

L'indagine Multiscopo I cittadini e il tempo libero

La rilevazione "I cittadini e il tempo libero" fa parte di un sistema integrato di indagini sociali (Indagini Multiscopo sulle Famiglie). L'indagine nasce nel 1995 con cadenza quinquennale e in questi anni ha permesso di descrivere in modo sistematico un campo complesso e frammentario qual è quello del tempo libero e dei rapporti che esistono tra quest'ultimo e la partecipazione culturale.

L'indagine si concentra sulle concezioni, gli atteggiamenti e i comportamenti della popolazione riconducibili alla sfera del tempo libero e in particolare all'insieme di comportamenti e di attività relativi alla partecipazione culturale, alla pratica sportiva e alle attività più direttamente legate alla sfera del sé e dell'autorealizzazione.

I temi oggetto dell'indagine vengono analizzati sia sotto gli aspetti più tradizionali sia sotto quelli emergenti, dando ampio spazio ad approfondimenti nelle varie sezioni dei questionari. Si va dalla fruizione della televisione alle letture, dall'uso delle nuove tecnologie agli aspetti legati alla formazione. L'indagine permette di capire i profondi cambiamenti che stanno avvenendo nella nostra società relativamente al tempo libero, contestualizzato rispetto al più ampio tessuto della vita quotidiana.

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

L'indagine è stata realizzata a Maggio del 2006 su un campione di circa 24mila famiglie (per un totale di circa 54mila individui) distribuite in 853 Comuni italiani di diversa ampiezza demografica (cfr. cap. 2 in questa appendice).

I dati vengono raccolti con la tecnica del questionario cartaceo. I questionari utilizzati sono disponibili [on line](http://www.istat.it/strumenti/rispondenti/indagini/famiglia_societa/cultura/) all'indirizzo: http://www.istat.it/strumenti/rispondenti/indagini/famiglia_societa/cultura/

I dati pubblicati in questa statistica in breve si riferiscono alle sezione 6 del questionario ROSSO.

1. Popolazione di riferimento

Tavola 1 - Popolazione di riferimento per sesso, classe di età, regione, ripartizione geografica e tipo di comune - Anno 2006 (a) (dati in migliaia)

	Famiglie	Individui								
		Maschi			Femmine			Maschi e femmine		
		Totale	6 anni e più	15 anni e più	Totale	6 anni e più	15 anni e più	Totale	6 anni e più	15 anni e più
REGIONI										
Piemonte	1.865	2.093	1.979	1.809	2.211	2.103	1.946	4.304	4.082	3.755
Valle d'Aosta	55	61	57	52	63	59	54	123	116	106
Lombardia	3.879	4.608	4.328	3.952	4.813	4.548	4.202	9.421	8.876	8.154
Trentino-Alto Adige	392	481	448	400	496	465	418	977	913	818
<i>Bozano-Bozen</i>	188	237	220	195	242	226	202	479	446	397
<i>Trento</i>	204	244	228	205	254	239	216	498	466	421
Veneto	1.833	2.310	2.170	1.981	2.391	2.258	2.072	4.701	4.428	4.053
Friuli-Venezia Giulia	509	578	547	502	617	588	547	1.195	1.135	1.049
Liguria	742	759	722	662	837	801	748	1.595	1.523	1.409
Emilia-Romagna	1.724	2.029	1.913	1.753	2.136	2.027	1.883	4.165	3.941	3.636
Toscana	1.484	1.739	1.644	1.514	1.861	1.772	1.656	3.600	3.416	3.170
Umbria	348	418	395	363	445	423	390	864	819	753
Marche	591	742	701	638	780	741	685	1.522	1.442	1.323
Lazio	2.132	2.526	2.372	2.135	2.740	2.593	2.373	5.266	4.965	4.509
Abruzzo	497	632	598	541	667	635	581	1.300	1.233	1.122
Molise	122	155	147	133	164	156	142	319	304	275
Campania	1.936	2.804	2.608	2.294	2.965	2.778	2.469	5.769	5.386	4.762
Puglia	1.451	1.969	1.847	1.642	2.085	1.968	1.770	4.054	3.815	3.412
Basilicata	219	290	274	247	301	286	260	591	559	507
Calabria	742	971	914	815	1.020	966	871	1.991	1.880	1.686
Sicilia	1.907	2.414	2.259	1.996	2.578	2.430	2.181	4.992	4.689	4.177
Sardegna	628	809	767	694	838	799	734	1.647	1.567	1.429
RIPARTIZIONI GEOGRAFICHE										
Italia nord-occidentale	6.541	7.520	7.086	6.474	7.923	7.511	6.950	15.443	14.597	13.424
Italia nord-orientale	4.458	5.398	5.078	4.635	5.640	5.338	4.921	11.038	10.416	9.556
Italia centrale	4.555	5.424	5.112	4.651	5.827	5.530	5.104	11.251	10.641	9.755
Italia meridionale	4.967	6.822	6.388	5.672	7.201	6.788	6.092	14.023	13.176	11.764
Italia insulare	2.535	3.223	3.026	2.690	3.416	3.230	2.916	6.639	6.255	5.606
TIPICI DI COMUNE										
Comune centro dell'area metropolitana	3.870	4.240	4.014	3.658	4.578	4.375	4.078	8.818	8.389	7.736
Periferia dell'area metropolitana	2.626	3.459	3.244	2.895	3.698	3.453	3.101	7.157	6.698	5.996
Fino a 2.000 abitanti	1.416	1.748	1.648	1.501	1.788	1.704	1.562	3.536	3.352	3.063
Da 2.001 a 10.000 abitanti	5.601	7.206	6.769	6.095	7.542	7.086	6.454	14.748	13.855	12.549
Da 10.001 a 50.000 abitanti	5.718	7.192	6.738	6.123	7.408	7.024	6.439	14.599	13.762	12.561
50.001 abitanti e più	3.827	4.543	4.275	3.851	4.992	4.754	4.348	9.535	9.029	8.199
Italia	23.057	28.387	26.689	24.122	30.007	28.396	25.982	58.394	55.085	50.105

(a) Popolazione al 01/04/2006.

2. Strategia di campionamento e livello di precisione dei risultati

2.1. Obiettivi conoscitivi

La popolazione di interesse dell'indagine in oggetto, 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 (a 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.

