

Nota metodologica

1 Disegno di campionamento

1.1 Gli obiettivi conoscitivi

Da un punto di vista generale, la popolazione oggetto di indagine è costituita dall'insieme delle famiglie residenti in Italia e dagli individui che le compongono, al netto dei membri permanenti delle convivenze. Dal momento che Eu-Silc è un'indagine ripetuta a cadenza annuale con la finalità di stimare sia parametri di natura trasversale sia parametri di natura longitudinale, la popolazione d'interesse deve essere ricondotta nei due concetti di popolazione trasversale per i parametri di natura trasversale e longitudinale per i parametri di natura longitudinale.

La popolazione trasversale è costituita dalle unità esistenti ad una prefissata data di ciascun anno di rilevazione. Per quanto riguarda la popolazione longitudinale, invece, a fronte di una definizione teorica di popolazione dinamica che include tutte le unità che entrano ed escono in un intervallo di tempo e che tuttavia non è pienamente osservabile, esistono in letteratura tre definizioni operative a cui corrispondono popolazioni osservabili in un intervallo di tempo; esse sono la popolazione unione, popolazione intersezione e popolazione a un punto prefissato nell'intervallo. Tali aspetti, tuttavia, verranno approfonditi nei volumi dei prossimi anni, quando saranno diffuse per la prima volta le stime longitudinali.

L'indagine fornisce sia dati trasversali su reddito, povertà e esclusione sociale e condizioni di vita, sia dati longitudinali su reddito, lavoro e su alcuni indicatori non-monetari di esclusione sociale.

In particolare la progettazione del disegno d'indagine è stata guidata, per rispondere ai requisiti Eurostat, principalmente dai seguenti parametri:

- ad ogni occasione di indagine t :
 1. la percentuale di individui poveri, ossia la percentuale di individui il cui reddito netto familiare equivalente¹ è inferiore al 60 per cento del valore mediano della distribuzione;
 2. il reddito medio annuo familiare netto
- tra due occasioni di indagine consecutive t e $t+1$:
 1. la variazione netta della percentuale di famiglie povere;
 2. la variazione netta del reddito familiare medio;
 3. la percentuale di famiglie che permangono nello stato di povertà per due o più occasioni di indagine.

Tra gli obiettivi di interesse per l'Istat ricordiamo le variazioni lorde tra due occasioni di indagine relativamente allo stato occupazionale (ad esempio il numero di individui che passano dallo stato di occupato allo stato di disoccupato) e alla classe di reddito.

I domini territoriali di studio, ossia gli ambiti rispetto ai quali sono riferiti i parametri di popolazione oggetto di stima, sono:

- l'intero territorio nazionale;
- le cinque ripartizioni geografiche (Italia nord-occidentale, nord-orientale, centrale, meridionale, insulare);
- le regioni geografiche (a eccezione del Trentino-Alto Adige le cui stime sono prodotte separatamente per le province di Bolzano e Trento);

1.2 Il disegno di rilevazione

La scelta del disegno di campionamento è stata determinata principalmente dalla necessità di ottenere stime corrette ed efficienti dei parametri trasversali e longitudinali ricordati nel paragrafo precedente, tenendo conto di alcuni requisiti indicati in sede europea. Il sistema di indagini prevede infatti sia una componente trasversale (per fornire stime di livello e variazioni nette) sia una componente longitudinale (per la valutazione delle variazioni lorde o a livello micro); tuttavia, la componente trasversale è maggiormente rilevante in quanto la priorità di Eu-Silc è fornire stime a livello trasversale, tempestive e di alta qualità, comparabili tra i diversi paesi della Unione europea.

¹ Il reddito equivalente è ottenuto utilizzando la scala di equivalenza Ocse modificata (si veda il Glossario).

I requisiti Eurostat che hanno condizionato la definizione del disegno di rilevazione sono i seguenti: l'indagine deve avere cadenza annuale e la durata di ogni panel deve essere di almeno 4 anni (per la natura dei parametri longitudinali sulle permanenze).

Per quanto riguarda l'articolazione delle componenti trasversale e longitudinale era possibile ipotizzare due rilevazioni indipendenti oppure un'unica rilevazione integrata. Si è scelto di adottare un'indagine integrata, per evitare duplicazioni di informazioni.

L'indagine integrata è basata su un campione di tipo panel ruotato, che segue lo schema riportato nel prospetto 1, in cui il campione relativo a ogni occasione d'indagine è costituito da quattro gruppi di rotazione, ciascuno dei quali rimane nel campione per quattro anni consecutivi. Ad esempio, con riferimento all'occasione t+3, le unità del gruppo di rotazione D4 vengono intervistate per la quarta volta, le unità di E3 per la terza, le unità di F2 per la seconda e le unità di G1 entrano nel campione nell'anno t+3 per la prima volta.

Prospetto 1. Schema di rotazione

	A	B	C	D	E	F	G	H	I
T	A4	B3	C2	D1					
T+1		B4	C3	D2	E1				
T+2			C4	D3	E2	F1			
t+3				D4	E3	F2	G1		
t+4					E4	F3	G2	H1	
t+5						F4	G3	H2	I1

1.3 Definizione della numerosità campionaria

Per la componente trasversale, Eurostat ha fissato la numerosità minima in termini di famiglie dei campioni nazionali in base alla precisione della stima della percentuale di famiglie povere a livello nazionale. Sotto l'ipotesi di un campione casuale semplice di unità finali, la numerosità campionaria minima per il campione trasversale di ogni anno risulta pari a 5 mila famiglie per ogni paese. Tenuto conto del peso demografico dell'Italia, la numerosità campionaria minima assegnata all'indagine italiana è risultata pari a 7.250 famiglie. A tale numerosità corrisponde, per una stima di riferimento del tasso di povertà pari a 15 per cento, un errore relativo del 2.8 per cento, ottenuto sulla base di un campione casuale semplice.

A partire da tale numerosità riferita al campione casuale semplice, è stata definita la numerosità campionaria effettiva sulla base dell'effetto del disegno campionario prescelto, che tiene in considerazione l'impatto sull'efficienza delle stime prodotto dalla stratificazione e dal clustering.

Infine, la numerosità del campione da selezionare è stata definita tenendo conto dei tassi di risposta attesi.

1.4 Descrizione dello schema di selezione delle unità

Lo schema di selezione è di tipo complesso e si avvale di due differenti piani di campionamento. Nell'ambito di ognuno dei domini definiti dalla regione geografica i comuni sono suddivisi in due sottoinsiemi sulla base della popolazione residente:

- l'insieme dei comuni Auto Rappresentativi (che indicheremo d'ora in avanti come comuni Ar) costituito dai comuni di maggiore dimensione demografica;
- l'insieme dei comuni Non Auto Rappresentativi (o Nar) costituito dai rimanenti comuni.

Nell'ambito dell'insieme dei comuni Ar, ciascun comune viene considerato come uno strato a se stante e viene adottato un disegno noto con il nome di campionamento a grappoli. Le unità primarie di campionamento sono rappresentate dalle famiglie anagrafiche, estratte in modo sistematico dall'anagrafe del comune stesso; per ogni famiglia anagrafica inclusa nel campione vengono rilevate le caratteristiche oggetto di indagine di tutti i componenti di fatto.

