

SN162

# Indagine sulle spese delle famiglie (2014)

Istat

Versione: 1.0 - 27/06/2017



## UniData

Bicocca Data Archive

Website: [www.unidata.unimib.it](http://www.unidata.unimib.it)

E-mail: [unidata@unimib.it](mailto:unidata@unimib.it)

Tel.: +39 02 6448 7513

Fax: +39 02 6448 7561

La presente documentazione è distribuita da UniData secondo la [licenza CC-BY 3.0](#).  
La fonte che ha prodotto i dati e UniData che li ha distribuiti non rispondono per alcun  
utilizzo improprio dei dati e delle elaborazioni pubblicate.

*This documentation is distributed by UniData under the [CC-BY 3.0 License](#).  
Neither the depositor nor UniData bear any responsibility for the analysis or  
interpretation of the data produced by the user.*



Università degli Studi di Milano-Bicocca  
Via Bicocca degli Arcimboldi 8  
20126 - Milano (Italia)

# **Note metodologiche**

## *Methodological Notes*

### **INDICE**

#### *TABLE OF CONTENTS*

Aspetti metodologici dell'indagine  
*Study's methodological issues*

---

p. 3

Descrizione del file  
*File description*

---

p. 24

**Aspetti metodologici dell'indagine**  
*Study's methodological issues*



**File ad uso pubblico  
micro.STAT**

## **Indagine sulle spese delle famiglie**

**Periodo di riferimento: anno 2014**

### **Aspetti metodologici dell'indagine**

## INDICE

1. Introduzione .....	3
2. La popolazione di riferimento .....	4
3. Il disegno campionario .....	4
4. La rilevazione e il trattamento dei dati.....	9
5. La metodologia di calcolo dei pesi campionari.....	10
6. La diffusione dei risultati dell'indagine.....	16
7. Glossario.....	17
8. Riferimenti bibliografici.....	19
9. Contatti .....	19

## 1. Introduzione

L'indagine sulle spese delle famiglie rileva i comportamenti di spesa delle famiglie residenti in Italia. Tale indagine sostituisce la precedente indagine sui consumi e le sostanziali modifiche introdotte in tutte le fasi del processo hanno reso necessario ricostruire le serie storiche dei principali aggregati di spesa a partire dal 1997. I confronti temporali possono dunque essere effettuati esclusivamente con i dati in serie storica pubblicati insieme al comunicato stampa dell'8 luglio 2015 (<http://www.istat.it/it/archivio/164313>) nonché con quelli rilasciati sulla banca dati del sito dell'Istituto (<http://dati.istat.it/>), mentre non possono essere effettuati con i dati precedentemente diffusi.

Le definizioni e le metodologie sono armonizzate alle più recenti direttive europee (<http://www.istat.it/it/archivio/71980>).

Oggetto della rilevazione sono tutte le spese sostenute dalle famiglie residenti per acquistare beni e servizi destinati al consumo familiare o per effettuare regali a persone esterne alla famiglia. In tale definizione rientrano anche i beni provenienti dal proprio orto o dalla propria azienda agricola direttamente consumati dalla famiglia (autoconsumi) o regalati, i beni e servizi forniti dal datore di lavoro ai dipendenti a titolo di salario o per prestazioni di servizio, i fitti stimati delle abitazioni occupate dai proprietari o godute a titolo gratuito (fitti figurativi). Ogni altra spesa effettuata dalla famiglia per scopo diverso dal consumo è esclusa dalla rilevazione (ad es., l'acquisto di una casa e di terreni, il pagamento delle imposte, le spese connesse con attività professionale). Va, infine, precisato che le spese relative al mutuo per l'acquisto di abitazioni e alla restituzione di prestiti contratti con banche o società finanziarie o con parenti o amici, seppur rilevate nell'indagine, non rientrano nel concetto economico di consumo (in quanto rappresentano forme di investimento) e non vengono quindi incluse nel calcolo della spesa.

Alla rilevazione della spesa è associata la rilevazione delle principali caratteristiche socio-economiche degli individui che compongono la famiglia.

L'indagine rappresenta, inoltre, la base informativa per le stime ufficiali della povertà assoluta e relativa in Italia. Nella stima della povertà, la spesa considerata esclude, oltre alle spese per mutui e restituzione prestiti, anche le spese relative alla manutenzione straordinaria dell'abitazione e le spese relative ai premi pagati per assicurazioni vita e rendite vitalizie (assicurazione vita, rendite vitalizie). Per quanto riguarda il calcolo

dell'incidenza di povertà assoluta, poiché le relative soglie non sono direttamente ricostruibili a partire dalle variabili contenute nei dati dell'indagine, è stata introdotta la variabile "povertà assoluta" che consente di classificare le famiglie in povere assolute e non povere assolute.

## **2. La popolazione di riferimento**

La popolazione di riferimento è costituita dalle famiglie residenti e dagli individui che le compongono.

L'unità di rilevazione è la famiglia di fatto, intesa come insieme di persone coabitanti, legate da vincoli di matrimonio o parentela, affinità, adozione, tutela o affettivi e che partecipano alla spesa familiare e/o condividono il reddito familiare.

Sono quindi escluse dalla popolazione di riferimento le persone che vivono permanentemente nelle comunità (caserme, ospedali, istituti religiosi, eccetera).

## **3. Il disegno campionario**

Le innovazioni apportate in termini di campionamento hanno riguardato l'introduzione della tipologia comunale come variabile di stratificazione delle unità di primo stadio e la modifica dello schema di assegnazione casuale del periodo di compilazione del diario a seguito del suo ampliamento da 7 a 14 giorni.

