



**File di microdati
per la ricerca**

**Condizione e Integrazione sociale dei
Cittadini Stranieri**
2011-2012

Aspetti metodologici dell'indagine

INDICE

1. Introduzione	3
2. La popolazione di riferimento.....	4
3. Il disegno campionario	4
4. La rilevazione e il trattamento dei dati	10
5. La metodologia di calcolo dei pesi campionari	12
6. La diffusione dei risultati dell'indagine	23
7. Glossario	23
8. Riferimenti bibliografici.....	25
9. Contatti	25

1. Introduzione

Nel 2011-2012 l'Istat ha condotto, per la prima volta, la rilevazione statistica sulla "Condizione e Integrazione sociale dei cittadini stranieri"¹ per ampliare il quadro informativo sul fenomeno migratorio e sulla presenza straniera in Italia e rilevare informazioni su numerosi aspetti che interessano le condizioni di vita e il processo di integrazione dei cittadini stranieri in Italia.

Molti sono i temi affrontati, riferiti a diversi aspetti della vita dei cittadini stranieri, come la famiglia, i matrimoni, i figli, la formazione scolastica e professionale, l'appartenenza religiosa e linguistica, la storia migratoria, la storia lavorativa, le attuali condizioni di lavoro, le condizioni di salute, l'utilizzo e l'accessibilità ai servizi sanitari, gli stili di vita, le relazioni sociali, la partecipazione sociale, le esperienze di discriminazione vissuta, la sicurezza, le esperienze di vittimizzazione subite, le condizioni abitative, ecc.

Per la varietà e ricchezza dei temi trattati, l'indagine segna un passaggio rilevante da parte della statistica ufficiale sulla conoscenza della presenza straniera in Italia, ampliando il panorama delle informazioni disponibili ad ambiti precedentemente inesplorati o esplorati solo parzialmente, e comunque con minore dettaglio esplicativo, nonché su comportamenti, atteggiamenti e opinioni dei cittadini stranieri, ad integrazione e complemento dei dati di fonte amministrativa, correntemente prodotti dall'Istat.

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

Per maggiori informazioni sulla rilevazione è possibile consultare la pagina web <http://www.istat.it/it/archivio/10825>

¹ L'indagine è stata realizzata con l'interesse e il contributo finanziario di varie Istituzioni quali: il Dipartimento per le Pari Opportunità della Presidenza del Consiglio dei Ministri, il Ministero della Salute e il Dipartimento per le Libertà Civili e l'Immigrazione del Ministero dell'Interno, in qualità di Autorità responsabile per l'Italia del Fondo Europeo per l'Integrazione dei cittadini dei Paesi terzi.

2. La popolazione di riferimento

La popolazione di riferimento dell'indagine "Condizione e Integrazione sociale dei Cittadini Stranieri" è costituita dalle famiglie residenti in Italia con almeno un cittadino straniero (residente) e dai cittadini stranieri che le compongono.

Per famiglia si intende la famiglia di fatto, ossia un insieme di persone coabitanti e legate da vincoli di matrimonio, parentela, affinità, adozione, tutela o affettivi. Sono quindi esclusi dalla popolazione di riferimento i cittadini stranieri che vivono permanentemente nelle comunità (ospedali, istituti religiosi, brefotrofi, carceri, ecc.) e quelli presenti, ma non residenti sul territorio nazionale.

I cittadini stranieri sono identificati sul principio della cittadinanza² e non del luogo di nascita. Sono oggetto della rilevazione anche le persone di cittadinanza italiana per acquisizione (straniere alla nascita), di seguito definite naturalizzati, conviventi in famiglia con persone straniere.

I nazionali, vale a dire i cittadini italiani sin dalla nascita, presenti tra i componenti delle famiglie campionate non sono stati intervistati, ma per questi sono state comunque raccolte informazioni circa le caratteristiche strutturali, di tipo sociodemografico (sesso, età, cittadinanza, stato di nascita, titolo di studio, ecc.). Questa parte di popolazione, infatti, non essendo oggetto della rilevazione, è stata investigata soltanto con riferimento alle informazioni necessarie a ricostruire la composizione familiare, così da disporre di un set completo di dati anche per le famiglie miste, cioè composte da italiani e stranieri.

La popolazione straniera cui fanno riferimento le stime è stata calcolata a partire da quella residente al Censimento 2011.

3. Il disegno campionario

3.1. Definizione della dimensione campionaria

Il disegno campionario dell'indagine sugli stranieri presenta le caratteristiche generali dei disegni utilizzati per le indagini ISTAT sulle famiglie condotte con intervista diretta e selezione dalle anagrafi comunali.

In particolare, si tratta di un disegno a due stadi di selezione, dove le unità di primo stadio sono i comuni e le unità di secondo stadio sono le famiglie. I comuni sono stratificati per regione e tipologia comunale e selezionati con probabilità proporzionale alla popolazione

² Rientrano nell'accezione di cittadino straniero anche gli apolidi, cioè persone che non sono riconosciute cittadini da alcuno Stato. Tuttavia, data l'esiguità di questa categoria di persone, non si sono verificati casi di questo tipo nelle famiglie campione.

straniera residente. Per garantire che le diverse nazionalità siano opportunamente rappresentate nel campione di comuni estratti al primo stadio, è stata effettuata una selezione bilanciata dei comuni sulla base delle nazionalità presenti.

La lista di selezione disponibile per le unità di primo stadio è l'archivio dei comuni italiani, che contiene, per ciascun comune, il numero degli stranieri residenti per sesso e cittadinanza.

Nella progettazione dello schema di selezione dei comuni si è dovuto tenere conto del fatto che la popolazione degli stranieri residenti presenta una distribuzione molto disomogenea sul territorio, concentrata soprattutto nelle regioni del Centro-Nord (tabella 1). Al momento della selezione delle unità risultava che in alcuni comuni non erano presenti cittadini stranieri (circa 80 comuni) e molti altri ne avevano un numero molto esiguo.

In considerazione del fatto che nella progettazione di un disegno campionario a due stadi è necessario fissare il numero minimo di interviste da effettuare in ogni comune, tenendo conto, inoltre, dell'esigenza di prevedere un adeguato numero di famiglie sostitutive per le famiglie non rispondenti, è stata individuata una soglia per permettere di escludere dalla lista di selezione quei comuni con popolazione straniera nulla o eccessivamente esigua per la gestione dell'indagine sul campo. Inoltre, la distribuzione molto disomogenea della presenza straniera nei comuni del Centro – Nord rispetto a quella del Sud e Isole ha fatto anche propendere per una soglia differenziata per area geografica, dal momento che la soglia ottimale per il Centro-Nord (intorno alle 100 unità) avrebbe comportato una copertura non accettabile per il Sud e le Isole (inferiore al 90%).

