

NOTA METODOLOGICA

Il disagio socio-economico degli individui e delle famiglie a livello sub-comunale: primi risultati

Anno 2021

Introduzione

L'Istat ha condotto, insieme ad alcuni Comuni¹, un progetto per misurare il disagio socio-economico degli individui e delle famiglie a livello sub-comunale (Carbonetti *et al.*, 2025a; Carbonetti *et al.*, 2025b). L'obiettivo è quello di mettere a disposizione un insieme di indicatori e un indice composito del disagio, per domini territoriali sub-comunali, evidenziando le aree in cui si osserva la maggiore diffusione del fenomeno.

Dato il carattere multidimensionale del disagio socio-economico, sono stati dapprima definiti nove indicatori elementari, rappresentativi delle componenti più rilevanti del fenomeno. Successivamente, tali indicatori sono stati sintetizzati nell'Indice composito sul Disagio Socio-Economico (IDISE), calcolato a livello di sezione di censimento e restituito per aree sub-comunali risultanti da specifiche aggregazioni di sezioni. Per la costruzione dell'indice si è utilizzata la metodologia dell'*Adjusted Mazziotta-Pareto Index* (AMPI⁺) (Mazziotta e Pareto, 2016).

Lo studio si basa sull'integrazione dei risultati del Censimento Permanente della Popolazione e delle Abitazioni (CPPA) con le informazioni provenienti da fonti amministrative non anagrafiche e sui dati prodotti con i Registri statistici di base e tematici dell'Istituto. Le varie fonti sono geo-codificate alle sezioni di censimento 2021.

Gli indicatori elementari di disagio socio-economico

Il Disagio socio-economico è definito come *“Condizione in cui gli individui sperimentano difficoltà a soddisfare adeguatamente le loro necessità di base a causa della carenza o insufficienza delle risorse e delle opportunità di tipo sociale, economico, lavorativo ed educativo”*. A partire da tale definizione, sono stati individuati nove indicatori elementari che misurano le componenti più rilevanti del disagio socio-economico.

Per ciascun indicatore elementare, vengono di seguito riportati il metodo di calcolo, le componenti del fenomeno rappresentate e le fonti utilizzate. Tutti gli indicatori sono calcolati con riferimento alla popolazione residente in famiglia mentre sono escluse le persone residenti in convivenza, spesso caratterizzate da peculiarità del tutto differenti e che godono di specifiche tutele e garanzie rispetto al disagio così come sopra definito.

- **DIS1** – *Incidenza percentuale di individui di età pari o superiore a 70 anni che vivono da soli e non possiedono una casa di proprietà* sul totale della popolazione di 70 anni e oltre. L'indicatore rappresenta il disagio sociale delle persone anziane (70 anni e più) che non risiedono in convivenza e sperimentano la solitudine e le possibili difficoltà economiche derivanti dalla mancanza di una casa di proprietà. Fonte: CPPA e Catasto delle unità immobiliari (Agenzia delle Entrate).

¹ I Comuni coinvolti nella fase sperimentale dello studio (cfr. *Nota Informativa - nota 1*) sono: Bari, Bologna, Cagliari, Carpi, Catania, Firenze, Genova, Gorizia, Messina, Milano, Modena, Napoli, Olbia, Padova, Palermo, Parma, Perugia, Prato, Reggio di Calabria, Roma, Taranto, Torino, Trieste, Venezia e Verona.

- **DIS2** – *Incidenza percentuale di individui in famiglie in cui nessun membro è occupato o riceve una pensione da lavoro²* sul totale della popolazione residente in famiglia. L'indicatore rappresenta il disagio socio-economico delle famiglie dovuto alla mancata partecipazione dei suoi componenti, attuale o passata, al mercato del lavoro. L'indicatore assume un significato socio-economico, considerando la partecipazione al mercato del lavoro come un indicatore di inclusione sociale e non solo come fonte di reddito. Fonte: CPPA e Registro Tematico sui Redditi.
- **DIS3** – *Incidenza percentuale di individui in famiglie a basso reddito equivalente³* sul totale della popolazione residente in famiglia. L'indicatore misura il disagio economico dovuto alla carenza di reddito; si fa specifico riferimento agli individui che vivono in famiglie con un livello di reddito familiare disponibile equivalente al di sotto del 60% della mediana della distribuzione individuale del reddito disponibile equivalente a livello nazionale. Fonte: CPPA, Registro Tematico sui Redditi (Modulo Banca Dati Reddittuale Integrata – BDR-I⁴) e Registro Base degli Individui (RBI).
- **DIS4** – *Tasso di occupazione 25-64 anni*, misura l'incidenza percentuale della popolazione occupata che vive in famiglia, di età compresa tra 25 e 64 anni, sul totale della popolazione di età 25-64 anni residente in famiglia. L'indicatore rappresenta l'impiego della popolazione adulta nel mercato del lavoro ed è discorde rispetto al disagio socio-economico, ovvero un valore più alto descrive un minor disagio. Fonte: CPPA.
- **DIS5** – *Incidenza percentuale di individui di età compresa tra 0 e 64 anni che vivono in famiglie con bassa intensità lavorativa⁵*. L'indicatore misura il livello di partecipazione dei diversi componenti della famiglia al mercato del lavoro, nel corso dell'anno. Fonti: RBI e Registro Tematico del Lavoro (occupati regolari ad esclusione dei residenti in Italia che lavorano all'estero).
- **DIS6** – *Incidenza percentuale di individui occupati di età compresa tra 25 e 64 anni che vivono in famiglia e con occupazione "non stabile" durante l'anno* sul totale della popolazione di età 25-64 anni residente in famiglia con un segnale di lavoro nell'anno. L'indicatore rappresenta la componente del disagio, della popolazione adulta, dovuta allo stato di precarietà dell'attività lavorativa. Il lavoro precario, infatti, è spesso associato a bassa remunerazione e può impedire alle famiglie di raggiungere una sicurezza economica, nonché generare incapacità di investimento. Vengono considerati "non stabili" i lavoratori dipendenti con contratto a tempo determinato e i lavoratori non dipendenti collaboratori o lavoratori occasionali. Fonti: RBI e Registro Tematico del Lavoro (occupati regolari ad esclusione dei residenti in Italia che lavorano all'estero).
- **DIS7** – *Incidenza percentuale di individui che vivono in famiglia di età compresa tra 25 e 64 anni con basso livello di istruzione* sul totale della popolazione di età 25-64 anni residente in famiglia. L'indicatore rappresenta il potenziale disagio educativo dovuto a un livello di istruzione non superiore al diploma di scuola secondaria di I grado (licenza media o di avviamento professionale). Fonte: CPPA.

