

SN229

Aspetti della vita quotidiana (2020)

Aspects of daily life (2020)

Istat

Versione/Version: 1.0 - 12/05/2022



UniData

Bicocca Data Archive

Website: www.unidata.unimib.it

E-mail: 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
Survey's methodological issues

p. 3

Descrizione del file
File description

p. 26

Aspetti metodologici dell'indagine
Survey's methodological issues



File di microdati per la ricerca

Aspetti della vita quotidiana
Periodo di riferimento: anno 2020

Aspetti metodologici dell'indagine

INDICE

1. Introduzione	3
2. Obiettivi conoscitivi	3
3. Strategia di campionamento	4
4. La rilevazione e il trattamento dei dati	7
5. La metodologia di calcolo dei pesi campionari	9
6. Valutazione del livello di precisione delle stime	12
7. La diffusione dei risultati dell'indagine	21
8. Riferimenti bibliografici.....	22
9. Contatti	22

1. Introduzione

L'indagine campionaria “Aspetti della vita quotidiana” fa parte del sistema integrato di Indagini Multiscopo sulle famiglie avviato dal 1993 con l'obiettivo di produrre informazioni sugli individui e sulle famiglie. L'indagine viene svolta ogni anno e le informazioni raccolte consentono di conoscere le abitudini dei cittadini e i problemi che essi affrontano ogni giorno. Aree tematiche variegata si susseguono nei questionari, permettendo di capire come vivono gli individui e se sono soddisfatti del funzionamento di quei servizi di pubblica utilità che devono contribuire al miglioramento della qualità della vita. Scuola, lavoro, vita familiare e di relazione, abitazione e zona in cui si vive, tempo libero, partecipazione politica e sociale, salute, stili di vita e rapporto con i servizi sono indagati in un'ottica in cui oggettività dei comportamenti e soggettività delle aspettative, delle motivazioni, dei giudizi contribuiscono a definire l'informazione sociale.

L'indagine rientra tra quelle comprese nel [Programma statistico nazionale](#), che raccoglie l'insieme delle rilevazioni statistiche necessarie al Paese.

2. Obiettivi conoscitivi

La popolazione di interesse dell'indagine multiscopo “Aspetti della vita quotidiana”, ossia l'insieme delle unità statistiche intorno alle quali si intende investigare, è costituita dalle famiglie residenti in Italia e dai membri che le compongono; sono pertanto esclusi i membri permanenti delle convivenze. La famiglia è intesa come famiglia di fatto, ossia un insieme di persone coabitanti e legate da vincoli di matrimonio, parentela, affinità, adozione, tutela o affettivi.

Il periodo di riferimento è prevalentemente costituito dai dodici mesi che precedono l'intervista, anche se per alcuni quesiti il riferimento è al momento dell'intervista.

I domini 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, Italia nord-orientale, Italia centrale, Italia meridionale, Italia insulare);
- le regioni geografiche (ad eccezione del Trentino-Alto Adige le cui stime sono prodotte separatamente per le province di Bolzano e Trento);
- la tipologia comunale ottenuta suddividendo i comuni italiani in sei classi formate in base a caratteristiche socio-economiche e demografiche:
 - A) comuni appartenenti all'area metropolitana suddivisi in:

- A1, comuni centro dell’area metropolitana: Torino, Milano, Venezia, Genova, Bologna, Firenze, Roma, Napoli, Bari, Palermo, Catania, Cagliari;
- A2, comuni che gravitano intorno ai comuni centro dell’area metropolitana;
- B) comuni non appartenenti all’area metropolitana suddivisi in:
 - B1, comuni aventi fino a 2.000 abitanti;
 - B2, comuni con 2.001-10.000 abitanti;
 - B3, comuni con 10.001-50.000 abitanti;
 - B4, comuni con oltre 50.000 abitanti.

3. Strategia di campionamento

3.1. Descrizione generale del disegno di campionamento

Per l’indagine AVQ 2020 il campione è stato integrato con il disegno campionario seguito per il Master Sample del Censimento permanente. Nel caso specifico, i comuni campione per la corrente indagine sono stati individuati come sotto-campione del campione di 2850 comuni del Master Sample utilizzato per il 2018. A tale scopo, lo schema campionario classico utilizzato per le indagini sulle famiglie, di seguito descritto, è stato implementato sul sotto-universo dei comuni rilevati per il Censimento Permanente a ottobre del 2018.

Nell’ambito di ogni area ottenuta dall’incrocio delle regioni con le sei tipologie comunali di cui sopra, i comuni universo sono stati suddivisi in due sottoinsiemi: i comuni di maggiore dimensione demografica costituiscono strati a sé stante e vengono definiti Auto Rappresentativi (AR); i rimanenti comuni sono definiti Non Auto Rappresentativi (NAR) e sono suddivisi, sulla base della dimensione demografica, in strati di uguale ampiezza; da tali strati i comuni campione (due per ogni strato) sono stati selezionati con probabilità proporzionali alla loro dimensione.

Per ognuno dei comuni coinvolti nell’indagine (AR e NAR), viene effettuato un campionamento a grappoli: i grappoli - le famiglie - sono selezionati in maniera casuale dalla lista anagrafica e tutti i componenti che appartengono alla famiglia di fatto vengono sottoposti a rilevazione. La numerosità minima di famiglie campione per ciascun comune è stata posta pari a 24.

Le famiglie sono selezionate per ciascun comune campione a partire dal campione teorico selezionato per il Master Sample; per ogni famiglia inclusa nel campione vengono rilevate

le caratteristiche oggetto di indagine di tutti i componenti di fatto appartenenti alla famiglia medesima.

3.2. Definizione della dimensione campionaria

Per un'indagine ad obiettivi plurimi, come quella in esame, è poco realistico pensare di poter disegnare una strategia campionaria che assicuri prefissati livelli di precisione di tutte le stime prodotte. La questione è complicata dal fatto che l'indagine ha la finalità di determinare stime per livelli territoriali differenti, il che comporta l'adozione di soluzioni di tipo ottimale diverse e contrastanti. Ad esempio, se l'unico ambito territoriale di pubblicazione delle stime fosse quello nazionale, una soluzione approssimativamente ottimale sarebbe quella di determinare la numerosità nazionale e ripartirla tra le regioni in modo proporzionale alla loro dimensione demografica; viceversa, avendo la finalità di produrre stime con uguale attendibilità a livello regionale, una soluzione approssimativamente ottimale sarebbe quella di selezionare un campione uguale in tutte le regioni. Quest'ultima soluzione, però, è poco efficiente per le stime a livello nazionale. Per affrontare questo problema, conformemente a quanto fatto in altri paesi, si è fatto ricorso ad una strategia che perviene alla definizione della numerosità campionaria attraverso approssimazioni successive.

In base alle considerazioni precedenti si è deciso di adottare un'ottica mista basata sia su criteri di costo ed organizzativi, sia su una valutazione degli errori campionari delle principali stime a livello nazionale e con riferimento a ciascuno dei domini territoriali di interesse.

I criteri seguiti possono essere sintetizzati nei seguenti punti:

- la dimensione del campione teorico in termini di famiglie, prefissata a livello nazionale essenzialmente in base a criteri di costo ed operativi, è pari a circa 24.000 famiglie;
- il numero di comuni campione interessati non deve essere superiore a 900 in modo da consentire un buon lavoro di controllo e supervisione.

L'allocazione del campione di famiglie e di comuni tra le varie regioni è stata quindi calcolata adottando un criterio di compromesso tale da garantire sia l'affidabilità delle stime a livello nazionale sia quella delle stime a livello di ciascuno dei domini territoriali descritti nel precedente paragrafo.

3.3. 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 un comune campione nell'ambito di ciascuno strato definito sui comuni dell'insieme N_r ;
- 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 dominio territoriale individuato dalle aree A1, A2, B1, B2, B3 e B4 di ciascuna 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 A_r , 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 A_r e N_r : i comuni di dimensione superiore o uguale a ${}_r\lambda$ sono definiti come comuni A_r e i rimanenti come N_r ;
- suddivisione dei comuni dell'insieme N_r in strati aventi dimensione, in termini di popolazione residente, approssimativamente costante e all'incirca pari alla soglia ${}_r\lambda$.

Effettuata la stratificazione, i comuni A_r sono inclusi con certezza nel campione; per quanto riguarda, invece, i comuni N_r , nell'ambito di ogni strato viene estratto un comune campione con probabilità proporzionale alla dimensione demografica, mediante la procedura di selezione sistematica proposta da Madow ¹.

¹ Madow, W.G. "On the theory of systematic sampling II", *Annals of Mathematical Statistics*, 20, (1949): 333-354.

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 ordinate le 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:

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

Nel prospetto 1 viene riportata la distribuzione regionale dell'universo e del campione dei comuni, delle famiglie e degli individui.

