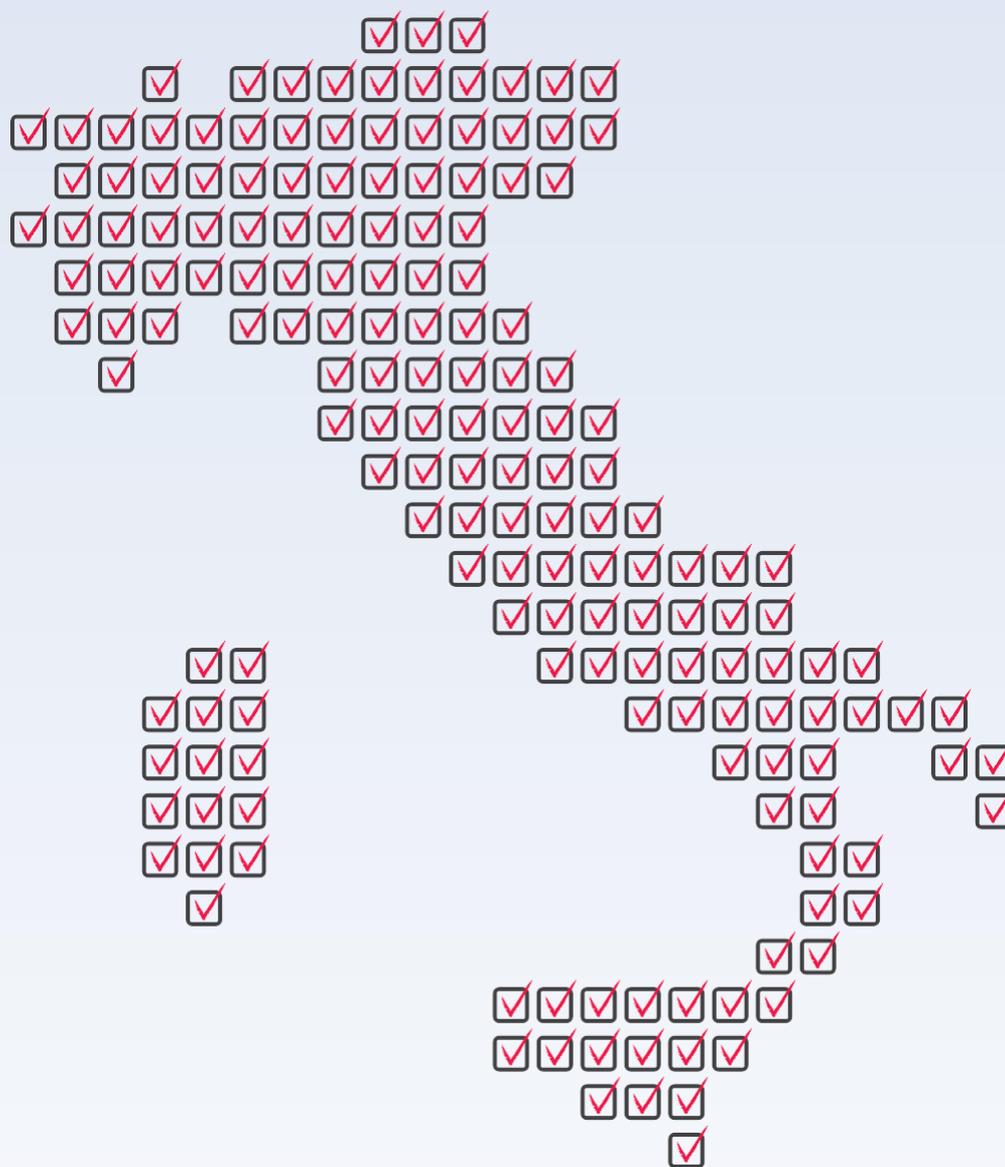


Atti del 15° Censimento generale della popolazione e delle abitazioni



2 I metodi campionari e le tecniche di rilevazione



15° CENSIMENTO GENERALE
DELLA POPOLAZIONE
E DELLE ABITAZIONI
2011

 Istat

Atti a cura di: Giuseppe Sindoni

Fascicolo a cura di: Giancarlo Carbonetti, Marco Fortini, Donatella Zindato

Atti del 15° Censimento generale della popolazione e delle abitazioni
2 - I METODI CAMPIONARI E LE TECNICHE DI RILEVAZIONE

© 2023

Istituto nazionale di statistica Via Cesare Balbo, 16 – Roma

ISBN 978-88-458-2127-1

Salvo diversa indicazione la riproduzione è libera,
a condizione che venga citata la fonte.

Immagini, loghi (compreso il logo dell'Istat),
marchi registrati e altri contenuti di proprietà di terzi
appartengono ai rispettivi proprietari e
non possono essere riprodotti senza il loro consenso.



INDICE

	pag.
Presentazione dell'opera	6
Il 15° Censimento generale della popolazione e delle abitazioni	8
Introduzione	12
Capitolo 1 – La strategia di rilevazione del Censimento 2011	14
1.1 Le innovazioni del Censimento 2011: dalla rilevazione “porta a porta” al censimento assistito da lista	14
1.2 Il censimento assistito da lista	16
1.3 Uso di nuove tecniche di rilevazione	18
1.4 Una strategia di rilevazione modulare	20
1.5 Strategia <i>short form/long form</i>	23
1.6 Uso di nuovi strumenti territoriali: la Rilevazione dei numeri civici (Rnc)	24
1.7 Il Sistema di gestione e monitoraggio della rilevazione (Sgr)	25
Capitolo 2 – Predisposizione delle liste (Lac, Lifa, Rnc)	28
2.1 Produzione e uso delle Liste anagrafiche comunali (Lac)	28
2.1.1 <i>Le sperimentazioni</i>	29
2.1.1.1 <i>Dal primo approccio alla rilevazione pilota nel 2009</i>	29
2.1.1.2 <i>La rilevazione sperimentale delle Lac nel 2010</i>	31
2.1.1.3 <i>La scelta dei Comuni campione</i>	31
2.1.2 <i>La prima acquisizione delle Lac per il Censimento 2011</i>	32
2.1.3 <i>La seconda acquisizione delle Lac e il calcolo delle variazioni anagrafiche</i>	34
2.2 Produzione e uso della Lista integrativa da fonti ausiliarie (Lifa)	36
2.2.1 <i>L'archivio dei Permessi di soggiorno (PS)</i>	37
2.2.2 <i>L'identificazione ed eliminazione dei record PS duplicati</i>	39
2.2.3 <i>L'integrazione dei record di stranieri con permesso di soggiorno con le Lac e la costruzione delle Lifa</i>	40
2.2.4 <i>L'utilizzo e il rendimento della Lifa nella fase di recupero della sotto-copertura delle Lac durante il Censimento 2011</i>	42
2.3. Produzione e uso della lista degli interni potenzialmente abitativi	43
Capitolo 3 – Disegno di campionamento e selezione del campione	45
3.1 Introduzione	45
3.2 La strategia di campionamento tramite <i>short/long form</i>	45
3.3 I Comuni e le unità eleggibili alla rilevazione campionaria	46
3.3.1 <i>I Comuni rilevati a campione tramite short/long form</i>	46
3.3.2 <i>I domini di campionamento: le aree di censimento</i>	47
3.3.3 <i>Le unità eleggibili al campionamento</i>	49
3.4 Le liste di campionamento	49
3.4.1 <i>Premessa</i>	49
3.4.2 <i>La geocodifica delle famiglie presenti nelle Liste anagrafiche comunali</i>	49
3.4.3 <i>Le Lac “ibride”</i>	51