2.2. Strategia di campionamento

2.2.1 Descrizione generale del disegno di campionamento

Il disegno di campionamento è di tipo complesso e si avvale di due differenti schemi di campionamento. Nell'ambito di ognuno dei domini definiti dall'incrocio della regione geografica con le sei aree A₁, A₂, B₁, B₂, B₃ e B₄, i comuni sono suddivisi in due sottoinsiemi sulla base della popolazione residente:

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

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

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

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

2.2.2 *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 che quella delle stime a livello di ciascuno dei domini territoriali descritti nel paragrafo 1.

2.2.3 *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 Nar;
- scelta di un numero minimo di famiglie da intervistare in ciascun comune campione; tale numero è stato posto pari a 23;
- 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 A_1 , A_2 , B_1 , B_2 , B_3 e B_4 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 Nar : i comuni di dimensione superiore o uguale a ${}_r\lambda$ sono definiti come comuni A_r e i rimanenti come Nar ;
- suddivisione dei comuni dell'insieme Nar in strati aventi dimensione, in termini di popolazione residente, approssimativamente costante e all'incirca pari alla soglia ${}_r\lambda$.

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

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

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

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

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

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

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

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

REGIONI	Comuni		Famiglie		Individui	
	Universo	Campione	Universo (a)	Campione	Universo (a)	Campione
Piemonte	1.206	60	1.861.408	1.403	4.303.650	3.314
Valle d'Aosta	74	18	55.090	410	123.163	899
Lombardia	1.545	82	3.871.487	1.666	9.420.525	4.056
<i>Bolzano Bozen</i>	<i>116</i>	<i>22</i>	<i>187.344</i>	<i>588</i>	<i>478.960</i>	<i>1.517</i>
<i>Trento</i>	<i>223</i>	<i>26</i>	<i>204.097</i>	<i>529</i>	<i>497.945</i>	<i>1.331</i>
Veneto	581	53	1.831.721	1.199	4.701.139	3.108
Friuli-Venezia Giulia	219	32	509.258	707	1.194.916	1.685
Liguria	235	25	740.822	821	1.595.460	1.755
Emilia-Romagna	341	48	1.723.738	1.170	4.165.010	2.809
Toscana	287	52	1.479.477	1.172	3.600.090	2.823
Umbria	92	22	347.724	597	863.521	1.474
Marche	246	37	590.820	825	1.521.631	2.091
Lazio	378	35	2.129.517	1.100	5.265.717	2.762
Abruzzo	305	38	496.313	877	1.299.727	2.270
Molise	136	24	121.878	601	318.970	1.562
Campania	551	55	1.938.256	1.354	5.768.614	4.002
Puglia	258	49	1.448.127	1.129	4.054.436	3.153
Basilicata	131	26	219.149	624	590.712	1.677
Calabria	409	42	740.879	989	1.990.874	2.718
Sicilia	390	54	1.904.510	1.312	4.991.885	3.395
Sardegna	377	40	623.682	848	1.647.077	2.168
Italia	8.100	840	23.025.296	19.921	58.394.022	50.569

(a) Stima Indagine Multiscopo, dati in migliaia.

2.2.4 Procedimento per il calcolo delle stime

Le stime prodotte dall'indagine sono essenzialmente stime di frequenze assolute e relative, riferite alle famiglie e agli individui.

Le stime sono ottenute mediante uno stimatore di ponderazione vincolata, che è il metodo di stima adottato per la maggior parte delle indagini Istat sulle imprese e sulle famiglie.

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. Se, per esempio, a un'unità campionaria viene attribuito un peso pari a 30, allora questa unità rappresenta se stessa e altre 29 unità della popolazione che non sono state incluse nel campione.

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

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

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 unità² 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 18 totali noti, che si riferiscono alla distribuzione della popolazione regionale per sesso e sei classi di età³ e della popolazione regionale nelle sei aree A_1, A_2, B_1, B_2, B_3 e B_4 . Indicando, quindi, con ${}_kX$ ($k=1, \dots, 18$) 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 hij , la condizione sopra descritta è espressa dalla seguente uguaglianza

$${}_kX = \sum_{h=1}^H \sum_{i=1}^{n_h} \sum_{j=1}^{m_{hi}} W_{hijk} X_{hij} \quad (k=1, \dots, 18)$$

in cui H indica il numero complessivo di strati definiti nella regione. Se, ad esempio, ${}_6X$ indica il numero di maschi di età maggiore o uguale a sessantacinque anni, la variabile ausiliaria ${}_6X_{hij}$ rappresenta il numero di maschi di età maggiore o uguale a sessantacinque anni della famiglia hij .

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) 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).

² Al fine di ottenere stime coerenti per individui e famiglie i pesi finali sono definiti in modo tale che a ciascuna famiglia hij e a tutti i componenti della stessa sia assegnato un medesimo peso finale W_{hij} .

³ Le classi di età considerate sono: 0-5 anni, 6-13 anni, 14-24 anni, 25-44 anni, 45-64 anni, 65 anni e più.

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 vincolata.⁴ 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 nel paragrafo 3, 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.

2.3. Valutazione del livello di precisione delle stime

2.3.1 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 $\hat{\text{Var}}(\hat{Y}_d)$ 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:

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

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 nel paragrafo 2.4, 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 $\hat{\text{Var}}(\hat{Y}_d)$ 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 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:

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

dove Z_{hij} è la variabile linearizzata espressa come $Z_{hij} = Y_{hij} - \mathbf{X}'_{hij} \beta$, essendo $\mathbf{X}_{hij} = (x_{hij1}, \dots, x_{hijK})'$ il vettore contenente i valori delle K (K=18) 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

⁴ Nella letteratura in lingua anglosassone sull'argomento tali stimatori sono noti come *calibration estimators*.

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

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, $\hat{\text{Var}}(\hat{Z}_h)$, della stima \hat{Z}_h è differente a seconda che lo strato sia Ar oppure Nar. Possiamo, quindi scomporre come segue

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

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:

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

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:

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

dove le quantità sono espresse come:

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

Utilizzando le espressioni (8) e (9) è possibile, infine, calcolare la varianza di campionamento, $\hat{\text{Var}}(\hat{Y}_d)$, 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 $k=1.96$.

2.3.2 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 3.1 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.

2.3.3. 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 seconda riga del prospetto 2 alla voce Persone (a = 8,886722, b = -1,121521).

