

Allegato 1

Nota metodologica¹

Base integrata di microdati amministrativi “Persons and Places 2016”

PSN 2661 - Popolazioni che usano un territorio e mobilità per studio e lavoro

Sommario

1	Definizioni, classificazioni	2
1.1	L’universo, le unità statistiche e le principali variabili.....	3
2	L’equazione della popolazione insistente.....	5
3	Metodologia di costruzione della base informativa.....	6
4	Variabili rilasciate nella base finale.....	10
5	Controlli e correzioni.....	11
5.1	Trattamento di errori non campionari.....	11
5.2	Controllo territoriale rispetto all’elenco dei comuni al 31 dicembre 2016 e trattamento degli errori 12	
5.3	Copertura universo degli individui	13
5.4	Controllo della componente “Residente” nella Popolazione Insistente rispetto al benchmark “Bilancio demografico”	13
5.5	Controllo della componente “Residenti stranieri” nella Popolazione Insistente rispetto al benchmark “Bilancio demografico”	14
5.6	Copertura del campo di osservazione universo degli studenti e degli universitari.....	15
5.7	Missing e valori non dovuti	16
5.8	Valori anomali	17

¹ A cura di Roberta Vivio vivio@istat.it e Gaia Rocchetti grocchetti@istat.it

1 Definizioni, classificazioni

La Popolazione Insistente è costituita dall'unione degli individui italiani e stranieri che risultano iscritti in anagrafe in Italia alla data di riferimento - sia in famiglia che in convivenza - e da quelli (sia iscritti in anagrafe che non iscritti ma con domicilio fiscale) che hanno segnali amministrativi di lavoro o studio sul territorio italiano.

In particolare:

- a livello nazionale è costituita da: tutti gli individui residenti in Italia, tutti gli individui che lavorano o studiano in Italia (residenti, e non residenti con domicilio fiscale), dagli stranieri presenti per motivi occasionali (es. turisti). Non comprende gli italiani i residenti all'estero.
- a livello comunale è costituita da: tutti gli individui che risiedono nel comune senza segnali di lavoro o studio, tutti gli individui residenti nel comune con mobilità per studio o lavoro all'interno dello stesso comune di residenza, tutti gli individui non residenti nel comune in entrata da un altro comune per studio o lavoro. Infine comprende i non residenti nel comune che sono presenti per motivi occasionali (es. turisti). Esclude i residenti/domiciliati nel comune che lavorano/studiano in altro comune.

Nell'ambito del progetto è stata studiata una tassonomia che classifica gli individui in cinque tipologie in funzione del *motivo*, della *frequenza*, e del *verso della mobilità casa/lavoro (casa/studio)*.

Tav. 1 - Popolazione insistente per tipologia di individuo e caratteristiche dello spostamento

Tipologia di individuo		Spostamento		
strato		Motivo(*)	Verso	Frequenza
1	Individui Statici senza mobilità per lavoro/studio	Nessuna attività di Studio, lavoro	-	--
2	Individui Dinamici con mobilità in entrata per studio o lavoro	Studio, lavoro	In entrata da altra area	Giornaliera o Periodica
3	Individui Dinamici con mobilità in uscita per studio o lavoro	Studio, lavoro	In uscita verso altra area	Giornaliera o Periodica
4	Individui Dinamici con mobilità all'interno per studio o lavoro	Studio, lavoro	All'interno dell'area	Giornaliera o Periodica
5	City users	Turismo, eventi sportivi, salute, ecc.	In entrata da altra area	Occasionale

(*) per studio si intende la formazione scolastica di qualsiasi livello (scuola primaria, scuola secondaria, università).

Per il motivo e il verso si fanno le seguenti assunzioni: che gli spostamenti avvengano dal luogo di residenza (origine) al luogo di destinazione (studio/lavoro), che ogni lavoratore/studente si sposti

per svolgere la sua attività effettuando uno ed un solo spostamento, e che uno studente si sposti per studio, un lavoratore per lavoro².

Allo stato attuale non sono disponibili informazioni utili a distinguere tra frequenza giornaliera e periodica, né quelle per la individuazione dei *city users*. L'output riguarda quindi le prime quattro tipologie di individuo.

1.1 L'universo, le unità statistiche e le principali variabili

L'universo target è costruito integrando gli individui iscritti all'Anagrafe della Popolazione residente (Liste Anagrafiche Comunali – LAC - allo 01/01/2017), con gli individui che hanno segnali di lavoro/studio in Italia nel mese di dicembre dell'anno 2016 (Tav. 2a) e che siano registrati nell'anagrafe fiscale.

Le unità:

Lavoratore: individuo con almeno un segnale di presenza negli archivi/registri seguenti:

- Istat – Registro ASIA DB Occupazione
- MIUR - Archivio del personale universitario
- MIUR - Archivio del personale delle scuole statali
- INPDAP - Posizioni degli assicurati iscritti
- MEF - NOIPA - Cedolini stipendiali
- INPS - Rapporti di lavoro domestico
- INPS - Lavoratori autonomi in agricoltura.

Tali archivi/registri contengono informazioni relative ad occupati di imprese, enti, aziende, ecc. localizzate in Italia e facenti parte del sistema economico italiano.

Si assume che il segnale sia indicativo della condizione di “lavoratore che si reca nella sede di lavoro”, ovvero che l'informazione sia riferibile ad una persona che effettua degli spostamenti casa/lavoro.

Studente: individui con un segnale di presenza nell'archivio:

- MIUR - Anagrafe degli studenti 2016/17

L'archivio contiene le iscrizioni scolastiche in strutture italiane sul territorio italiano. Dalle scuole dell'infanzia a quelle superiori di secondo grado. Sono considerati eleggibili soltanto gli studenti per i quali è indicato che “frequentano”. Si assume che l'informazione sia riferibile ad una persona che effettua degli spostamenti casa/scuola.

Studente universitario: individui con un segnale di presenza nell'archivio:

- MIUR - Archivio degli iscritti e delle iscrizioni universitarie 2016/17.

² Per gli universitari lavoratori si assume predominante la motivazione “lavoro”.