3.4.4	<i>Mancata geocodifica ed errata geocodifica</i>	52
3.4.5	<i>Integrazione tra i dati delle Liste anagrafiche comunali e le aree di censimento</i>	53
3.4.6	<i>Lo sdoppiamento delle sezioni di censimento</i>	54
3.5	Le procedure di estrazione dei campioni e la validazione dei risultati	54
3.5.1	<i>Caricamento dei dati di input e di output</i>	54
3.5.2	<i>Preparazione della lista di campionamento</i>	55
3.5.3	<i>Estrazione dei campioni di famiglie</i>	55
3.5.4	<i>Controlli effettuati</i>	55
3.5.5	<i>Output prodotto</i>	55
3.6	Risultati	56
3.7	Considerazioni conclusive	57
Capitolo 4 – Le tecniche di rilevazione		59
4.1	Introduzione	59
4.2	La consegna dei questionari	59
4.3	La restituzione multicanale	61
4.3.1	<i>Il monitoraggio della restituzione multicanale tramite Sgr</i>	63
4.4	Il completamento della rilevazione sul campo	63
4.4.1	<i>Il recupero delle mancate risposte</i>	64
4.4.2	<i>Il recupero della sotto-copertura</i>	65
4.4.2.1	<i>Il recupero della sotto-copertura nei Comuni Rnc</i>	65
4.4.2.2	<i>Il recupero della sotto-copertura nei Comuni non Rnc</i>	66
4.5	Casi particolari della rilevazione della popolazione	66
4.5.1	<i>La rilevazione delle convivenze</i>	66
4.5.1.1	<i>La rilevazione delle convivenze in convenzione</i>	67
4.5.1.2	<i>La rilevazione delle convivenze non in convenzione</i>	67
4.5.2	<i>La rilevazione delle persone senza tetto e delle persone senza fissa dimora</i>	68
4.5.3	<i>La rilevazione delle minoranze linguistiche</i>	69
4.5.3.1	<i>La rilevazione censuaria nella Provincia autonoma di Trento</i>	69
4.5.3.2	<i>La rilevazione censuaria nella Provincia autonoma di Bolzano</i>	70
4.5.3.3	<i>La rilevazione censuaria nella Regione Friuli-Venezia Giulia</i>	72
4.5.4	<i>La rilevazione nei Comuni del “cratere sismico”</i>	73
Capitolo 5 – I pesi di riporto all’universo per la stima delle frequenze		76
5.1	Introduzione	76
5.2	Lo stimatore di ponderazione vincolata	76
5.2.1	<i>Aspetti metodologici</i>	76
5.2.2	<i>La riponderazione</i>	77
5.2.3	<i>Alcune proprietà dello stimatore di ponderazione vincolata</i>	78
5.2.4	<i>Criteri per il disegno di un sistema di vincoli in ambito censuario</i>	78
5.3	Analisi preliminari per la definizione della strategia di calcolo dei pesi calibrati	80
5.3.1	<i>Possibili approcci al disegno del sistema dei vincoli</i>	80
5.3.1.1	<i>Approccio A - Sistema di pesi unico</i>	80
5.3.1.2	<i>Approccio B - Sistema di pesi a blocchi</i>	82
5.3.1.3	<i>Approccio C - Sistema di pesi primario e secondario</i>	85
5.3.2	<i>Test sulle procedure di calibrazione</i>	85
5.3.2.1	<i>Riponderazione multipla e procedura bottom-up</i>	86
5.3.2.2	<i>Riponderazione unica</i>	87
5.4	Il disegno definitivo del sistema dei vincoli	88

5.4.1 Oggetto “popolazione”	88
5.4.2 Oggetto “famiglie”	95
5.4.3 Oggetto “alloggi ed edifici”	96
5.5 Il calcolo dei pesi di riporto all’universo	100
5.5.1 Introduzione	100
5.5.2 La preparazione dei dati di input	100
5.5.3 Domini di riponderazione	102
5.5.3.1 Oggetto “popolazione”	102
5.5.3.2 Oggetto “famiglie”	103
5.5.3.3 Oggetto “alloggi ed edifici”	104
5.5.4 La coerenza delle stime calibrate	104
5.5.5 Le soluzioni informatiche	106
5.5.6 I risultati del calcolo dei pesi finali	107
5.6 Considerazioni conclusive	111
Capitolo 6 – Errori di campionamento	113
6.1 Introduzione	113
6.2 Gli errori campionari	113
6.3 Definizione dei modelli di rappresentazione sintetica degli errori campionari	115
6.3.1 La strategia di costruzione dei modelli di errore	116
6.3.2 Alcuni risultati	117
6.3.3 Il calcolo degli errori di campionamento per il <i>Quality report</i> di Eurostat	118
Conclusioni	119
Riferimenti bibliografici	120
Allegato A – Elenco dei Comuni sottoposti a campionamento al 15° Censimento generale della popolazione e delle abitazioni	123
Allegato B – Modelli sintetici di rappresentazione degli errori campionari	134

Presentazione dell'opera

Gli atti del 15° Censimento generale della popolazione e delle abitazioni, data la complessità e l'importanza dell'argomento, sono articolati in sei volumi che documentano l'operazione censuaria attraverso la normativa di riferimento e le principali soluzioni organizzative, metodologiche e tecniche. Ciascun volume è autonomo nella trattazione tematica e affronta in maniera dettagliata ed esaustiva il proprio argomento specifico, ma restano evidenti le interconnessioni delle fasi e dei processi di progettazione e produzione. La collana è corredata da un insieme di documenti di riferimento, disponibili sul sito web dell'Istat¹.

