

NOTA METODOLOGICA

La matrice del pendolarismo per motivi di studio relativa all'edizione 2021 del Censimento permanente della popolazione e delle abitazioni.

1. Introduzione

La matrice origine-destinazione per motivi di studio 2021 contiene i dati riferiti ai flussi intra ed extra-comunali di individui residenti - in famiglia o in convivenza - che si spostano per almeno tre giorni a settimana tra il luogo della propria dimora abituale e il luogo abituale di studio. Si precisa che la popolazione di riferimento è costituita dagli individui non occupati; restano pertanto esclusi gli studenti-lavoratori. Le stime tengono esclusivamente conto degli spostamenti sul territorio italiano escludendo gli spostamenti con l'estero.

Il profondo cambiamento introdotto dal Censimento permanente della popolazione e delle abitazioni ha implicato una revisione concettuale e metodologica del processo di costruzione degli aggregati di diffusione. La nuova strategia mira a massimizzare l'integrazione tra le informazioni contenute nelle indagini campionarie del Censimento permanente con quelle presenti nei registri statistici e negli archivi amministrativi, così da assicurare, oltre a una solida base inferenziale, anche la completa coerenza del sistema delle statistiche censuarie.

A differenza del processo adottato nel 2011¹, basato sul conteggio completo della popolazione che si sposta per motivi di studio, a partire dal 2018 è stato utilizzato un approccio di tipo predittivo².

L'approccio predittivo si fonda sull'utilizzo di modelli statistici in grado di integrare informazioni provenienti da fonti diverse, comprese quelle di natura amministrativa, e di sfruttare le relazioni osservate tra variabili al fine di stimare con la maggiore precisione possibile gli indicatori di interesse. L'intero processo ha richiesto due passaggi principali (oltre ad una serie di raffinamenti di minor portata ma fondamentali): da un lato la stima dei flussi provincia-provincia effettuata tramite un modello di stima per piccole aree; dall'altro l'utilizzo di un modello multinomiale per la stima degli ammontari dei pendolari interni ed esterni a livello comunale.

¹ Si veda <https://www.istat.it/produzione-editoriale/atti-del-15-censimento-generale-della-popolazione-e-delle-abitazioni/>

² https://www.istat.it/wp-content/uploads/2021/05/Censimento_spostamenti_pendolari.pdf

2. Fonti

Per il processo di stima della matrice di origine destinazione per motivo di studio è stato utilizzato un modello basato sull'integrazione tra i dati campionari e le seguenti fonti di dati:

- l'indagine campionaria Master Sample del Censimento permanente della popolazione e delle abitazioni;
- il Registro Base degli Individui per ciò che concerne le variabili demografiche sesso, età, cittadinanza e residenza;
- archivi amministrativi (fonti: MIM e MUR): Anagrafe Nazionale degli Studenti delle scuole, Anagrafe delle scuole statali e non statali, degli istituti comprensivi di ogni ordine e grado, Anagrafe Nazionale degli studenti dei corsi post laurea, Anagrafe Nazionale degli studenti universitari, Anagrafe delle università;
- la Banca Dati Integrata da Archivi Amministrativi sui Percorsi d'Istruzione e Titolo di Studio;
- la variabile relativa all'occupazione prodotta dal Censimento permanente³ per l'identificazione degli individui non occupati;
- la matrice di distanza intercomunale in termini di tempi di percorrenza e distanze chilometriche stradali⁴;
- variabili comunali ausiliarie caratterizzanti il comune di residenza: tipologia comunale (capoluogo di regione, provincia, non capoluogo), grado di urbanizzazione⁵ (zone densamente popolate, zone a densità intermedia di popolazione, zone scarsamente popolate), classe del comune in termini di distanza dai poli con un'offerta congiunta delle tipologie di servizio salute, istruzione e mobilità (polo, polo intercomunale, cintura, intermedio, periferico, ultra-periferico)⁶, presenza di scuole per tipologia di istituto a livello comunale.

3. Fasi del processo di stima

Le fasi del processo di stima sono:

- A. Pretrattamento delle stime dirette dei flussi origine-destinazione tra province;
- B. Stima da modello dei flussi origine-destinazione tra province;
- C. Stima del totale di pendolari a livello comunale, distinti per pendolari entro il comune e fuori del comune (entro il territorio italiano);
- D. Stima dei flussi origine-destinazione tra coppie di comuni

³ Si veda <https://www.istat.it/it/files//2020/12/NOTA-TECNICA-CENSIPOP.pdf>

⁴ Si veda <https://www.istat.it/notizia/matrici-di-contiguita-distanza-e-pendolarismo/>

⁵ Si veda <https://www.istat.it/comunicato-stampa/geografie-funzionali-per-lanalisi-territoriale/>

⁶ Si veda <https://situas.istat.it/web/#/home>; <https://www.istat.it/comunicato-stampa/la-geografia-delle-aree-interne-nel-2020-vasti-territori-tra-potenzialita-e-debolezze/>

A. Pretrattamento delle stime dirette dei flussi origine-destinazione tra province

Per ogni coppia di province, le stime dirette dei flussi origine–destinazione per motivi di studio sono state ricavate utilizzando il comune di residenza e il comune di studio rilevati nell'indagine Master Sample. Considerato che, per numerose coppie di province, tali stime derivano da un numero molto limitato di unità campionarie, è stato implementato un processo di aggiustamento volto a evitare che l'applicazione diretta dei pesi di riporto all'universo generasse valori del flusso eccessivamente amplificati.