Nell'ambito dei comuni Nar viene adottato un disegno a due stadi con stratificazione delle unità primarie. Le Unità primarie (Up) sono i comuni, le Unità secondarie sono le famiglie anagrafiche; per ogni famiglia anagrafica inclusa nel campione vengono rilevate le caratteristiche oggetto di indagine di tutti i componenti di fatto.

I comuni vengono selezionati con probabilità proporzionali alla loro dimensione demografica e senza reimmissione, mentre le famiglie vengono estratte con probabilità uguali e senza reimmissione.

Stratificazione e selezione delle unità campionarie

L'obiettivo della stratificazione è quello di formare gruppi (o strati) di unità caratterizzate, relativamente alle variabili oggetto d'indagine, da massima omogeneità interna agli strati e massima eterogeneità fra gli strati. Il raggiungimento di tale obiettivo si traduce in termini statistici in un guadagno nella precisione delle stime, ossia in una riduzione dell'errore campionario a parità di numerosità campionaria.

Nell'indagine in esame, i comuni vengono stratificati in base alla loro dimensione demografica e nel rispetto delle seguenti condizioni:

- autoponderazione del campione a livello regionale;
- selezione di $\bar{n} = 4$ comuni campione nell'ambito di ciascuno strato definito sui comuni dell'insieme Nar ;
- scelta di un numero minimo di famiglie da intervistare in ciascun comune campione; tale numero è stato posto pari a 24;
- formazione di strati aventi ampiezza approssimativamente costante in termini di popolazione residente.

Il procedimento di stratificazione, attuato all'interno di ogni regione geografica, si articola nelle seguenti fasi:

- ordinamento dei comuni del dominio in ordine decrescente secondo la loro dimensione demografica in termini di popolazione residente;
- determinazione di una soglia di popolazione per la definizione dei comuni Ar , mediante la relazione:

$${}_r\lambda = \frac{{}_r\bar{m} \cdot {}_r\delta}{{}_r f}$$

in cui per la generica regione geografica r si è indicato con: ${}_r\bar{m}$ il numero minimo di famiglie da intervistare in ciascun comune campione; ${}_r\delta$ il numero medio di componenti per famiglia; ${}_r f$ la frazione di campionamento;

- suddivisione di tutti i comuni nei due sottoinsiemi Ar e Nar : i comuni di dimensione superiore o uguale a ${}_r\lambda$ sono definiti come comuni Ar e i rimanenti come Nar ;
- suddivisione dei comuni dell'insieme Nar in strati aventi dimensione, in termini di popolazione residente, approssimativamente costante e all'incirca pari a $\bar{n} \cdot {}_r\lambda$.

Effettuata la stratificazione, i comuni Ar sono inclusi con certezza nel campione; per quanto riguarda, invece, i comuni Nar , nell'ambito di ogni strato viene estratto un comune campione con probabilità proporzionale alla dimensione demografica, mediante la procedura di selezione sistematica proposta da Madow².

La selezione delle famiglie da intervistare in ogni comune campione viene effettuata dalla lista anagrafica di ciascun comune senza reimmissione e con probabilità uguali.

In particolare, la tecnica di selezione è di tipo sistematico e, nell'ambito di ogni comune viene attuata attraverso le seguenti fasi:

- vengono messi in sequenza i fogli delle famiglie dell'anagrafe del comune;
- si calcola il passo di campionamento e_{hi} , come rapporto tra il numero delle famiglie residenti nel comune i dello strato h e il corrispondente numero di famiglie campione, $e_{hi} = M_{hi}/m_{hi}$;
- si selezionano le m_{hi} famiglie che nella sequenza costruita al punto 1) occupano le seguenti posizioni:

$$1, 1+e_{hi}, 1+2e_{hi}, \dots, 1+(m_{hi}-1)e_{hi}.$$

Per quanto concerne, infine, l'articolazione del campione di famiglie nei quattro gruppi di rotazione (si veda Prospetto 1) si è proceduto come segue. Il campione di ciascun comune Ar è stato suddiviso in quattro sottocampioni di pari dimensione, ognuno assegnato casualmente a un gruppo di rotazione. Nella parte Nar , invece, l'assegnazione ai gruppi ha riguardato l'intero campione di famiglie di ogni comune campione; pertanto, ciascun comune estratto in ogni strato Nar è associato casualmente a un unico gruppo di rotazione.

2 La costruzione delle stime trasversali

L'indagine, nella sua componente trasversale, deve produrre sia le stime riferite al numero di individui (o famiglie) che nella popolazione di riferimento possiedono una certa caratteristica, sia il livello di una quantità misurata sugli individui (o famiglie), come ad esempio il reddito.

² Madow, W.G. "On the theory of systematic sampling II", *Annals of Mathematical Statistics*, 20, (1949): 333-354.

Per il calcolo dei coefficienti di riporto all'universo si utilizza una procedura generalizzata di stima che si basa sull'uso di una famiglia di stimatori, noti in letteratura come *calibration estimator* (stimatori di ponderazione vincolata). La metodologia alla base di tali stimatori consente la determinazione di un unico coefficiente di riporto all'universo in grado di produrre stime coerenti a totali noti, desunti da fonti esterne, sia per individui sia per famiglia assegnando, cioè, lo stesso coefficiente di riporto all'universo a tutti gli individui della stessa famiglia.

La strategia adottata per la costruzione dei coefficienti di riporto all'universo si sviluppa attraverso le fasi tipiche utilizzate per la costruzione degli stimatori nelle varie indagini campionarie dell'Istituto. In particolare possiamo distinguere:

1. la determinazione della probabilità di inclusione di ogni unità statistica e del relativo peso base;
2. calcolo dei coefficienti di correzione per mancata risposta totale;
3. determinazione dei coefficienti di riporto all'universo finali vincolati ai totali noti desunti da fonti esterne all'indagine.

2.1 La probabilità di inclusione e il peso base

Il principio su cui è basato ogni metodo di stima campionaria è che le unità appartenenti al campione rappresentino anche le unità della popolazione che non sono incluse nel campione stesso. A tale scopo, ad ogni unità campionaria viene attribuito un peso finale, o coefficiente di riporto all'universo, che indica quante unità della popolazione sono rappresentate, rispettivamente, da ogni unità presente nel campione.

Senza perdere di generalità e per chiarire gli argomenti trattati, definiamo la seguente simbologia:

U popolazione di riferimento oggetto di indagine;

y_k valore della variabile Y assunto dalla k -esima osservazione della popolazione;

π_k probabilità, assegnata da un generico disegno di campionamento, che l'unità k -esima ha di essere inclusa nel campione s ;

Il totale di una generica variabile Y , calcolato sull'intera popolazione, assume la seguente forma:

$$Y = \sum_{k \in U} y_k \quad (1)$$

In linea teorica il disegno di campionamento assegna le probabilità di inclusione ad ogni unità del campione in modo tale che

$$\hat{Y} = \sum_{j \in s} y_j \frac{1}{\pi_j} \quad (2)$$

sia uno stimatore corretto della (1).

