16 anni.

In tutte le fasi che hanno riguardato la costruzione dell'impianto metodologico alla base delle previsioni, l'Istat si è avvalsa della concreta cooperazione di Francesco Billari e Rebecca Graziani dell'Università Bocconi di Milano.

Questionario degli esperti e modello probabilistico

Il metodo probabilistico adottato poggia sulle opinioni degli esperti (*expert-based model*) per definire l'evoluzione futura dei più importanti indicatori demografici e rientra nella classe più ampia dei modelli random scenario. Tale modello, utilizzato per la definizione degli scenari probabilistici a livello nazionale, si fonda sull'elicitazione di una serie di parametri da cui viene derivata la futura evoluzione stocastica di ciascuna componente demografica. 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 necessarie sono sintetizzate attraverso i seguenti indicatori: 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.

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 = 2019", "t1 = 2050", "t2 = 2080", generando in questo modo due sotto-intervalli, 2019-2050 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.

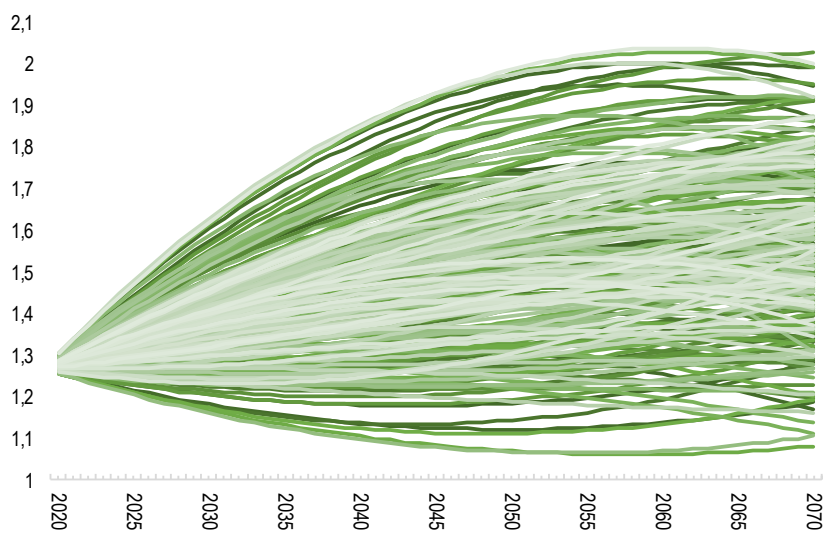
Gli indicatori demografici sono assunti, per semplicità di ipotesi, tra loro indipendenti (ad esempio, la propensione media di fecondità espressa dalle donne non è influenzata dal livello delle migrazioni e viceversa), benché il modello permetta nella sua versione generalizzata la possibilità di far 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 3000 simulazioni col metodo Markov Chain Monte Carlo¹.

¹ La scelta è caduta sugli anni 2050 e 2080 al fine di individuare due intervalli temporali di analoga lunghezza. Per quanto la disponibilità di informazioni raccolte presso gli esperti permetta di estendere l'orizzonte delle previsioni fino al 2080, si è optato di arrestare l'esercizio iterativo di simulazione fino al 2070 onde ricavare un periodo previsivo di 50 anni precisi.

PROSPETTO A1. VALORI MEDI, VARIANZE E CORRELAZIONI OTTENUTE SOTTO LE IPOTESI DEGLI ESPERTI PER TIPO DI INDICATORE. Anni 2019, 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 2019					
Valore osservato	1,27	81,1	85,4	333	180
Anno 2050					
Ipotesi media	1,51	84,7	88,1	256	131
Ipotesi alta	1,75	86,3	89,5	343	172
Varianza	0,034	1,441	1,309	4.593	1.042
Anno 2080					
Ipotesi media condizionata all'ipotesi media 2050	1,55	87,1	90,0	240	127
Ipotesi media condizionata all'ipotesi alta 2050	1,74	88,4	91,0	305	158
Ipotesi alta condizionata all'ipotesi media 2050	1,75	88,5	91,5	348	186
Varianza	0,044	2,180	2,002	7.523	2.675
Correlazione 2050-2080					
Coefficiente di correlazione	0,68	0,66	0,54	0,51	0,46