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

ZONE TERRITORIALI	Famiglie			Persone		
	a	b	R ² (%)	a	b	R ² (%)
ITALIA	8,5366571	-1,0937177	97,6276497	9,49947323	-1,159486	92,5635167
RIPARTIZIONI GEOGRAFICHE						
Nord-ovest	8,34598436	-1,0705	97,0731309	9,41064682	-1,1597838	91,2210678
Nord-est	7,74912049	-1,0528472	96,5718251	9,08977909	-1,161757	90,4865028
Centro	8,18498033	-1,0809565	97,2715278	9,55820314	-1,1919948	91,4326293
Sud	7,9268767	-1,0711444	95,8930991	8,55897167	-1,1198254	91,1083532
Isole	7,74741851	-1,0568943	95,5428964	8,5920424	-1,1283838	90,8634959
TIPI DI COMUNE						
A1	8,7369616	-1,1111095	97,9158277	9,81868483	-1,198316	94,0191813
A2	7,98141709	-1,0568878	94,8495626	9,55344841	-1,1897098	91,6521832
B1	6,26525231	-0,940206	94,7993493	7,73734474	-1,0662611	87,2113286
B2	7,57440514	-1,0257911	95,3917938	8,59707203	-1,1071637	90,553972
B3	8,04283568	-1,0679424	95,7848534	8,85196918	-1,1253693	91,1012546
B4	8,07344069	-1,0947271	97,7100428	9,31938485	-1,1926978	93,7236983
REGIONI						
Piemonte	8,11214816	-1,0955922	97,6104938	8,64522886	-1,1406719	92,3600753
Valle d'Aosta	5,29288259	-1,0796114	92,7388762	6,11246563	-1,1699455	89,5738863
Lombardia	8,5809095	-1,0728186	96,4489689	9,38704015	-1,1436839	90,0258024
<i>Bolzano</i>	6,12085009	-1,0497789	92,2812458	7,0608333	-1,1448238	87,0189691
<i>Trento</i>	6,63679423	-1,0876609	94,4761017	7,00604689	-1,1307736	89,1462479
Veneto	7,87634701	-1,0477542	95,1694638	8,66118005	-1,1163312	88,6027492
Friuli-Venezia Giulia	7,44502643	-1,0943009	95,4189159	8,01482998	-1,1454262	89,1992622
Liguria	7,28223545	-1,0580607	94,0488365	8,02934534	-1,1353029	92,7089792
Emilia-Romagna	8,35508805	-1,110843	97,6152868	9,00059557	-1,1632615	92,1053756
Toscana	8,142712	-1,0996061	97,4854203	9,1090618	-1,1855081	90,8934985
Umbria	7,20762463	-1,1029301	95,6394905	7,93744541	-1,1702792	92,2819119
Marche	7,56049498	-1,1169678	96,8749273	7,86146014	-1,1323397	92,1719706
Lazio	8,58473138	-1,0972903	96,444381	9,43720352	-1,1663421	90,834998
Abruzzo	6,90130204	-1,0491355	93,6788976	7,37842868	-1,0952489	90,8826233
Molise	5,77473492	-1,0619637	93,8476036	6,39583043	-1,1229709	89,208113
Campania	8,11911147	-1,0734607	95,09408	9,50005204	-1,1833922	91,6016117
Puglia	7,73350229	-1,0590572	94,361328	8,0774015	-1,0852111	91,1047576
Basilicata	6,37932227	-1,0580226	93,1824367	6,97217708	-1,1204364	88,8676114
Calabria	7,36827914	-1,0676314	94,7946171	7,83729734	-1,1008724	89,0832225
Sicilia	8,10264277	-1,0793527	95,5408474	8,69529752	-1,1250359	91,2589306
Sardegna	7,23472288	-1,0623731	95,9230766	7,70718132	-1,1036304	92,6136024

(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

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

STIME	Piemonte	Valle d'Aosta	Lombardia	Bolzano	Trento	Veneto	Friuli-Venezia Giulia	Liguria	Emilia-Romagna	Toscana	Umbria
20.000	25,4	6,7	36,0	11,8	12,7	28,6	18,3	20,2	26,6	25,3	15,6
30.000	20,4	5,4	29,0	9,5	10,1	23,2	14,7	16,3	21,3	20,3	12,5
40.000	17,4	4,6	24,8	8,2	8,7	19,9	12,6	14,0	18,1	17,3	10,6
50.000	15,4	4,1	22,0	7,3	7,7	17,7	11,1	12,5	16,0	15,3	9,4
60.000	13,9	3,7	20,0	6,6	7,0	16,1	10,1	11,3	14,5	13,8	8,5
70.000	12,8	-	18,4	6,1	6,4	14,9	9,2	10,4	13,3	12,7	7,8
80.000	11,9	-	17,1	5,7	6,0	13,9	8,6	9,7	12,3	11,8	7,3
90.000	11,2	-	16,1	5,4	5,6	13,0	8,1	9,1	11,6	11,1	6,8
100.000	10,5	-	15,2	5,1	5,3	12,3	7,6	8,6	10,9	10,5	6,4
200.000	7,2	-	10,5	-	-	8,6	5,2	6,0	7,4	7,1	4,4
300.000	5,8	-	8,4	-	-	6,9	4,2	4,8	5,9	5,7	3,5
400.000	4,9	-	7,2	-	-	6,0	3,6	4,1	5,0	4,9	-
500.000	4,4	-	6,4	-	-	5,3	3,2	3,7	4,5	4,3	-
750.000	3,5	-	5,2	-	-	4,3	-	-	3,6	3,5	-
1.000.000	3,0	-	4,4	-	-	3,7	-	-	3,0	2,9	-
2.000.000	2,0	-	3,0	-	-	-	-	-	-	-	-