I sei volumi che formano la collana sono i seguenti:

1. *Le norme, l'organizzazione e la progettazione finanziaria*
2. *I metodi campionari e le tecniche di rilevazione*
3. *Il processo di controllo, correzione e validazione dei dati*
4. *Gli strumenti di informazione territoriale*
5. *I sistemi informatici a supporto delle operazioni censuarie*
6. *La valutazione della qualità*

Il primo volume *Le norme, l'organizzazione e la progettazione finanziaria* contiene le norme, sia internazionali che nazionali, sulla base delle quali si è progettato e realizzato il Censimento. Vengono illustrati e descritti nei loro aspetti salienti il Piano generale di Censimento, il quadro finanziario in termini di progettazione e costi, l'organizzazione della rete di rilevazione, gli interventi di formazione e gli strumenti di comunicazione operativa adottati, i processi di gestione dei servizi in appalto, ed infine le strategie di comunicazione con i cittadini.

Il secondo volume *I metodi campionari e le tecniche di rilevazione* partendo da una trattazione generale della strategia di rilevazione assistita da liste amministrative, descrive nel dettaglio il processo di preparazione delle liste anagrafiche di base e delle liste ausiliarie per il recupero della sotto-copertura anagrafica. Vengono illustrati il disegno di campionamento per la selezione delle famiglie che hanno ricevuto il questionario in forma ridotta, gli aspetti salienti delle tecniche di rilevazione adottate e i processi di calcolo dei pesi di riporto all'universo per le variabili campionarie e di stima degli errori.

Il terzo volume *Il processo di controllo, correzione e validazione dei dati* descrive i contenuti informativi acquisiti tramite i questionari e le unità che sono state oggetto della rilevazione, i processi di controllo, correzione e validazione che hanno portato alla produzione degli aggregati di diffusione, i processi messi in atto per la pubblicazione dei primi risultati, dei dati di popolazione legale e sua struttura demografica e dei dati socio-economici, secondo il Piano di diffusione nazionale e quello del Regolamento europeo.

Il quarto volume *Gli strumenti di informazione territoriale* descrive l'insieme di informazioni territoriali prodotte prima, durante e dopo le operazioni censuarie e finalizzate sia all'organizzazione delle operazioni sul campo che alla produzione dell'informazione resa al paese. Vengono illustrati il processo di aggiornamento delle basi territoriali, i metodi di definizione delle aree di censimento per la strategia di campionamento, la preparazione degli archivi di numeri civici utilizzati a supporto delle operazioni censuarie, la Rilevazione dei numeri civici, che ha portato alla creazione di un archivio di indirizzi ed edifici nelle città maggiori e il censimento degli edifici e delle abitazioni.

Il quinto volume *I sistemi informatici a supporto delle operazioni censuarie* descrive l'infrastruttura tecnico-informatica a supporto della raccolta dei dati e del monitoraggio delle operazioni sul campo. Vengono descritte le applicazioni web attraverso le quali i Comuni hanno potuto trasmettere all'Istat le liste anagrafiche e quelle per effettuare la Rilevazione dei numeri civici, il "sistema integrato" di acquisizione dei dati, gestione della rilevazione e *repository* della documentazione, i sistemi per l'acquisizione e la verifica dei dati

¹ <http://www.istat.it/it/archivio/189483>

provenienti dalla lettura ottica, il sistema per la gestione della validazione finale dei dati, il sistema di rendicontazione delle spese censuarie e i sistemi di diffusione dei risultati.

Il sesto volume *La valutazione della qualità* descrive la *Post enumeration survey* (Pes) ossia la rilevazione campionaria areale volta a stimare gli indicatori di qualità del censimento richiesti dal regolamento europeo. La rilevazione post censuaria, infatti, ha avuto il principale obiettivo di stimare il numero di individui abitualmente dimoranti nel periodo di riferimento, il tasso di sotto-copertura e il tasso di sovra-copertura del Censimento. Il volume descrive gli aspetti organizzativi, metodologici e i risultati della rilevazione campionaria.

Il 15° Censimento generale della popolazione e delle abitazioni

A cura di Andrea Mancini

già Direttore del Dipartimento per i censimenti e gli archivi amministrativi e statistici dell'Istat

Per la prima volta, nel 2011, tutti gli stati membri dell'Unione europea hanno condotto il censimento della popolazione seguendo le regole comuni stabilite da un nuovo specifico Regolamento del Consiglio e del Parlamento². Questo faceva proprie le nuove Raccomandazioni della Conferenza degli statistici europei (Ces)³, ma introduceva anche vincoli cogenti nei tempi di esecuzione, di riferimento dei dati e di trasmissione dei risultati. Ulteriori norme comuni venivano stabilite dai regolamenti di attuazione, emanati dalla Commissione europea⁴, soprattutto in materia di maggior dettaglio delle classificazioni, definizioni e nomenclature che i paesi membri erano chiamati a rispettare.

Destinato a svolgersi in occasione del 150° anniversario dell'Unità d'Italia, il 15° Censimento generale della popolazione e delle abitazioni è stato progettato dall'Istat con il supporto di un Comitato consultivo⁵ costituito da Luigi Biggeri, allora Presidente dell'Istat, e composto da rappresentanti dei Ministeri interessati, del Cisis-Regioni, dell'Upi, dell'Anci, dell'Usci e dell'Anpci, delle associazioni professionali degli ufficiali di anagrafe (Anusca e DeA), da esperti nominati dal Presidente dell'Istat, oltre che da molti colleghi delle varie direzioni dell'Istituto. Il censimento è stato progettato in modo da rispettare le regole stabilite dai Regolamenti dell'Unione e in particolare di riuscire a diffondere i risultati con maggiore tempestività delle precedenti tornate e comunque entro i prescritti ventisette mesi dalla fine del 2011, ma anche di ottenere elevati standard di qualità dei dati, di mantenere la confrontabilità dei dati rispetto alle precedenti edizioni censuarie, di contenere al massimo i costi di produzione e il fastidio statistico delle famiglie.

Per cogliere questi obiettivi, parzialmente contrastanti, l'Istat ha deciso di effettuare il 15° Censimento introducendo numerose e rilevanti innovazioni nei metodi, nelle tecniche e nell'organizzazione della rilevazione. Le novità sono descritte nei vari fascicoli che compongono gli "Atti del censimento", che, come tradizione dell'Istituto, intendono garantire la memoria storica della rilevazione insieme alle informazioni che ne rappresentano il prodotto.