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 2020-2070

L'ultimo passo di stima è finalizzato al calcolo dei valori di ciascun parametro negli anni intermedi i due intervalli 2019-2050 e 2050-2080. Tale attività è realizzata, per ciascuna delle 3000 simulazioni, mediante interpolazione con curve di tipo quadratico, passanti per i punti noti in corrispondenza degli anni 2019, 2050 e 2080. Si è così pervenuti alla definizione di 3000 curve stocastiche per ogni indicatore demografico considerato a livello nazionale. A titolo di esempio, la Figura A1 descrive il fascio di curve relativo al numero di figli per donna, ricavato dalla procedura sopra descritta.

La scelta di considerare un numero di 3000 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 l'odierna 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.

Relazione tra le previsioni nazionali e regionali

Il modello probabilistico fornisce un set di 3000 simulazioni nazionali per ciascun indicatore demografico di sintesi. Poiché l'obiettivo delle previsioni Istat è anche quello di fornire indicazioni a livello territoriale, proseguendo la tradizione del modello multiregionale, 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, si vedrà più avanti, è di natura *bottom-up* dal lato della produzione degli output finali.

L'azione principale è quella di derivare 3000 scenari regionali stocastici dai 3000 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 maschile e femminile, 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 3000 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 3000 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 3000 vettori di fecondità, e gli altrettanti sulle immigrazioni e le emigrazioni con l'estero e, infine, con le 3000 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 3000 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) 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 (2120, ben oltre l'ultimo anno delle previsioni) tra le regioni per ciascuna componente demografica fondamentale. Ciò comporta che i 3000 scenari stocastici regionali raffigurano 3000 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 (2021-2070). 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 1977-2019;
- per la mortalità, le probabilità di morte per sesso ed età del periodo 1974-2019;
- per le migrazioni interne e internazionali, i trasferimenti di residenza per sesso ed età del 2015-2019.

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* (dal termine *nowcast*=previsione del presente, contrapposto al più noto termine *forecast*=previsione del futuro). 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 (effetto *jump-off*), pur preservando quella che è la potenziale variabilità del fenomeno studiato. Questo genere di operazione 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. Nessun modello previsivo applicato alle serie storiche menzionate al paragrafo precedente, e di cui si dà documentazione in quelli successivi, avrebbe potuto essere in grado di prevedere con precisione lo shock determinato dalla pandemia. Non certamente gli oltre 740 mila decessi riscontrati per tutte le cause nel 2020 (circa 100 mila in più di quanto atteso), ma nemmeno il calo ulteriore delle nascite (405 mila) nel contesto di un quadro complessivo già di per sé compromesso dalla ben nota contrazione dei comportamenti riproduttivi. Né, infine, poteva risultare prevedibile la forte contrazione registrata nelle migrazioni, conseguenti le misure intraprese a livello nazionale per il contenimento della diffusione del virus (*lockdown*).

Dal momento che la popolazione base delle previsioni è quella rilevata al 1° gennaio 2021, si sono rese necessarie operazioni di correzione a breve termine degli input previsivi che hanno interessato i primi anni di previsione. Con ciò si vuole, infatti, non solo tenere conto degli eventi eccezionali che hanno influenzato il 2021, ma anche quelli successivi entro i quali si presuppone che gli effetti pandemici possano andare ad esaurirsi e a far sì di raccordare la serie prevista degli input di breve termine con quelli di medio e lungo².

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:

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

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 2021 i fattori di correzione sono stati costruiti rapportando i dati del bilancio demografico provvisorio di ciascuna regione, rilasciati nel marzo 2022 dall’Istat³, alle previsioni preliminarmente prodotte per tale anno.