STIME	Marche	Lazio	Abruzzo	Molise	Campania	Puglia	Basilicata	Calabria	Sicilia	Sardegna
20.000	17,4	31,9	17,5	9,3	28,5	25,2	12,9	20,1	27,4	19,3
30.000	13,8	25,6	14,1	7,5	22,9	20,3	10,4	16,2	22,0	15,6
40.000	11,8	21,8	12,1	6,5	19,6	17,5	8,9	13,9	18,9	13,4
50.000	10,4	19,3	10,8	5,7	17,4	15,5	7,9	12,3	16,7	11,9
60.000	9,4	17,5	9,8	5,2	15,8	14,1	7,2	11,2	15,2	10,8
70.000	8,6	16,1	9,1	4,8	14,5	13,0	6,6	10,3	14,0	9,9
80.000	8,0	14,9	8,4	4,5	13,5	12,1	6,2	9,6	13,0	9,3
90.000	7,5	14,0	7,9	4,2	12,7	11,4	5,8	9,0	12,2	8,7
100.000	7,1	13,2	7,5	4,0	12,0	10,8	5,5	8,5	11,5	8,2
200.000	4,8	9,0	5,2	-	8,3	7,5	3,8	5,9	7,9	5,7
300.000	3,8	7,2	4,2	-	6,7	6,0	-	4,7	6,4	4,6
400.000	3,3	6,2	3,6	-	5,7	5,2	-	4,1	5,4	3,9
500.000	2,9	5,5	-	-	5,1	4,6	-	3,6	4,8	-
750.000	-	4,4	-	-	4,1	3,7	-	-	3,9	-
1.000.000	-	3,7	-	-	3,5	3,2	-	-	3,3	-
2.000.000	-	2,6	-	-	2,4	-	-	-	-	-

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

STIME	Marche	Lazio	Abruzzo	Molise	Campania	Puglia	Basilicata	Calabria	Sicilia	Sardegna
20.000	18,7	34,8	17,7	9,4	33,0	26,3	12,7	21,6	29,4	20,0
30.000	14,9	27,4	14,1	7,5	25,9	21,1	10,1	17,3	23,4	16,0
40.000	12,6	23,2	12,1	6,4	21,9	18,1	8,6	14,7	19,9	13,6
50.000	11,1	20,4	10,7	5,6	19,2	16,0	7,6	13,0	17,6	12,0
60.000	10,0	18,3	9,7	5,1	17,2	14,5	6,9	11,8	15,9	10,9
70.000	9,2	16,7	8,9	4,7	15,7	13,3	6,3	10,8	14,5	10,0
80.000	8,5	15,5	8,3	4,3	14,5	12,4	5,9	10,1	13,5	9,3
90.000	8,0	14,5	7,7	4,0	13,5	11,6	5,5	9,4	12,6	8,7
100.000	7,5	13,6	7,3	3,8	12,7	11,0	5,2	8,9	11,9	8,2
200.000	5,1	9,1	5,0	2,6	8,4	7,5	3,5	6,1	8,1	5,6
300.000	4,0	7,2	4,0	2,1	6,6	6,1	2,8	4,9	6,4	4,5
400.000	3,4	6,1	3,4	-	5,6	5,2	2,4	4,2	5,5	3,8
500.000	3,0	5,3	3,0	-	4,9	4,6	2,1	3,7	4,8	3,4
750.000	2,4	4,2	2,4	-	3,9	3,7	-	2,9	3,8	2,7
1.000.000	2,0	3,6	2,1	-	3,3	3,2	-	2,5	3,3	2,3
2.000.000	-	2,4	-	-	2,2	2,2	-	1,7	2,2	-
3.000.000	-	1,9	-	-	1,7	1,7	-	-	1,8	-
4.000.000	-	1,6	-	-	1,4	1,5	-	-	1,5	-
5.000.000	-	1,4	-	-	1,3	-	-	-	1,3	-

2.3.4 Esempi di calcolo degli errori campionari

2.3.4.1 Esempi relativi alle stime delle persone

Esempio 1

Nel 2006, le persone in Italia nella classe di età compresa tra i 20 e 24 anni che dichiarano di conoscere la lingua inglese sono 2.320.000. Si cerca il livello di stima che più si avvicina a 2.320.000 nella prima colonna del prospetto 4 per il totale Italia.

L'errore relativo percentuale della stima considerata è pari a 2,6 %.

L'errore assoluto sarà: $\sigma(2.320.000) = 0,026 \times 2.320.000 = 60.320$

L'intervallo di confidenza avrà come estremi : $2.320.000 - (1,96 \times 60.320) = 2.201.773$
 $2.320.000 + (1,96 \times 60.320) = 2.438.227$

Esempio 2

Considerando la stima precedente si possono ottenere valori più precisi dell'errore di campionamento operando mediante interpolazione lineare dei due livelli di stima consecutivi tra i quali è compreso il valore della stessa.

Tali livelli sono 2.000.000 e 3.000.000 ai quali corrispondono i valori percentuali 2,6 e 2,0.

L'errore relativo corrispondente a 2.320.000 è pari a :

$$\sigma(2.320.000) = 2,6 - (2,6 - 2,0) / (3.000.000 - 2.000.000) \times (2.320.000 - 2.000.000) = 2,41\%$$

Il corrispondente errore assoluto è $0,0241 \times 2.320.000 = 55.912$
e l'intervallo di confidenza avrà come estremi: $2.320.000 - (1,96 \times 55.912) = 2.210.413$
 $2.320.000 + (1,96 \times 55.912) = 2.429.588$

Esempio 3

Il calcolo dell'errore può essere effettuato, direttamente, tramite la funzione interpolante:

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

i cui parametri, riportati nel prospetto 2 alla riga Italia, sono i seguenti:

$$a = 9,49947323 \quad b = -1,159486$$

Per $\hat{Y} = 2.320.000$ si ha:

$$\hat{\varepsilon}(\hat{Y}) = \sqrt{\exp(9,49947323 - 1,159486 \times \log(2.320.000))} = 0,0235742$$

L'errore relativo percentuale è quindi pari al 2,34% e il calcolo dell'errore assoluto e dell'intervallo di confidenza è del tutto analogo a quello degli esempi 1 e 2.

3. Indicatori di qualità

Di seguito vengono riportati alcuni degli indicatori sullo svolgimento delle interviste relative all'indagine.

