

16 Giugno 2026

## Stime per piccole aree delle persone con gravi limitazioni funzionali: un'applicazione per provincia, sesso e classe di età<sup>1</sup>

### 1. Obiettivi del lavoro

Il presente lavoro descrive una sperimentazione metodologica finalizzata alla produzione di stime territoriali ad alta disaggregazione per l'indicatore di Prevalenza di Persone con Gravi Limitazioni Funzionali (GLF) rilevato attraverso l'ultima edizione disponibile dell'Indagine Europea sulla Salute (EHIS). L'obiettivo è stimare con affidabilità la prevalenza del fenomeno a livello provinciale, con disaggregazione per sesso e per due macro classi di età (15–64 anni e 65 anni e oltre), facendo ricorso a tecniche di stima indiretta basate su modelli per piccole aree (Small Area Estimation – SAE). Tali metodi consentono di superare i limiti delle sole stime dirette, spesso non sufficientemente precise nei piccoli domini, offrendo un'informazione più solida e dettagliata per l'identificazione dei bisogni di salute e la programmazione dei servizi assistenziali nei territori.

La crescente esigenza di dati granulari per supportare le decisioni pubbliche, unita alla disponibilità di nuove fonti ausiliarie a livello subregionale, ha spinto l'Istat ad intensificare lo sviluppo di soluzioni metodologiche avanzate, volte a migliorare la capacità descrittiva del sistema statistico e a potenziare la produzione di stime a livelli territoriali più dettagliati. Le indagini campionarie, come quelle sulla salute, costituiscono la base per costruire indicatori sanitari, ma la loro capacità di fornire stime affidabili è limitata al livello regionale, a causa della necessità di bilanciare costi e ampiezza campionaria. Tuttavia, in ambito sanitario, è spesso necessario disporre di dati più dettagliati per monitorare l'equità all'interno delle regioni e programmare l'offerta di servizi, specialmente in un contesto di invecchiamento demografico. In Italia, la percentuale di over-75enni (12,6%) supera la quota di tutti i paesi Ue, la cui media è pari al 10,2% (dato provvisorio 2024), rendendo ancor più rilevante la possibilità di avere informazioni statistiche granulari sulla salute e sull'autonomia della popolazione anziana, per prevedere l'impatto sul welfare, tradizionalmente sostenuto dalla famiglia.

Le indagini campionarie, come l'EHIS, rilevano indicatori in linea con la classificazione ICF (International Classification of Functioning, Disability and Health) e adottano un approccio

bio-psico-sociale, che considera anche i fattori ambientali che influenzano la partecipazione sociale. Tuttavia, i limiti campionari impediscono di ottenere stime affidabili a livello sub-regionale, in particolare nelle aree più piccole con scarsa numerosità campionaria.

Il metodo di stima adottato utilizza dati aggregati a livello provinciale, consentendo di produrre stime affidabili nel pieno rispetto delle normative vigenti sulla protezione dei dati personali e sul segreto statistico, come previsto dall'art. 9 del d.lgs. n. 322/1989. La normativa italiana e europea limita inoltre l'utilizzo dei dati relativi a categorie particolari di dati personali, tra cui informazioni sulla salute, che possono essere trattati solo se previsto da specifiche disposizioni di legge o regolamento, con misure adeguate a tutelare i diritti degli interessati (art. 2-sexies, d.lgs. n. 196/2003; art. 9 del Regolamento UE 2016/679).

Per rispettare questi vincoli, le informazioni raccolte con l'indagine EHIS sono utilizzate esclusivamente in forma aggregata. L'integrazione di dati aggregati provenienti dal Sistema Integrato dei Registri (SIR) consente di applicare metodi di stima indiretta (Small Area Estimation, SAE), in linea con le raccomandazioni europee (Regolamento IESS n. 2019/1700, art. 9), producendo stime più dettagliate e precise senza collegare dati personali a livello elementare.

Questo approccio permette di ottenere valori interpretabili e confrontabili, evitando il trattamento di dati sensibili e garantendo che la condizione di disabilità, in particolare, non sia direttamente identificabile a livello individuale. La sperimentazione dei metodi SAE rappresenta un'opportunità per migliorare la qualità delle stime dell'indicatore GLF dell'EHIS, pur rimanendo pienamente entro i vincoli normativi che regolano l'accesso e l'uso dei dati personali.

L'indagine EHIS (European Health Interview Survey), condotta in modo armonizzato in tutti i Paesi dell'Unione Europea, raccoglie un'ampia gamma di indicatori comparabili a livello internazionale, relativi allo stato di salute, all'utilizzo dei servizi sanitari, alla prevenzione e ai principali determinanti della salute. Tra questi figurano le *limitazioni funzionali* — relative a vista, udito, mobilità e capacità cognitive — che costituiscono l'obiettivo di questa prima sperimentazione, nonché la *riduzione dell'autonomia nelle attività della vita quotidiana (ADL e IADL)*, per la quale si stanno sperimentando metodologie di stima SAE che potrebbero essere integrate successivamente in questa statistica sperimentale. Tali indicatori, integrati con tecniche di stima capaci di produrre risultati attendibili anche a livello provinciale, forniscono un contributo rilevante al rafforzamento dell'offerta informativa territoriale. In particolare, consentono di individuare con maggiore precisione le aree che presentano maggiori criticità, favorendo una più efficace allocazione delle risorse e un miglior supporto alle politiche sociali, basato su evidenze riferite a contesti territoriali più disaggregati rispetto alle tradizionali analisi regionali o nazionali.

In questo quadro, le informazioni derivate dall'indagine EHIS potrebbero contribuire ciclicamente ad ampliare significativamente anche l'offerta informativa sul benessere dei territori (BesT), il sistema sviluppato da Istat per monitorare il benessere equo e sostenibile a livello subregionale (Istat, 2024).

L'indicatore del bisogno di assistenza non soddisfatto tra gli anziani fragili potrebbe rappresentare un'ulteriore area di interesse, anche per il monitoraggio di aspetti legati alla nuova normativa sull'assistenza agli anziani (d.lgs. n. 29 del 15 marzo 2024). In sintesi, questa sperimentazione che si concentra al momento sull'indicatore *Gravi Limitazioni Funzionali*, mira a fornire un esempio su come sia possibile ampliare il potenziale informativo delle indagini sulla salute, fornendo dati di qualità a livello territoriale più dettagliato, utili a meglio supportare le politiche di programmazione sanitaria e sociale in un contesto demografico in evoluzione e al fine di calibrare tali interventi sulla base di informazioni statistiche efficienti e granulari.

Il risultato finale di questa sperimentazione consente di alimentare il Sistema Informativo "Disabilità in Cifre" dell'Istat con le stime indirette dell'indicatore *Gravi Limitazioni Funzionali* a livello provinciale, per genere e per macro classi di età, al di fine renderle disponibili a decisori, ricercatori ed utenti interessati. Le stime presentate evidenziano un'eterogeneità infra-regionale, offrendo un approfondimento del quadro informativo oltre il livello medio regionale.

### 1.1. Disegno campionario dell'Indagine EHIS

Il disegno di campionamento presenta una struttura generale tipicamente utilizzata nella maggior parte delle indagini sociali sulle famiglie condotte dall'Istituto Nazionale di Statistica. Questo disegno si basa infatti su uno schema a due stadi comuni-famiglie, con una stratificazione dei comuni in base alla loro ampiezza demografica. Il disegno dell'indagine del 2019 è stato inoltre integrato con quello dell'indagine del Censimento Permanente Popolazione e Abitazioni. I comuni selezionati costituiscono un sottocampione dei 2850 comuni presenti nel campione da lista selezionato per la tornata del Censimento permanente relativa al 2018. Le unità di secondo stadio sono le famiglie, estratte con criterio di scelta casuale dalle liste anagrafiche per i comuni campione con meno di 1000 abitanti e dalla lista delle famiglie selezionate per il Censimento Permanente del 2018 per i comuni campione con 1000 abitanti e oltre. Il campione finale è composto da circa 22.800 famiglie, che risiedono in 835 comuni di diverse dimensioni e sono distribuite su tutto il territorio nazionale.

I domini pianificati comprendono le cinque macro-aree geografiche italiane (Nord-Ovest, Nord-Est, Centro, Sud e Isole) secondo la classificazione NUTS 1, le regioni italiane (NUTS 2) e le due province autonome di Bolzano e Trento. Accanto a questi domini amministrativi sono state incluse anche le cosiddette "aree vaste", aggregazioni territoriali sub-regionali costituite da più comuni contigui che condividono caratteristiche economiche, sociali o di servizi. Queste aree, spesso corrispondenti ad aggregazioni di

ASL, presentano una maggiore omogeneità funzionale e risultano particolarmente utili per finalità di pianificazione e monitoraggio in ambito socio-sanitario. Questo concetto, infatti, è spesso usato in ambito sanitario e socioeconomico per ottimizzare la gestione dei servizi pubblici e per scopi di monitoraggio statistico e di programmazione (ISTAT, 2018, Atlante statistico dei comuni, <https://aster.istat.it/>). Il disegno campionario ha quindi previsto la stratificazione per Area vasta all'interno di ciascuna regione, per un totale di 67 aree su tutto il territorio nazionale (ad eccezione di alcune regioni, che nel tempo hanno concentrato tutti i distretti sanitari delle singole ASL in una sola ASL: Marche, Molise e Sardegna) in modo da poter eseguire eventuali confronti subregionali con l'indagine sulla salute del 2012, in cui l'area vasta era un dominio pianificato. Per ulteriori dettagli sulla strategia campionaria adottata, si può fare riferimento alla nota metodologica (Istat, 2021).