Per gli anni successivi al 2021 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:

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

con ϵ quantità arbitraria, scelta per assicurare che in media regionale il numero di anni per garantire il rientro dalle previsioni di breve a quelle di medio-lungo termine non superi i 5 anni. A questo punto, i livelli dei fattori di correzione per gli anni successivi a “b”, per un totale di “Y” anni, sono dati da:

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

Intervalli di confidenza e scenario mediano

Una volta lanciata la procedura di calcolo inerente le 3000 simulazioni regionali, il margine di incertezza viene calcolato per tutti i possibili livelli informativi, dalla popolazione prevista ai dati di flusso, sulla base delle

² Tenendo peraltro presente il meccanismo iterativo dato dal metodo coorti componenti, ovvero un meccanismo di continua interazione stock-flussi nel tempo, la correzione imposta nei primi anni ha effetti anche sui risultati di tutti quelli successivi, fino al 2070 in questo caso.

³ Cfr.: Istat, la dinamica demografica – anno 2021, <https://www.istat.it/it/archivio/267834>.

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 3000 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 sesso.

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 3000 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 2040, 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 1977-2019. 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 stimatane la serie nel periodo 1977-2019. 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 2120. In particolare il vincolo di convergenza prevede che, dal 2020 al 2120, 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.

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 1974-2019, 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 2070 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 1974-2019 con la stessa metodologia e, successivamente, facendo convergere ogni parametro regionale al corrispondente parametro nazionale al 2120. 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, vale a dire sul 2015-2019. Questa necessità, considerando la complessità di prevedere flussi migratori internazionali ricorrendo ad analisi di lunghe serie storiche, porta all'uso di un modello molto semplificato. 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. 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 (2120) al medesimo livello, ossia alla semisomma iniziale dei due valori.

Una volta determinati i totali dei flussi in ingresso e in uscita fino al 2070, le distribuzioni per sesso ed età associate vengono derivate applicando il modello Castro-Rogers (Rogers and Castro, 1981) alla serie 2015-2019. 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 in pieno), 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.

La probabilità di migrazione specifica per età (110), sesso (2), regione di origine (21) e di destinazione (21) costituisce la componente elementare della matrice O/D composta di $110 \times 2 \times 21 \times 21 = 97020$ celle per ciascun anno di calendario. Le probabilità sono stimate sulla base dei livelli osservati nelle singole annualità del periodo 2015-2019. I vettori di probabilità così ottenuti, a livello di ciascuna annualità, sono successivamente perequati utilizzando la funzione Castro-Rogers.

Pertanto, indicando con

$$m_{x,s,t}^{i,j}$$

la generica probabilità prospettiva di migrare per un individuo di età "x" e sesso "s" tra la regione "i" e la regione "j" relativa all'annualità "t" (t=2015, ..., 2019), 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:

$$\mu_{x,s}^{i,j} = E(m_{x,s,t}^{i,j})$$

$$\sigma_{x,s}^{i,j} = E(m_{x,s,t}^{i,j} - \mu_{x,s}^{i,j})^2$$

Dalle suddette variabili casuali vengono estratti casualmente 3000 valori per ciascuno dei 97020 elementi della matrice di O/D, dando così luogo alla creazione casuale di 3000 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 sesso, 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. L'ipotesi alla base del modello si fonda, infatti, sul mantenimento per tutto il periodo previsivo di una propensione alla mobilità che rimanga costante. Ciò comporta che i flussi migratori interni evolvano nel tempo solo in virtù delle variazioni che interessano livello e struttura per età della popolazione esposta al rischio di migrare.

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 2020 e 2021.

In primo luogo va rilevata una differenza piuttosto contenuta tra la popolazione totale base 2021 (59 milioni 236mila) e quella che era stata stimata nello scenario mediano alla stessa data dalle previsioni in base 2020 (59 milioni 249mila).

Dal lato dei flussi previsti nel tratto di comune proiezione (2021-2070) si coglie una valutazione lievemente più positiva nelle previsioni in base 2021, dove ad esempio si prevedono 19,6 milioni di nascite contro le 19,5 dell'esercizio precedente. Anche le altre poste di bilancio, per quanto non tali da stravolgere gli esiti tendenziali emersi dalle previsioni in base 2020, sono più favorevoli per l'esercizio in base 2021. Quest'ultimo, infatti, presenta nel comune tratto proiettivo 69mila decessi in meno, 65 mila immigrati dall'estero in più e 19mila emigrati per l'estero in meno.

