

Nota metodologica¹

Condizioni socio-economiche delle famiglie 2016 – ARCH.I.M.E.DE

1 Base dati integrata

La base dati statistica “Condizioni socio-economiche delle famiglie – ARCH.I.M.E.DE” deriva dall’integrazione di vari archivi amministrativi ed è costruita con l’intento di essere uno strumento di conoscenza di alcune delle connessioni esistenti in una famiglia tra aspetti socio-demografici ed aspetti economici. Per questa finalità, le famiglie residenti sono descritte attraverso i tratti salienti della loro struttura demografica, di alcune caratteristiche relative alla partecipazione al mercato del lavoro e ai percorsi di istruzione dei componenti e ai redditi dei percettori. Nella base sono compresenti sia variabili di tipo sociale sia variabili di natura economica, riferite alle famiglie e agli individui che le compongono. Per motivi di qualità, i dati forniti nell’output sono relativi ai soli comuni con più di 5.000 abitanti².

Il riferimento del progetto nel Piano Statistico Nazionale è IST- 02663.

2 Archivi utilizzati per la costruzione della base dati integrata

Nella Tabella 1 sono riportati gli archivi amministrativi elementari ed i relativi contenuti informativi utilizzati nel processo.

Tabella 1 - Archivi di base e principali variabili acquisite. Anno di riferimento delle Fonti: 2016 (dove non indicato diversamente)

Fonti di dati	Ente titolare	Contenuti informativi
Lista Anagrafiche Comunali (LAC) allo 01/01/2017	Comuni italiani	Caratteristiche anagrafiche dei componenti della famiglia; dimensione e tipologia familiare
Banca Dati Reddituale (BDR)	Ministero dell'Economia e delle Finanze (MEF)	Selezione di informazioni presenti nelle dichiarazioni dei redditi (modelli Unico Persone Fisiche, 730, 770)
Casellario dei pensionati	Istituto Nazionale della Previdenza Sociale (INPS)	Importi delle prestazioni pensionistiche
Rapporti di lavoro domestico	INPS	Stima della retribuzione percepita dai lavoratori domestici
Uniemens	INPS	Importi degli assegni al nucleo familiare per i dipendenti del settore privato
Modello Unico Persone Fisiche	Agenzia delle Entrate	Reddito derivante dall’esercizio di attività commerciali o dall’esercizio di arti e professioni per determinate tipologie di contribuenti
Trattamenti non pensionistici	INPS	Erogazioni temporanee in denaro a individui/famiglie (es: disoccupazione;

¹ A cura di Sara Casacci – casacci@istat.it

² L’informazione relativa alla numerosità della popolazione residente utilizzata è derivata dai dati ufficiali della Rilevazione Istat sulla “Popolazione residente comunale per sesso, anno di nascita e stato civile”. Dal momento che i codici dei comuni della citata rilevazione sono riferiti al 31/12/2016, mentre i codici delle LAC utilizzate per la costruzione della base dati sono riferiti allo 01/01/2017, si registrano delle discrepanze tra le due fonti relativamente ai comuni per cui sono intercorse delle variazioni territoriali tra il 31/12/2016 e il 01/01/2017. Queste discrepanze determinano l’esclusione di alcuni comuni dalla base dati pubblicata. Per ovviare, almeno in parte, a questo problema, i comuni delle province di Sassari e del Sud Sardegna per cui risultano più di 5000 abitanti sono stati comunque reintegrati nella base dati rilasciata.

		Assegni al nucleo familiare)
Censimento Popolazione 2011	Istat	Livello di Istruzione
Anagrafe degli studenti	Ministero dell’Istruzione, dell’Università e della Ricerca (MIUR)	Livello di Istruzione, partecipazione a percorsi formativi scolastici
Archivio degli iscritti e delle iscrizioni universitarie	MIUR	Livello di Istruzione, partecipazione a percorsi formativi universitari
“Precarietà lavorativa – ARCH.I.M.E.DE”	Istat	Occupazione

3 Universo e unità di analisi

L’universo di riferimento è costituito dall’insieme delle famiglie i cui componenti risultano iscritti nelle Liste Anagrafiche dei Comuni (LAC) alla data di riferimento (01/01/2017 per la base dati del 2016).