Nel caso del campione di Eu-Silc, come nel caso dei disegni di campionamento tradizionalmente utilizzati dall'Istat per le principali indagini su famiglie e individui, l'inverso della probabilità di inclusione assume la seguente forma:

$$\frac{1}{\pi_{hi}} = n_h \frac{P_h}{P_{hi}} \frac{M_{hi}}{m_{hi}} \quad (3)$$

dove :

h denota l'indice di strato;

i è l'indice di comune;

n_h indica il numero di comuni campione dello strato h ;

P_h indica il totale della popolazione residente nello strato h ;

P_{hi} il totale della popolazione residente nel comune i dello strato h ;

M_{hi} indica il totale di famiglie residenti nel comune i dello strato h ;

m_{hi} indica il numero di famiglie campione nel comune i dello strato h .

In tal senso, ogni famiglia estratta da un certo comune campione si vede assegnata dal disegno di campionamento la medesima probabilità di inclusione³.

Pertanto la (2) può essere scritta come:

$$\hat{Y} = \sum_{h=1}^H \sum_{i=1}^{m_h} y_{hi} \frac{1}{\pi_{hi}} \quad (4)$$

2.2 La correzione per mancata risposta

Nel corso della fase di raccolta delle informazioni presso le famiglie che formano il campione, come accade per ogni indagine statistica, alcune di queste si trovano nell'impossibilità di partecipare all'indagine. Quando il meccanismo che genera la mancata partecipazione, e quindi la mancata risposta, è *ignorabile* si assume l'ipotesi che il comportamento dei rispondenti sia del tutto simile a quello dei non rispondenti con conseguente riduzione casuale della numerosità del campione teorico di partenza e riduzione della precisione delle stime prodotte. In tale circostanza, il correttore per mancata risposta assume la forma dell'inverso della probabilità di risposta (δ_{hi}):

$$\frac{1}{\delta_{hi}} = \frac{m_{hi}}{m_{hi}^r} \quad (5)$$

in cui m_{hi}^r rappresenta il numero di famiglie rispondenti nel comune i dello strato h mentre il coefficiente di riporto all'universo corretto per mancata risposta risulta essere:

$$k_{hi} = \frac{1}{\pi_{hi}} \frac{1}{\delta_{hi}} = n_h \frac{P_h}{P_{hi}} \frac{M_{hi}}{m_{hi}} \frac{m_{hi}}{m_{hi}^r} = n_h \frac{P_h}{P_{hi}} \frac{M_{hi}}{m_{hi}^r} \quad (6)$$

Quando il meccanismo che genera la mancata risposta è *non ignorabile*, come nella maggior parte delle situazioni reali, l'ipotesi di uguaglianza di comportamento tra chi partecipa e chi non partecipa all'indagine viene meno ed è quindi necessario ricorrere ad altre ipotesi e strumenti in modo da poter tener conto dei reali meccanismi che generano la mancata risposta.

Per l'indagine Eu-Silc si è pensato di adottare una strategia che ripercorre i criteri delle *celle di ponderazione* al fine di individuare delle sottopopolazioni dove è ipotizzabile una uguaglianza di comportamento di risposta tra coloro che hanno partecipato all'indagine e coloro che non hanno partecipato. La metodologia sottesa alle *celle di ponderazione* rientra nel novero della modellizzazione esplicita per la riduzione della distorsione della mancata risposta totale. In particolare, si determina la probabilità di risposta in funzione delle variabili che determinano le celle.

In Eu-Silc, utilizzando le informazioni da fonte anagrafica disponibili già in fase di estrazione del campione, è stata operata una suddivisione in celle omogenee secondo le variabili per le quali si è avuto un tasso di risposta differenziato. In particolare, per ognuno dei quattro campioni che compongono il campione complessivo trasversale, sono state utilizzate le seguenti informazioni:

- dimensione demografica del comune in classi (D , 5 modalità);
- cittadinanza della persona di riferimento (T , 2 modalità);
- regione di residenza (L , 21 modalità);
- distribuzione delle famiglie per numero di componenti (C , 4 modalità).

³ Nel caso di Eu-Silc, in ogni campione longitudinale, il numero di comuni campione in ogni strato è pari a 1. Inoltre, l'espressione (3) diventa $1/\pi_{hi} = M_h/m_{hi}$ nel caso dei comuni Ar che costituiscono uno strato a se.

Tabella 2.1 - Modalità e riferimento temporale delle variabili utilizzate per la costruzione delle celle di ponderazione

Variabile	Modalità	Fonte e riferimento temporale
Regione	21 modalità: 19 Regioni (escluso il Trentino Alto Adige) più le Province Autonome di Trento e Bolzano (Eurostat Nuts II level)	Informazioni desunte dalle anagrafi comunali Maggio 2004
Famiglie per numero componenti	- uno - due - tre - quattro e +	
Dimensione demografica del comune di residenza	- Meno di 10.000 abitanti - Da 10.001 a 50.000 abitanti - Da 50.001 a 250.000 abitanti - Da 250.001 a 500.000 abitanti - Più di 500.000 abitanti	
Cittadinanza della persona di riferimento	- Italiana - Straniera	

La tabella 2.1 riporta il dettaglio delle informazioni utilizzate per la realizzazione delle celle di ponderazione.

Sia Z la partizione delle famiglie campione composta dagli elementi (celle) generati dal prodotto cartesiano ($D \times T \times L \times C$) delle modalità delle variabili considerate. Per ogni cella z della partizione Z è stata calcolata la probabilità di risposta (η) come rapporto tra il numero di famiglie rispondenti e il numero di famiglie desunte dal campione teorico.

In simboli:

$$\eta_z = \frac{m_z^r}{m_z} \quad (7)$$

L'inverso di η è il correttore per mancata risposta e ad ogni famiglia del comune i dello strato h , appartenente alla cella z viene moltiplicato il proprio peso base per tale correttore, pertanto il coefficiente di riporto all'universo corretto per mancata risposta totale diviene:

$${}_{hi} k_z^* = \frac{1}{\pi_{hi}} \frac{1}{\eta_z} = n_h \frac{P_h}{P_{hi}} \frac{M_{hi}}{m_{hi}} \frac{m_z}{m_z^r} \quad (8)$$

Nel caso in cui in una cella z^* si presenta una scarsa numerosità (≤ 13) di famiglie rispondenti e/o un elevato valore del correttore ($1/\eta > 2,4$)⁴ allora si procede al *collassamento* delle celle nel modo seguente:

- si determina una nuova partizione Q generata dal prodotto cartesiano ($D \times T \times L$) delle modalità delle variabili considerate escluse quelle relative alla distribuzione delle famiglie per numero componenti;
- si individua la cella q (elemento della partizione Q) che contiene la cella z^* (elemento di Z) e si calcola il rapporto η_q ;
- si attribuisce il valore η_q a tutte le famiglie della cella z^* (sempre che per la cella q siano soddisfatti i vincoli sopra descritti)⁵.