Prospetto 1 – Distribuzione regionale dei comuni, delle famiglie e degli individui nell'universo e nel campione - Anno 2020

REGIONI	Comuni			Famiglie			Individui	
	Campione effettivo	Campione teorico	Universo master sample	Campione effettivo	Campione teorico	Universo (a)	Campione effettivo	Universo (a)
Piemonte	57	57	359	1375	1879	1986	3019	4291
Valle d'Aosta - Vallée d'Aoste	20	20	35	457	632	60	983	124
Lombardia	78	78	499	1655	2272	4368	3845	10013
Trentino-Alto Adige	44	44	108	1.065	1.444	455	2489	1065
<i>Bolzano - Bozen</i>	21	21	51	521	700	226	1206	527
<i>Trento</i>	23	23	57	544	744	229	1283	538
Veneto	51	51	206	1073	1379	2031	2526	4866
Friuli-Venezia Giulia	30	30	76	729	969	536	1624	1200
Liguria	25	25	78	810	1110	748	1679	1526
Emilia-Romagna	45	45	144	1065	1379	1952	2432	4430
Toscana	48	48	120	1023	1471	1649	2313	3691
Umbria	21	21	39	529	712	377	1244	873
Marche	36	36	82	751	966	640	1780	1504
Lazio	31	31	147	1147	1802	2602	2565	5832
Abruzzo	33	33	96	754	989	562	1717	1296
Molise	20	20	42	542	664	126	1262	298
Campania	51	51	220	1277	1627	2206	3330	5755
Puglia	46	46	125	1017	1275	1599	2432	3982
Basilicata	24	24	54	592	684	243	1362	552
Calabria	38	39	130	817	1092	815	1895	1909
Sicilia	49	49	174	1109	1492	1999	2692	4932
Sardegna	35	35	121	742	1022	719	1642	1617
Italia	782	783	2.855	18.529	24.860	25.673	42.831	59.756

(a) Stima Indagine multiscopo "Aspetti della vita quotidiana", dati in migliaia.

4. La rilevazione e il trattamento dei dati

La rilevazione, di tipo campionario, è condotta con cadenza annuale in genere nel mese di Marzo.

Il disegno di indagine prevede in generale che a tutte le famiglie campione sia proposta dapprima l'intervista via web e successivamente, alle famiglie non rispondenti, viene inviato il rilevatore per l'intervista.

L'intervista prevede l'utilizzo di due questionari: uno per intervista diretta e uno per autocompilazione.

Il primo è il questionario base della rilevazione che viene somministrato mediante intervista faccia a faccia. Questo modello è composto: da una “Scheda Generale”, in cui si rilevano le relazioni di parentela ed altre informazioni di natura socio-demografica e socio-economica relative ai componenti della famiglia; da quattro “Schede Individuali”, una per ciascun componente della famiglia e da un “Questionario familiare” che contiene quesiti familiari ai quali risponde un solo componente adulto.

Il secondo è un modello somministrato per autocompilazione. Il modello viene consegnato dal rilevatore a ciascun componente della famiglia e contiene quesiti che possono essere agevolmente compilati in autonomia dal rispondente anche senza l'intervento diretto del rilevatore. Dal 2019 la parte per intervista diretta, che in passato utilizzava un modello cartaceo, è stata realizzata mediante tecnica assistita da computer (CAPI), mentre il questionario per autocompilazione è stato somministrato in PAPI, per cui la tecnica è divenuta CAWI/CAPI-PAPI.

Le informazioni vengono fornite direttamente da tutti gli individui di 14 anni e più, mentre i bambini e i ragazzi al di sotto dei 14 anni vengono intervistati in modalità proxy, ciò significa che è un genitore o un componente maggiorenne a fornire le informazioni in loro vece. Taluni quesiti della rilevazione, per la sensibilità dell'argomento trattato, prevedono la facoltà di non rispondere.

Trattandosi di un'indagine con una componente PAPI, i questionari autocompilati sono sottoposti a registrazione. A conclusione della registrazione dei dati, o meglio contestualmente ad essa poiché la registrazione procede per lotti distinti di questionari, prende avvio la fase di controllo della qualità dei dati raccolti e di validazione degli stessi, che ha il duplice obiettivo di garantire la qualità delle stime prodotte e produrre un archivio di dati elementari privo di incoerenze.

Questi obiettivi vengono perseguiti attraverso un complesso e reiterato processo:

- di esplorazione dei dati, basato su una reportistica che ne evidenzia anomalie e incoerenze;

- di correzione delle incompatibilità rilevate tramite l'applicazione di opportuni interventi di correzione, sia deterministica, sia probabilistica.

Tutte le procedure di correzione sono poi valutate mediante analisi delle distribuzioni semplici e congiunte, con la determinazione dell'impatto delle procedure sulle stime finali, con le analisi di indicatori sulla frequenza di attivazione delle regole di compatibilità e di indicatori sulla frequenza di correzione per le variabili e con la valutazione delle tipologie di errore individuate (mancate risposte parziali, errori sistematici, errori casuali, valori anomali).

5. La metodologia di calcolo dei pesi campionari

Le stime prodotte dall'indagine sono di frequenze assolute e relative, riferite alle famiglie e agli individui o stime di totali di variabili quantitative. Sono ottenute mediante uno stimatore di ponderazione vincolata. 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. Questo principio viene realizzato attribuendo a ogni unità campionaria un peso che indica il numero di unità della popolazione rappresentata dall'unità medesima. Per esempio, se a un'unità campionaria viene attribuito un peso pari a 30, ciò indica che questa unità rappresenta se stessa e altre 29 unità della popolazione non incluse nel campione.

La procedura che consente di costruire i pesi finali da attribuire alle unità campionarie rispondenti, è articolata in generale nelle seguenti fasi:

- 1) si calcolano i pesi diretti come reciproco della probabilità di inclusione delle unità;
- 2) si calcolano i fattori correttivi per mancata risposta totale, come l'inverso del tasso di risposta in opportuni sottoinsiemi di unità e si ottengono i pesi base, o pesi corretti per mancata risposta totale, moltiplicando i pesi diretti per i corrispondenti fattori correttivi per mancata risposta totale;
- 3) si costruiscono i fattori correttivi che consentono di soddisfare, a livello regionale, la condizione di uguaglianza tra i totali noti di alcune variabili ausiliarie e le corrispondenti stime campionarie;
- 4) si calcolano, infine, i pesi finali mediante il prodotto dei pesi base per i fattori correttivi ottenuti al passo 3.

I fattori correttivi del passo 3 sono ottenuti dalla risoluzione di un problema di minimo vincolato, in cui la funzione da minimizzare è una funzione di distanza (opportunamente prescelta) tra i pesi base e i pesi finali e i vincoli sono definiti dalla condizione di uguaglianza

tra stime campionarie dei totali noti di popolazione e valori noti degli stessi. La funzione di distanza prescelta è la funzione logaritmica troncata; l'adozione di tale funzione garantisce che i pesi finali siano positivi e contenuti in un predeterminato intervallo di valori possibili, eliminando in tal modo i pesi positivi estremi (troppo grandi o troppo piccoli).

Le variabili ausiliarie considerate a livello regionale sono: tipologia comunale, classi di età, sesso, cittadinanza (italiani/stranieri), titolo di studio (totali stimati dall'indagine su Forze di Lavoro).

Per l'indagine in oggetto il calcolo dei pesi diretti del passo 1 ha tenuto conto del fatto che il campione è stato selezionato come sotto-campione del campione del Master Sample del Censimento della popolazione.

Per il calcolo dei pesi finali è stato utilizzato il software ISTAT ReGenesees² basato su R.

Al fine di rendere più chiara la successiva esposizione, introduciamo la seguente simbologia: d , indice di livello territoriale di riferimento delle stime; i , indice di comune; j , indice di famiglia; p , indice di componente della famiglia; h , indice di strato di comuni; y , generica variabile oggetto di indagine; Y_{hijp} , valore di y osservato sul componente p della famiglia j del comune i dello strato h ; P_{hij} , numero di componenti della famiglia j del comune

i dello strato h ; $Y_{hij} = \sum_{p=1}^{P_{hij}} Y_{hijp}$, totale della variabile y osservato sulla famiglia j del comune i

dello strato h ; M_{hi} , numero di famiglie residenti nel comune i dello strato h ; m_{hi} , campione di famiglie nel comune i dello strato h ; N_h , totale di comuni nello strato h ; n_h , numero di comuni campione nello strato h (nell'indagine in oggetto si ha $n_h=1$); H_d , numero totale di strati nel generico dominio territoriale d .

Ipotizziamo di voler stimare, con riferimento ad un generico dominio d , il totale della generica variabile y oggetto di indagine, espresso dalla seguente relazione

$$Y_d = \sum_{h=1}^{H_d} \sum_{i=1}^{N_h} \sum_{j=1}^{M_{hi}} Y_{hij} \quad (1)$$

La stima del totale (1) è data da

$$\hat{Y}_d = \sum_{h=1}^{H_d} \hat{Y}_h, \text{ essendo } \hat{Y}_h = \sum_{i=1}^{n_h} \sum_{j=1}^{m_{hi}} W_{hij} Y_{hij}, \quad (2)$$

² Per maggiori informazioni si veda <https://www.istat.it/it/metodi-e-strumenti/metodi-e-strumenti-it/elaborazione/strumenti-di-elaborazione/regenesees>.

in cui W_{hij} è il peso finale da attribuire a tutti i componenti della famiglia j del comune i dello strato h .