Tavola 1 - Persone secondo le modalità di risposta ai quesiti del questionario per intervista, sesso e classe di età - Anno 2006 (composizione percentuale)

CLASSI DI ETÀ	Ha risposto direttamente ai quesiti			Totale
	SI	No	Non indicato	
MASCHI				
0-14	22,1	61,9	16,0	100,0
15-24	58,9	38,4	2,7	100,0
25-34	64,8	32,5	2,7	100,0
35-44	72,6	24,3	3,1	100,0
45-54	74,1	23,2	2,8	100,0
55-64	79,3	18,7	2,1	100,0
65-74	83,7	14,2	2,1	100,0
75 e più	82,9	15,6	1,5	100,0
Totale	65,3	30,1	4,6	100,0
FEMMINE				
0-14	23,4	60,0	16,7	100,0
15-24	66,7	30,6	2,7	100,0
25-34	78,5	19,0	2,5	100,0
35-44	88,4	9,0	2,6	100,0
45-54	87,3	10,2	2,5	100,0
55-64	88,5	8,8	2,7	100,0
65-74	89,0	8,9	2,1	100,0
75 e più	81,3	15,3	3,4	100,0
Totale	75,5	20,0	4,5	100,0
MASCHI E FEMMINE				
0-14	22,7	61,0	16,3	100,0
15-24	62,7	34,6	2,7	100,0
25-34	71,7	25,7	2,6	100,0
35-44	80,5	16,7	2,8	100,0
45-54	80,8	16,5	2,6	100,0
55-64	84,0	13,6	2,4	100,0
65-74	86,6	11,3	2,1	100,0
75 e più	81,9	15,4	2,7	100,0
Totale	70,5	24,9	4,5	100,0

Tavola 2 - Persone secondo le modalità di risposta ai quesiti del questionario per intervista e regione - Anno 2006
(composizione percentuale)

REGIONI	Ha risposto direttamente ai quesiti			Totale
	Si	No	Non indicato	
Piemonte	71,6	25,1	3,3	100,0
Valle d'Aosta	67,1	27,3	5,7	100,0
Lombardia	71,5	23,2	5,3	100,0
Trentino-Alto Adige	65,9	31,9	2,2	100,0
<i>Bozano-Bozen</i>	68,3	27,6	4,2	100,0
<i>Trento</i>	63,1	36,9	-	100,0
Veneto	73,0	22,5	4,5	100,0
Friuli-Venezia Giulia	73,2	20,8	6,0	100,0
Liguria	75,2	20,8	4,0	100,0
Emilia-Romagna	75,2	22,1	2,8	100,0
Toscana	70,9	24,2	4,9	100,0
Umbria	71,5	23,4	5,1	100,0
Marche	68,1	28,9	2,9	100,0
Lazio	67,8	26,1	6,0	100,0
Abruzzo	70,9	25,3	3,8	100,0
Molise	70,4	25,1	4,5	100,0
Campania	66,9	26,4	6,6	100,0
Puglia	71,3	24,8	3,9	100,0
Basilicata	67,3	28,3	4,4	100,0
Calabria	73,3	20,7	6,0	100,0
Sicilia	66,5	29,2	4,3	100,0
Sardegna	74,9	20,9	4,2	100,0
Italia	70,5	24,9	4,5	100,0

Tavola 3 - Persone per modalità di compilazione del questionario autocompilato modello azzurro, sesso e classe di età - Anno 2006 (composizione percentuale)

CLASSI DI ETÀ	Compilato durante l'intervista				Modalità della compilazione				Motivo della mancata compilazione			
	No	Sì	Non indicato	Totale	Intervista alla persona	Intervista a un familiare	Compilato personalmente dal componente	Compilato da un familiare	Rifiuto	Assenza del rispondente	Non indicato	Totale
MASCHI												
0-14	42,0	49,8	8,2	100,0	9,0	32,5	11,5	38,3	1,1	0,1	7,7	100,0
15-24	48,2	46,4	5,3	100,0	19,9	15,8	52,2	6,5	1,0	0,6	4,0	100,0
25-34	43,0	52,5	4,5	100,0	26,5	14,0	50,5	3,8	0,8	0,7	3,8	100,0
35-44	41,6	54,7	3,7	100,0	30,8	11,0	50,5	3,3	1,2	0,4	2,8	100,0
45-54	44,5	51,7	3,8	100,0	31,5	9,8	50,7	3,8	1,0	0,2	3,1	100,0
55-64	38,9	58,1	3,0	100,0	39,3	9,0	44,5	3,7	1,1	..	2,4	100,0
65-74	26,7	70,2	3,1	100,0	55,0	9,1	27,8	4,8	0,9	0,1	2,3	100,0
75 e più	17,8	79,7	2,5	100,0	64,9	12,6	13,0	6,0	0,8	0,2	2,5	100,0
Totale	39,3	56,2	4,5	100,0	31,9	14,7	38,9	9,5	1,0	0,3	3,7	100,0
FEMMINE												
0-14	42,3	49,2	8,5	100,0	8,5	32,5	14,0	36,6	0,8	0,1	7,6	100,0
15-24	48,8	45,8	5,4	100,0	18,7	12,5	59,5	3,4	1,1	0,4	4,4	100,0
25-34	46,1	49,5	4,4	100,0	28,8	8,2	56,1	2,4	0,8	0,3	3,5	100,0
35-44	44,7	51,2	4,1	100,0	33,5	4,1	56,0	1,6	1,0	0,3	3,4	100,0
45-54	43,9	52,0	4,1	100,0	35,7	5,2	52,1	2,5	0,9	0,2	3,3	100,0
55-64	34,7	61,9	3,4	100,0	48,8	5,3	39,0	3,3	0,8	0,1	2,8	100,0
65-74	20,3	76,5	3,2	100,0	65,4	6,0	19,8	5,2	1,0	-	2,6	100,0
75 e più	13,6	83,6	2,8	100,0	69,6	12,1	6,7	8,4	0,6	0,1	2,5	100,0
Totale	37,5	58,0	4,5	100,0	37,9	10,6	38,7	8,0	0,9	0,2	3,8	100,0
MASCHI E FEMMINE												
0-14	42,1	49,5	8,3	100,0	8,7	32,5	12,7	37,5	0,9	0,1	7,7	100,0
15-24	48,5	46,1	5,4	100,0	19,3	14,2	55,8	5,0	1,0	0,5	4,2	100,0
25-34	44,6	51,0	4,5	100,0	27,7	11,1	53,3	3,1	0,8	0,5	3,6	100,0
35-44	43,2	52,9	3,9	100,0	32,2	7,6	53,2	2,5	1,1	0,4	3,1	100,0
45-54	44,2	51,8	4,0	100,0	33,6	7,4	51,4	3,1	1,0	0,2	3,2	100,0
55-64	36,7	60,1	3,2	100,0	44,2	7,1	41,7	3,5	0,9	..	2,6	100,0
65-74	23,3	73,6	3,1	100,0	60,6	7,4	23,5	5,0	1,0	..	2,5	100,0
75 e più	15,2	82,1	2,7	100,0	67,8	12,3	9,2	7,4	0,7	0,2	2,5	100,0
Totale	38,4	57,1	4,5	100,0	35,0	12,6	38,8	8,7	0,9	0,2	3,8	100,0

(..) I valori non raggiungono la metà della cifra dell'ordine minimo considerato.