Nel caso del collassamento delle celle, il numero delle famiglie campione non coincide con il numero delle famiglie campione dato come somma dei correttori per mancata risposta ($1/\eta_z$)⁶. Per tale motivo, per ognuno dei quattro campioni longitudinali, si procede a un riproporzionamento, di un fattore λ dei coefficienti appena calcolati in modo che sia soddisfatta la seguente uguaglianza:

$$m = \lambda \sum_{z \in Z} \frac{1}{\eta_z} \quad (9)$$

Il peso base corretto per mancata risposta assume, quindi, la seguente forma:

⁴ L'evidenza empirica e le varie prove effettuate hanno contribuito a determinare tali limiti.

⁵ Qualora il collassamento produca celle non utilizzabili, si procede ad una nuova operazione di collassamento togliendo prima la cittadinanza, poi le informazioni sulla dimensione demografica del comune.

⁶ La somma dei correttori per mancata risposta fornisce come risultato il numero teorico di unità campionarie; nel caso di Eu-Silc tale numero è pari a m per ognuno dei 4 campioni longitudinali.

$${}_{hi}\psi_z = {}_{hi}k_z^* \cdot \lambda \quad (10)$$

mentre la stima del totale espressa dalla (4) diviene:

$$\hat{Y} = \sum_{j=1}^{m'} y_j \psi_j \quad (11)$$

dove m' è il numero complessivo di famiglie rispondenti.

Dal processo di validazione sono emerse, anche per Eu-Silc, discrepanze, in alcuni casi anche accentuate, tra la stima di aggregati di interesse e i dati provenienti da altre fonti usati come riferimento nel processo di validazione. Tali differenze sono dovute verosimilmente al diverso comportamento di risposta in funzione del livello e della tipologia di reddito. Tale fenomeno, già riscontrato in altre indagini sui redditi come quella condotta dalla Banca d'Italia⁷, porta a possibili distorsioni nella stima degli aggregati e delle distribuzioni di interesse.

La soluzione adottata per la prima *wave* di Eu-Silc, soluzione che possiamo definire transitoria, è stata quella di inserire un coefficiente di correzione ν calcolato utilizzando la metodologia degli stimatori di ponderazione vincolata e le informazioni desunte dalla Rilevazione Continua sulle Forze di Lavoro. L'utilizzo di tali stimatori per limitare gli effetti dovuti alla mancata risposta totale rappresenta un esempio di modellizzazione implicita del processo di mancata risposta.

Nello specifico, l'informazione utilizzata è la posizione nella professione delle persone occupate opportunamente riclassificata desunta dalla Rcfl del IV trimestre 2004⁸.

I totali noti applicati sono:

- distribuzione delle famiglie per numero di componenti;
- popolazione per sesso e classi d'età⁹;
- popolazione per posizione nella professione.

L'introduzione dei totali noti riferiti alla popolazione per sesso e classi d'età e alla distribuzione delle famiglie per numero componenti sono serviti per dare "stabilità" ai correttori in funzione sia della costruzione delle celle di ponderazione sia della "calibrazione" finale che porta ai coefficienti di riporto all'universo finali¹⁰.

Con riferimento alla (10), per ogni famiglia campione del comune i appartenente allo strato h e per ogni cella di ponderazione z elemento della partizione Z , il correttore completo per mancata risposta assume la seguente forma:

$${}_{hi}\psi_{zg}^* = {}_{hi}\psi_z \times \nu_g \quad (12)$$

dove l'indice g esprime la generica combinazione dei totali noti inseriti nel passo di calibrazione.

⁷ Il Supplemento al Bollettino statistico della Banca d'Italia del Gennaio 2006, a pagina 35 recita: "...La difficoltà a ottenere l'intervista è crescente al crescere del reddito e del titolo di studio del capofamiglia; minori difficoltà si incontrano con le famiglie con un ridotto numero di componenti, con capofamiglia pensionato o residenti in comuni di piccole dimensioni...".

⁸ Il riferimento al IV trimestre 2004 della RCFL, che coincide con il periodo di rilevazione dei dati di Eu-Silc, è perché in Eu-Silc la rilevazione dettagliata delle caratteristiche della condizione è riferita a tale periodo.

⁹ Le classi d'età utilizzate sono: 0-14, 15-24, 25-44, 45-64, 65 e più al netto delle convivenze istituzionali..

¹⁰ In genere, quando si operano calibrazioni successive è necessario inserire in ogni passo dei vincoli altamente correlati con i vincoli dell'ultimo passo. Trattandosi di procedure indipendenti, infatti, c'è il rischio che ciò che si intende imporre in un generico passo venga poi annullato nel passo successivo di calibrazione previsto. Nel nostro caso si è preferito inserire la distribuzione per sesso e classe d'età (più ampie di quelle finali) in modo da "indirizzare" la calibrazione utilizzata per correggere la mancata risposta nella direzione dei vincoli finali che nel caso di Eu-Silc, hanno il solo obiettivo di l'accuratezza delle stime.

Tabella 2.2 - Modalità e riferimento temporale delle variabili utilizzate nel primo passo di calibrazione

Variabile	Modalità	Fonte e riferimento temporale
Famiglie per numero componenti	- uno - due - tre - quattro e +	Rilevazione continua sulla forze di lavoro IV trimestre 2004
Popolazione per ripartizione, sesso e classi d'età	- Nord-ovest - Nord-est - Centro - Sud - Isole	
	- Maschi - Femmine - 0-14 - 15-24 - 25-44 - 45-64 - 65+	
Popolazione di 15 anni e più per ripartizione e posizione nella professione ricodificata	- Nord-ovest - Nord-est - Centro - Sud - Isole - Dirigente, quadro, impiegato - Operaio e assimilato - Imprenditore e libero professionista - Lavoratore in proprio, socio di cooperativa e coadiuvante nell'azienda familiare - Co.co.co e prestatore d'opera occasionale - Disoccupato - Inattivo	

2.3 I totali noti e i coefficienti di riporto all'universo finali

Per il calcolo dei coefficienti di riporto all'universo finali, come già evidenziato nel corso del lavoro, si adottano gli stimatori di ponderazione vincolata (*calibration estimator*) metodologia implementata in un software, denominato Genesees, utilizzato in numerose altre indagini dell'istituto. La metodologia si basa sull'utilizzo di opportune informazioni ausiliarie, sintetizzate in totali noti, che, correlate con le variabili principali oggetto di indagine, hanno la funzione di aumentare l'accuratezza delle stime. La scelta delle informazioni ausiliarie, inoltre, è anche legata alla disponibilità periodica di tali fonti: proprio per questo Eurostat suggerisce ma non impone l'utilizzo delle informazioni da utilizzare per la costruzione delle stime finali.

