

Nota metodologica

Premessa

L'indagine sulle *Condizioni di salute e ricorso ai servizi sanitari* ha l'obiettivo di analizzare i comportamenti relativi alla salute e all'utilizzo dei servizi sanitari secondo le caratteristiche demografiche e socio-economiche dei cittadini. Attraverso tali informazioni raccolte direttamente presso gli individui è possibile costruire indicatori su condizioni di salute e qualità della vita, presenza di disabilità, stili di vita prevenzione, ricorso ai servizi sanitari, uso dei farmaci. Ciò rende possibile individuare segmenti di popolazione a rischio, studiare le disuguaglianze nella salute e nell'accesso ai servizi; conoscere i profili degli utilizzatori dei servizi sanitari e le modalità di fruizione. Integrando con tali informazioni le fonti di dati amministrativi è possibile arricchire la base informativa necessaria per la pianificazione socio-sanitaria e la promozione della salute pubblica, sia a livello nazionale che a livello locale.

Le indagini sui temi della salute realizzate tramite intervista (HIS - *Health Interview Survey*) sono peraltro condotte in tutti i paesi dell'Unione Europea, che ne ha promosso l'armonizzazione per consentire la comparabilità delle informazioni utili alla programmazione di politiche sanitarie comunitarie.

L'indagine si inserisce nel Sistema delle Indagini Multiscopo sulle famiglie avviato nel 1993 ed è stata ripetuta con cadenza pressoché quinquennale.

L'indagine è di tipo campionario (Cfr. Appendice: 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 da Ministero della Salute e Regioni.

Principali definizioni

Quozienti standardizzati - La maggior parte dei fenomeni rilevati nell'indagine sono influenzati dalla struttura per età della popolazione. Se una popolazione ha un'alta proporzione di anziani il tasso grezzo, ad esempio della maggior parte delle malattie croniche, risulterà più elevato che in una popolazione della stessa numerosità ma con un numero maggiore di giovani. Ciò è rilevante per la pianificazione di interventi di sanità pubblica, ma non fornisce un'informazione adeguata per analisi epidemiologiche. Per rendere confrontabili popolazioni con diversa composizione per età è necessario calcolare dei tassi che rimuovano gli effetti della struttura demografica sui fenomeni oggetto di studio. Il metodo maggiormente utilizzato è quello della standardizzazione diretta dei tassi usando una popolazione arbitraria di riferimento definita "standard". La popolazione standard utilizzata è quella del Censimento 2011 per classi di età quinquennali, in modo da poter confrontare sia i dati del 2005 con quelli del 2013, che per la comparazione sul territorio (regioni e ripartizione) e gli aspetti socio-economici.

Salute percepita - è rilevata sulla base del quesito: "Come va in generale, la sua salute?". Le modalità predefinite di risposta sono: molto bene – bene – discretamente – male – molto male. N.B. La scala delle modalità è invertita rispetto alla precedente rilevazione 1999-2000, per armonizzarla a livello internazionale. E' stato inoltre somministrato il questionario SF12, derivante da una versione più estesa dell'SF36, mediante il quale sono stati costruiti un indice di stato di salute fisico (PCS) ed un indice di stato di salute psicologico (MCS). Inoltre con l'inserimento di ulteriori quesiti del questionario SF36 è stato possibile indagare la dimensione della salute mentale mediante l'indice Mental Health (MHI). Per l'interpretazione degli indici, si precisa che a

valori più elevati del punteggio medio corrisponde un migliore stato di salute. Per saperne di più vedi <http://crc.marionegri.it/qdv/index.php?page=sf36>.

Malattie croniche - sono rilevate attraverso una batteria di quesiti riferiti a 22 items in cui sono riportate le principali patologie croniche o di lunga durata; ad essi è stato aggiunto un quesito aperto per rilevare le "altre malattie croniche". Le 22 malattie cronico – degenerative sono: Asma, asma di tipo allergico, allergia (escluso asma allergico), diabete, ipertensione, Disturbi del comportamento alimentare (bulimia, anoressia), celiachia, infarto del miocardio, angina pectoris, altre malattie del cuore, ictus-emorragia cerebrale, bronchite cronica – enfisema, artrosi - artrite, osteoporosi, tumore maligno (incluso linfoma e leucemia), cefalea o emicrania ricorrente, ansietà cronica, depressione, Alzheimer – demenze senili, parkinsonismo, cirrosi epatica, malattie della tiroide, insufficienza renale cronica.