L'indagine deve fornire stime trimestrali a livello nazionale, mentre i domini territoriali di riferimento per le stime a livello annuale sono:

- l'intero territorio nazionale;
- le cinque grandi ripartizioni geografiche (Nord-ovest : Piemonte, Valle d'Aosta, Lombardia, Liguria; Nord-est: Trentino-Alto Adige, Veneto, Friuli-Venezia Giulia, Emilia-Romagna; Centro: Toscana, Umbria, Marche, Lazio; Sud: Abruzzo, Molise, Campania, Puglia, Basilicata, Calabria; Isole: Sicilia, Sardegna);
- la tipologia comunale, ottenuta suddividendo i comuni italiani in tre classi, formate in base a caratteristiche socio-economiche e demografiche (sulla base della popolazione residente per l'anno di riferimento), che consta delle seguenti modalità:

- T1: comuni centro dell'area metropolitana: Torino, Milano, Venezia, Genova, Bologna, Firenze, Roma, Napoli, Bari, Palermo, Catania e Cagliari;
- T2: comuni che gravitano intorno al centro dell'area metropolitana e comuni con più di 50.000 abitanti,
- T3: rimanenti comuni.

Il disegno d'indagine è stato progettato, inoltre, per fornire stime a livello regionale sull'intero anno, distinguendo le province di Bolzano e Trento. La base di campionamento adottata, ossia la lista di selezione delle unità campionarie, è costituita dalle LAC, un archivio unico delle famiglie residenti nei comuni italiani, proveniente dalle liste anagrafiche comunali.

### **3.1. Descrizione generale**

Il tipo di disegno adottato è a due stadi di selezione (comuni, famiglie). Ciò è determinato dal fatto che l'indagine utilizza l'intervista diretta con tecnica CAPI, e pertanto è necessario, per ragioni sia di costo e sia soprattutto organizzative, che il campione sia concentrato in un numero limitato di comuni.

Le unità di primo stadio, i comuni, sono stratificati all'interno di ciascuna regione geografica secondo le due variabili tipologia comunale e dimensione demografica. La prima è definita sulla base dei domini illustrati nel precedente paragrafo, la seconda variabile di stratificazione – espressa in termini di popolazione residente – viene utilizzata per costruire una soglia in base alla quale suddividere i comuni in due sottoinsiemi: (i) gli Auto Rappresentativi o AR, sono i comuni con popolazione al di sopra della soglia e costituiscono ognuno uno strato a sé; (ii) i Non Auto Rappresentativi o NAR, sono i restanti comuni. Per i comuni AR si adotta un disegno ad uno stadio stratificato, mentre per i NAR si usa un disegno a due stadi con stratificazione delle unità di primo stadio.

Da ogni comune campione si estrae un campione di famiglie anagrafiche che costituiscono le unità finali di campionamento.

Ogni trimestre si intervistano circa 7.000 famiglie, per un totale di 28.000 famiglie all'anno. Il disegno di campionamento, definito con riferimento a un generico trimestre dell'anno, viene replicato in modo identico per i quattro trimestri e viene anche effettuata una stratificazione mensile del campione trimestrale. Di conseguenza, la dimensione temporale può essere considerata un'ulteriore variabile di stratificazione del campione.



Essendo l'indagine basata su una rilevazione trimestrale, il disegno di campionamento viene definito con riferimento al generico trimestre di un anno e il campione di comuni risultante è coinvolto quattro volte in modo identico nella rilevazione nell'anno di riferimento. Inoltre, viene attuata una stratificazione mensile suddividendo il campione di ciascuno strato della popolazione nei tre mesi che compongono ciascun trimestre. In particolare, la partecipazione dei comuni inclusi nel campione nell'arco dell'anno di riferimento è la seguente:

- ogni comune AR è coinvolto nell'indagine tutti i mesi (e il numero di famiglie campione è suddiviso nei mesi);
- ciascuno dei tre comuni campione di ogni strato NAR partecipa all'indagine quattro mesi nell'anno a distanza di tre mesi secondo lo schema seguente:

<b>comuni</b>	<b> mese di rilevazione</b>			
<b>1</b>	gennaio	aprile	luglio	ottobre
<b>2</b>	febbraio	maggio	agosto	novembre
<b>3</b>	marzo	giugno	settembre	dicembre

### **3.2. Stratificazione e selezione del campione**

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:

- auto-ponderazione del campione al livello di regione;
- definizione di un numero minimo di famiglie da intervistare per comune;
- stratificazione dei comuni sulla base dell'ampiezza demografica;
- formazione di strati di comuni di ampiezza approssimativamente costante in termini di popolazione residente.

Il procedimento di stratificazione è attuato all'interno di ogni dominio territoriale individuato, per ciascuna regione geografica, dalle tre tipologie comunali considerate.

Con riferimento al generico incrocio di regione e tipologia comunale, indicato nel seguito con il termine *dominio*, si denoti con:  $h$  ( $h=1,\dots,H$ ), l'indice di strato di comuni;  $i$  ( $i=1,\dots,N$ ), l'indice di comune;  $M_h$ , il numero di famiglie residenti nello strato  $h$ ;  $M_{hi}$  il numero di famiglie residenti nel comune  $i$  dello strato  $h$ ;  $P_h$ , il numero di individui residenti nello strato  $h$ ;  $P_{hi}$  il numero di individui residenti nel comune  $i$  dello strato  $h$ ;  $n$ , il numero di comuni campione in ogni strato.

Per la definizione del campione di comuni e di famiglie relativo ad un trimestre sono state dapprima effettuate le seguenti scelte:

- (a) definizione del numero complessivo di famiglie campione a livello nazionale;
- (b) definizione del numero  $m_r$  di famiglie campione per ciascuna regione;
- (c) distribuzione delle numerosità regionali tra i domini definiti dalla tipologia comunale in misura proporzionale alla popolazione residente;
- (d) scelta del numero minimo di famiglie,  $\bar{m}_r$ , da intervistare in ciascun comune campione.