In Eu-Silc, come in ogni indagine armonizzata a livello di Unione europea, la strategia di ponderazione è "guidata" al fine di garantire la comparabilità delle stime di ogni paese membro, oltre che per l'aumento dell'accuratezza delle stime. In particolare, il documento Eu-Silc 134-rev/04 - *Cross-sectional weighting: first year each sub-sample* - suggerisce, tra l'altro, i totali noti da utilizzare per la costruzione dello stimatore di ponderazione vincolata. Ognuno dei quattro campioni longitudinali è vincolato a:

- popolazione residente¹¹ per ripartizione territoriale, sesso e classi d'età (0-15, 16-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60-64, 65-69, 70-74, 75 e più) al 31 dicembre dell'anno di riferimento del reddito (anno t-1);
- il numero di famiglie residenti per regione riferite alla data di rilevazione (31 dicembre dell'anno t).

¹¹ Si intende al netto delle convivenze istituzionali.

Tabella 2.3 - Modalità e riferimento temporale delle variabili utilizzate nel secondo passo di calibrazione

Variabile	Modalità	Riferimento temporale	
Vincoli Eurostat applicati a ciascuno dei 4 campioni longitudinali			
Popolazione per ripartizione, sesso e classe di età	- Nord-ovest - Nord-est - Centro - Sud - Isole	31/12/2003	
	- Maschi - Femmine		
	- 0-15		- 45-49
	- 16-19		- 50-54
	- 20-24		- 55-59
	- 25-29		- 60-64
	- 30-34		- 65-69
	- 35-39		- 70-74
- 40-44	- 75+		
Famiglie per Regione	- 21 modalità: 19 Regioni (escluso il Trentino Alto Adige) più le Province Autonome di Trento e Bolzano (Eurostat Nuts II level)	31/12/2004	
Vincoli per esigenze informative interne applicati al campione trasversale nel complesso			
Popolazione per regione, sesso e classe di età	- 21 modalità: 19 Regioni (escluso il Trentino Alto Adige) più le Province Autonome di Trento e Bolzano (Eurostat Nuts II level)	31/12/2003	
	- Maschi - Femmine		
	- 0-15		
	- 16-24		
	- 25-44		
- 45-64			
- 65+			
Popolazione straniera per ripartizione e sesso	- Nord-ovest - Nord-est - Centro - Sud - Isole	31/12/2003	
	- Maschi - Femmine		
Popolazione straniera per ripartizione e cittadinanza	- Nord-ovest - Nord-est - Centro - Sud - Isole	31/12/2003	
	- UE - Non UE		
Popolazione straniera per ripartizione e cittadinanza	- Nord-ovest - Nord-est - Centro - Sud - Isole	31/12/2003	
	- Maggiorenni - Minorenni		
	- Maschi - Femmine		
	- 0-14		
	- 15-24		
- 25-44			
- 45-64			
- 65+			

Per esigenze informative nazionali, le stime prodotte con il campione *trasversale* (la somma dei 4 campioni longitudinali) sono vincolate anche a:

- popolazione residente per regione, sesso e classi d'età (0-14, 15-24, 25-44, 45-64, 65 e più) al 31 dicembre dell'anno di riferimento del reddito (anno t-1);
- popolazione straniera residente per ripartizione e sesso al 31 dicembre dell'anno di riferimento del reddito (anno t-1);
- popolazione straniera residente per ripartizione e provenienza (UE e non UE) 31 dicembre dell'anno di riferimento del reddito (anno t-1).

popolazione straniera residente per ripartizione ed età (maggioresnni e minoresnni) 31 dicembre dell'anno di riferimento del reddito (anno t-1).

3 Valutazione del livello di precisione delle stime

Gli errori di campionamento consentono di valutare il grado di precisione delle stime ottenute da un'indagine campionaria.

Con riferimento alla stima \hat{Y}_d nel dominio d , possono essere calcolati l'errore di campionamento assoluto espresso come:

$$\hat{\sigma}(\hat{Y}_d) = \sqrt{\hat{V}\hat{a}r(\hat{Y}_d)} \quad (13)$$

e l'errore di campionamento relativo espresso come:

$$\hat{\varepsilon}(\hat{Y}_d) = \frac{\hat{\sigma}(\hat{Y}_d)}{\hat{Y}_d} \quad (14)$$

Nella (13) $\hat{V}\hat{a}r(\hat{Y}_d)$ denota la stima della varianza campionaria di \hat{Y}_d .

L'errore di campionamento assoluto permette inoltre di costruire un intervallo di confidenza, nel quale con una certa probabilità si trova il parametro oggetto di stima:

$$\Pr\{\hat{Y}_d - k\hat{\sigma}(\hat{Y}_d) \leq Y_d \leq \hat{Y}_d + k\hat{\sigma}(\hat{Y}_d)\} = P \quad (15)$$

Nella (15) il valore di k dipende dal valore fissato per la probabilità P ; ad esempio, per $P=0.95$ si ha $k=1.96$.

Il calcolo degli errori di campionamento è stato eseguito mediante la procedura generalizzata di stima della varianza campionaria correntemente utilizzata in Istat.

Nel Prospetto 2 sono riportati l'errore campionario relativo, assoluto e l'intervallo di confidenza per il reddito familiare netto (esclusi i fitti imputati) per regione, ripartizione geografica e Italia.

Nel Prospetto 3 sono riportati l'errore campionario relativo, assoluto e l'intervallo di confidenza per il reddito individuale netto da lavoro (dipendente ed autonomo) per regione, ripartizione geografica e Italia; infine, nel Prospetto 4 sono riportati l'errore campionario relativo, assoluto e l'intervallo di confidenza per il reddito individuale netto da pensione per regione, ripartizione geografica e Italia.

Prospetto 2 - Reddito familiare netto (esclusi i fitti imputati), errore campionario relativo ed assoluto e intervallo di confidenza per regione e ripartizione geografica - Anno 2006

REGIONI E RIPARTIZIONI GEOGRAFICHE	Reddito familiare netto (media)	Errore		Intervallo di confidenza (p=0,95)	
		Relativo (%)	Assoluto	Estremo inferiore	Estremo superiore
Piemonte	29.599	1,9	563	28.496	30.702
Valle d'Aosta/Vallée d'Aoste	30.575	3,8	1164	28.294	32.856
Lombardia	32.852	1,6	538	31.797	33.907
Trentino-Alto Adige	33.039	1,9	632	31.801	34.277
<i>Bolzano/Bozen</i>	35.964	2,8	990	34.023	37.904
<i>Trento</i>	30.376	2,6	802	28.803	31.948
Veneto	30.479	1,5	467	29.563	31.395
Friuli-Venezia Giulia	29.305	1,8	522	28.282	30.328
Liguria	26.190	2,0	517	25.178	27.203
Emilia-Romagna	32.587	2,0	641	31.332	33.843
Toscana	31.107	1,7	525	30.077	32.137
Umbria	30.787	2,0	626	29.559	32.015
Marche	30.496	1,9	566	29.388	31.605
Lazio	31.504	2,6	814	29.908	33.100
Abruzzo	27.223	4,4	1196	24.880	29.567
Molise	23.420	3,6	840	21.774	25.066
Campania	24.254	2,5	600	23.078	25.429
Puglia	24.547	2,3	568	23.433	25.661
Basilicata	23.355	2,6	615	22.150	24.560
Calabria	22.623	3,5	798	21.059	24.187
Sicilia	21.463	2,9	624	20.240	22.687
Sardegna	27.172	3,3	894	25.420	28.923
<i>Nord-Ovest</i>	31.165	1,2	363	30.453	31.876
<i>Nord-Est</i>	31.394	1,0	326	30.756	32.033
<i>Centro</i>	31.189	1,4	427	30.353	32.026
<i>Sud</i>	24.341	1,4	336	23.683	24.999
<i>Isole</i>	22.903	2,3	518	21.888	23.919
Italia	28.872	0,6	174	28.531	29.212

