

Nota metodologica

Premessa

L'indagine sulle Condizioni di salute e ricorso ai servizi sanitari rileva lo stato di salute della popolazione, il ricorso ai principali servizi sanitari, alcuni fattori di rischio per la salute.

I risultati dell'indagine sono stati in parte presentati nel comunicato "Tutela della salute e accesso alle cure" (<http://www.istat.it/it/archivio/128176>), nella statistica report "Gravidanza, parto e allattamento al seno" (<http://www.istat.it/it/archivio/141431>), nelle tavole di dati "Condizioni di salute, fattori di rischio e prevenzione" (<http://www.istat.it/it/archivio/144093>) e "La cura e il ricorso ai servizi sanitari" (<http://www.istat.it/it/archivio/156420>).

Oggi si propongono i risultati relativi al ricorso alle cure odontoiatriche e alla salute dei denti in Italia evidenziandone anche i cambiamenti rispetto alla precedente edizione dell'indagine (2005).

L'indagine è di tipo campionario (Cfr. Strategia di campionamento e livello di precisione delle stime) ed è condotta mediante quattro distinte rilevazioni a cadenza trimestrale, anche per tener conto dell'effetto stagionale dei fenomeni, di particolare rilievo per le problematiche della salute. Con la finalità di soddisfare i bisogni informativi a livello territoriale e consentire stime regionali e sub-regionali, la numerosità campionaria è stata notevolmente ampliata grazie al contributo messo a disposizione dal Ministero della Salute e dalle Regioni, con il Piemonte in qualità di regione capofila.

Strategia di campionamento e livello di precisione dei risultati¹

Obiettivi dell'indagine e domini di stima

La popolazione di interesse dell'indagine su '*Condizioni di salute e ricorso ai servizi sanitari*' – ossia l'insieme delle unità statistiche oggetto di investigazione – è costituita dalle famiglie residenti in Italia e dagli individui che le compongono, al netto dei membri permanenti delle convivenze. L'unità di rilevazione è la famiglia, intendendo per famiglia la *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 dell'indagine è costituito dai dodici mesi che vanno da luglio 2012 a giugno 2013, mentre il periodo di riferimento dei fenomeni indagati varia da quesito a quesito.

Il disegno campionario è stato definito alla luce delle evidenze emerse riguardo alla necessità che i domini di stima di principale interesse per la programmazione sanitaria regionale, le ASL, venissero tenuti in considerazione nella definizione degli obiettivi dell'indagine e del disegno.

Per soddisfare i bisogni informativi a livello territoriale e consentire stime regionali e sub-regionali utili alla programmazione sanitaria locale, anche nel 2012-2013 - come per le due precedenti indagini del 1999/2000 e del 2004/2005 - è stata notevolmente ampliata la numerosità campionaria grazie al contributo del Ministero della Salute e delle Regioni, passando da una numerosità campionaria teorica complessiva nell'anno di 24.000 famiglie ad una di 60.000.

Per la progettazione del disegno campionario dell'indagine 2012-2013, i domini di studio, ossia gli ambiti territoriali ai quali sono riferiti i parametri di popolazione oggetto di stima, sono:

- i domini tradizionali delle indagini Multiscopo, ossia le cinque ripartizioni geografiche (Italia Nord-Occidentale, Italia Nord-Orientale, Italia Centrale, Italia Meridionale, Italia Insulare); le regioni (ad eccezione del Trentino Alto Adige le cui stime sono prodotte distintamente per le province autonome di Bolzano e Trento).
- domini di stima sub-regionali, nel seguito indicati come Aree Vaste, costituiti da aggregati territoriali di interesse per la programmazione sanitaria a livello locale e definiti in relazione allo specifico contesto informativo dell'indagine sulle condizioni di salute.

¹ La nota è a cura di Claudia De Vitiis, Francesca Inglese e Diego Moretti.

Questi ultimi domini sono stati definiti partendo dalla considerazione che, sebbene le unità amministrative territoriali di prevalente interesse per la programmazione sanitaria sono le Aziende Sanitarie Locali (ASL), tuttavia non era possibile progettare, per vincoli di costo, un disegno campionario che garantisse stime attendibili a tale livello di dettaglio. Pertanto, si è proceduto alla definizione di domini di stima ottenuti dalla aggregazione delle ASL. La definizione di tali macro-aree è stata effettuata sulla base del criterio di ampiezza demografica, tenendo conto del vincolo di numerosità complessiva e di rappresentatività del campione. La dimensione media di popolazione delle Aree Vaste è di circa 850.000 abitanti.

Rispetto alla rilevazione del 2004-2005, questa indagine ha beneficiato del fatto che negli ultimi anni le ASL hanno subito un processo di aggregazione, passando dalle 197 del 2004 a 145 nel 2010. Tale riduzione ha determinato una distribuzione della popolazione nelle ASL più omogenea che ha reso possibile la definizione di un nuovo disegno campionario che tenesse conto delle ASL in modo esplicito nella costruzione dei domini di stima sub-regionali. In tal modo la stratificazione dei comuni è stata effettuata all'interno delle ASL e tale soluzione garantisce che le ASL di maggiore dimensione abbiano una dimensione campionaria (nel campione riferito all'intero anno di rilevazione) sufficiente a garantire livelli di precisione delle stime accettabili.