Tavola 4 - Persone per modalità di compilazione del questionario autocompilato modello azzurro e regione - Anno 2006 (composizione percentuale)

REGIONI	Compilato durante l'intervista				Modalità della compilazione				Motivo della mancata compilazione			
	No	Sì	Non indicato	Totale	Intervista alla persona	Intervista a un familiare	Compilato personalmente dal componente	Compilato da un familiare	Rifiuto	Assenza del rispondente	Non indicato	Totale
Piemonte	35,5	60,8	3,7	100,0	36,8	12,5	37,9	8,2	0,9	0,2	3,4	100,0
Valle d'Aosta	47,6	49,1	3,3	100,0	27,4	8,9	47,4	11,8	1,6	0,3	2,7	100,0
Lombardia	43,3	50,3	6,5	100,0	30,9	9,1	44,6	8,4	1,5	0,2	5,4	100,0
Trentino-Alto Adige	52,8	44,9	2,2	100,0	29,8	8,8	45,6	11,3	1,9	0,6	2,0	100,0
<i>Bolzano-Bozen</i>	<i>37,3</i>	<i>59,3</i>	<i>3,4</i>	<i>100,0</i>	<i>37,6</i>	<i>11,1</i>	<i>35,7</i>	<i>11,5</i>	<i>0,8</i>	<i>0,3</i>	<i>3,1</i>	<i>100,0</i>
<i>Trento</i>	<i>70,5</i>	<i>28,5</i>	<i>1,0</i>	<i>100,0</i>	<i>20,9</i>	<i>6,3</i>	<i>56,9</i>	<i>11,2</i>	<i>3,1</i>	<i>1,0</i>	<i>0,7</i>	<i>100,0</i>
Veneto	46,8	51,3	2,0	100,0	32,2	8,8	48,0	9,3	0,2	0,1	1,4	100,0
Friuli-Venezia Giulia	48,2	47,2	4,6	100,0	31,3	7,7	44,4	9,6	3,4	0,4	3,2	100,0
Liguria	36,5	61,1	2,3	100,0	36,6	9,7	43,6	7,3	0,3	0,2	2,3	100,0
Emilia-Romagna	42,5	55,1	2,4	100,0	30,1	10,4	48,6	8,2	0,6	0,1	2,1	100,0
Toscana	36,0	58,8	5,2	100,0	35,2	11,2	39,7	8,6	0,7	0,4	4,2	100,0
Umbria	40,2	54,7	5,0	100,0	35,9	10,7	40,1	8,0	0,5	0,4	4,4	100,0
Marche	44,1	51,5	4,4	100,0	31,8	12,8	41,0	10,9	0,1	0,1	3,3	100,0
Lazio	41,6	53,4	4,9	100,0	32,5	13,0	39,8	8,9	1,4	0,1	4,2	100,0
Abruzzo	35,1	61,5	3,4	100,0	42,5	14,9	31,6	8,1	0,0	0,3	2,6	100,0
Molise	38,0	56,5	5,5	100,0	38,3	13,3	34,3	9,7	1,2	0,3	3,0	100,0
Campania	23,8	70,3	5,9	100,0	40,9	23,0	25,3	5,3	0,2	..	5,3	100,0
Puglia	39,6	55,3	5,1	100,0	33,0	12,5	38,8	10,1	2,1	0,3	3,2	100,0
Basilicata	30,4	65,3	4,4	100,0	36,6	15,1	34,0	9,8	-	0,1	4,4	100,0
Calabria	31,0	63,2	5,8	100,0	41,7	12,6	29,4	9,7	0,4	0,1	6,1	100,0
Sicilia	30,0	64,8	5,2	100,0	39,5	18,1	29,9	7,3	0,3	0,1	4,7	100,0
Sardegna	36,5	57,4	6,1	100,0	31,2	10,2	42,3	8,6	2,1	0,6	4,9	100,0
Italia	38,4	57,1	4,5	100,0	35,0	12,6	38,8	8,7	0,9	0,2	3,8	100,0

(..) I valori non raggiungono la metà della cifra dell'ordine minimo considerato.

Tavola 5 - Persone per modalità di compilazione del questionario autocompilato modello giallo, sesso e classe di età - Anno 2006 (composizione percentuale)