Prospetto 3 - Reddito individuale netto da lavoro, errore relativo ed assoluto e intervallo di confidenza per regione e ripartizione geografica - Anno 2006

REGIONI E RIPARTIZIONI GEOGRAFICHE	Reddito individuale netto da lavoro (media)	Errore		Intervallo di confidenza (p=0,95)	
		Relativo (%)	Assoluto	Estremo inferiore	Estremo superiore
Piemonte	18.439	2,4	440	17.576	19.302
Valle d'Aosta/Vallée d'Aoste	18.559	5,0	925	16.747	20.371
Lombardia	19.940	2,0	395	19.165	20.715
Trentino-Alto-dige	17.774	2,6	454	16.885	18.663
<i>Bolzano/Bozen</i>	<i>19.185</i>	<i>3,3</i>	<i>642</i>	<i>17.926</i>	<i>20.444</i>
<i>Trento</i>	<i>16.471</i>	<i>3,9</i>	<i>640</i>	<i>15.217</i>	<i>17.725</i>
Veneto	17.252	2,1	361	16.544	17.960
Friuli-Venezia Giulia	17.442	2,5	442	16.575	18.309
Liguria	16.826	2,6	445	15.954	17.698
Emilia-Romagna	18.396	2,5	460	17.495	19.297
Toscana	18.259	2,4	435	17.407	19.111
Umbria	17.498	3,4	589	16.344	18.652
Marche	16.065	2,4	387	15.306	16.824
Lazio	19.493	3,6	704	18.113	20.873
Abruzzo	16.100	4,7	756	14.617	17.583
Molise	14.766	7,1	1043	12.722	16.810
Campania	15.213	3,2	486	14.261	16.165
Puglia	14.487	3,3	480	13.547	15.427
Basilicata	15.064	3,9	586	13.916	16.212
Calabria	13.886	5,5	770	12.377	15.395
Sicilia	14.705	4,0	587	13.554	15.856
Sardegna	16.331	4,2	686	14.986	17.676
<i>Nord-Ovest</i>	<i>19.224</i>	<i>1,4</i>	<i>278</i>	<i>18.680</i>	<i>19.768</i>
<i>Nord-Est</i>	<i>17.765</i>	<i>1,4</i>	<i>242</i>	<i>17.291</i>	<i>18.239</i>
<i>Centro</i>	<i>18.447</i>	<i>1,9</i>	<i>357</i>	<i>17.747</i>	<i>19.147</i>
<i>Sud</i>	<i>14.887</i>	<i>1,9</i>	<i>275</i>	<i>14.347</i>	<i>15.427</i>
<i>Isole</i>	<i>15.181</i>	<i>3,0</i>	<i>462</i>	<i>14.276</i>	<i>16.086</i>
Italia	17.487	0,8	139	17.215	17.759

Prospetto 4 - Reddito individuale netto da pensione, errore campionario relativo ed assoluto e intervallo di confidenza per regione e ripartizione geografica - Anno 2006

REGIONI E RIPARTIZIONI GEOGRAFICHE	Reddito individuale netto da lavoro (media)	Errore		Intervallo di confidenza (p=0,95)	
		Relativo (%)	Assoluto	Estremo inferiore	Estremo superiore
Piemonte	12.054	2,2	268	11.529	12.578
Valle d'Aosta/Vallée d'Aoste	11.781	4,1	481	10.839	12.723
Lombardia	12.612	2,1	265	12.093	13.131
Trentino-Alto Adige	11.828	3,0	353	11.136	12.521
<i>Bolzano/Bozen</i>	<i>12.360</i>	<i>3,9</i>	<i>484</i>	<i>11.412</i>	<i>13.309</i>
<i>Trento</i>	<i>11.336</i>	<i>4,5</i>	<i>512</i>	<i>10.333</i>	<i>12.339</i>
Veneto	11.232	2,7	302	10.641	11.824
Friuli-Venezia Giulia	12.535	3,1	384	11.783	13.286
Liguria	12.492	3,4	430	11.650	13.335
Emilia-Romagna	11.988	2,3	273	11.452	12.524
Toscana	11.833	2,0	237	11.369	12.298
Umbria	11.584	2,9	333	10.932	12.236
Marche	11.396	3,6	408	10.596	12.195
Lazio	13.876	3,6	495	12.905	14.848
Abruzzo	9.888	5,2	515	8.878	10.898
Molise	9.208	5,1	467	8.292	10.123
Campania	10.475	3,3	347	9.795	11.156
Puglia	10.584	3,0	316	9.965	11.203
Basilicata	8.738	5,5	479	7.800	9.676
Calabria	9.622	4,7	450	8.741	10.504
Sicilia	10.433	3,5	363	9.722	11.144
Sardegna	11.915	5,1	603	10.733	13.098
<i>Nord-Ovest</i>	<i>12.423</i>	<i>1,4</i>	<i>179</i>	<i>12.072</i>	<i>12.774</i>
<i>Nord-Est</i>	<i>11.745</i>	<i>1,5</i>	<i>171</i>	<i>11.410</i>	<i>12.080</i>
<i>Centro</i>	<i>12.610</i>	<i>1,9</i>	<i>234</i>	<i>12.152</i>	<i>13.069</i>
<i>Sud</i>	<i>10.192</i>	<i>1,8</i>	<i>184</i>	<i>9.832</i>	<i>10.552</i>
<i>Isole</i>	<i>10.806</i>	<i>2,9</i>	<i>311</i>	<i>10.196</i>	<i>11.415</i>
Italia	11.681	0,8	92	11.500	11.862

Avvertenze

Per ciò che concerne i prospetti e le tavole, si precisa che, salvo diversa indicazione, i valori monetari sono espressi in euro e gli altri valori assoluti in migliaia di unità; inoltre, i totali possono non quadrare per effetto di arrotondamento. Le percentuali e i quozienti, essendo calcolati sui dati assoluti prima dell'arrotondamento, possono non coincidere con i risultati che si ottengono elaborando i dati in migliaia. In generale i quozienti sono espressi per 100 persone o per 100 famiglie, salvo diversa indicazione;

Nei prospetti e nelle tavole l'assenza di valori numerici in una riga o in una colonna di dati è segnalata attraverso i seguenti segni convenzionali:

- Linea (-): indica che il fenomeno non esiste oppure che il fenomeno esiste e viene rilevato ma i casi non si sono verificati.
- Quattro puntini (...): indicano che il fenomeno esiste ma i dati non si conoscono per qualsiasi ragione
- Due puntini (..): sono utilizzati per i dati in percentuale ed indicano i numeri che non raggiungono la metà della cifra dell'ordine minimo considerato. Per l'indagine Eu-Silc, tale segno convenzionale indica una numerosità campionaria inferiore alle 20 unità (Regolamento comunitario n. 1982/2003).