Dalla scelta di  $\bar{m}_r$  e  $m_r$  dipende la suddivisione dei comuni in AR e NAR e la formazione degli strati attraverso i seguenti passi:

- (1) calcolo della frazione di campionamento regionale  $f_r = m_r/M_r$ , essendo  $M_r$  il numero di famiglie residenti nella regione  $r$ ;
- (2) determinazione del valore della *soglia*<sup>1</sup>  $\lambda_r$ , mediante la relazione

$$\lambda_r = \frac{\bar{m}_r \delta_r}{f_r},$$

in cui  $\delta_r$  è il numero medio di componenti per famiglia a livello regionale; risulta evidente da tale espressione che la soglia per la definizione dei comuni AR cresce al crescere di  $\bar{m}_r$ ;

- (3) suddivisione dei comuni in AR e NAR sulla base della soglia  $\lambda_r$ ;
- (4) ordinamento decrescente dei comuni NAR all'interno di ogni dominio in funzione della loro dimensione demografica;
- (5) suddivisione dei comuni NAR in strati la cui dimensione è approssimativamente uguale al prodotto  $n \times \lambda_r$ ;

---

<sup>1</sup> Dal momento che il campione di famiglie di ogni regione è distribuito in modo proporzionale tra i domini e, di conseguenza, la frazione di campionamento è costante all'interno della regione, la soglia risulta anch'essa definita a livello regionale, così come la suddivisione dei comuni in AR e NAR.

(6) selezione di  $n$  comuni campione da ciascuno strato  $h$  ( $h=1,\dots,H$ ) con probabilità proporzionale all'ampiezza; per il generico comune  $c$  tale probabilità è espressa dalla formula:

$$z_{hi} = n \cdot P_{hi} / P_h ;$$

(7) definizione del numero  $m_{hi}$  di famiglie da intervistare in ogni comune; dalla condizione di autoponderazione a livello regionale

$$\frac{n \cdot P_{hi}}{P_h} \frac{m_{hi}}{M_{hi}} = f_r ,$$

in cui il primo membro rappresenta la probabilità d'inclusione delle famiglie del comune  $c$  dello strato  $h$  (essendo le due frazioni rispettivamente la probabilità d'inclusione di primo e di secondo stadio), si ottiene

$$m_{hi} = \frac{f_r}{n} \frac{P_h \cdot M_{hi}}{P_{hi}} .$$

Al fine di raggiungere la numerosità campionaria desiderata in termini di comuni nel rispetto della condizione di auto-ponderazione a livello regionale, il numero minimo di interviste per comune è stato posto pari a 14 per le regioni con popolazione superiore a 1.500.000 di abitanti e 12 per le regioni più piccole.

Relativamente alla quantità  $m$ , per consentire la stratificazione mensile del campione di comuni, si è scelta una numerosità per ogni strato NAR pari a 3.

Nella tabella seguente vengono riportate le numerosità campionarie in termini di comuni.

	Comuni campione		
	AR	NAR	TOTALI
<b>Mese</b>	54	147	201
<b>Trimestre</b>	54	441	495
<b>Anno</b>	54	441	495

L'estrazione delle famiglie campione viene effettuata, mediante selezione casuale, dalla lista delle anagrafi comunali; per ciascun comune campione vengono selezionate quindi complessivamente  $4m_{hi}$  famiglie che vengono ripartite nell'anno come segue:

- il campione di famiglie di ciascun comune AR viene suddiviso in 12 campioni mensili;
- il campione di famiglie di ciascun comune NAR viene suddiviso in 4 campioni mensili;
- ciascun campione mensile viene suddiviso in due gruppi, uno per ciascun dei due periodi di rilevazione di 14 giorni di ogni mese.

Nel prospetto 1 è riportato, per ogni regione, l'universo di riferimento e il numero di Comuni, famiglie e individui intervistati.

**Prospetto 1 – Distribuzione regionale dei comuni, delle famiglie e degli individui nell'universo e nel campione - Anno 2014**

REGIONI	Comuni		Famiglie		Individui	
	Campione	Universo	Campione	Universo (a)	Campione	Universo (a)
Piemonte / Valle d'Aosta - Vallée d'Aoste	46	1.280	1.553	2.075	3.390	4.528
Liguria	17	235	724	782	1.548	1.579
Lombardia	45	1.544	1.511	4.394	3.427	9.919
Trentino-Alto Adige	18	333	519	443	1.270	1.041
Veneto	33	581	1.236	2.048	3.163	4.889
Friuli-Venezia Giulia	21	218	820	561	1.881	1.219
Emilia-Romagna	31	348	1.014	1.988	2.400	4.419
Toscana	33	287	828	1.638	1.850	3.733
Umbria	15	92	440	381	1.038	891
Marche	21	239	806	644	2.079	1.545
Lazio	24	378	997	2.629	2.350	5.828
Abruzzo	21	305	516	558	1.251	1.329
Molise	12	136	529	131	1.162	313
Campania	31	551	1.325	2.148	3.244	5.855
Puglia	27	258	1.146	1.577	2.994	4.078
Basilicata	14	131	612	232	1.500	576
Calabria	22	409	887	793	1.989	1.973
Sicilia	27	390	819	2.033	2.188	5.075
Sardegna	15	377	522	712	1.224	1.657
<b>Italia</b>	<b>473</b>	<b>8.092</b>	<b>16.804</b>	<b>25.768</b>	<b>39.948</b>	<b>60.448</b>

(a) Dati in migliaia.

## 4. La rilevazione e il trattamento dei dati

La rilevazione è condotta con due diverse tecniche utilizzate nelle tre fasi di raccolta dati:

- l'intervista iniziale diretta condotta dal rilevatore in modalità CAPI (*Computer Assisted Personal Interview*) attraverso la quale vengono rilevate le caratteristiche dei componenti della famiglia di fatto e dell'abitazione in cui vive, nonché alcune spese periodiche per l'abitazione;
- l'autocompilazione del diario cartaceo, sul quale la famiglia registra le spese alimentari e quelle per beni e servizi di largo consumo per un periodo di 14 giorni;
- l'intervista finale diretta condotta dal rilevatore in modalità CAPI per rilevare le altre spese familiari meno frequenti o eccezionali.

