Istat
Istituto Nazionale di Statistica

Indagine sul reddito e le condizioni di vita delle famiglie (EU-SILC) – dati longitudinali

2014-2017

Aspetti metodologici dell'indagine

Maggio 2019

INDICE

1. Introduzione 3
2. La popolazione di riferimento 3
3. Il disegno campionario 3
4. La rilevazione e il trattamento dei dati 5
5. La metodologia di calcolo dei pesi campionari 7
6. La diffusione dei risultati dell'indagine 12
7. Glossario 12
8. Riferimenti bibliografici 13
9. Contatti 13

1. Introduzione

L'indagine sul reddito e le condizioni di vita delle famiglie è inserita all'interno di un più ampio sistema informativo europeo denominato "Statistics on Income and Living Conditions" (EU-SILC) deliberato con Regolamento del Parlamento Europeo, n. 1177/2003 e coordinato da Eurostat. Tale sistema risponde alla sempre più ampia e dettagliata richiesta di informazione statistica su argomenti come redditi, povertà, esclusione sociale.

L'indagine ha quindi come obiettivo prioritario quello di fornire, usando definizioni e metodi armonizzati, dati comparabili tra Paesi, sia a livello trasversale che longitudinale, per l'analisi della distribuzione dei redditi, della diseguaglianza e della povertà, della deprivazione, dell'esclusione sociale e della qualità della vita delle famiglie.

2. La popolazione di riferimento

La popolazione di riferimento è costituita da tutte le famiglie residenti in Italia al momento dell'indagine e dai relativi componenti. Sono escluse le persone che vivono in istituzioni. Per famiglia si intende un insieme di persone che dimorano abitualmente nella stessa abitazione e legate da vincoli di parentela, affinità, adozione, tutela, affetto o amicizia. Tutti i componenti della famiglia sono rilevati ma solo le persone di 16 anni o più sono intervistate.

3. Il disegno campionario

L'indagine è condotta su un campione annuale di circa 22.000 famiglie (per un totale di quasi 50.000 individui), distribuite in circa 650 Comuni italiani di diversa ampiezza demografica e selezionate secondo un disegno campionario che le rende statisticamente rappresentative della popolazione residente in Italia. Sebbene il Regolamento EU-SILC richieda che i principali indicatori siano statisticamente significativi a livello nazionale, in Italia l'indagine è stata disegnata per assicurare stime trasversali affidabili anche a livello regionale.

Il disegno adottato è un campionamento casuale composito a due stadi con stratificazione delle unità di primo stadio: le unità di primo stadio sono i Comuni, quelle di secondo stadio sono le famiglie. La stratificazione dei Comuni per dimensione demografica (all'interno di

ciascuna regione) definisce sia strati formati da un solo Comune auto-rappresentativo (AR), sia strati formati da più Comuni (NAR). Per gli strati AR il campionamento è a un solo stadio con selezione casuale delle famiglie campione dalla lista anagrafica comunale mediante scelta casuale e senza re-immissione. Per gli strati NAR vengono selezionati casualmente 4 Comuni campione (o soltanto 2 negli strati formati soltanto da 2 o 3 Comuni) con probabilità di inclusione proporzionale alla dimensione demografica dei Comuni stessi; le famiglie vengono selezionate all'interno dei Comuni campione come nel caso AR.

Il disegno di indagine deve integrare una componente trasversale e una componente longitudinale. Per l'Italia, il campione relativo a ogni occasione d'indagine è un panel ruotato costituito da quattro gruppi rotazionali (ognuno di dimensione pari a un quarto della numerosità campionaria complessiva teorica). Ogni gruppo rimane nel campione per quattro anni consecutivi e ogni anno il campione si rinnova con l'entrata di un nuovo gruppo che sostituisce il gruppo entrato nel campione quattro anni prima.