Si assume che tale presenza sia indicativa della condizione “universitario che frequenta corsi accademici”, ovvero che l’informazione sia riferibile ad una persona che effettua degli spostamenti casa/università.

Nota:

- Gli archivi di cui sopra non contengono gli individui con iscrizione anagrafica in Italia che lavorano o studiano all’estero, ovvero oltre confine o in territori fuori giurisdizione (es. organizzazioni o università internazionali su territorio italiano). Tali individui pertanto sono conteggiati nella popolazione insistente come residenti, ma non sono conteggiati nella popolazione con mobilità per lavoro/studio. Nella stessa fattispecie rientrano anche i pensionati, i disoccupati, le casalinghe.
- le duplicazioni di segnali sono state risolte durante il processo con regole deterministiche.

Le variabili della base intermedia:

Sesso, luogo nascita e data nascita

Le variabili sesso, luogo nascita e data nascita, sono estratte dalla base dati anagrafica dell’Anagrafe tributaria, se presenti, altrimenti per ciascuna di esse è stata selezionata la modalità più frequente presente in SIM INDIVIDUI.

Le modalità fuori range sono state imputate con “NULL”.

Residenza e Domicilio fiscale

La residenza anagrafica è quella della LAC al 31/12/2016. Il domicilio fiscale è quello dell’Anagrafe Tributaria al 1/1/2017.

Cittadinanza e codice famiglia

Sono estratte dall’Anagrafe della popolazione residente.

Data di decesso

E’ estratta dalla base dati anagrafica dell’Anagrafe tributaria.

Comune di Origine della Mobilità per studio/lavoro: si assume che per l’individuo con un segnale di lavoro/studio sia il comune di iscrizione anagrafica, se è iscritto nelle LAC riferite al 1.1.2017; e sia il comune di domicilio fiscale riportato nella base dati dell’Anagrafe Tributaria al 1.1.2017 per chi non lo è.

- Liste anagrafiche comunali (LAC) al 01.01.2017;
- Agenzia delle Entrate -Anagrafe delle Persone Fisiche al 1.1.2017

Per gli individui appartenenti ai comuni le cui liste anagrafiche comunali non sono entrate nel processo di produzione è stata imputata l’iscrizione anagrafica del 2015 (1.1.2016) se presente, e in subordine il domicilio fiscale del 2016.

Comune di destinazione della Mobilità: si assume che per l’individuo con un segnale di studio sia il comune registrato in uno dei seguenti archivi (subordinatamente al tipo di unità di analisi):

- MIUR - Anagrafe delle scuole statali e non statali
- MIUR - Anagrafica delle strutture universitarie

- MIUR - Anagrafica dei corsi universitari

Si assume che per l'individuo con un segnale di lavoro sia il comune registrato in uno dei seguenti archivi (subordinatamente al tipo di unità di analisi):

- Istat - ASIA_Unità Locali
- MEF - NOIPA - Cedolini stipendiali
- MEF - NOIPA - Uffici di Servizio
- MIUR – Personale Universitario
- MIUR – Personale Scuole
- INAIL – Posizioni assicurative territoriali
- INPS_Rapporti di Lavoro domestico
- INPS_Lavoratori Autonomi Agricoltura

Per l'identificazione del luogo di lavoro è stata studiata una procedura ad hoc di tipo gerarchico.

2 L'equazione della popolazione insistente

Si definisce *Popolazione insistente* (PI) in una data area di riferimento j :

$$\begin{aligned}
 PI_j &= NO_LUS_j + LUS_{jj}^{\leftrightarrow} + LUS_{ij}^{\rightarrow} + CU_{ij}^{\rightarrow} = \\
 &= Pop. residente - LUS_{ji}^{\leftarrow} + LUS_{ij}^{\rightarrow} + CU_{ij}^{\rightarrow} \\
 i &= 1, \dots, n; j = 1, \dots, n; i \neq j \\
 (n &= \text{numero di comuni}; i = \text{comune di origine}; j = \text{comune di destinazione})
 \end{aligned}$$

dove:

- (1) NO_LUS_j = *Individui statici senza attività di lavoro/studio*: individui non lavoratori, non studenti e non universitari con residenza anagrafica nel comune j (es. bambini in età pre-scolare, pensionati, casalinghe, disoccupati).
- (2) LUS_{ij}^{\rightarrow} = *Individui Dinamici con mobilità in entrata (in j) per studio o lavoro*: lavoratori, studenti o universitari con residenza anagrafica/domicilio fiscale in un comune i diverso da quello di lavoro/studio. La mobilità è in entrata nel comune di studio/lavoro j , in uscita da quello di residenza/domicilio i .
- (3) LUS_{ij}^{\leftarrow} = *Individui Dinamici con mobilità in uscita (da j) per studio o lavoro*: lavoratori, studenti o universitari con residenza anagrafica/domicilio fiscale in un comune diverso da quello di lavoro/studio. La mobilità è in entrata nel comune di studio/lavoro i , in uscita da quello di residenza/domicilio j .
- (4) $LUS_{jj}^{\leftrightarrow}$ = *Individui Dinamici con mobilità all'interno (di j) per studio o lavoro*: lavoratori, studenti o universitari con residenza anagrafica/domicilio fiscale nello stesso comune j di lavoro/studio. La mobilità è interna al comune.

(5) $CU_{ij} \rightarrow$ = *City users*: individui con residenza anagrafica/domicilio fiscale in un comune *i* diverso da quello di destinazione verso cui si spostano per turismo, salute, sport, ecc. La mobilità è in entrata nel comune di destinazione, in uscita da quello di residenza/domicilio. La frequenza è occasionale.