Si ricorda che i risultati dell'indagine riguardano la popolazione residente al netto dei membri permanenti delle convivenze, poiché si tratta di un'indagine campionaria effettuata presso le famiglie campione che rappresentano le unità di rilevazione.

Le **ripartizioni geografiche** costituiscono una suddivisione geografica del territorio e sono così articolate:

- **Nord-ovest** comprende: Piemonte, Valle d'Aosta, Lombardia, Liguria;
- **Nord-est** comprende: Trentino-Alto Adige, Veneto, Friuli-Venezia Giulia, Emilia-Romagna;
- **Centro** comprende: Toscana, Umbria, Marche, Lazio;
- **Sud** comprende: Abruzzo, Molise, Campania, Puglia, Basilicata, Calabria;
- **Isole** comprende: Sicilia, Sardegna.

I comuni italiani sono suddivisi, secondo il **tipo di comune**, nelle seguenti classi:

- **comuni centro delle aree metropolitane**: si tratta dei comuni di Torino, Milano, Venezia, Genova, Bologna, Firenze, Roma, Napoli, Bari, Palermo, Catania, Cagliari;
- **comuni appartenenti alla periferia delle aree metropolitane**: costituiscono i comuni delle cinture urbane;
- **altri comuni**: suddivisi per dimensione demografica (fino a 2.000 abitanti, da 2.001 a 10 mila, da 10.001 a 50.000 e oltre i 50.000).

Si precisa che la soglia dei 2 mila abitanti costituisce la dimensione demografica suggerita dagli organismi internazionali per identificare uno stile di vita tipico dei piccoli centri.

Glossario

I dati generali individuali fanno riferimento alle caratteristiche delle persone all'epoca dell'intervista. In particolare:

- **l'età** è espressa in anni compiuti al 31 dicembre dell'anno di riferimento del reddito (2006);

- **il titolo di studio** è quello più elevato conseguito;

- **la condizione** è quella dichiarata come unica o prevalente dalle persone di 15 anni e più.

Si precisa inoltre che per:

occupato si intende chi svolge un'attività lavorativa (in proprio o alle dipendenze) da cui trae un profitto o una retribuzione (salario, stipendio, eccetera). Sono da considerare occupati:

- chi collabora con un familiare che svolge attività lavorativa in conto proprio (coadiuvante familiare), anche se non retribuita;
- i lavoratori in Cassa Integrazione Guadagni;
- le persone assunte con contratto di formazione lavoro;
- gli apprendisti retribuiti.

persona in cerca di occupazione si intende chi ha perduto una precedente occupazione o chi non ha mai esercitato un'attività lavorativa ed è alla ricerca attiva di un'occupazione che è in grado di accettare se gli viene offerta;

casalinga è chi si dedica prevalentemente alle faccende domestiche;

studente è chi si dedica prevalentemente allo studio;

ritirato dal lavoro è chi ha cessato un'attività lavorativa per raggiunti limiti di età, invalidità o altra causa; la figura del ritirato dal lavoro non coincide necessariamente con quella del pensionato in quanto, non sempre, il ritirato dal lavoro gode di una pensione;

in altra condizione è colui che si trova in condizione diversa da quelle sopra elencate (militare, inabile al lavoro, benestante, detenuto, eccetera)

- **la posizione nella professione** è quella dichiarata come unica o prevalente dagli occupati di 15 anni e più che viene aggregata nel modo seguente:

Lavoratore dipendente:

dirigente: chi ricopre un ruolo caratterizzato da un elevato grado di professionalità, autonomia e potere decisionale ed esplica la sua funzione al fine di promuovere, coordinare e gestire la realizzazione degli obiettivi dell'impresa o dell'ente, sia esso privato o pubblico;

quadro: chi svolge nel campo tecnico o scientifico e amministrativo, con diverso grado di responsabilità, discrezionalità di poteri e autonomia, funzioni direttive e /o coordinamento nell'ambito di un servizio o di un ufficio;

impiegato: chi collabora, a volte con adeguata autonomia operativa o nei limiti di principio o procedure, allo svolgimento dell'attività compresa nel campo in cui opera, o che svolge attività di mera esecuzione;

operaio: chi guida e controlla, con apporto di adeguata competenza tecnico-pratica, gruppi di altri operai esercitando a volte un certo potere di iniziativa per la condotta delle lavorazioni. Chi esercita un'attività per la cui esecuzione si richiedono conoscenze specialistiche o cognizioni tecnico-pratiche;

apprendista: chi, nell'ambito della normativa ufficiale, esercita un'attività lavorativa finalizzata all'apprendimento di un mestiere, l'addestramento al lavoro ed alla acquisizione di specifiche competenze che consentono l'ottenimento di una qualificazione professionale;

lavorante a domicilio per conto di imprese: chi lavora nel proprio domicilio esclusivamente o prevalentemente su commissione di imprese (industriali, commerciali, artigiane) e non direttamente per conto di consumatori.

Lavoratore autonomo:

imprenditore: chi gestisce in proprio un'impresa (agricola, industriale, commerciale, di servizi, eccetera);

libero professionista: chi esercita in conto proprio una professione o arte liberale (notaio, avvocato, medico, dentista, ingegnere edile, eccetera);

lavoratore in proprio: chi gestisce un'azienda agricola, una piccola azienda industriale o commerciale, una bottega artigiana, un negozio o un esercizio pubblico partecipandovi col proprio lavoro manuale. In tale posizione sono compresi anche i coltivatori diretti, i mezzadri e simili; chi lavora nel proprio domicilio direttamente per conto dei consumatori e non su commissione di imprese;

socio di cooperativa di produzione e/o di prestazione di servizi: chi è membro attivo di una cooperativa di produzione di beni e/o di prestazione di servizi indipendentemente dalla specie di attività in cui la cooperativa è operante, cioè chi, come corrispettivo all'opera prestata non percepisce una remunerazione regolata da contratti di lavoro (nel qual caso dovrà essere classificato nell'ambito della categoria 'dipendenti'), ma un compenso proporzionato alla prestazione e/o una quota parte degli utili dell'impresa;

coadiuvante nella ditta di un familiare: chi collabora con un familiare che svolge un'attività in conto proprio (es. moglie che aiuta il marito negoziante, figlio che aiuta il padre agricoltore, eccetera).