Disegno di campionamento

Struttura generale del disegno

Il disegno di campionamento ha una struttura generale che ricalca quella degli schemi campionari della maggior parte delle indagini sulle famiglie, ossia un disegno a più stadi comuni-famiglie, con stratificazione dei comuni.

Nell'ambito di ogni Area Vasta i comuni universo sono stati suddivisi in due sottoinsiemi separatamente per ciascuna ASL: i comuni di maggiore dimensione demografica costituiscono strato 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 (quattro per ogni strato) vengono 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 - vengono selezionati in maniera sistematica 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 30.

Definizione della numerosità campionaria e allocazione tra i domini

Per la presente indagine, che rientra nel sistema delle Indagini Multiscopo sulle famiglie, è stato effettuato un ampliamento della numerosità campionaria, rispetto a quella standard delle altre indagini del sistema, a seguito di una convenzione cui partecipano il Ministero della Salute, Regioni e ISTAT. È stato pertanto necessario ridefinire la numerosità campionaria complessiva e la sua allocazione tra i diversi domini territoriali.

Tuttavia, per un'indagine con molteplici obiettivi di stima come quella in esame, è necessario individuare le stime e i domini territoriali di stima per i quali si richiede che gli errori campionari non siano superiori a certi limiti prefissati. Infatti, non è realistico pensare di poter disegnare una strategia campionaria che assicuri certi livelli di precisione a tutte le stime prodotte, considerando anche il fatto che le stime vengono prodotte con riferimenti territoriali differenti. L'allocazione ottimale delle unità del campione con riferimento a un dato tipo di dominio può risultare, infatti, contrastante con l'allocazione ottimale con riferimento a un altro tipo di dominio. 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 o di area vasta, una soluzione approssimativamente ottimale sarebbe quella di selezionare un campione uguale in tutti i domini. Quest'ultima soluzione, però, è poco efficiente per le stime a livello nazionale.

La dimensione complessiva del campione è stata fissata in circa 60.000 interviste, delle quali 24.000 corrispondenti al campione base dell'indagine Multiscopo annuale. Le restanti 36.000

costituiscono l'ampliamento, finanziato con fondi del Ministero della Salute. L'assegnazione della numerosità campionaria complessiva ai domini di stima (regioni e aree vaste) è stata effettuata in passi successivi, seguendo un'ottica di compromesso tra un'allocazione uniforme tra i diversi domini di stima e un'allocazione proporzionale alla popolazione.

Nel prospetto 1 è presentata la distribuzione del campione, in termini di famiglie e comuni, relativa alle Regioni e alle Aree Vaste. Sulla base dei numeri contenuti nella seconda colonna, si può osservare come la dimensione del campione di ogni regione sia il risultato di un compromesso tra la dimensione della regione in termini di popolazione e il numero di Aree Vaste che per la regione sono state definite. Per quanto riguarda le numerosità campionarie delle Aree Vaste, è immediato constatare come la variabilità della distribuzione sia alquanto contenuta; ciò è coerente con il criterio di privilegiare un'allocazione uniforme del campione tra i domini di stima sub-regionali.

PROSPETTO 1. – DISTRIBUZIONE REGIONALE DELLA POPOLAZIONE, DEL CAMPIONE DI FAMIGLIE E DI COMUNI E AREE VASTE

Regioni	Numero Aree vaste	Comuni campione	Famiglie campione teorico	Famiglie campione per area vasta:		Individui popolazione	Individui intervistati
				Minimo	Massimo		
Piemonte	5	104	4.344	800	976	4.443.892	7.976
Valle d'Aosta – Vallée d' Aoste	1	26	1.080	-	-	128.350	1.649
Liguria	3	43	2.564	800	920	1.601.910	4.373
Lombardia	8	173	6.596	688	1.164	9.996.322	12.125
Bolzano-Bozen	1	25	1.048	-	-	510.050	2.266
Trento	1	29	1.108	-	-	531.636	1.986
Veneto	5	106	3.972	764	820	4.931.841	8.375
Friuli-Venezia Giulia	2	50	2.112	988	1.124	1.226.714	4.132
Emilia-Romagna	5	96	4.260	740	912	4.451.888	7.301
Toscana	3	91	3.596	1.100	1.280	3.758.835	6.153
Umbria	1	31	1.420	-	-	906.510	2.906
Marche	2	52	2.144	1.044	1.100	1.563.894	4.233
Lazio	4	82	4.220	940	1.320	5.807.181	7.856
Abruzzo	2	53	2.140	1.052	1.088	1.343.325	4.003
Molise	1	28	1.104	-	-	317.132	2.375
Campania	6	111	4.464	708	770	5.818.893	10.657
Puglia	5	89	3.672	648	820	4.073.626	8.283
Basilicata	1	36	1.336	-	-	582.786	3.076
Calabria	2	65	2.308	1.080	1.228	1.999.809	4.934
Sicilia	6	102	4.340	652	792	5.025.552	9.438
Sardegna	3	64	2.540	788	916	1.668.198	4.976
Italia	67	1.456	60.368	-	-	60.688.343	119.073

Stratificazione e selezione dei comuni

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 sono stratificati, nell'ambito di ciascun dominio (indicato nel seguito come d) definito dalla ASL, in base alla loro dimensione demografica e nel rispetto delle seguenti condizioni:

- autoponderazione del campione a livello di ogni dominio d; tale condizione assicura che venga assegnata la stessa probabilità di selezione a ogni unità finale di campionamento appartenente al dominio d;

- scelta del numero, \bar{n} , di comuni campione da estrarre da ciascuno strato NAR: tale parametro è stato posto pari a quattro, in modo tale che ognuno dei quattro comuni campione di ogni strato svolga la rilevazione in uno dei quattro trimestri d'indagine; i comuni AR, invece, svolgono l'indagine in tutti e quattro i trimestri;
- scelta di un numero minimo di famiglie da intervistare in ciascun comune campione; tale minimo è stato posto pari a 30;
- formazione di strati aventi ampiezza approssimativamente costante in termini di popolazione residente.