3 Metodologia di costruzione della base informativa

Il processo di produzione è ingegnerizzato. Gli input sono unità e variabili di fonti amministrative e registri statistici, integrati con linkage deterministico basato sulla chiave *codice_individuo_SIM*. Per la valutazione della qualità sono stati effettuati controlli di coerenza e copertura rispetto alle fonti ufficiali e controlli sugli errori non campionari nei microdati (missing, dati anomali, fuori dominio, errori di classificazione). Sono state effettuate esclusivamente correzioni deterministiche per il recupero di valori mancanti e la ricodifica di valori anomali. Per la tutela della riservatezza sono state realizzate aggregazioni e soppressione di alcune modalità delle variabili, sono state limitate al minimo le classificazioni incrociate e, infine, sopprese le celle a rischio di violazione. I risultati della valutazione della qualità hanno guidato la scelta del livello territoriale e delle variabili per l'output presentato.

I passi del metodo seguito sono stati sinteticamente i seguenti:

1. Ricognizione tra le fonti amministrative disponibili in Istat nel Sistema Integrato di Microdati (SIM) volta all'individuazione di quelle contenenti le unità e le variabili di interesse: ciascuna fonte è stata esaminata sia per il possibile contributo in termini di unità di rilevazione e/o di analisi, sia per quello in termini di variabili dirette, derivate, di controllo (Tav. 2). Si è constatato che non esistono informazioni amministrative utili a quantificare la categoria degli occasionali e a determinare la frequenza degli spostamenti che consentirebbe di distinguere tra *giornalieri* ed i *periodici*. Anche la categoria dei *city users* non è al momento quantificabile attraverso i dati amministrativi³.

Nella selezione delle fonti è stato seguito il principio di pertinenza e non ridondanza che ha portato a concludere per una scelta limitata ad alcune variabili estratte da 21 archivi. Dal momento che le fonti amministrative, se considerate separatamente, non contengono tutte le variabili di interesse e, nella maggior parte dei casi, si riferiscono a una parte della popolazione target, si rende necessaria la loro integrazione. L'integrazione consente di ottenere sia una completezza delle variabili di interesse (per esempio, per arrivare alla copertura della variabile "comune di lavoro" occorre acquisire la stessa informazione da una decina di archivi, unificarla e armonizzarla) sia una copertura pressoché totale dell'universo di riferimento (ottenuta unificando il segnale di presenza sul territorio da 13 archivi).

2. Utilizzo combinato di archivi e variabili guidato dai risultati della ricognizione e dalla progettazione: alcuni entrano nella fase finalizzata all'individuazione della popolazione residente, altri in quella dei non residenti con segnali di attività in Italia (studenti, studenti

³ Come sviluppo futuro è possibile studiare modalità di integrazione di dati aggregati a livello comunale rilevati da indagine.

universitari, lavoratori), altri nella fase di individuazione dei comuni di origine/destinazione degli spostamenti (Tav. 2a).

3. Validazione a livello di microdato: controllo e correzione sugli errori non campionari (fuori dominio, missing, ecc.⁴).
4. Validazione a livello macro: controllo rispetto a fonti di benchmark ufficiali⁵.
5. Realizzazione della base integrata finale in cui ogni record corrisponde ad uno strato di unità. Gli strati corrispondono alle combinazioni delle modalità delle variabili finali. In nessuno strato è presente una numerosità inferiore alla soglia che permette l'identificazione delle unità.

⁴ Vedi par. 5.1.

⁵ Vedi parr. 5.4 - 5.8.

Tav. 2 – Fonti e variabili e loro principale utilizzo. Anno 2016

Titolare o detentore dell'archivio	Denominazione dell'archivio	Denominazione variabili estratte	Utilizzo
Istat	Rilevazione delle liste anagrafiche comunali (LAC)	presenza nell'archivio; residenza (indirizzo, comune, provincia), cittadinanza	individui residenti; comune_residenza; cittadinanza
Agenzia delle entrate	Base dati dell'anagrafe tributaria	sex, luogo di nascita, domicilio (indirizzo, comune, provincia)	individui non residenti; sesso; comune_nascita; data di nascita; comune_domicilio
INAIL	Denunce nominative assicurative	presenza nell'archivio	segnale di lavoro
INAIL	Lavoratori parasubordinati	sede di lavoro	comune_lavoro
INAIL	Lavoratori interinali	localizzazione della sede di lavoro	comune_lavoro
INPS	Posizioni degli assicurati iscritti (ex INPDAP - Gestione dipendenti pubblici)	sede dell'ente (indirizzo, comune, provincia), ragione sociale	segnale di lavoro; comune_lavoro
INPS	Rapporti di lavoro domestico	presenza nell'archivio; sede di lavoro (indirizzo, comune, provincia);	segnale di lavoro; comune_lavoro
INPS	Lavoratori autonomi in agricoltura	presenza nell'archivio; tipo azienda, comune, provincia, sede dell'azienda (indirizzo, comune, provincia)	segnale di lavoro; comune_lavoro
INAIL	Archivio delle Posizioni assicurative territoriali	luogo di svolgimento dell'attività (indirizzo, comune, provincia)	comune_lavoro
MEF - Ministero dell'Economia e delle finanze	Banca dati reddituale	presenza nella base, residenza anagrafica (indirizzo, comune, provincia)	segnale di lavoro; comune_residenza
MEF - Ministero dell'economia e delle finanze	NOIPA - Anagrafica, cedolini e Uffici di Servizio	sede di servizio (indirizzo, comune, provincia)	segnale di lavoro; comune_lavoro
MIUR	Anagrafe del personale docente e non docente universitario	sede di servizio (indirizzo, comune, provincia); tipo di contratto	segnale di lavoro; comune_lavoro
MIUR	Anagrafe del personale del comparto scuola	sede di servizio (indirizzo, comune, provincia); tipo contratto	segnale di lavoro; comune_lavoro
Istat	Registro Asia Occupazione	presenza nell'archivio; qualifica, tipo di contratto, tipo di azienda	segnale di lavoro
Istat	Registro delle imprese e delle unità locali ASIA	sede dell'UL (indirizzo, comune, provincia)	comune_lavoro
MIUR	Anagrafe degli alunni delle scuole primarie e secondarie di primo e secondo grado	presenza nell'archivio	segnale di studio
MIUR	Anagrafe delle scuole statali primarie e secondarie di primo grado	sede della scuola (indirizzo, comune, provincia)	comune_studio
MIUR	Anagrafe delle scuole statali secondarie di secondo grado	sede della scuola (indirizzo, comune, provincia)	comune_studio
MIUR	Anagrafe nazionale degli studenti universitari - ANS	presenza nell'archivio; facoltà, corso, anno di corso	segnale di studio; comune_studio
MIUR	Anagrafica delle Università	denominazione, sede (indirizzo, comune, provincia)	comune_studio
MIUR	Anagrafica dei corsi	denominazione corso, ateneo, sede (indirizzo, comune, provincia)	comune_studio