• **Famiglia e nucleo familiare**

la **famiglia** è costituita dall'insieme delle persone coabitanti legate da vincoli di matrimonio o parentela, affinità, adozione, tutela o affettivi;

il **nucleo** è l'insieme delle persone che formano una coppia con figli celibi o nubili, una coppia senza figli, un genitore solo con figli celibi o nubili;

una famiglia può coincidere con un nucleo, può essere formata da un nucleo più altri membri aggregati, da più nuclei (con o senza membri aggregati), o da nessun nucleo (persone sole, famiglie composte ad esempio da due sorelle, da un genitore con figlio separato, divorziato o vedovo, eccetera);

• **Fitto imputato**

Il valore del fitto imputato viene stimato dal proprietario in base al prezzo che secondo lui si dovrebbe pagare per vivere in affitto nella sua abitazione. Da questa stima, vengono detratti gli eventuali interessi pagati sul mutuo casa. La parte di spesa per il mutuo destinata a rimborsare il capitale prestato, invece, non viene sottratta perché corrisponde a una riduzione del debito, cioè a un aumento del patrimonio della famiglia. Per analogia, agli inquilini che pagano un affitto agevolato viene imputata la differenza fra il prezzo di mercato e il fitto pagato. Il fitto imputato è considerato al netto di eventuali spese di manutenzione ordinaria, che corrispondono all'ammortamento del patrimonio abitativo.

• Indice di Gini

L'indice di concentrazione di Gini è una misura sintetica del grado di disuguaglianza della distribuzione del reddito: è pari a zero nel caso di una perfetta equità della distribuzione dei redditi, quando tutte le famiglie ricevono lo stesso reddito; è invece pari a uno nel caso di totale disuguaglianza, quando il reddito totale è percepito da una sola famiglia. Il metodo utilizzato per il calcolo dell'indice di Gini sui dati campionari dell'indagine Eu-Silc deriva dalla formula, valida per i grandi campioni:

$$G = \frac{2 \operatorname{cov}[y_i, F(y_i)]}{\bar{y}}$$

Dove y_i è il reddito della famiglia, \bar{y} è il reddito medio e $F(y_i)$ è il *rango frazionario*, cioè il posto (numero ordinale) della famiglia nella distribuzione dei redditi ordinati dal più basso al più alto, successivamente diviso per il numero totale di famiglie. In questa sede, conformemente alle convenzioni adottate in sede europea, l'indice di Gini è stato calcolato attribuendo ad ogni individuo il reddito (familiare) equivalente della famiglia di appartenenza e considerando come unità statistiche gli individui.

• Reddito netto

Il reddito netto familiare considerato dall'indagine è pari alla somma dei redditi da lavoro dipendente e autonomo, di quelli da capitale reale e finanziario, delle pensioni e degli altri trasferimenti pubblici e privati al netto delle imposte personali, dell'Ici e dei contributi sociali a carico dei lavoratori dipendenti ed autonomi. Da questa somma vengono sottratti anche i trasferimenti versati ad altre famiglie (per esempio, gli assegni di mantenimento per un ex-coniuge). I redditi da lavoro dipendente comprendono il valore figurativo dell'auto aziendale concessa per uso privato ma non i buoni-pasto e gli altri *fringe benefits* non-monetari. Non sono compresi gli eventuali beni prodotti dalla famiglia per il proprio consumo (autoconsumo). Nella definizione armonizzata adottata da Eurostat per le prime tre edizioni dell'indagine Eu-Silc (2004-2006), i redditi da capitale reale non comprendono il reddito figurativo delle abitazioni occupate dai proprietari (cioè l'affitto imputato). In base al Regolamento, l'inserimento di questa ultima componente del reddito sarà comunque obbligatorio a partire dall'edizione 2007. Data l'importanza della proprietà dell'abitazione nel contesto italiano, si è deciso di riportare le statistiche relative ai redditi familiari sia al netto, sia al lordo dei fitti imputati. Il valore dell'affitto imputato viene stimato dal proprietario in base al prezzo che secondo lui si dovrebbe pagare per vivere in affitto nella sua abitazione. Da questa stima, vengono detratti gli eventuali interessi pagati sul mutuo-casa. La parte di spesa per il mutuo destinata a rimborsare il capitale prestato, invece, non viene sottratta perché corrisponde ad una riduzione del debito, cioè ad un aumento del patrimonio della famiglia. Per analogia, agli inquilini che pagano un affitto agevolato viene imputata la differenza fra il prezzo di mercato e l'affitto pagato. L'affitto imputato è considerato al netto di eventuali spese di manutenzione ordinaria, che corrispondono all'ammortamento del patrimonio abitativo.

Il reddito netto familiare stimato in base all'indagine campionaria Eu-Silc non è perfettamente comparabile con il reddito disponibile aggregato del settore Famiglie, riportato nei Conti nazionali. L'indagine campionaria Eu-Silc riguarda le famiglie residenti registrate nelle Anagrafi comunali, mentre il settore Famiglie di Contabilità nazionale include tutte le famiglie presenti da più di un anno sul territorio nazionale (quindi, per esempio, anche gli immigrati irregolari). In secondo luogo, il reddito disponibile delle famiglie di Contabilità nazionale include una stima dell'economia 'sommersa' che, per ovvie ragioni, non è possibile rilevare attraverso un'indagine campionaria condotta presso le famiglie. In generale, nella esperienza della maggior parte dei paesi, le indagini campionarie sottostimano una parte dei redditi per effetto della scarsa memoria o della reticenza di alcuni intervistati. In particolare, risulta particolarmente difficile la rilevazione dei redditi da attività finanziarie e di una parte dei redditi da lavoro autonomo.

- **Altri trasferimenti pubblici**

Includono le prestazioni monetarie non pensionistiche tra cui, le indennità di disoccupazione, la Cassa Integrazione Guadagni, gli assegni familiari, le indennità di maternità, nonché i redditi da borse lavoro e da borse di studio.

- **Scala di equivalenza**

Una semplice operazione preliminare necessaria per misurare la disuguaglianza dei redditi consiste nell'ordinare le famiglie in base al reddito, dal più basso al più alto. Questo ordinamento richiede il confronto del reddito di ogni famiglia con quello di tutte le altre. Per confrontare i redditi, tuttavia, è necessario tener conto del fatto che le famiglie più numerose hanno maggiori bisogni. Un modo estremo di considerare la diversa composizione familiare è quello di ordinare le famiglie in base al reddito *pro-capite*, cioè al reddito familiare diviso per il numero di componenti. Una simile soluzione, tuttavia, ignora completamente le 'economie di scala', cioè il fatto che i costi di una famiglia non sono perfettamente proporzionali al numero di componenti. Per esempio, la bolletta del gas di una famiglia di quattro persone non è, in generale, precisamente pari a quattro volte la spesa di una persona sola. Per questa ragione, si utilizzano le scale di equivalenza. Una scala di equivalenza è un insieme di parametri che vengono utilizzati per dividere il reddito familiare in modo da ottenere un reddito 'equivalente', che tiene conto della diversa composizione delle famiglie. In questa sede, per ogni famiglia, il parametro utilizzato per calcolare il reddito 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 minore di 14 anni). Questa scala di equivalenza, raccomandata dall'Ocse, è attualmente impiegata da Eurostat per il calcolo degli indicatori di disuguaglianza compresi nelle statistiche ufficiali dell'Unione europea.