Dalla precedente relazione si desume, quindi, che per ottenere la stima del totale (1) occorre moltiplicare il valore della variabile y assunto da ciascuna unità campionaria per il peso di tale unità³ ed effettuare, a livello del dominio di interesse, la somma dei prodotti così ottenuti. Il peso da attribuire alle unità campionarie è ottenuto per mezzo di una procedura complessa che:

- corregge l'effetto distorsivo della mancata risposta totale dovuta all'impossibilità di intervistare alcune delle famiglie selezionate per irreperibilità o per rifiuto all'intervista;
- tiene conto della conoscenza di totali noti di importanti variabili ausiliarie (disponibili da fonti esterne all'indagine), nel senso che le stime campionarie dei totali noti delle variabili ausiliarie devono coincidere con i valori noti degli stessi.

Nell'indagine in oggetto vengono definiti per ciascuna regione geografica 24 totali noti, che si riferiscono alla distribuzione della popolazione regionale per sesso e otto classi di età⁴, della popolazione regionale nelle sei aree A_1, A_2, B_1, B_2, B_3 e B_4 e della popolazione straniera residente in Italia per regione e sesso. Indicando, quindi, con ${}_kX$ ($k=1, \dots, 24$) il totale noto della k -esima variabile ausiliaria per la generica regione geografica e con ${}_kX_{hij}$ il valore assunto dalla k -esima variabile ausiliaria per la famiglia rispondente h_{ij} , la condizione sopra descritta è espressa dalla seguente uguaglianza

$${}_kX = \hat{X}_k = \sum_{h=1}^H \sum_{i=1}^{n_h} \sum_{j=1}^{m_{hi}} W_{hij} {}_kX_{hij} \quad (k=1, \dots, 24)$$

in cui H indica il numero complessivo di strati definiti nella regione. Se, ad esempio, ${}_8X$ indica il numero di maschi di età maggiore o uguale a sessantacinque anni, la variabile ausiliaria ${}_8X_{hij}$ rappresenta il numero di maschi di età maggiore o uguale a sessantacinque anni della famiglia h_{ij} .

La procedura che consente di costruire i pesi finali da attribuire alle unità campionarie rispondenti, è articolata nelle seguenti fasi:

- 1) si calcolano i pesi diretti come reciproco della probabilità di inclusione delle unità;
- 2) si calcolano i fattori correttivi per mancata risposta totale, come l'inverso del tasso di risposta del comune cui ciascuna unità appartiene;

³ Al fine di ottenere stime coerenti per individui e famiglie i pesi finali sono definiti in modo tale che a ciascuna famiglia h_{ij} e a tutti i componenti della stessa sia assegnato un medesimo peso finale W_{hij} .

⁴ Le classi di età considerate sono: 0-5 anni, 6-13 anni, 14-24 anni, 25-34 anni, 35-44 anni, 45-54 anni, 55-64 anni, 65 anni e più.

- 3) si ottengono i pesi base, o pesi corretti per mancata risposta totale, moltiplicando i pesi diretti per i corrispondenti fattori correttivi per mancata risposta totale;
- 4) si costruiscono i fattori correttivi che consentono di soddisfare, a livello regionale, la condizione di uguaglianza tra i totali noti delle variabili ausiliarie e le corrispondenti stime campionarie;
- 5) si calcolano, infine, i pesi finali mediante il prodotto dei pesi base per i fattori correttivi ottenuti al passo 4.

I fattori correttivi del passo 4 sono ottenuti dalla risoluzione di un problema di minimo vincolato, in cui la funzione da minimizzare è una funzione di distanza (opportunamente prescelta) tra i pesi base e i pesi finali e i vincoli sono definiti dalla condizione di uguaglianza tra stime campionarie dei totali noti di popolazione e valori noti degli stessi. La funzione di distanza prescelta è la funzione logaritmica troncata; l'adozione di tale funzione garantisce che i pesi finali siano positivi e contenuti in un predeterminato intervallo di valori possibili, eliminando in tal modo i pesi positivi estremi (troppo grandi o troppo piccoli).

Tutti i metodi di stima che scaturiscono dalla risoluzione di un problema di minimo vincolato del tipo sopra descritto rientrano in una classe generale di stimatori nota come stimatori di ponderazione vincolata⁵. Un importante stimatore appartenente a tale classe, che si ottiene utilizzando la funzione di distanza euclidea, è lo stimatore di regressione generalizzata. Come verrà chiarito meglio in seguito, tale stimatore riveste un ruolo centrale perché è possibile dimostrare che tutti gli stimatori di ponderazione vincolata convergono asintoticamente, all'aumentare della numerosità campionaria, allo stimatore di regressione generalizzata.

6. Valutazione del livello di precisione delle stime

6.1. Metodologia di calcolo degli errori campionari

Le principali statistiche di interesse per valutare la variabilità campionaria delle stime prodotte da un'indagine sono l'errore di campionamento assoluto e l'errore di campionamento relativo. Indicando con $\hat{V} \text{ ar } (\hat{Y}_d)$ la stima della varianza della generica stima \hat{Y}_d , la stima dell'errore di campionamento assoluto di \hat{Y}_d si può ottenere mediante la seguente espressione:

$$\hat{\sigma}(\hat{Y}_d) = \sqrt{\hat{V} \text{ ar } (\hat{Y}_d)}; \quad (3)$$

⁵ Nella letteratura in lingua anglosassone sull'argomento tali stimatori sono noti come *calibration estimators*.

la stima dell'errore di campionamento relativo di \hat{Y}_d è invece definita dall'espressione:

$$\hat{\varepsilon}(\hat{Y}_d) = \frac{\hat{\sigma}(\hat{Y}_d)}{\hat{Y}_d}. \quad (4)$$

Come è stato descritto in precedenza, le stime prodotte dall'indagine sono state ottenute mediante uno stimatore di ponderazione vincolata definito in base a una funzione di distanza di tipo logaritmico troncato. Poiché, lo stimatore adottato non è funzione lineare dei dati campionari, per la stima della varianza $\hat{V}ar(\hat{Y}_d)$ si è utilizzato il metodo proposto da Woodruff; in base a tale metodo, che ricorre all'espressione linearizzata in serie di Taylor, è possibile ricavare la varianza di ogni stimatore non lineare (funzione regolare di totali) calcolando la varianza dell'espressione linearizzata ottenuta. In particolare, per la definizione dell'espressione linearizzata dello stimatore ci si è riferiti allo stimatore di regressione generalizzata, sfruttando la convergenza asintotica di tutti gli stimatori di ponderazione vincolata a tale stimatore, poiché nel caso di stimatori di ponderazione vincolata che utilizzano funzioni di distanza differenti dalla distanza euclidea (che conduce allo stimatore di regressione generalizzata) non è possibile derivare l'espressione linearizzata dello stimatore.

L'espressione linearizzata dello stimatore (2) è data, quindi, da:

$$\hat{Y}_d \cong \hat{Z}_d = \sum_{h=1}^{H_d} \hat{Z}_h, \text{ essendo } \hat{Z}_h = \sum_{i=1}^{n_h} \sum_{j=1}^{m_{hj}} Z_{hij} W_{hij} \quad (5)$$

dove Z_{hij} è la variabile linearizzata espressa come $Z_{hij} = Y_{hij} - \mathbf{X}'_{hij} \beta$, essendo $\mathbf{X}_{hij} = (x_{1hij}, \dots, x_{K,hij}, \dots, x_{K,hij})'$ il vettore contenente i valori delle K ($K=24$) variabili ausiliarie, osservati per la generica famiglia hij e $\hat{\beta}$, il vettore dei coefficienti di regressione del modello lineare che lega la variabile di interesse y alle K variabili ausiliarie x . In base alla (5), si ha, quindi, che la stima della varianza della stima \hat{Y}_d è ottenuta mediante la seguente relazione:

$$\hat{V}ar(\hat{Y}_d) \cong \hat{V}ar(\hat{Z}_d) = \sum_{h=1}^{H_d} \hat{V}ar(\hat{Z}_h). \quad (6)$$

Dalla (6) risulta che la stima della varianza della stima \hat{Y}_d viene calcolata come somma della stima delle varianze dei singoli strati, Ar e Nar , appartenenti al dominio d . La formula di calcolo della varianza, $\hat{V}ar(\hat{Z}_h)$, della stima \hat{Z}_h è differente a seconda che lo strato sia Ar oppure Nar . Possiamo, quindi scomporre come segue:

$$\hat{V}ar(\hat{Y}_d) \cong \hat{V}ar(\hat{Z}_d) = \sum_{h=1}^{H_{AR}} \hat{V}ar(\hat{Z}_h) + \sum_{h=1}^{H_{NAR}} \hat{V}ar(\hat{Z}_h) \quad , (7)$$

in cui H_{AR} e H_{NAR} indicano rispettivamente il numero di strati Ar e Nar appartenenti al dominio d .

Negli strati Ar (in cui ciascun comune fa strato a sé e $N_h = n_h = 1$, l'indice i di comune diviene superfluo e viene omesso) la varianza è stimata mediante la seguente espressione:

$$\sum_{h=1}^{H_{AR}} \hat{V}ar(\hat{Z}_h) = \sum_{h=1}^{H_{AR}} M_h^2 \frac{(M_h - m_h)}{m_h(m_h - 1)} \sum_{j=1}^{m_h} (Z_{hj} - \bar{Z}_h)^2 \quad , (8)$$

dove si è posto, $M_h = M_{hi}$, $m_h = m_{hi}$, $Z_{hj} = Z_{hij}$ e $\bar{Z}_h = \frac{1}{m_h} \sum_{j=1}^{m_h} Z_{hj}$.