Le stime prodotte dall'indagine sono essenzialmente stime di frequenze assolute e relative, riferite alle famiglie e agli individui. Le stime sono ottenute mediante uno stimatore calibrato (Devaud D., Tillé Y., 2019; Deville J. C., Särndal C. E., 1992) con vincoli su:

- distribuzione della popolazione nelle 21 regioni italiane (19 regioni più le provincie autonome di Trento e Bolzano) per sesso e sette classi d'età (0-14, 15-24, 25-44, 45-54, 55-64, 65-74, 75+);
- distribuzione della popolazione nelle 5 ripartizioni territoriali per sesso e nove classi d'età (0-14, 15-24, 25-34, 35-44, 45-54, 55-64, 65-74, 75-84, 85+);
- distribuzione della popolazione per cittadinanza (famiglia di cittadini italiani, famiglia di cittadini stranieri, famiglia mista).

## 1.2. Alcuni risultati diffusi

I risultati a livello regionale dell'indicatore selezionato evidenziano una geografia piuttosto articolata. Le prevalenze delle *Gravi Limitazioni Funzionali* (sensoriali, motorie o cognitive) tra le persone di 15-64 anni sono relativamente contenute, pari a **3,5%**, corrispondente a circa **1,345 milioni** di individui, mentre tra gli over 65 gli individui con tali gravi difficoltà nelle funzioni di base sono circa **3,86 milioni** (28,4% della popolazione over 65). Rilevanti sono anche le differenze regionali: le incidenze più elevate si riscontrano in Calabria, Sardegna e Sicilia (35,6%, 35,2% e 34,6%), a fronte del valore minimo del Friuli Venezia Giulia (21,7%).

Tra i circa 3,9 milioni di individui over 65 con *Gravi Limitazioni Funzionali*:

- 2,833 milioni (20,9%) hanno gravi difficoltà nel camminare, salire o scendere le scale senza aiuto di una persona o ausili;
- 1,874 milioni (13,8%) riferiscono gravi difficoltà nell'udito o nella vista, anche con ausili;

- 1,113 milioni (8,2%) presentano gravi difficoltà nella memoria o nella concentrazione.

## 2. Metodologia utilizzata

Le stime a livello provinciale per sesso e per le due macro-classi di età (15-64 e 65 anni e oltre) dell'indicatore relativo alle *Gravi Limitazioni Funzionali* sono state ottenute mediante l'applicazione di una metodologia consolidata di stima per piccole aree basata su un modello ad effetti misti definito a livello di area (Fay-Herriot, 1979). Questo metodo permette di combinare le informazioni provenienti da indagini campionarie con dati ausiliari disponibili a livello aggregato, migliorando significativamente la precisione delle stime laddove le sole informazioni dirette non siano sufficientemente affidabili.

Supponiamo di voler stimare un indicatore per  $D$  sottopopolazioni o domini. Per ciascun dominio  $d$  ( $d = 1, 2, \dots, D$ ), disponiamo di una stima diretta dell'indicatore basata sull'indagine campionaria, indicata con  $\hat{\theta}_d$  ed un vettore di variabili ausiliarie  $X_d$  disponibile da fonti esterne (es. dati amministrativi o censuari). Il modello di Fay-Herriot si basa su due componenti:

1. Un **modello di campionamento**, che descrive la stima diretta come funzione del vero valore dell'indicatore e di un errore casuale di campionamento:

$$\hat{\theta}_d = \theta_d + e_d$$

dove  $e_d$  rappresenta l'errore di campionamento, con media zero e varianza  $\sigma_e^2$  nota o stimata.

2. Un **modello di regressione a livello di area**, che esprime il vero valore dell'indicatore come funzione lineare delle variabili ausiliarie, più un termine casuale di area:

$$\theta_d = X_d \beta + u_d$$

dove  $\beta$  è il vettore dei coefficienti da stimare e  $u_d$  è l'effetto casuale dell'area  $d$  con media zero, varianza  $\sigma_u^2$  ed indipendente dall'errore di campionamento.

Combinando i due modelli, si ottiene il modello complessivo a effetti misti:

$$\hat{\theta}_d = X_d \beta + u_d + e_d. \quad (1)$$

Lo **stimatore SAE** risultante è una combinazione pesata tra la stima diretta e quella sintetica,

$$\hat{\theta}_d^{sae} = \hat{\gamma}_d \hat{\theta}_d + (1 - \hat{\gamma}_d) \mathbf{X}_d^T \hat{\beta}, \quad (2)$$

in cui  $\hat{\beta}$  è la stima di massima verosimiglianza di  $\beta$  ottenuta sotto l'ipotesi di normalità per la distribuzione dell'errore di campionamento e dell'effetto casuale di area e  $\hat{\gamma}_d$  è il peso attribuito allo stimatore diretto, dato da:

$$\hat{\gamma}_d = \hat{\sigma}_u^2 / (\hat{\sigma}_e^2 + \hat{\sigma}_u^2), \quad (3)$$

dove:

- $\hat{\sigma}_e^2$  è la varianza stimata delle stime dirette;
- $\hat{\sigma}_u^2$  è la varianza stimata con il metodo della massima verosimiglianza dell'effetto casuale di area.

L'equazione (2) rappresenta la versione empirica del miglior predittore lineare non distorto di  $\theta_d$  – EBLUP – sotto il modello (1), ottenuta sostituendo le stime di massima verosimiglianza  $\hat{\beta}$  e  $\hat{\sigma}_u^2$  nella formula del miglior predittore lineare non distorto di  $\theta_d$ . Quest'ultimo è derivato minimizzando l'errore quadratico medio di predizione (Rao e Molina, 2015, capp. 5 e 6). In sintesi, il modello adottato consente di migliorare in modo significativo la qualità delle stime per piccole aree, grazie ad un uso efficace delle informazioni ausiliarie disponibili nei sottodomini di interesse, integrate con i dati diretti dell'indagine.

Nell'applicazione del metodo sono state selezionate variabili ausiliarie secondo criteri di coerenza tematica con il fenomeno delle gravi limitazioni funzionali, disponibilità a livello provinciale e disaggregazione per sesso e per le due classi di età considerate, qualità e stabilità della fonte, variabilità territoriale e significatività statistica. La selezione è stata condotta combinando valutazioni teoriche, analisi di correlazione, verifiche sulla capacità predittiva delle covariate e procedure stepwise implementate con il pacchetto R embi, al fine di rafforzare la capacità esplicativa del modello e affinare le previsioni nei diversi segmenti della popolazione:

- il numero di persone titolari di una pensione legata alla condizione di disabilità (Invalidità, Indennitarie, Invalidità civile, Indennità di accompagnamento, guerra), rapportato alla popolazione, utilizzato come proxy della presenza territoriale di popolazione con disabilità riconosciuta, essendo all'epoca assente un registro nazionale integrato sulla disabilità;
- l'indice di attrattività ospedaliera, derivato dai dati delle Schede di Dimissione Ospedaliera (SDO) elaborate dal Ministero della Salute, quale indicatore sintetico dell'offerta e della domanda di servizi sanitari complessi. Tale indice misura la capacità di un'area di attrarre pazienti da altre province o regioni, riflettendo i flussi di mobilità

ospedaliera e quindi rappresenta un indicatore di accessibilità e qualità dell'offerta sanitaria.

Oltre alle variabili di contesto a livello di sotto-dominio di interesse, l'integrazione tra dati demografici e informazioni reddituali provenienti da fonti amministrative ha permesso di arricchire la caratterizzazione della popolazione di riferimento, fornendo ulteriori elementi utili alla specificazione del modello. In particolare, sono state considerate covariate riferite ai sotto-domini, derivate dalle seguenti informazioni:

- distribuzione della popolazione per 7 classi di età;
- distribuzione della popolazione per 3 livelli di istruzione (primaria, secondaria, titolo universitario);
- tasso di popolazione a rischio di povertà;
- quintili del reddito equivalente calcolati a livello nazionale, regionale e provinciale;
- distribuzione della popolazione per fonte di reddito (lavoro, pensione, capitale), raggruppata in 5 classi di ampiezza crescente;
- distribuzione della popolazione per 4 classi del numero medio di settimane lavorate, ottenute suddividendo l'anno in trimestri.

L'integrazione tra dati demografici e informazioni reddituali provenienti da fonti amministrative ha permesso di arricchire ulteriormente la caratterizzazione della popolazione di riferimento e di migliorare la specificazione del modello. L'insieme delle variabili selezionate fornisce un quadro coerente delle caratteristiche socio-demografiche e territoriali potenzialmente associate alla distribuzione delle gravi limitazioni funzionali.

### **3. Implementazione del metodo ed analisi dei risultati per l'indicatore GLF**

L'analisi presentata in questo paragrafo descrive il processo di stima dell'indicatore delle *Gravi Limitazioni Funzionali (GLF)* per domini non pianificati, a partire dai dati raccolti nell'ambito dell'ultima edizione disponibile dell'European Health Interview Survey (EHIS), riferita all'anno 2019. L'obiettivo è quello di produrre stime affidabili per domini territoriali e socio-demografici non rappresentati direttamente dall'indagine, con un livello di dettaglio sufficiente a soddisfare le crescenti esigenze informative legate alla pianificazione delle politiche sanitarie e sociali locali.

La disponibilità di stime più disaggregate è fondamentale per analizzare in modo più accurato le disuguaglianze socio-sanitarie tra territori, in particolare nelle aree più vulnerabili; valutare il carico assistenziale e i bisogni della popolazione anziana fragile, in relazione al livello di autonomia funzionale; supportare una programmazione territoriale basata su evidenze empiriche, soprattutto in relazione ai Livelli Essenziali di Assistenza (LEA) e agli interventi previsti dal Piano Nazionale di Ripresa e Resilienza (PNRR).

Tuttavia, le indagini campionarie tradizionali come l'EHIS, pur offrendo una solida base informativa, presentano limiti legati alla ridotta numerosità campionaria, che impedisce la produzione di stime affidabili per domini territoriali di piccola ampiezza, come le province, soprattutto quando si richiedono disaggregazioni per sesso e classi di età. Questo limite si è rivelato particolarmente critico nell'edizione 2019 dell'EHIS, dove l'affidabilità delle stime dirette si mantiene accettabile solo a livello regionale, mentre a livello provinciale l'incertezza aumenta sensibilmente.