Anche la differenza tra le popolazioni finali dei due distinti esercizi previsivi è di lieve entità (136 mila unità al 1° gennaio 2070 per lo scenario mediano base 2021), confermando la sostanziale tenuta delle previsioni in base 2020, nonostante il cambio di popolazione base e gli aggiustamenti di breve termine sulle componenti di bilancio. Sotto questo punto di vista il prospetto A2 mette in risalto come il processo di rivisitazione delle ipotesi per tutte le componenti demografiche abbia interessato soltanto i primi anni di previsione.

PROSPETTO A2. CONFRONTO TRA LE IPOTESI SUI PRINCIPALI INDICATORI SINTETICI DELLO SCENARIO MEDIANO IN BASE 2020 E DELLO SCENARIO MEDIANO IN BASE 2021. Anni 2021, 2030, 2050 e 2070.

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 2021					
Base 2020	1,21	80,1	84,9	271	130
Base 2021	1,25	80,0	84,6	286	129
Anno 2030					
Base 2020	1,37	82,2	86,2	279	146
Base 2021	1,37	82,2	86,2	281	145
Anno 2050					
Base 2020	1,50	84,7	88,1	258	131
Base 2021	1,50	84,7	88,1	258	131
Anno 2070					
Base 2020	1,55	86,5	89,5	244	126
Base 2021	1,55	86,5	89,5	244	126

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 2019, 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. In questo caso le ultime previsioni disponibili sono in base 2021 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, in primo luogo, il fatto che le previsioni di Eurostat hanno come base di riferimento il 1° gennaio 2019, ovvero proiettano in partenza una popolazione non allineata alle risultanze del censimento 2020 e, in particolare, rispetto a queste ultime una popolazione significativamente più alta. In secondo luogo, va menzionato 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).

Per tutte le componenti demografiche in gioco le ipotesi adottate sono in partenza molto differenziate tra Istat e Eurostat/UNPD. Ciò si deve al fatto che, a differenza dello scenario Istat, le previsioni Eurostat non scontano lo shock demografico del 2020 prodotto dalla pandemia Covid-19. Le previsioni UNPD, a loro volta, 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. Nel medio e lungo termine, di fatto, le ipotesi continuano a risultare piuttosto differenziate tra i vari enti produttori. In particolare, nei confronti dei movimenti migratori, laddove rispetto a una UNPD che si dimostra piuttosto cauta sull'Italia, si contrappone Eurostat con una visione molto più ottimistica. Tale evidenza si deve in parte alla metodologia di Eurostat, la quale, oltre a prevedere l'evoluzione di fondo delle migrazioni nette, incorpora nel modello una componente additiva di *replacement-migration*⁴.

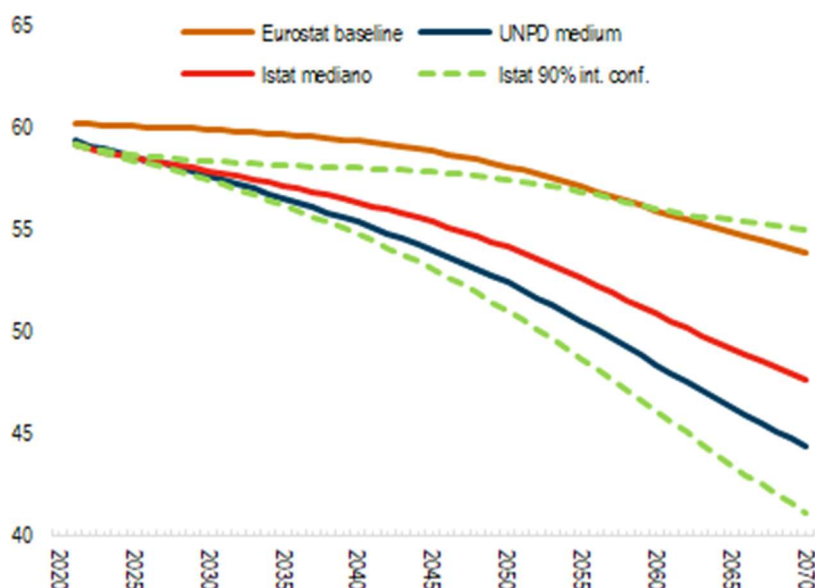
⁴ Tale componente assegna in ogni anno di previsione una quota aggiuntiva di migranti netti nella misura pari al 10% della riduzione riscontrata nella popolazione in età attiva (15-64 anni).