L’unità di analisi è quindi la famiglia anagrafica, intesa come insieme degli individui residenti in famiglia (non sono considerati gli individui residenti in convivenza) e definita come *l’insieme di persone legate da vincoli di matrimonio, parentela, affinità, adozione, tutela o affettivi, coabitanti ed aventi dimora abituale nello stesso comune. Una famiglia può essere costituita anche da una sola persona*³.

Nelle indagini campionarie sulle famiglie condotte dall’Istat⁴ l’unità di rilevazione è costituita, invece, dalla famiglia di fatto, intesa come un *insieme di persone coabitanti e legate da vincoli affettivi, di matrimonio, parentela, affinità, adozione e tutela*.

4 Variabili di output

Le principali variabili contenute nella base dati rilasciata sono relative:

- alla tipologia familiare, ricostruita a partire dalle informazioni anagrafiche dei componenti;
- all’istruzione (dalla scuola primaria fino all’università);
- alla partecipazione al mercato del lavoro;
- al reddito lordo familiare ed equivalente.

Tali informazioni consentono all’utente di segmentare l’universo delle famiglie in modo flessibile, enucleando le caratteristiche più funzionali alle analisi, identificando e qualificando sottopopolazioni specifiche.

5 Metodologia utilizzata per la costruzione della collezione di dati

La popolazione di riferimento della base dati è costituita dalle famiglie anagrafiche; tuttavia, allo scopo di collezionare in modo organizzato dati sui caratteri della popolazione di riferimento, si è partiti dalla raccolta di informazioni sull’individuo, che si configura come unità di rilevazione. La base dati familiare “Condizioni socio-economiche delle famiglie – ARCH.I.M.E.DE” rappresenta quindi un output di secondo livello rispetto ad una base dati individuale intermedia, costruita dall’integrazione delle fonti amministrative. La costruzione della base dati individuale intermedia è descritta nel paragrafo 5.2, mentre le principali procedure per la derivazione della base dati familiare sono riportate nel par. 5.3.

³ Regolamento Anagrafico (legge 24 Dicembre 1954, n. 1228; DPR 30 Maggio 1989, n.223).

⁴ Istat (2009). Navigando tra le fonti demografiche e sociali.

5.1 Informazioni sul campo di osservazione

L'input del processo è costituito dalle Liste Anagrafiche Comunali (LAC), dalle quali vengono selezionati gli individui che soddisfano la seguente *condizione di elegibilità*: “Un individuo è considerato elegibile se ha un codice identificativo valido, e se appartiene ad una famiglia in cui tutti i componenti hanno un codice identificativo valido”. Per il 2016, gli individui non elegibili sono lo 0,4 per cento del totale.

La base di dati sulle famiglie è derivata da quella individuale: i componenti di una famiglia sono identificati dalla medesima combinazione delle variabili codice provincia, codice comune e codice famiglia.

5.2 Informazioni sulla procedura per la determinazione delle variabili della base dati individuale intermedia

Dalle LAC sono tratte o derivate le variabili riferite alle caratteristiche demografiche delle famiglie (numero, età e cittadinanza dei componenti, tipologia familiare).

Le variabili reddituali sono il risultato di un'integrazione effettuata a partire da una selezione di variabili presenti negli archivi disponibili in Istat, che sono:

- Ministero dell'Economia e Finanze - Banca Dati Reddittuale
- INPS - Casellario dei Pensionati⁵
- INPS - Archivio dei lavoratori domestici
- Agenzia delle Entrate - Unico Persone Fisiche
- INPS - Archivio dei Trattamenti Monetari Non Pensionistici
- INPS - Archivio Uniemens

La Banca Dati Reddittuale del Ministero dell'Economia e Finanze costituisce l'archivio principale da cui sono ricavate le informazioni relative alle voci di reddito. L'integrazione di questi archivi consente, da un lato, di recuperare alcune voci di reddito che altrimenti risulterebbero sottostimate (ad es. redditi esenti, una stima della retribuzione dei lavoratori domestici, il reddito da lavoro autonomo dei contribuenti minimi, alcuni trasferimenti pubblici non soggetti a tassazione) e, dall'altro, di riclassificare alcuni importi. Nell'utilizzo delle informazioni reddituali è opportuno tenere presente che le voci di reddito sono al lordo della tassazione e non colgono il sommerso.