Il procedimento di stratificazione si articola nelle seguenti fasi:

- 1) ordinamento dei comuni del dominio d in ordine decrescente secondo la loro dimensione demografica in termini di popolazione residente;
- 2) determinazione di una soglia di popolazione ${}_d\lambda$ per la definizione dei comuni AR, mediante la relazione:

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

in cui, per il generico dominio d si indica con: ${}_d\bar{m}$ il numero minimo di famiglie da intervistare in ciascun comune campione; ${}_d\delta$ il numero medio di componenti per famiglia; ${}_d f$ la frazione di campionamento;

- 3) suddivisione dei comuni nei due sottoinsiemi AR e NAR: i comuni di dimensione superiore o uguale a ${}_d\lambda$ vengono definiti AR; quelli di dimensione inferiore vengono definiti NAR;
- 4) suddivisione dei comuni dell'insieme NAR in strati aventi dimensione, in termini di popolazione residente, approssimativamente costante e pari all'incirca a ${}_d\lambda \cdot \bar{n}$, essendo \bar{n} il numero di comuni campione da estrarre da ciascuno strato, posto pari a quattro.

Effettuata la stratificazione, i comuni AR sono inclusi con certezza nel campione; per quanto riguarda, invece, i comuni NAR, nell'ambito di ogni strato vengono estratti \bar{n} comuni campione con probabilità proporzionale alla dimensione demografica, mediante la procedura di selezione sistematica proposta da Madow (1949)².

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 occupano le seguenti posizioni:

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

Procedimento per il calcolo delle stime

Le stime prodotte dall'indagine sono stime di frequenze (assolute e relative) e stime del numero totale e medio di eventi. Alcune stime hanno come riferimento le famiglie, altre gli 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.

² Madow, W.G. (1949) "On the theory of systematic sampling II", Ann. Math. Stat., 20, 333-354.

Questo principio viene realizzato attribuendo a ogni unità campionaria un peso che indica il numero di unità della popolazione rappresentate 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 ; 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 ; H_d , numero totale di strati nel generico dominio territoriale d . Indichiamo poi con Y_{nij} il totale della generica variabile y osservato sulla famiglia j del comune i dello strato h :

$$Y_{nij} = \sum_{p=1}^{P_{hij}} Y_{hijp}$$

Ipotizziamo di voler stimare, con riferimento ad un generico dominio d , il totale della 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

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

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 demografiche 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 dei totali noti riferiti a diverse sottopopolazioni: la popolazione regionale per sesso e otto classi di età⁴; la popolazione regionale per area vasta, sesso e cinque classi di età⁵; la popolazione straniera totale per sesso; la popolazione residente a livello totale per ASL.

Indicando, quindi, con ${}_k X$ il totale noto della k -esima variabile ausiliaria per la generica regione geografica e con ${}_k X_{hij}$ il valore assunto dalla k -esima variabile ausiliaria per la famiglia rispondente hij , la condizione sopra descritta è espressa dalla seguente uguaglianza

$${}_k X = \hat{{}_k X} = \sum_{h=1}^H \sum_{i=1}^{n_h} \sum_{j=1}^{m_{hi}} {}_k X_{hij}$$

in cui H indica il numero complessivo di strati definiti nella regione.

3 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} .

4 Le classi di età considerate a livello regionale sono: 0-5, 6-13, 14-24, 25-34, 35-44, 45-64, 65-74, 75 e più.

5 Le classi di età per le aree vaste sono: 0-13, 14-24, 25-44, 45-64, 65 e più.

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 nel campione;
- 2) si calcolano i fattori correttivi per mancata risposta totale, definiti come inverso del tasso di caduta nel comune a cui ciascuna unità appartiene, oppure, nel caso di caduta totale di comuni auto rappresentativi, redistribuendo il peso del comune a livello regionale;
- 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 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 (opportunosamente 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 lineare; 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 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 vincolata⁶. 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 4, tale stimatore riveste un ruolo centrale in quanto è possibile dimostrare⁷ che tutti gli stimatori di ponderazione vincolata convergono asintoticamente, all'aumentare della numerosità campionaria, allo stimatore di regressione generalizzata.

Valutazione del livello di precisione delle stime

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 (o coefficiente di variazione). Indicando con $\hat{V}ar(\hat{Y}_d)$ 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

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

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 nel paragrafo 3, le stime prodotte dall'indagine sono state ottenute mediante uno stimatore di ponderazione vincolata definito in base ad 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

⁶ Nella letteratura in lingua anglosassone sull'argomento tali stimatori sono noti come *calibration estimators*.