Per superare queste criticità è stata sperimentata l'applicazione dei metodi di stima SAE all'indicatore GLF, al fine di produrre stime a livello provinciale, distinte per sesso e per due macro-classi di età (15–64 anni e 65 anni e oltre), per un totale di 428 domini. L'approccio adottato si caratterizza per alcune innovazioni metodologiche: la produzione di stime per domini non pianificati, ovvero in assenza di rappresentatività diretta; l'integrazione di variabili ausiliarie provenienti da fonti amministrative e demografiche (INPS, censimenti, registri); l'impiego di trasformazioni funzionali per migliorare l'aderenza del modello alle ipotesi di normalità; e l'utilizzo del benchmarking per garantire la coerenza delle stime indirette con quelle ufficiali regionali.

Le stime dirette ottenute mostrano un errore accettabile solo a livello regionale, mentre a livello provinciale — soprattutto quando si considera la disaggregazione per sesso e per le due macro-classi di età — l'incertezza diventa troppo elevata. Per questo motivo, l'applicazione dei metodi SAE consente notevoli miglioramenti in termini di efficienza e precisione. Per valutare l'affidabilità delle stime dirette a livelli così disaggregati, è stato adottato il seguente criterio basato sul coefficiente di variazione (CV %):

- stime pubblicabili senza restrizioni:  $CV \% \leq 16,6$ ;
- stime pubblicabili con riserva:  $16,6 < CV \% \leq 33,3$ ;
- stime non raccomandabili:  $CV \% > 33,3$ .

L'analisi delle stime dirette ottenute dall'indagine EHIS 2019 per l'indicatore GLF evidenzia importanti criticità in termini di precisione. In particolare, molte stime presentano coefficienti di variazione elevati, tali da comprometterne l'affidabilità e la possibilità di diffusione ufficiale. La classificazione delle stime dirette in base al CV riportate nella Tabella 1 mostra che:

- solo 83 stime rientrano nella fascia considerata pubblicabile ( $CV \leq 16,5\%$ );
- 154 stime sono pubblicabili con cautela ( $16,5\% < CV \leq 33,3\%$ );
- ben 166 stime risultano non pubblicabili ( $CV > 33,3\%$ );
- per 25 domini, non è stato possibile calcolare alcuna stima diretta per assenza di campione.

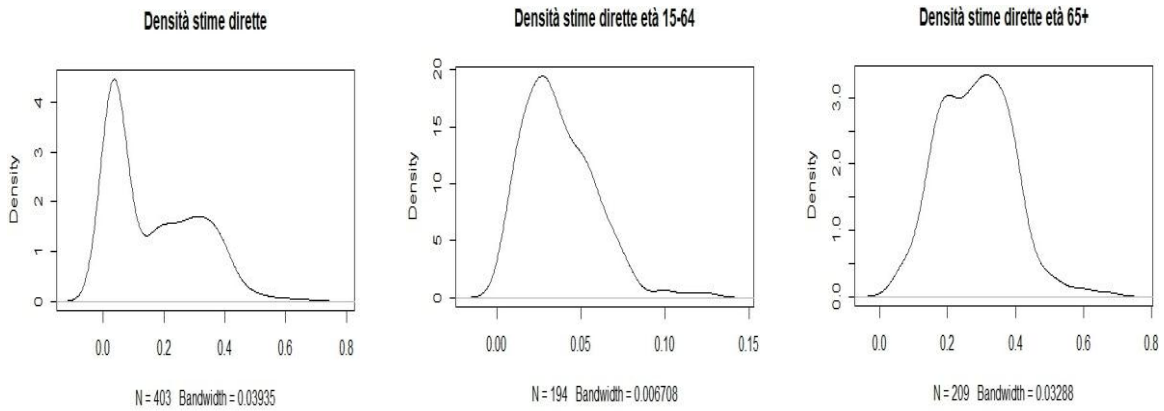
**Tabella 1** Stime dirette per la variabile GLF a livello provinciale Anno 2019, classificate in base al grado di precisione

CV % delle stime	Valutazione	Stime
[0; 16,6]	Pubblicabili	83
(16,6; 33,3]	Pubblicabili con cautele	154
(33,3, ed oltre]	Non pubblicabili	166
Non disponibile		25

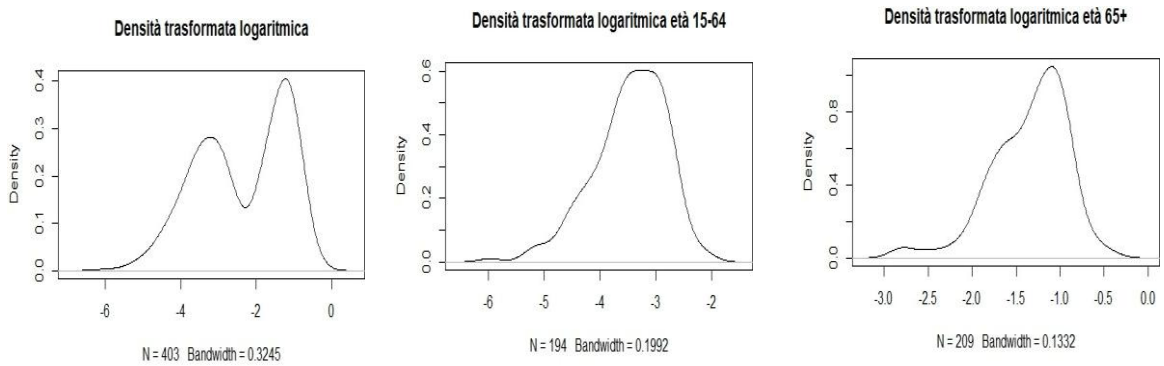
Questi risultati evidenziano l'incapacità dei metodi di stima diretta a fornire stime attendibili ed affidabili nei domini disaggregati, rendendo indispensabile il ricorso a metodi di stima per piccole aree. In questo contesto, l'applicazione dei metodi SAE consente un sostanziale miglioramento della copertura e della qualità delle informazioni disponibili per il monitoraggio della salute nei territori, offrendo un supporto concreto alla costruzione di politiche sanitarie e sociali basate su evidenze. I risultati ottenuti nell'ambito di questa sperimentazione sono potenzialmente replicabili anche per la prossima edizione dell'indagine sulla salute (EHIS 2025), contribuendo così a consolidare l'utilizzo sistematico dei metodi SAE nella produzione ufficiale e nel rafforzamento del sistema informativo nazionale.

Dopo aver delineato le motivazioni e i benefici legati all'uso dei metodi SAE, si passa ora all'analisi statistico-metodologica dei risultati ottenuti. Dall'analisi delle stime (Figura 1), emerge che i valori nella popolazione anziana (65+) sono generalmente più elevati rispetto a quelli relativi alla fascia 15-64 anni, con una parziale sovrapposizione tra le code delle distribuzioni. La distribuzione delle stime appare quindi come una combinazione di due distribuzioni distinte per le due classi di età. La Figura 2 mostra la distribuzione della variabile dopo una trasformazione logaritmica, che, pur mantenendo un'asimmetria legata alle classi di età, risulta complessivamente più adeguata rispetto alla distribuzione su scala originale.

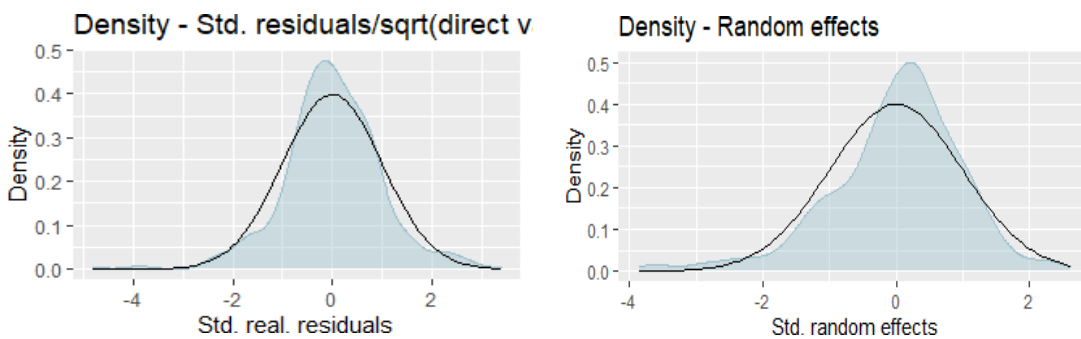
Lo stimatore di Fay-Herriot si basa sulla specificazione del modello ad effetti misti descritto nel paragrafo 2, in cui si assume che gli errori di campionamento e gli effetti casuali siano a media nulla, indipendenti e distribuiti normalmente. Tuttavia, le analisi diagnostiche condotte sui residui standardizzati e sugli effetti casuali mostrano significative deviazioni da tali ipotesi. In particolare, si osservano asimmetrie marcate (skewness = -0,45 per i residui e -1,56 per gli effetti casuali di area), code pesanti (curtosi pari rispettivamente a 5,24 e 8,64) e valori molto bassi dei p-value nei test di Shapiro-Wilk (rispettivamente  $< 1e-7$  e  $< 1e-13$ ), che indicano una chiara violazione dell'assunzione di normalità.