Tav. 2a – Fonti e variabili estratte per principale utilizzo. Anno 2016

Titolare o detentore dell'archivio	Denominazione dell'archivio	Universo di riferimento	sesso, età, luogo di nascita	cittadinanza	Comune di origine	Comune di destinazione	
					comune di residenza/ domicilio fiscale	comune di lavoro	comune di studio
Istat	Rilevazione delle liste anagrafiche comunali (LAC)	x	x	x	x		
Agenzia delle entrate	Base dati dell'anagrafe tributaria	x			x		
INAIL	Denunce nominative assicurative	x					
INAIL	Lavoratori parasubordinati	x				x	
INAIL	Lavoratori interinali					x	
INPS	Posizioni degli assicurati iscritti (ex INPDAP - Gestione dipendenti pubblici)					x	
INPS	Rapporti di lavoro domestico					x	
INPS	Lavoratori autonomi in agricoltura					x	
INAIL	Archivio delle Posizioni assicurative territoriali	x				x	
MEF - Ministero dell'Economia e delle finanze	Banca dati reddituale						
MEF - Ministero dell'economia e delle finanze	NOIPA - Anagrafica, cedolini e Uffici di Servizio					x	
MIUR	Anagrafe del personale docente e non docente universitario					x	
MIUR	Anagrafe del personale del comparto scuola					x	
Istat	Registro Asia Occupazione	x					
Istat	Registro delle imprese e delle unità locali ASIA	x				x	
MIUR	Anagrafe degli alunni delle scuole primarie e secondarie di primo e secondo grado						
MIUR	Anagrafe delle scuole statali primarie e secondarie di primo grado						x
MIUR	Anagrafe delle scuole statali secondarie di secondo grado						x
MIUR	Anagrafe nazionale degli studenti universitari - ANS						
MIUR	Anagrafica delle Università	x					x
MIUR	Anagrafica dei corsi						x

4 Variabili rilasciate nella base finale

Le variabili e le relative modalità della base finale sono:

Var.1. **Comune di origine della mobilità:** residenza, domicilio fiscale, null;

Var.2. **Comune di destinazione della mobilità:** comune di lavoro, di studio, null;

Var.3. **Tipo di attività individuo:** lavoratore, studente, universitario, altro;

Var.4. **Tipologia di insidenza dell'individuo:**

1. Individui Statici (senza mobilità per lavoro/studio),
2. Individui Dinamici con mobilità in entrata (per studio o lavoro),
3. Individui Dinamici con mobilità in uscita (per studio o lavoro),
4. Individui Dinamici con mobilità all'interno dell'area (per studio o lavoro).

Per ogni combinazione di modalità viene rilasciata la relativa numerosità di strato, ovvero il numero di individui che appartengono allo strato.

Per garantire una maggiore qualità del dato, i codici dei comuni di origine e destinazione che hanno meno di 5000 abitanti sono posti a *null*, mantenendo solo il codice provincia.

5 Controlli e correzioni

La base dati grezza è stata controllata al fine di enucleare le situazioni di incoerenza al suo interno. Quelli qui presentati sono solo una parte dei controlli di qualità effettuati.

Sulla base dati individuale sono stati effettuati controlli sui microdati (missing, fuori dominio, incoerenze tra variabili, ecc.) e controlli aggregati rispetto a fonti ufficiali di benchmark, (sintetizzati in indici di errore). Tali controlli hanno messo in luce e quantificato le incoerenze presenti, gli errori di classificazione, di misurazione, di rappresentazione (sovra - sotto copertura della popolazione di riferimento), brevemente descritti di seguito. Infine sono stati valutati i valori estremi ed anomali.

Gli oggetti da controllare e le fonti di benchmark utilizzate sono elencate nello schema seguente:

Oggetto nella base	Fonte di benchmark
Popolazione residente	Bilancio demografico - Istat
Popolazione residente straniera	Bilancio demografico - Istat
Studenti scuole primarie/secondarie	Anagrafe studenti - Miur
Studenti universitari	Anagrafe studenti - Miur
Comuni	Classificazione ufficiale dei Comuni - Istat

5.1 Trattamento di errori non campionari

Sono state definite alcune regole di controllo per individuare alcune fattispecie di errore o situazioni non ammissibili e le relative regole di intervento. Alcuni esempi:

- **Molteplicità di presenza:** nel caso di individui duplicati nelle LAC è stata scelta l'unità con data di iscrizione in anagrafe più recente.
- **Valore mancante per il comune di iscrizione anagrafica:**
 - agli individui residenti nei comuni le cui liste anagrafiche comunali non sono entrate nel processo di produzione ma catturati da segnali di lavoro/studio, è stato applicato il seguente processo di recupero del comune di iscrizione anagrafica:
 - ⊖ dell'anno precedente ove presente,
 - ⊖ associata al CAP per i comuni multicap.
- **Molteplicità di stato:**
 - nel caso di individui con segnali di lavoro nel mese di dicembre, e studio, è stato privilegiato lo stato “studente” per i minori, e lo stato di “lavoratore” per i maggiorenni.

- se l’individuo esercita più attività di lavoro/studio durante l’anno, è considerata quella svolta a dicembre;
- se un lavoratore ha più rapporti di lavoro con imprese diverse, per poter assegnare a ogni individuo un unico riferimento geografico si attribuisce a ogni soggetto un unico rapporto di lavoro seguendo il criterio del contratto gerarchicamente superiore.