⁷ Deville J.C., Sarndal C.E. (1992) "Calibration Estimators in Survey Sampling", Journal of the American Statistical Association, vol. 87, pp. 376-382.

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, in quanto 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

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

dove Z_{hij} è la variabile linearizzata espressa come $Z_{hij} = Y_{hij} - \mathbf{X}'_{hij} \beta$, essendo $\mathbf{X}_{hij} = (x_{1hij}, \dots, x_{K h ij})'$ il vettore contenente i valori delle 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 può essere 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 effettuare la seguente scomposizione:

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

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

Per l'insieme degli 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

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

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

Per l'insieme degli strati NAR la varianza viene stimata invece mediante la formula seguente

$$\sum_{h=1}^{H_{NAR}} \hat{\text{var}}(\hat{Z}_h) = \sum_{h=1}^{H_{NAR}} \frac{n_h}{n_h - 1} \sum_{i=1}^{n_h} \left(\hat{Z}_{hi} - \frac{\hat{Z}_h}{n_h} \right)^2 \quad (9)$$

dove le quantità sono espresse come

$$\hat{Z}_{hi} = \sum_{j=1}^{m_{hi}} Z_{hij} W_{hij} \quad \text{e} \quad \hat{Z}_h = \sum_{i=1}^{n_h} \sum_{j=1}^{m_{hi}} Z_{hij} W_{hij}.$$

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

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

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

Presentazione sintetica degli errori campionari

Poiché a ciascuna stima ${}_d\hat{Y}$ corrisponde un errore campionario relativo $\hat{\varepsilon}({}_d\hat{Y})$, per consentire un uso corretto delle informazioni prodotte dall'indagine sarebbe necessario pubblicare, per ogni stima, anche il corrispondente errore di campionamento relativo. Tuttavia sia per limiti di tempo e di costi di elaborazione, sia perché le tavole di pubblicazione risulterebbero appesantite e di non facile consultazione per l'utente finale, non è possibile pubblicare tutti gli errori di campionamento delle stime fornite. Inoltre, non sarebbero comunque disponibili gli errori delle stime non pubblicate, che l'utente può ricavare in modo autonomo.

Al fine di permettere comunque una valutazione della variabilità campionaria di tutte le stime d'interesse, si ricorre a una presentazione sintetica degli errori relativi basata su *modelli regressivi*; ossia fondata sulla determinazione di una funzione matematica che mette in relazione ciascuna stima con il proprio errore di campionamento.

L'approccio utilizzato per la costruzione dei modelli è differente a seconda che la variabile oggetto di stima sia qualitativa o quantitativa. Infatti, per le stime di frequenze assolute (o relative) riferite alle modalità di variabili qualitative, è possibile utilizzare modelli che hanno un fondamento teorico, secondo cui gli errori relativi delle stime di frequenze assolute sono funzione decrescente dei valori delle stime stesse; per le stime di totali di variabili quantitative, invece, il problema è piuttosto complesso, dal momento che non è stata ancora elaborata un'adeguata base teorica per l'interpolazione degli errori campionari delle stime in questione. L'approccio adottato per trattare il caso di variabili quantitative è pertanto di tipo empirico ed è fondato sull'evidenza sperimentale che l'errore assoluto di un totale è una funzione crescente del totale stesso. Si tratta pertanto di individuare la relazione matematica che meglio si *adatta*⁸ alla nuvola di punti costituita dalle coppie di valori $({}_d\hat{Y}, \hat{\varepsilon}({}_d\hat{Y}))$, per un numero il più possibile elevato di stime, separatamente per i diversi livelli territoriali di pubblicazione delle stime.

E' bene precisare che i modelli di interpolazione degli errori sono validi, oltre che per le stime assolute di frequenze e di totali, anche per le stime di frequenze relative e di medie di variabili quantitative riferite all'intera popolazione del dominio di riferimento (ripartizione, regione o tipologia comunale), come ad esempio il "numero medio di accertamenti diagnostici per abitante nel Piemonte". Se si vuole calcolare l'errore relativo di una stima riferita a una sottopopolazione differente (ad esempio la popolazione di coloro che presentano una certa modalità di una variabile di interesse) è necessario ricorrere ad un'approssimazione. Infatti, la stima di una frequenza relativa o di una media specifica (o di un qualunque indicatore) riferita ad un sottogruppo di famiglie o persone, è ottenibile come rapporto tra due quantità entrambe stimate:

$$\hat{R}_d = \frac{\hat{N}_d}{\hat{P}_d},$$

in cui \hat{P}_d è la stima del numero di persone che presentano la caratteristica c nel dominio d , \hat{N}_d è la stima del totale della variabile quantitativa n sulle persone con la caratteristica c e \hat{R}_d è l'indicatore definito come rapporto tra \hat{N}_d e \hat{P}_d (per esempio: numero medio di visite effettuate da persone con malattie croniche nel Molise).

⁸ L'adattamento del modello alla nuvola di punti viene valutato in termini di indice di determinazione R^2 .

Una valutazione approssimata⁹ dell'errore di \hat{R}_d , valida sotto l'ipotesi di incorrelazione tra \hat{R}_d e \hat{P}_d , si può ottenere come:

$$\hat{\varepsilon}(\hat{R}_d) = \sqrt{\hat{\varepsilon}^2(\hat{N}_d) - \hat{\varepsilon}^2(\hat{P}_d)}, \quad (11)$$

in cui $\hat{\varepsilon}(\hat{N}_d)$ e $\hat{\varepsilon}(\hat{P}_d)$ si possono calcolare utilizzando il modello (12).