SCHEMA ROTAZIONALE DEL CAMPIONE EU-SILC

Gruppo entrante	Anno di indagine										
	T-10	*T-9*	*T-8*	*T-7*	*T-6*	*T-5*	*T-4*	*T-3*	*T-2*	*T-1*	*T*
1	w1										
2	w1	w2									
3	w1	w2	w3								
4	w1	w2	w3	w4							
5		w1	w2	w3	w4						
6			w1	w2	w3	w4					
7				w1	w2	w3	w4				
8					w1	w2	w3	w4			
9						w1	w2	w3	w4		
10							w1	w2	w3	w4	
11								w1	w2	w3	w4
12									w1	w2	w3
13										w1	w2
14											w1

Campione trasversale *T*

Campione longitudinale *T-3 - T*

Tutti gli individui di almeno 14 anni appartenenti alle famiglie estratte ed effettivamente intervistate sono definiti "individui campione" e tale stato rimane identico per tutta la permanenza dell'individuo nel panel; la famiglia campione è definita come quella contenente almeno un individuo campione. Pertanto dalla occasione di indagine (wave)

successiva alla prima, qualora un individuo campione lasci la famiglia di origine e formi una nuova famiglia nell'ambito del territorio nazionale (detta famiglia split), sarà seguito e la sua famiglia entrerà a far parte del campione. Le famiglie di origine e split continueranno ad essere intervistate per tutta la durata del panel, fintanto che contengono almeno un individuo campione. Dalla seconda wave, gli eventuali componenti aggiuntivi delle famiglie campione sono definiti individui coresidenti e continuano a essere intervistati nelle wave successive finché fanno parte di una famiglia campione. Un individuo campione esce definitivamente dal campione in caso di trasferimento in istituzione, di trasferimento all'estero o di decesso, in quanto non più rappresentativo della popolazione obiettivo (individui presenti in famiglie residenti in Italia).

Oltre ai Comuni del campione base, in ogni anno di indagine è stata attivata la rilevazione anche in altri Comuni per effetto dei trasferimenti delle famiglie o di singoli individui campione sul territorio nazionale.

4. La rilevazione e il trattamento dei dati

L'indagine è realizzata mediante interviste sulla situazione familiare e interviste individuali. Dal 2011 la realizzazione delle interviste a domicilio avviene con la tecnica CAPI (Computer Assisted Personal Interview). Dal 2015 è stata introdotta per una parte delle interviste la tecnica CATI (Computer Assisted Telephone Interview).

Le informazioni sono raccolte mediante un questionario elettronico strutturato in tre parti:

a) la scheda generale, in cui vengono inserite le informazioni demografiche di base di tutti i componenti della famiglia (sesso, data e luogo di nascita, cittadinanza, ecc.) e alcune informazioni sui minori di 16 anni di età (frequenza scolastica e affidamento a servizi di cura formali o informali);

b) il questionario familiare, in cui vengono raccolte informazioni su condizioni abitative, spese per l'abitazione, situazione economica, deprivazione materiale, rete di aiuto informale e su alcune tipologie di reddito ricevute a livello familiare;

c) il questionario individuale, per ciascun componente familiare di almeno 16 anni di età, in cui vengono registrate informazioni sul livello di istruzione e formazione, sulle condizioni di salute, sulla condizione professionale (attuale o trascorsa) e sui redditi percepiti nell'anno solare precedente l'intervista (da lavoro dipendente, autonomo e da collaborazione, da trasferimenti pensionistici e non pensionistici, da capitale reale o finanziario, da trasferimenti privati).

Il questionario è progettato in modo da definire automaticamente i percorsi di risposta e di attivare delle avvertenze in presenza di incoerenze riscontrabili nelle risposte fornite; questo permette di sanare alcune incompatibilità direttamente in fase di intervista, senza bisogno di un ritorno sul campo, consentendo così di ridurre ulteriormente l'errore non campionario e di migliorare quindi la qualità dei dati rilevati. Il questionario prevede inoltre dei quesiti a conferma per le interviste successive alla prima al fine di ridurre il carico sul rispondente e di migliorare la qualità delle informazioni rilevate negli anni precedenti.
I dati rilevati sono sottoposti ad un processo di controllo, correzione ed imputazione dei valori mancanti, al fine di risolvere alcune incompatibilità sostanziali e di riconciliare le informazioni raccolte tramite l'intervista con altre provenienti da archivi amministrativi.