Nell'attuale disegno di indagine, particolare attenzione è stata dedicata alle modalità di raccolta del dato al fine di catturare meglio il comportamento di spesa di ciascuna famiglia e fornire stime di qualità a livello micro e macro. Sono stati pertanto ampliati i periodi di riferimento delle spese ed è stata adottata la più recente classificazione europea dei beni e servizi oggetto di consumo (*Classification of Individual COntsumption by Purpose*). Sono state inserite nuove voci di spesa e, soprattutto, è stata utilizzata una disaggregazione più dettagliata delle voci preesistenti: il numero delle voci di spesa è passato da circa 260 a circa 480. Le modifiche sostanziali introdotte in tutte le fasi del processo (Grassi, Pannuzi (a cura di); in corso di stampa) hanno ovviamente comportato differenze nelle stime: gli attuali capitoli di spesa differiscono, sia per livello sia per composizione, da quelli pubblicati fino al 2013. Per tale motivo i confronti temporali possono essere effettuati esclusivamente con i dati ricostruiti in serie storica e non con quelli precedentemente rilasciati.

Il nuovo disegno campionario (e il sistema di ponderazione), permette di analizzare anche i comportamenti di consumo delle famiglie che risiedono in aree metropolitane e di quelle con componenti stranieri.

L'indagine rileva le spese con periodi di riferimento molto diversi; il calcolo della spesa totale richiede dunque una standardizzazione temporale del dato rilevato, ovvero un riferimento temporale comune (il mese), in modo che tutte le spese possano essere sommate tra loro.

In particolare, le spese rilevate tramite diario a 14 giorni sono state riportate al mese applicando un coefficiente ottenuto come rapporto tra il numero medio di giorni nel mese e il numero di giorni di rilevazione e pari a 2,14. Per le spese rilevate con periodo di riferimento superiore al mese, la spesa è stata divisa per il numero di mesi compresi nel periodo di rilevazione (ad esempio, le spese rilevate con riferimento a tre mesi sono state divise per tre, mentre quelle rilevate con riferimento a dodici mesi per dodici).

## **5. La metodologia di calcolo dei pesi campionari**

L'ampliamento dei contenuti informativi ha consentito infine di migliorare la fase di produzione delle stime, sia per quanto concerne il sistema di ponderazione – con l'inserimento di un più ampio numero di vincoli e la revisione dei metodi di trattamento degli outliers – sia per quanto riguarda le metodologie di stima, con particolare riferimento

al metodo di valorizzazione monetaria degli autoconsumi e al periodo di riferimento per la diffusione delle stime.

Dato il carattere campionario dell'indagine, per poter ottenere stime per l'intera popolazione di riferimento (le famiglie residenti) è necessario utilizzare opportuni coefficienti di riporto all'universo. Il valore di tale coefficiente indica il numero di unità nella popolazione rappresentate dall'unità campionaria. L'indagine produce stime trimestrali e stime annuali.

I coefficienti di riporto all'universo per i dati annuali non si ottengono come semplice media aritmetica dei pesi trimestrali, ma tengono conto degli aggiornamenti anagrafici sulla popolazione residente.

Il calcolo dei coefficienti di riporto all'universo, siano essi trimestrali o annuali, utilizza una procedura generalizzata basata 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, che siano correlati alle principali variabili oggetto di indagine.

La strategia adottata per la costruzione dei coefficienti di riporto all'universo è simile a quella utilizzata dalle varie indagini campionarie sulle famiglie dell'Istituto e si sviluppa in tre fasi:

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

## 5.1 La probabilità di inclusione e il peso diretto

Il principio su cui è basato ogni metodo di stima campionaria è che le unità appartenenti al campione rappresentino anche le unità della popolazione non incluse nel campione stesso. Per questo motivo ad ogni unità campionaria è possibile attribuire un peso, il coefficiente di riporto all'universo, che indica quante unità della popolazione essa rappresenta.

Senza perdere di generalità, definiamo la seguente simbologia:

$U$  popolazione di riferimento oggetto di indagine;

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

$y_j$  valore della variabile Y assunto dalla j-esima osservazione della popolazione;

$\pi_j$  probabilità, assegnata dal disegno di campionamento, che l'unità j-esima sia inclusa nel campione S;

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

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

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 disegno di campionamento dell'Indagine sulle spese, la probabilità di inclusione di un generico individuo è data dalla probabilità di estrazione del suo comune di residenza (direttamente proporzionale all'ampiezza demografica dei comuni all'interno dello strato) e dalla probabilità di estrazione della sua famiglia tra le famiglie eleggibili del comune. Per la natura del disegno campionario, la probabilità di inclusione, e dunque il peso base, è uguale per tutte le famiglie appartenenti allo stesso comune.

Per una generica famiglia eleggibile  $j$ , nel comune  $i$  dello strato  $h$ , il peso diretto  $d_{hij}$ , inverso della probabilità di inclusione  $\pi_{hij}$ , assume la seguente forma:

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

dove :

$h$  denota l'indice di strato;

$i$  è l'indice di comune;

$j$  denota l'indice della famiglia;

$P_h$  indica il totale della popolazione residente nello strato  $h$ ;

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

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

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

I pesi base vengono di volta in volta calcolati sui dati anagrafici più aggiornati per popolazione e famiglie residenti.

## 5.2 La correzione per mancata risposta

Nel corso della fase di raccolta delle informazioni, come sempre accade nelle indagini, alcune unità campionarie non partecipano (per irreperibilità, rifiuto, interruzione definitiva o altri motivi). Tale fenomeno, noto come mancata risposta totale, implica che la numerosità campionaria teorica e quella effettivamente raggiunta siano diverse. Nella fase di calcolo dei coefficienti di riporto all'universo si tiene conto di tale fenomeno e viene introdotto un correttore per mancata risposta totale: sotto l'ipotesi che in determinati gruppi (ottenuti come aggregazioni di strati appartenenti alla stessa regione e tipologia comunale) il comportamento dei rispondenti sia simile a quello dei non rispondenti, il correttore assume la forma dell'inverso del tasso di risposta ( $\delta_g$ ):

$$\frac{1}{\delta_g} = \frac{m_g}{m_g^r} \quad (4)$$

in cui  $m_g^r$  rappresenta il numero di famiglie rispondenti nel determinato gruppo  $g$ .