Negli strati Nar, in cui viene estratto un solo comune campione da ogni strato, per stimare la varianza di campionamento si ricorre alla tecnica di collassamento degli strati. Questa tecnica consiste nel formare G gruppi contenenti ciascuno L_g ($L_g \geq 2$) strati; la varianza viene stimata mediante la formula seguente:

$$\sum_{h=1}^{H_{NAR}} \hat{V}ar(\hat{Z}_h) = \sum_{g=1}^G \hat{V}ar(\hat{Z}_g) = \sum_{g=1}^G \frac{L_g}{L_g - 1} \sum_{h=1}^{L_g} \left(\hat{Z}_{hg} - \frac{\hat{Z}_g}{L_g} \right)^2 \quad (9)$$

dove le quantità sono espresse come:

$$\hat{Z}_{hg} = \sum_{j=1}^{m_{hi}} Z_{hij} W_{hij} \quad e \quad \hat{Z}_g = \sum_{h=1}^{L_g} \sum_{j=1}^{m_{hi}} Z_{hij} W_{hij} \quad .$$

Utilizzando le espressioni (8) e (9) è possibile, infine, calcolare la varianza di campionamento, $\hat{V}ar(\hat{Y}_d)$, in base alla (7) e calcolare, quindi, in base alla (3) ed alla (4) rispettivamente l'errore di campionamento assoluto e l'errore di campionamento relativo.

Gli errori campionari espressi dalla (3) e dalla (4) consentono di valutare il grado di precisione delle stime; inoltre, l'errore assoluto permette di costruire un intervallo di confidenza, che, con livello di fiducia P contiene il parametro oggetto di stima, l'intervallo viene espresso come:

$$\left\{ \hat{Y}_d - k_p \hat{\sigma}(\hat{Y}_d) \leq Y_d \leq \hat{Y}_d + k_p \hat{\sigma}(\hat{Y}_d) \right\} \quad (10)$$

Nella (10) il valore di k_p dipende dal valore fissato per la probabilità P ; ad esempio, per $P=0.95$ si ha $k=1.96$.

6.2. Fondamenti statistici della procedura per il calcolo degli errori campionari

Per il calcolo degli errori di campionamento delle indagini condotte dall'Istat sulle famiglie e sulle imprese viene correntemente utilizzata una procedura informatica sviluppata nell'ambito dell'Istituto. Nel paragrafo precedente è stata descritta la metodologia, implementata dalla procedura, per il calcolo degli errori di campionamento delle stime prodotte dall'indagine mentre, nel presente paragrafo, vengono discussi i fondamenti statistici e i limiti della metodologia medesima.

Negli strati Ar, nei quali si adotta un disegno di campionamento a grappoli e in cui le unità primarie (le famiglie) vengono selezionate senza reimmissione e probabilità uguali, la procedura consente di ottenere stime della varianza campionaria che risultano corrette.

Negli strati Nar, per i quali si adotta un disegno di campionamento a due stadi con selezione delle unità primarie (comuni) senza reimmissione e probabilità variabili, la procedura consente di ottenere stime corrette della varianza campionaria qualora:

- in ciascuno strato sono selezionate due o più unità primarie;
- le unità primarie sono scelte mediante estrazioni indipendenti.

La prima condizione non viene soddisfatta in quanto, nell'indagine in oggetto, da ciascuno strato viene selezionato un solo comune campione e per stimare la varianza di campionamento si ricorre alla tecnica di collassamento degli strati. Questa tecnica, che consiste nel formare superstrati contenenti ciascuno un numero di strati maggiore di uno, conduce in generale ad una sovrastima della varianza di campionamento effettiva.

La seconda ipotesi implica che la selezione delle unità primarie venga effettuata con reimmissione. Anche questa assunzione non è soddisfatta per i comuni Nar e ciò comporta una sovrastima della varianza. Si osservi, tuttavia, che tale sovrastima dipende dalla frazione di campionamento di ciascuno strato Nar: è di entità trascurabile negli strati nei quali la frazione di campionamento è piccola, mentre viceversa può risultare di entità più cospicua per quegli strati in cui la frazione di campionamento è maggiore.

Tale metodologia è implementata nel software ISTAT Regenesees che è stato impiegato anche per il calcolo degli errori campionari e per la presentazione sintetica degli errori campionari.

6.3. Presentazione sintetica degli errori campionari

Ad ogni stima \hat{Y}_d corrisponde un errore di campionamento relativo $\hat{\varepsilon}(\hat{Y}_d)$; ciò significa che per consentire una lettura corretta delle tabelle pubblicate sarebbe necessario presentare per ogni stima pubblicata il corrispondente errore di campionamento relativo. Ciò, tuttavia, non è possibile sia per limiti di tempo e di costi di elaborazione, sia perché le tavole della pubblicazione risulterebbero appesantite e di non facile consultazione per l'utente finale. Inoltre, non sarebbero comunque disponibili gli errori delle stime non pubblicate, che l'utente può ricavare in modo autonomo.

Per le ragioni sopra esposte, si ricorre frequentemente a una presentazione sintetica degli errori relativi, basata sul metodo dei modelli regressivi. Questo metodo si basa sulla determinazione di una funzione matematica che mette in relazione ciascuna stima con il proprio errore relativo.

Nella presente indagine, il modello utilizzato per le stime di frequenze assolute e relative, è del tipo seguente:

$$\log(\hat{\varepsilon}^2(\hat{Y}_d)) = a + b \log(\hat{Y}_d) \quad (11)$$

dove i parametri a e b vengono stimati utilizzando il metodo dei minimi quadrati.

Nel prospetto 2 sono riportati i valori dei coefficienti a e b e dell'indice di determinazione R^2 del modello utilizzato per l'interpolazione degli errori campionari di stime di frequenze assolute e relative, per totale Italia, ripartizione geografica, tipologia comunale e regione.

Sulla base delle informazioni contenute in tale prospetto, è possibile calcolare la stima dell'errore di campionamento relativo di una determinata stima di frequenza assoluta \hat{Y}_d mediante la formula:

$$\hat{\varepsilon}(\hat{Y}_d) = \sqrt{\exp(a + b \log(\hat{Y}_d))} \quad (12)$$

che si ricava facilmente dalla (11).

Se, per esempio, la stima \hat{Y}_d si riferisce agli individui dell'Italia Nord occidentale, l'errore relativo corrispondente si ottiene introducendo nella (12) i valori dei parametri a e b riportati nella riga del Nord-ovest del prospetto 2 alla voce Persone.

I prospetti 3 e 4, presentati in aggiunta, consentono di rendere più agevole il calcolo degli errori campionari. Essi riguardano, rispettivamente, le famiglie e gli individui e hanno la seguente struttura:

a) in fiancata sono elencati i valori crescenti di stima (20.000, 30.000, ..., 25.000.000);

b) le colonne successive contengono gli errori di campionamento relativo, per ciascun dominio territoriale di interesse, calcolati mediante la formula (12), corrispondenti alle stime di frequenze assolute della prima colonna.

Le informazioni contenute in tali prospetti permettono di calcolare l'errore relativo di una generica stima di frequenza assoluta (o relativa) mediante due procedimenti che risultano di facile applicazione, anche se conducono a risultati meno precisi di quelli ottenibili mediante l'espressione (12). Il primo metodo consiste nell'individuare, nella prima colonna del prospetto, il livello di stima che più si avvicina alla stima di interesse e nel considerare come errore relativo il valore che si trova sulla stessa riga, nella colonna corrispondente al dominio territoriale di riferimento.

Con il secondo metodo, l'errore campionario della stima \hat{Y}_d si ricava mediante la seguente espressione:

$$\hat{\varepsilon}(\hat{Y}_d) = \hat{\varepsilon}(\hat{Y}_d^{k-1}) - \frac{\hat{\varepsilon}(\hat{Y}_d^{k-1}) - \hat{\varepsilon}(\hat{Y}_d^k)}{\hat{Y}_d^k - \hat{Y}_d^{k-1}} (\hat{Y}_d - \hat{Y}_d^{k-1}) \quad (13)$$

dove \hat{Y}_d^{k-1} e \hat{Y}_d^k sono i valori delle stime, riportati nella prima colonna, entro i quali è compresa la stima di interesse \hat{Y}_d , ed $\hat{\varepsilon}(\hat{Y}_d^{k-1})$ e $\hat{\varepsilon}(\hat{Y}_d^k)$ i corrispondenti errori relativi.