Sotto il profilo metodologico, il censimento del 2011 ha conservato l'enumerazione completa e simultanea mediante rilevazione di campo, ma ha utilizzato, per la prima volta in Italia, una lista di partenza delle unità di rilevazione personalizzata in base ai dati nominativi delle Liste anagrafiche comunali (Lac) di tutti i Comuni italiani, riferite al 31 dicembre 2010 e acquisite mediante apposita applicazione elettronica gestita sul web. Dunque si è trattato di una rilevazione attuata sul campo, ma largamente supportata fin dall'inizio dall'uso di dati amministrativi. Ciò ha consentito di introdurre due ulteriori innovazioni di rilievo, costituite dalla spedizione postale di circa 25 milioni di plichi contenenti i questionari di famiglia, indirizzati all'intestatario del foglio di famiglia sulla base dei dati anagrafici, nonché dalla predisposizione di un complesso Sistema di gestione della rilevazione (Sgr) completamente elettronico, in grado di controllare in tempo reale l'andamento della rilevazione su tutto il territorio nazionale con riferimento a ciascuna singola unità di rilevazione e a ciascun singolo operatore di censimento. Questo sistema, completamente sviluppato in Istat, ha comportato ulteriori opportunità di innovazione sul piano della tecnica di rilevazione, che infatti è divenuta "multicanale", dando la possibilità a ciascuna famiglia di compilare il questionario in forma elettronica⁶ o cartacea⁷. In questo modo, rispetto al censimento del 2001, il numero dei rilevatori è diminuito di circa il 40 per cento, essendo il

² Regolamento (Ce) n. 763/2008 del Parlamento europeo e del Consiglio del 9 luglio 2008 relativo ai censimenti della popolazione e delle abitazioni.

³ Conference of European statisticians, recommendations for the 2010 Censuses of population and housing.

⁴ Commission regulation (Ec) n. 1201/2009 of 30 November 2009 implementing Regulation (Ec) n. 763/2008 of the European Parliament and of the Council on population and housing censuses as regards the technical specifications of the topics and of their breakdowns; Commission regulation (Eu) n. 519/2010 of 16 June 2010 adopting the programme of the statistical data and of the metadata for population and housing censuses provided for by Regulation (Ec) n. 763/2008 of the European Parliament and of the Council; Commission regulation (Eu) n. 1151/2010 of 8 December 2010 implementing Regulation (Ec) No 763/2008 of the European Parliament and of the Council on population and housing censuses, as regards the modalities and structure of the quality reports and the technical format for data transmission.

⁵ Comitato consultivo per la preparazione a livello comunale del 15° Censimento generale della popolazione e delle abitazioni di cui alla Deliberazione del Presidente dell'Istat n. 673/PER del 6 giugno 2007.

⁶ Con restituzione via web e rilascio automatico della ricevuta.

⁷ Con restituzione ad uno qualsiasi degli Uffici postali sparsi sul territorio oppure ad uno dei centri comunali di raccolta predisposti da ciascun Ufficio comunale di censimento (Ucc).

loro compito limitato al recupero dei questionari delle famiglie non ancora rispondenti, alla ricerca di persone e famiglie non ancora iscritte nell'anagrafe del Comune alla data di riferimento del censimento (9 ottobre 2011) e alla enumerazione delle abitazioni non occupate. I risultati sono stati positivi sotto i profili della tempestività, della qualità dei dati raccolti e del risparmio di spesa, in conseguenza di tassi di risposta del 22,6 per cento via Uffici postali, del 31,7 per cento presso i centri comunali di raccolta, del 12,3 per cento attraverso i rilevatori e del 33,4 per cento via web. Circa 8,5 milioni di questionari sono stati raccolti attraverso l'applicazione *online*. La cifra assoluta di questionari compilati *online* in Italia è la seconda più alta a livello mondiale nella tornata censuaria 2010. Nei primi due mesi di operazioni il carico medio su Sgr è stato di circa 115 questionari completati al minuto, con un picco di 500 nei periodi di massimo carico.

Inoltre Sgr ha consentito a ciascun Ufficio comunale di censimento di effettuare le operazioni di confronto tra censimento e anagrafe in modo del tutto simultaneo alla rilevazione in atto. Ciò ha prodotto tre rilevanti risultati innovativi. In primo luogo, ha dato modo all'Istat di conoscere in tempo reale i risultati del confronto a livello di singolo nominativo nel rispetto della privacy, grazie al fatto che i risultati nominativi del confronto individuale effettuato dall'Ufficio comunale di censimento venivano memorizzati all'interno di Sgr stesso. In secondo luogo, ha consentito all'Istat di usare i risultati del confronto per determinare con grande tempestività i dati definitivi di popolazione legale, distinti per sesso, anno di nascita e cittadinanza (italiana e straniera). Il 18 ottobre 2012, a circa un anno dalla data di riferimento del censimento e a cinque mesi dalla fine della rilevazione, Enrico Giovannini, allora Presidente dell'Istat, ha potuto inviare al governo questi dati per la pubblicazione in Gazzetta ufficiale, avvenuta poi il 18 dicembre 2012, contemporaneamente alla diffusione sul sito dell'Istituto delle prime tavole di dati definitivi. In terzo luogo, ha permesso di fornire tempestivamente ai servizi demografici di ciascun Comune le liste nominative degli irreperibili al censimento e dei censiti non iscritti in anagrafe per la verifica post censuaria dei registri locali di popolazione, così come prescritto dal regolamento anagrafico nazionale.

Di nuovo sotto il profilo del metodo statistico, il censimento del 2011 si è anche caratterizzato per l'uso di campioni di famiglie residenti in aree di censimento urbane alle quali somministrare un questionario in forma breve (35 quesiti in luogo degli 84 contenuti nel questionario in forma lunga). Dunque, una parte non marginale dei risultanti dati è costituita da stime statistiche che, tuttavia, hanno rispettato errori campionari molto contenuti⁸. Anche questa novità è stata resa possibile dall'uso iniziale delle Lac e da Sgr. contribuendo al contenimento dei costi e del fastidio statistico di circa otto milioni di famiglie, alle quali è pervenuto il questionario in forma breve.

Molte delle innovazioni brevemente ricordate sono state oggetto di sperimentazioni, anche ripetute, condotte tra il 2008 e il 2010 dall'Istat in comune con le amministrazioni locali che avrebbero poi agito come organi di censimento e segnatamente con gli uffici di statistica e i servizi demografici dei Comuni. Di esse non si dà conto nei fascicoli degli atti del censimento, ma un'ampia documentazione è comunque pubblicata sul sito dell'Istituto⁹, in particolare con riferimento alla rilevazione pilota, che si è svolta alla fine del 2009 coinvolgendo un campione di 78 mila famiglie di 31 Comuni, e all'acquisizione sperimentale delle liste anagrafiche, che si è tenuta nel 2010 in 1.855 Comuni. Di fatto queste sperimentazioni, insieme con le attività del progetto di informazione e formazione ai Comuni sulle innovazioni censuarie¹⁰, hanno svolto un ruolo assai rilevante nel preparare la rete di rilevazione degli organi territoriali di censimento ben prima dell'inizio delle operazioni e in un clima di diffusa cooperazione e condivisione con le strutture centrali del Sistema statistico nazionale che sicuramente ha poi favorito l'efficacia della rilevazione e la capacità di affrontare e gestire le emergenze in corso d'opera.