Il disegno di campionamento è di tipo complesso e si avvale di due differenti schemi di campionamento. Nell'ambito di ognuno dei domini definiti dall'incrocio della regione geografica con quattro aree che identificano la tipologia comunale, i comuni italiani sono suddivisi in due sottoinsiemi sulla base della popolazione straniera residente:

- l'insieme dei comuni Auto Rappresentativi (Ar) costituito dai comuni di maggiore dimensione demografica;
- l'insieme dei comuni Non Auto Rappresentativi (o Nar) costituito dai rimanenti comuni.

Per i comuni Ar si adotta un disegno ad uno stadio stratificato, mentre per i Nar si usa un disegno a due stadi con stratificazione delle unità di primo stadio. Nell'ambito dell'insieme dei comuni Ar, ciascun comune viene considerato come uno strato a sé stante

Da ogni comune campione si estrae un campione di famiglie anagrafiche in cui è presente almeno un cittadino straniero, selezionate casualmente con passo sistematico

dall'anagrafe del comune stesso, che costituiscono le unità finali di campionamento; in ogni famiglia campione sono rilevate le caratteristiche oggetto di indagine su tutti i cittadini stranieri che ne fanno parte e, come già accennato (par.2), soltanto le caratteristiche sociodemografiche sui cittadini nazionali.

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

3.2. La numerosità campionaria

La numerosità campionaria teorica è di circa 12.000 famiglie e il disegno campionario a due stadi. Tale numerosità campionaria garantisce l'attendibilità di stime di frequenze a livello dei domini di stima pianificati, ovvero l'intero territorio nazionale, sei ripartizioni geografiche e quattro tipologie comunali (cfr. Glossario).

E' bene precisare che il livello territoriale per il quale è possibile produrre stime attendibili dipende dal livello delle stime stesse e dalla procedura di stima che è possibile mettere in atto sulla base dei risultati conseguiti sul campo.

Nella tabella 1 è illustrata l'allocazione del campione di famiglie tra le regioni, le ripartizioni e le tipologie comunali individuata a partire dalla distribuzione della popolazione straniera residente disponibile alla data di estrazione. Poiché nell'archivio dei comuni le informazioni sugli stranieri residenti sono presenti soltanto con riferimento agli individui, per calcolare il numero di famiglie sono stati utilizzati i dati sul numero di famiglie con almeno uno straniero desunti dai bilanci demografici relativi agli stranieri residenti.

L'allocazione del campione è avvenuta in due fasi.

1. Allocazione del campione tra le sei ripartizioni geografiche. E' stata definita un'allocazione del campione tra le ripartizioni geografiche in un'ottica di compromesso tra l'allocazione uniforme e l'allocazione proporzionale, attribuendo un peso pari a 0,8 a quella uniforme e 0,2 a quella proporzionale, nell'ottica di privilegiare l'attendibilità delle stime a livello di ripartizione.
2. Allocazione del campione delle ripartizioni tra le regioni. All'interno di ciascuna ripartizione l'allocazione tra le regioni è stata ottenuta nella stessa ottica della prima fase, ma con pesi invertiti, cioè attribuendo un peso pari a 0,2 all'allocazione uniforme e 0,8 a quella proporzionale. In tal modo le regioni con un numero esiguo di stranieri residenti si vedono assegnato un campione molto piccolo.

Tabella 1. Allocations del campione tra i domini territoriali

Dominio territoriale	Popolazione straniera (*)	Campione FAMIGLIE	% di copertura in corrispondenza delle soglie di popolazione straniera residente nei comuni	
			80	100
Piemonte	310.543	579	92,9	90,9
Valle D'Aosta/Vallée d'Aoste	6.604	142	79,8	73,1
Lombardia	815.335	1.358	98,0	97,2
Bolzano	32.945	166	93,0	91,4
Trento	37.889	145	87,5	83,2
Veneto	403.985	779	99,1	98,6
Friuli Venezia Giulia	83.306	237	96,1	94,7
Liguria	90.881	288	94,9	93,9
Emilia-Romagna	365.687	751	99,6	99,3
Toscana	275.149	666	99,5	99,1
Umbria	75.631	256	99,3	98,4
Marche	115.299	311	97,7	96,5
Lazio	390.993	1.008	98,6	98,1
Abruzzo	59.749	759	90,5	88,3
Molise	6.271	210	63,3	59,0
Campania	114.792	1.287	91,3	89,6
Puglia	63.868	742	93,4	90,7
Basilicata	9.595	209	74,1	69,5
Calabria	50.871	582	85,9	82,2
Sicilia	98.152	1.281	93,4	91,0
Sardegna	25.106	514	79,3	74,5
Nord-Ovest	1.223.363	2.367	96,4	95,3
Nord-Est	923.812	2.078	98,3	97,7
Centro	857.072	2.240	98,8	98,2
Sud-Est	129.888	1.711	90,6	88,0
Sud-Ovest	175.258	2.079	88,8	86,3
Isole	123.258	1.794	90,5	87,7
Comuni metropolitani	679.551	2.940	100,0	100,0
Comuni della cintura metropolitana	342.367	2.544	99,0	98,5
Comuni fino a 10.000 ab.	936.155	2.826	88,3	84,9
Comuni oltre 10.000 ab.	1.474.578	3.495	99,9	99,8
Italia	3.432.651	12.269	96,7	95,6

(*) Anno 2009

3.3. Stratificazione e selezione delle unità campionarie

L'obiettivo della stratificazione è quello di formare gruppi (o strati) di unità caratterizzate, relativamente alle variabili oggetto d'indagine, da massima omogeneità interna agli strati e massima eterogeneità fra gli strati. Il raggiungimento di tale obiettivo si traduce in termini statistici in un guadagno nella precisione delle stime, ossia in una riduzione dell'errore campionario a parità di numerosità campionaria.

Nell'indagine in oggetto, i comuni vengono stratificati in base alla loro dimensione in termini di stranieri residenti e nel rispetto delle seguenti condizioni:

- autponderazione del campione a livello regionale;
- scelta di un numero minimo di famiglie da intervistare in ciascun comune campione;

- scelta del numero, \bar{n} , di comuni campione da estrarre da ciascuno strato Nar: tale parametro è stato posto pari a 3;
- formazione di strati aventi ampiezza approssimativamente costante in termini di popolazione residente.