Nei casi in cui non è possibile assumere l'ipotesi di incorrelazione tra \hat{R}_d e \hat{P}_d , è necessario ricorrere alla linearizzazione di \hat{R}_d e calcolare gli errori sulla variabile linearizzata¹⁰ Z_i definita, per ogni unità campionaria i del dominio d , come:

$$Z_i = \frac{1}{\hat{P}_d} (N_i - \hat{R}_d \cdot P_i),$$

essendo N_i il valore della variabile n presentato dall'unità i e P_i una variabile dicotomica che assume il valore 1 se l'unità i presenta la caratteristica c e 0 altrimenti. Questo procedimento è stato utilizzato per il calcolo degli errori campionari degli indici di stato psicofisico, per i quali l'ipotesi di incorrelazione tra \hat{R}_d e \hat{P}_d non è stata ritenuta valida.

Presentazione sintetica degli errori campionari per stime di frequenze

Il modello utilizzato per le stime di frequenze assolute, con riferimento al generico dominio d , è il seguente:

$$\log \hat{\varepsilon}^2({}_d\hat{Y}) = a + b \log({}_d\hat{Y}) \quad (12)$$

in cui i parametri a e b vengono stimati, separatamente per ogni dominio d , utilizzando il metodo dei minimi quadrati.

Il prospetto 2 riporta i valori dei coefficienti a e b e dell'indice di determinazione R^2 delle funzioni utilizzate per l'interpolazione degli errori campionari delle stime di frequenze, separatamente per le famiglie e per le persone, per totale Italia, ripartizione geografica, tipologia comunale e regione.

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

$$\hat{\varepsilon}(\hat{Y}_d) = \sqrt{\exp(a + b \log(\hat{Y}_d))} \quad (13)$$

che si ricava facilmente dalla (12).

Se, per esempio, la stima di frequenza assoluta \hat{Y}_d si riferisce agli individui dell'Italia Nord Occidentale, l'errore relativo corrispondente si ottiene introducendo nella (13) i valori dei parametri a e b riportati nella seconda riga del prospetto 2 alla voce PERSONE ($a = 6,626544$, $b = -1,00398$).

I prospetti 4 e 5 consentono, inoltre, di rendere più agevole la valutazione degli errori campionari. Essi presentano la seguente struttura: a) in fiancata sono elencati i valori crescenti di stima (20.000, 30.000, ..., 1.000.000); b) le colonne successive contengono gli errori di campionamento relativo, per ciascun dominio territoriale di interesse, calcolati mediante l'espressione (13), corrispondenti alle stime della prima colonna.

Le informazioni contenute in tali prospetti permettono di calcolare l'errore relativo di una generica stima (di frequenza assoluta o di un totale) mediante due procedimenti che risultano di facile applicazione, anche se conducono a risultati meno precisi di quelli ottenibili mediante l'espressione (13). Il primo metodo consiste nell'individuare, nella prima colonna del prospetto, il

⁹ Si veda: P.D. Falorsi, S. Falorsi (1996) 'Indagine sulle forze di lavoro: descrizione della strategia di campionamento e valutazione dell'errore campionario dei principali indicatori provinciali del mercato del lavoro', 1996, ISTAT-Documenti).

¹⁰ Secondo il metodo di linearizzazione di Woodruff, è possibile calcolare la varianza di uno stimatore non lineare approssimandolo mediante la formula di Taylor. Si veda 'Manuale di tecniche d'indagine' - vol. 5 - Note e relazioni - ISTAT 1989.

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.

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

$$\hat{\varepsilon}(\hat{Y}_d) = \hat{\varepsilon}(\hat{Y}_d^{k-1}) - \frac{\hat{\varepsilon}(\hat{Y}_d^{k-1}) - \hat{\varepsilon}(\hat{Y}_d^k)}{\hat{Y}_d^k - \hat{Y}_d^{k-1}} (\hat{Y}_d - \hat{Y}_d^{k-1}) \quad (14)$$

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.

Presentazione sintetica degli errori campionari per stime di totali

Il modello utilizzato per le stime di totali di variabili quantitative, con riferimento al generico dominio d, è il seguente:

$$\sigma_d(\hat{Y}) = a + b {}_d\hat{Y} + c {}_d\hat{Y}^2 \quad (15)$$

dove i parametri a, b e c vengono stimati, mediante il metodo dei minimi quadrati, adattando il modello (15) ad una nuvola di punti costituita dal maggior numero di coppie $(\sigma_d(\hat{Y}), {}_d\hat{Y})$.

I prospetti 7 e 8 riportano i valori dei coefficienti a, b, c e dell'indice di determinazione R^2 delle funzioni utilizzate per l'interpolazione degli errori campionari delle stime di totali riferite alle persone, per tutte le aree territoriali considerate.