Il coefficiente di riporto all'universo, corretto per mancata risposta, per una generica famiglia eleggibile  $j$ , nel comune  $i$  dello strato  $h$ , del gruppo  $g$  (dato come intersezione tra regione e tipologia comunale), risulta pertanto essere:

$$k_{ghij} = d_{hij} \frac{1}{\delta_g} = \frac{P_h}{P_{hi}} \frac{M_{hi}}{m_{hi}} \frac{m_g}{m_g^r} \quad (5)$$

## 5.3 La calibrazione a fonti esterne

Per il calcolo dei coefficienti di riporto all'universo finali si adottano gli stimatori di ponderazione vincolata (*calibration estimator*). La metodologia si basa sull'utilizzo di informazioni ausiliarie, sintetizzate in totali noti, che, essendo correlate con le variabili principali oggetto di indagine, hanno la funzione di aumentare l'accuratezza delle stime. I pesi finali si ottengono risolvendo un problema di minimo vincolato, in cui la funzione da minimizzare è una funzione di distanza tra i pesi diretti corretti per la mancata risposta ( $\mathbf{k}$ ) e i pesi finali ( $\mathbf{w}$ ) delle famiglie del campione rispondente ( $S^r$ ); i vincoli sono proprio le



condizioni di uguaglianza tra le stime campionarie delle variabili ausiliarie e i rispettivi totali noti desunti da fonti esterne all'indagine<sup>2</sup>. Formalmente:

$$\begin{cases} \text{Min} \left\{ \sum_{j \in S^*} \text{dist}(k_j, w_j) \right. \\ \left. \sum_{j \in S^*} x_j * w_j = t \right. \end{cases}$$

dove  $t$  è il vettore dei totali noti e  $x_j$  è il vettore delle variabili ausiliarie osservate sulla  $j$ -esima unità campionaria appartenente al campione rispondente ( $S^*$ ). La funzione di distanza utilizzata è la logaritmica troncata.

I totali noti, introdotti come vincoli nel calcolo dei pesi finali, consentono di migliorare l'accuratezza delle stime: quanto più le variabili ausiliarie considerate sono correlate con le variabili oggetto d'indagine, tanto più si riduce la distorsione delle stime. Nello specifico dell'Indagine sulle spese delle famiglie, le stime campionarie sono state vincolate ai seguenti totali noti:

1. Popolazione residente per ripartizione, sesso e classi di età (0-14, 15-29, 30-44, 45-59, 60-74, 75 e più);
2. Popolazione residente per regione;
3. Famiglie residenti per regione;
4. Popolazione residente per ripartizione e tipologia di comune (Comuni metropolitani, comuni periferia dei comuni metropolitani e comune oltre 50.000 abitanti, altri comuni);
5. Famiglie residenti per ripartizione e tipologia di comune (Comuni metropolitani, comuni della cintura metropolitana e comuni con più di 50.000 abitanti, altri comuni);
6. Popolazione straniera residente per ripartizione e sesso;
7. Popolazione di 15 anni e più per ripartizione, condizione e posizione nella professione (Dirigenti, quadri e impiegati; operai e assimilati; imprenditori e liberi professionisti; lavoratori in proprio e altri autonomi; disoccupati; inattivi);
8. Popolazione per ripartizione e mese;
9. Famiglie per ripartizione e mese.

---

<sup>2</sup> La calibrazione è una calibrazione integrata, vale a dire che i coefficienti di riporto all'universo sono stati determinati in modo da poter essere utilizzati, indifferentemente, per la costruzione delle stime relative alle famiglie e agli individui. Tutti gli individui di una stessa famiglia hanno il medesimo coefficiente di riporto della famiglia stessa.

I vincoli ai punti 1, 2 e 3 sono gli stessi utilizzati nella vecchia indagine sui consumi delle famiglie (al netto delle classi 30-44 anni e 45-59 anni che prima erano unite in un'unica classe d'età).

I vincoli ai punti 4 e 5 sono stati inseriti per rispettare il nuovo disegno di campionamento e considerare i differenti livelli e stili di spesa associati alle diverse tipologie di comune di residenza.

Il vincolo al punto 6 è stato inserito per tener conto del peso che, nel corso del tempo, ha assunto la popolazione straniera sul totale della popolazione residente, e del fatto che questa sottopopolazione ha stili e livelli di consumo generalmente differenti<sup>3</sup>.

Il vincolo al punto 7 è stato introdotto in considerazione del fatto che la condizione e la posizione professionale impattano sui livelli di reddito, e quindi di spesa, della popolazione. Questo totale è desunto dalla distribuzione proveniente dall'indagine Istat sulle forze di lavoro.

I vincoli 8 e 9 permettono di ripartire uniformemente la popolazione e le famiglie nei diversi mesi (tre mesi nel trimestre o dodici nell'anno), al fine di tenere conto del carattere fortemente stagionale di alcune tipologie di spesa.

I vincoli imposti sono gli stessi sia per le stime trimestrali che per le stime annuali. La differenza è che per le stime annuali le informazioni sui totali noti sono più aggiornate. Questo vale sia per le informazioni anagrafiche su popolazione e famiglie, ma anche per i totali provenienti dall'Indagine sulle forze di lavoro. Per motivi di disponibilità temporale, infatti, nelle stime trimestrali i vincoli provenienti dall'Indagine sulle forze di lavoro sono sfasati di un trimestre. Ad esempio, per il secondo trimestre di un determinato anno dell'Indagine sulle spese si fa riferimento al primo trimestre dello stesso anno dell'Indagine sulle forze di lavoro. Questo non vale per le stime annuali, per le quali si utilizza come totale noto la media d'anno dell'Indagine sulle forze di lavoro.