Il procedimento di stratificazione, attuato all'interno di ogni dominio territoriale individuato dalle quattro aree rispetto alle quali sono stati classificati i comuni (cfr Glossario) di ciascuna regione geografica, si articola nelle seguenti fasi:

- ordinamento dei comuni del dominio in ordine decrescente secondo la loro dimensione demografica in termini di popolazione straniera residente;
- determinazione di una soglia di popolazione per la definizione dei comuni Ar, mediante la relazione:

$${}_r\lambda = \frac{{}_r\bar{m} \cdot {}_r\delta}{{}_rf}$$

in cui per la generica regione geografica r si è indicato con: ${}_r\bar{m}$ il numero minimo di famiglie da intervistare in ciascun comune campione; ${}_r\delta$ il numero medio di componenti per famiglia (nel caso specifico è la dimensione media delle famiglie di stranieri); ${}_rf$ la frazione di campionamento, definita dal rapporto tra la dimensione campionaria e la popolazione straniera;

- suddivisione di tutti i comuni nei due sottoinsiemi Ar e Nar: i comuni di dimensione superiore o uguale a ${}_r\lambda$ sono definiti come comuni Ar e i rimanenti come Nar;
- suddivisione dei comuni dell'insieme Nar in strati aventi dimensione, in termini di popolazione straniera residente, approssimativamente costante e all'incirca pari \bar{n} volte la soglia ${}_r\lambda$.

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 3 comuni campione con probabilità proporzionale alla dimensione demografica e seguendo lo schema di selezione bilanciata descritto nel paragrafo seguente.

Il numero minimo di interviste per comune, ${}_r\bar{m}$, è stato posto a 10 per le regioni del Centro-Nord e a 8 per le regioni del Sud e delle Isole. Sulla base di questa scelta e tenendo in considerazione l'esigenza di disporre di un numero sufficiente di famiglie per le sostituzioni, la soglia di popolazione straniera per l'inclusione dei comuni è stata fissata a 100 per le regioni del Centro-Nord e 80 per le regioni del Sud e delle Isole. In tal modo

l'universo di selezione è costituito da 4033 comuni, che garantiscono una copertura della popolazione degli stranieri residenti di circa il 96% (tabella 1).

3.3.1 Selezione dei comuni bilanciata rispetto alle nazionalità

Per l'estrazione dei comuni all'interno degli strati è stata studiata una selezione bilanciata, allo scopo di conseguire una maggiore rappresentatività delle nazionalità straniere presenti sul territorio in modo molto disomogeneo. Si è cercato di tenere conto in tal modo della distribuzione della presenza straniera nei comuni in termini di nazionalità, sebbene non fosse possibile effettuare una stratificazione per nazionalità. In altri termini, si è posto un vincolo sulla distribuzione risultante dei comuni selezionati, realizzando un campione bilanciato (Deville e Tillé 2004)³.

In particolare, in ogni generica area geografica a , $m(a)$ comuni campione sono selezionati dagli $M(a)$ comuni universo mediante un campione bilanciato, con probabilità di inclusione definite all'interno di ciascuno strato in modo proporzionale alla popolazione straniera residente. Le equazioni di bilanciamento impongono che le stime dirette dei totali di popolazione relativi alle N prefissate nazionalità presenti nei comuni coincidano con i corrispondenti totali noti; in simboli:

$$\sum_{c=1}^{m(a)} \frac{\mathbf{x}_c}{\pi_c} = \sum_{c=1}^{M(a)} \mathbf{x}_c$$

in cui \mathbf{x}_c è la probabilità di inclusione del comune c e

$$\mathbf{x}'_c = ({}_1P_c, \dots, {}_n P_c, \dots, {}_N P_c, \pi_c)$$

è il vettore di variabili ausiliarie riferito al generico comune c , in cui ${}_n P_c$ indica il numero di stranieri nazionalità n residenti nel comune c , nota dall'archivio dei comuni.

Poiché non era possibile tenere in considerazione tutte le nazionalità, anche quelle con pochissime presenze sul territorio nazionale, è stato scelto di realizzare un *bilanciamento* basato sulle nazionalità più numerose individuate a livello nazionale. Dopo un'analisi delle possibilità concrete di estrazione di un campione bilanciato, si è scelto di procedere ad un bilanciamento rispetto alle prime 15 nazionalità⁴ individuate a livello nazionale, che coprono circa il 72% della popolazione straniera totale, vincolate su tre ripartizioni geografiche (Nord, Centro e Sud-Isole).

³ Deville, J.C. and Tillé, Y. (2004). Efficient Balanced Sampling: The Cube Method. *Biometrika*, 91, 893-912.

⁴ Le 15 nazionalità più presenti al momento dell'estrazione del campione sono: Romania, Albania, Marocco, Repubblica Popolare Cinese, Ucraina, Filippine, Tunisia, Polonia, India, Moldavia, Macedonia- ex Rep. Jugoslavia, Ecuador, Perù, Egitto, Bangladesh.

3.3.2 Secondo stadio di campionamento: selezione delle famiglie

Una volta estratti i comuni campione, al secondo stadio di campionamento vengono selezionate le famiglie campione dalla lista delle famiglie in cui è presente almeno uno straniero residente. Come già accennato, tutti gli individui stranieri appartenenti a tali famiglie vengono intervistati poiché rappresentano la popolazione di riferimento. Gli individui con cittadinanza italiana, invece, non vengono intervistati poiché non appartengono alla popolazione di riferimento (par.2).

Il numero delle famiglie estratte è stata tre volte superiore la numerosità campionaria teorica, così da consentire la costituzione di *quartine* di nominativi campione da utilizzare per la sostituzione delle famiglie non rispondenti: ad ogni unità campione sono quindi state associate tre unità di riserva. La costituzione di quartine di nominativi campione offre maggiori garanzie affinché la numerosità campionaria individuata in fase di progettazione sia rispettata fino alla conclusione dell'indagine. Per ridurre il rischio di effetti distorsivi del campione, nella formazione delle quartine si è proceduto secondo un criterio di omogeneità per cittadinanza (del capofamiglia), per alfabetico di via e per dimensione anagrafica della famiglia.