Malattie croniche gravi – l'indicatore è stato definito aggregando nella categoria 'gravi' nove gruppi di patologie selezionate in base all'ipotesi che la presenza di queste patologie implichi peggiori condizioni di salute e un elevato livello di limitazioni. Le malattie croniche gravi individuate sono: diabete; infarto del miocardio; angina pectoris; altre malattie del cuore; ictus, emorragia cerebrale; bronchite cronica, enfisema; cirrosi epatica; tumore maligno (inclusi linfoma/leucemia); parkinsonismo; Alzheimer, demenze senili. Queste patologie sono state considerate per confrontare l'indicatore rispetto al 2005. Nella categoria delle gravi è stata inclusa anche l'insufficienza renale cronica, rilevata solo per l'edizione del 2012-2013, quando non si effettuano confronti temporali.

Invalidità permanenti – Sono rilevate le invalidità permanenti di tipo motorio, sensoriale (cecità, sordomutismo e sordità), le invalidità da insufficienza mentale e da malattia mentale o disturbi del comportamento.

Abitudine al fumo - E' definito fumatore una persona che dichiara di fumare attualmente, sia tutti i giorni che occasionalmente. Il fumatore abituale è colui che dichiara di fumare regolarmente tutti i giorni. Ex fumatore è una persona che ha dichiarato di aver fumato in passato, sia tutti i giorni che occasionalmente.

Indice di massa corporea - L'IMC (Body Mass Index - BMI) è un indice pondo – statura le dato dal rapporto tra il peso corporeo di un individuo, espresso in chilogrammi, ed il quadrato della sua statura, espressa in metri. Secondo i criteri stabiliti dall'Organizzazione mondiale della sanità il valore soglia dell'IMC per stabilire se un individuo possa considerarsi obeso è pari a 30, sono sottopeso le persone con valori di IMC inferiori a 18,5; sono normopeso le persone con valori di IMC compresi nell'intervallo 18,5-24,9; sono in sovrappeso le persone con valori di IMC compresi nell'intervallo da 25-29,9.

Attività fisico-sportiva si intende l'esercizio di attività sportiva o motoria, sia con rilevante sforzo fisico, che di tipo moderato o leggero, praticata almeno una volta a settimana nel tempo libero. Tra le attività fisico – sportive vengono incluse: sport agonistici e non, palestra, jogging, passeggiate in bicicletta, giardinaggio, passeggiate a piedi per almeno un chilometro. L'OMS nel 2010 ha pubblicato le raccomandazioni sui livelli di attività fisica protettivi per la salute (*Health Enhancing Physical Activity – HEPA, Fonte: WHO (2010) Global recommendations on Physical Activity for Health*) specifici per bambini, ragazzi, adulti e anziani, sulla base di evidenze scientifiche che hanno dimostrato gli effetti benefici sullo stato di salute prodotti da una attività fisica di moderata intensità, svolta con regolarità:

- per i bambini e ragazzi dai 5 ai 17 anni 60 minuti al giorno di attività fisico-sportiva da moderata a intensiva (da svolgere nei diversi ambiti della vita quotidiana)
- per gli adulti 18-64 anni sono 150 minuti di attività fisico-sportiva moderata a settimana, oppure 75 minuti di attività fisico-sportiva intensiva a settimana, oppure una combinazione equivalente di attività fisico-sportiva intensiva e moderata
- per le persone anziane le stesse quantità raccomandate per gli adulti (con particolare attenzione alla pratica di attività fisico-sportive più adatte per questa fascia d'età)

Uso regolare di farmaci prescritti – Necessità di un uso regolare di farmaci prescritti da un medico durante tutto l'anno, sia giornaliero, che con minor frequenza (escluso i contraccettivi).

Terapie non convenzionali - Trattamenti di cura e rimedi alternativi e/o complementari rispetto alla medicina tradizionale. Sono in particolare rilevati il ricorso a: agopuntura, omeopatia, fitoterapia e trattamenti manuali (esclusi i massaggi estetici), nonché il ricorso ad altri tipi di terapie non convenzionali mediante un quesito aperto.