In particolare, le fasi del processo possono essere cosi sintetizzate (per approfondimenti Istat, 2008):

- controllo delle chiavi identificative delle famiglie e degli individui, necessarie anche per la definizione del campione da intervistare nuovamente l'anno successivo;
- correzione delle variabili anagrafiche;
- correzione della struttura delle famiglie di fatto e loro classificazione nelle differenti tipologie;
- imputazione da donatore dei questionari individuali mancanti, tramite una procedura basata sul donatore di distanza minima;
- correzione degli errori sistematici con tecniche di tipo deterministico;
- integrazione dei dati campionari con dati di fonte amministrativa (per approfondimenti Istat, 2009), con l'obiettivo di ridurre al minimo l'errore di stima del reddito;
- correzione e imputazione degli errori stocastici delle variabili qualitative, tramite procedure automatizzate di controllo e correzione di tipo probabilistico, che applicano la metodologia di Fellegi-Holt, basata sul principio del minimo cambiamento e implementata nel software Concord realizzato dall'Istat;
- individuazione ed eliminazione degli outliers per le variabili quantitative, tramite una procedura generalizzata basata sul metodo di Hidiroglou-Berthelot;
- imputazione delle variabili quantitative (laddove mancanti o ritenute non valide perché incoerenti o anomale), tramite una tecnica basata sulla costruzione di modelli di regressioni sequenziali multiple implementata nel software IVEware;
- costruzione delle variabili sui redditi lordi mediante l'utilizzo di modelli di microsimulazione delle imposte e dei contributi sociali (per approfondimenti Istat, 2011);

- costruzione delle variabili finali definite secondo le linee guida Eurostat;
- controllo e correzione delle variabili demografiche e di alcune variabili di istruzione e lavoro per garantire la coerenza dei percorsi longitudinali.

5. La metodologia di calcolo dei pesi campionari

Dato il disegno campionario integrato, finalizzato alla produzione di stime trasversali e longitudinali, viene prodotto un sistema di pesi che consente di usare in modo flessibile i microdati rilasciati, secondo gli obiettivi conoscitivi.

5.1 Pesi trasversali

Il calcolo dei pesi per la componente trasversale è una procedura piuttosto articolata, che può essere sintetizzata nei seguenti passi:

- Correzione dei pesi diretti (da disegno) per mancata risposta.

  La probabilità di risposta è stimata con modelli logistici, distinti per tecnica di rilevazione (CATI/CAPI) e occasione di indagine (wave) a cui si trova l'individuo. Per la prima wave le covariate sono: ripartizione geografica; tipologia di Comune; numero componenti; sesso; classe di età; cittadinanza dell'intestatario della scheda familiare. Nelle wave successive le covariate includono anche informazioni rilevate nell'indagine precedente: ripartizione geografica; tipologia di Comune; numero componenti; tipologia familiare; fonte principale di reddito della famiglia; titolo di godimento dell'abitazione; quinto di reddito equivalente; numero di sintomi di deprivazione; rischio di povertà, grave deprivazione, intensità lavorativa molto bassa, sovraffollamento dell'abitazione, e grave deprivazione abitativa; presenza di reddito da lavoro dipendente, da lavoro autonomo, da pensione; sesso, classe di età, livello di istruzione; condizione professionale; cittadinanza.
- 1° trimming dei pesi corretti.
- 1° passo di calibrazione su alcuni totali noti provenienti dalla rilevazione sulle Forze di Lavoro (relativi a livello di istruzione e condizione ILO per sesso e classi di età).
- 2° trimming dei pesi corretti.
- 2° passo di calibrazione su alcuni totali noti di popolazione:
  - Per l'intero campione si impongono i seguenti vincoli: i) Popolazione per ripartizione geografica, sesso e 14 classi quinquennali di età al 31 dicembre dell'anno precedente la rilevazione; ii) Popolazione per regione, sesso e 5

classi di età al 31 dicembre dell'anno precedente la rilevazione; iii) Popolazione di cittadinanza italiana per ripartizione geografica al 31 dicembre dell'anno precedente la rilevazione; iv) Popolazione per ripartizione geografica e tipologia di Comune al 31 dicembre dell'anno precedente la rilevazione; v) Numero di famiglie per regione al periodo di rilevazione.