**Figura 1:** densità delle stime dirette nei domini di interesse



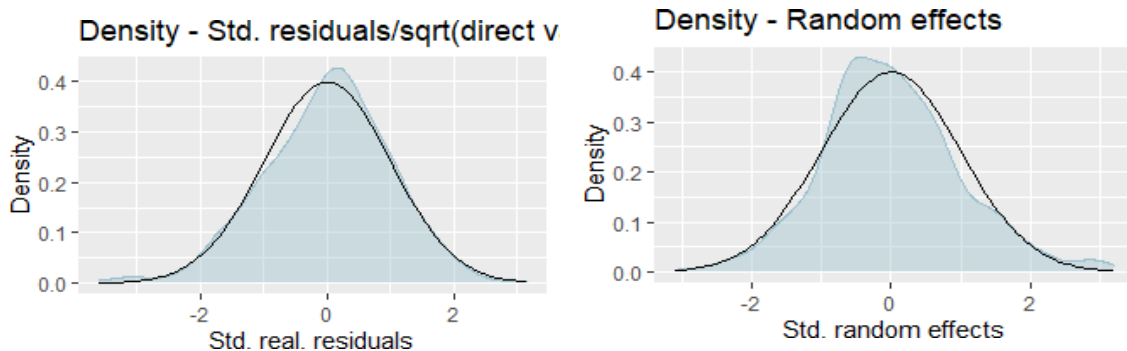
**Figura 2:** densità della trasformata logaritmica delle stime dirette nei domini di interesse



**Figura 3:** Distribuzione dei residui e degli effetti casuali standardizzati del modello standard.

Per migliorare l'aderenza alle ipotesi di normalità e stabilizzare la varianza, è stata applicata una trasformazione logaritmica alle stime dirette. Al fine di attenuare gli effetti distorsivi determinati dalla trasformazione inversa necessaria per tornare alla scala originale, è stato impiegato il metodo di Rao e Molina (2015, cap.6) implementato nel pacchetto emdi del software R utilizzato per le procedure di stima. Tale trasformazione ha contribuito a ottenere una distribuzione più simmetrica degli errori e degli effetti casuali, riducendo l'impatto degli outlier e migliorando la qualità dell'adattamento del

modello. I risultati confermano l'efficacia dell'intervento: sia i residui che gli effetti casuali mostrano valori di skewness prossimi allo zero e curtosi più contenute. Inoltre, i test di normalità di Shapiro-Wilk evidenziano un miglioramento sostanziale: il p-value per i residui standardizzati sale a 0,066 (non significativo), mentre per gli effetti casuali, sebbene il p-value rimanga significativo (0,002), si osserva un progresso molto rilevante rispetto al modello originario. Questi risultati sono illustrati ed apprezzabili anche in Figura 4.



**Figura 4:** Distribuzione dei residui e degli effetti casuali standardizzati del modello con trasformata logaritmica

Anche le misure di bontà dell'adattamento confermano il miglioramento ottenuto con l'applicazione della trasformazione logaritmica. In particolare, l' $R^2$  aggiustato, che tiene conto del numero di covariate incluse nel modello e fornisce una valutazione più robusta della capacità predittiva, mostra un incremento significativo: si passa da 0,847 nel modello non trasformato a 0,920 nel modello trasformato. Questo aumento indica che una quota maggiore della variabilità osservata nelle stime dirette viene spiegata dal modello, a conferma di una migliore coerenza tra i dati e le ipotesi del modello stesso. L'incremento dell' $R^2$  aggiustato riflette non solo un miglioramento nell'aderenza del modello ai dati, ma anche una maggiore efficienza nella stima dei parametri, grazie alla maggiore stabilità della varianza e alla riduzione dell'influenza dei valori anomali.

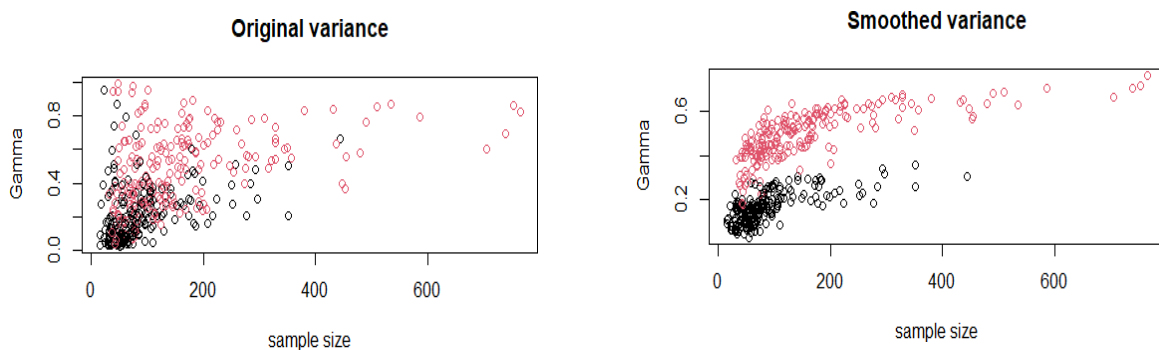
Nei domini, in particolare quando la numerosità campionaria è ridotta, la stima della varianza associata alle stime dirette tende a essere instabile e poco affidabile. Questa instabilità è dovuta al fatto che, con pochi dati disponibili, le varianze campionarie possono risultare casualmente sovrastimate o sottostimate, generando valori fortemente fluttuanti da un'area all'altra. Tale condizione ha effetti significativi nel modello Fay-Herriot, dove i pesi  $\gamma$  (definiti nell'equazione 3) regolano il bilanciamento tra la stima diretta e la componente modellata dello stimatore composito (equazione 2). In presenza di una varianza stimata più bassa, il peso  $\gamma$  tende ad aumentare l'influenza della stima diretta; viceversa, una varianza più elevata attribuisce maggiore peso alla componente da modello. Tuttavia, se le varianze sono stimate in modo instabile, i pesi risultanti possono essere estremi o incoerenti, riducendo la robustezza e la precisione complessiva dello stimatore. Per questo motivo, è fondamentale adottare strategie correttive per

stabilizzare i pesi e migliorare l'affidabilità delle stime prodotte, attraverso un modello di smussamento delle varianze associate alle stime dirette. Dopo una procedura di selezione, il modello impiegato è stato il seguente,

$$\log[\text{VAR}(\hat{\theta}_d)^2] = \beta_0 + \beta_1 \log(n) + \beta_2 \log(\hat{\theta}_d) \quad (4)$$

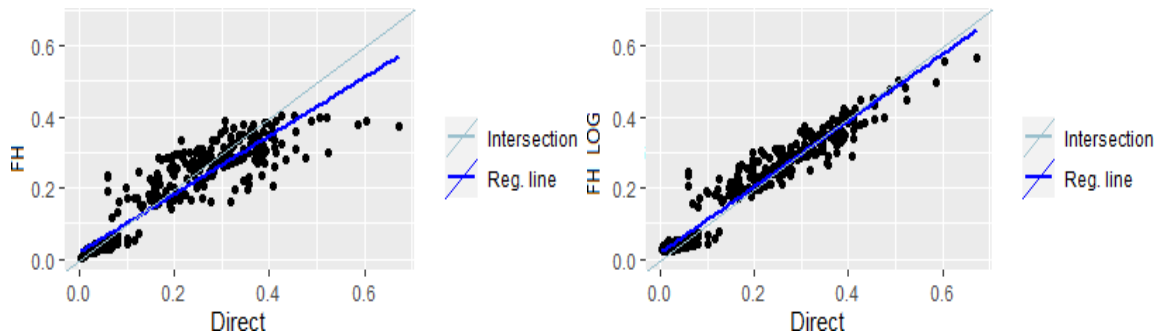
in cui si mette in relazione varianza campionaria con la numerosità campionaria dell'area e all'intensità del fenomeno osservato, permettendo di predire una varianza più stabile e meno influenzata dal rumore campionario. In altre parole, invece di utilizzare direttamente le varianze osservate, spesso troppo instabili, si ottiene una stima "smussata" che riduce la variabilità casuale e riflette meglio la struttura sottostante dei dati.

Come mostrato in Figura 5, l'impiego delle varianze smussate contribuisce a stabilizzare la distribuzione dei pesi  $\gamma$  tra le due macro-classi di età considerate. Questo porta a una composizione dei pesi meno suscettibile a fluttuazioni casuali, migliorando così l'adattabilità del modello e aumentando l'affidabilità e la precisione delle stime prodotte nelle piccole aree.

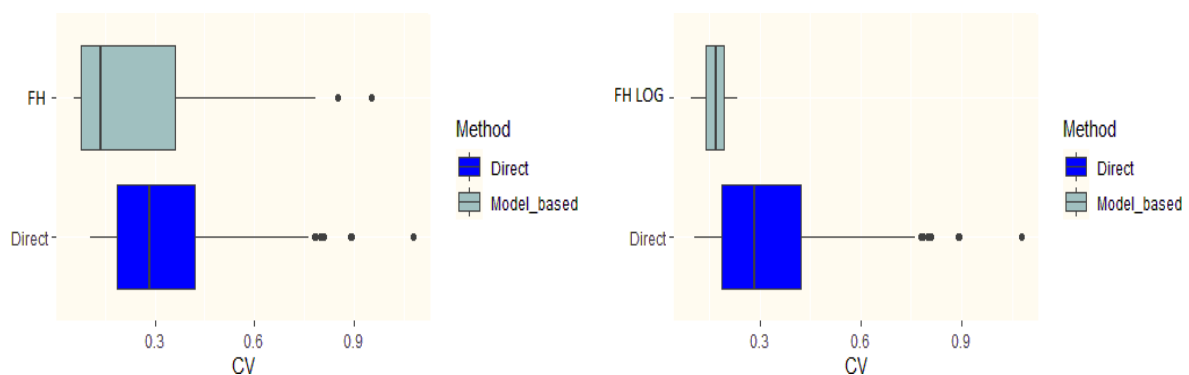


**Figura 5:** Valori dei coefficienti gamma (neri classe 15-64, rossi classe over 65) rispetto alla dimensione campionaria del dominio per il modello FH basato su trasformata logaritmica: varianza originale (sinistra) e varianze smussate (destra).

Il confronto tra il modello Fay-Herriot standard e quello con trasformazione logaritmica (Figura 6) evidenzia che la trasformazione logaritmica migliora la coerenza delle stime, in particolare nelle aree con ridotta numerosità campionaria. Le stime ottenute con il modello log-trasformato, inoltre, permettono ulteriori guadagni di efficienza, rispetto all'applicazione del metodo su scala originaria, garantendo maggiore stabilità e affidabilità delle stime (Figura 7).



