

Nota metodologica¹

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

Anno 2013

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 delle connessioni tra aspetti socio-demografici di una famiglia ed aspetti più economici. Per questa finalità le famiglie residenti in un determinato territorio sono descritte attraverso la loro struttura demografica, il lavoro, lo studio ed i redditi dei componenti. Nella base sono compresenti sia variabili di tipo sociale sia variabili di natura economica, riferite alle famiglie e agli individui che la compongono. Per motivi di riservatezza i dati forniti sono relativi ai soli comuni con più di 5.000 abitanti.

Il riferimento del progetto nel Piano statistico nazionale è Psn: IST- 02663.

2 Archivi utilizzati per la costruzione della base dati integrata

Gli archivi amministrativi elementari entrati nel processo ed i relativi contenuti informativi utilizzati sono riportati nella Tabella 1.

Tabella 1 - Archivi di base e principale utilizzo. Anno di riferimento delle Fonti: 2013 (dove non indicato diversamente)

Fonti di dati	Ente titolare	Contenuti informativi
Lista Anagrafiche Comunali (LAC) allo 01/01/2014	Comuni italiani	Caratteristiche anagrafiche dei componenti della famiglia; dimensione e tipologia familiare
Banca Dati Reddittuale (BDR)	Ministero dell'Economia e delle Finanze (MEF)	Redditi
Casellario dei pensionati	Istituto Nazionale della Previdenza Sociale (INPS)	Redditi
Rapporti di lavoro domestico	INPS	Redditi
Uniemens	INPS	Redditi
Modello Unico Persone Fisiche	Agenzia delle Entrate	Redditi
Trattamenti non pensionistici	INPS	Redditi; percettori di indennità di Disoccupazione/Mobilità
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
Archivio degli iscritti e delle iscrizioni universitarie	MIUR	Livello di Istruzione, partecipazione a percorsi formativi
“Precarietà lavorativa – ARCH.I.M.E.DE”	Istat	Occupazione

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

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 (01/01/2014 per la base dati del 2013). 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. La definizione di famiglia è quella contenuta nel Regolamento Anagrafico (legge 24 Dicembre 1954, n. 1228; DPR 30 Maggio 1989, n.223). Nelle indagini campionarie sulle famiglie condotte dall'Istat² 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 Principali variabili di interesse

Le principali variabili contenute nella base dati descrivono la struttura demografica della famiglia, il lavoro, lo studio e i redditi: sono relative alla tipologia familiare, all'istruzione (dalla scuola primaria fino all'università), alla partecipazione al mercato del lavoro, e al reddito lordo familiare e equivalente. Tali informazioni consentono all'utilizzatore 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'unità 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.

5.1 Informazioni sul campo di osservazione

L'input del processo è costituito dalle LAC, dalle quali vengono selezionati gli individui ritenuti "eleggibili" (ovvero gli individui appartenenti a famiglie in cui nessun componente è sprovvisto di codici identificativi). Il file di dati sulle famiglie è derivato da quello 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 costruzione delle variabili del file individuale intermedio

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

Le variabili reddituali sono il risultato di un'integrazione effettuata a partire da una selezione di variabili presenti negli archivi disponibili in Istat. 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; gli altri

² Istat (2009). Navigando tra le fonti demografiche e sociali.

archivi utilizzati sono: il Casellario dei Pensionati dell'Inps³, l'archivio dei lavoratori domestici dell'Inps, l'Unico Persone Fisiche dell'Agenzia delle Entrate, l'archivio dei Trattamenti Monetari Non Pensionistici dell'Inps e l'archivio Uniemens dell'Inps. L'integrazione di questi archivi consente, da un lato, di recuperare alcune voci di reddito che altrimenti risulterebbero sottostimate (ad es. le pensioni 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 (ad es. trasferimenti pubblici quali disoccupazione e mobilità sono scorporati dal reddito da lavoro dipendente e sommati al reddito da trasferimenti pubblici). 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 *frequenza scolastica ed universitaria* derivano rispettivamente da Anagrafe degli studenti e Archivio degli iscritti e delle iscrizioni universitarie. La variabile relativa al *titolo di studio* del capofamiglia deriva dal Censimento della Popolazione del 2011, e dai successivi aggiornamenti derivati da informazioni del MIUR relativi a: 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 "Precarietà lavorativa – ARCH.I.M.E.DE". L'*intensità lavorativa* fa riferimento alla sola occupazione regolare derivata dalle fonti amministrative.

5.2 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: unipersonale; coppie (coniugate e non) senza figli; coppie (coniugate e non) con figli; 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 "per anno" 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, al suo primo rilascio, 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

³ Preventivamente trattato dalla Direzione Centrale per le Statistiche Sociali e il Censimento della Popolazione, servizio Sistema integrato salute, assistenza, previdenza e giustizia.

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.

I successivi rilasci della base saranno soggetti a processi di correzione e le informazioni sulla qualità saranno accompagnate da misurazioni puntuali.

7 Copertura

Nella base dati si registra una sottocopertura per la popolazione residente⁴ pari a circa l'1%, dovuta ai seguenti motivi:

- a) le famiglie di 147 comuni non sono incluse nella base dati in quanto 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 680 mila individui);
- 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;
- c) le famiglie residenti in comuni con codici territoriali non corretti sono state escluse.

Sui dati anagrafici utilizzati nella base dati non sono stati effettuati interventi di *editing*, pertanto sono presenti mancate cancellazioni anagrafiche, specialmente nelle fasce di età superiori ai 65 anni.

Si registra, infine, una sottocopertura per una parte degli studenti delle scuole primarie e secondarie, in quanto non sono disponibili le informazioni relative ad alcune scuole della provincia autonoma di Bolzano e della regione Valle d'Aosta. 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.

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; tale numero risulta sovrastimato rispetto ai cosiddetti "NEET" a causa della mancanza di fonti informative su formazione professionale, AFAM, 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

⁴ "Popolazione residente comunale per sesso, anno di nascita e stato civile"

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 Canberra⁵, 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)⁶, mentre imposte e contributi sociali vengono calcolati mediante integrazione con dati amministrativi e stime da modello di micro-simulazione⁷. 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 della base "Condizioni socio-economiche delle famiglie" invece risulta più elevata, dato che viene calcolata utilizzando valori di reddito lordi. Si individuano quindi situazioni di povertà più diffuse ma generalmente meno critiche rispetto a Eu-Silc.

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.

⁵ Unece (2011). Canberra Group Handbook on Household Income Statistics.

⁶ Istat (2009). Integrazione di dati campionari Eu-Silc con dati di fonte amministrativa. Metodi e norme.

⁷ Istat (2011). La metodologia di stima dei redditi lordi nell'indagine Eu-Silc. Metodi e norme.