Le variabili relative alla iscrizione scolastica ed universitaria derivano rispettivamente da:

- MIUR - Anagrafe degli studenti 2016/2017
- MIUR - Archivio degli iscritti e delle iscrizioni universitarie 2016/2017

La variabile relativa al titolo di studio del capofamiglia deriva dal Censimento della Popolazione del 2011 e dai successivi aggiornamenti derivati da informazioni del MIUR⁶ relativi a:

- MIUR - Esiti, Anagrafe degli studenti, Iscrizioni universitarie, Lauree, Titoli pregressi associati alle Iscrizioni universitarie e Lauree.

Le variabili relative alla partecipazione al mercato del lavoro sono derivate dalla base dati Istat:

- ISTAT - Precarietà lavorativa – ARCH.I.M.E.DE⁷

Deriva dalla stessa base anche l'*intensità lavorativa*, che fa riferimento alla sola occupazione regolare derivata dalle fonti amministrative.

⁵ Preventivamente trattato dalla Direzione Centrale per le Statistiche Sociali e il Censimento della Popolazione, servizio Sistema integrato salute, assistenza, previdenza e giustizia.

⁶ In particolare, viene utilizzata la Base dati integrata su Istruzione e Titoli di studio – BIT prodotta dal Servizio Fonti amministrative e integrazione dei registri della Direzione Centrale per la Raccolta Dati.

⁷ Si tratta, in particolare, di una elaborazione aggiuntiva effettuata sui dati di Precarietà Lavorativa, relativa soprattutto alla tipologia di lavoratori domestici. Per questo motivo, l'identificazione degli individui con segnali di occupazione regolare negli Archivi Amministrativi tra le due basi di dati non è del tutto sovrapponibile.

5.3 Informazioni sulla procedura per la costruzione delle variabili familiari

Il file di dati elementari riferito alle famiglie è derivato per aggregazione dai dati elementari riferiti agli individui. La variabile tipologia familiare adotta una classificazione a sei modalità: famiglia unipersonale; coppie (coniugate e non) senza figli; coppie (coniugate e non) con figli; famiglia monogenitore; altro; non classificabile. Tale classificazione è stata ottenuta implementando un algoritmo che utilizza le informazioni relative alla relazione di parentela e allo stato civile dei singoli membri delle famiglie.

Il *reddito familiare lordo* è ottenuto sommando le relative voci di reddito percepite da tutti i componenti delle famiglie; il *reddito equivalente (lordo)* è calcolato sulla base dei componenti della famiglia anagrafica.

L'intensità lavorativa rappresenta il rapporto fra il totale di mesi lavorati dai componenti della famiglia durante l'anno di riferimento, e il totale di mesi potenzialmente disponibili per attività lavorative. L'intensità lavorativa è una misura annuale e assume valori compresi tra 0 e 1 (rispettivamente: totale assenza di segnali lavorativi nell'anno, e partecipazione continuativa nel mercato del lavoro nel corso dell'anno). La variabile assume valore 0 anche nel caso in cui i componenti siano inattivi (pensionati, bambini, casalinghe). Per questo motivo, al fine di consentire una corretta individuazione della popolazione di riferimento per il calcolo di eventuali indicatori territoriali relativi all'intensità lavorativa, nella base dati è stata inserita la variabile *numero di persone di età compresa tra i 18 e i 59 anni al netto degli studenti tra i 18 e i 24 anni*. L'intensità lavorativa è infine ricodificata in classi.