**Figura 6:** Andamento delle stime FH e FH LOG rispetto le stime dirette



**Figura 7:** Distribuzione dei coefficienti di variazione delle stime FH e FH LOG rispetto le stime dirette.

Un aspetto essenziale nella produzione di stime per piccole aree è la garanzia di coerenza tra i risultati ottenuti a diversi livelli di disaggregazione territoriale. Per raggiungere questo obiettivo, le stime indirette prodotte tramite modello sono state sottoposte a un processo di benchmarking, ovvero un riallineamento rispetto alle corrispondenti stime dirette già pubblicate a livello regionale. Questo intervento non solo assicura l'allineamento con le stime ufficiali diffuse dall'istituto nei domini pianificati, ma contribuisce anche a rafforzare la robustezza delle stime prodotte con metodi SAE nei domini non pianificati. Inoltre, un ulteriore elemento da considerare è l'impatto del benchmarking sulla precisione delle stime. La Tabella 2 riporta la distribuzione dei coefficienti di variazione per tre stimatori — diretto, SAE, e SAE con benchmarking — suddivisi nelle usuali tre classi di precisione:

- $CV \leq 16,6\%$  (buona precisione),
- $16,6\% < CV \leq 33,3\%$  (accettabile),
- $CV > 33,3\%$  (bassa precisione).

La tabella evidenzia chiaramente l'efficacia del modello SAE, in particolare di quello basato su una trasformazione logaritmica del parametro di interesse. Infatti, tutte le stime calcolate sulla base di questo modello (sia SAE che SAE con benchmarking) risultano entro

la soglia di accettabilità del CV (< 33,3%), a differenza delle stime dirette che presentano 166 casi con CV superiori a tale soglia.

**Tabella 2:** Distribuzione dei CV% delle stime classificati in base a tre livelli di precisione

Stimatore	0-16,6%	16,6%-33,3%	>33,3%
Diretto	83	154	166
SAE	197	231	0
SAE_BENCH	152	276	0

Questo risultato conferma che la trasformazione logaritmica migliora significativamente la precisione nelle stime per piccole aree, rendendo la distribuzione dei CV più concentrata nelle classi di maggiore affidabilità. È importante sottolineare che il processo di messa a coerenza ha introdotto variazioni molto contenute tra le stime iniziali (SAE) e quelle finali (SAE\_BENCH), a riprova della capacità del modello adottato di fornire stime robuste ed efficienti, mantenendo una buona coerenza con le stime dirette.

Il lieve incremento dei CV osservato dopo il benchmarking è del tutto coerente: deriva dall'aggiunta, nel calcolo dell'MSE, di una componente di extra-variabilità introdotta per tenere conto dell'incertezza associata all'allineamento delle stime SAE a quelle dirette. Questo rappresenta un compromesso metodologicamente accettabile, finalizzato a garantire la coerenza tra vari domini di stima senza compromettere eccessivamente la precisione delle stime.

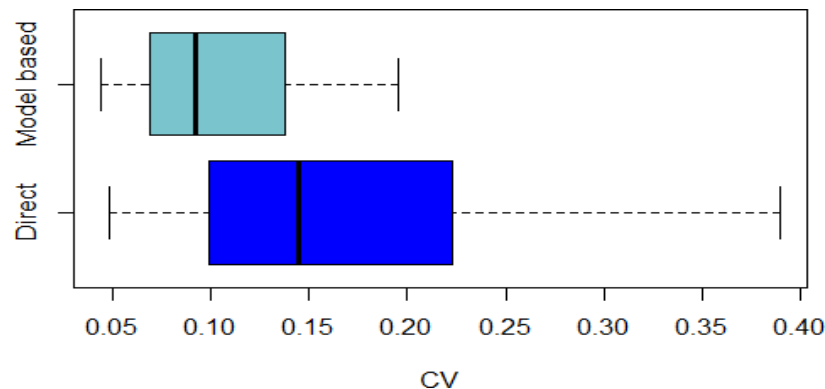
In sintesi, il processo metodologico si è articolato come segue:

- **Modello Fay-Herriot:** è stato adottato un modello lineare a effetti misti, definito su una specifica trasformazione del parametro di interesse, al fine di migliorare il rispetto delle ipotesi di normalità dei residui e di stabilizzare la varianza. Questo accorgimento ha consentito di aumentare la robustezza e l'affidabilità delle stime prodotte.
- **Smussamento della varianza:** per affrontare l'instabilità delle varianze campionarie, particolarmente marcata nelle piccole aree con bassa numerosità campionaria, è stato implementato un modello di smussamento delle varianze associate alle stime dirette. Tale modello lega la varianza stimata alla numerosità e all'intensità del fenomeno osservato, permettendo di ottenere stime di varianza più stabili ed a definire pesi del dello stimatore SAE più coerenti, migliorandone l'adattamento e la precisione.
- **Integrazione ed utilizzo di fonti ausiliarie:** sono state incorporate variabili ausiliarie disponibili a livello provinciale e classificate per sesso e le due classi di età, selezionate per la loro significatività esplicativa nei confronti del fenomeno analizzato. Tra queste figurano indicatori demografici, socio-economici e sanitari provenienti da fonti ufficiali.
- **Benchmarking delle stime:** per garantire coerenza e compatibilità tra i diversi livelli territoriali, le stime indirette SAE sono state sottoposte a un processo di calibrazione

(benchmarking) rispetto alle corrispondenti stime dirette disponibili a livello regionale. Questo passaggio ha assicurato la conservazione dell'additività tra le stime provinciali e regionali, favorendo un'integrazione armonica dei risultati con il sistema informativo ufficiale.

Per completare il quadro informativo, è possibile ottenere stime marginali aggregando quelle prodotte per domini più dettagliati definiti dall'incrocio tra provincia, sesso e classe di età. Questo procedimento permette di fornire stime affidabili anche per domini non pianificati, ottenuti tramite aggregazione lungo una o più dimensioni (ad esempio per sola provincia, solo sesso o sola classe di età), garantendo la coerenza con il modello di stima adottato.

Un caso particolarmente rilevante riguarda i domini REG × Sesso × Classe di età (15-64; 65+), per i quali le stime dirette mostrano coefficienti di variazione (CV) relativamente elevati (CV per la classe di età 15-64 fra 11% e 39%; CV per la classe di età 65 ed oltre fra 5% e 26%). Per migliorare l'affidabilità anche a questo livello marginale, è possibile aggregare le stime provinciali per piccole aree finali, ottenendo così stime regionali distinte per sesso e macro-classi di età, più efficienti e coerenti con il metodo e il modello di stima utilizzati per le corrispondenti stime provinciali (Figura 8).



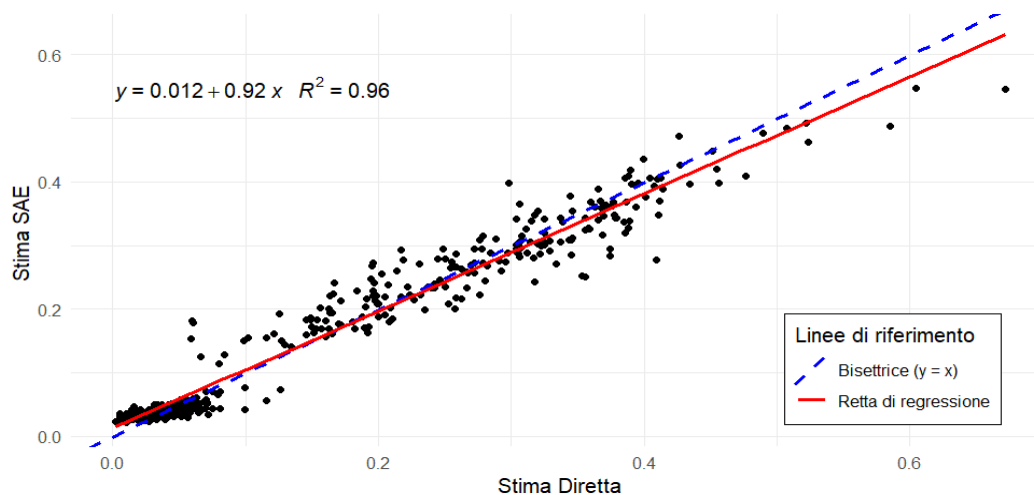
**Figura 8:** distribuzione del CV% per le stime SAE – Model based – e dirette per i domini ottenuti come REGIONE x SESSO x macro-classe d'età.

Questo approccio consente di mantenere la coerenza tra le stime SAE a diversi livelli di disaggregazione, evitando l'uso di modelli diversi. Le stime finali sono state ottenute come medie ponderate delle stime provinciali per ciascuna regione, ed il calcolo dell'errore quadratico medio (MSE) ha tenuto conto della covarianza tra le stime provinciali all'interno della stessa regione. Come mostrato nella Figura 8 le stime per i domini marginali così ottenute risultano caratterizzate da CV compresi tra il 4% e il 20%, evidenziando un netto miglioramento rispetto alle stime dirette.

#### 4. Descrizione dei risultati, valutazione e validazione delle stime SAE

Di seguito si presentano i risultati delle analisi condotte per valutare e validare le stime finali prodotte per piccole aree. Per verificare l'eventuale presenza di distorsioni riconducibili alle ipotesi del modello, le stime SAE sono state confrontate con le corrispondenti stime dirette, che — pur affette da elevata variabilità — rappresentano una stima corretta in base al disegno di campionamento.