Le ipotesi sulla fecondità sono abbastanza simili, per quanto nel medio-lungo termine Eurostat e UNPD producano previsioni meno favorevoli rispetto a Istat. Le ipotesi sulla sopravvivenza sono anch'esse non particolarmente distanti, tuttavia Eurostat e soprattutto UNPD evidenziano aspettative molto favorevoli circa l'allungamento della vita media, che nel modello Istat si intravedono solo in parte.

PROSPETTO A3. CONFRONTO TRA LE IPOTESI DEI PRINCIPALI INDICATORI SINTETICI TRA LO SCENARIO MEDIANO ISTAT IN BASE 2021, LO SCENARIO EUROSTAT IN BASE 2019 E LO SCENARIO UNPD IN BASE 2021. Anni 2021, 2030, 2050 e 2070.

Scenario	Numero medio di figli per donna	Speranza di vita alla nascita – Uomini	Speranza di vita alla nascita – Donne	Saldo migratorio estero (migliaia)
Anno 2021				
Istat Mediano	1,25	80,0	84,6	157
Eurostat Baseline	1,33	81,4	85,8	178
UNPD Medium	1,28	80,5	85,1	28
Anno 2030				
Istat Mediano	1,37	82,2	86,2	136
Eurostat Baseline	1,37	82,6	86,9	224
UNPD Medium	1,35	83,2	87,1	58
Anno 2050				
Istat Mediano	1,50	84,7	88,1	127
Eurostat Baseline	1,45	84,9	89,0	214
UNPD Medium	1,44	85,8	89,6	58
Anno 2070				
Istat Mediano	1,55	86,5	89,5	118
Eurostat Baseline	1,52	87,0	90,9	207
UNPD Medium	1,48	88,2	91,9	58

FIGURA A2. POPOLAZIONE TOTALE PREVISTA SECONDO GLI SCENARI ISTAT, EUROSTAT E UNPD. Anni 2021-2070, milioni di residenti



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. Purtroppo, lo scarto iniziale dovuto alla differente base di popolazione delle previsioni Eurostat rende piuttosto spurio il confronto con lo scenario Istat. A loro volta, le previsioni UNPD, pur allineate in partenza a quelle dell'Istat, danno un'evoluzione della popolazione molto più pessimista. 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. Lo scenario Eurostat, stante il significativo impatto di una previsione riguardo al saldo migratorio con l'estero più sostenuta, risulta più ottimista. Fino al punto da mantenere una popolazione più ampia persino del limite superiore dell'intervallo di confidenza al 90% delle previsioni Istat per gran parte dell'orizzonte previsivo. Lo scenario UNPD, al contrario, tende a collocarsi circa a metà tra lo scenario mediano Istat e il suo intervallo di confidenza inferiore.

Diffusione dati e termini di utilizzo

Il quadro dettagliato delle ipotesi sottostanti le previsioni e dei principali risultati è consultabile sia sul sito generalista di Istituto dati.istat.it (tema: Popolazione e famiglie > Previsioni demografiche) sia sul sito tematico demo.istat.it.

La diffusione dei dati 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 riproduzione delle informazioni contenute nella presente nota e nelle banche dati dati.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).

Glossario

Campo di variazione (*range*): misura della variabilità di un fenomeno quantitativo definita dalla differenza tra il valore massimo e il valore minimo osservato.

Coppia: due persone legate da relazione affettiva e sentimentale. Può essere formata da persone di sesso opposto o dello stesso sesso. I vincoli tra persone in coppia possono essere formali (coppia *de iure*: coniugati, uniti civilmente o conviventi di fatto ai sensi della legge 76/2016) o informali (coppia *de facto*).