6 Errori non campionari

Come detto in precedenza, la base dati, è frutto di integrazione di microdati di natura amministrativa su cui non sono stati effettuati processi di calibratura o interventi di correzione. Le procedure di controllo hanno evidenziato le incoerenze presenti, gli errori di misclassificazione, di misurazione, di rappresentazione (sovra - sotto copertura della popolazione di riferimento), brevemente descritti nel seguito. Tali errori tuttavia non compromettono il livello qualitativo generale della base dati statistica.

7 Copertura

Nella base dati si registra una sottocopertura per la popolazione residente⁸ pari allo 0,4 per cento dovuta ai seguenti motivi:

- a) per cinque comuni le LAC non erano disponibili al momento dell'inizio della produzione (in questi comuni risiedono, secondo i dati della Rilevazione Istat sulla "Popolazione residente comunale per sesso, anno di nascita e stato civile", circa 16,7 mila individui). Quattro di questi comuni hanno, comunque, una popolazione inferiore ai 5000 abitanti e non figurano nella base dati pubblicata;
- b) le famiglie in cui almeno un componente era sprovvisto dei codici necessari all'integrazione delle fonti impiegate nel processo sono state escluse dal file di dati.

Si registra, inoltre, una sottocopertura per una parte degli studenti delle scuole primarie e secondarie, in quanto non sono disponibili le informazioni relative ad alcune scuole della regione Valle d'Aosta⁹. Per tale ragione, le variabili relative al numero di studenti e al numero di persone di età compresa tra i 15 e i 29 anni che non studiano e non lavorano non sono utilizzabili per questi territori. Infine, si segnalano situazioni di sottocopertura relativamente agli iscritti a scuole pubbliche gestite da enti diversi dallo Stato, a scuole ospedaliere e ad istituti carcerari.

⁸ "Popolazione residente comunale per sesso, anno di nascita e stato civile".

⁹ Nella collezione di microdati riferita al 2013 non erano disponibili, oltre alle informazioni relative ad alcune scuole della regione Valle d'Aosta, anche le informazioni relative alla provincia autonoma di Bolzano.

8 Chiarezza

Relativamente alle variabili reddituali è opportuno tenere presente che le fonti amministrative non coprono alcune tipologie di reddito:

- a) i redditi da fabbricati e dei terreni essendo derivati dalle dichiarazioni dei redditi (inclusi nel *reddito da capitale reale*) sono sottostimati in quanto alcuni contribuenti sono esentati da tale obbligo.
- b) i redditi disponibili nella base dati, soprattutto i *redditi da capitale*, non includono redditi soggetti ad imposta sostitutiva (ad es. interessi sui BOT o sugli altri titoli del debito pubblico) e redditi soggetti a ritenuta alla fonte a titolo di imposta (ad es. interessi sui conti correnti bancari o postali). Non sono inoltre inclusi alcuni redditi esenti (ad es. somme percepite a titolo di indennità risarcitorie).

Inoltre le voci di reddito sono al lordo della tassazione.

Le variabili indicanti il numero di persone di età compresa tra i 15 e i 29 anni che non studiano e non lavorano fanno riferimento ai soli giovani non inseriti in un percorso scolastico/universitario e non impegnati in un'attività lavorativa. Rispetto ai cosiddetti "NEET", si evidenzia la mancanza di fonti informative su formazione professionale, AFAM (Alta formazione Artistica, Musicale e Coreutica) di I e di II livello, dottorati di ricerca e tirocini.

Per quanto riguarda il *titolo di studio del capofamiglia* è stata rilevata una sottocopertura. In particolare, le informazioni sui titoli di studio conseguiti all'estero dopo il 2011, sui diplomati AFAM e sui qualificati dei percorsi di istruzione e formazione gestiti dalle Regioni non sono disponibili. Inoltre, l'informazione è mancante per parte della popolazione in LAC non censita.