4. La rilevazione e il trattamento dei dati

La rilevazione è stata condotta attraverso interviste dirette condotte mediante tecnica CAPI (*Computer Assisted Personal Interviewing*). Gli individui al di sotto dei 14 anni nelle famiglie campione sono stati intervistati in modalità *proxy*, cioè le risposte sono state fornite da un genitore o un componente maggiorenne della famiglia. La stessa modalità di intervista è stata utilizzata per le persone temporaneamente assenti.

Considerato il target di popolazione cui è stata rivolta la rilevazione, costituito prevalentemente da individui stranieri, per facilitare la comunicazione durante l'intervista e garantire un maggior livello di comprensione delle domande il questionario è stato tradotto in 10 lingue⁵.

Le operazioni di rilevazione si sono svolte da maggio del 2011 a novembre 2012 su un campione effettivo di 9.553 famiglie con almeno un cittadino straniero residenti in 833 comuni italiani. La rilevazione ha riguardato complessivamente 20.379 individui con cittadinanza straniera, 696 naturalizzati e 4.251 individui con cittadinanza italiana dalla nascita (nazionali).

⁵ Inglese, francese, spagnolo, tedesco, romeno, albanese, polacco, russo, arabo, cinese.

Il ricorso alla tecnica CAPI è sembrata la risposta migliore per far fronte alla complessità metodologica e organizzativa dell'indagine. Per la complessità del questionario, in cui sono presenti percorsi molto articolati e diversificati per i diversi target di popolazione, la possibilità per il rilevatore di essere guidato nell'intervista dal questionario informatizzato è apparso un fattore essenziale per la coerenza e completezza dell'intervista.

Inoltre, l'utilizzo della tecnica Capi consente di anticipare alla fase di acquisizione dell'informazione una parte considerevole dei controlli, ottenendo dati grezzi meno affetti da errore; le risposte fornite vengono sottoposte ad un piano di verifica implementato nel questionario elettronico e quindi risolte sin dal momento della rilevazione. Tale piano si basa sul controllo dei domini delle variabili, dei percorsi del questionario e delle incongruenze logiche tra le informazioni raccolte.

Il questionario elettronico prevede sia controlli di tipo hard, sia controlli di tipo soft. Nel primo caso, a fronte di un'informazione incoerente, la registrazione del dato errato viene inibita e quindi l'intervista non può proseguire finché esso non viene corretto; nel secondo caso, la registrazione del dato errato viene consentita previa segnalazione dell'errore.

Generalmente, un controllo di tipo hard viene attivato se l'informazione risulta assolutamente inaccettabile, un controllo di tipo soft quando essa è altamente improbabile.

La strategia di controllo e correzione successiva alla fase di raccolta dati cerca di salvaguardare il più possibile l'informazione rilevata, partendo dal presupposto che l'insieme dei controlli inseriti in fase di acquisizione fornisca dati grezzi di elevata qualità.

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

L'Istat è tenuto per legge a rispettare il segreto statistico (art. 9 del decreto legislativo n. 322/1989). I dati raccolti in occasione dell'indagine possono quindi essere utilizzati, esclusivamente a fini statistici, dall'Istat e dagli altri enti ed uffici di statistica facenti parte del Sistema statistico nazionale. I medesimi dati possono inoltre essere utilizzati, per sole finalità di ricerca scientifica, da soggetti non facenti parte del citato Sistema statistico nazionale nei limiti e secondo le modalità stabilite dall'art. 7 del Codice di deontologia e di buona condotta per i trattamenti di dati personali a scopi statistici e di ricerca scientifica effettuati nell'ambito del Sistema statistico. La diffusione dei dati e la loro comunicazione avviene sotto forma aggregata in modo tale che non sia possibile risalire alle persone che le forniscono, assicurando così la massima riservatezza. Ai sensi della stessa normativa

(art.7, comma 2) la persona intervistata potrà decidere se rispondere o meno ad alcuni quesiti di natura "sensibile" contenuti nelle varie sezioni del questionario e indicati sul retro dei questionari. Ai sensi della legge che disciplina la tutela della privacy (D.Lgs.196/2003), titolare delle rilevazioni è l'Istat.

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

5. La metodologia di calcolo dei pesi campionari

L'indagine deve produrre le stime riferite al numero di individui che nella popolazione di riferimento possiedono una certa caratteristica o il livello di una quantità misurata sugli individui.

Per il calcolo dei coefficienti di riporto all'universo si utilizza una procedura generalizzata di stima, basata sull'uso di una famiglia di stimatori, noti in letteratura come calibration estimator (stimatori di ponderazione vincolata). La metodologia alla base di tali stimatori consente la determinazione di un unico coefficiente di riporto all'universo in grado di produrre stime coerenti a totali noti, desunti da fonti esterne, e correlati alle principali variabili oggetto di indagine.

La famiglia di stimatori di ponderazione vincolata coincide asintoticamente con lo stimatore di regressione generalizzato: per campioni sufficientemente grandi, quindi, tali stimatori hanno approssimativamente le stesse proprietà, ovvero sono corretti, consistenti e con la stessa varianza campionaria.

La strategia adottata per la costruzione dei coefficienti di riporto all'universo si sviluppa attraverso le fasi tipiche utilizzate per la costruzione degli stimatori nelle varie indagini campionarie dell'Istituto. In particolare possiamo distinguere:

- la determinazione della probabilità di inclusione di ogni unità statistica e del relativo peso diretto, pari all'inverso della probabilità di inclusione;
- calcolo dei coefficienti di correzione per mancata risposta totale;
- determinazione dei coefficienti di riporto all'universo finali vincolati ai totali noti desunti da fonti esterne all'indagine.

5.1 La probabilità di inclusione e il peso diretto

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 stesso. A tale scopo, ad ogni unità campionaria viene attribuito un peso, o coefficiente di riporto all'universo, che indica quante unità della popolazione sono rappresentate, rispettivamente, da ogni unità presente nel campione.

Senza perdere di generalità, definiamo la seguente simbologia:

U popolazione di riferimento oggetto di indagine;

y_k valore della variabile Y assunto dalla k -esima osservazione della popolazione;

y_j valore della variabile Y assunto dalla j -esima osservazione della popolazione;

π_j probabilità, assegnata dal disegno di campionamento, che l'unità j -esima sia inclusa nel campione S ;

Il totale di una generica variabile Y , calcolato sull'intera popolazione, assume la seguente forma:

$$Y = \sum_{k \in U} y_k \quad (1)$$

Il disegno di campionamento assegna le probabilità di inclusione ad ogni unità del campione in modo tale che:

$$\hat{Y} = \sum_{j \in s} y_j \frac{1}{\pi_j} \quad (2)$$

sia uno stimatore corretto della (1).