Crescita naturale (tasso di): differenza tra il tasso di natalità e il tasso di mortalità.

Crescita totale (tasso di): somma del tasso migratorio netto totale e del tasso di crescita naturale.

Coorti componenti (modello): algoritmo di calcolo continuo che in modalità iterativa simula l'evoluzione dell'equazione fondamentale della popolazione per classi di età, consentendo di determinare le poste demografiche oggetto di previsione e di ottenere la popolazione superstite alla fine di ogni anno.

Decesso: cessazione di ogni segno di vita in un qualsiasi momento successivo alla nascita vitale.

Dipendenza anziani (indice di): rapporto tra la popolazione di 65 anni e più e la popolazione di 15-64 anni, moltiplicato per 100.

Dipendenza strutturale (indice di): rapporto tra la popolazione in età non attiva (0-14 anni e 65 anni e più) e la popolazione in età attiva (15-64 anni), moltiplicato per 100.

Emigratorio interno (tasso): rapporto tra il numero di cancellati per l'interno e l'ammontare medio della popolazione residente, moltiplicato per 1.000.

Emigratorio per l'estero (tasso): rapporto tra il numero di cancellati per l'estero e l'ammontare medio della popolazione residente, moltiplicato per 1.000.

Età media: età media della popolazione detenuta a una certa data espressa in anni e decimi di anno.

Età media al parto: età media al parto delle madri espressa in anni e decimi di anno, calcolata considerando i soli nati vivi.

Famiglia: insieme di persone legate da vincoli di matrimonio, parentela, affinità, adozione, tutela, affettivi, coabitanti e aventi dimora abituale nello stesso Comune. Può essere costituita anche da una sola persona.

Famiglie con nuclei: comprende le coppie con figli, le coppie senza figli, i genitori soli, le famiglie con due o più nuclei.

Famiglie senza nuclei: comprende famiglie formate da una sola persona e le famiglie multipersonali, ossia che non costituiscono un nucleo familiare anche se composte da più persone.

Fecondità per età (tasso specifico di): il rapporto tra il numero di nati vivi da donne di età compresa tra x e $x+1$ e il numero medio di donne residenti di tali età in un dato anno.

Immigratorio dall'estero (tasso): rapporto tra il numero di iscritti dall'estero e l'ammontare medio della popolazione residente, moltiplicato per 1.000.

Immigratorio interno (tasso): rapporto tra il numero di iscritti dall'interno e l'ammontare medio della popolazione residente, moltiplicato per 1.000.

Intervallo predittivo (o di confidenza): intervallo associato a una variabile casuale ancora da osservare, con una specifica probabilità che la variabile casuale vi ricada all'interno.

Iscrizione e cancellazione anagrafica per trasferimento di residenza: iscrizione riguarda le persone trasferitesi in un Comune da altri Comuni o dall'estero; la cancellazione riguarda le persone trasferitesi in altro Comune o all'estero.

Migratorio netto con l'estero (tasso): differenza tra il tasso immigratorio dall'estero e il tasso emigratorio con l'estero.

Migratorio netto con l'interno (tasso): differenza tra il tasso immigratorio dall'interno e il tasso emigratorio per l'interno.

Migratorio netto totale (tasso): somma del tasso migratorio netto con l'interno e del tasso migratorio netto con l'estero.

Mortalità (tasso di): rapporto tra il numero dei decessi nell'anno e l'ammontare medio della popolazione residente, moltiplicato per 1.000.

Nato vivo: prodotto del concepimento che, una volta espulso o completamente estratto dal corpo materno, indipendentemente dalla durata della gestazione, respiri o manifesti altro segno di vita.

Natalità (tasso di): rapporto tra il numero dei nati vivi nell'anno e l'ammontare medio della popolazione residente, moltiplicato per 1.000.