Peraltro le innovazioni introdotte con il 15° Censimento sono state rese possibili anche dalla tempestiva rideterminazione delle basi territoriali o, più propriamente, dal complesso degli adempimenti precensuari in materia ecografica, di onomastica stradale e numerazione civica previsti dalla legge anagrafica (L. 24 dicembre 1954, n. 1228) e dal suo regolamento di attuazione (D.P.R. 30 maggio 1989 n. 223). Nel caso del censimento del 2011 questi adempimenti hanno assunto un precipuo valore propedeutico alle progettate innovazioni di metodo e tecniche della rilevazione, poiché alla rideterminazione delle basi territoriali è stata per la prima volta affiancata nei centri abitati dei Comuni con più di 20 mila abitanti una rilevazione esaustiva dei numeri civici

⁸ Dell'entità degli errori si parla diffusamente nel volume 2.

⁹ <https://www.istat.it/it/censimenti-permanenti/censimenti-precedenti/popolazione-e-abitazioni/popolazione-2011>

¹⁰ Progetto INFO/FORCENS http://www.istat.it/it/files/2011/04/doc_1_2010.pdf

con l'obiettivo di verificare territorialmente la presenza di numeri civici, edifici e relative unità abitative e di riconciliare su mappa le relative informazioni. Nel quarto fascicolo degli Atti sono documentate le attività di rideterminazione delle basi territoriali, di rilevazione dei numeri civici e di loro georiferimento alle sezioni di censimento, queste due ultime anche con riferimento alle operazioni avvenute durante il censimento in tutti i Comuni con meno di 20 mila abitanti. Di fatto si è trattato di un investimento pubblico in una infrastruttura informativa nazionale che, anche grazie all'intensa collaborazione con l'Agenzia del territorio poi confluita nell'Agenzia delle entrate, ha consentito all'Istat di realizzare, come ricaduta del censimento e con effetti duraturi nel tempo, l'Archivio nazionale di numeri civici e strade urbane.

Condurre un censimento basato su tecniche di distribuzione postale, con scelta autonoma dei rispondenti in merito alla modalità di restituzione del questionario e con necessità di gestire servizi *online* 24 ore su 24 per un periodo di tempo non breve, ha comportato la necessità di predisporre un'organizzazione più articolata rispetto a quella adottata nelle precedenti tornate censuarie. Anche nel 2011 l'organizzazione si è fondata sui due tradizionali pilastri della rete territoriale di rilevazione da un lato e dei servizi censuari in appalto dall'altro lato, ma le interconnessioni tra i due pilastri sono state di necessità assai più delicate e complesse. La rete territoriale ha continuato a essere incentrata sugli Uffici comunali di censimento, ma questi hanno dovuto rafforzare tutte le funzioni di *back office*, cioè quelle deputate a usare le numerose funzionalità di Sgr e a effettuare il confronto tra censimento e anagrafe in tempo reale, in cambio di un significativo contenimento delle funzioni tradizionali dei rilevatori. Per contro i servizi censuari in appalto hanno assunto dimensioni e complessità assai superiori a quelle delle volte precedenti. Ciò ha comportato uno sforzo assai rilevante da parte dell'amministrazione dell'Istat e segnatamente del Servizio per il coordinamento del supporto amministrativo ai censimenti (Sac), appositamente creato per specificare le caratteristiche dei servizi in appalto e per controllare la corretta esecuzione di tutti i contratti stipulati con società esterne a seguito di gare pubbliche. I principali hanno riguardato il servizio di web hosting per la restituzione via internet dei questionari compilati e per la gestione di Sgr, il servizio di distribuzione postale alle famiglie iscritte in Lac dei plichi contenenti i questionari censuari e di loro raccolta presso gli Uffici postali una volta compilati, il servizio di stampa dei questionari e di registrazione in lettura ottica e tradizionale e di gestione delle immagini, il servizio di sportello telefonico per i rispondenti e per gli organi di censimento della rete territoriale. Il secondo e il quinto fascicolo degli Atti danno conto delle caratteristiche organizzative di questi servizi e delle numerose interconnessioni operative tra ciascuno di essi e con le altre funzioni svolte direttamente dall'Istat o affidate agli organi della rete territoriale nelle varie fasi di conduzione del censimento. Si vuole qui sottolineare come i risultati ottenuti dal censimento siano largamente dipesi dall'efficacia dell'azione svolta dalla direzione amministrativa dell'Istat e dal Servizio Sac.

Il sesto fascicolo degli Atti illustra in dettaglio i metodi, le tecniche e l'organizzazione della rilevazione campionaria areale di controllo della copertura del Censimento (Pes), condotta da un gruppo di lavoro del tutto autonomo dal Dipartimento dell'Istat responsabile del Censimento. I risultati testimoniano della qualità dei risultati ottenuti dal censimento, con un tasso di sotto-copertura a livello nazionale pari all'1,45 per cento della popolazione censita e un tasso di sovra-copertura pari allo 0,38 per cento. Per la popolazione di cittadinanza italiana il primo tasso si riduce allo 0,66 per cento, mentre sale all'11,23 per cento per la popolazione straniera a ulteriore conferma della difficoltà di enumerare questo segmento di popolazione residente anche a causa della sua maggiore mobilità sul territorio. Peraltro le innovazioni di metodo, tecnica e organizzazione hanno consentito di soddisfare gli obiettivi di maggiore tempestività richiesti dal Regolamento europeo e stabiliti dal Piano generale di Censimento. Le operazioni di censimento sul campo si sono concluse a febbraio del 2012 in tutti i Comuni, ad eccezione dei nove ai quali l'Istat aveva concesso la possibilità di registrare in proprio i questionari compilati in forma cartacea. Entro aprile 2012 si sono concluse le operazioni di confronto tra dati censuari e dati delle anagrafi. Entro ottobre 2012 erano pronti i dati definitivi della struttura demografica della popolazione residente per singolo Comune, sesso, anno di nascita e cittadinanza (italiana e straniera), poi pubblicati il 18 dicembre 2012 nel Dpcm relativo alla popolazione cosiddetta legale e diffusi in pari data sul sito dell'Istat. La trasmissione ad Eurostat di tutti gli ipercubi di dati per il Census Hub creato da Eurostat è stata completata entro la scadenza regolamentare della fine di marzo 2014, con aggiornamento di alcune celle a settembre dello stesso anno, come consentito dal Regolamento della Commissione.