Visite mediche - Le visite mediche sono rilevate con riferimento a quelle effettuate nelle quattro settimane precedenti l'intervista. Non sono comprese quelle effettuate presso la Guardia Medica, il Pronto Soccorso, durante un ricovero o in regime di Day Hospital, sono invece incluse le visite effettuate per svolgere attività sportiva e/o lavorativa. Le visite mediche generiche sono quelle effettuate dal medico di famiglia. Tra le visite specialistiche sono rilevate le visite: geriatriche, cardiologiche, ostetrico-ginecologiche, odontoiatriche, otorino-laringoiatra, ortopediche, neurologiche, psichiatriche-psicologiche, urologiche, gastroenterologiche, dietologiche, dermatologiche, altro tipo di visita specialistica.

Accertamenti diagnostici – Sono esami di tipo clinico per la valutazione dello stato di salute. Sono esclusi quelli effettuati durante un ricovero ospedaliero o in Day Hospital. Per accertamenti di laboratorio si intendono analisi del sangue e delle urine. Negli accertamenti specialistici sono inclusi gli altri tipi di accertamenti, quali: radiografie, ecografie, risonanza magnetica, TAC, elettrocardiogramma, pap-test, ecc.

Ricoveri ospedalieri – Sono rilevati i ricoveri con almeno un pernottamento in ospedale, istituto di cura convenzionato o in casa di cura privata. Si precisa che non vanno considerati i lungo-degenti quando la durata della degenza supera i due anni.

Servizi di riabilitazione – Si considera il ricorso a trattamenti di riabilitazione quali: fisioterapia, riabilitazione al linguaggio, riabilitazione della vista, massoterapia, laserterapia, ionoforesi, ecc.).

Come è rilevata la presenza di limitazioni funzionali

Nell'indagine sulla salute condotta nel biennio 2012-2013, per rilevare la presenza di limitazioni funzionali, è stata inserita la stessa batteria di quesiti¹, utilizzata nelle precedenti indagini Istat sulla salute, mediante la quale è stata stimata, fino all'edizione del 2004-2005, la presenza di disabilità.

Il cambiamento di denominazione dell'aggregato individuato mediante questo strumento è dovuto alla necessità di accogliere la nuova definizione di disabilità proposta con la classificazione ICF (*International Classification of Functioning, Disability and Health*) dell'Organizzazione Mondiale della Sanità (<http://www.who.int/classifications/icf/en/>). Con questa nuova classificazione la disabilità non è più concepita come riduzione delle capacità funzionali determinata da una malattia o menomazione, ma come la risultante di una interazione tra condizioni di salute e fattori contestuali (personali e ambientali).

Con l'ICF si pone l'accento sulle "limitazioni delle attività" ad indicare le difficoltà che un individuo può incontrare nell'eseguire delle attività ed il termine "restrizioni della partecipazione" per definire i problemi che un individuo può (ma non necessariamente deve) incontrare nelle diverse "aree di vita". Una persona con limitazioni funzionali non è ineluttabilmente costretta a sperimentare restrizioni alla partecipazione sociale. Con l'ICF si intende valutare, non la riduzione di capacità in sé, ma quanto l'individuo è in grado di "fare"; gli ostacoli da rimuovere o gli interventi da effettuare perché l'individuo possa raggiungere il massimo della propria auto-realizzazione.

¹ La batteria di quesiti è stata predisposta da un gruppo di lavoro dell'OECD sulla base della classificazione ICIDH (*International Classification of Impairments, Disabilities and Handicaps*)¹ dell'Organizzazione Mondiale della Sanità. Nella batteria di quesiti è inclusa la scala per la misurazione del livello di difficoltà nelle attività quotidiane (ADL - *Activities of Daily Living*) proposta inizialmente da Katz negli anni '60.

La disabilità, quindi, non è più considerata una condizione della persona, ma il risultato negativo dell'interazione tra singolo individuo e ambiente.

Si parla quindi di *persone con limitazioni funzionali* e non più di *persone con disabilità* per riferirsi alla popolazione che presenta le difficoltà in alcune specifiche dimensioni: la dimensione fisica, riferibile alle funzioni del movimento e della locomozione; la sfera di autonomia nelle funzioni quotidiane che si riferisce alle attività di cura della persona; la dimensione della comunicazione che riguarda le funzioni della vista, dell'udito e della parola.

Ad ogni dimensione corrisponde una batteria di quesiti con modalità di risposta che identificano diversi gradi di difficoltà (da una totale autonomia alla presenza di qualche difficoltà, ad un livello di difficoltà maggiore fino ad arrivare all'inabilità di adempiere la funzione senza l'aiuto di altre persone).