Prospetto 2 – Valori dei coefficienti a, b e dell'indice di determinazione R² (%) delle funzioni utilizzate per le interpolazioni degli errori campionari delle stime riferite alle famiglie e alle persone per totale Italia, ripartizione geografica, tipo di comune e regione – Anno 2020

ZONE TERRITORIALI	Famiglie			Persone		
	a	b	R ² (%)	a	b	R ² (%)
ITALIA	8,212615	-1,060818	98,9%	8,829211	-1,119203	96,0%
RIPARTIZIONI GEOGRAFICHE						
Nord	8,382654	-1,073316	98,4%	8,961163	-1,129625	96,7%
<i>Nord-ovest</i>	<i>8,547210</i>	<i>-1,081377</i>	98,0%	<i>9,250112</i>	<i>-1,150531</i>	96,1%
<i>Nord-est</i>	<i>8,342273</i>	<i>-1,092148</i>	97,6%	<i>8,895853</i>	<i>-1,151189</i>	96,0%
Centro	8,664331	-1,106974	97,3%	9,118254	-1,157560	95,0%
Mezzogiorno	7,656624	-1,036649	97,7%	8,152824	-1,088544	94,5%
<i>Sud</i>	<i>7,635715</i>	<i>-1,039811</i>	97,3%	<i>8,132216</i>	<i>-1,093745</i>	94,5%
<i>Isole</i>	<i>7,778735</i>	<i>-1,045615</i>	96,2%	<i>8,281355</i>	<i>-1,101482</i>	93,8%
TIPI DI COMUNE						
A1	7,888720	-1,033479	99,0%	8,021168	-1,048233	99,0%
A2	7,886910	-1,021829	97,5%	8,079220	-1,041755	97,8%
B1	5,829384	-0,873280	93,1%	6,063806	-0,905561	93,0%
B2	7,256548	-0,989044	97,3%	7,393681	-1,003267	96,9%
B3	7,673211	-1,018777	98,2%	7,661707	-1,022458	97,8%
B4	7,943757	-1,048205	98,8%	8,042788	-1,059157	98,7%
REGIONI						
Piemonte	8,258402	-1,100029	97,3%	8,924846	-1,173814	95,4%
Valle d'Aosta - Vallée d'Aoste	5,526678	-1,111248	96,8%	5,999951	-1,190512	95,7%
Liguria	7,770538	-1,108077	98,5%	8,082137	-1,150165	97,0%
Lombardia	8,969708	-1,100083	98,1%	9,486954	-1,153919	96,7%
Trentino-Alto Adige	6,685618	-1,079525	97,6%	6,971403	-1,119557	96,2%
<i>Bolzano - Bozen</i>	<i>6,703295</i>	<i>-1,085753</i>	96,5%	<i>7,139956</i>	<i>-1,146777</i>	95,0%
<i>Trento</i>	<i>6,806639</i>	<i>-1,103628</i>	97,1%	<i>7,205714</i>	<i>-1,159283</i>	95,1%
Veneto	8,211951	-1,070932	96,6%	8,729361	-1,132079	94,7%
Friuli-Venezia Giulia	7,329523	-1,093411	98,0%	8,065098	-1,185578	96,4%
Emilia-Romagna	8,597542	-1,114721	97,3%	9,230112	-1,183955	95,6%
Toscana	8,221385	-1,086946	97,1%	8,640694	-1,141695	94,6%
Umbria	7,161503	-1,076439	97,7%	7,995767	-1,181835	95,3%
Marche	7,794960	-1,122101	97,6%	8,514753	-1,208431	96,0%
Lazio	8,698281	-1,097324	96,0%	9,414074	-1,169741	94,4%
Abruzzo	7,102745	-1,058087	96,4%	7,587579	-1,121260	94,8%
Molise	5,794380	-1,056660	95,6%	6,126949	-1,109170	93,2%
Campania	7,875469	-1,040861	97,2%	8,611206	-1,119626	95,7%
Puglia	7,764101	-1,045287	97,0%	8,328175	-1,111048	95,5%
Basilicata	6,491832	-1,070798	95,6%	6,890533	-1,132904	92,6%
Calabria	6,784531	-0,990616	97,1%	7,353946	-1,062713	93,1%
Sicilia	8,009645	-1,056712	96,2%	8,135755	-1,074373	94,5%
Sardegna	7,483716	-1,068058	96,4%	7,944751	-1,129092	94,4%

(a) Italia nord-occidentale: Piemonte, Valle d'Aosta, Lombardia, Liguria; Italia nord-orientale: Bolzano, Trento, Veneto, Friuli-Venezia Giulia, Emilia-Romagna; Italia centrale: Toscana, Umbria, Marche, Lazio; Italia meridionale: Abruzzo, Molise, Campania, Puglia, Basilicata, Calabria; Italia insulare: Sicilia, Sardegna.

(b) Comuni tipo A1: Area urbana centro; Tipo A2: Area urbana periferia; Tipo B1: comuni fino a 2.000 abitanti; Tipo B2: da 2.001 a 10.000 abitanti; Tipo B3: da 10.001 a 50.000 abitanti; Tipo B4: oltre 50.000 abitanti.

Prospetto 3 – Valori interpolati degli errori campionari relativi percentuali delle stime riferite alle famiglie per totale Italia, ripartizione geografica, tipo di comune e regione – Anno 2020

STIME	Italia	Nord	Nord-ovest	Nord-est	Centro	Mezzogiorno	Sud	Isole	A1	A2	B1	B2	B3	B4
20.000	31,8	32,5	33,9	29,0	31,7	27,1	26,4	27,6	30,9	32,7	24,4	28,1	29,9	29,6
30.000	25,6	26,2	27,2	23,3	25,3	22,0	21,4	22,3	25,1	26,6	20,5	23,0	24,3	23,9
40.000	22,0	22,4	23,3	19,9	21,6	18,9	18,4	19,2	21,6	23,0	18,0	19,9	21,0	20,6
50.000	19,5	19,9	20,7	17,6	19,1	16,9	16,4	17,1	19,3	20,5	16,4	17,9	18,7	18,3
60.000	17,7	18,0	18,7	15,9	17,3	15,3	14,9	15,5	17,5	18,7	15,1	16,3	17,1	16,6
70.000	16,3	16,6	17,2	14,6	15,8	14,2	13,8	14,3	16,2	17,3	14,1	15,1	15,8	15,3
80.000	15,2	15,5	16,0	13,6	14,7	13,2	12,9	13,4	15,1	16,1	13,3	14,2	14,7	14,3
90.000	14,3	14,5	15,0	12,8	13,8	12,4	12,1	12,6	14,2	15,2	12,7	13,4	13,9	13,4
100.000	13,5	13,7	14,2	12,1	13,0	11,8	11,4	11,9	13,5	14,4	12,1	12,7	13,2	12,7
200.000	9,4	9,4	9,8	8,3	8,9	8,2	8,0	8,3	9,4	10,1	8,9	9,0	9,2	8,8
300.000	7,6	7,6	7,8	6,6	7,1	6,7	6,5	6,7	7,6	8,2	7,5	7,4	7,5	7,2
400.000	6,5	6,5	6,7	5,7	6,0	5,7	5,6	5,8	6,6	7,1	6,6	6,4	6,5	6,2
500.000	5,8	5,8	6,0	5,0	5,3	5,1	5,0	5,1	5,9	6,3	6,0	5,7	5,8	5,5
750.000	4,6	4,6	4,8	4,0	4,3	4,1	4,0	4,1	4,8	5,1	5,0	4,7	4,7	4,4
1.000.000	4,0	4,0	4,1	3,4	3,6	3,6	3,5	3,6	4,1	4,4	4,4	4,1	4,1	3,8
2.000.000	2,8	2,7	2,8	2,3	2,5	2,5	2,4	2,5	2,9	3,1	3,3	2,9	2,9	2,6
3.000.000	2,2	2,2	2,3	1,9	2,0	2,0	2,0	2,0	2,3	2,5	2,7	2,4	2,3	2,1
4.000.000	1,9	1,9	1,9	1,6	1,7	1,7	1,7	1,7	2,0	2,2	2,4	2,0	2,0	1,8
5.000.000	1,7	1,7	1,7	1,4	1,5	1,6	1,5	1,5	1,8	1,9	2,2	1,8	1,8	1,6
7.500.000	1,4	1,4	1,4	1,1	1,2	1,3	1,2	1,2	1,4	1,6	1,8	1,5	1,5	1,3
10.000.000	1,2	1,2	1,2	1,0	1,0	1,1	1,0	1,1	1,2	1,4	1,6	1,3	1,3	1,1
15.000.000	0,9	0,9	0,9	0,8	0,8	0,9	0,8	0,9	1,0	1,1	1,4	1,1	1,0	0,9
20.000.000	0,8	0,8	0,8	0,7	0,7	0,8	0,7	0,7	0,9	1,0	1,2	0,9	0,9	0,8
25.000.000	0,7	0,7	0,7	0,6	0,6	0,7	0,6	0,7	0,8	0,9	1,1	0,8	0,8	0,7