² Si fa presente che in tali famiglie potrebbero essere percepite altre forme di pensioni (es. pensioni di invalidità civile, pensioni e assegni sociali, pensioni ai superstiti per decesso di un familiare lavoratore o in pensione da lavoro) o altri trasferimenti non pensionistici (es. reddito di cittadinanza, indennità di disoccupazione Naspi) o redditi da capitale (es. affitti di immobili, rendite da investimenti finanziari).

³ Il reddito equivalente è calcolato dividendo il valore del reddito netto familiare per un opportuno coefficiente di correzione (scala di equivalenza), che permette di tener conto dell'effetto delle economie di scala e di rendere direttamente confrontabili i livelli di reddito di famiglie diversamente composte. La scala di equivalenza (definita "OCSE modificata" e utilizzata anche a livello europeo) è pari alla somma di più coefficienti individuali (1 per il primo adulto, 0,5 per ogni altro adulto e 0,3 per ogni minore di 14 anni). Tutti i membri della stessa famiglia possiedono lo stesso reddito (individuale) equivalente netto.

⁴ La fonte BDR-I include redditi di fonte amministrativa fiscale, previdenziale e assistenziale. Il reddito disponibile individuale è calcolato come differenza tra il reddito al lordo delle imposte (incluse le componenti non imponibili e i trasferimenti inter-familiari) e le imposte. Il reddito al lordo delle imposte contiene: il reddito da lavoro al netto dei contributi sociali, il reddito da capitale e i trasferimenti monetari pensionistici, assicurativi (es. Cassa Integrazione Guadagni, indennità di disoccupazione Naspi) e assistenziali (es. Reddito di Cittadinanza, Assegni al nucleo familiare). La fonte BDR-I non include i redditi finanziari non presenti nelle fonti fiscali soggetti all'imposizione alla fonte, quelli soggetti a imposizione separata (es. arretrati da lavoro), i redditi irregolari e quelli prodotti all'estero.

⁵ Bassa intensità lavorativa è la condizione in cui le persone vivono in famiglie dove il rapporto tra le settimane lavorate nell'anno da tutti i componenti in età lavorativa e le settimane complessive teoricamente disponibili nell'anno è inferiore a 0,20.

- **DIS8** – Incidenza percentuale di individui che vivono in famiglia di età compresa tra 15 e 29 anni che non sono occupati e non sono iscritti ad alcun corso di studi sul totale della popolazione di età 15-29 anni residente in famiglia. L'indicatore misura il disagio sociale e educativo dei giovani dovuto all'uscita dai percorsi di istruzione e alla mancata occupazione. Fonte: CPPA, Ministero dell'Istruzione e del Merito (MIM) e Ministero dell'Università e della Ricerca (MUR).
- **DIS9** – Incidenza percentuale di studenti che vivono in famiglia che abbandonano la scuola o ripetono l'anno. L'indicatore è calcolato come rapporto percentuale tra gli studenti della scuola secondaria di I e II grado⁶ che abbandonano⁷ o che non hanno superato l'anno⁸ e gli studenti con segnali di frequenza nell'anno scolastico $t/t+1$; numeratore e denominatore fanno riferimento agli studenti che vivono in famiglia. L'indicatore rappresenta il disagio educativo degli studenti nella scuola secondaria di I e II grado, dovuto alle difficoltà nel percorso scolastico, misurato dagli eventi di abbandono scolastico o di mancato superamento dell'anno (esito finale negativo). Si tratta di un indicatore *proxy* della dispersione nella scuola secondaria superiore. Fonte: CPPA e Registro Tematico su Istruzione e Formazione.

L'Indice composito di Disagio Socio-Economico (IDISE)

L'IDISE è un indice composito costruito come combinazione di nove indicatori elementari che rappresentano le componenti socio-economiche più rilevanti del fenomeno e misurabili con elevata disaggregazione territoriale.

Tutti gli indicatori elementari sono stati selezionati con polarità positiva, in quanto concordi rispetto al disagio socio-economico, tranne il tasso di occupazione 25-64 anni per il quale un valore più alto dell'indicatore descrive una minore intensità del disagio.

La metodologia seguita per la costruzione dell'indice (*Adjusted Mazziotta-Pareto Index - AMPI*)⁹ produce un indice parzialmente non compensativo basato sui seguenti aspetti concettuali:

- il processo di normalizzazione (*Constrained Min-Max Method*) degli indicatori elementari permette il confronto dei valori delle unità statistiche, sia nel tempo che nello spazio, rispetto ad un riferimento comune che non cambia nel tempo;
- la funzione di aggregazione è una media penalizzata tramite una funzione di variabilità che evita l'effetto compensazione tipico delle trasformazioni lineari;
- la sequenza di queste due tecniche (prima la normalizzazione e poi l'aggregazione) permette di ottenere misure robuste che rispettano le proprietà attese per un indice composito.

La procedura di normalizzazione trasforma gli indicatori elementari rispetto a due valori, denominati *goal/posts*, e pone uguale a 100 un'unità statistica di riferimento ad un anno posto come base. I valori ottenuti cadono all'incirca nell'intervallo (70;130), dove 100 rappresenta il dato di riferimento.