5.2 Controllo territoriale rispetto all’elenco dei comuni al 31 dicembre 2016 e trattamento degli errori

L’unione di informazioni territoriali da archivi amministrativi di diversi gestori necessiterebbe di un lavoro di armonizzazione rispetto all’*Elenco dei comuni al 31 dicembre 2016*⁶. I controlli (Tav. 3) mettono in luce e quantificano la presenza di disallineamento tra i comuni presenti negli archivi di input - e quindi nella base dati integrata intermedia riferita a dicembre 2016 – e la classificazione vigente.

Tav. 3 Comuni italiani nella base integrata per tipo di relazione dell’individuo con il comune e validità rispetto alla classificazione 31.12.2016.

Relazione individuo/comune	Comuni presenti nei dati di input	di cui non validi al 31.12.2016	Individui nei comuni non validi
Comune di residenza	7.978	163	621.194
Comune di domicilio fiscale	8.120	277	673.033
Comune di studio	6.934	45	61.137
Comune di lavoro	8.272	278	74.842
Comune di iscrizione universitaria	267	0	0

Per la produzione della base 2016 sono state acquisite le LAC disponibili per 7.978 comuni, su 7.998 comuni ufficiali al 31.12.2016. Il disallineamento è interamente dovuto alla fattispecie “cessazione del codice per fusione di diversi comuni o cambiamento di provincia dei comuni stessi”. Nel dettaglio, i 163 codici_comune presenti nella LAC e non nella lista dei comuni ufficiali al 31/12/2016, corrispondono ai 183 che sono invece presenti tra i comuni ufficiali e non nelle LAC. Il gap è, quindi, solo apparente ed è attribuibile alla diversa data di registrazione dell’elenco comuni nelle due fonti.

Per i comuni fuori classificazione si è optato per un intervento conservativo di pulizia del dato individuale, piuttosto che di correzione: i codici comune non presenti nella lista ufficiale del comune al 31/12/2016, per un totale di 640.208 individui, sono stati mantenuti nella base seppur valorizzati con un codice di anomalia.

⁶ Elenco dei comuni al 31 dicembre 2016 (<http://www.istat.it/it/archivio/6789>)

5.3 Copertura universo degli individui

Le regole di eleggibilità della popolazione insistente prevedono che:

Sono considerati eleggibili:

- tutti gli individui con codice SIM
- gli individui appartenenti ad almeno una delle LAC pervenute all'Istat, validate entro la data di inizio del processo di produzione e acquisite nel sistema SIM⁷
- gli individui che abbiano segnali di lavoro/studio provenienti dagli archivi e registri statistici considerati che siano presenti in AT e/o in LAC.

Non sono eleggibili:

- gli individui non residenti e non occupati, i non residenti iscritti a università straniera sul territorio italiano, i non residenti non regolarmente presenti sul territorio italiano.
- gli individui non presenti in LAC e non presenti in AT.

La popolazione insistente così determinata al 31 dicembre 2016 è di 61,331 milioni (Tav. 4). Sono circa 33 milioni gli individui che hanno segnali amministrativi di lavoro, studio, di iscrizione universitaria a dicembre, e tra essi 464 mila sono stranieri non residenti. In totale in Italia sono 636 mila gli stranieri non residenti presenti a fine anno che hanno comunque avuto un segnale di attività durante l'anno.

Tav. 4 – Popolazione insistente al 31.12.2016 per tipo di individuo

Popolazione insistente	Lavoratori, studenti, universitari	Altri	Totale
con iscrizione anagrafica e/o domicilio fiscale	32.612.832	28.082.075	60.694.907
solo con domicilio fiscale (stranieri non residenti)	464.541	172.186	636.727
Totale	33.077.373	28.254.261	61.331.634

5.4 Controllo della componente “Residente” nella Popolazione Insistente rispetto al benchmark “Bilancio demografico”

La popolazione residente nel Bilancio Demografico (BD) al 31.12.2016 (pari a 60,589 milioni) è il benchmark rispetto a cui è stato effettuato il controllo della popolazione residente nella base dati della Popolazione Insistente. In particolare sono stati calcolati l'errore assoluto medio (AARE)

⁷ Le LAC che arrivano oltre la dead line stabilita da SIM sono archiviate fuori SIM.

rispetto alla distribuzione della popolazione residente per comune e l'indice di correlazione (Tav. 5).

$$\text{Average Absolute Relative Error} = AARE(\hat{\theta}) = \frac{1}{D} \sum_{d=1}^D ARE_d = \frac{1}{D} \left| \frac{\hat{\theta}_d}{\theta_d} \right| - 1$$

Tav. 5 - Errore assoluto medio della popolazione residente nella Popolazione insistente rispetto al Bilancio Demografico e indice di correlazione

Errore tra le distribuzioni per comune	0,0315
Indice di correlazione tra le distribuzioni per comune	0,9999

Valutando la distribuzione per ampiezza dell'errore, risulta che il 95,5% dei comuni differisce per meno del 5%, ed il 65,2% è al di sotto dell'1% (Tav.6).

Tav. 6 - Comuni per classe di differenza relativa tra popolazione residente nella base rispetto al Bilancio Demografico - totale residenti italiani

Classe di differenza r. %	N comuni	% comuni	% cumulata comuni
0	308	3,9	3,9
]0-1]	4.910	61,4	65,2
]1-5]	2.422	30,3	95,5
]5-10]	130	1,6	97,1
>10	228	2,9	100,0
Totale complessivo	7.998	100,0	

5.5 Controllo della componente “Residenti stranieri” nella Popolazione Insistente rispetto al benchmark “Bilancio demografico”

La Popolazione straniera nel Bilancio Demografico ammonta a 5,023 milioni. Nella base dati integrata è pari a 5,049 milioni. Anche in questo caso sono stati calcolati l'errore e la correlazione (Tav. 7).

Tav. 7 - Errore assoluto medio della popolazione residente straniera nella Popolazione insistente rispetto al Bilancio Demografico e indice di correlazione

Errore tra le distribuzioni per comune	0,0366
Indice di correlazione tra le distribuzioni per comune	0,9999

Valutando la distribuzione per ampiezza dell'errore, si trova che il 82,7% dei comuni differisce per meno del 5%.