Per una corretta interpretazione dei dati è bene sottolineare che, per ciascuna domanda, l'intervistato risponde secondo la valutazione soggettiva del proprio livello di autonomia. Tale valutazione può variare non solo in funzione delle diverse condizioni di salute e di livello di autonomia, ma anche per differenze culturali e cognitive o per disparità nel disporre di sostegni materiali e relazionali di cui il disabile può avvalersi per far fronte alle proprie limitazioni.

Si definisce *persona con limitazioni funzionali* quella che, escludendo le condizioni riferite a limitazioni temporanee, dichiara il massimo grado di difficoltà in almeno una delle funzioni rilevate con ciascuna domanda, pur tenendo conto dell'eventuale ausilio di apparecchi sanitari (protesi, bastoni, occhiali, ecc.).

A seconda della sfera di autonomia funzionale compromessa, sono state costruite quattro tipologie di limitazioni funzionali: *confinamento*, *difficoltà nel movimento*, *difficoltà nelle funzioni della vita quotidiana*, *difficoltà della comunicazione*.

- Per *confinamento* si intende la costrizione permanente a letto, su una sedia, o nella propria abitazione per motivi fisici o psichici; coloro che risultano confinati rispondono solo ai quesiti sull'attività motoria compatibili con la relativa tipologia di confinamento.
- Le persone con *difficoltà nel movimento* hanno problemi nel camminare (riescono solo a fare qualche passo senza aver bisogno di fare soste o non sono in grado di camminare), non sono in grado di salire e scendere da soli una rampa di scale senza fermarsi, non riescono a chinarsi per raccogliere oggetti da terra.
- Le *difficoltà nelle funzioni della vita quotidiana* riguardano la completa assenza di autonomia nello svolgimento delle essenziali attività quotidiane o di cura della persona, quali mettersi a letto o sedersi da soli, vestirsi da soli, lavarsi o farsi il bagno o la doccia da soli, mangiare da soli anche tagliando il cibo.
- Nelle *difficoltà della comunicazione* sono infine comprese le limitazioni nel *sentire* (non riuscire a seguire una trasmissione televisiva anche alzando il volume e nonostante l'uso di apparecchi acustici); limitazioni nel *vedere* (non riconoscere un amico ad un metro di distanza); difficoltà nella *parola* (non essere in grado di parlare senza difficoltà).

Non sono inclusi nella popolazione osservata le persone residenti permanentemente in istituzioni in quanto la stima derivante dall'indagine si riferisce alle persone che vivono in famiglia. Tra questi ultimi sono senz'altro presenti quote non trascurabili di persone presumibilmente con elevati livelli di gravità di limitazioni nelle dimensioni studiate (basti pensare alle Residenze Sanitarie Assistenziali per anziani non autosufficienti). Tuttavia, poiché in Italia le persone con limitazioni funzionali vivono molto frequentemente in famiglia, l'effetto in termini di sottostima della consistenza del fenomeno risulta abbastanza contenuto². Inoltre non sono inclusi i bambini di età inferiore ai 6 anni in quanto la batteria di quesiti non è idonea per rilevare la presenza di limitazioni funzionali nei bambini. Lo strumento inoltre consente solo in modo parziale di cogliere le limitazioni funzionali connesse a patologie psichiatriche e ad insufficienze mentali.

² Si stima che il numero di istituzionalizzati in Italia sia pari a circa 370.000 persone. Fonte: ISTAT – CISIS, *Rilevazione sui presidi residenziali socio-assistenziali, anno 2011*

Appendice

Strategia di campionamento e livello di precisione dei risultati

1. 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.

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.

2. Disegno di campionamento

2.1. 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.

2.2. 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: minimo	Famiglie campione per area vasta: massimo	Individui popolazione	Individui intervistati
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

2.3 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 pari 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)³.

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:

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

3. 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.

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_{hij} il totale della generica variabile y osservato sulla famiglia j del comune i dello strato h :

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

³ Madow, W.G. (1949) "On the theory of systematic sampling II", Ann. Math. Stat., 20, 333-354

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

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

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 unità⁴ 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 età⁵; la popolazione regionale per area vasta, sesso e cinque classi di età⁶; 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.

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.

⁴ 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} .

⁵ 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.

⁶ Le classi di età per le aree vaste sono: 0-13, 14-24, 25-44, 45-64, 65 e più.