Le stime dirette sono state pertanto sottoposte a un procedimento di smoothing basato sulla numerosità osservata per ciascuna coppia origine-destinazione e sulla distanza tra le province coinvolte, al fine di garantire una maggiore stabilità e coerenza delle stime.

B. Stima da modello dei flussi origine-destinazione tra province

Le stime dirette, pur corrette secondo il processo descritto al punto A, presentano in diversi casi livelli di variabilità residua non compatibili con gli standard di precisione richiesti. Per tale motivo si è proceduto all'applicazione di metodi di stima per piccole aree, con l'obiettivo di migliorare l'accuratezza delle stime rispetto a quanto ottenibile mediante le sole informazioni campionarie. In questo quadro, i domini di stima sono costituiti dalle coppie origine–destinazione di province.

La metodologia adottata si basa su modelli a livello di area, mediante un'estensione del modello ad effetti misti di Fay–Herriot (Fay e Herriot, 1979; Datta e Mandal, 2015; Tang et al., 2018). Sono stati stimati modelli distinti per i flussi intra-provincia e per i flussi inter-provincia, impiegando come variabili ausiliarie i flussi rilevati nel precedente censimento e quelli derivati da fonti amministrative, corretti in funzione della distanza o dei tempi di percorrenza tra province. Al fine di attenuare eventuali eterogeneità strutturali e ridurre possibili effetti distorsivi, la procedura di stima è stata articolata in modelli separati per le province di origine appartenenti alle macro-aree Nord, Centro e Sud.

C. Stima del numero di pendolari a livello comunale (dentro/fuori il comune)

Per garantire la coerenza tra la matrice comune-comune e la matrice provinciale dei flussi, in continuità con quanto realizzato per gli anni 2018 e 2019, è stato stimato un modello a livello di individuo in grado di sfruttare in modo ottimale le potenzialità dell'informazione ausiliaria amministrativa. Analogamente alle procedure adottate negli anni precedenti, la stima del numero di pendolari interni ed esterni al comune per motivi di studio è stata ottenuta mediante modelli logistici multinomiali, considerando la sottopopolazione dei non occupati 2021 individuata dal Censimento della popolazione, che costituisce il vincolo di bilancio dei flussi.

I modelli integrano le informazioni provenienti dalle fonti di dati sopra elencate e includono:

- variabili individuali quali età, sesso, cittadinanza (italiana/straniera), il comune o i comuni in cui è osservato nel corso dell'anno il segnale di studio desunto dalle fonti amministrative, nonché la distanza tra il comune di residenza e il luogo di studio;

- variabili comunali quali la tipologia comunale, le informazioni derivanti dal Censimento 2011 e la presenza di scuole per tipologia di istituto.

Per contenere possibili effetti distorsivi nelle stime, i modelli sono stati stimati separatamente per ciascuna regione e distintamente per le classi di età 6-16 anni e no 6-16 anni.

Non sono state prodotte le stime degli spostamenti verso l'estero, poiché non è stato possibile addestrare in modo adeguato i modelli di stima in quanto i dati campionari disponibili non permettevano di raggiungere il livello di accuratezza richiesto.

D. Stima dei flussi origine destinazione tra coppie di comuni

Le stime dei flussi finali di origine destinazione a livello comunale sono state ottenute attraverso un processo a due stadi. Al primo stadio (D1), i flussi tra province stimati al punto B sono stati ripartiti tra i comuni appartenenti a ciascuna coppia di province, sfruttando il contributo dei flussi amministrativi. La procedura di redistribuzione è stata definita in modo da riflettere le differenti probabilità di spostamento quotidiano verso il luogo di studio, determinate dall'intensità del segnale amministrativo tra il comune di residenza e il luogo di studio. Tale approccio consente di rappresentare in maniera più aderente le dinamiche effettive di mobilità, tenendo conto dei singoli territori in termini di diffusione scolastico-universitaria e propensione/necessità di spostamento. Al secondo stadio (D2), le stime comunali ottenute al punto D1 sono state ricalibrate per garantire l'allineamento con le stime comunali di pendolari interni ed esterni al comune ottenute al punto C.

Riferimenti Bibliografici

Datta, G.S. e Mandal, A. (2015). Small area estimation with uncertain random effects. *Journal of the American Statistical Association*, 110(512), 1735–1744.

Fay, R.E. e Herriot, R.A. (1979). Estimates of income for small places: an application of James-Stein procedures to census data. *Journal of the American Statistical Association*, 74(366a), 269–277.

Tang, X., Ghosh, M., Ha, N.S. e Sedransk, J. (2018). Modeling random effects using global-local shrinkage priors in small area estimation. *Journal of the American Statistical Association*, 113(1), 1476–1489.