La probabilità di inclusione di un generico individuo è data: dalla probabilità di estrazione del comune di residenza (direttamente proporzionale all'ampiezza demografica dei comuni all'interno dello strato); e dalla probabilità di estrazione della famiglia di appartenenza tra le famiglie eleggibili del comune.

Per una generica famiglia eleggibile j , nel comune i dello strato h , il peso diretto d_{hij} , inverso della probabilità di inclusione π_{hij} , assume la seguente forma:

$$d_{hij} = \frac{1}{\pi_{hij}} = \frac{1}{c_h} \frac{P_h}{P_{hi}} \frac{M_{hi}}{m_{hi}} \quad (3)$$

dove :

- h denota l'indice di strato;
- i è l'indice di comune;
- j denota l'indice della famiglia;
- c_h indica il numero di comuni campione dello strato h ;
- P_h indica il totale della popolazione residente nello strato h ;
- P_{hi} il totale della popolazione residente nel comune i dello strato h ;
- M_{hi} indica il totale di famiglie eleggibili nel comune i dello strato h ;
- m_{hi} indica il numero di famiglie campione nel comune i dello strato h .

5.2. La correzione per mancata risposta

Nel corso della fase di raccolta delle informazioni presso le unità che formano il campione, come accade per tutte le indagini statistiche, alcune di queste si trovano nell'impossibilità di partecipare all'indagine. Questo aspetto comporta che al termine della rilevazione, la numerosità campionaria teorica individuata nel disegno e quella effettiva differiscono numericamente (mancata risposta totale). Nell'indagine in questione, l'utilizzo delle quartine ha fatto sì che il problema della mancata risposta totale si mantenesse a livelli bassi: il campione finale è infatti pari a 9.553 unità rispetto alle 12.269 previste in fase di definizione del disegno nel caso del campione nazionale. Per ovviare alla mancata partecipazione di alcune unità del campione teorico all'indagine, nella fase di calcolo dei coefficienti di riporto all'universo viene introdotto, come di norma, un correttore per mancata risposta che, sotto l'ipotesi che il comportamento dei rispondenti sia simile a quello dei non rispondenti all'interno dello stesso strato, assume la forma dell'inverso del tasso di risposta (δ_h):

$$\frac{1}{\delta_h} = \frac{m_h}{m_h^r} \quad (4)$$

in cui m_h^r rappresenta il numero di famiglie rispondenti nello strato h .

In questa maniera, il coefficiente di riporto all'universo corretto per mancata risposta, da assegnare al campione rispondente, risulta essere:

$$k_{hij} = d_{hij} \frac{1}{\delta_h} = \frac{1}{c_h} \frac{P_h}{P_{hi}} \frac{M_{hi}}{m_{hi}} \frac{m_h}{m_h^r} \quad (5)$$

5.3. La calibrazione a fonti esterne

Per il calcolo dei coefficienti di riporto all'universo finali si adottano gli stimatori di ponderazione vincolata (*calibration estimator*). La metodologia si basa sull'utilizzo di opportune informazioni ausiliarie, sintetizzate in totali noti, che, correlate con le variabili principali oggetto di indagine, hanno la funzione di aumentare l'accuratezza delle stime. I pesi finali si ottengono risolvendo un problema di minimo vincolato, in cui la funzione da minimizzare è una funzione di distanza tra i pesi diretti corretti per la mancata risposta (\mathbf{k}) e i pesi finali (\mathbf{w}) delle famiglie del campione rispondente (S^r), e i vincoli sono proprio le condizioni di uguaglianza delle stime campionarie di alcune variabili ausiliarie con i rispettivi totali noti desunti da fonti esterne all'indagine⁶

$$\begin{cases} \text{Min} \left\{ \sum_{j \in S^r} \text{dist}(k_j, w_j) \right. \\ \left. \sum_{j \in S^r} x_j * w_j = t \right. \end{cases}$$

dove t è il vettore dei totali noti e x_j è il vettore delle variabili ausiliarie osservate sulla j -esima unità campionaria appartenente al campione rispondente (S^r). La funzione di distanza utilizzata è la logaritmica troncata.

I totali noti introdotti come vincoli nel calcolo dei pesi finali consentono di migliorare l'accuratezza delle stime, poiché quanto più le variabili ausiliarie considerate sono correlate con le variabili oggetto d'indagine, tanto più si riduce la distorsione delle stime. Nello specifico dell'indagine Istat 'Condizione e integrazione sociale dei cittadini stranieri residenti in Italia' le stime campionarie sono state vincolate ai seguenti totali noti:

⁶ La calibrazione è una calibrazione integrata, vale a dire che tutti gli individui della medesima famiglia hanno lo stesso coefficiente di riporto all'universo.

- stranieri residenti in Italia secondo le 15 cittadinanze più rappresentate (Romania, Albania, Marocco, Repubblica Popolare Cinese, Ucraina, Filippine, Tunisia, Polonia, India, Moldavia, Macedonia- ex Rep. Jugoslavia, Ecuador, Perù, Egitto, Bangladesh)
- stranieri residenti per ripartizione (Nord – Ovest, Nord – Est, Centro, Sud – Ovest, Sud – Est, Isole) e 8 gruppi di cittadinanze (Unione Europea; Europa Centro-Orientale; Africa Settentrionale; Africa Occidentale; Asia Centro-Meridionale; Asia Orientale; America Centro-Meridionale; Altro)
- stranieri residenti per ripartizione, sesso e classi d'età (0-5, 6-15, 16-24, 25-34, 35-44, 45-64, 65 e più);
- stranieri residenti per ripartizione e tipologia del comune di residenza (Comuni metropolitani, comuni periferia dei comuni metropolitani, altri comuni fino a 10.000 abitanti, altri comuni con più di 10.000 abitanti).

5.4. I dati campionari e il loro utilizzo

Il file di microdati per la ricerca è composto da un file individui, composto da un record per ogni individuo componente la famiglia intervistata.

