

## NOTA METODOLOGICA

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

### 1. *Introduzione*

La matrice origine-destinazione degli spostamenti per motivi di lavoro si riferisce alla popolazione residente in famiglia o in convivenza. Il file contiene i dati, riferiti al 31/12/2021, sul numero di individui che si spostano tra comuni o all'interno dello stesso comune, per almeno tre giorni a settimana, al luogo abituale di lavoro partendo dall'alloggio di residenza nel quale rientrano giornalmente. Le stime tengono esclusivamente conto degli spostamenti sul territorio italiano escludendo gli spostamenti con l'estero.

La matrice è stata utilizzata per la definizione dei sistemi locali del lavoro 2021. 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 2011, basato sul conteggio completo della popolazione che si sposta per motivi di lavoro, a partire dal 2018 è stato utilizzato un approccio di tipo predittivo.

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 maggiore precisione gli indicatori di interesse. Stimatori basati su modelli a livello di area sono stati utilizzati per stimare i flussi aggregati a livello provinciale, modelli a livello individuale sono stati invece utilizzati per stimare la probabilità di ciascun occupato di essere pendolare e, tramite aggregazione di questa, il numero di pendolari entro e fuori ciascun comune.

### 2. *Fonti*

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

dati:

- il Registro Base degli Individui per ciò che concerne le variabili demografiche sesso, età, cittadinanza e residenza;
- dati amministrativi sul lavoro da fonti previdenziali (gestioni INPS dei dipendenti pubblici e privati) e fonti fiscali (Certificazione Unica e Anagrafe Tributaria) per ciò che concerne l'individuazione del comune di lavoro;
- la variabile relativa all'occupazione prodotta dal Censimento permanente<sup>1</sup>;
- la matrice di distanza intercomunale in termini di tempi di percorrenza e distanze chilometriche stradali<sup>2</sup>;
- variabili comunali ausiliarie caratterizzanti il comune di residenza, quali la tipologia comunale (capoluogo di regione, provincia, non capoluogo), il grado di urbanizzazione (alto, medio, basso) e la classe del comune in termini di distanza dai poli di attrazione (polo, polo intercomunale, cintura, intermedio, periferico, ultra-periferico)<sup>3</sup>.

### **3. Fasi del processo di stima**

Le fasi del processo di stima sono:

- A. Pretrattamento dei dati amministrativi;
- B. Stima 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 (ma entro il territorio italiano);
- D. Stima dei flussi origine-destinazione tra coppie di comuni

#### **A. Pretrattamento dei dati amministrativi**

Per ogni individuo, le traiettorie amministrative degli spostamenti per motivi di lavoro sono state calcolate considerando il comune di residenza, come indicato nel Registro Base degli Individui. Soltanto gli individui residenti sono stati considerati. Il comune di lavoro è stato calcolato a partire dall'insieme dei dati amministrativi contenenti informazioni sul luogo di lavoro alla data del 31/12/2021. Nel caso di individui che svolgono lavori in più comuni simultaneamente, un solo comune di lavoro è stato selezionato, tenendo conto delle caratteristiche dei rapporti di lavoro (durata, full-time/part-time, retribuzione).

La pseudo-matrice degli spostamenti per motivi di lavoro è stata calcolata a partire dai dati del censimento e dalle traiettorie amministrative. Utilizzando le informazioni sul carattere pendolare di ogni rispondente al Censimento della popolazione 2021 e le distanze chilometriche fra comuni italiani, è stato stimato il tasso di pendolarismo per

---

<sup>1</sup> si veda <https://www.istat.it/it/files//2020/12/NOTA-TECNICA-CENSIPOP.pdf>

<sup>2</sup> si veda <https://www.istat.it/notizia/matrici-di-contiguita-distanza-e-pendolarismo/>

<sup>3</sup> si veda <https://www.istat.it/classificazione/principali-statistiche-geografiche-sui-comuni/>

Divisioni della classificazione delle attività economiche (ATECO) e classe di distanza fra comuni di origine e destinazione (ampiezza di classe pari a 50 km di distanza). La pseudo matrice di pendolarismo è stata, quindi, calcolata sommando, per ogni coppia di comune-origine-comune destinazione, gli spostamenti registrati nei dati amministrativi corretti con i corrispondenti tassi di pendolarismo per classi ATECO e classe dimensionale. Infine, per ridurre il numero delle traiettorie caratterizzate da flussi di intensità esigua, la pseudo matrice di pendolarismo per motivi di lavoro è stata sottoposta ad una procedura statistica di concentrazione dei flussi in base alla loro intensità e ai tempi di percorrenza (Lenormand *et al.*, 2014).