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 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 vincolata⁷. 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 dimostrare⁸ che tutti gli stimatori di ponderazione vincolata convergono asintoticamente, all'aumentare della numerosità campionaria, allo stimatore di regressione generalizzata.

4. Valutazione del livello di precisione delle stime

4.1. 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 (o coefficiente di variazione). 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 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 $\hat{\text{Var}}(\hat{Y}_d)$ 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, 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

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

⁷ Nella letteratura in lingua anglosassone sull'argomento tali stimatori sono noti come *calibration estimators*.

⁸ Deville J.C., Sarndal C.E. (1992) "Calibration Estimators in Survey Sampling", Journal of the American Statistical Association, vol. 87, pp. 376-382.

dove Z_{hij} è la variabile linearizzata espressa come $Z_{hij} = Y_{hij} - \mathbf{X}'_{hij}\beta$, essendo $\mathbf{X}_{hij} = (X_{hij,1}, \dots, X_{hij,k}, \dots, X_{hij,K})'$ 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 $n_h = n_{hi} = 1$, l'indice i di comune diviene superfluo e viene omissso) 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

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

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

4.2. Presentazione sintetica degli errori campionari

Poiché a ciascuna stima \hat{Y}_d corrisponde un errore campionario relativo $\hat{\varepsilon}(\hat{Y}_d)$, 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*⁹ 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).

Una valutazione approssimata¹⁰ dell'errore di \hat{R}_d , valida sotto l'ipotesi di incorrelazione tra \hat{N}_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{N}_d e \hat{P}_d , è necessario ricorrere alla linearizzazione di \hat{R}_d e calcolare gli errori sulla variabile linearizzata¹¹ Z_i definita, per ogni unità campionaria i del dominio d , come:

⁹ L'adattamento del modello alla nuvola di punti viene valutato in termini di indice di determinazione R^2 .

¹⁰ 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)

¹¹ 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

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

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 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:

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

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 $(\sigma_d(\hat{Y}), \hat{Y})$.

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 R^2 (%) 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
ITALIA	6,72447	-1,0148	99,1	8,31901	-1,1239	93,9
RIPARTIZIONI GEOGRAFICHE						
Nord-ovest	6,69057	-1,0039	97,4	8,49283	-1,1366	93,3
Nord-est	7,71706	-1,0962	99,1	8,13474	-1,1381	93,4
Centro	8,11316	-1,1135	98,3	8,27562	-1,1322	94,4
Sud	7,7246	-1,0965	98,2	7,33388	-1,0711	91,8
Isole	7,15322	-1,0696	99,0	6,90353	-1,0482	95,2
REGIONI						
Piemonte	6,94938	-1,0406	99,2	7,46963	-1,0925	94,0
Valle d'Aosta	5,40937	-1,1307	96,8	5,3654	-1,1415	90,9
Lombardia	8,18132	-1,0908	98,0	8,66378	-1,1408	92,3
- Bolzano	5,69305	-1,003	91,5	6,32844	-1,0862	86,8
- Trento	5,76922	-1,0124	94,5	6,67769	-1,1135	90,6
Veneto	7,88814	-1,11	97,7	8,05351	-1,1349	92,8
Friuli-Venezia Giulia	7,06773	-1,1119	97,8	7,49975	-1,169	90,1
Liguria	6,89404	-1,0846	99,1	7,183	-1,1191	94,6
Emilia-Romagna	7,28908	-1,0599	98,5	7,84535	-1,116	92,8
Toscana	7,50956	-1,076	98,7	7,63078	-1,0927	93,0
Umbria	7,01736	-1,1252	98,3	6,88337	-1,1202	94,7
Marche	6,46869	-1,0469	98,2	6,63503	-1,0727	92,4
Lazio	8,30606	-1,1201	97,8	8,43165	-1,1376	94,3
Abruzzo	7,07001	-1,0998	97,3	6,65037	-1,0676	90,1
Molise	6,17642	-1,1194	97,6	5,75895	-1,0911	87,3
Campania	7,62996	-1,0811	95,3	7,38651	-1,0678	89,9
Puglia	8,03394	-1,1348	98,6	7,38668	-1,0805	92,1
Basilicata	5,28015	-0,9989	98,3	5,2639	-1,0017	90,1
Calabria	6,17984	-0,9853	98,6	5,84248	-0,9608	90,7
Sicilia	7,20898	-1,0684	98,9	6,88294	-1,0389	95,1
Sardegna	6,74985	-1,0725	98,4	6,86458	-1,087	92,5