Per un corretto utilizzo dei dati si fa presente che a seconda della selezione che si opera sui dati è possibile effettuare elaborazioni sulle seguenti unità di analisi:

- Individui.* Ogni componente è individuato dal numero progressivo della famiglia (IDFAM) e dal numero d'ordine all'interno della stessa (IDIND). Per selezionare i componenti della stessa famiglia si dovrà procedere a considerare tutti i record individuali che hanno lo stesso progressivo della famiglia.
- Famiglie.* Nel procedere ad analisi sulle famiglie è sufficiente selezionare solo il primo componente di ciascuna utilizzando il numero d'ordine individuale "01" (IDIND=01), sempre presente in ogni famiglia, che identifica il primo componente della famiglia.

Per agevolare l'utente nell'uso dei dati sono state costruite alcune variabili di uso corrente. E' il caso, per esempio, della variabile *Target group* (TG), che individua specifici sottoinsiemi di popolazione costruiti a partire dalle seguenti informazioni individuali, rilevate per ciascun componente della famiglia campione: paese di nascita, cittadinanza alla nascita e cittadinanza attuale. I target group così individuati in base alle risultanze campionarie sono:

- S1 (TG=01), *Stranieri immigrati*: persone nate all'estero, di cittadinanza straniera, sia alla nascita sia al momento della rilevazione;

- S2 (TG=02), *Stranieri di seconda generazione, non naturalizzati*: persone nate in Italia, di cittadinanza straniera sia alla nascita sia al momento della rilevazione;
- S3 (TG=03), *Immigrati stranieri, naturalizzati di prima generazione*: nati all'estero, di cittadinanza straniera alla nascita e italiani al momento della rilevazione;
- S4 (TG=04), *Stranieri naturalizzati di seconda generazione*: nati in Italia, di cittadinanza straniera alla nascita e italiani al momento della rilevazione;
- N1 (TG=07), *Stranieri ex-nazionali nati all'estero* (discendenti di emigrati italiani): nati all'estero, di cittadinanza italiana alla nascita e stranieri al momento della rilevazione;
- N2 (TG=08), *Stranieri ex-nazionali nati in Italia* (emigrati Italiani che hanno perso la cittadinanza italiana): nati in Italia, di cittadinanza italiana alla nascita e stranieri al momento della rilevazione;
- N3 (TG=20), *Nazionali nati all'estero*: nati all'estero, di cittadinanza italiana alla nascita e italiani al momento della rilevazione;
- N4 (TG=21), *Nazionali nati in Italia*: nati in Italia, di cittadinanza italiana alla nascita e italiani al momento della rilevazione.

La tabella 2 riporta nel dettaglio la numerosità per ciascuna tipologia di gruppo del campione.

Tabella 2. Individui componenti le famiglie campione per tipologia di target group

Tipologia	Target group	Numerosità campione	Stima nella popolazione
S1	Stranieri immigrati	17.544	3.452.625
S2	Stranieri di seconda generazione, non naturalizzati	2.821	582.868
S3	Immigrati stranieri, naturalizzati di prima generazione	497	176.789
S4	Stranieri naturalizzati di seconda generazione	199	74.920
N1	Stranieri ex-nazionali nati all'estero	1	275
N2	Stranieri ex-nazionali nati in Italia	13	947
N3	Nazionali nati all'estero	154	32.741
N4	Nazionali nati in Italia	4.097	1.123.258
Totale		25.326	5.444.424

Si fa presente che le stime riferite ai cittadini naturalizzati (S3 e S4) e ai nazionali (N3 e N4) sono stime “dirette”, cioè non sono calibrate; per questi individui è disponibile soltanto il peso diretto poiché non sono disponibili totali noti da altre fonti. A questi individui si applica quindi lo stesso coefficiente di riporto all'universo dei componenti stranieri della famiglia.

Il coefficiente di riporto all'universo (COEFFIN) è stato determinato in modo da poter essere usato indifferentemente per costruire stime relative agli individui e alle famiglie. Per l'elaborazione delle stime di indagine il COEFFIN deve essere diviso per 100.000.000. Per esempio, se in un record individuale il COEFFIN vale 109.783.297.628 significa che l'individuo rappresenta 1.097,83297628 ($109.783.297.628/100.000.000$) individui nella rispettiva popolazione residente in Italia, nel periodo di riferimento. Ponderando, quindi, le informazioni presenti nel record individuale per il valore del COEFFIN, le caratteristiche relative a tale individuo saranno stimate per 1.097,83297628 individui nella popolazione di riferimento. Analogamente, saranno stimate per 1.097,83297628 famiglie se il record in analisi è riferito alla famiglia.

Il periodo di riferimento dei fenomeni indagati varia da quesito a quesito.

I domini di studio, ossia gli ambiti territoriali rispetto ai quali sono riferiti i parametri di popolazione oggetto di stima, sono:

- l'intero territorio nazionale;
- le quattro ripartizioni geografiche (Nord Ovest, Nord Est, Centro, Sud e Isole);
- i comuni italiani suddivisi in:
 - A1) comuni centro e periferia di area metropolitana: Torino, Milano, Venezia, Genova, Bologna, Firenze, Roma, Napoli, Bari, Palermo, Catania, Cagliari;
 - A3) comuni non appartenenti all'area metropolitana aventi fino a 10.000 abitanti;
 - A4) comuni non appartenenti all'area metropolitana con oltre 10.000 abitanti.

L'aggregazione di alcuni domini di stima rispetto a quelli individuati in fase progettuale dal disegno di campionamento (cfr. par. 3) riflette le scelte metodologiche effettuate per il controllo del rischio di violazione della riservatezza (cfr. documento Descrizione del file).

La dicitura "RISERVATO ISTAT" è utilizzata nei casi in cui l'informazione può compromettere la riservatezza del dato oppure l'informazione non è statisticamente significativa (variabili d'appoggio, variabili non attendibili dal punto di vista campionario, ecc.).

Trattandosi di un'indagine di natura campionaria, l'utilizzatore dei dati dovrà valutare l'attendibilità delle stime ottenibili, sfruttando tutte le informazioni e gli strumenti di calcolo messi a disposizione, per procedere ad una corretta interpretazione dei dati. A tale scopo, i paragrafi successivi riportano informazioni relative alla valutazione degli errori di stima, con esempi di calcolo.

5.5. Valutazione del livello di precisione delle stime

Le stime prodotte da un'indagine campionaria sono sempre affette da errore. Questo si distingue in errore campionario, che deriva dall'incertezza derivante dall'aver osservato la variabile di interesse solo su una parte (campione) della popolazione; ed errore non campionario, che deriva da errori nelle liste della popolazione utilizzate per selezionare le unità del campione; mancate risposte parziali dovute a risposte mancanti o non ammissibili a causa di errori di rilevazione o di registrazione; in generale, da tutto ciò che ha a che fare con le tecniche di indagine utilizzate e i comportamenti dei rilevatori.

