



**Aspetti della vita quotidiana**  
2013

**Aspetti metodologici dell'indagine**

## INDICE

1. Introduzione .....	3
2. La popolazione di riferimento.....	3
3. Il disegno campionario.....	5
4. La rilevazione e il trattamento dei dati .....	9
5. La metodologia di calcolo dei pesi campionari .....	10
6. La diffusione dei dati dell'indagine .....	15
7. Glossario .....	15
8. Riferimenti bibliografici.....	16

## 1. Introduzione

L'indagine campionaria "Aspetti della vita quotidiana", che fa parte di un sistema integrato di indagini sociali - le **Indagini Multiscopo sulle famiglie** - rileva le informazioni fondamentali relative alla vita quotidiana degli individui e delle famiglie. A partire dal 1993, 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. La popolazione di riferimento

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. Il disegno campionario**

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 è considerato come uno strato a sé 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.

#### **3.1. 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.

### ***3.2. 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 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  $Nar$ : i comuni di dimensione superiore o uguale a  $r$  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 <sup>1</sup>.

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:

---

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

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

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 2013**

REGIONI	Comuni		Famiglie		Individui	
	Campione	Universo	Campione	Universo (a)	Campione	Universo (a)
Piemonte / Valle d'Aosta - Vallée d'Aoste	84	1.280	1.837	2.031	4.043	4.563
Liguria	26	235	805	775	1.668	1.599
Lombardia	83	1.544	1.657	4.238	3.948	9.984
Trentino-Alto Adige	47	333	1.106	426	2.720	1.042
Veneto	54	581	1.146	2.003	2.842	4.925
Friuli-Venezia Giulia	32	218	701	538	1.637	1.223
Emilia-Romagna	47	348	1.044	1.891	2.426	4.444
Toscana	46	287	1.014	1.608	2.362	3.750
Umbria	22	92	571	370	1.371	905
Marche	34	239	703	610	1.795	1.561
Lazio	36	378	1.197	2.489	2.691	5.791
Abruzzo	37	305	794	537	1.962	1.341
Molise	23	136	601	127	1.491	317
Campania	55	551	1.359	2.082	3.738	5.813
Puglia	50	258	1.067	1.528	2.857	4.068
Basilicata	26	131	587	231	1.485	582
Calabria	43	409	946	792	2.339	1.996
Sicilia	50	390	1.219	1.972	3.086	5.019
Sardegna	39	377	797	705	1.854	1.665
<b>Italia</b>	<b>834</b>	<b>8.092</b>	<b>19.151</b>	<b>24.952</b>	<b>46.315</b>	<b>60.588</b>

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

## 4. La rilevazione e il trattamento dei dati

Le informazioni vengono raccolte tramite due questionari: uno ROSA somministrato per intervista e uno VERDE che è compilato personalmente da ogni componente della famiglia. Il questionario ROSA, che rappresenta il questionario base della rilevazione, contiene i quesiti familiari e quattro schede individuali, una per ogni componente. Qualora i componenti siano più di quattro sono previste delle schede individuali aggiuntive di colore BIANCO. Occasionalmente sono presenti dei questionari aggiuntivi di approfondimento su temi specifici.

Per quanto riguarda, invece, i metadati dell'Indagine, nonché il contenuto informativo, le attività di prevenzione, controllo e valutazione dell'errore e gli indicatori di copertura e mancata risposta, si può consultare SIQual, il sistema di visualizzazione della documentazione delle Indagini Istat presente in SIDI (Sistema Informativo di Documentazione delle Indagini).

L'Istat è tenuto per legge a rispettare il segreto statistico (art. 9 del decreto legislativo n. 322/1989). I dati raccolti in occasione dell'indagine possono quindi essere utilizzati, esclusivamente a fini statistici, dall'Istat e dagli altri enti ed uffici di statistica facenti parte del Sistema statistico nazionale. I medesimi dati possono inoltre essere utilizzati, per sole finalità di ricerca scientifica, da soggetti non facenti parte del citato Sistema statistico nazionale nei limiti e secondo le modalità stabilite dall'art. 7 del Codice di deontologia e di buona condotta per i trattamenti di dati personali a scopi statistici e di ricerca scientifica effettuati nell'ambito del Sistema statistico. La diffusione dei dati e la loro comunicazione avviene sotto forma aggregata in modo tale che non sia possibile risalire alle persone che le forniscono, assicurando così la massima riservatezza. Ai sensi della stessa normativa (art.7, comma 2) la persona intervistata potrà decidere se rispondere o meno ad alcuni quesiti di natura "sensibile" contenuti nelle varie sezioni del questionario e indicati sul retro dei questionari. Ai sensi della legge che disciplina la tutela della privacy (D.Lgs.196/2003), titolare delle rilevazioni è l'Istat.

Usufruiscono dei dati Istituzioni, Enti di ricerca Nazionali e Internazionali, l'Ufficio di Statistica delle Comunità europee (Eurostat), Università, studenti e tutti i cittadini che ne sono interessati. Titolare del trattamento dei dati personali raccolti con la presente indagine è l'Istat - Istituto nazionale di statistica; responsabile del trattamento è il Direttore centrale delle statistiche socio-demografiche e ambientali al quale è possibile rivolgersi per quanto riguarda l'esercizio dei diritti degli interessati.

## 5. La metodologia di calcolo dei pesi campionari

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 unità<sup>2</sup> 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 20 totali noti, che si riferiscono alla distribuzione della popolazione regionale per sesso e sei classi di età<sup>3</sup>, 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  ${}_kX$  ( $k=1, \dots, 20$ ) 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

$${}_kX = \hat{X} = \sum_{h=1}^H \sum_{i=1}^{n_h} \sum_{j=1}^{m_{hi}} W_{hijk} X_{hij} \quad (k=1, \dots, 20)$$

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  $h_{ij}$ .

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

<sup>2</sup> 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  $W_{hij}$ .

<sup>3</sup> 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.

- 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 (opportuna 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 vincolata<sup>4</sup>. 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.

## **5.2. 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

---

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

stima  $\hat{Y}_d$ , la stima dell'errore di campionamento assoluto di  $\hat{Y}_d$  si può ottenere mediante la seguente espressione:

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

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  $\hat{V}\text{ar}(\hat{Y}_d)$  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 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  $\mathbf{X}_{hij} = (X_{hij,1}, \dots, X_{hij,K})'$  il vettore contenente i valori delle K (K=20) 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

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

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:

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

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

## 6. La diffusione dei risultati dell'indagine

I principali risultati dell'indagine vengono resi disponibili sul sito dell'Istat attraverso sia il Datawarehouse I.stat sia le statistiche report pubblicate nei settori con argomento: "Opinioni dei cittadini", "Salute e sanità", "Cultura, comunicazione, tempo libero", "Partecipazione sociale". I dati d'indagine vengono resi disponibili mediante il rilascio di file di microdati. Ricercatori e studiosi possono 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.

I principali risultati dell'Indagine vengono anche divulgati nei volumi generali dell'Istat. 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 comunicati 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.

## 7. 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** (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).

- **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;

## 8. Riferimenti bibliografici

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

## Contatti

### **Struttura e dinamica sociale**

Unità operativa

Qualità della vita quotidiana e benessere

**Sante Orsini**

Tel. +39 06 4673.7256

Email [orsini@istat.it](mailto:orsini@istat.it)

**Silvia Montecolle**

Tel. +39 06 4673.7361

Email [montecol@istat.it](mailto:montecol@istat.it)

## Curatori dei capitoli

Il documento è stato redatto da Silvia Montecolle