Tav. 8 - Comuni per classe di differenza relativa tra la Popolazione residente straniera nella base e Popolazione residente straniera nel Bilancio Demografico Istat

Classe di differenza %	N comuni	% comuni	% cumulata comuni
0	2.444	30,6	30,6
]0-1]	1.847	23,1	53,7
]1-5]	2.325	29,1	82,7
]5-10]	647	8,1	90,8
>10	520	6,5	97,3
Non calcolabile (a)	215	2,7	100,0
Totale complessivo	7.998	100,0	

(a) Denominatore (pop. Straniera in DEMO uguale a 0)

5.6 Copertura del campo di osservazione universo degli studenti e degli universitari

Ai fini della popolazione insistente per studio sono considerati eleggibili gli individui che risultano iscritti e frequentanti dall'anagrafe degli studenti del Miur⁸.

Per le scuole primarie e secondarie va detto che gli archivi escludono:

- i dati degli alunni degli istituti non paritari iscritti all'apposito albo, considerati validi per l'assolvimento dell'obbligo, ma non per il rilascio di titolo riconosciuto;
- i dati degli alunni di istituti di istruzione stranieri con sede in Italia;
- i dati degli alunni degli istituti statali e paritari di Bolzano, per i quali vengono trasmessi i soli dati individuali relativi ai nominativi dei promossi con lode, in quanto è previsto un premio;
- i dati degli alunni degli istituti paritari di Trento. Una situazione simile relativa agli alunni degli istituti di Aosta è in via di superamento.

Sussistono problemi di sottocopertura relativamente ai dati dei soggetti che assolvono l'obbligo scolastico presso riformatori o case circondariali, nonché agli alunni delle scuole militari (alle quali ci si iscrive a 16 anni, dopo l'obbligo e che sono sede di esami di stato). Nella Popolazione Insistente sono stati considerati eleggibili soltanto gli studenti frequentanti.

Per il sistema Universitario sono stati considerati gli iscritti a un corso universitario sia del vecchio che del nuovo ordinamento risultanti dall'anagrafe degli studenti universitari per l'anno accademico

⁸ Fonti: **Scuole primarie statali e non statali, Scuole secondarie di I° grado statali e non statali:** Elaborazione e diffusione dei risultati delle rilevazioni effettuate dal Ministero dell'istruzione, dell'università e della ricerca. Le informazioni riguardano consistenza e caratteristiche delle scuole, del relativo personale e degli iscritti.

Scuole secondarie di II° grado statali e non statali: Elaborazione e diffusione dei risultati della rilevazione sulle scuole secondarie di secondo grado, effettuata dal Ministero dell'istruzione, dell'università e della ricerca. Le informazioni riguardano consistenza e caratteristiche delle scuole, del relativo personale e degli iscritti. Dall'a.s. 2006/2007 all'a.s. 2009/10, tra gli iscritti della P.A. di Bolzano sono inclusi anche gli iscritti alle prime e seconde classi delle scuole professionali, riconosciute dalla Legge Finanziaria 2007 (comma 623) per l'adempimento dell'obbligo di istruzione (fonte ASTAT).

2015/2016. Da tale archivio sono esclusi gli studenti di sedi distaccate in Italia di università straniere. L'informazione sulla frequenza ai corsi non è disponibile pertanto tutti gli studenti universitari sono considerati eleggibili. Inoltre, si fa l'assunzione che ogni studente universitario insista sul comune presso il quale ha sede il suo Ateneo di iscrizione. Gli studenti lavoratori ai fini della mobilità sono considerati lavoratori.

5.7 Missing e valori non dovuti

È stata effettuata l'analisi di tutti i valori mancanti nella base intermedia, al fine di distinguere tra *missing* e valori non dovuti.

Per la Cittadinanza si registrano 640 mila missing, a fronte di 636 mila unità non in Lac. L'archivio da cui viene attinta la Cittadinanza è la LAC, che è esaustiva (a meno di 3283 casi per tutti gli individui che ad essa fanno riferimento), quindi sono casi recuperabili da altri archivi.

Per il Luogo di nascita, che è una variabile dovuta per chi è nato in Italia, il 64% dei casi di missing riguarda proprio cittadini stranieri. Il rimanente 36% è relativo a cittadini italiani che potrebbero essere nati all'estero. Il 92% coincide con l'ammontare di stranieri.

Tavola 9 – Valori mancanti su alcune variabili Core per tipo di individuo. Valori assoluti.

Variabile e tipo di individuo	Frequenza
Sesso	79
Data di nascita	0
Cittadinanza	640.010
<i>Tipo di individuo:</i>	
• Non in Anagrafe	636.727
• In anagrafe	3.283
• Universitari	30.447
• Studenti	55.081
• Lavoratori	369.634
Provincia / Comune NASCITA	6.006.783
<i>Tipo di individuo:</i>	
• Non in Anagrafe	480.320
• In anagrafe	5.526.463
• Universitari	94.377
• Studenti	293.011
• Lavoratori	2.735.522
• Cittadinanza straniera	3.882.812
• Cittadinanza italiana	1.642.496

5.8 Valori anomali

Per l'individuazione dei valori anomali nella base si è utilizzato il metodo del campo di variazione interquartile. L'intervallo di accettazione di una generica osservazione (Hidioglou e Berthelot, 1986) sarà dato da:

$$A = (q_{50} - c_{inf} d_{inf}; q_{50} + c_{sup} d_{sup}) = (A_{inf}; A_{sup})$$

dove c_{inf} e c_{sup} sono parametri arbitrari; d_{inf} e d_{sup} sono le distanze dalla mediana del primo quartile e del terzo quartile. Per $c=1$ l'intervallo coincide con lo scarto interquartile.

Ciò premesso, si definiscono anomali i punti esterni all'intervallo:

$$A = (q_{50} - 3(q_{50} - q_{25}); q_{50} + 3(q_{75} - q_{50})) = (A_{inf}; A_{sup})$$

In particolare, sono stati creati 42 strati di popolazione determinati dalla combinazione delle variabili:

- tipo di attività dell'individuo: lavoratore, studente, universitario, altro individuo.
- cittadinanza: italiana, non italiana, null.
- tipo di residenza: iscritto soltanto in LAC, iscritto sia in LAC che in AT, iscritto soltanto in AT.