L'indice composito è ottenuto aggregando gli indicatori elementari normalizzati con peso uguale, mediante media aritmetica semplice corretta aggiungendo alla media stessa un fattore (coefficiente di penalità) che dipende dalla variabilità dei valori normalizzati di ciascuna unità (variabilità orizzontale), ossia dalla variabilità degli indicatori rispetto ai valori di riferimento utilizzati per la normalizzazione.

⁶ Sono esclusi quelli in sezioni ospedaliere, carcerarie e serali e coloro che hanno segnali solo agli esami di stato.

⁷ Studenti che hanno segnali di abbandono nell'a.s. $t/t+1$ o di trasferimento senza avere segnali di reinscrizione in altra scuola del sistema pubblico e privato paritario nello stesso anno oppure studenti che hanno completato l'anno e che non risultano iscritti nel sistema di istruzione nazionale nell'a.s. successivo.

⁸ Studenti che, non avendo superato l'anno, si reinscrivono nella stessa classe l'anno successivo.

⁹ Per maggiori approfondimenti sugli indici compositi e sulla metodologia AMPI si veda: De Muro *et al.*, 2011; Mazziotta e Pareto, 2016, 2017, 2024.

L'indice composito della i -esima unità si ottiene applicando con penalità positiva la versione corretta del metodo delle penalità per coefficiente di variazione ($AMPI^+$) dove:

$$AMPI_i^+ = M_{r_i} + S_{r_i} cv_i$$

M_{r_i} e S_{r_i} sono rispettivamente la media aritmetica e lo scostamento quadratico medio dei valori normalizzati degli indicatori dell'unità i e $cv_i = S_{r_i}/M_{r_i}$ è il coefficiente di variazione dei valori normalizzati degli indicatori dell'unità i .

L'IDISE è calcolato per tutte le sezioni di censimento eleggibili all'analisi, aggregando i nove indicatori elementari normalizzati con peso uguale e utilizzando come base di riferimento (posta pari a 100) i valori medi comunali. Per le aree sub-comunali, ottenute dall'aggregazione di sezioni di censimento, l'indice è calcolato invece come media aritmetica ponderata dei valori dell'IDISE delle sezioni che compongono l'area, con peso pari alla popolazione residente in famiglia in ciascuna sezione. In questo modo si garantisce la confrontabilità dei valori dell'IDISE all'interno dello stesso comune, tra aree sub-comunali diverse e tra aree sub-comunali e sezioni di censimento.

Ai fini delle analisi di tipo comparativo, i valori dell'IDISE sono quindi confrontabili soltanto tra aree appartenenti allo stesso comune e non tra aree ricadenti in comuni diversi.

Domini territoriali di *output* a livello sub-comunale

L'IDISE e i relativi indicatori elementari sono prodotti con riferimento alla popolazione residente in famiglia al 31/12/2021 nelle sezioni di censimento 2021 di centro abitato e in cui sono presenti edifici ad uso prevalentemente residenziale¹⁰.

I dati sono diffusi per domini di *output* riferiti a due tipologie di aggregazioni di sezioni di censimento:

- ASC – Aree Sub-Comunali;
- ADU – Aree di Disagio socio-economico in ambito Urbano.

Le ASC sono suddivisioni del territorio comunale in unità amministrative e/o toponomastiche definite nell'ambito delle Basi Territoriali 2021¹¹. Per i comuni aventi più livelli di ASC, è stata considerata quella di massima disaggregazione territoriale¹².

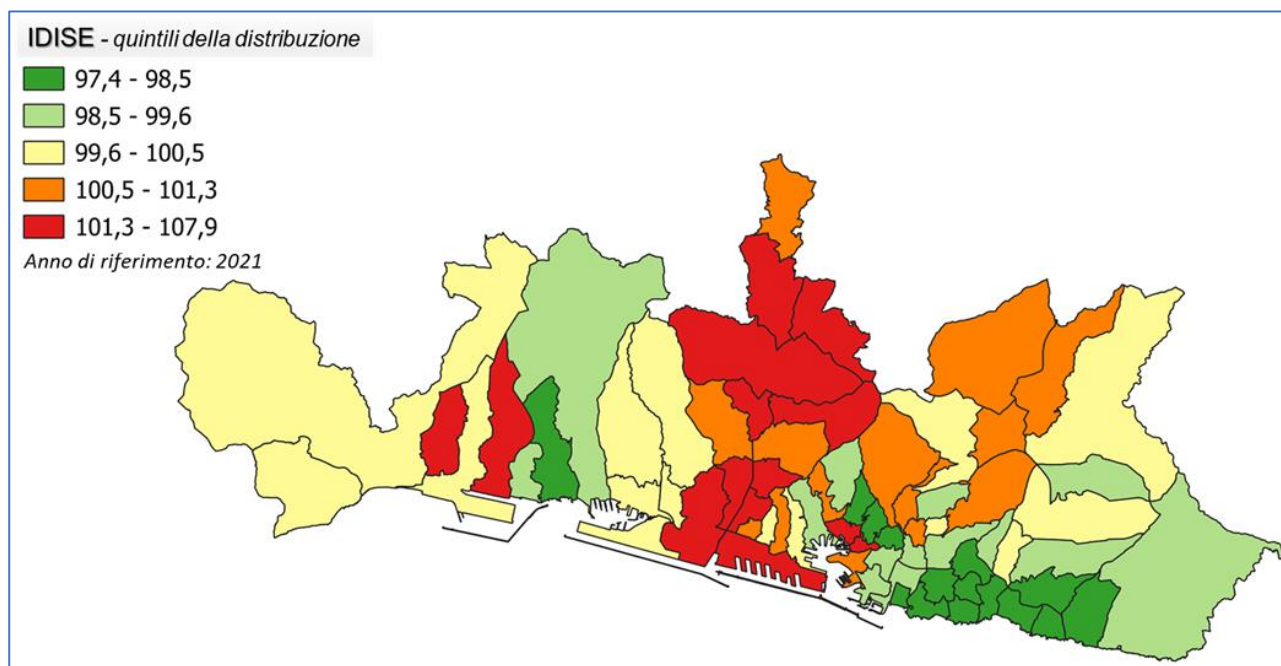
Nella Figura 1 è riportata, a titolo di esempio, la mappatura delle ASC prese in considerazione per il Comune di Genova (Zone Urbanistiche), secondo il valore dell'IDISE.