In questo paragrafo si descrivono le metodologie e le tecniche utilizzate per la valutazione dell'errore campionario associato alle stime prodotte.

Le principali statistiche per valutare l'errore campionario sono l'errore di campionamento assoluto e l'errore di campionamento relativo. La stima dell'errore di campionamento assoluto e relativo di una generica stima \hat{Y} sono definite dalle seguenti espressioni:

$$\sigma(\hat{Y}) = \sqrt{\widehat{Var}(\hat{Y})}$$

$$\hat{e}(\hat{Y}) = \frac{\sigma(\hat{Y})}{\hat{Y}}$$

Poiché le stime sono ottenute mediante uno stimatore di calibrazione con funzione di distanza logaritmico troncata, lo stimatore adottato non è funzione lineare dei dati campionari. Per la stima della varianza $\widehat{Var}(\hat{Y})$ si è quindi utilizzata un'espressione linearizzata in serie di Taylor, con la quale è possibile ricavare la varianza di ogni stimatore non lineare calcolando la varianza dell'espressione linearizzata ottenuta. In particolare, si fa riferimento allo stimatore di regressione generalizzata, sfruttando la convergenza asintotica di tutti gli stimatori di calibrazione a tale stimatore (Deville, Särndal, 1992)⁷.

Conoscendo la stima \hat{Y} di un parametro Y della popolazione e la stima dell'errore assoluto $\sigma(\hat{Y})$ ad essa associato, è possibile costruire un intervallo di confidenza che, con livello di fiducia α , contiene al suo interno il valore del parametro Y oggetto di stima; tale intervallo è:

$$\{\hat{Y} - k_{\alpha}\sigma(\hat{Y}) \leq Y \leq \hat{Y} + k_{\alpha}\sigma(\hat{Y})\}$$

⁷ Deville, J.-C. and C.-E. Sarndal (1992), Calibration estimators in survey sampling, Journal of the American Statistical Association 87, 376-382.

dove il valore di k_{α} dipende dalla forma della distribuzione campionaria dello stimatore e dal valore scelto per il livello di confidenza α ; per grandi campioni si fa comunemente riferimento alla distribuzione normale e si ha ad esempio, per $\alpha=0,05$, che $k=1,96$.

5.6. Presentazione sintetica degli errori campionari

Ad ogni stima generica corrisponde una stima dell'errore campionario relativo che consente di valutarne la precisione; pertanto, per consentire una corretta interpretazione delle stime prodotte, sarebbe necessario presentare contestualmente a ciascuna stima anche il corrispondente errore campionario stimato. Ciò, tuttavia, non è possibile quando le stime prodotte sono in numero molto elevato. Per questi motivi si ricorre frequentemente ad una presentazione sintetica delle stime degli errori campionari, basata sul metodo dei modelli regressivi. Questo metodo si basa sulla determinazione di una semplice funzione matematica che mette in relazione ciascuna stima con il proprio errore campionario relativo stimato.

Il modello utilizzato per le stime di frequenze assolute e relative riferite agli individui è il seguente:

$$\log(\hat{\epsilon}^2(\hat{Y})) = a + b * \log(\hat{Y})$$

dove i parametri a e b sono stimati con il metodo dei minimi quadrati. I modelli regressivi del tipo descritto, che permettono la presentazione sintetica degli errori di campionamento, sono stati ottenuti tramite un software generalizzato messo a punto dall'Istat.

Nel prospetto 3 sono riportati i valori dei coefficienti a e b e del coefficiente di determinazione R^2 dei modelli stimati per l'interpolazione degli errori campionari relativi delle stime di frequenze assolute e relative per il totale Italia e per le diverse ripartizioni geografiche.

Utilizzando gli opportuni coefficienti è possibile calcolare una stima dell'errore campionario relativo di una generica stima di una frequenza \hat{Y} applicando la seguente formula:

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

Prospetto 3 - Valori dei coefficienti a, b e R² delle funzioni utilizzate per l'interpolazione degli errori campionari delle stime

	a	b	R²
ITALIA	9,513502	-1,29689	97,22
RIPARTIZIONE GEOGRAFICA			
Nord – Ovest	9,387337	-1,28158	97,60
Nord – Est	8,41149	-1,24568	95,74
Centro	9,511312	-1,34720	95,21
Sud e Isole	6,269312	-1,18334	90,81
TIPO DI COMUNE			
Centro e periferia dei comuni metropolitani	8,072002	-1,21964	94,78
Altri comuni fino a 10.000 abitanti	8,896622	-1,25646	96,88
Altri comuni con più di 10.000 abitanti	9,04692	-1,28708	97,02

Infine, il prospetto 4 ha lo scopo di rendere più agevole e immediata la valutazione degli errori campionari. In testata sono elencati valori crescenti di stima di frequenze relative percentuali (0,5%, 1%, 2%, ..., 40%, 50%); in fiancata sono riportati i domini di riferimento delle stime; le celle interne contengono gli errori campionari relativi percentuali stimati mediante la formula precedente. Consultando queste tavole è possibile disporre di una valutazione immediata (anche se meno precisa rispetto all'applicazione della formula precedente) dell'errore campionario di una generica stima di una frequenza relativa percentuale (o assoluta, ricavabile moltiplicando la frequenza relativa al totale degli stranieri nel dominio di riferimento), cercando nella testata il valore che più si avvicina alla stima di interesse e in fiancata il dominio di riferimento.

Prospetto 4 - Valori interpolati degli errori campionari relativi percentuali delle stime

	STIME DI FREQUENZA RELATIVA PERCENTUALE								
	0,5%	1%	2%	5%	10%	20%	30%	40%	50%
ITALIA	15,5	9,9	6,3	3,5	2,2	1,4	1,1	0,9	0,8
RIPARTIZIONE GEOGRAFICA									
Nord-Ovest	30,2	19,4	12,4	6,9	4,4	2,8	2,2	1,8	1,6
Nord-Est	26,4	17,2	11,2	6,3	4,1	2,7	2,1	1,7	1,5
Centro	31,6	19,8	12,4	6,7	4,2	2,6	2,0	1,7	1,4
Sud e Isole	17,8	11,8	7,8	4,6	3,0	2,0	1,6	1,3	1,2
TIPO DI COMUNE									
Centro e Periferia di area metropolitana	25,1	16,5	10,8	6,2	4,1	2,7	2,1	1,7	1,5
Comuni oltre 10.000 abitanti	30,3	19,6	12,7	7,1	4,6	3,0	2,3	1,9	1,7
Comuni fino a 10.000 abitanti	21,8	14,0	8,9	5,0	3,2	2,0	1,6	1,3	1,1

5.7. Esempi di calcolo degli errori campionari

Esempio 1

Nel 2011/2012, l'indagine stima che sono 112.261 i cittadini stranieri di 6 anni e più che dichiarano di “aver subito eventi di discriminazione nel loro percorso di studi in Italia”.