CLASSI DI ETÀ	Compilato durante l'intervista				Modalità della compilazione				Motivo della mancata compilazione			
	No	Sì	Non indicato	Totale	Intervista alla persona	Intervista a un familiare	Compilato personalmente dal componente	Compilato da un familiare	Rifiuto	Assenza del rispondente	Non indicato	Totale
MASCHI												
18-24	45,4	47,1	7,5	100,0	18,4	16,5	51,7	5,9	1,0	0,7	5,8	100,0
25-34	42,3	52,4	5,3	100,0	27,1	14,1	49,3	4,1	0,9	0,5	3,9	100,0
35-44	40,4	55,4	4,2	100,0	31,5	11,2	49,1	3,4	0,9	0,4	3,5	100,0
45-54	43,7	52,0	4,3	100,0	31,4	9,9	50,1	3,8	1,1	0,1	3,4	100,0
55-64	37,9	58,4	3,8	100,0	39,4	9,3	43,8	3,4	1,1	0,1	3,0	100,0
65-74	25,7	71,0	3,4	100,0	54,8	9,3	28,1	4,4	0,9	0,1	2,5	100,0
75 e più	16,9	80,5	2,6	100,0	66,2	12,8	12,0	5,6	0,8	0,3	2,3	100,0
Totale	37,4	58,3	4,4	100,0	36,8	11,6	42,7	4,2	1,0	0,3	3,5	100,0
FEMMINE												
18-24	47,2	46,1	6,6	100,0	19,1	12,5	57,9	3,6	0,9	0,3	5,6	100,0
25-34	45,5	49,9	4,6	100,0	28,6	8,6	55,7	2,4	0,7	0,3	3,8	100,0
35-44	43,1	51,8	5,1	100,0	34,6	4,1	54,5	1,4	0,9	0,3	4,1	100,0
45-54	42,6	52,8	4,6	100,0	36,8	5,5	50,5	2,5	0,8	0,1	3,7	100,0
55-64	33,9	62,4	3,7	100,0	48,9	5,5	38,5	3,2	0,7	0,1	3,1	100,0
65-74	20,0	76,8	3,2	100,0	66,2	5,9	19,3	5,2	0,8	-	2,7	100,0
75 e più	12,5	84,1	3,3	100,0	70,3	12,4	6,3	7,8	0,5	0,1	2,7	100,0
Totale	35,3	60,3	4,4	100,0	43,8	7,3	40,9	3,5	0,8	0,2	3,6	100,0
MASCHI E FEMMINE												
18-24	46,3	46,6	7,1	100,0	18,7	14,5	54,8	4,8	1,0	0,5	5,7	100,0
25-34	43,9	51,1	5,0	100,0	27,9	11,3	52,5	3,2	0,8	0,4	3,8	100,0
35-44	41,7	53,6	4,7	100,0	33	7,7	51,8	2,4	0,9	0,3	3,8	100,0
45-54	43,1	52,4	4,5	100,0	34,2	7,7	50,3	3,2	1,0	0,1	3,6	100,0
55-64	35,8	60,4	3,7	100,0	44,3	7,3	41,1	3,3	0,9	0,1	3,1	100,0
65-74	22,6	74,1	3,3	100,0	60,9	7,4	23,4	4,8	0,8	..	2,6	100,0
75 e più	14,2	82,7	3,1	100,0	68,7	12,5	8,5	6,9	0,6	0,2	2,5	100,0
Totale	36,3	59,3	4,4	100,0	40,5	9,3	41,8	3,8	0,9	0,2	3,5	100,0

(..) I valori non raggiungono la metà della cifra dell'ordine minimo considerato.

Tavola 6 - Persone per modalità di compilazione del questionario autocompilato modello giallo e regione - Anno 2006 (composizione percentuale)

REGIONI	Compilato durante l'intervista				Modalità della compilazione				Motivo della mancata compilazione			
	No	Sì	Non indicato	Totale	Intervista alla persona	Intervista a un familiare	Compilato personalmente dal componente	Compilato da un familiare	Rifiuto	Assenza del rispondente	Non indicato	Totale
Piemonte	34,1	61,9	4,0	100,0	42,1	9,2	41,4	3,1	0,7	0,2	3,4	100,0
Valle d'Aosta	42,2	54,6	3,2	100,0	34,9	6,8	49,3	3,7	1,9	0,4	3,1	100,0
Lombardia	42,1	51,9	6,0	100,0	36,1	6,1	48,0	2,8	1,8	0,2	5,1	100,0
Trentino-Alto Adige	49,1	48,5	2,5	100,0	36,9	5,8	49,5	3,8	1,4	0,7	2,0	100,0
<i>Bolzano-Bozen</i>	<i>34,4</i>	<i>63,0</i>	<i>2,7</i>	<i>100,0</i>	<i>45,9</i>	<i>6,6</i>	<i>39,5</i>	<i>4,5</i>	<i>0,9</i>	<i>0,3</i>	<i>2,3</i>	<i>100,0</i>
<i>Trento</i>	<i>65,1</i>	<i>32,6</i>	<i>2,3</i>	<i>100,0</i>	<i>27,0</i>	<i>5,0</i>	<i>60,4</i>	<i>2,9</i>	<i>1,9</i>	<i>1,1</i>	<i>1,7</i>	<i>100,0</i>
Veneto	43,2	54,3	2,6	100,0	37,8	6,9	50,4	2,5	0,2	0,1	2,0	100,0
Friuli-Venezia Giulia	47,5	48,4	4,1	100,0	36,4	5,0	47,2	4,6	2,9	0,3	3,5	100,0
Liguria	34,4	63,2	2,4	100,0	41,3	7,8	45,8	2,9	0,3	0,3	1,6	100,0
Emilia-Romagna	39,6	58,2	2,2	100,0	35,2	7,3	52,0	3,0	0,5	0,1	2,0	100,0
Toscana	33,3	61,8	4,9	100,0	40,2	9,4	41,8	3,8	0,7	0,4	3,8	100,0
Umbria	37,4	58,7	3,9	100,0	42,2	7,4	41,9	4,2	0,6	0,4	3,4	100,0
Marche	42,7	53,0	4,3	100,0	36,5	9,6	44,3	6,0	0,1	0,1	3,4	100,0
Lazio	39,4	56,2	4,3	100,0	38,5	10,2	42,4	4,0	1,3	0,1	3,5	100,0
Abruzzo	31,5	64,9	3,6	100,0	48,8	10,9	33,6	3,8	0,1	0,3	2,6	100,0
Molise	36,7	59,2	4,2	100,0	42,5	11,5	36,0	6,2	1,5	0,5	2,0	100,0
Campania	24,3	71,1	4,6	100,0	48,0	17,2	28,1	2,4	0,2	-	4,0	100,0
Puglia	38,1	56,6	5,3	100,0	35,7	9,0	44,9	4,9	1,9	0,4	3,4	100,0
Basilicata	28,2	66,1	5,6	100,0	42,3	12,3	36,4	4,4	-	0,1	4,5	100,0
Calabria	29,7	63,4	7,0	100,0	47,0	8,1	32,2	5,2	0,4	0,1	7,0	100,0
Sicilia	26,8	67,6	5,7	100,0	46,5	13,9	31,6	3,5	0,2	0,1	4,2	100,0
Sardegna	33,1	61,4	5,5	100,0	36,7	7,2	44,0	5,5	1,7	0,4	4,5	100,0
Italia	36,3	59,3	4,4	100,0	40,5	9,3	41,8	3,8	0,9	0,2	3,5	100,0

Tavola 7 - Famiglie con almeno due componenti per numero di componenti che hanno fornito le notizie sulla scheda familiare e regione - Anno 2006 (composizione percentuale)