¹⁰ Lo studio ha considerato, come campo di osservazione, le sezioni con popolazione residente in famiglia di centro abitato e identificate dalla presenza di edifici ad uso prevalentemente residenziale. Per approfondimenti sulle Basi territoriali 2021 e sulle caratteristiche territoriali delle sezioni di censimento 2021 in base al loro uso/copertura del suolo si rimanda ai seguenti link: <https://www.istat.it/notizia/basi-territoriali-e-variabili-censuarie/>; <https://www.istat.it/notizia/caratteristiche-territoriali-sezioni-censimento-2021-raggruppate-in-macroaree/>

¹¹ L'Istat ha il compito di verificare il rispetto dei criteri stabiliti per il disegno delle ASC predisposte dai Comuni. Una volta validate, la mappatura delle ASC è resa disponibile nelle Basi Territoriali dell'Istat (ultimo rilascio ufficiale: 2021). Per approfondimenti: <https://www.istat.it/notizia/basi-territoriali-e-variabili-censuarie/>.

¹² Il livello di dettaglio delle ASC considerate è riportato nella Nota Informativa (Tavola 1).

FIGURA 1. IDISE per ASC. Zone Urbanistiche del Comune di Genova (ASC3). Anno 2021



Fonte: Istat

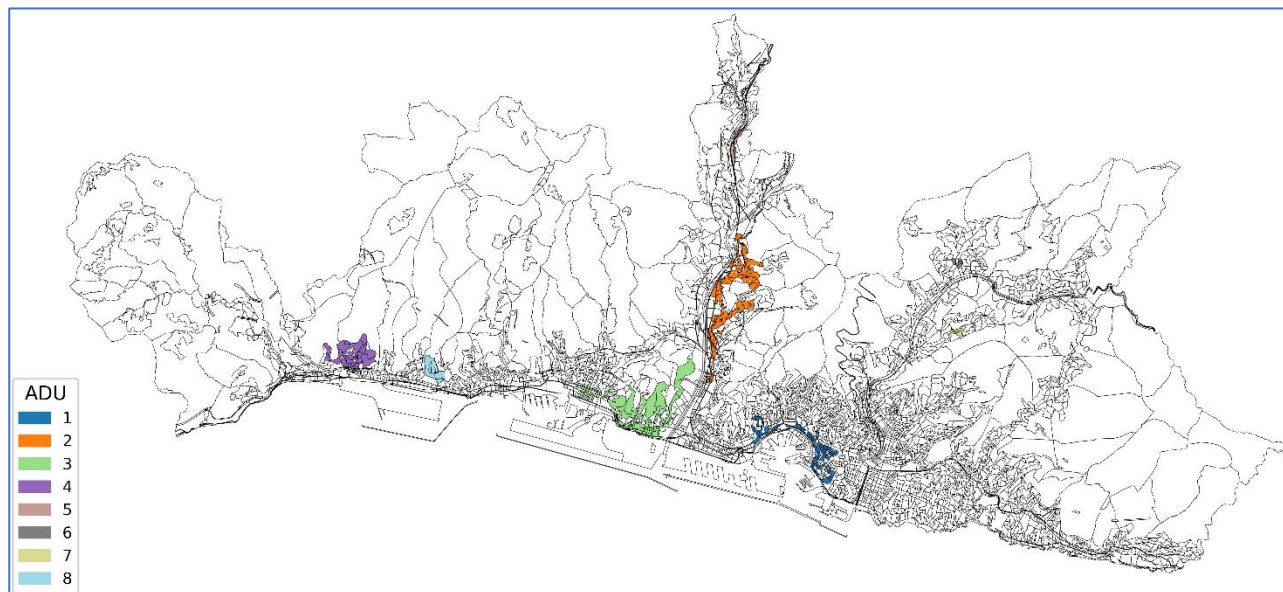
Le ADU sono aree sub-comunali caratterizzate dall'elevata presenza di condizioni di disagio socio-economico, appositamente definite per le finalità dello studio. Tali aree sono individuate tramite una procedura di tipo sequenziale sviluppata dall'Istat¹³ che, a partire dalle sezioni di censimento con i valori più alti dell'indice di disagio, aggrega sezioni contigue e omogenee rispetto ai valori dell'IDISE in maniera iterativa e con regole di arresto legate a predefiniti parametri dimensionali riferiti alla popolazione minima e massima dell'ADU e al livello di disagio misurato dall'IDISE (cfr. Allegato). I risultati per ADU si riferiscono pertanto ad aree sub-comunali più circoscritte rispetto alle ASC e le cui misure evidenziano condizioni di disagio particolarmente rilevanti.

Le ADU si contraddistinguono per dimensione e per livello di IDISE, a seconda del macro-ciclo iterativo in cui vengono identificate dalla procedura. Sono possibili tre tipologie di ADU: quelle di tipo 1 coincidono con le più ampie in termini di sezioni coinvolte e popolazione residente e presentano un livello di IDISE medio-alto; quelle di tipo 2 sono di dimensioni medie e hanno un IDISE alto; quelle di tipo 3 sono quelle più piccole, aggregano poche sezioni e mostrano un valore dell'IDISE molto alto. Le ADU di tipo 3 si riferiscono, quindi, a specifiche aree, solitamente molto circoscritte, dove gli indicatori elementari e l'indice composito mostrano valori molto elevati del disagio.

Nella Figura 2 è rappresentata, sempre a titolo di esempio, la mappatura delle ADU del Comune di Genova, distinguibili tra loro per la differente colorazione.