È probabile che il 15° Censimento sia stato di transizione tra quelli precedenti, condotti secondo il metodo tradizionale dell'enumerazione mediante rilevazione generale sul campo, e quelli futuri, conducibili sulla base di sistemi di microdati integrati desunti da una pluralità di fonti amministrative. Nel 2011 l'uso delle Liste

anagrafiche comunali ha segnato una tappa decisiva per realizzare un registro statistico della popolazione residente, così come la sistematica acquisizione di dati sui numeri civici ha consentito di impiantare un archivio generalizzato tramite il quale georeferenziare qualsiasi informazione di fonte amministrativa o statistica dotata di indirizzo. Nel futuro sarà dunque possibile fondare il censimento italiano sull'uso dei dati amministrativi, così come accade già da tempo in altri paesi europei. La rilevazione sul campo potrà allora assumere un disegno campionario funzionale da un lato alla correzione degli errori di copertura delle fonti anagrafiche e dall'altro all'acquisizione delle variabili censuarie non disponibili nelle fonti amministrative. Grazie alla conseguente riduzione dei costi, sarà anche possibile aumentare la frequenza dell'offerta informativa statistica di dati riferiti a livelli territoriali locali e aggiornare annualmente le informazioni censuarie finora prodotte una volta ogni decennio. Questo è l'obiettivo di lungo periodo che ha influenzato le scelte dell'Istat in materia di metodi e tecniche per il censimento del 2011.

Introduzione¹¹

In Italia, come in molti paesi, i censimenti hanno sempre rappresentato un'occasione di innovazione e di forte evoluzione tecnologica. Il Censimento della popolazione e delle abitazioni del 2011 è stato fondamentale ai fini del passaggio dal Censimento tradizionale, basato sulla rilevazione “porta a porta”, esaustiva e simultanea, condotta su tutto il territorio nazionale, al Censimento permanente, basato sull'integrazione di dati di fonte amministrativa e dati rilevati sul campo.

Pur nell'ambito di una metodologia ancora sostanzialmente convenzionale (rilevazione sul campo esaustiva e simultanea), il censimento del 2011 si è basato fortemente sull'uso delle fonti amministrative. In particolare, nel 2011 le fonti amministrative sono state utilizzate per guidare la rilevazione (le Liste anagrafiche comunali per la spedizione postale dei questionari alle famiglie iscritte in anagrafe, e la Lista integrativa da fonti ausiliarie, per rilevare la popolazione abitualmente dimorante non ancora iscritta in anagrafe).

La progettazione del censimento 2011 ha tenuto conto di diversi fattori (Benassi *et al.*, 2014), primo fra tutti la necessità di coniugare la disponibilità di dati dettagliati a livello territoriale fine alla tempestività nella diffusione degli stessi, al fine di garantirne la rilevanza e rispettare i tempi di diffusione previsti dal Regolamento dell'Unione europea sui censimenti della popolazione e delle abitazioni (Regolamento (Ce) n. 763/2008 del Parlamento europeo e del Consiglio del 9 luglio 2008 relativo ai censimenti della popolazione e delle abitazioni).

Inoltre, la strategia tradizionale si inseriva in un contesto ambientale di crescente difficoltà. Il mutamento negli stili di vita e nella struttura della popolazione, soprattutto nei grandi Comuni, rendeva la rilevazione “porta a porta” sempre meno compatibile con il rispetto dei tempi di rilevazione necessari per una diffusione tempestiva dei risultati. Un altro elemento di criticità era rappresentato dal crescente carico statistico sui rispondenti.

La non sostenibilità della strategia tradizionale, sia in termini finanziari che organizzativi, era emersa anche dall'analisi dei dati del sistema di monitoraggio del censimento 2001. In particolare, erano emersi forti ritardi nella fase di rilevazione (solo il 2 per cento dei Comuni aveva concluso la rilevazione nei tempi stabiliti dal piano di rilevazione), difficoltà legate alla gestione dell'enorme rete di rilevazione richiesta dalla rilevazione esaustiva “porta a porta” (quali la difficoltà di trovare risorse adeguatamente qualificate e l'elevato *turnover* delle risorse sul campo), all'elevato numero di attori (a partire dai circa 8 mila Uffici comunali di censimento a cui era affidata la rilevazione sul campo) e all'elevata variabilità degli stessi in quanto a dimensioni, capacità e risorse. A conferma di ciò, l'impatto organizzativo delle operazioni di censimento sui Comuni era risultato fortemente dipendente dalla sua dimensione demografica (Fortini *et al.*, 2007).

A partire dall'analisi delle criticità della strategia tradizionale, e dallo studio delle esperienze estere (Abbatini *et al.*, 2007), è stata quindi progettata una strategia di censimento fortemente innovativa, basata su un approccio modulare e flessibile e su un uso più efficace dei dati delle anagrafi comunali.

Gli obiettivi principali della nuova strategia censuaria erano:

- la riduzione del carico di lavoro degli Uffici comunali di censimento attraverso l'inversione del rapporto tra *front office* (rilevatori sul campo) e *back office* comunale;
- la diversificazione delle tecniche di compilazione/restituzione del questionario per garantire ai rispondenti la massima flessibilità e al contempo incrementare il tasso di restituzione spontanea (riducendo così il ricorso ai rilevatori);
- la riduzione del carico statistico sui rispondenti attraverso la rilevazione su base campionaria di una parte delle informazioni censuarie.

Questi obiettivi sono stati raggiunti attraverso una strategia basata sull'uso dei dati delle anagrafi comunali per guidare la rilevazione (che ha trasformato la rilevazione “porta a porta” in un censimento “assistito da lista”) e sull'adozione di soluzioni differenziate in base all'ampiezza demografica dei Comuni.

Inoltre, al fine di ridurre il carico statistico complessivo sulle famiglie e produrre un risparmio sui costi della rilevazione, è stata adottata una strategia *short form/long form* per la rilevazione su base campionaria delle principali variabili socio-economiche.

¹¹ L'introduzione è a cura di Donatella Zindato.

In particolare, l'utilizzo delle 8.092 Liste anagrafiche comunali (Lac) come liste di famiglie (e di indirizzi) ha reso possibile la spedizione postale dei questionari (*mail-out*) e l'adozione di un sistema di restituzione multicanale, consentendo così una drastica riduzione del numero degli operatori sul campo (65 mila rilevatori contro i 100 mila del 2001).

Scegliendo la modalità che più soddisfaceva le proprie esigenze, le famiglie potevano optare per la compilazione del questionario *online* oppure compilare il questionario cartaceo e restituirlo presso uno dei 14 mila Uffici postali sul territorio italiano o presso uno dei Centri comunali di raccolta appositamente istituiti sul territorio comunale. Un quarto canale di restituzione era rappresentato dai rilevatori che, al termine della prima fase della rilevazione (la cosiddetta fase di "restituzione spontanea"), venivano inviati sul campo per il recupero mirato delle mancate risposte.