STIME	Piemonte	Valle d'Aosta - Vallée d'Aoste	Liguria	Lombardia	Trentino-Alto Adige	Bolzano	Trento	Veneto	Friuli-Venezia Giulia	Emilia-Romagna	Toscana	Umbria
20.000	26,8	6,5	20,2	38,2	13,5	13,2	12,7	30,2	17,4	29,5	28,0	17,4
30.000	22,2	5,2	16,1	30,6	10,8	10,6	10,2	24,3	13,9	23,5	22,5	14,0
40.000	19,2	4,4	13,7	26,1	9,3	9,1	8,7	20,8	11,9	20,0	19,2	12,0
50.000	17,1	3,9	12,1	23,1	8,2	8,0	7,7	18,5	10,5	17,7	17,0	10,6
60.000	15,6	3,5	11,0	20,9	7,5	7,3	6,9	16,8	9,5	16,0	15,4	9,6
70.000	14,4	3,2	10,1	19,2	6,9	6,7	6,4	15,4	8,8	14,7	14,2	8,9
80.000	13,4	3,0	9,4	17,8	6,4	6,2	5,9	14,4	8,1	13,6	13,2	8,2
90.000	12,7	2,8	8,8	16,7	6,0	5,8	5,5	13,5	7,6	12,8	12,4	7,7
100.000	12,0	2,6	8,3	15,8	5,7	5,5	5,2	12,8	7,2	12,0	11,7	7,3
200.000	8,4	1,8	5,6	10,8	3,9	3,8	3,6	8,8	4,9	8,2	8,0	5,0
300.000	6,8	1,4	4,5	8,6	3,1	3,0	2,9	7,1	4,0	6,5	6,4	4,0
400.000	5,9	1,2	3,8	7,4	2,7	2,6	2,4	6,1	3,4	5,6	5,5	3,5
500.000	5,3	1,1	3,4	6,5	2,4	2,3	2,2	5,4	3,0	4,9	4,9	3,1
750.000	4,3	0,9	2,7	5,2	1,9	1,8	1,7	4,3	2,4	3,9	3,9	2,5
1.000.000	3,7	0,7	2,3	4,4	1,6	1,6	1,5	3,7	2,0	3,3	3,3	2,1
2.000.000	2,6	0,5	1,6	3,0	1,1	1,1	1,0	2,6	1,4	2,3	2,3	1,5

STIME	Marche	Lazio	Abruzzo	Molise	Campania	Puglia	Basilicata	Calabria	Sicilia	Sardegna
20.000	19,0	33,8	18,5	9,7	29,6	27,4	12,8	22,0	29,3	21,3
30.000	15,2	27,1	14,9	7,8	24,0	22,2	10,3	18,0	23,6	17,1
40.000	12,9	23,1	12,8	6,7	20,7	19,1	8,8	15,6	20,3	14,7
50.000	11,4	20,4	11,4	6,0	18,4	17,0	7,8	14,0	18,1	13,1
60.000	10,3	18,5	10,3	5,4	16,7	15,4	7,1	12,8	16,4	11,8
70.000	9,4	17,0	9,5	5,0	15,4	14,2	6,5	11,8	15,1	10,9
80.000	8,7	15,8	8,9	4,7	14,4	13,3	6,1	11,1	14,1	10,2
90.000	8,2	14,8	8,3	4,4	13,5	12,5	5,7	10,5	13,2	9,5
100.000	7,7	14,0	7,9	4,1	12,8	11,8	5,4	9,9	12,5	9,0
200.000	5,2	9,6	5,5	2,9	8,9	8,2	3,7	7,0	8,7	6,2
300.000	4,2	7,7	4,4	2,3	7,2	6,7	3,0	5,8	7,0	5,0
400.000	3,5	6,5	3,8	2,0	6,2	5,7	2,6	5,0	6,0	4,3
500.000	3,1	5,8	3,4	1,8	5,5	5,1	2,3	4,5	5,3	3,8
750.000	2,5	4,6	2,7	1,4	4,5	4,1	1,8	3,7	4,3	3,1
1.000.000	2,1	4,0	2,3	1,2	3,9	3,5	1,6	3,2	3,7	2,6
2.000.000	1,4	2,7	1,6	0,8	2,7	2,5	1,1	2,3	2,6	1,8

Prospetto 4 – Valori interpolati degli errori campionari relativi percentuali delle stime riferite alle persone per totale Italia, ripartizione geografica, tipo di comune e regione – Anno 2020

STIME	Italia	Nord	Nord- ovest	Nord- est	Centro	Mezzogiorno	Sud	Isole	A1	A2	B1	B2	B3	B4
20.000	32,4	32,9	34,2	28,6	30,9	26,9	25,9	26,9	30,7	32,7	23,4	28,1	29,2	29,4
30.000	25,8	26,1	27,1	22,6	24,5	21,6	20,8	21,5	24,8	26,4	19,5	22,9	23,7	23,7
40.000	22,0	22,2	23,0	19,2	20,7	18,4	17,7	18,4	21,4	22,8	17,1	19,8	20,5	20,4
50.000	19,4	19,6	20,2	16,9	18,2	16,3	15,7	16,2	19,0	20,3	15,5	17,7	18,3	18,1
60.000	17,5	17,7	18,2	15,2	16,4	14,8	14,2	14,7	17,3	18,4	14,2	16,2	16,6	16,4
70.000	16,1	16,2	16,7	13,9	15,0	13,6	13,1	13,5	15,9	17,0	13,3	15,0	15,4	15,2
80.000	14,9	15,0	15,4	12,9	13,9	12,6	12,1	12,5	14,9	15,9	12,5	14,0	14,4	14,1
90.000	14,0	14,0	14,4	12,0	13,0	11,9	11,4	11,7	14,0	14,9	11,8	13,2	13,5	13,3
100.000	13,2	13,2	13,6	11,3	12,2	11,2	10,8	11,1	13,2	14,1	11,3	12,5	12,8	12,5
200.000	8,9	8,9	9,1	7,6	8,2	7,7	7,4	7,6	9,2	9,8	8,3	8,8	9,0	8,7
300.000	7,1	7,1	7,2	6,0	6,5	6,2	5,9	6,1	7,4	8,0	6,9	7,2	7,3	7,0
400.000	6,1	6,1	6,1	5,1	5,5	5,3	5,0	5,2	6,4	6,9	6,0	6,2	6,3	6,0
500.000	5,3	5,3	5,4	4,5	4,8	4,7	4,5	4,6	5,7	6,1	5,4	5,6	5,6	5,4
750.000	4,3	4,2	4,3	3,5	3,8	3,7	3,6	3,7	4,6	4,9	4,5	4,6	4,6	4,3
1.000.000	3,6	3,6	3,6	3,0	3,2	3,2	3,1	3,1	4,0	4,3	4,0	3,9	3,9	3,7
2.000.000	2,5	2,4	2,4	2,0	2,2	2,2	2,1	2,1	2,7	3,0	2,9	2,8	2,8	2,6
3.000.000	2,0	1,9	1,9	1,6	1,7	1,8	1,7	1,7	2,2	2,4	2,4	2,3	2,3	2,1
4.000.000	1,7	1,6	1,6	1,4	1,4	1,5	1,4	1,5	1,9	2,1	2,1	2,0	1,9	1,8
5.000.000	1,5	1,5	1,4	1,2	1,3	1,3	1,3	1,3	1,7	1,8	1,9	1,8	1,7	1,6
7.500.000	1,2	1,2	1,1	0,9	1,0	1,1	1,0	1,0	1,4	1,5	1,6	1,4	1,4	1,3
10.000.000	1,0	1,0	1,0	0,8	0,8	0,9	0,9	0,9	1,2	1,3	1,4	1,2	1,2	1,1
15.000.000	0,8	0,8	0,8	0,6	0,7	0,7	0,7	0,7	1,0	1,0	1,2	1,0	1,0	0,9
20.000.000	0,7	0,7	0,6	0,5	0,6	0,6	0,6	0,6	0,8	0,9	1,0	0,9	0,9	0,8
25.000.000	0,6	0,6	0,6	0,5	0,5	0,6	0,5	0,5	0,7	0,8	0,9	0,8	0,8	0,7

STIME	Piemonte	Valle d'Aosta - Vallée d'Aoste	Liguria	Lombardia	Trentino- Alto Adige	Bolzano	Trento	Veneto	Friuli- Venezia Giulia	Emilia- Romagna	Toscana	Umbria
20.000	25,9	5,5	19,1	37,9	12,8	12,1	11,8	28,9	15,9	28,7	26,4	15,7
30.000	20,4	4,3	15,1	30,0	10,2	9,6	9,3	23,0	12,5	22,6	20,9	12,3
40.000	17,3	3,7	12,8	25,4	8,7	8,2	7,9	19,5	10,6	19,1	17,8	10,4
50.000	15,1	3,2	11,3	22,3	7,6	7,2	6,9	17,2	9,2	16,7	15,6	9,1
60.000	13,6	2,9	10,2	20,1	6,9	6,5	6,2	15,5	8,3	15,0	14,1	8,2
70.000	12,4	2,6	9,3	18,4	6,3	5,9	5,7	14,2	7,6	13,7	12,9	7,5
80.000	11,5	2,4	8,6	17,0	5,9	5,5	5,3	13,2	7,0	12,6	12,0	6,9
90.000	10,7	2,3	8,1	15,9	5,5	5,1	4,9	12,3	6,5	11,8	11,2	6,4
100.000	10,1	2,1	7,6	15,0	5,2	4,8	4,6	11,6	6,1	11,1	10,5	6,0
200.000	6,7	1,4	5,1	10,0	3,5	3,2	3,1	7,9	4,1	7,3	7,1	4,0
300.000	5,3	1,1	4,0	7,9	2,8	2,6	2,5	6,2	3,2	5,8	5,6	3,2
400.000	4,5	0,9	3,4	6,7	2,4	2,2	2,1	5,3	2,7	4,9	4,8	2,7
500.000	3,9	0,8	3,0	5,9	2,1	1,9	1,8	4,7	2,4	4,3	4,2	2,3
750.000	3,1	0,6	2,4	4,7	1,7	1,5	1,4	3,7	1,9	3,4	3,3	1,8
1.000.000	2,6	0,5	2,0	4,0	1,4	1,3	1,2	3,2	1,6	2,8	2,8	1,6
2.000.000	1,7	0,4	1,4	2,7	1,0	0,9	0,8	2,1	1,0	1,9	1,9	1,0
3.000.000	1,4	0,3	1,1	2,1	0,8	0,7	0,6	1,7	0,8	1,5	1,5	0,8
4.000.000	1,2	0,2	0,9	1,8	0,7	0,6	0,5	1,4	0,7	1,2	1,3	0,7
5.000.000	1,0	0,2	0,8	1,6	0,6	0,5	0,5	1,3	0,6	1,1	1,1	0,6