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	14,2	24,8	14,8	8,6	21,5	20,1	10,0	16,7	18,5	14,4
30.000	11,5	19,8	11,8	6,8	17,2	16,0	8,1	13,7	14,9	11,6
40.000	9,9	16,8	10,1	5,8	14,8	13,6	7,0	11,9	12,8	10,0
50.000	8,8	14,9	8,9	5,1	13,1	12,0	6,3	10,6	11,4	8,8
60.000	8,0	13,4	8,1	4,6	11,9	10,8	5,8	9,7	10,3	8,0
70.000	7,4	12,3	7,4	4,3	10,9	9,9	5,3	9,0	9,5	7,4
80.000	6,9	11,4	6,9	-	10,1	9,2	5,0	8,4	8,8	6,9
90.000	6,5	10,7	6,5	-	9,5	8,6	4,7	8,0	8,3	6,4
100.000	6,1	10,1	6,1	-	9,0	8,1	4,5	7,6	7,8	6,1
200.000	4,3	6,8	4,2	-	6,2	5,5	-	5,4	5,4	4,2
300.000	3,4	5,4	3,3	-	5,0	4,3	-	4,4	4,4	3,4
400.000	3,0	4,6	2,8	-	4,3	3,7	-	3,8	3,7	2,9
500.000	2,6	4,1	-	-	3,8	3,2	-	3,4	3,3	-
750.000	-	3,3	-	-	3,0	2,6	-	-	2,7	-
1.000.000	-	2,8	-	-	2,6	2,2	-	-	2,3	-
2.000.000	-	1,9	-	-	1,8	-	-	-	-	-

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	24,5	25,1	20,8	23,0	19,5	17,6
30.000	19,5	19,9	16,5	18,3	15,7	14,2
40.000	16,6	16,9	14,0	15,6	13,4	12,2
50.000	14,7	14,9	12,4	13,7	11,9	10,9
60.000	13,2	13,5	11,2	12,4	10,8	9,9
70.000	12,1	12,3	10,2	11,3	9,9	9,1
80.000	11,3	11,4	9,5	10,5	9,3	8,5
90.000	10,5	10,7	8,9	9,8	8,7	8,0
100.000	9,9	10,1	8,3	9,3	8,2	7,6
200.000	6,7	6,8	5,6	6,3	5,7	5,3
300.000	5,4	5,4	4,5	5,0	4,6	4,3
400.000	4,6	4,6	3,8	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	3,0	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	1,7	1,6
3.000.000	1,5	1,5	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	-	-	0,7	-
15.000.000	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	Lombardia	Bolzano	Trento	Veneto	Friuli-Venezia Giulia	Liguria	Emilia Romagna	Toscana	Umbria
20.000	15,9	4,3	22,6	9,3	9,6	17,1	10,9	12,0	17,0	17,2	10,3
30.000	15,0	4,1	21,3	8,8	9,1	16,2	10,3	11,3	16,0	16,3	9,7
40.000	12,8	3,5	18,0	7,5	7,7	13,7	8,7	9,7	13,7	13,9	8,3
50.000	11,4	3,0	15,9	6,6	6,8	12,1	7,6	8,5	12,1	12,3	7,3
60.000	10,3	2,7	14,3	6,0	6,2	10,9	6,9	7,7	10,9	11,1	6,6
70.000	9,4	2,5	13,1	5,5	5,7	10,0	6,3	7,1	10,0	10,2	6,0
80.000	8,8	2,3	12,2	5,1	5,3	9,3	5,8	6,6	9,3	9,5	5,6
90.000	8,2	2,2	11,4	4,8	4,9	8,7	5,4	6,1	8,7	8,9	5,2
100.000	7,8	2,0	10,7	4,6	4,6	8,2	5,1	5,8	8,2	8,4	4,9
200.000	5,3	-	7,2	3,1	3,2	5,5	3,4	3,9	5,6	5,8	3,4
300.000	4,3	-	5,7	2,5	2,5	4,4	2,7	3,1	4,4	4,6	2,7
400.000	3,6	-	4,9	2,1	2,1	3,7	2,3	2,7	3,8	3,9	2,3
500.000	3,2	-	4,3	-	-	3,3	2,0	2,3	3,3	3,5	2,0
750.000	2,6	-	3,4	-	-	2,6	1,6	1,9	2,7	2,8	1,6
1.000.000	2,2	-	2,9	-	-	2,2	1,3	1,6	2,3	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	-	-	-	-	-
5.000.000	-	-	1,1	-	-	-	-	-	-	-	-