Sono stati calcolati anche gli indici di posizione classici (media, Mediana), di variazione (scarto quadratico medio), e di asimmetria (Fisher):

$$\text{Indice di asimmetria di Fisher} \quad A = \frac{m - Me}{s}$$

Gli strati sono stati quindi ordinati per dimensione, e i valori anomali sono stati calcolati nei primi 13 strati, che complessivamente coprono il 98,5% della Popolazione Insistente (Tav.20). In totale si è trovato che i casi fuori intervallo sono circa 375 mila (0,62%).

Le incidenze maggiori di valori anomali si riscontrano negli strati 2,4, 5, 9 e 10. Per questi strati si procede ad un approfondimento, ovvero alla ricerca di questi casi all'interno di altri archivi amministrativi. I risultati sono presentati nelle tavole seguenti.

Tav. 10 - Presenza dei record anomali dello strato S2 "Lavoratori tra 15 e 74 anni, italiani, residenti e domiciliati" per presenza in archivi amministrativi

Frequenza assoluta	Incidenza% nello strato	Denominazione Archivio/Registro
112.719	72,8	DB_occupazione
154.910	100,0	Agenzia delle Entrate - Anagrafe delle Persone Fisiche
154.810	99,9	LAC
153.779	99,3	MEF - Banca dati reddituale
147.239	95,0	INPS - Casellario dei Pensionati
114.211	73,7	ACI - Parco veicoli circolanti
110.894	71,6	CCIAA - Archivio Persone d'Impresa - Iscrizione della società nel Registro delle Imprese
63.520	41,0	INPS - Archivio dei lavoratori autonomi: artigiani e commercianti

In questo strato l'anomalia scatta oltre i 69 anni e riguarda 154 mila individui. Si riscontra che il 95% di essi è Pensionato, ed il 72,8% ha un lavoro (il 71,6% è una persona d'impresa, ed il 41,0% è un lavoratore autonomo). Tutti sono presenti nella Banca dati reddituale. Si ritiene dunque che questo sottoinsieme di individui sia costituito da Occupati con reddito da pensione (in Italia nel 2016 sono stati stimati in 436 mila unità⁹).

Tav. 11 - Presenza dei record anomali dello strato S4 "Altri individui, non italiani, residenti e domiciliati " per presenza in archivi amministrativi

Frequenza assoluta	Incidenza% nello strato	Denominazione Archivio/Registro
51.323	100,0	Agenzia delle Entrate - Anagrafe delle Persone Fisiche
49.073	95,6	LAC
28.760	56,0	Ministero degli interni - Permessi di soggiorno
17.828	34,7	INPS - Casellario dei Pensionati
7.797	15,2	MEF - Banca dati reddituale
5.742	11,2	ACI - Parco veicoli circolanti

In S4 la soglia superiore di anomalia si fissa a 74 anni (nello strato l'età media e mediana è di 35 anni), ed interessa il 2,07% degli individui. Un terzo delle 50 mila persone in situazione di anomalia risultano essere nel Casellario dei pensionati, e più della metà ha un permesso di soggiorno. Si tratta verosimilmente di stranieri disoccupati o con un reddito da pensione.

Tav. 12 - Presenza dei record anomali dello strato S5 "Lavoratori tra 15 e 74 anni, non italiani, residenti e domiciliati" per presenza in archivi amministrativi

Frequenza assoluta	Incidenza% nello strato	Denominazione Archivio/Registro
27.283	100,0	Agenzia delle Entrate - Anagrafe delle Persone Fisiche
26.191	96,0	LAC
20.094	73,7	INAIL - Denunce Nominative Assicurative
19.093	70,0	INPS - Rapporti di lavoro domestico
18.050	66,2	Ministero degli interni - Permessi di soggiorno
14.895	54,6	MEF - Banca dati reddituale

Anche in S5 la soglia di allarme è oltre i 60 anni: 63 per la precisione, contro i 39 e i 40 anni dell'età mediana e media. L'incidenza è del 1,43%, ed in questo caso si tratta per lo più di lavoratori domestici stranieri.

⁹ https://www.istat.it/it/files/2017/12/Focus_pensionati_DEF.pdf

Tav. 13 - Presenza dei record anomali dello strato S9 "Universitari, italiani, residenti e domiciliati" per presenza in archivi amministrativi

Frequenza assoluta	Incidenza% nello strato	Denominazione Archivio/Registro
103.365	100,0	Agenzia delle Entrate - Anagrafe delle Persone Fisiche
103.365	100,0	MIUR - Archivio degli iscritti e delle iscrizioni universitarie
103.130	99,8	LAC
41.683	40,3	ACI - Parco veicoli circolanti
40.006	38,7	MEF - Banca dati reddituale

La soglia è a 28 anni, e determina un 7,92% di casi di anomalie nello strato. Non ci sono segnali di redditi, né di pensioni. E' verosimile che si tratti di studenti fuori corso.

Tav. 14 - Presenza dei record anomali dello strato S10 "Lavoratori tra 15 e 74 anni, universitari, residenti e domiciliati, Italiani" per presenza in archivi amministrativi

Frequenza assoluta	Incidenza% nello strato	Denominazione Archivio/Registro
14.142	100,0	Agenzia delle Entrate - Anagrafe delle Persone Fisiche
14.142	100,0	MIUR - Archivio degli iscritti e delle iscrizioni universitarie
14.122	99,9	LAC
14.050	99,3	MEF - Banca dati reddituale

In questo caso, in cui la numerosità è di 319 casi, l'anomalia scatta dopo i 26/29 anni. I 14 mila casi rappresentano il 4,43% del totale. Il 99,3% di essi è presente anche nella Base dati reddituale, cosa che rende plausibile l'ipotesi che siano studenti/lavoratori.