- Per il campione entrante (wave 1) si impongono i seguenti vincoli: i) Popolazione per ripartizione geografica, sesso e 5 classi di età al 31 dicembre dell'anno precedente la rilevazione; ii) Popolazione di cittadinanza italiana per ripartizione geografica al 31 dicembre dell'anno precedente la rilevazione; iii) Popolazione per ripartizione geografica e tipologia di Comune al 31 dicembre dell'anno precedente la rilevazione; iv) Numero di famiglie per ripartizione geografica al periodo di rilevazione.
- Per gli altri gruppi rotazionali (wave 2, 3 e 4) si impongono i seguenti vincoli: i) Popolazione per ripartizione geografica al 31 dicembre dell'anno precedente la rilevazione; ii) Numero di famiglie per ripartizione geografica al periodo di rilevazione.

5.2 Pesi longitudinali

In primo luogo, per ogni panel e per ogni wave viene stimato un peso (RB060 in R-file) che consente di rappresentare la popolazione longitudinale alla prima wave (coincidente con quella trasversale del medesimo anno di rilevazione) e sopravvivente alle wave successive (Figura 1).

Figura 1: Somma dei pesi longitudinali (RB060) per panel e wave

	2014	2015	2016	2017
P2			60.500.229	60.200.753
P3		60.843.061	60.532.761	60.233.124
P4	60.623.575	60.370.775	60.062.884	59.765.572

Si procede in modo diverso per la prima wave e le successive:

- Wave 1
  - I pesi longitudinali della prima wave coincidono con i pesi trasversali finali stimati per lo stesso anno di rilevazione, moltiplicati per un fattore di espansione che assicura il riporto al totale di popolazione del sottocampione longitudinale.
- Wave 2, 3 e 4
  - Correzione per mancata risposta – Per ottenere i pesi delle wave successive, si parte dai pesi della prima wave e si correggono per la mancata risposta (attrition), tenendo conto anche della probabilità di sopravvivenza delle unità campionarie (ovvero della probabilità di continuare a far parte della popolazione obiettivo di studio e quindi di rimanere in vita, non emigrare all'estero, non trasferirsi in istituzione) dopo 1, 2 o 3 anni, a seconda della wave. Entrambe le probabilità vengono stimate con modelli logistici e covariate relative al primo anno di osservazione (sesso, classi di età, ripartizione geografica, tipologia di Comune, livello di istruzione, condizione professionale autodichiarata, fonte principale di reddito, titolo di godimento dell'abitazione, numero di sintomi di deprivazione, quinto di reddito equivalente, numero di componenti della famiglia).
  - Pre-calibrazione – Si approssima la popolazione longitudinale alla wave 2, 3 o 4, aggiungendo le nascite e sottraendo i decessi e i trasferimenti all'estero alla popolazione residente alla prima wave. I pesi corretti per mancata risposta vengono quindi espansi affinché la loro somma riproduca questo totale "noto".
  - Calibrazione dei pesi – Vengono considerate, per ciascuna coppia di anni, le transizioni della condizione professionale oggettiva ILO risultanti dalle stime sul campione longitudinale della rilevazione sulle Forze di Lavoro (in particolare si fa riferimento alle transizioni tra le seguenti condizioni occupazionali: occupato, disoccupato, altro inattivo). Il peso longitudinale finale è ottenuto effettuando la calibrazione dei pesi corretti per mancata risposta e pre-calibrati su questa distribuzione. Contemporaneamente si impone ai pesi di preservare alcune caratteristiche della popolazione longitudinale, come risultanti dalla stima pre-calibrazione: ripartizione geografica, sesso, classe di età (0-16, 16-24, 25-44, 45-65, 65+) e i principali indicatori (rischio di povertà, grave deprivazione materiale, intensità lavorativa molto bassa, rischio di povertà ed esclusione sociale). Questa strategia è adottata al fine di non alterare la struttura della