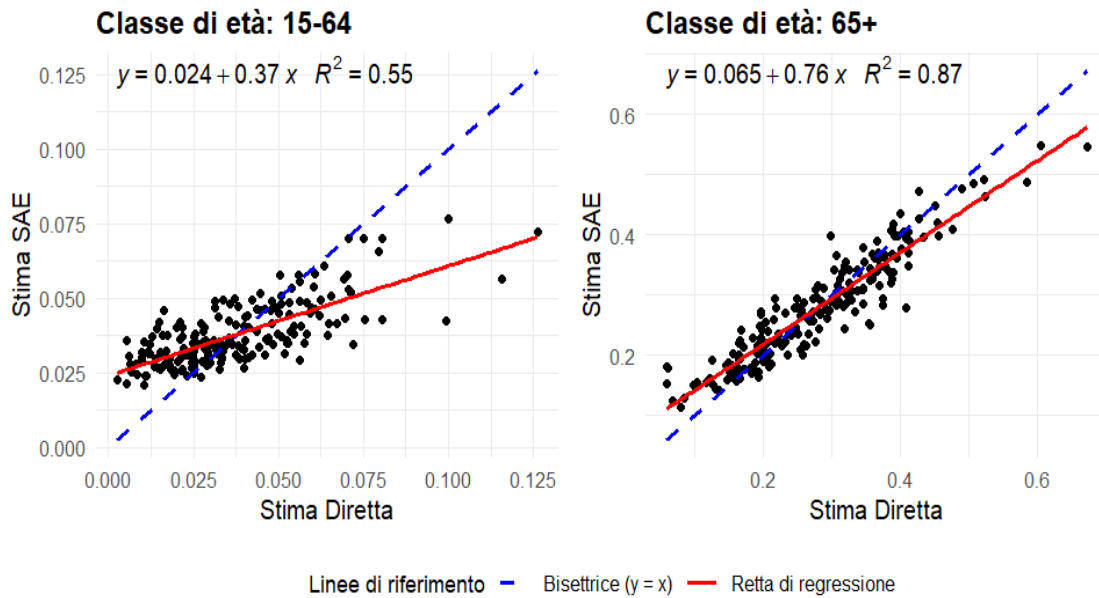
L'analisi di regressione lineare tra le stime SAE e quelle dirette evidenzia una relazione forte e significativa: il coefficiente angolare è pari a 0,922, l'intercetta è prossima allo zero e il coefficiente di determinazione  $R^2$  è pari a 0,956, a conferma di un'elevata coerenza dei risultati ottenuti con i due approcci di stima. Come mostrato in Figura 9, si distinguono chiaramente i due sottogruppi relativi alle due macro-classi di età considerate (15–64 e 65+). Le stime SAE tendono a concentrarsi attorno alla bisettrice, con una leggera compressione dei valori estremi, caratteristica tipica dell'effetto di shrinkage introdotto dal modello. Questo comportamento è coerente con quanto atteso: le stime per piccole aree possono presentare una lieve attenuazione nei valori molto alti o molto bassi rispetto alle stime dirette, senza, tuttavia, evidenza di bias sistematico.



**Figura 9** - Confronto tra stime SAE e stime dirette.

Analizzando separatamente le due classi di età (figura 10), si osserva che nella fascia 15-64 anni la relazione tra le stime SAE e quelle dirette è meno forte: il coefficiente angolare è pari a 0,367 e l' $R^2$  scende a 0,55, indicando una maggiore variabilità rispetto al modello complessivo. Al contrario, nella fascia 65 anni e più, la coerenza tra le due stime risulta decisamente più elevata, con un coefficiente angolare di 0,763 e un  $R^2$  pari a 0,872. Questa differenza di performance è in parte spiegabile con il diverso grado di precisione delle stime dirette nelle due classi di età. Nella fascia 15-64 anni, infatti, le stime dirette tendono a essere più instabili non solo a causa della minore numerosità campionaria, ma anche perché in questa sottopopolazione il fenomeno analizzato è assai meno frequente

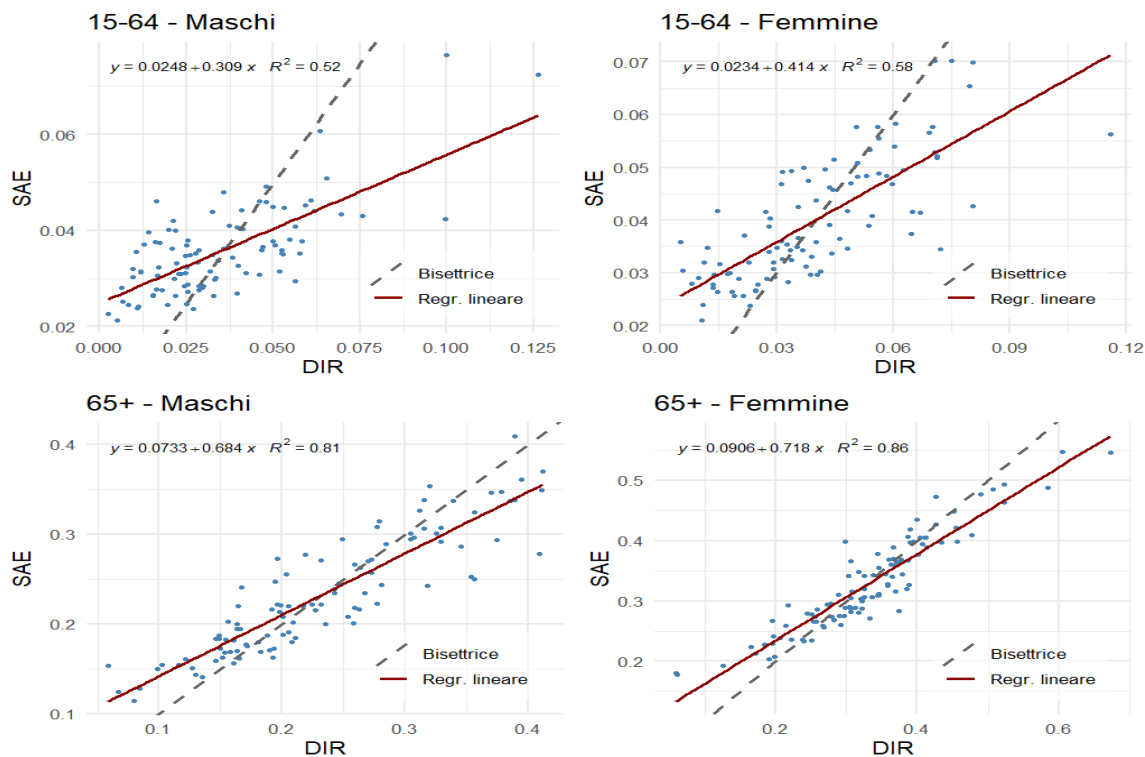
rispetto alla sotto-popolazione 65+. La combinazione di questi due fattori – dimensione campionaria ridotta e bassa prevalenza del fenomeno – accresce l'incertezza statistica, rendendo più complesso il confronto con le stime basate su modello. Al contrario, nella popolazione anziana, dove il fenomeno è più diffuso e il segnale statistico più forte, le stime risultano generalmente più stabili e la corrispondenza tra i due approcci di stima è più elevata.



**Figura 10:** Confronto tra stime SAE e stime dirette nelle due macro-classi di età.

L'analisi, ulteriormente disaggregata per sesso, è riportata nella figura 11. Quando i modelli vengono separati per sesso, per i maschi nella fascia 15-64 anni il coefficiente di regressione è 0,309 con un  $R^2$  di 0,516, mentre nella fascia 65+ il coefficiente sale a 0,684 con un  $R^2$  più alto, pari a 0,806. Le femmine mostrano una tendenza simile: per la fascia 15-64 anni il coefficiente è 0,414 con un  $R^2$  di 0,579, mentre per la fascia 65+ il coefficiente aumenta a 0,718 e l' $R^2$  raggiunge 0,864.

In tutti i confronti, il p-value del coefficiente angolare del modello è estremamente significativo, confermando che la relazione tra le stime SAE e quelle dirette è solida in tutti i sottogruppi. Questo risultato suggerisce che le stime SAE possano essere un buon sostituto delle stime dirette, con lievi differenze a seconda della classe di età e del sesso. Le variazioni nella variabilità ( $R^2$ ) tra i sottogruppi suggeriscono che la qualità delle stime potrebbe essere leggermente diversa, ma il modello generale si conferma molto valido, soprattutto per la fascia 65+, dove l'adattamento del modello risulta particolarmente forte.



**Figura 11** Confronto tra stime SAE e stime dirette nelle due classi di età distinte per sesso.

Quando si applicano metodi di stima per piccole aree è normale osservare differenze tra stime dirette e stime ottenute tramite modello, soprattutto nei domini caratterizzati da elevata instabilità campionaria. Tale instabilità può derivare sia dalla minore numerosità campionaria sia dalla minore diffusione del fenomeno nella popolazione considerata, condizioni che tendono a produrre stime dirette con coefficienti di variazione elevati.

In questi casi è quindi atteso che le stime SAE possano discostarsi maggiormente dalle stime dirette; tale comportamento non rappresenta di per sé un segnale di distorsione del modello, ma una caratteristica intrinseca dei metodi di stima per piccole aree. Se l'analisi si basa esclusivamente sul confronto grafico tra stime dirette e SAE, eventuali segnali di distorsione risultano interpretabili in modo più informativo soprattutto nei domini in cui le stime dirette sono relativamente più precise (cioè con coefficienti di variazione più contenuti). Per questo motivo alla precedente valutazione grafica è stata affiancata una verifica statistica formale, applicando il test di Wald *goodness-of-fit* (Brown et al., 2001). Tale test consente di verificare formalmente l'eventuale presenza di una componente di distorsione sistematica riconducibile al modello utilizzato in quanto consente di valutare se le differenze tra stime dirette e stime SAE siano compatibili con l'incertezza statistica associata alle due stime. In particolare, per ciascuna area è stato calcolato il contributo:

$$GOF = (SAE - DIR)^2 / (Var(DIR) + MSE(SAE))$$

dove  $\text{Var}(\text{DIR})$  rappresenta la varianza della stima diretta e  $\text{MSE}(\text{SAE})$  l'errore quadratico medio della stima SAE. La somma di tali contributi segue approssimativamente una distribuzione  $\chi^2$  con un numero di gradi di libertà pari al numero di aree considerate. Valori elevati della statistica del test indicherebbero la presenza di discrepanze tra le due tipologie di stima non spiegabili dalla loro variabilità statistica e quindi potenzialmente riconducibili a una componente distorsiva del modello.