Nucleo familiare: insieme delle persone che formano una relazione di coppia o di tipo genitore-figlio. Si intende la coppia coniugata, unita civilmente o convivente, senza figli o con figli celibi o nubili, o anche un solo genitore assieme ad uno o più figli mai sposati. Nell'ambito di una famiglia possono esistere uno o più nuclei familiari, ma può anche non esservene nessuno, come nel caso delle famiglie formate da un membro isolato (famiglie monocomponenti) o più membri isolati (altre persone residenti).

Numero medio di figli per donna: numero di figli che una donna metterebbe al mondo nel caso in cui, nel corso nella propria vita riproduttiva, fosse sottoposta al calendario di fecondità (sotto forma di tassi specifici di fecondità per età) dell'anno di osservazione.

Previsione: sviluppo atteso nel futuro.

Previsione demografica: elaborazione che mostra lo sviluppo futuro di una popolazione quando vengono assunte determinate ipotesi riguardo al futuro corso della mortalità, della fecondità e della migratorietà.

Previsione demografica deterministica: elaborazione sul futuro sviluppo di una popolazione, riassumibile in una singola serie di valori ottenuti da un singolo set di ipotesi demografiche, che non riporta alcuna misura riguardo all'incertezza usualmente associabile ai risultati.

Previsione demografica probabilistica: elaborazione sul futuro sviluppo di una popolazione, riassumibile in un set di valori o in una distribuzione di probabilità, in cui le variabili utilizzate sono di natura casuale che non possono essere previste con certezza e in cui non tutte le ipotesi sono egualmente probabili.

Popolazione residente: costituita in ciascun Comune (e analogamente per altre ripartizioni territoriali) delle persone aventi dimora abituale nel Comune stesso. Non cessano di appartenere alla popolazione residente le persone temporaneamente dimoranti, in altro Comune o all'estero, per l'esercizio di occupazioni stagionali o per causa di durata limitata.

Probabilità (prospettiva) di migrazione interregionale: probabilità che un individuo di età x (in anni compiuti al 1° gennaio) sposti residenza tra due regioni prima che si concluda l'anno.

Probabilità (classica) di morte: probabilità che un individuo di età precisa x muoia prima del compimento del compleanno $x+1$.

Probabilità (prospettiva) di morte: probabilità che un individuo di età x (in anni compiuti al 1° gennaio) non sopravviva entro l'anno.

Saldo migratorio con l'estero: differenza tra il numero degli iscritti per trasferimento di residenza dall'estero e il numero dei cancellati per trasferimento di residenza all'estero.

Saldo migratorio interno: differenza tra il numero degli iscritti per trasferimento di residenza da altro Comune e il numero dei cancellati per trasferimento di residenza in altro Comune.

Saldo migratorio totale: somma del saldo migratorio con l'estero e del saldo migratorio interno.

Saldo naturale (o dinamica naturale): differenza tra il numero di nascite e il numero di decessi.

Saldo totale: somma del saldo naturale e del saldo migratorio totale.

Scenario previsivo: descrizione del contesto, anche concettuale, nel quale la popolazione viene proiettata. In un approccio deterministico normalmente si riferisce all'ipotesi definita principale o centrale. In uno stocastico può riferirsi all'ipotesi identificata come media o mediana.

Simulazione: implementazione quantitativa di un singolo set di ipotesi demografiche da lanciare nel modello coorti-componenti al fine di ottenere un singolo set di previsioni demografiche.

Speranza di vita alla nascita (o vita media): numero medio di anni che una persona può contare di vivere dalla nascita nell'ipotesi in cui, nel corso della propria esistenza, fosse sottoposta ai rischi di mortalità per età dell'anno di osservazione.

Speranza di vita all'età "x": numero medio di anni che una persona di età compiuta "x" può contare di sopravvivere nell'ipotesi in cui, nel corso della successiva esistenza, fosse sottoposta ai rischi di mortalità per età (dall'età "x" in su) dell'anno di osservazione.

Tipologia familiare: classificazione in base alla presenza o meno di almeno un nucleo familiare e per tipo di nucleo.

Vecchiaia (indice di): rapporto tra la popolazione di 65 anni e più e la popolazione di età 0-14 anni, moltiplicato per 100.