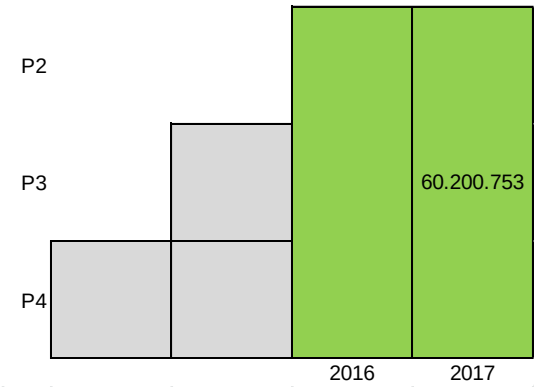
popolazione longitudinale, secondo le principali caratteristiche, a seguito dell'imposizione dei vincoli sulle transizioni delle condizioni occupazionali.

I modelli di correzione della mancata risposta e di sopravvivenza vengono stimati a livello individuale, tenendo conto delle caratteristiche di ciascuna unità campionaria. Per questo motivo, a partire dalla seconda wave, individui appartenenti alla stessa famiglia possono avere pesi diversi. Quindi, non vengono prodotti pesi longitudinali familiari.

Al fine di sfruttare in modo flessibile la struttura del disegno campionario, vengono rilasciati, solo per l'ultimo anno di rilevazione, altri pesi che consentono di ottenere stime sulle transizioni osservate nei panel intervistati per due, tre o quattro anni:

- RB062: può essere usato per la stima delle transizioni osservate negli ultimi due anni di rilevazione, utilizzando l'insieme delle unità campionarie appartenenti ai tre panel intervistati in questa coppia di anni. Questi pesi sono stimati a partire dai pesi RB060 dell'ultimo anno di rilevazione, riproporzionati in modo che le unità appartenenti ai tre panel intervistati nel 2016 e nel 2017, e compresenti nei due anni di rilevazione, siano rappresentative di tutta la popolazione longitudinale estratta nel 2016 e sopravvissuta fino al 2017 (Figura 2).

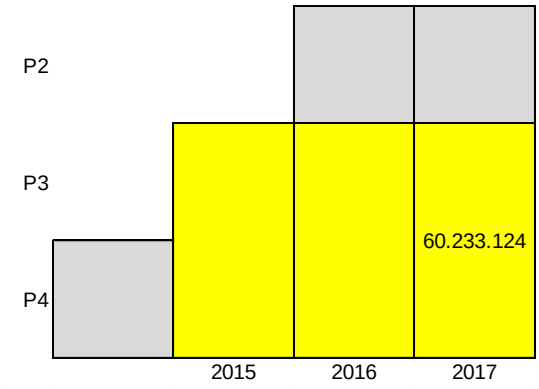
  Figura 2: Somma dei pesi longitudinali per le stime sui compresenti negli ultimi due anni di osservazione (RB062)



- RB063: può essere usato per la stima delle transizioni osservate negli ultimi tre anni di rilevazione, utilizzando l'insieme delle unità campionarie appartenenti ai due panel intervistati tra il 2015 e il 2017. Tali pesi vengono stimati a partire dai pesi RB060 dell'ultimo anno di rilevazione, riproporzionati in modo tale che le sole unità appartenenti ai due panel intervistati in questi anni, e compresenti nei tre anni di

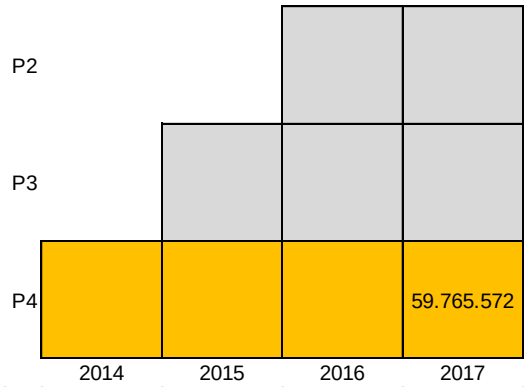
rilevazione, siano rappresentative di tutta la popolazione longitudinale estratta nel 2015 e sopravvissuta fino al 2017 (Figura 3).