Il calcolo dell'errore può essere effettuato direttamente tramite la funzione interpolante i cui parametri sono riportati nel prospetto 3, alla riga Italia, e sono i seguenti:

$$a = 9,513502$$

$$b = -1,29689$$

per Y stimato pari a 112.261 si ha:

$$e(\hat{Y}) = \sqrt{\exp(9,513502 - 1,29689 * \log(112,261))} = 0,05872$$

L'errore relativo è quindi pari a 5,872% mentre il calcolo dell'errore assoluto e dell'intervallo di confidenza errore assoluto viene effettuato nel modo seguente.

L'errore assoluto sarà:

$$\sigma(112.261) = 0,05872 * 112.261 = 6.591,966$$

L'intervallo di confidenza (per $P=0,95$ e $K=1,96$) avrà i seguenti estremi:

$$\text{Estremo inferiore: } 112.261 - (1,96 * 6.591,966) = 99.340,75$$

$$\text{Estremo superiore: } 112.261 + (1,96 * 6.591,966) = 125.181,3$$

Esempio 2

La valutazione dell'errore campionario di stime di frequenze relative percentuali può essere ottenuta mediante un metodo approssimato utilizzando i dati del prospetto 4.

Considerando, ad esempio, la stima pari al 39,2% riferita alla quota di cittadini stranieri residenti in Italia di 6 anni e più che dichiarano di non avere nessuna difficoltà con la lingua italiana, sulla base dei dati del prospetto 4 risulta che l'errore relativo di tale stima è pari a circa 0,9%. Nel prospetto 4, infatti, 40% è la frequenza relativa percentuale che a livello Italia più si avvicina al valore stimato del 39,2%, cui corrisponde un errore relativo pari a 0,9%

L'errore assoluto della frequenza relativa percentuale sarà:

$$\sigma(39,2) = 0,9/100 * 39,2 = 0,3528$$

L'intervallo di confidenza (per $P=0,95$ e $K=1,96$) avrà i seguenti estremi:

$$\text{Estremo inferiore: } 39,2 - (1,96 * 0,3528) = 39,2 - 0,6915 = 38,5 \text{ estremo inferiore della percentuale}$$

$$\text{Estremo superiore: } 39,2 + (1,96 * 0,3528) = 39,2 + 0,6915 = 39,9 \text{ estremo superiore della percentuale}$$

6. La diffusione dei risultati dell'indagine

I principali risultati dell'indagine vengono resi disponibili sul sito dell'Istat attraverso le statistiche report pubblicate su temi diversi. I risultati sono diffusi anche attraverso il datawarehouse I.Stat (<http://dati.istat.it/>) nella sezione Popolazione e famiglie→ Stranieri e immigrati.

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

I principali risultati dell'Indagine sono divulgati anche in volumi a carattere generale (ad esempio il Rapporto annuale). I volumi curati dall'Istat sono consultabili nel Catalogo editoriale.

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

7. Glossario

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

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

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

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

casalinga è chi si dedica prevalentemente alle faccende domestiche;

studente è chi si dedica prevalentemente allo studio;

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

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

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

- **dirigenti, imprenditori, liberi professionisti;**
- **direttivi, quadri, impiegati;**
- **capo operai, operai** (inclusi apprendisti, lavoratori a domicilio per conto di imprese);
- **lavoratori in proprio, coadiuvanti** (inclusi soci di cooperative di produzione di beni e/o prestazioni di servizio).

• **famiglia e nucleo familiare**

Si precisa che:

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

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

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

Nord-Ovest (Piemonte, Valle d'Aosta, Lombardia, Liguria);

Nord-Est (Trento, Bolzano, Veneto, Friuli Venezia Giulia, Emilia Romagna)

Centro (Toscana, Umbria, Marche, Lazio);

Sud-Ovest (Campania, Basilicata e Calabria)

Sud-Est (Abruzzo, Molise e Puglia);

Isole (Sicilia, Sardegna)

• **il tipo di comune:** i comuni italiani sono suddivisi nelle seguenti classi:

A1) comuni centro dell'area metropolitana: Torino, Milano, Venezia, Genova, Bologna, Firenze, Roma, Napoli, Bari, Palermo, Catania, Cagliari;

A2) comuni che gravitano intorno ai comuni centro dell'area metropolitana;

A3) comuni non appartenenti all'area metropolitana aventi fino a 10.000 abitanti;

A4) comuni non appartenenti all'area metropolitana con oltre 10.000 abitanti.

• **gli indici PCS MCS MHI**, diversi quesiti presenti nel questionario (SF12) consentono di costruire 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 (MH) e l'Indice di Vitalità (VT). 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>. 53

• **malattie croniche gravi:** l'indicatore è stato definito aggregando nella categoria 'gravi' dieci patologie (o gruppi) 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, e insufficienza renale cronica, rilevata solo per l'ultima edizione dell'indagine.

8. Riferimenti bibliografici

Deville J. C., Särndal C. E., (1992), "Calibration Estimators in Survey Sampling", Journal of the American Statistical Association, vol. 87, n.418, 1992.

9. Contatti

Unità operativa SDS/B "Integrazione sociale degli stranieri"

Monica Perez

Tel. +39 06 4673.7321

Email: perez@istat.it

Paola Muccitelli

Tel. +39 06 4673.7423

Email: muccitel@istat.it

Curatori dei capitoli

Il documento è a cura di Monica Perez

Si devono a:

Andrea Cutillo i paragrafi 5.1, 5.2, 5.3

Claudia De Vitiis, Nicoletta Cibella e Emanuela Scavalli i paragrafi 3.1, 3.2, 3.3

Paola Muccitelli i paragrafi 5.4, 5.6, 6 e 7

Monica Perez i paragrafi 1, 2, 4