L'adozione di un sistema di restituzione multicanale ha richiesto lo sviluppo di un sistema *online* di gestione della rilevazione, che consentisse di gestire in contemporanea le diverse opzioni per la restituzione del questionario. Il Sistema di gestione della rilevazione (Sgr), alimentato con i dati relativi a oltre 25 milioni di famiglie provenienti dalle Liste anagrafiche comunali e a quasi 9 milioni di individui potenzialmente presenti sul territorio ma non ancora iscritti in anagrafe, ha consentito alla rete territoriale l'esecuzione di tutte le operazioni previste dal Piano generale di Censimento, rappresentando di fatto la spina dorsale dell'intero processo di rilevazione.

Questa modalità operativa ha prodotto benefici in termini di tempestività, costi e capacità di controllo centralizzato della rete censuaria. La tracciabilità di qualsiasi operazione effettuata sul sistema ha permesso all'Istat, così come agli Uffici regionali e provinciali di censimento, di avere costantemente un quadro completo relativo all'andamento delle operazioni e allo stato di avanzamento delle varie attività presso ciascun Ufficio comunale di censimento. Nel contempo, ha reso gli operatori della rete comunale indipendenti nella gestione delle operazioni loro affidate.

Il volume "I metodi campionari e le tecniche di rilevazione", secondo fascicolo degli *Atti del 15° Censimento generale della popolazione e delle abitazioni*, descrive in dettaglio la strategia censuaria del 2011 e alcuni processi di importanza fondamentale utili alla sua implementazione. In particolare, il capitolo 1 ripercorre il processo di progettazione e test e presenta le innovazioni di metodi e tecniche che hanno caratterizzato la nuova strategia censuaria; il capitolo 2 descrive i processi di predisposizione delle liste utilizzate per guidare la rilevazione e per il recupero della sotto-copertura anagrafica; il capitolo 3 illustra la strategia di rilevazione *short form/long form* e il disegno di campionamento per la selezione, nei Comuni di maggiori dimensioni, di campioni di famiglie a cui somministrare il questionario in forma estesa (*long form*); il capitolo 4 descrive le tecniche e le fasi della rilevazione e le strategie *ad hoc* predisposte per la rilevazione di particolari popolazioni (le persone in convivenza, le minoranze linguistiche, le persone senza tetto e senza fissa dimora, i residenti nei Comuni del "cratere sismico"); il capitolo 5 documenta i processi di calcolo dei pesi di riporto all'universo per le variabili rilevate su base campionaria; infine, il capitolo 6 illustra il processo di calcolo degli errori campionari associati alle stime e i relativi modelli di errore.

Insieme agli altri fascicoli della serie dedicata al 15° Censimento generale della popolazione e delle abitazioni, il volume fornisce un contributo prezioso alla conoscenza del processo di progettazione e realizzazione del Censimento 2011.

CAPITOLO 1

LA STRATEGIA DI RILEVAZIONE DEL CENSIMENTO 2011¹²

1.1 Le innovazioni del Censimento 2011: dalla rilevazione “porta a porta” al censimento assistito da lista

Gli approcci adottati per condurre i censimenti della popolazione e delle abitazioni variano da paese a paese e si evolvono nel tempo, grazie all'introduzione di metodologie e tecniche di rilevazione innovative. Anche in Italia il Censimento della popolazione ha sempre rappresentato un'occasione di innovazione e di forte evoluzione tecnologica. Dal punto di vista della strategia di rilevazione, tuttavia, è stato caratterizzato fino al 2001, dalla rilevazione esaustiva sul campo e dall'individuazione dei rispondenti per mezzo della tecnica “porta a porta”. Il rilevatore percorreva a tappeto il territorio della sezione di censimento, esplorando tutti gli accessi esterni individuati lungo il percorso senza l'ausilio di liste precompilate. Una volta individuate le unità di rilevazione (ovvero le persone dimoranti abitualmente presso gli alloggi individuati lungo il percorso), il rilevatore consegnava il questionario cartaceo che, auto-compilato dai rispondenti, veniva successivamente ritirato dallo stesso rilevatore.

All'Istat spettavano, oltre alla progettazione e al coordinamento delle attività di rilevazione, le attività di elaborazione e diffusione dei dati; mentre la rilevazione sul campo era affidata ai Comuni (incluso il reclutamento e la formazione dei rilevatori), che effettuavano anche la revisione dei questionari compilati, il confronto censimento-anagrafe e, a rilevazione conclusa, la revisione anagrafica. La complessa e onerosa macchina organizzativa era costituita da 103 Uffici provinciali di censimento (Upc), più di 8 mila Uffici comunali di censimento (Ucc), 10 mila coordinatori e 100 mila rilevatori (questi i numeri del censimento 2001), ai quali era affidato il compito di rilevare 25 milioni di famiglie e abitazioni.

Al momento della progettazione del censimento del 2011, era pressante l'esigenza di sviluppare una strategia alternativa alla rilevazione “porta a porta”, che garantisse una diffusione più tempestiva dei risultati¹³ e che tenesse conto dei cambiamenti verificatisi nella società italiana nel decennio intercensuario, e in particolare dei mutamenti negli stili di vita e nella struttura della popolazione¹⁴, che rendevano sempre più problematica la consegna e il ritiro dei questionari da parte del rilevatore, soprattutto nei grandi Comuni.

A partire dall'analisi del contesto e delle criticità della strategia tradizionale, e sulla base dello studio delle esperienze estere (Abbatini *et al.*, 2007), è stata quindi elaborata una strategia di censimento fortemente innovativa, basata sull'uso dei dati delle anagrafi comunali per guidare la rilevazione, sulla flessibilità delle tecniche di rilevazione e sull'adozione di soluzioni differenziate in base all'ampiezza demografica dei Comuni.

La non sostenibilità della strategia tradizionale era evidenziata innanzitutto dall'analisi dei dati del sistema di monitoraggio del censimento 2001. Emergevano tra i principali punti critici dell'organizzazione, oltre alla difficoltà di rispettare le date previste dal piano di rilevazione¹⁵, il grosso sforzo organizzativo ed economico richiesto ai Comuni¹⁶; la difficoltà di trovare risorse umane qualificate e di formarle adeguatamente, anche a fronte dell'alto *turnover*; l'elevato numero di attori, caratterizzati da sovrapposizione di funzioni e forte variabilità in quanto a dimensioni, capacità, e risorse. Altra evidenza fondamentale, ai fini dell'elaborazione della strategia censuaria 2011, era la forte dipendenza dell'impatto organizzativo delle operazioni censuarie

¹² Il Capitolo 1 è stato curato da Donatella Zindato, che ha redatto anche i paragrafi 1.1, 1.2, 1.3 e 1.4. Il paragrafo 1.5 è stato redatto da Giancarlo Carbonetti. Il paragrafo 1.6 è stato redatto da Damiano Abbatini. Il paragrafo 1.7 è stato redatto da Maura Giacommo.