Si è adottata una procedura di individuazione delle osservazioni anomale, vale a dire osservazioni con livelli e comportamenti di spesa che, seppur veri, sono molto lontani o dissimili dal collettivo socio-economico al quale appartengono. Tali unità campionarie rappresentano quindi un numero ristretto di famiglie dell'universo ed è quindi opportuno che abbiano associato un coefficiente di riporto abbastanza basso per non influenzare la stima dal punto di vista sia della distorsione che della varianza. I gruppi socio-economici all'interno dei quali individuare comportamenti di spesa anomali sono stati definiti dall'incrocio tra ripartizione di appartenenza, tipologia di comune e numerosità familiare;

---

<sup>3</sup> Senza vincolo, la popolazione straniera risulterebbe ampiamente sottostimata. Per diversi motivi, infatti, le famiglie con stranieri hanno generalmente tassi di risposta più bassi rispetto alle famiglie di soli italiani.

all'interno di ogni gruppo, le famiglie anomale vengono individuate tramite una *cluster analysis* che considera sia il livello complessivo di spesa che la sua composizione tra i diversi capitoli. Alle famiglie individuate come anomale viene attribuito, nella fase precedente alla calibrazione finale, un peso corretto per la mancata risposta più basso rispetto a quello che avrebbero avuto tramite la procedura generalizzata. Questa procedura, permette di mantenere nel campione le osservazioni anomale e rende possibile l'uso degli stimatori classici. Il numero di osservazioni anomale è comunque limitato, in media meno di 50 famiglie a trimestre.

L'introduzione di nuovi vincoli di popolazione e la minore numerosità campionaria effettiva avrebbero comportato un campo di variazione dei coefficienti di riporto più ampio. Per non avere famiglie con coefficienti troppo elevati è stata quindi adottata una procedura di calibrazione ricorsiva, individuando le famiglie con valori dei coefficienti esterni a un intervallo di accettazione (inferiori al quinto percentile o superiori al 95esimo) e imponendo a queste valori dei coefficienti pari ai limiti del suddetto intervallo. La procedura di calibrazione viene nuovamente eseguita in modo da rispettare i vincoli imposti.

## **6. La diffusione dei risultati dell'indagine**

- Le stime riferite all'anno precedente vengono diffuse nel mese di luglio, nella Statistica Report "La spesa per consumi delle famiglie".
- Tutti i risultati dell'indagine vengono diffusi attraverso il datawarehouse dell'Istituto I.Stat (<http://dati.istat.it>) nella sezione Condizioni economiche delle famiglie e disuguaglianze.
- I microdati sono disponibili, su richiesta soggetta ad autorizzazione, secondo diverse modalità:
  - File ad uso pubblico, scaricabili direttamente dal sito Istat;
  - File per la ricerca, rilasciabili a studiosi di università o enti di ricerca a seguito della presentazione di un progetto di ricerca;
  - File per il Sistan, rilasciabili esclusivamente agli uffici di statistica del Sistema Statistico Nazionale.
  - File standard, rilasciabili a seguito di richiesta motivata, con finalità di studio (disponibili fino al 2013);
  - Laboratorio per l'Analisi dei Dati ELEMENTARI (ADELE), ricercatori e studiosi possono accedere ai file di microdati di tutte le indagini dell'Istat, nonché a

particolari file risultanti dall'integrazione di dati provenienti da più indagini (File integrati). Anche l'accesso al Laboratorio ADELE è gratuito.

## 7. Glossario

1.

**Classificazione dei consumi individuali secondo lo scopo (Coicop):** è una classificazione armonizzata a livello internazionale delle voci di spesa secondo lo scopo (Classification of Individual Consumption According by Purpose). Predisposta dalla Divisione statistica delle Nazioni unite per consentire il confronto dei comportamenti di consumo tra paesi, la Coicop è costituita da 14 capitoli di spesa, a loro volta articolati in categorie e in gruppi.

**Consumi delle famiglie:** l'insieme delle spese sostenute per tutti i beni e servizi destinati al soddisfacimento dei bisogni familiari e individuali. Include: i) la stima del valore del canone mensile delle abitazioni di proprietà a disposizione della famiglia comprensive delle pertinenze (fitti figurativi) in quanto rappresentano il costo del servizio "abitare" che la famiglia consuma nel corso del tempo; ii) il valore monetario dei prodotti autoconsumati dalla famiglia, provenienti cioè dalla propria impresa (individuale o familiare), dalla coltivazione diretta (inclusa la raccolta di prodotti spontanei), dall'allevamento (incluse la caccia e la pesca) o ricevuti come compenso per la propria attività lavorativa; iii) i beni e servizi, acquistati o autoprodotti, regalati a persone esterne alla famiglia. Esclude spesa per consumi tutte quelle spese sostenute dalle famiglie per i) attività professionale; ii) trasferimenti o donazioni; iii) accrescimento del patrimonio; iv) tasse e tributi.

**Consumi finali:** rappresentano il valore dei beni e servizi impiegati per soddisfare direttamente i bisogni umani, siano essi individuali o collettivi. Sono utilizzati due concetti: la spesa per consumi finali e i consumi finali effettivi. La differenza fra i due concetti sta nel trattamento riservato ad alcuni beni e servizi che sono finanziati dalle amministrazioni pubbliche o dalle istituzioni senza scopo di lucro al servizio delle famiglie, ma che sono forniti alle famiglie come trasferimenti sociali in natura; questi beni sono compresi nel consumo effettivo delle famiglie, mentre sono esclusi dalla loro spesa finale (Sistema europeo dei conti, Sec 2010).

**Fitto figurativo:** è una componente non-monetaria del reddito delle famiglie o della spesa per consumi delle famiglie che vivono in case di loro proprietà; rappresenta il costo che queste dovrebbero sostenere per prendere in affitto, ai prezzi vigenti sul mercato

immobiliare, un'unità abitativa con caratteristiche identiche a quella in cui vivono (al netto delle spese di condominio, riscaldamento, accessorie e con riferimento a una casa non ammobiliata). Negli studi sulla povertà e sulla distribuzione del reddito, il concetto viene esteso anche alle famiglie in usufrutto o in uso gratuito e agli inquilini con affitti agevolati, cioè inferiori ai prezzi di mercato.