Tavola 15 – Strati di popolazione e valori anomali per strato

Tipo attività dell'individuo							Percentili, Mediana, media, Stdev							Soglie dei valori anomali			Numero di casi - Valori assoluti nello strato		Numero di casi - Incidenza % sullo strato	
Strato	Lavoratore	Studente	Universitario	Tipo_Residenza	Tipo_cittadinanza	Numero individui nello strato	D10	D25	Me	D75	D90	μ	STDEV	INF	SUP	I.Fisher	< INF	> SUP	< INF	> SUP
1	altro	altro	altro	LD	Italiano	27.100.145	19	40	63	75	83	56	24	-6	99	-0,29	0	17.334	0,00	0,06
2	OCC15_74	altro	altro	LD	Italiano	19.993.543	28	36	45	53	60	45	12	18	69	0,00	6.232	154.910	0,03	0,77
3	altro	Stu_primaria	altro	LD	Italiano	2.499.099	6	7	8	9	10	8	1	5	11	0,00	0	0	0,00	0,00
4	altro	altro	altro	LD	Non Italiano	2.481.318	5	23	35	48	60	35	19	-1	74	0,00	0	51.323	0,00	2,07
5	OCC15_74	altro	altro	LD	Non Italiano	1.908.871	26	32	39	47	55	40	11	18	63	0,09	1.103	27.283	0,06	1,43
6	altro	Stu_second1	altro	LD	Italiano	1.547.443	11	11	12	13	13	12	1	9	15	0,00	0	0	0,00	0,00
7	altro	Stu_second2	altro	LD	Italiano	1.413.660	16	16	17	18	18	17	1	14	20	0,00	0	0	0,00	0,00
8	altro	Stu_infanzia	altro	LD	Italiano	1.347.626	3	3	4	5	5	4	1	1	7	0,00	0	0	0,00	0,00
9	altro	altro	Universitario	LD	Italiano	1.305.280	19	20	22	24	27	23	5	16	28	0,20	0	103.365	0,00	7,92
10	OCC15_74	altro	Universitario	LD	Italiano	319.143	20	22	26	34	44	29	9	14	50	0,33	0	14.142	0,00	4,43
11	altro	Stu_primaria	altro	LD	Non Italiano	249.370	6	7	8	9	10	8	1	5	11	0,00	0	0	0,00	0,00
12	altro	Stu_infanzia	altro	LD	Non Italiano	156.388	3	3	4	5	5	4	1	1	7	0,00	0	0	0,00	0,00
13	altro	Stu_second1	altro	LD	Non Italiano	117.926	11	11	12	13	13	12	1	9	15	0,00	0	0	0,00	0,00
TOTALE																	7.335	368.357	0,01	0,61

LETTERATURA DI RIFERIMENTO

- Istat, Movimento turistico in Italia (2018), <https://www.istat.it/it/archivio/224376>
- Istat, Spostamenti quotidiani e nuove forme di mobilità (2018), <https://www.istat.it/it/archivio/224469>
- Basile, R., Girardi, A., Mantuano, M., Russo, G., (2017), “Interregional migration, human capital externalities and unemployment dynamics: evidence from Italian provinces”
<http://hdl.handle.net/10419/168560>
- City of Melbourne. (2017) “Daily population estimates and forecasts”
- De Angelis, I., Mariani, V., Torrini, R., (2017) “New evidence on interregional mobility of students in tertiary education: the case of Italy”. *Questioni di economia e finanza*
- Faggian, A., Rajbhandari, I., Dotzel, K., (2017) “The interregional migration of human capital and its regional consequences: a review”, *Regional studies*, 51:1, 128-143 DOI: 10.1080/00343404.2016.1263388
- Istat (2017), “Forme, livelli e dinamiche dell’urbanizzazione”. *Lecture statistiche*. ISBN: 978-88-458-1916-2
<http://www.istat.it/it/archivio/1>
- Statistics New Zealand, (2017), “Experimental population estimates from linked administrative data”
- Vivio, R., (a cura di) (2016b) “Studenti e Bacini Universitari”, *Istat Lecture Statistiche*.
- Istat, (2015) “La nuova geografia dei sistemi locali”
- Salvucci, G., Abbate, C., (2015) “Densità demografica notturna e diurna”. Conferenza ESRI ITALIA.
- Istat (2014), “Gli spostamenti quotidiani per motivi di studio e di lavoro”
- Vivio, R. (2014), “Il sottoprogetto sperimentale: popolazione insistente”. In Garofalo: *Il Progetto ARCHIMEDE obiettivi e risultati sperimentali*. Istat Working Paper, 1/2014
- Eurostat, (2013), “Feasibility studies on the usually resident population and vital event”. Working group on population statistics, Luxembourg
- Eurostat, “Statistics on commuting patterns at regional level” http://ec.europa.eu/eurostat/statistics-explained/index.php/Statistics_on_commuting_patterns_at_regional_level
- Biagi, B., Faggian, A., McCann, P. (2011) “Long and Short Distance Migration in Italy: The Role of Economic, Social and Environmental Characteristics”, 6:1, pp.111-131.
- City of Melbourne.(2009) “Melbourne city users Estimates and Forecasts”, 2004-2020 (di Silk. B., Bell, J.A.)
- Istat (2009), “Atlante di geografia statistica ed amministrativa”.
- Palvarini, P., (2009) “Popolazioni nuove e forme dell’abitare”.
- Cristaldi, F., (a cura di) (2003) “Le mille popolazioni metropolitane, un’analisi geografica dell’area romana”
- Nuvolati, G. (2003) “Resident e Non Resident”, *The journal of regional analysis and policy*, 33:2, pp. 67-83.

Gismondi, R (2002), “Un confronto tra metodi di identificazione di osservazioni outlier in indagini longitudinali finalizzate alla stima di una variazione: proposte teoriche e verifiche empiriche”, in *Statistica*, anno LXIII, n2, 2002.

Martinotti, G. (1994) “New social morphology”, UNESCO/MOST Wien, 10-12 February 1994.

Martinotti (1993), “Metropoli”, Il Mulino, Bologna.