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	11,6	20,4	12,0	6,8	17,3	16,2	8,4	13,8	15,6	12,1
30.000	11,0	19,2	11,3	6,4	16,4	15,3	8,0	13,1	14,8	11,4
40.000	9,4	16,3	9,7	5,5	14,0	13,1	6,9	11,4	12,7	9,8
50.000	8,3	14,4	8,6	4,9	12,5	11,6	6,2	10,3	11,3	8,6
60.000	7,6	13,0	7,8	4,4	11,3	10,5	5,6	9,4	10,3	7,8
70.000	7,0	11,9	7,2	4,0	10,4	9,7	5,2	8,7	9,5	7,2
80.000	6,5	11,0	6,7	3,8	9,7	9,0	4,9	8,2	8,9	6,7
90.000	6,1	10,3	6,3	3,5	9,1	8,5	4,6	7,7	8,3	6,3
100.000	5,7	9,7	6,0	3,3	8,6	8,0	4,4	7,4	7,9	5,9
200.000	4,0	6,5	4,1	2,3	5,9	5,5	3,1	5,3	5,5	4,1
300.000	3,2	5,2	3,3	1,8	4,8	4,4	2,5	4,3	4,5	3,3
400.000	2,7	4,4	2,8	-	4,1	3,8	2,2	3,8	3,8	2,8
500.000	2,4	3,9	2,5	-	3,6	3,4	1,9	3,4	3,4	2,5
750.000	1,9	3,1	2,0	-	2,9	2,7	-	2,8	2,8	2,0
1.000.000	1,7	2,6	1,7	-	2,5	2,3	-	2,4	2,4	1,7
2.000.000	-	1,8	-	-	1,7	1,6	-	1,7	1,7	-
3.000.000	-	1,4	-	-	1,4	1,3	-	-	1,3	-
4.000.000	-	1,2	-	-	1,2	-	-	-	1,2	-
5.000.000	-	1,0	-	-	1,1	-	-	-	1,0	-

Prospetto 5 - Valori dei coefficienti a, b, c e dell'indice di determinazione R² (%) delle funzioni utilizzate per l'interpolazione degli errori campionari delle stime di TOTALI i riferite alle *persone* per totale Italia, ripartizione geografica, tipo di comune e regione

ZONE TERRITORIALI	a	b	c	R ²
ITALIA	22062,800	0,007	-0,00000000012	83,8
RIPARTIZIONI GEOGRAFICHE				
Nord-ovest	12570,650	0,014	-0,00000000096	79,2
Nord-est	8921,840	0,014	-0,00000000100	79,6
Centro	10240,390	0,015	-0,00000000100	84,0
Sud	9383,200	0,014	-0,00000000100	86,2
Isole	5818,960	0,021	-0,00000000300	89,5
REGIONI				
Piemonte	5361,250	0,024	-0,00000000600	82,4
Valle d'Aosta	311,580	0,049	-0,00000052300	76,3
Lombardia	11080,970	0,019	-0,00000000200	76,5
- <i>Bolzano</i>	1229,610	0,044	-0,00000008000	74,5
- <i>Trento</i>	1382,840	0,040	-0,00000006700	72,8
Veneto	5811,320	0,024	-0,00000000600	79,5
Friuli-Venezia Giulia	2219,710	0,031	-0,00000003500	69,9
Liguria	2392,700	0,033	-0,00000002700	84,7
Emilia-Romagna	6014,190	0,023	-0,00000000600	79,2
Toscana	5332,730	0,029	-0,00000001000	82,7
Umbria	1712,860	0,036	-0,00000004400	83,6
Marche	2640,070	0,028	-0,00000001300	79,9
Lazio	7934,480	0,024	-0,00000000500	84,2
Abruzzo	2595,250	0,032	-0,00000002100	78,2
Molise	654,370	0,047	-0,00000017600	79,1
Campania	6280,720	0,026	-0,00000000500	84,9
Puglia	5327,270	0,026	-0,00000000600	83,5
Basilicata	1039,920	0,044	-0,00000008800	83,8
Calabria	2830,430	0,042	-0,00000002100	87,9
Sicilia	5208,320	0,026	-0,00000000600	89,1
Sardegna	2476,920	0,033	-0,00000002400	84,0