Figura 3: Somma dei pesi longitudinali per le stime sui compresenti negli ultimi tre anni di osservazione (RB063)



- RB064: può essere usato per la stima di transizioni e indicatori (quali la persistenza a rischio di povertà) relativi alla popolazione osservata per quattro anni. Tale peso ha valore non nullo solo per gli individui presenti in tutte e quattro le occasioni di indagine (dal 2014 al 2017), viene stimato a partire dai pesi RB060 dell'ultimo anno di rilevazione, riproporzionati in modo che le sole unità compresenti nei quattro anni di rilevazione siano rappresentative di tutta la popolazione longitudinale estratta nel 2014 e sopravvissuta fino al 2017 (Figura 4).

Figura 4: Somma dei pesi longitudinali per le stime sui compresenti in quattro anni di osservazione (RB064)



Per ulteriori dettagli, si veda l'allegato "Weighting procedure" al Quality Report.

6. La diffusione dei risultati dell'indagine

Le stime dell'indagine sono consultabili nel datawarehouse I.Stat all'indirizzo http://dati.istat.it (tema: "Condizioni economiche delle famiglie e disuguaglianze", argomenti "Reddito", "Povertà", "Condizioni abitative").

Annualmente è diffusa una Comunicazione per i media della collana Statistiche Report con gli indicatori più significativi: Condizioni di vita, reddito e carico fiscale - Anno 2017.

I microdati dell'indagine relativi alle edizioni dal 2004 al 2015 sono stati rilasciati sotto forma di File Standard. Il materiale disponibile e le modalità per richiederlo sono consultabili all'indirizzo https://www.istat.it/it/archivio/4152 per la componente trasversale e all'indirizzo https://www.istat.it/it/archivio/4204 per la componente longitudinale.

In adempimento al Regolamento europeo n. 1177/2003, i dati dell'indagine sono trasmessi annualmente a Eurostat. I principali indicatori, archiviati nel database di Eurostat, sono consultabili al link: http://ec.europa.eu/eurostat/data/database (Tema "*Population and social conditions*", argomento "*Income and living conditions*").

Dati riepilogativi su reddito e condizioni economiche delle famiglie sono inoltre diffusi nel volume Noi Italia, nel Rapporto Bes: il benessere equo e sostenibile in Italia e nel sistema informativo sugli obiettivi di sviluppo sostenibile delle Nazioni Unite.

Il Quality Report nazionale è disponibile all'indirizzo tramite il sito di Eurostat http://ec.europa.eu/eurostat/web/income-and-living-conditions/quality/eu-and-national-quality-reports.

Per ulteriori informazioni è possibile accedere al sito dell'Istat all'indirizzo: http://www.istat.it/it/archivio/5663.

7. Glossario

Per le definizioni adottate, vedere i seguenti documenti allegati:

"Methodological Guidelines And Description Of EU-SILC Target Variables_2017 operation.pdf"

"Silc Disclosure Control Rules_Year 2017_Longitudinal Data.pdf"

8. Riferimenti bibliografici

Istat (2008), *L'indagine europea sui redditi e le condizioni di vita delle famiglie (EU-SILC)*, Metodi e Norme n. 37, 2008.

Istat (2009), *Integrazione dei dati campionari EU-SILC con dati di fonte amministrativa*, Metodi e Norme n. 38, 2009.

Istat (2011), *La metodologia di stima dei redditi lordi nell'indagine EU-SILC – Indagine europea sui redditi e le condizioni di vita delle famiglie*, Metodi e Norme n. 49, 2011.

Istat (2015), *Indagine sul reddito e le condizioni di vita (EU-SILC),* in "Navigando tra le fonti sociali", http://schedefontidati.istat.it/index.php/Indagine_sul_reddito_e_le_condizioni_di_vita_(EU-SILC) .

9. Contatti

Francesca Lariccia

francesca.lariccia@istat.it

Daniela Lo Castro

daniela.locastro@istat.it

Curatori dei capitoli

Francesca Lariccia

Daniela Lo Castro