¹³ Una prima versione sperimentale dell'algoritmo è illustrata in Ballin *et al.*, 2025. Successivamente la procedura è stata rielaborata secondo tre macro-cicli, così come illustrato nell'Allegato.

FIGURA 2. Mappa del disagio comunale per ADU. Comune di Genova. Anno 2021



Fonte: Istat

Gli indicatori di contesto socio-demografico

I file rilasciati contengono anche i dati di un insieme di indicatori di contesto sulle caratteristiche socio-demografiche della popolazione residente nelle aree di output. Tali indicatori sono calcolati prevalentemente sulla base dei risultati del Censimento Permanente della Popolazione e delle Abitazioni (edizione 2021).

Così come per gli indicatori elementari di disagio, tutti gli indicatori di contesto hanno come riferimento la popolazione residente in famiglia al 31/12/2021 (sono escluse le convivenze), tranne CON1 (densità della popolazione), che invece è riferito alla popolazione totale per coerenza statistica con il denominatore. Di seguito vengono riportate, per ciascun indicatore, la descrizione e la metodologia di calcolo.

- **CON1** – *Densità della popolazione*. Rapporto tra la popolazione totale e la superficie (Kmq).
- **CON2** – *Incidenza percentuale della popolazione giovane*. Rapporto tra la popolazione 0-24 anni e il totale della popolazione, moltiplicato per 100.
- **CON3** – *Incidenza percentuale della popolazione anziana*. Rapporto tra la popolazione di 65 anni e oltre e il totale della popolazione, moltiplicato per 100.
- **CON4** – *Incidenza percentuale della popolazione straniera*. Rapporto tra la popolazione di cittadinanza straniera e il totale della popolazione, moltiplicato per 100.
- **CON5** – *Incidenza percentuale della popolazione anziana, straniera*. Rapporto tra la popolazione di cittadinanza straniera di 65 anni e oltre e il totale della popolazione straniera, moltiplicato per 100.
- **CON6** – *Incidenza percentuale degli studenti*. Rapporto tra la popolazione iscritta a un corso di studio¹⁴ e il totale della popolazione, moltiplicato per 100.

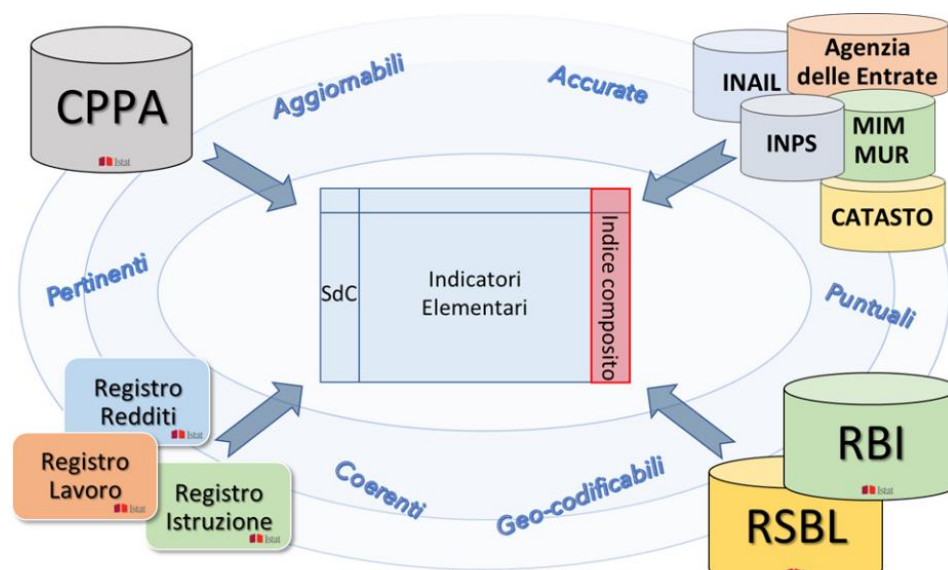
¹⁴ Per l'identificazione degli alunni iscritti a un corso di studio si utilizzano le seguenti fonti: MIM - Studenti delle scuole; MUR - Archivio degli studenti universitari - Iscritti e iscrizioni; MUR - Anagrafe Nazionale degli Studenti dei corsi Post laurea - Iscritti e iscrizioni (Dottorati, Master e Specializzazioni).

- **CON7** – *Incidenza percentuale dei laureati di età compresa tra 25 e 64 anni*. Rapporto tra la popolazione di 25-64 anni in possesso di un diploma di laurea e il totale della popolazione di 25-64 anni, moltiplicato per 100.
- **CON8** – *Tasso di occupazione 15-64 anni*. Rapporto tra la popolazione occupata di età compresa tra 15 e 64 anni e il totale della popolazione di 15-64 anni, moltiplicato per 100.
- **CON9** – *Rapporto di mascolinità*. Rapporto tra la popolazione maschile e la popolazione femminile, moltiplicato per 100.
- **CON10** – *Rapporto tra la popolazione giovane (0-24 anni) e la popolazione anziana (65 e più)* moltiplicato per 100.
- **CON11** – *Rapporto tra la popolazione straniera e la popolazione italiana* moltiplicato per 100.
- **CON12** – *Indice di ricambio della popolazione potenzialmente attiva*. Rapporto tra la popolazione di 20-24 anni e la popolazione di 60-64 anni, moltiplicato per 100.
- **CON13** – *Incidenza percentuale delle famiglie unipersonali*. Rapporto tra il numero delle famiglie costituite da una sola persona e il totale delle famiglie, moltiplicato per 100.
- **CON14** – *Incidenza percentuale delle famiglie con 5 componenti e più*. Rapporto tra il numero delle famiglie con almeno 5 componenti e il totale delle famiglie, moltiplicato per 100.