STIME	Marche	Lazio	Abruzzo	Molise	Campania	Puglia	Basilicata	Calabria	Sicilia	Sardegna
20.000	17,8	33,8	17,2	8,8	29,0	26,2	11,5	20,5	28,6	19,8
30.000	13,9	26,7	13,7	7,0	23,1	21,0	9,1	16,5	23,0	15,8
40.000	11,7	22,5	11,7	6,0	19,7	17,9	7,8	14,2	19,7	13,4
50.000	10,2	19,8	10,3	5,3	17,4	15,8	6,8	12,6	17,5	11,8
60.000	9,2	17,8	9,3	4,8	15,7	14,3	6,2	11,4	15,8	10,7
70.000	8,3	16,2	8,5	4,4	14,4	13,1	5,6	10,5	14,6	9,8
80.000	7,7	15,0	7,9	4,1	13,3	12,2	5,2	9,8	13,6	9,1
90.000	7,2	14,0	7,4	3,8	12,5	11,4	4,9	9,2	12,7	8,5
100.000	6,7	13,2	7,0	3,6	11,8	10,7	4,6	8,7	12,0	8,0
200.000	4,4	8,8	4,7	2,5	8,0	7,3	3,1	6,0	8,3	5,4
300.000	3,5	6,9	3,8	2,0	6,4	5,8	2,5	4,9	6,7	4,3
400.000	2,9	5,9	3,2	1,7	5,4	5,0	2,1	4,2	5,7	3,7
500.000	2,5	5,1	2,8	1,5	4,8	4,4	1,9	3,7	5,1	3,2
750.000	2,0	4,1	2,3	1,2	3,8	3,5	1,5	3,0	4,1	2,6
1.000.000	1,7	3,4	1,9	1,0	3,2	3,0	1,3	2,6	3,5	2,2
2.000.000	1,1	2,3	1,3	0,7	2,2	2,0	0,8	1,8	2,4	1,5
3.000.000	0,9	1,8	1,0	0,5	1,8	1,6	0,7	1,4	1,9	1,2
4.000.000	0,7	1,5	0,9	0,5	1,5	1,4	0,6	1,2	1,7	1,0
5.000.000	0,6	1,3	0,8	0,4	1,3	1,2	0,5	1,1	1,5	0,9

6.4. Esempi di calcolo per la costruzione dell'intervallo di confidenza

Nelle tabelle seguenti sono illustrate le modalità di calcolo per la costruzione dell'intervallo di confidenza. Nel primo esempio (tabella 1), l'intervallo è calcolato per una stima sulle famiglie, l'errore campionario è da ricercare nel prospetto 3. Nel secondo esempio (tabella 2), il calcolo è fatto per una stima di individui, l'errore di riferimento è nel prospetto 4.

Tabella 1- Esempio per il calcolo degli errori campionari nel caso di stime riferite alle famiglie

	Famiglie in Sicilia che dichiarano “adeguate” le proprie risorse economiche
Stima puntuale:	1.159.000
Errore relativo (CV)	3,7/100=0,037
Stima intervallare	
Semi ampiezza dell'intervallo:	42.883
Limite inferiore dell'intervallo di confidenza:	1.074.949
Limite superiore dell'intervallo di confidenza:	1.243.051

Tabella 2- Esempio per il calcolo degli errori campionari nel caso di stime riferite alle persone

	Persone in Italia molto soddisfatte della propria salute
Stima puntuale:	8.915.000
Errore relativo (CV)	1,2/100=0,012
Stima intervallare	
Semi ampiezza dell'intervallo	106.980
Limite inferiore dell'intervallo di confidenza	8.705.319
Limite superiore dell'intervallo di confidenza	9.124.681

Per avere un intervallo più preciso, l'errore campionario può essere calcolato direttamente con la funzione interpolante:

$$\hat{\varepsilon}(\hat{Y}) = \sqrt{\exp(a + b \log(\hat{Y}))}$$

utilizzando i parametri riportati nel prospetto 2.

7. La diffusione dei risultati dell'indagine

I principali risultati dell'indagine vengono resi disponibili sul sito dell'Istat attraverso statistiche report pubblicate nei settori con argomento: “Opinioni dei cittadini”, “Salute e sanità”, “Cultura, comunicazione, tempo libero”, “Partecipazione sociale”. I risultati sono diffusi sempre sul sito anche attraverso il datawarehouse I.Stat.

I dati d'indagine vengono resi disponibili mediante il rilascio di file di microdati (MFR e micro.stat). Ricercatori e studiosi possono, inoltre, accedere al Laboratorio di Analisi dei Dati Elementari (“ADELE”) per effettuare di persona le proprie analisi statistiche sui microdati dell'indagine, nel rispetto delle norme sulla riservatezza dei dati personali.

Ogni anno, inoltre, i dati raccolti vengono analizzati e pubblicati anche su volumi a carattere generale (Rapporto annuale, Annuario statistico italiano, Noi Italia, Italia in cifre). I volumi curati dall’Istat sono consultabili nel [Catalogo editoriale](#).

I dati diffusi sono privi degli elementi identificativi del soggetto al quale si riferiscono, nonché di ogni altro elemento che consenta, anche indirettamente, il collegamento con le famiglie o gli individui intervistati.

8. Riferimenti bibliografici

Il sistema di indagini sociali multiscopo, Metodi e Norme, n. 31, Anno 2006.

9. Contatti

**Servizio Registro della popolazione,
statistiche demografiche e condizioni di vita**

Sante Orsini
Tel. +39 06 4673.7256
Email orsini@istat.it

Descrizione del file
File description



**File ad uso pubblico
micro.STAT**

Aspetti della vita quotidiana
Periodo di riferimento: anno 2020

Descrizione del file

INDICE

Introduzione	3
Le misure di protezione adottate	4
Riferimenti bibliografici	10

Introduzione

I file di microdati ad uso pubblico (micro.STAT) sono collezioni di dati elementari, liberamente scaricabile via web (https://www.istat.it/it/dati-analisi-e-prodotti/microdati#file_uso_pubblico), per le quali, a partire dal 2013, sono stati sviluppati anche i corrispondenti file per la ricerca (MFR¹). Il contenuto informativo di questi ultimi è superiore rispetto a quello del file ad uso pubblico².

Per la predisposizione del file micro.STAT relativo all'indagine *Aspetti della vita quotidiana*, periodo di riferimento 2020, 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 statistiche 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.

¹ 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 (https://www.istat.it/it/dati-analisi-e-prodotti/microdati#file_ricerca).

² 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 *Aspetti della vita quotidiana*, periodo di riferimento 2020, sono state adottate le misure di protezione appresso specificate.

Il progressivo famiglia univoco a livello indagine [*profam*] è stato sostituito con un numero fittizio.

Le variabili peso [*peso*, *pesomin*] e statura [*stat*, *statmin*] sono state sostituite con le variabili indice di massa corporea:

1. Indice di massa corporea per le persone di 18 anni e più [*bmi*], con le modalità *sottopeso*, *normopeso*, *sovrappeso*, *obeso*;
2. Indice di massa corporea per le persone di 2-17 anni [*bmimin*], con le modalità *non in eccesso di peso*, *in eccesso di peso*.