9 Confrontabilità con altre fonti

L'Istat produce dati ufficiali rispetto alle condizioni economiche delle famiglie e alla povertà assoluta e relativa attraverso alcune indagini, tra cui l'indagine sul reddito e le condizioni di vita delle famiglie Eu-Silc. Quest'ultima è un'indagine campionaria che fornisce statistiche a livello trasversale e longitudinale producendo stime fino al livello regionale, mentre la base dati "Condizioni socio-economiche delle famiglie – ARCH.I.M.E.DE" è frutto della sola integrazione di dati amministrativi ed è volta all'analisi territoriale (anche comunale) trasversale.

È opportuno sottolineare che i microdati della base dati "Condizioni socio-economiche delle famiglie – ARCH.I.M.E.DE" non sono confrontabili con quelli diffusi da Eu-Silc. In primo luogo, la definizione di reddito (e la relativa classificazione in macro-voci) adottata da Eu-Silc rappresenta un adattamento al contesto italiano di quella internazionale riportata nel manuale di Canberra¹⁰, mentre i redditi presenti nella base dati presentano dei disallineamenti rispetto alla definizione ufficiale, risentendo quindi di una diversa quantificazione. I principali disallineamenti sono i seguenti:

1) La base dati include soltanto voci di reddito al lordo della tassazione. Eu-Silc, invece, rileva i redditi netti attraverso interviste dirette, successivamente integrati con alcuni dati di fonte amministrativa (Agenzia delle Entrate, Inps)¹¹, mentre imposte e contributi sociali vengono calcolati mediante integrazione con dati amministrativi e stime da modello di micro-simulazione¹². In Eu-Silc la percentuale di famiglie con un reddito equivalente inferiore al 60 per cento del reddito mediano equivalente viene calcolata utilizzando il reddito netto e quindi risente dell'effetto redistributivo dell'imposta; l'analoga percentuale calcolata sui dati

¹⁰ Unece (2011). Canberra Group Handbook on Household Income Statistics.

¹¹ Istat (2009). Integrazione di dati campionari Eu-Silc con dati di fonte amministrativa. Metodi e norme.

¹² Istat (2011). La metodologia di stima dei redditi lordi nell'indagine Eu-Silc. Metodi e norme.

della base “Condizioni socio-economiche delle famiglie” invece risulta più elevata, dato che viene calcolata utilizzando valori di reddito lordi.

2) Si evidenzia una minore copertura dei redditi nelle fonti amministrative (specialmente dei redditi esenti, a tassazione separata o soggetti ad imposta sostitutiva), come già indicato nel paragrafo 8. Inoltre i redditi della base “Condizioni socio-economiche delle famiglie” non includono i redditi da trasferimenti privati (da altre famiglie) ad eccezione degli assegni periodici percepiti da coniugi o ex-coniugi, alcuni trasferimenti non pensionistici, contributi per affitti/mutui/utenze, *fringe benefits* erogati ai lavoratori dipendenti come l’auto aziendale.

3) Il reddito equivalente è calcolato sulla base dei componenti della famiglia anagrafica e non della famiglia di fatto come avviene in Eu-Silc.

Glossario

Famiglia anagrafica: insieme di persone legate da vincoli di matrimonio, parentela, affinità, adozione, tutela o affettivi, coabitanti ed aventi dimora abituale nello stesso comune (anche se non sono ancora iscritte nell’anagrafe della popolazione residente del comune medesimo). Una famiglia può essere costituita anche da una sola persona. L’assente temporaneo non cessa di appartenere alla propria famiglia sia che si trovi presso altro alloggio (o convivenza) dello stesso comune, sia che si trovi in un altro comune.

Lavoratori: individui con segnali di occupazione regolare negli Archivi Amministrativi nel mese di ottobre 2016¹³.

Intensità lavorativa: rappresenta una misura della partecipazione dei membri della famiglia al mercato del lavoro (solo occupazione regolare derivata dalle fonti amministrative) nel corso dell’anno. L’informazione viene rilevata a livello individuale e poi riportata sulla famiglia.