REGIONI	Numero di rispondenti			Totale
	Un componente	Più componenti	Non indicato	
Piemonte	49,0	49,7	1,4	100,0
Valle d'Aosta	47,3	50,8	1,9	100,0
Lombardia	45,7	51,1	3,2	100,0
Trentino-Alto Adige	51,5	45,7	2,8	100,0
<i>Bozano-Bozen</i>	<i>50,7</i>	<i>48,3</i>	<i>1,0</i>	<i>100,0</i>
<i>Trento</i>	<i>52,4</i>	<i>42,9</i>	<i>4,7</i>	<i>100,0</i>
Veneto	43,0	56,5	0,5	100,0
Friuli-Venezia Giulia	42,0	56,3	1,8	100,0
Liguria	51,8	46,9	1,3	100,0
Emilia-Romagna	50,9	48,4	0,7	100,0
Toscana	46,3	51,2	2,5	100,0
Umbria	51,9	47,5	0,7	100,0
Marche	51,7	45,6	2,7	100,0
Lazio	57,9	39,5	2,6	100,0
Abruzzo	48,3	50,8	0,9	100,0
Molise	51,6	47,7	0,7	100,0
Campania	48,9	48,7	2,4	100,0
Puglia	48,0	49,4	2,6	100,0
Basilicata	51,4	48,0	0,6	100,0
Calabria	48,3	49,2	2,5	100,0
Sicilia	59,8	37,7	2,5	100,0
Sardegna	45,0	53,9	1,1	100,0
Italia	49,5	48,6	1,9	100,0

Tavola 8 - Famiglie per disponibilità all'intervista e regione - Anno 2006 (composizione percentuale)

REGIONI	Disponibilità all'intervista							Totale
	Sempre scarsa	Scarsa all'inizio poi migliore	Buona all'inizio poi scarsa	Sufficiente	Buona	Molto interessata	Non indicato	
Piemonte	4,1	3,8	3,5	14,9	51,0	12,2	10,5	100,0
Valle d'Aosta	1,7	0,6	2,1	13,0	73,4	3,8	5,5	100,0
Lombardia	2,4	4,9	2,5	13,1	50,9	14,3	11,9	100,0
Trentino-Alto Adige	2,6	3,2	1,9	21,5	49,0	15,4	6,4	100,0
<i>Bozano-Bozen</i>	2,4	4,1	3,4	31,7	39,9	9,8	8,6	100,0
<i>Trento</i>	2,7	2,2	0,2	9,8	59,4	21,7	3,9	100,0
Veneto	3,6	5,3	3,6	11,6	51,0	15,3	9,7	100,0
Friuli-Venezia Giulia	3,3	1,9	3,2	10,7	58,6	12,8	9,5	100,0
Liguria	1,5	2,5	1,5	16,2	56,3	11,2	10,8	100,0
Emilia-Romagna	3,5	5,1	2,4	15,6	51,1	11,0	11,4	100,0
Toscana	1,7	4,5	3,3	14,4	51,1	13,9	11,1	100,0
Umbria	1,6	3,0	1,8	19,3	54,5	7,4	12,5	100,0
Marche	2,3	5,6	1,1	13,4	48,3	18,1	11,1	100,0
Lazio	3,0	6,2	2,7	15,0	47,1	12,2	13,8	100,0
Abruzzo	1,4	3,7	4,1	21,3	51,6	10,4	7,6	100,0
Molise	3,1	4,0	5,5	16,8	42,3	15,7	12,6	100,0
Campania	3,1	7,8	4,0	21,7	40,3	8,1	14,8	100,0
Puglia	3,6	3,7	1,8	23,2	43,9	10,6	13,2	100,0
Basilicata	2,0	4,1	5,3	16,2	39,7	22,2	10,4	100,0
Calabria	2,1	5,8	3,7	23,6	39,3	12,9	12,6	100,0
Sicilia	2,9	8,4	2,4	21,8	36,5	12,2	15,7	100,0
Sardegna	2,6	4,0	3,6	18,1	43,8	17,1	10,9	100,0
Italia	2,7	4,8	3,0	17,4	47,7	12,9	11,5	100,0

Glossario

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

- **l'età** è espressa in anni compiuti;
- **il titolo di studio** è quello più elevato conseguito;
- **la condizione** è quella dichiarata come unica o prevalente dalle persone di 15 anni e più.
Si precisa inoltre che per:

occupato si intende chi possiede un'occupazione in proprio o alle dipendenze da cui trae un profitto o una retribuzione (utile, onorario, stipendio, salario) o chi collabora con un familiare che svolge un'attività lavorativa in conto proprio senza avere un regolare contratto di lavoro (coadiuvante);

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

casalinga è chi si dedica prevalentemente alle faccende domestiche;

studente è chi si dedica prevalentemente allo studio;

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

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

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

dirigenti, imprenditori, liberi professionisti;

direttivi, quadri, impiegati;

capo operai, operai, subalterni (inclusi apprendisti, lavoratori a domicilio per conto di imprese);

lavoratori in proprio, coadiuvanti (inclusi soci di cooperative di produzione di beni e/o prestazioni di servizio).

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

Nord-ovest comprende: Piemonte, Valle d'Aosta, Lombardia, Liguria.

Nord-est comprende: Trentino-Alto Adige, Veneto, Friuli-Venezia Giulia, Emilia-Romagna.

Centro comprende: Toscana, Umbria, Marche, Lazio.

Sud comprende: Abruzzo, Molise, Campania, Puglia, Basilicata, Calabria.

Isole comprende: Sicilia, Sardegna.

- **il tipo di comune**

I comuni italiani sono suddivisi nelle seguenti classi:

comuni centro delle aree metropolitane: si tratta dei comuni di Torino, Milano, Venezia, Genova, Bologna, Firenze, Roma, Napoli, Bari, Palermo, Catania, Cagliari;

comuni appartenenti alla periferia delle aree metropolitane: costituiscono i comuni delle cinture urbane;

altri comuni: suddivisi per dimensione demografica (fino a 2.000 abitanti, da 2.001 a 10.000, da 10.001 a 50.000 e oltre i 50.000);

Si precisa che la soglia dei 2.000 abitanti costituisce la dimensione demografica suggerita dagli organismi internazionali per identificare uno stile di vita tipico dei piccoli centri.

- **famiglia e nucleo familiare**

Si precisa che:

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

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

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