PROSPETTO 2. VALORI DEI COEFFICIENTI A, B E DELL'INDICE DI DETERMINAZIONE R² (%) DELLE FUNZIONI UTILIZZATE PER L'INTERPOLAZIONE DEGLI ERRORI CAMPIONARI DELLE STIME DI FREQUENZE 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
RIPARTIZIONI GEOGRAFICHE						
Nord-ovest	6,67896	-1,0051	97,5857	8,49213	-1,1383	93,5971
Nord-est	7,72887	-1,0998	99,1377	8,09244	-1,1367	93,6095
Centro	8,16173	-1,1188	98,6372	8,27677	-1,1335	94,5875
Sud	7,71646	-1,0965	98,2305	7,32488	-1,0708	91,7796
Isole	7,07539	-1,0621	98,9064	6,83058	-1,0417	95,0011
REGIONI						
Piemonte	6,88055	-1,0368	99,3201	7,42024	-1,109	94,1224
Valle d'Aosta	5,51009	-1,1462	96,896	5,42477	-1,1495	91,4376
Lombardia	6,91828	-1,0903	99,2114	7,14659	-1,1182	94,722
Liguria	8,22823	-1,097	98,411	8,67932	-1,1438	92,6299
- Bolzano	5,97389	-1,0376	93,0425	6,38715	-1,095	88,2905
- Trento	5,76696	-1,0139	94,6235	6,65009	-1,1123	90,701
Veneto	7,8614	-1,1102	97,8442	7,9812	-1,1307	93,1317
Friuli-Venezia Giulia	7,02412	-1,1088	97,658	7,47961	-1,1678	90,179
Emilia-Romagna	7,35044	-1,0683	98,4052	7,84094	-1,1177	92,9838
Toscana	7,59662	-1,0841	99,1453	7,63446	-1,0934	93,393
Umbria	6,96028	-1,1248	98,528	6,83019	-1,1182	94,7658
Marche	6,48554	-1,0507	98,5805	6,66058	-1,0768	92,475
Lazio	8,30749	-1,1224	97,8106	8,4292	-1,1392	94,4575
Abruzzo	6,92388	-1,0888	97,2656	6,57312	-1,062	90,0849
Molise	6,19995	-1,1257	97,703	5,74158	-1,0916	87,5926
Campania	7,64456	-1,0828	95,5643	7,37796	-1,0674	90,0989
Puglia	7,99215	-1,1311	98,5049	7,38203	-1,0799	92,0187
Basilicata	5,35677	-1,0082	98,3769	5,31888	-1,0081	90,2151
Calabria	6,15734	-0,9855	98,7247	5,83179	-0,9618	90,9188
Sicilia	7,14325	-1,0612	98,7643	6,81981	-1,0327	94,8641
Sardegna	6,71702	-1,071	98,451	6,8437	-1,0865	92,9889
ITALIA	6,69121	-1,014	99,1346	8,29063	-1,1231	94,0552

PROSPETTO 3. VALORI INTERPOLATI DEGLI ERRORI CAMPIONARI RELATIVI PERCENTUALI DELLE STIME DI FREQUENZE RIFERITE ALLE FAMIGLIE. Per totale Italia, ripartizione geografica, tipo di comune e regione.

STIME	Italia	Nord-ovest	Nord-est	Centro	Sud	Isole
20.000	18,8	19,5	20,6	23,2	20,8	17,9
30.000	15,3	15,9	16,5	18,5	16,6	14,4
40.000	13,2	13,7	14,1	15,8	14,2	12,4
50.000	11,8	12,3	12,4	13,9	12,6	11,0
60.000	10,8	11,2	11,2	12,6	11,4	10,0
70.000	9,9	10,4	10,3	11,5	10,5	9,2
80.000	9,3	9,7	9,6	10,7	9,7	8,6
90.000	8,8	9,1	9,0	10,0	9,1	8,0
100.000	8,3	8,7	8,5	9,4	8,6	7,6
200.000	5,8	6,1	5,8	6,4	5,9	5,3
300.000	4,8	5,0	4,6	5,1	4,7	4,2
400.000	4,1	4,3	4,0	4,4	4,0	3,6
500.000	3,7	3,9	3,5	3,8	3,6	3,2
750.000	3,0	3,1	2,8	3,1	2,8	2,6
1.000.000	2,6	2,7	2,4	2,6	2,4	2,2
2.000.000	1,8	1,9	1,6	1,8	1,7	1,5
3.000.000	1,5	1,6	1,3	1,4	1,3	1,2
4.000.000	1,3	1,4	1,1	1,2	1,1	-
5.000.000	1,1	1,2	1,0	1,1	1,0	-
7.500.000	0,9	-	-	-	-	-
10.000.000	0,8	-	-	-	-	-
15.000.000	0,7	-	-	-	-	-

PROSPETTO 3 (SEGUE). VALORI INTERPOLATI DEGLI ERRORI CAMPIONARI RELATIVI PERCENTUALI DELLE STIME DI FREQUENZE RIFERITE ALLE FAMIGLIE. Per totale Italia, ripartizione geografica, tipo di comune e regione.