**Incidenza della povertà:** si ottiene dal rapporto tra il numero di famiglie con spesa media mensile per consumi pari o al di sotto della soglia di povertà e il totale delle famiglie residenti.

**Paniere di povertà assoluta:** rappresenta l'insieme dei beni e servizi che, nel contesto italiano, vengono considerati essenziali per una determinata famiglia per conseguire uno standard di vita minimamente accettabile.

**Persona di riferimento:** Persona rispetto alla quale sono definite le relazioni di parentela, generalmente corrispondente all'intestatario della scheda anagrafica familiare.

**Scala di equivalenza:** sono utilizzate nelle indagini statistiche per uniformare unità di analisi eterogenee (esempio consumi e redditi delle famiglie) mediante l'utilizzo di appropriati coefficienti di correzione. Nell'indagine sui consumi delle famiglie i coefficienti sono utilizzati per determinare la soglia di povertà quando le famiglie hanno un numero di componenti diverso da due. La soglia di povertà per una famiglia di una persona è pari a 0,60 volte quella di 2 persone, per una famiglia di 3 persone il coefficiente è pari a 1,33, per quattro persone a 1,63, per cinque a 1,90, per una famiglia di sei persone è pari a 2,16, per una di 7 persone o più è pari a 2,40 (scala Carbonaro – Prospetto 1). La scala di equivalenza in uso nelle indagini EU-Silc, come da Regolamento comunitario, è la cosiddetta Ocse modificata. Essa assegna il valore di 1 al primo componente adulto del nucleo familiare; si aggiunge 0,5 per ogni adulto in più e 0,3 per ciascun minore (individui di età inferiore ai 14 anni) presente nella famiglia.

**Prospetto 1– Scala di equivalenza Carbonaro**

Ampiezza della famiglia	Scala di equivalenza
1	0,60
2	1,00
3	1,33
4	1,63
5	1,90
6	2,16
7 o più	2,40

**Soglia di povertà assoluta:** rappresenta la spesa minima necessaria per acquisire i beni e servizi inseriti nel paniere di povertà assoluta. La soglia di povertà assoluta varia, per costruzione, in base alla dimensione della famiglia, alla sua composizione per età, alla ripartizione geografica e alla dimensione del comune di residenza.

**Soglia di povertà relativa:** per una famiglia di due componenti è pari alla spesa media procapite nel Paese.

**Spesa equivalente:** è calcolata dividendo il valore familiare della spesa per il coefficiente della scala di equivalenza e permette di rendere direttamente confrontabili i livelli di spesa di famiglie di ampiezza diversa.

**Spesa media mensile:** è calcolata dividendo la spesa totale per il numero delle famiglie residenti in Italia.

**Spesa mediana mensile:** è il valore di spesa per consumi che divide la distribuzione di frequenza in due parti uguali (il 50% delle famiglie presenta un valore di spesa per consumi inferiore o pari alla mediana, il 50% un valore superiore). Poiché la spesa per consumi ha una distribuzione asimmetrica, la mediana risulta sempre inferiore al valore medio.

**Spesa media per persona (procapite):** si ottiene dividendo la spesa totale per consumi delle famiglie per il numero totale dei componenti.

**Tipo di comune:** *area metropolitana* - comuni con oltre 250.000 abitanti; *grandi comuni* - comuni della periferia dell'area metropolitana e comuni con 50.000 abitanti e più; *piccoli comuni* - comuni con meno di 50.000 abitanti.

## 8. Riferimenti bibliografici

Grassi D., Pannuzi N. (a cura di). In corso di stampa. "La nuova indagine sulle spese per consumi in Italia". Roma: Istat (collana Metodi).

Istat. 2015. La spesa per consumi delle famiglie, Anno 2014. Roma: Istat (Statistica Report, 8 luglio 2015).

Istat. 2015. La povertà in Italia, Anno 2014. Roma: Istat (Statistica Report, 15 luglio 2015).

## 9. Contatti

Per informazioni riguardanti la rilevazione è possibile contattare:

Sabrina Barcherini e-mail: [sabrina.barcherini@istat.it](mailto:sabrina.barcherini@istat.it)  
Donatella Grassi e-mail: [donatella.grassi@istat.it](mailto:donatella.grassi@istat.it)  
Ilaria Arigoni e-mail: [ilaria.arigoni@istat.it](mailto:ilaria.arigoni@istat.it)  
Claudio Falorni e-mail: [claudio.falorni@istat.it](mailto:claudio.falorni@istat.it)

Link alla rilevazione:

<http://www.istat.it/it/archivio/71980>

## **Curatori dei capitoli**

Il documento è stato curato da Sabrina Barcherini.

Si devono a:

Andrea Cutillo i paragrafi 5, 5.1, 5.2 e 5.3

Claudia De Vitiis, Monica Russo e Marco D. Terribili i paragrafi 3, 3.1 e 3.2.

**Descrizione del file**  
*File description*





**File ad uso pubblico  
micro.STAT**

## **Indagine sulle spese delle famiglie**

Periodo di riferimento: anno 2014

### **Descrizione del file**

# INDICE

Introduzione .....	3
Le misure di protezione adottate.....	4
Riferimenti bibliografici.....	6

## Introduzione

I file di microdati ad uso pubblico (micro.STAT) sono collezioni di dati elementari, liberamente scaricabile via web (<http://www.istat.it/it/archivio/microdati-ad-uso-pubblico>), per le quali, a partire dal 2014, sono stati sviluppati anche i corrispondenti file per la ricerca (MFR<sup>1</sup>). Il contenuto informativo di questi ultimi è superiore rispetto a quello del file ad uso pubblico<sup>2</sup>.

Per la predisposizione del file micro.STAT relativo all'Indagine sulle spese delle famiglie, periodo di riferimento 2014, è stata adottata una opportuna metodologia, allo scopo di limitare il rischio di violazione della riservatezza. Gli scenari di intrusione considerati sono stati:

- identificazione attraverso archivi esterni, ovvero attraverso il collegamento con i dati rilasciati da altre fonti pubbliche;
- identificazione spontanea, ovvero conseguente a conoscenze *a priori* dell'utente che potrebbero permettere di attribuire correttamente i dati rilasciati alle unità della popolazione rilevata.