Le fonti utilizzate

Gli indicatori di disagio e di contesto sono elaborati utilizzando i dati (Istat, 2024a, 2024b) del Censimento Permanente della Popolazione e delle Abitazioni (CPPA) – integrati con alcuni archivi amministrativi e con le informazioni contenute nel Registro Statistico di Base dei Luoghi (RSBL) – e dei Registri statistici Tematici sviluppati dall'Istat su reddito, lavoro, istruzione e formazione. Le diverse fonti sono utilizzate in modo integrato con il Registro Base degli Individui (RBI) e sono geo-codificate¹⁵ alle sezioni di censimento 2021. La Figura 3 riassume il processo di integrazione delle diverse fonti di dati utilizzate per la produzione degli indicatori elementari di disagio e di contesto.

FIGURA 3. Integrazione e qualità dei dati a livello di sezione di censimento (SdC).



¹⁵ Sono possibili piccoli errori di geo-codifica che tuttavia producono effetti distorsivi di scarso rilievo, così come messo in luce dai numerosi test sperimentali effettuati.

I dati impiegati sono caratterizzati da elevati livelli di qualità in termini di accuratezza, puntualità, pertinenza e coerenza. Tali dati sono inoltre aggiornabili di anno in anno, così da permettere la replicazione dei risultati con cadenza annuale. Questo offrirà la possibilità di condurre analisi non solo da un punto di vista spaziale, ma anche in serie storica, di modo da aprire nuove frontiere per la ricerca tematica e metodologica dello studio dei fenomeni demografici, economici e sociali ad elevato dettaglio territoriale.

Riferimenti bibliografici

Ballin M., Bianchi G., Carbonetti G., Lorusso P. (2025). An approach for determining areas of socio-economic deprivation at sub-municipal level. *Italian Journal of Economic, Demographic and Statistical Studies – RIEDS* (*forthcoming*).

Carbonetti G., Biasciucci F., Cutillo A., Mazziotta M., Quondamstefano V., Tamburrano M.T., Tronu D. (2025a). Measuring socio-economic deprivation at sub-municipal level through the integration of census and administrative data. *Italian Journal of Economic, Demographic and Statistical Studies – RIEDS*, Volume LXXIX n.1 Gennaio-Marzo 2025.

Carbonetti G., Biasciucci F., Cutillo A., Mazziotta M., Quondamstefano V., Tronu D. (2025b). An innovative approach for the analysis of socio-economic phenomena at sub-municipal level: the household deprivation study project. *Data, Statistics and AI for Well-Being of People and Organizations*. Book of Short Papers of the ASA Rome Conference, pp. 121-126. Padua: Cleup. <https://doi.org/10.26398/asaproc.0090>

De Muro P., Mazziotta M., Pareto A. (2011). Composite Indices of Development and Poverty: An Application to MDGs. *Social Indicators Research*, Vol. 104 (1), pp.1-18.

Istituto Nazionale di Statistica (2024a). L'informazione statistica microterritoriale: l'esperienza del Registro statistico di base dei luoghi. E-book a cura di Ferruzza A., Di Zio M. Capitolo 13, Il Censimento della Popolazione e delle abitazioni e il Registro RSBL, pp. 95-100.

<https://www.istat.it/produzione-editoriale/informazione-statistica-microterritoriale-l'esperienza-del-registro-statistico-dei-luoghi/>.

Istituto Nazionale di Statistica (2024b). Nota metodologica di accompagnamento ai dati per sezione. <https://www.istat.it/wp-content/uploads/2023/06/Nota-Metodologica-sezioni-censimento-2021.pdf>

Mazziotta M., Pareto A. (2024). *Statistica per gli indici compositi*. Giappichelli Editore, Torino.

Mazziotta M., Pareto A. (2017). Synthesis of indicators: the composite indicators approach. In: *Complexity in Society: From Indicators Construction to their Synthesis*. Social Indicators Research Series. Filomena Maggino Editors, Springer, pp.159-191.

Mazziotta M., Pareto A. (2016). On a Generalized Non-compensatory Composite Index for Measuring Socio-economic Phenomena. *Social Indicators Research*, Vol. 127 (3), pp. 983-1003.

Per informazioni tecniche e metodologiche

Giancarlo Carbonetti, Andrea Cutillo, Elena Marchesich, Debora Tronu
progettodisagio@istat.it

Allegato - La procedura di costruzione delle ADU

1. Introduzione

La procedura di costruzione delle ADU (*Aree di Disagio socio-economico in ambito Urbano*) segue una logica di tipo aggregativo e sequenziale (Ballin *et al.*, 2025). Dopo aver individuato una sezione di censimento (SdC) con alto valore dell'IDISE (*Indice di Disagio Socio-Economico*) quale punto di partenza dell'algoritmo, si uniscono a questa altre SdC confinanti e omogenee rispetto alla misura del disagio IDISE calcolato per SdC. Il processo continua fino a che non sono violate predefinite regole di arresto dell'algoritmo, o non sono più identificabili SdC potenzialmente aggregabili, o non sono più identificabili nuovi punti di partenza della procedura.

La procedura è interamente parametrizzata: le regole di arresto e di individuazione delle SdC potenzialmente aggregabili sono determinate da quantità che possono essere modificate in relazione alle caratteristiche statistiche e dimensionali che si desiderano per le ADU finali.

La procedura prevede tre “macro-cicli” di iterazione. Alla fine del primo macro-ciclo, quando la procedura non individua più nuove ADU, i parametri vengono modificati per individuare nuove ADU mediamente più piccole, e che quindi possono individuare livelli più elevati di disagio. La stessa operazione viene implementata alla fine del secondo macro-ciclo.