STIME	Piemonte	Valle d'Aosta-Vallée d'Aoste	Liguria	Lombardia	Bolzano-Bozen	Trento	Veneto	Friuli-Venezia Giulia	Emilia Romagna	Toscana	Umbria
20.000	18,4	5,4	14,4	26,8	11,6	11,8	20,9	13,8	19,9	20,8	12,4
30.000	14,9	4,3	11,5	21,4	9,4	9,6	16,7	11,0	16,0	16,7	9,8
40.000	12,8	3,6	9,9	18,3	8,1	8,3	14,2	9,4	13,7	14,3	8,4
50.000	11,4	3,2	8,7	16,2	7,2	7,4	12,6	8,3	12,2	12,7	7,4
60.000	10,4	2,9	7,9	14,7	6,6	6,8	11,3	7,5	11,1	11,5	6,7
70.000	9,6	-	7,3	13,5	6,1	6,3	10,4	6,9	10,2	10,6	6,1
80.000	9,0	-	6,8	12,5	5,7	5,8	9,7	6,4	9,5	9,8	5,7
90.000	8,4	-	6,3	11,7	5,3	5,5	9,1	6,0	8,9	9,2	5,3
100.000	8,0	-	6,0	11,1	5,0	5,2	8,5	5,7	8,4	8,7	5,0
200.000	5,6	-	4,1	7,6	3,5	3,7	5,8	3,9	5,8	6,0	3,4
300.000	4,5	-	3,3	6,1	-	-	4,6	3,1	4,7	4,8	2,7
400.000	3,9	-	2,8	5,2	-	-	4,0	2,6	4,0	4,1	-
500.000	3,5	-	2,5	4,6	-	-	3,5	2,3	3,6	3,6	-
750.000	2,8	-	2,0	3,7	-	-	2,8	-	2,9	2,9	-
1.000.000	2,4	-	-	3,1	-	-	2,4	-	2,5	2,5	-
1.500.000	2,0	-	-	2,5	-	-	1,9	-	2,0	2,0	-
2.000.000	1,7	-	-	2,1	-	-	1,6	-	1,7	-	-

PROSPETTO 3 (SEGUE). VALORI INTERPOLATI DEGLI ERRORI CAMPIONARI RELATIVI PERCENTUALI DELLE STIME DI FREQUENZE RIFERITE ALLE FAMIGLIE. Per totale Italia, ripartizione geografica, tipo di comune e regione.

STIME	Marche	Lazio	Abruzzo	Molise	Campania	Puglia	Basilicata	Calabria	Sicilia	Sardegna
20.000	14,1	24,6	14,5	8,4	21,4	20,1	9,9	16,5	18,6	14,3
30.000	11,4	19,6	11,6	6,7	17,2	16,0	8,1	13,5	15,0	11,5
40.000	9,8	16,6	10,0	5,7	14,7	13,6	7,0	11,7	12,9	9,9
50.000	8,7	14,7	8,8	5,0	13,1	12,0	6,2	10,5	11,4	8,8
60.000	7,9	13,3	8,0	4,5	11,8	10,8	5,7	9,6	10,4	7,9
70.000	7,3	12,2	7,3	4,2	10,9	9,9	5,3	8,9	9,6	7,3
80.000	6,8	11,3	6,8	3,9	10,1	9,2	4,9	8,3	8,9	6,8
90.000	6,4	10,6	6,4	3,6	9,5	8,6	4,6	7,9	8,4	6,4
100.000	6,0	10,0	6,0	3,4	9,0	8,1	4,4	7,5	7,9	6,0
200.000	4,2	6,7	4,1	-	6,2	5,5	3,1	5,3	5,5	4,2
300.000	3,4	5,4	3,3	-	4,9	4,3	-	4,3	4,4	3,4
400.000	2,9	4,6	2,8	-	4,2	3,7	-	3,8	3,8	2,9
500.000	2,6	4,0	2,5	-	3,8	3,3	-	3,4	3,4	2,6
750.000	-	3,2	-	-	3,0	2,6	-	2,8	2,7	-
1.000.000	-	2,7	-	-	2,6	2,2	-	-	2,3	-
1.500.000	-	2,2	-	-	2,1	1,7	-	-	1,9	-
2.000.000	-	1,9	-	-	1,8	-	-	-	1,6	-

PROSPETTO 4. VALORI INTERPOLATI DEGLI ERRORI CAMPIONARI RELATIVI PERCENTUALI DELLE STIME DI FREQUENZE RIFERITE ALLE PERSONE. Per totale Italia, ripartizione geografica, tipo di comune e regione.

STIME	Italia	Nord-ovest	Nord-est	Centro	Sud	Isole
20.000	24,3	24,9	20,5	22,9	19,4	17,5
30.000	19,3	19,8	16,3	18,2	15,6	14,2
40.000	16,4	16,8	13,9	15,5	13,4	12,2
50.000	14,5	14,8	12,2	13,6	11,9	10,9
60.000	13,1	13,3	11,0	12,3	10,8	9,9
70.000	12,0	12,2	10,1	11,3	9,9	9,1
80.000	11,1	11,3	9,3	10,4	9,2	8,5
90.000	10,4	10,6	8,7	9,8	8,7	8,0
100.000	9,8	10,0	8,2	9,2	8,2	7,6
200.000	6,7	6,7	5,6	6,2	5,7	5,3
300.000	5,3	5,3	4,4	4,9	4,6	4,3
400.000	4,5	4,5	3,7	4,2	3,9	3,7
500.000	4,0	4,0	3,3	3,7	3,5	3,3
750.000	3,2	3,2	2,6	2,9	2,8	2,6
1.000.000	2,7	2,7	2,2	2,5	2,4	2,3
2.000.000	1,8	1,8	1,5	1,7	1,6	1,6
3.000.000	1,5	1,4	1,2	1,3	1,3	1,3
4.000.000	1,2	1,2	1,0	1,1	1,1	1,1
5.000.000	1,1	1,1	0,9	1,0	1,0	1,0
7.500.000	0,9	0,9	0,7	0,8	0,8	-
10.000.000	0,7	0,7	0,6	0,7	0,7	-
15.000.000	0,6	0,6	-	-	0,6	-
20.000.000	0,5	-	-	-	-	-
25.000.000	0,4	-	-	-	-	-

PROSPETTO 4 (SEGUE). VALORI INTERPOLATI DEGLI ERRORI CAMPIONARI RELATIVI PERCENTUALI DELLE STIME DI FREQUENZE RIFERITE ALLE PERSONE. Per totale Italia, ripartizione geografica, tipo di comune e regione.