## B. Stima dei flussi origine – destinazione tra province

Il metodo di stima dei flussi provinciali è una estensione del modello ad effetti misti a livello di area di Fay Herriot (Fay e Herriot, 1979; Datta e Mandal, 2015; Tang *et al.*, 2018). Sono stati adattati modelli distinti per stimare i flussi intra provincia e i flussi tra province<sup>4</sup>.

I modelli area level sono solitamente utilizzati per fornire stime affidabili per i domini di stima quando sono disponibili stime dirette da indagine per ogni dominio di stima e quando sono disponibili variabili ausiliarie note a livello di area e supposte correlate con il fenomeno di interesse. Nel caso in questione le piccole aree sono individuate dalle coppie di province origine-destinazione. Tra i modelli per piccole aree, il modello Fay-Herriot è tra i più comunemente impiegati; tale modello combina le stime dirette campionarie con variabili ausiliarie disponibili a livello aggregato. Nel caso in questione, sono state utilizzate come variabili ausiliarie i flussi derivati da dati amministrativi e flussi relativi al precedente censimento. Al fine di ridurre eventuali effetti distorsivi, sono stati stimati modelli separati per le province di origine del Nord, Centro e Sud.

La matrice provinciale viene stimata a partire da informazioni ausiliarie disponibili solo a livello aggregato, perdendo gran parte della ricchezza informativa originariamente contenuta nei dati individuali.

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

Per far sì che la matrice comune-comune derivata da quella provinciale sia coerente le stime del numero di pendolari diffuso per gli anni 2018 e 2019, è stato stimato un modello a livello di individuo in grado di sfruttare al meglio le potenzialità dell'informazione ausiliaria amministrativa. Analogamente a quanto fatto nel 2018 e nel 2019<sup>5</sup>, la stima del numero di pendolari interni ed esterni al comune per motivo di lavoro è stata ottenuta utilizzando modelli logistici multinomiali, prendendo in considerazione la sottopopolazione degli occupati individuati dal Censimento della popolazione.

I modelli tengono conto delle informazioni provenienti dalle fonti di dati sopra elencate e, in particolare: variabili individuali quali l'età, il sesso, la cittadinanza (italiana

---

<sup>4</sup> si veda presentazione al *Third Workshop on methodologies for Official Statistics*, Roma, ISTAT, 4-5/12/2024, [https://www.istat.it/wp-content/uploads/2024/11/2\\_solari\\_invited.pdf](https://www.istat.it/wp-content/uploads/2024/11/2_solari_invited.pdf)

<sup>5</sup> si veda <https://www.istat.it/comunicato-stampa/gli-spostamenti-per-motivi-di-studio-o-lavoro-nel-2019-secondo-il-censimento-permanente-della-popolazione>

/straniera), il comune o i comuni in cui viene osservato nel corso dell'anno il segnale di lavoro desunto dalle fonti amministrative, le distanze dei luoghi di lavoro dal comune di residenza; variabili comunali quali la tipologia comunali e le informazioni relative al censimento 2011. Per limitare effetti distorsivi nelle stime, i modelli sono stati stimati separatamente per ciascuna regione.

Poiché i dati amministrativi forniscono informazioni soltanto per i comuni italiani, le stime tengono esclusivamente conto degli spostamenti sul territorio italiano escludendo gli spostamenti con l'estero.

#### D. Stima dei flussi origine destinazione tra coppie di comuni

I flussi tra province stimati nel passo B sono redistribuiti tra i comuni di ogni coppia di province, tenendo conto della pseudo matrice di pendolarismo (si veda il passo A).

In altre parole, la redistribuzione dei flussi è stata realizzata in modo tale da tenere conto delle diverse probabilità di spostamento quotidiano verso il luogo di lavoro, a seconda della tipologia di attività economica in cui ciascun individuo è coinvolto e della corrispondente distanza stradale. Questo approccio permette di affinare ulteriormente la stima del fenomeno del pendolarismo, garantendo che la distribuzione tra i comuni rifletta in maniera più precisa le reali dinamiche di mobilità, tenendo conto delle specifiche caratteristiche occupazionali e del diverso comportamento di pendolarismo associato a ciascun settore economico.

La distribuzione dei flussi provinciali tra i comuni di ciascuna coppia di province è stata attuata attraverso la selezione probabilistica delle tratte comunali utilizzando la distribuzione multinomiale con probabilità proporzionali alle intensità delle tratte della pseudo matrice del passo A.

La coerenza con le stime comunali del numero di pendolari interni ed esterni al comune ottenute al punto C è stata garantita riproporzionando i flussi comunali ottenuti mediante la selezione dalla variabile multinomiale con le stime degli spostamenti comunali interni ed esterni.

#### 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.

Lenormand, M., Huet, S. e Gargiulo, F. (2014). Generating French virtual commuting networks at the municipality level. *Journal of Transport and Land Use*, 7(1), 43–55. <https://doi.org/10.5198/jtlu.v7i1.360>.

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.