Le variabili coinvolte nel processo di protezione sono quelle che possono consentire l'associazione tra le informazioni e i rispondenti, ossia:

- gli *identificativi diretti*, che individuano in maniera univoca le unità di *rilevazione/analisi* (come ad esempio indirizzo e codice fiscale);
- gli *identificativi indiretti*, o *variabili chiave* (come ad esempio il numero di componenti della famiglia), che permettono di circoscrivere la popolazione cui appartengono i rispondenti e che, da sole o in combinazione con altre, possono portare alla re-identificazione di uno o più record.

Mentre i primi vengono cancellati dal file, i secondi vengono trattati allo scopo di limitare il rischio di violazione della riservatezza. Per il significato delle variabili e delle rispettive modalità si rinvia al questionario di indagine contenuto nella cartella metadati.

---

<sup>1</sup> MFR è l'acronimo per il file per la ricerca (Microdata File for Research). La documentazione metodologica di tali file è consultabile sul sito Istat (<http://www.istat.it/it/archivio/file-per-la-ricerca>).

<sup>2</sup> I ricercatori, che necessitano di informazioni maggiormente dettagliate, possono, su richiesta motivata e previa autorizzazione del Presidente dell'Istituto, richiedere il file di microdati per la ricerca.

## Le misure di protezione adottate

Per la predisposizione del file ad uso pubblico per l'Indagine sulle spese delle famiglie, periodo di riferimento 2014, sono state adottate le misure di protezione appresso specificate.

### *Soppressione di variabili*

1. Codice famiglia
2. Codice fiscale dei componenti la famiglia
3. Percentuale di reddito percepito dai componenti la famiglia sul totale dei redditi familiari: sostituita dal principale percettore
4. Comune e provincia di residenza
5. Tipologia di comune

### *Ricodifiche di variabili categoriali o quantitative discrete*

1. Età dei componenti la famiglia: [0,18), [18,35],[35,65],[65 e oltre)
2. Numero di componenti: top-coding dal sesto componente (6 e più componenti)
3. Relazione di parentela dei componenti la famiglia
  - 1=persona di riferimento (PR)
  - 2=coniuge/convivente di PR
  - 3=figlio di PR
  - 4=altra relazione
4. Titolo di studio dei componenti la famiglia
  - 1=nessun titolo
  - 2=scuola primaria (elementari)
  - 3=scuola secondaria di primo grado
  - 4= scuola secondaria di secondo grado
  - 5=laurea e post laurea
5. Attività economica dei componenti la famiglia
  - 1=settore primario
  - 2=settore secondario
  - 3=settore terziario
6. Posizione nella professione dei componenti la famiglia
  - 1=dirigenti e impiegati
  - 2=operai e assimilati
  - 3=imprenditori e liberi professionisti

- 4=lavoratori in proprio
7. Fonte principale di reddito dei componenti la famiglia
- 1=reddito da lavoro
  - 2=reddito da pensione
  - 3=altri redditi
8. Condizione autodichiarata dei componenti la famiglia
- 1= occupato
  - 2= in cerca di occupazione
  - 3=casalinga/studente
  - 4=persona ritirata dal lavoro
  - 5=in altra condizione
9. Professione dei componenti la famiglia
- 1=imprenditori e alta dirigenza
  - 2=professioni della conoscenza
  - 3=professioni esecutive di ufficio
  - 4=professioni qualificate nel commercio e nei servizi
  - 5=professioni manuali qualificate
  - 6=professioni manuali non qualificate
  - 7=forze armate

*Ricodifiche di variabili quantitative continue*

1. Raggruppamento e top-coding delle variabili sulle spese delle famiglie relative a variabili sensibili:
- spesa per occhiali e lenti da vista, spesa per apparecchi acustici, spesa per protesi dentarie e apparecchi ortodontici, spesa per sedie e carrozzine e spesa per riparazione di attrezzature sanitarie sono state raggruppate insieme;
  - spesa per medico generico (certificati, vaccinazioni o altre prestazioni a pagamento), spesa per visite specialistiche, spesa per ricoveri in strutture pubbliche; e spesa per ricoveri in strutture private sono state raggruppate insieme;
  - spesa per analisi cliniche e esami diagnostici (analisi del sangue, radiografie, ecografie, elettrocardiogrammi, ecc.), spesa per infermieri e altri servizi paramedici (fisioterapisti, podologi, logopedisti, agopunturisti), spesa per cure termali, ginnastica correttiva e spesa per noleggio di attrezzature sanitarie o servizi di ambulanza sono state raggruppate insieme.

2. Per tutte le variabili relative alle spese sanitarie, sulla base dei box plot modificati per l'asimmetria (Hubert M. Vandervieren E., 2008), è stato effettuato un top-coding.

### *Soppressioni locali*

Sono stati inseriti valori mancanti in corrispondenza di tutte le variabili individuali dal settimo componente in poi. A causa delle soppressioni locali si possono verificare scostamenti rispetto ai dati pubblicati dall'Istat.

## **Riferimenti bibliografici**

- Hubert, M., Vandervieren, E. (2008). An adjusted boxplot for skewed distributions. *Computational Statistics and Data Analysis* 52, 5186-5201.
- Hundepool, A., Domingo-Ferrer, J., Franconi, L., Giessing, S., Nordholt, E. S., Spicer, K. e de Wolf, P.-P. (2012). *Statistical Disclosure Control*. Wiley.
- Willenborg, L. e de Waal, T. (1996). *Statistical Disclosure Control in Practice*. Lecture Notes in Statistics, 111, New York: Springer-Verlag.
- Willenborg, L. e de Waal, T. (2000). *Elements of statistical disclosure control*. Lecture Notes in Statistics, 115, New York: Springer-Verlag.

## **Curatori**

Il documento è stato redatto da

Sabrina Barcherini

Flavio Foschi

Luigi Virgili