STIME	Piemonte	Valle d'Aosta-Vallée d'Aoste	Liguria	Lombardia	Bolzano-Bozen	Trento	Veneto	Friuli-Venezia Giulia	Emilia Romagna	Toscana	Umbria
20.000	15,7	4,3	11,9	22,4	9,1	9,5	16,9	10,9	16,8	17,2	10,1
30.000	14,8	4,0	11,2	21,1	8,6	9,0	15,9	10,2	15,9	16,2	9,5
40.000	12,7	3,4	9,5	17,9	7,4	7,7	13,5	8,7	13,5	13,9	8,1
50.000	11,2	3,0	8,4	15,8	6,5	6,8	11,9	7,6	11,9	12,3	7,2
60.000	10,2	2,7	7,6	14,2	5,9	6,1	10,8	6,8	10,8	11,1	6,5
70.000	9,3	2,5	7,0	13,0	5,4	5,6	9,9	6,2	9,9	10,2	5,9
80.000	8,7	2,3	6,5	12,0	5,0	5,2	9,1	5,8	9,2	9,5	5,5
90.000	8,2	2,1	6,1	11,3	4,7	4,9	8,6	5,4	8,6	8,9	5,2
100.000	7,7	2,0	5,7	10,6	4,5	4,6	8,1	5,1	8,1	8,4	4,9
200.000	5,3	-	3,9	7,1	3,1	3,1	5,4	3,4	5,5	5,7	3,3
300.000	4,2	-	3,1	5,7	2,4	2,5	4,3	2,7	4,4	4,6	2,6
400.000	3,6	-	2,6	4,8	2,1	2,1	3,7	2,3	3,7	3,9	2,2
500.000	3,2	-	2,3	4,2	1,8	1,9	3,2	2,0	3,3	3,5	2,0
750.000	2,6	-	1,8	3,3	-	-	2,6	1,6	2,6	2,8	1,6
1.000.000	2,2	-	1,6	2,8	-	-	2,2	1,3	2,2	2,4	-
2.000.000	1,5	-	-	1,9	-	-	1,5	-	1,5	1,6	-
3.000.000	1,2	-	-	1,5	-	-	1,2	-	1,2	1,3	-
4.000.000	1,0	-	-	1,3	-	-	1,0	-	1,0	1,1	-
5.000.000	-	-	-	1,1	-	-	0,9	-	-	-	-

PROSPETTO 4 (SEGUE). VALORI INTERPOLATI DEGLI ERRORI CAMPIONARI RELATIVI PERCENTUALI DELLE STIME DI FREQUENZE 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	11,5	20,2	11,9	6,7	17,3	16,2	8,3	13,7	15,6	12,0
30.000	10,9	19,1	11,2	6,4	16,3	15,3	7,9	13,0	14,8	11,3
40.000	9,3	16,2	9,6	5,4	14,0	13,1	6,8	11,3	12,7	9,7
50.000	8,3	14,3	8,6	4,8	12,4	11,6	6,1	10,2	11,3	8,6
60.000	7,5	12,8	7,8	4,4	11,3	10,5	5,6	9,3	10,3	7,8
70.000	6,9	11,8	7,2	4,0	10,4	9,7	5,2	8,6	9,5	7,1
80.000	6,4	10,9	6,7	3,7	9,7	9,0	4,8	8,1	8,9	6,6
90.000	6,0	10,2	6,3	3,5	9,1	8,5	4,5	7,7	8,4	6,2
100.000	5,7	9,6	5,9	3,3	8,6	8,0	4,3	7,3	7,9	5,9
200.000	3,9	6,5	4,1	2,3	5,9	5,5	3,0	5,2	5,5	4,0
300.000	3,1	5,1	3,3	1,8	4,8	4,4	2,5	4,3	4,5	3,2
400.000	2,7	4,4	2,8	-	4,1	3,8	2,1	3,7	3,9	2,8
500.000	2,4	3,8	2,5	-	3,6	3,4	1,9	3,4	3,5	2,5
750.000	1,9	3,0	2,0	-	2,9	2,7	-	2,8	2,8	2,0
1.000.000	1,6	2,6	1,7	-	2,5	2,3	-	2,4	2,4	1,7
2.000.000	-	1,7	-	-	1,7	1,6	-	1,7	1,7	-
3.000.000	-	1,4	-	-	1,4	1,3	-	-	1,4	-
4.000.000	-	1,2	-	-	1,2	1,1	-	-	1,2	-
5.000.000	-	1,0	-	-	1,1	-	-	-	1,1	-