2. I parametri del processo

I parametri che possono essere valorizzati, e che sono differenziati per ampiezza demografica del comune, sono:

- ❖ P_{min} , P_{max} : dimensione minima e massima (in termini di popolazione residente in famiglia) di ciascuna ADU:
 - se la procedura aggregativa termina (per mancanza di sezioni ulteriormente aggregabili) con un'area la cui dimensione non arriva alla soglia minima P_{min} , tale area non viene accettata come ADU e viene “sciolta” (e le SdC coinvolte nella costruzione dell'ADU rifiutata tornano disponibili per un'altra aggregazione);
 - la costruzione di un'ADU termina quando l'aggiunta di un'ulteriore sezione porta al superamento della soglia P_{max} . In tal caso, se le altre condizioni sono soddisfatte, si accetta l'area ottenuta nel passo precedente, prima dell'aggiunta della sezione che porta al superamento della soglia;
- ❖ P_{tot} : popolazione complessiva contenuta nell'insieme delle ADU create (in termini percentuali sul totale della popolazione residente in famiglia);
- ❖ Min_{SdC} : numero minimo di sezioni che possono formare un'ADU; questo parametro è stato fissato nell'algoritmo pari a 3;
- ❖ ε : parametro di omogeneità (percentuale) che definisce la soglia minima dell'IDISE di una sezione affinché questa possa essere aggregata all'area in costruzione; a seguito di vari test, questo parametro è stato fissato nell'algoritmo pari al 10%;
- ❖ τ : valore soglia minimo dell'IDISE dell'area ottenuta affinché questa possa essere considerata un'ADU valida. Il valore di questo parametro dipende dalla variabilità dell'IDISE sulle sezioni del comune (cfr. paragrafo 5, Tabella 1).

3. Impatto della parametrizzazione sul disegno delle ADU

La scelta del parametro (ε) di omogeneità della misura IDISE, della popolazione minima (P_{min}) appartenente all'ADU e della soglia minima di disagio (τ) hanno inevitabili riflessi sia sul numero

finale delle ADU che sulla loro dimensione. Di seguito, si riportano alcune conseguenze sulle ADU finali derivanti dalla scelta dei parametri:

- ε - minore è questo valore e maggiore sarà l'omogeneità delle SdC interne all'ADU. Questo parametro esprime, infatti, la differenza che può esservi tra l'IDISE della sezione candidata all'aggregazione e quella dell'area in costruzione. Maggiore sarà il valore di ε e maggiore sarà la dimensione delle aree, poiché queste possono contenere sezioni con maggiore eterogeneità rispetto all'IDISE. Al contrario, minore è ε e minore sarà la dimensione delle aree, poiché aumenta l'omogeneità interna alle aree stesse;
- P_{min} - maggiore è questo valore e più ampia dovrà essere un'ADU – in termini di popolazione – per essere accettata. Se si desiderano ADU ampie, questo valore dovrà essere elevato. Se invece si desiderano ADU capaci di cogliere situazioni di disagio più circoscritte e riferite ad aggregazioni di sezioni con minore popolazione, questo valore dovrà essere basso;
- τ - è un valore di soglia minima riferito all'indice IDISE. Se questo valore è alto la procedura individuerà solo aree a forte disagio socio-economico.

Al fine di cogliere ADU che potrebbero sfuggire ad una singola parametrizzazione, la procedura è stata implementata in modo tale da operare secondo tre “macro-cicli”, in ciascuno dei quali viene adottata una diversa parametrizzazione; questa dipende anche dalla dimensione demografica del comune sul quale viene applicata la procedura (paragrafo 5).

Il risultato finale è l'unione delle ADU risultanti dalle replicazioni avvenute nell'ambito dei tre macro-cicli. Ad ogni replicazione vengono eliminate all'avvio le sezioni già assegnate nelle replicazioni precedenti. La procedura prevede che ad ogni passo sia calcolato l'indice IDISE dell'area in costruzione. L'IDISE sull'area è calcolato come media aritmetica ponderata dei valori IDISE nelle sezioni di censimento dell'area, con peso pari alla popolazione delle diverse SdC.¹⁶

4. Descrizione dell'algoritmo

Di seguito sono illustrati i passi seguiti dalla procedura, relativa all'avvio del primo macro-ciclo, e sviluppati nell'algoritmo computazionale:

- 1) Tra tutte le SdC che formano un comune, si definisce la lista L_E delle sezioni eleggibili all'analisi¹⁷.
- 2) Tra tutte le sezioni della lista L_E si prende quella con l'IDISE più alto (SdC_start1). Questa sezione diventa il punto iniziale di un'aggregazione.
- 3) Tra tutte le sezioni con IDISE superiore a 100 (media del comune), si individuano quelle contigue¹⁸ al punto di aggregazione SdC_start1 e che soddisfano il parametro di omogeneità ε ; ovvero, si individuano le sezioni contigue che hanno un IDISE > 100 e che differisce percentualmente per meno di ε dall'IDISE del punto di aggregazione.
- 4) Si creano delle aree aggregando il punto iniziale di aggregazione con ciascuna delle sezioni individuate al punto 3). Si ottengono tante nuove aree aggregate (“aggregazioni di sezioni di censimento”, d'ora in avanti definite “AGG”) quante sono le SdC individuate al punto 3).
- 5) Per ognuna di queste AGG si ricalcola l'IDISE e la popolazione complessiva e, successivamente, si effettuano i seguenti controlli:

¹⁶ Si noti che l'IDISE sull'area può essere calcolato anche secondo la metodologia AMPI*. Si è scelta l'opzione presentata nel testo per coerenza con il metodo di calcolo dell'IDISE adottato per le Aree Sub-Comunali (ASC) definite dai Comuni per motivi amministrativi, statistici o funzionali. Si evidenzia, comunque, che i valori prodotti con i due metodi sono molto simili.

¹⁷ Tale lista include tutte le sezioni di censimento con “TIPOLOC=1”, “COD_TIPO_S=1” e “pop_fam>0”. Quindi, possono formare le ADU solo le sezioni di “centro urbano”, con presenza di edifici ad “uso prevalentemente residenziale” e popolazione residente (in famiglia) censita.

¹⁸ La contiguità è definita in base ad un “buffer” di 25 metri costruito intorno a ciascuna sezione di censimento. Quindi, affinché due sezioni possano considerarsi contigue è sufficiente che le rispettive “aree di buffer” abbiano almeno un punto di sovrapposizione.