Applicando il test all'intero insieme delle aree territoriali non emerge alcuna evidenza di distorsione sistematica nelle stime SAE rispetto alle stime dirette. Il valore della statistica di Wald risulta pari a 273,7 con 403 gradi di libertà ( $p\text{-value} \approx 1$ ), indicando che le differenze osservate tra le due tipologie di stima sono pienamente coerenti con il livello di incertezza statistica associato alle stime.

L'analisi è stata quindi replicata separatamente per le due macro-classi di età considerate. Anche in questo caso non emergono evidenze di distorsione sistematica. Per la fascia 15–64 anni la statistica di Wald è pari a 207,9 con 194 gradi di libertà ( $p\text{-value} = 0,235$ ), mentre per la popolazione 65 anni e più la statistica è pari a 65,8 con 209 gradi di libertà ( $p\text{-value} \approx 1$ ).

L'analisi disaggregata per classe di età e sesso conferma tale risultato nella quasi totalità dei sottogruppi. In particolare, per le femmine 15–64 anni la statistica di Wald è pari a 84,1 con 97 gradi di libertà ( $p\text{-value} = 0,822$ ), mentre per la popolazione 65+ si ottengono valori pari a 31,3 ( $df = 105$ ) per le femmine e 34,6 ( $df = 104$ ) per i maschi, con  $p\text{-value}$  prossimi all'unità.

Un'unica deviazione dall'ipotesi di perfetta coerenza emerge per i maschi nella fascia di età 15–64 anni, per i quali la statistica di Wald è pari a 124 con 97 gradi di libertà ( $p\text{-value} = 0,035$ ). Si tratta tuttavia di un risultato borderline, non insolito in applicazioni di metodi di stima per piccole aree, soprattutto in presenza di stime dirette caratterizzate da elevata instabilità campionaria. Un'analisi diagnostica più approfondita dei contributi individuali alla statistica del test evidenzia che tale risultato è quasi interamente attribuibile a una singola area territoriale, la provincia di Pordenone, caratterizzata da una stima diretta particolarmente instabile e da un coefficiente di variazione molto elevato. Escludendo tale osservazione, la statistica del test per il gruppo maschi 15–64 anni si riduce a 113,5 con 96 gradi di libertà ( $p\text{-value} = 0,108$ ), non evidenziando quindi alcuna indicazione di distorsione sistematica del modello anche per questo sottogruppo.

Nel complesso, i risultati del test confermano che non emergono evidenze di distorsione sistematica delle stime SAE rispetto alle stime dirette. Le differenze osservate sono principalmente riconducibili alla maggiore instabilità delle stime dirette in alcuni domini caratterizzati da campioni ridotti e da una minore diffusione del fenomeno, condizioni nelle quali è fisiologico che le stime da modello si discostino maggiormente dalle stime dirette.

Alla luce dell'analisi diagnostica approfondita condotta, e in un'ottica di estrema cautela e correttezza metodologica, si è deciso di segnalare con uno specifico warning la sola stima relativa alla provincia di Pordenone per i maschi nella classe di età 15–64 anni. Tale segnalazione è introdotta a maggiore tutela nell'interpretazione dei risultati e al fine di raccogliere eventuali feedback utili a un ulteriore perfezionamento della metodologia di stima, anche in prospettiva di un eventuale passaggio alla diffusione ufficiale dell'Istituto. Oltre al confronto tra stime dirette e SAE, la validazione delle stime SAE si è anche basata sulla comparazione dei ranghi territoriali derivanti dalle stime dirette e dalle stime SAE. Il confronto dei ranghi consente di valutare non solo i valori degli indicatori, ma anche la loro posizione relativa all'interno della graduatoria nazionale. Le discrepanze nei ranghi rappresentano un indicatore utile ad identificare potenziali inefficienze nelle stime dirette, spesso caratterizzate da elevata variabilità nei piccoli domini a causa di campioni ridotti. Parallelamente, tali differenze possono evidenziare limiti o assunzioni non pienamente soddisfatte nei modelli SAE adottati, fornendo così elementi diagnostici utili per la valutazione e l'ottimizzazione delle metodologie di stima per piccole aree.

L'analisi delle correlazioni di rango (coefficiente di Spearman) tra le stime dirette e quelle ottenute mediante la metodologia SAE per l'indicatore GLF, sintetizzata nella Tabella 3, fornisce importanti indicazioni sulla coerenza e l'allineamento tra i due set di stima nei diversi sottogruppi di popolazione. Nel complesso, la correlazione complessiva tra i ranghi delle due tipologie di stima è pari a 0,806, indicando una buona concordanza. Questo suggerisce che la metodologia SAE è in grado di preservare in maniera coerente la distribuzione territoriale osservata nelle stime dirette.

Analizzando i dati per sesso, emerge una maggiore coerenza tra le femmine ( $p = 0,853$ ) rispetto ai maschi ( $p = 0,758$ ). Tale differenza potrebbe riflettere una maggiore affidabilità delle stime SAE per la popolazione femminile o, al contrario, una maggiore variabilità dei dati maschili che il modello riesce a catturare con minore precisione.

**Tabella 3** Correlazione dei ranghi complessiva e per sottogruppi.

Sesso	Classe di Età	Livello	Correlazione Ranghi (Spearman)
F	15-64	Sottogruppo	0,775089
F	65+	Sottogruppo	0,923917
M	15-64	Sottogruppo	0,587892
M	65+	Sottogruppo	0,906252
F	Tutte	Marginale per sesso	0,853387
M	Tutti	Marginale per sesso	0,757781
Tutti	15-64	Marginale per età	0,681853
Tutti	65+	Marginale per età	0,915176
Tutti	Tutte	Complessiva	0,805626

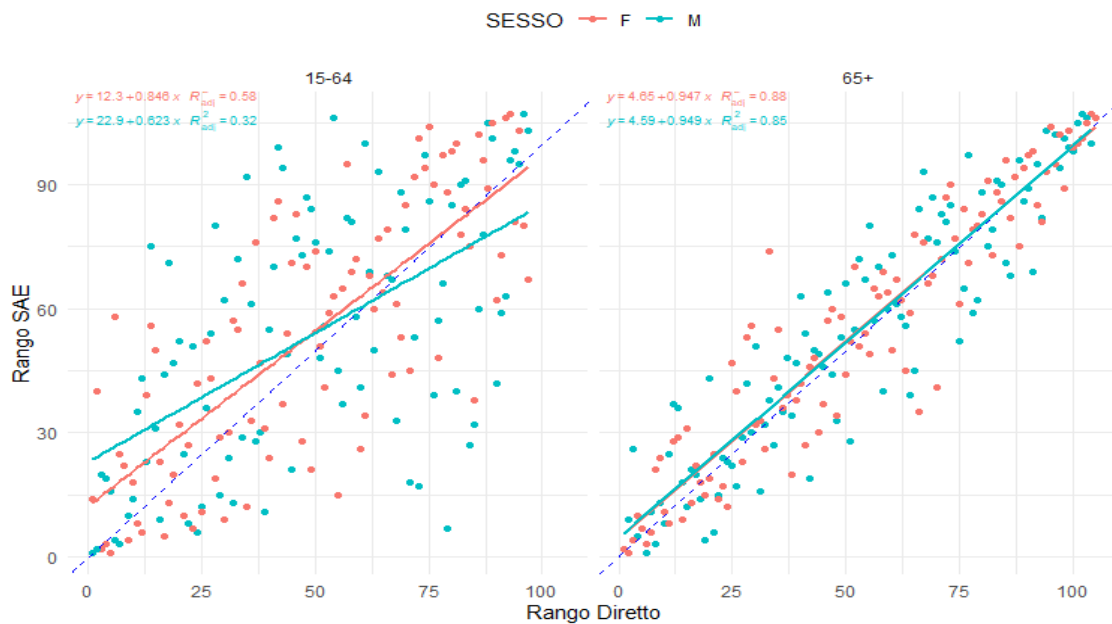
Il quadro più interessante emerge dalla disaggregazione per classe di età. La fascia 65+ mostra una correlazione molto elevata ( $\rho = 0,915$ ), a conferma della robustezza del modello SAE in un segmento di popolazione in cui l'indicatore GLF è più diffuso e le stime dirette risultano relativamente più affidabili. Al contrario, nella fascia 15–64 anni la correlazione scende a 0,682, indicando una minore sovrapposizione tra le classifiche prodotte dai due stimatori. Questa discrepanza può essere in parte attribuita a una maggiore instabilità e imprecisione delle stime dirette, dovuta alla più bassa prevalenza del fenomeno nelle età più giovani.

L'analisi per sottogruppi combinati di sesso ed età fornisce ulteriori elementi di riflessione. Le correlazioni più elevate si osservano tra gli anziani: 0,924 per le donne e 0,906 per gli uomini, a conferma della solidità delle stime SAE in questa fascia. Al contrario, tra i soggetti di età 15–64, la correlazione è ancora buona per le donne (0,775), mentre per gli uomini scende al 0,588. Quanto osservato emerge chiaramente anche nella figura 12, che mostra come le differenze di rango tra stime dirette e stime SAE siano più marcate nella prima classe di età, con un effetto più pronunciato per gli individui di sesso maschile all'interno della fascia 15-64."