La quota annuale calcolata a livello individuale è computata come media dell’intensità lavorativa mensile. Quest’ultima viene calcolata per tutti i mesi dell’anno e rappresenta una misura della quota del mese lavorata dal singolo individuo (varia tra 0 – nessuna quota del mese lavorata - e 1 – lavoro continuativo nel mese-). La presenza dell’occupato nelle fonti sul lavoro dipendente e parasubordinato è connessa al numero di settimane del mese per le quali si osserva un versamento di contributi previdenziali (a prescindere dall’orario giornaliero per i dipendenti); la presenza nelle fonti sul lavoro autonomo dipende dal periodo mensile di iscrizione dell’occupato in Anagrafe Tributaria e nelle Camere di commercio o dal versamento di contributi nelle specifiche Casse previdenziali di lavoro autonomo (Artigiani e commercianti, Agricoltura). Osservate tutte le posizioni lavorative che un soggetto può sostenere in un determinato periodo, l’intensità lavorativa mensile complessiva è calcolata sommando le intensità di tutte le posizioni (con somma massima unitaria); a livello annuale, l’intensità lavorativa è il rapporto tra la somma delle intensità mensili e il numero di mesi. Ai fini del calcolo della misura a livello familiare, si considerano i membri della famiglia di età compresa fra i 18 e i 59 anni, escludendo gli studenti nella fascia di età tra i 18 e i 24 anni.

Reddito lordo: il reddito complessivo al lordo della tassazione è inteso come somma di diverse componenti di natura economica percepite dai membri della famiglia anagrafica nel corso dell’anno.

Le principali voci incluse nel reddito lordo sono i redditi da lavoro dipendente e autonomo, da pensione, da capitale, da trasferimento pubblico e privato. La soluzione adottata prevede il recupero di voci di reddito dalle fonti amministrative utilizzate indipendentemente dal fatto che esse siano tassate o esenti e dall’eventuale tipologia di tassazione (ordinaria, separata). Alle variabili reddituali derivanti dall’utilizzo

¹³ Per informazioni aggiuntive sull’occupazione si veda la Nota metodologica della base integrata “Precarietà lavorativa – ARCH.I.M.E.DE”.

delle fonti corrisponde una definizione di tipo amministrativo permeata su concetti prevalentemente di tipo fiscale e retributivo, associabile a specifiche sotto-popolazioni di contribuenti e/o percettori di un reddito certificato. Per questo motivo la somma di queste componenti, che definisce la variabile di reddito lordo presenta dei disallineamenti con una definizione ufficiale di reddito (cfr. paragrafi 8 e 9).

Reddito lordo equivalente: per confrontare i redditi delle famiglie è necessario tener conto del fatto che le famiglie più numerose hanno maggiori bisogni. Per rendere equivalenti i redditi di famiglie diverse si usa una scala di equivalenza che tiene conto non solo del numero ma anche dell'età dei componenti della famiglia. La scala di equivalenza OCSE modificata utilizzata per calcolare il reddito familiare equivalente è pari alla somma di più coefficienti individuali (1 per il primo adulto, 0,5 per ogni altro adulto e 0,3 per ogni componente di età inferiore a 14 anni).

Titolo di studio: rappresenta il grado di istruzione degli individui ed è derivato prendendo come base il titolo di studio dichiarato al Censimento della Popolazione del 2011, aggiornato con i titoli di studio conseguiti successivamente al 2011 registrati nelle fonti MIUR. Non sono compresi i titoli di studio conseguiti all'estero dopo il 2011, i diplomi AFAM e le qualifiche dei percorsi di istruzione e formazione gestiti dalle Regioni.

Partecipazione a percorsi formativi scolastici o universitari: presenza dei componenti della famiglia anagrafica nell'Anagrafe degli studenti nell'anno scolastico 2016/2017 o nell'Archivio degli iscritti e delle iscrizioni universitarie nell'anno accademico 2016/2017. Non tutti i percorsi sono tracciabili nei dati amministrativi: sono esclusi gli individui inseriti in percorsi di formazione professionale, AFAM di I e II livello, dottorati di ricerca e tirocini (per ulteriori informazioni sulla copertura si veda il paragrafo 7).