Soppressione di variabili:

1. Provincia di residenza [*prov*]
2. Comune di residenza [*com*]
3. Ampiezza demografica del comune di residenza [*dom*]
4. Frequenza scolastica [*frsc*]

Ricodifiche di variabili categoriali:

1. Regione di residenza [*reg* nel file originario → *regMf* nel file mlcro.STAT]
 Aggregazione delle modalità *Provincia autonoma di Trento* e *Provincia autonoma di Bolzano* in un'unica modalità *Trentino-Alto Adige*
2. Stato civile [*stciv* → *stcivMf*]
 Aggregazione delle modalità *separato/a di fatto*, *separato/a legalmente*, *divorziato/a già in unione civile (per scioglimento unione)*:

Etichetta	Descrizione
1	celibe/nubile
2	coniugato/a coabitante con il coniuge; unito/a civilmente
3	separato/a di fatto; separato/a legalmente; divorziato/a; già in unione civile (per scioglimento unione)
6	vedovo/a; già in unione civile (per decesso del partner)
9	non disponibile

3. Titolo di studio [*istr* → *istrMf*]
 Aggregazione delle modalità originarie nelle seguenti:

Etichetta	Descrizione
01	laurea e post-laurea
07	diploma
09	licenza di scuola media
10	licenza di scuola elementare, nessun titolo di studio
99	non disponibile

4. Condizione occupazionale [*cond* → *condMi*]

Aggregazione delle modalità originarie nelle seguenti:

Etichetta	Descrizione
1	occupato
2	in cerca di occupazione
3	inattivo; altra condizione
9	non disponibile

5. Posizione nella professione [*posiz* → *posizMi*]

Aggregazione delle modalità originarie nelle seguenti:

Etichetta	Descrizione
01	dirigente, autonomo come imprenditore o libero professionista
02	direttivo, quadro o impiegato
03	capo operaio, operaio subalterno e assimilati, apprendista o lavorante a domicilio per conto d'impresa
04	lavoratore in proprio, socio cooperativa Produzione Beni e/o prestazioni di servizio, coadiuvante, collaborazione coordinata e continuativa (con o senza progetto) o prestazione d'opera occasionale
99	non disponibile

6. Settore di attività economica [*ateco* → *atecoMi*]

Aggregazione delle modalità originarie nelle seguenti:

Etichetta	Descrizione
01	agricoltura, silvicoltura e pesca
02	totale industria escluse costruzioni
03	costruzioni
04	totale servizi
99	non disponibile

7. Tipologia familiare [*tipfa2* → *tipfa2Mi*]

Aggregazione delle modalità:

Etichetta	Descrizione
01	senza nuclei
02	un nucleo senza isolati coppie
03	un nucleo senza isolati monogenitore
04	un nucleo con isolati coppie
05	un nucleo con isolati monogenitore
06	due nuclei senza isolati
07	due nuclei con isolati
08	tre o più nuclei

Ricodifiche di variabili quantitative

- Età [*eta* → *etaM*]

Aggregazione dell'età in anni compiuti nelle seguenti classi di età:

Etichetta	Descrizione
001	0 – 2 anni
002	3 – 5 anni
003	6 – 10 anni
004	11 – 13 anni
005	14 – 15 anni
006	16 – 17 anni
007	18 – 19 anni
008	20 – 24 anni
009	25 – 34 anni
010	35 – 44 anni
011	45 – 54 anni
012	55 - 59 anni
013	60 – 64 anni
014	65 – 74 anni
015	75 anni e più

Per la variabile:

- Numero di notti in cui è stato ricoverato negli ultimi tre mesi [*ggrica* → *ggricaM*]

i valori numerici sono stati aggregati secondo la seguente classificazione:

Etichetta	Descrizione
puntuali	Per i valori minori o uguali a 15
16	16 – 20
17	21 – 30
18	31 – 40
19	41 – 50
20	51 e oltre

Per la variabile:

- Numero di e-book letti negli ultimi 12 mesi [*nleboo* → *nlebooM*]

i valori numerici sono stati aggregati secondo la seguente classificazione:

Etichetta	Descrizione
puntuali	Per i valori minori o uguali a 15
20	16 – 20
21	21 – 25
22	26 - 30
23	31 – 40
24	41 - 50
25	51 e oltre

Per la variabile:

- Numero di volte in cui si è recato in biblioteca negli ultimi 12 mesi [*numbib* → *numbibM*]

i valori numerici sono stati aggregati secondo la seguente classificazione:

Etichetta	Descrizione
puntuali	Per i valori minori o uguali a 10
19	11 - 15
20	16 - 20
21	21 - 25
22	26 - 30
23	31 - 40
24	41 - 50
25	51 e oltre

- Per la variabile:
- Numero di volte in cui ha fatto ricorso all'assistenza domiciliare [*nassdo* → *nassdoM*]

i valori numerici sono stati aggregati secondo la seguente classificazione:

Etichetta	Descrizione
puntuali	Per i valori minori o uguali a 10
15	11 - 15
16	16 – 20
17	21 – 30
18	31 – 40
19	41 – 50
20	51 e oltre

Per le variabili:

- Numero di libri letti negli ultimi 12 mesi (considerare solo i libri letti per motivi non strettamente scolastici o professionali) [*nlibri* → *nlibriM*]
- Numero di libri cartacei letti [*nlibri_car* → *nlibri_carM*]

i valori numerici sono stati aggregati secondo la seguente classificazione:

Etichetta	Descrizione
puntuali	Per i valori minori o uguali a 20
21	21 - 25
22	26 - 30
23	31 - 40
24	41 - 50
25	51 e oltre

Per la variabile:

- Numero di bicchieri di vino o alcolici consumati abitualmente fuori dai pasti in una settimana [*bicfuori* → *bicfuoriM*]

i valori numerici sono stati aggregati secondo la seguente classificazione:

Etichetta	Descrizione
puntuali	Per i valori minori o uguali a 09
10	10 - 20
11	21 e oltre

Per la variabile:

- Numero di bicchieri al giorno di aperitivi alcolici, amari o superalcolici [*bicaltro* → *bicaltroM*]

i valori numerici sono stati aggregati secondo la seguente classificazione:

Etichetta	Descrizione
puntuali	Per i valori minori o uguali a 03
04	4 - 5
05	6 e oltre

Per la variabile:

- Numero di sigarette fumate in media al giorno [*nsigar* → *nsigarM*]

i valori numerici sono stati aggregati secondo la seguente classificazione:

Etichetta	Descrizione
puntuali	Per i valori minori o uguali a 15
16	16 – 20
17	21 e oltre

- Consumo di bevande alcoliche negli ultimi 12 mesi, numero bicchieri [*nbicalc* → *nbicalcM*]:

Etichetta	Descrizione
puntuali	Per i valori minori o uguali a 09
10	10 - 14
15	15 – 20
21	21 e oltre

Top coding di variabili quantitative

Per le variabili che seguono è stato applicato il metodo di *top-coding*, che consiste nel raggruppare le unità statistiche collocate lungo la coda dei valori più elevati della distribuzione.

- Numero segreterie telefoniche [*nsegt* → *nsegtM*];

i valori superiori a 1 sono stati ricodificati in 2, con significato “2 e oltre”.

- Numero di dvd [*ndvd* → *ndvdM*];
- Numero di hifi [*nhifi* → *nhifiM*];
- Numero di lettori di e-book [*nebook* → *nebookM*];

i valori superiori a 2 sono stati ricodificati in “3 e oltre”.

- Numero di motorini, scooter [*nmotor* → *nmotorM*];
- Numero di motociclette, moto [*nmoto* → *nmotoM*];
- Numero di carte di credito [*nccred* → *nccredM*];
- Numero di audiolibri ascoltati negli ultimi 12 mesi [*nlaudio* → *nlaudioM*];
- Numero di lettori mp3/mp4 [*nmp3* → *nmp3M*];
- Numero di volte in cui è stato ricoverato negli ultimi tre mesi [*nnrica* → *nnricaM*];

i valori superiori a 3 sono stati ricodificati in 4, con significato “4 e oltre”.

- Numero di volte in cui ha fatto ricorso al pronto soccorso negli ultimi tre mesi [*nproso* → *nprosoM*];
- Numero di volte in cui ha fatto ricorso alla guardia medica negli ultimi tre mesi [*ngumed* → *ngumedM*];

i valori superiori a 4 sono stati ricodificati in 5, con significato “5 e oltre”.

- Consumo abituale al giorno di birra, in bicchieri [*bicbirra* → *bicbirraM*];
- Consumo abituale al giorno di vino, in bicchieri [*bicvino* → *bicvinoM*];
- Numero di telefoni cellulari di cui si dispone [*ntelcel* → *ntelcelM*]

i valori superiori a 5 sono stati ricodificati in 6, con significato “6 e oltre”.

- Stanze di cui si compone l’abitazione [*stanze* → *stanzeM*];

i valori superiori a 12 sono stati ricodificati in 13, con significato “13 e oltre”.

Ulteriori interventi di protezione dei dati:

In alcuni record sono stati modificati i valori puntuali in corrispondenza delle variabili:

- Regione [*reg* → *regMf*]: sostituzione di alcuni valori con il corrispondente livello ripartizionale o con la modalità “non disponibile” (codice ‘99’)
- Stato civile [*stciv* → *stcivMi*], Cittadinanza [*citt* → *cittMi*], Condizione occupazionale [*cond* → *condMi*]: sostituzione di alcuni valori con il codice ‘9’ = “non disponibile”;
- Posizione nella professione [*posiz* → *posizMi*], Settore di attività economica [*ateco* → *atecoMi*], Titolo di studio [*istr* → *istrMi*]: sostituzione di alcuni valori con il codice ‘99’ = “non disponibile”.

Alcuni record, inoltre, sono stati soppressi.

Ai fini della tutela della riservatezza, le variabili individuate dal D.Lgs 30/06/03 sono state trattate mediante permutazione casuale di una parte delle osservazioni.

A causa delle misure di protezione adottate, si possono verificare scostamenti rispetto ai dati pubblicati dall’Istat.

Riferimenti bibliografici

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 Daniela Panaccione.