Per valutare se le differenze assolute di rango tra stime dirette e SAE dipendono dalla precisione delle stime, misurata tramite il coefficiente di variazione (CV), sono stati applicati un test non parametrico e un modello di regressione. Il test non parametrico di Kruskal-Wallis permette di confrontare la distribuzione delle differenze di rango tra tre classi di CV, definite dividendo il CV negli intervalli  $[0, 0,16]$ ,  $(0,16, 0,33]$ ,  $(\text{oltre } 0,33)$  e denominate rispettivamente "Basso", "Medio" e "Alto", verificando se almeno una classe differisce significativamente dalle altre. Parallelamente, il modello di regressione lineare con le differenze di rango come variabile risposta e predittori quali, la varianza stimata delle stime dirette, la dimensione del campione e la popolazione di riferimento permette di quantificare e isolare l'effetto di ciascun fattore sul cambiamento di rango. L'analisi dei risultati ci suggerisce che le differenze di rango tra stime dirette e stime SAE variano in modo significativo tra le diverse classi di CV, il test di Kruskal-Wallis ha un p-value molto basso ( $4.019e-07$ ), indicando che almeno una delle classi di CV ha differenze di rango significativamente diverse dalle altre. Il modello di regressione invece approfondisce questa relazione considerando anche altre variabili come la varianza campionaria delle stime dirette, il numero di unità campionarie, il valore della stima diretta e la popolazione indagata. Dai risultati si osserva che:

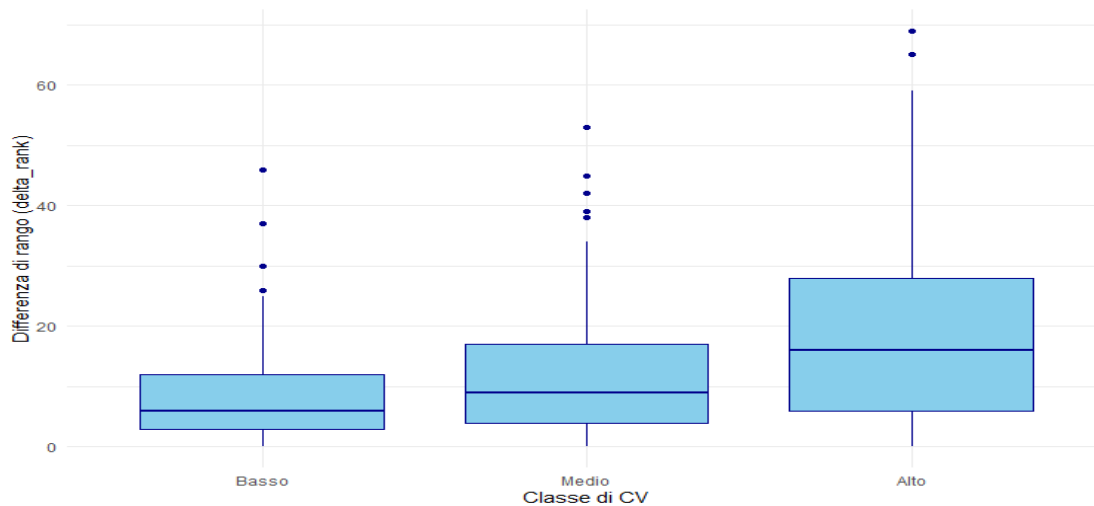
- La varianza campionaria ha un effetto positivo e significativo sulla differenza di rango (coefficiente 530,  $p = 0,0168$ ), confermando che una maggiore incertezza nelle stime dirette è associata a spostamenti più ampi nei ranghi.
- Il valore della stima diretta ha un effetto negativo e altamente significativo (coefficiente -36,25,  $p < 0,001$ ), suggerendo che stime dirette di proporzioni più elevate tendono a subire variazioni di rango inferiori rispetto a quelle più basse.
- Il numero di unità campionarie e la popolazione indagata non risultano significativamente associati con la differenza di rango.

In sintesi, sia il test non parametrico sia il modello di regressione indicano che la precisione delle stime dirette (rappresentata dal CV e/o dalla varianza) è un fattore chiave nel determinare quanto la posizione relativa dei domini cambia passando dalle stime dirette alle stime SAE. Stime dirette meno precise sono più soggette a variazioni di rango, mentre il valore della stima diretta gioca un ruolo inverso nel modificare queste differenze.



**Figura 12** Confronto Rango stime dirette vs stime SAE per classe di Età e Sesso.

Nella figura 13 è illustrata la variazione delle posizioni nella graduatoria tra le stime dirette e quelle SAE, suddivisa per le tre classi di coefficiente di variazione (CV). Si osserva chiaramente che i cambiamenti più marcati nei ranghi si verificano per le stime dirette con CV più elevati, ovvero quelle caratterizzate da una maggiore instabilità e minore precisione.



**Figura 13** Differenza di rango per classi di CV delle stime dirette.

In sintesi, l'applicazione della metodologia SAE consente di migliorare l'affidabilità delle stime nei piccoli domini mantenendo un'elevata coerenza con le stime dirette e preservando nel complesso la graduatoria territoriale. Le stime risultano particolarmente solide per la popolazione anziana, mentre in alcuni sottogruppi caratterizzati da maggiore instabilità campionaria — come gli uomini sotto i 65 anni — il confronto tra stime dirette e stime da modello può presentare una variabilità relativamente più elevata, senza evidenze di distorsioni sistematiche.

## 5 Conclusioni

La sperimentazione metodologica presentata costituisce un esempio rilevante di come sia possibile incrementare la granularità della produzione di statistica ufficiale per indicatori del benessere equo e sostenibile indicando una possibile evoluzione della statistica ufficiale, in questo caso nella dimensione sanitaria. L'applicazione dei metodi SAE all'indicatore delle Gravi Limitazioni Funzionali ha consentito di colmare un importante gap informativo, offrendo stime territoriali affidabili a livello provinciale e con disaggregazioni per sesso e classe di età.

In un contesto demografico caratterizzato da profondi cambiamenti e da un invecchiamento strutturale la disponibilità di dati attendibili sulla salute e sull'autonomia della popolazione anziana rappresenta un requisito fondamentale per orientare le politiche sanitarie e assistenziali. Le sole stime dirette da indagine, a causa dei limiti campionari nei piccoli domini, non sarebbero in grado di soddisfare questa esigenza crescente. L'esperienza qui condotta dimostra, invece, che l'adozione di tecniche di stima indiretta e l'integrazione con fonti ausiliarie — soprattutto di natura amministrativa — possono rafforzare e valorizzare il patrimonio informativo già raccolto, migliorando al tempo stesso la copertura territoriale e la qualità delle stime. Tutto questo nel pieno rispetto della normativa sulla privacy, grazie all'utilizzo esclusivo di dati aggregati.

L'innovazione introdotta è duplice. Sul piano del prodotto, vengono rese disponibili nuove evidenze statistiche di dettaglio, capaci di restituire differenze infra-regionali cruciali per la programmazione dei servizi. Sul piano del processo, si afferma l'uso sistematico di metodologie avanzate di stima, che aumentano la capacità adattiva del sistema statistico rispetto ai bisogni informativi emergenti.

Dal punto di vista istituzionale, la sperimentazione rafforza il ruolo dell'Istat nel rispondere a nuove sfide conoscitive e apre la strada a futuri sviluppi: i metodi SAE potranno essere estesi non solo ad altri indicatori di salute, come la perdita di autonomia nelle attività quotidiane (ADL/IADL), ma anche ad altri domini del benessere e della società, utilizzando dati provenienti dalle principali indagini Istat, quali Forze lavoro, EUSILC e AVQ, contribuendo così ad arricchire l'offerta informativa del sistema "Benessere dei Territori" (Best).

In prospettiva, l'uso regolare dei metodi SAE potrà costituire un pilastro stabile della produzione statistica ufficiale. Limitandosi agli indicatori sulla salute, tali tecniche potranno supportare in modo concreto la programmazione territoriale, il monitoraggio dei Livelli Essenziali di Assistenza (LEA) e la valutazione degli interventi previsti, ad esempio quelli finanziati dal PNRR. Questa sperimentazione assume, inoltre, un valore strategico in ambito europeo, ponendosi in piena coerenza con il Regolamento 2019/1700 (IESS), che incoraggia il ricorso a metodi innovativi per migliorare la copertura degli indicatori armonizzati a livello dell'Unione. L'esperienza maturata dall'Istat contribuisce così a rafforzare il ruolo dell'Italia nel contesto della statistica sociale europea. Infine, i risultati ottenuti sono potenzialmente replicabili già a partire dalla prossima edizione dell'indagine sulla salute (EHIS 2025), consolidando l'utilizzo dei metodi SAE nella produzione ufficiale e rafforzando in modo duraturo la capacità del sistema informativo nazionale di rispondere a sfide demografiche e sociali sempre più complesse.

## Riferimenti bibliografici

Brown G., Chambers R., Heady P., Heasman D. (2001) Evaluation of small area estimation methods: an application to unemployment estimates from the UK LFS. In: *Proceedings of statistics Canada symposium*. Statistics Canada, vol 2001, pp 1–10

Deville J. C., Särndal C. E. (1992), Calibration estimators in survey sampling, *Journal of the American Statistical Association*, 87(418), 376-382.

Devaud D., Tillé Y. (2019), Deville and Särndal's calibration: revisiting a 25-years-old successful optimization problem, *TEST*, 28(4), 1033-1065.

Fay R. 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(366), 269-277.

Eurostat (2018), European Health Interview Survey: Methodological manual, URL: <https://ec.europa.eu/eurostat/documents/3859598/8762193/KS-02-18-240-EN-N.pdf/5fa53ed4-4367-41c4-b3f5-260ced9ff2f6?t=1521718236000>

Istat (2021), Condizioni di salute e ricorso ai servizi sanitari in Italia e nell'Unione Europea - Indagine EHIS 2019 - Nota metodologica, URL: <https://www.istat.it/ws/fascicoloSidi/553/Nota%20metodologica.pdf>

Istat (2024), Il Benessere equo e sostenibile dei territori. Report regionali – Anno 2024, URL: <https://www.istat.it/comunicato-territoriale/il-benessere-equo-e-sostenibile-dei-territori-report-regionali-anno-2024/>

Ministero della Salute (2024), *Monitoraggio dei Livelli Essenziali di Assistenza (LEA)*, URL: <https://www.salute.gov.it/new/it/tema/livelli-essenziali-di-assistenza/il-sistema-di-monitoraggio/>

Rao, J. N., Molina, I. (2015). *Small area estimation*. John Wiley & Sons.

---

<sup>i</sup> La nota metodologica è stata curata da Michele D'Alò. Il paragrafo 1 è stato redatto da Lidia Gargiulo e Laura Iannucci; i paragrafi 2, 3, 4 e 5 sono stati redatti da Michele D'Alò.