- a. $Pop(AGG) \leq P_{max}$ (la popolazione dell'aggregazione non può eccedere la dimensione massima prevista);
- b. $IDISE(AGG) \geq \tau$, (l' $IDISE$ dell'aggregazione deve essere maggiore della soglia minima fissata per le ADU).

6) Tra le AGG che soddisfano entrambe le condizioni a. e b. del punto 5) si sceglie quella con il massimo disagio, ovvero con $IDISE$ più elevato. Questa AGG diventa la nuova area di aggregazione e si torna al passo 2).

I passi 2) - 6) vengono replicati fino a che le condizioni a. e b. del passo 5) sono soddisfatte e si trovano ancora sezioni contigue sufficientemente omogenee.

Quando i passi 2) - 6) non possono più essere iterati si fanno i seguenti controlli:

- Cond1: $Pop(AGG) \geq P_{min}$, ovvero si controlla se la popolazione risultante dall'ultima aggregazione utile abbia una dimensione sufficiente;
- Cond2: Il numero di SdC dell'AGG sia non inferiore al minimo ammesso (Min_{SdC}), pari a 3.

Se una delle due condizioni (Cond1 e Cond2) non è soddisfatta, l'aggregazione non è accettata e viene sciolta. In tal caso la procedura riparte da una nuova sezione: si torna al passo 2) e si riparte con un nuovo punto di aggregazione coincidente con la sezione con il secondo valore più elevato dell' $IDISE$ (SdC_start2).

Se entrambe le condizioni (Cond1 e Cond2) sono soddisfatte l'AGG creata è accettata come prima ADU. In questo caso si tolgono tutte le sezioni che fanno parte dell'ADU dalla lista L_E e si riparte dal passo 2), individuando come punto iniziale della successiva aggregazione la sezione con il più elevato valore dell' $IDISE$ tra quelle rimanenti (SdC_start2).

Dopo l'identificazione della prima ADU, si prosegue con successive iterazioni della procedura finora descritta, aggiungendo alla verifica delle condizioni Cond1 e Cond2 il controllo dell'ammontare P_{tot} . Se entrambe le condizioni (Cond1 e Cond2) sono soddisfatte e se aggiungendo la popolazione della nuova aggregazione a quella delle ADU già accettate si ha una popolazione totale inferiore a P_{tot} , la procedura prosegue con la ricerca di ulteriori ADU. Se, invece, entrambe le condizioni (Cond1 e Cond2) sono soddisfatte e se aggiungendo la popolazione della nuova aggregazione a quella delle ADU già accettate si ha una popolazione totale superiore a P_{tot} , l'aggregazione non è accettata, viene sciolta e la procedura si ferma.

La procedura si ferma anche nel caso in cui non siano più disponibili punti iniziali.

A questo punto la procedura passa al secondo macro-ciclo, con una diversa scelta dei parametri e la lista L_E ridotta delle sezioni che costituiscono le ADU individuate nel primo macro-ciclo. Vengono quindi ripetuti i passi 2) - 6) fino a quando la procedura non si ferma di nuovo. Dopodiché, in modo analogo, si passa al terzo macro-ciclo e si riapplicano i passi 2) - 6) con una terza parametrizzazione e una lista L_E ulteriormente ridotta o al più identica (nel caso in cui al secondo macro-ciclo non sia stata determinata alcuna ADU).

5. Le parametrizzazioni scelte

Ciascuno dei tre "macro-cicli" descritti in precedenza ha una propria parametrizzazione; di volta in volta si innalza la soglia τ e si abbassa la soglia P_{min} . In questo modo, si cercano ADU che abbiano via via un maggior livello di disagio, anche se caratterizzate da una minor dimensione demografica. Si tratta quindi di aree più "focalizzate" e "circoscritte" sul fenomeno in esame, sia in termini di misura del fenomeno che in termini di dimensione di popolazione. Con il passaggio dal primo al terzo macro-ciclo aumenta anche la dimensione del parametro P_{tot} .

Inoltre, i parametri fissati per i tre macro-cicli sono stati definiti in base alla dimensione del comune in esame¹⁹. Sono state fissate tre classi di popolazione: oltre 500mila abitanti; tra 100 e 500mila abitanti; fino a 100mila abitanti.

Nella Tabella 1 sono riportate le parametrizzazioni utilizzate nei tre macro-cicli della procedura. La scelta dei valori assegnati ai parametri è stato il risultato, da un lato, di un lungo percorso sperimentale e, dall'altro, dell'esito di vari incontri con alcuni Comuni in fase di validazione dei risultati preliminari per diversi scenari.

Tabella 1. Parametri fissati per i tre macro-cicli della procedura, definiti per tre classi demografiche dei comuni.

1° macro-ciclo		P_{min}	P_{max}	τ	P_{tot}
classe 1	> 500mila	5.000	15.000	$100 + \sigma$	10%
classe 2	100-500mila	2.500	7.500	$100 + \sigma$	10%
classe 3	<100mila	1.500	3.500	$100 + \sigma$	10%
2° macro-ciclo		P_{min}	P_{max}	τ	P_{tot}
classe 1	> 500mila	2.500	7.500	$100 + 1,5 \sigma$	20%
classe 2	100-500mila	1.500	3.500	$100 + 1,5 \sigma$	20%
classe 3	<100mila	750	2.000	$100 + 1,5 \sigma$	20%
3° macro-ciclo		P_{min}	P_{max}	τ	P_{tot}
classe 1	> 500mila	250	2.500	$100 + 2\sigma$	30%
classe 2	100-500mila	250	1.500	$100 + 2\sigma$	30%
classe 3	<100mila	250	750	$100 + 2\sigma$	30%

σ è lo scarto quadratico medio della distribuzione dei valori IDISE calcolati per sezione di censimento.

¹⁹ A riguardo è stato conteggiato il numero di unità statistiche che entrano nell'analisi: "popolazione residente in famiglia" censita nel 2021 nelle sezioni di censimento eleggibili allo studio del disagio ("TIPOLOC=1" & "COD_TIPO_S=1").