¹³ La necessità di migliorare la tempestività nasceva, oltre che dall'ovvia esigenza di garantire la rilevanza dei dati prodotti, anche dall'obbligo di rispettare i vincoli posti dal Regolamento europeo sui Censimenti della popolazione e delle abitazioni (Regolamento (Ce) 9 luglio 2008 n. 763/2008 del Parlamento europeo e del Consiglio), tra cui quelli relativi alla trasmissione dei dati alla Commissione (Eurostat) da parte di tutti gli Stati membri entro ventisette mesi dalla fine dell'anno di riferimento (cfr. art. 5).

¹⁴ Ad esempio, dato il numero crescente di famiglie unipersonali e di cosiddette *dink couples - double income no kids* - erano sempre più numerose le famiglie assenti dall'abitazione per l'intero arco della giornata, e allo stesso tempo erano in aumento le fasce di popolazione (popolazione anziana, popolazione straniera) per le quali era ipotizzabile la necessità di fornire assistenza “specializzata” alla compilazione del questionario.

¹⁵ Nel 2001, ad esempio, solo il 2 per cento dei comuni aveva concluso la rilevazione nei tempi stabiliti (Fortini *et al.*, 2007).

¹⁶ Si pensi al forte aggravio di carico di lavoro, repentino e concentrato nel tempo, cui i Comuni si trovavano a far fronte. Inoltre, in molti casi i comuni erano costretti a integrare con risorse finanziarie proprie i fondi statali ricevuti per il censimento (Fortini *et al.*, 2007).

dall'ampiezza demografica del Comune¹⁷. Ulteriori elementi di criticità erano rappresentati dal crescente carico statistico sui rispondenti (dovuto alla proliferazione di indagini, ufficiali e non), dalla disaffezione della popolazione nei confronti delle rilevazioni statistiche e dalle preoccupazioni relative alla tutela della confidenzialità delle informazioni fornite.

Veniva così delineandosi l'esigenza di una strategia basata sulle seguenti linee-guida:

- specializzazione delle funzioni, per evitare le sovrapposizioni che causavano inefficienze della macchina organizzativa e provocavano ritardi nelle operazioni;
- inversione del rapporto tra *front office* (rilevatori sul campo) e *back office* comunale, per diminuire il numero dei rilevatori necessari alla conduzione delle operazioni, garantendo al tempo stesso una maggiore qualità dei dati rilevati grazie al rafforzamento del *back office*;
- flessibilità dell'organizzazione della rilevazione, per rispondere alle esigenze diversificate dei Comuni;
- flessibilità delle tecniche di rilevazione, per rispondere alle esigenze diversificate dei rispondenti.

Ai tradizionali obiettivi censuari della determinazione della popolazione legale, della produzione di dati di elevato dettaglio territoriale sulle principali caratteristiche della popolazione, delle famiglie e degli alloggi, della revisione e dell'aggiornamento delle anagrafi comunali, si affiancavano quindi i seguenti obiettivi:

- coniugare la disponibilità di dati dettagliati a livello territoriale fine (garantita dalla strategia censuaria tradizionale) alla tempestività nella diffusione dei risultati, al fine di migliorare la rilevanza dei dati prodotti e rispettare i tempi di diffusione previsti dal Regolamento dell'Unione europea sui censimenti della popolazione e delle abitazioni (Unece, 2006);
- ridurre il carico di lavoro degli Ucc e rafforzarne le capacità di controllo e coordinamento delle operazioni sul campo grazie all'inversione del rapporto tra *front office* e *back office* e all'adozione di soluzioni differenziate in base alla classe di ampiezza demografica dei Comuni;
- diversificare le tecniche di compilazione/restituzione del questionario per garantire ai rispondenti la massima flessibilità e al contempo incrementare il tasso di restituzione spontanea (riducendo così il ricorso ai rilevatori);
- prendere in considerazione le aspettative dei cittadini relative alla diminuzione del carico statistico e alla tutela della privacy;
- contenere i costi, ma tenere conto dei costi effettivi del censimento, anche sulla base dei risultati dell'indagine condotta dall'Istat sui costi sostenuti dai Comuni per il censimento (Istat, 2017a);
- migliorare la coerenza tra dati di fonte anagrafica e dati censuari.

L'utilizzo delle Liste anagrafiche comunali (Lac) per guidare la rilevazione ha costituito la principale innovazione del censimento 2011. Insieme all'utilizzo di altri dati di fonte amministrativa a supporto della rilevazione, ha reso possibile la trasformazione della rilevazione "porta a porta" in un "censimento assistito da lista" (cfr. paragrafo 1.2.), seppur nell'ambito di una metodologia sostanzialmente convenzionale (Unece, 2014; Unece, 2015) ovvero basata sulla rilevazione esaustiva sul campo. L'utilizzo delle Lac come liste di famiglie e di indirizzi ha reso inoltre possibile la spedizione postale dei questionari e l'adozione di un sistema di restituzione multicanale (cfr. paragrafo 1.3.) che, seppur oneroso dal punto di vista organizzativo, appariva irrinunciabile dati i problemi di sostenibilità della strategia tradizionale.

L'uso delle tecnologie web ha costituito un'altra innovazione fondamentale del Censimento 2011 (Benassi *et al.*, 2014). Oltre al questionario elettronico per le famiglie (il Cawi - *Computer assisted web interviewing* - era uno dei canali di restituzione del questionario), sono stati sviluppati un sistema *online* per gestire in contemporanea le diverse fonti a supporto della rilevazione e le diverse opzioni per la compilazione/restituzione del questionario (cfr. paragrafo 1.7), e un sistema web dedicato alla rete di rilevazione, per rendere disponibili a tutti gli operatori le informazioni utili per lo svolgimento delle operazioni censuarie¹⁸.

Altra caratteristica fondamentale della nuova strategia è stato il suo carattere *modulare* ovvero la diversificazione di metodi e organizzazione per classe di ampiezza demografica del Comune (cfr. paragrafo

¹⁷ Le analisi condotte in relazione alle criticità di processo della precedente tornata censuaria avevano evidenziato come il principale fattore di rischio fosse costituito dalla dimensione del comune (Fortini *et al.*, 2007).

¹⁸ I tre sistemi sono stati sviluppati utilizzando tecnologie *open source*, e sono stati progettati per essere successivamente riutilizzati, secondo quanto previsto dalle linee guida dell'ente nazionale preposto alla digitalizzazione della pubblica amministrazione (Digit Pa) (Romano, 2017).

1.4). In particolare, nei Comuni di maggiori dimensioni¹⁹, il censimento è stato preceduto dalla *Rilevazione dei numeri civici – Rnc*²⁰, grazie alla quale sono stati censiti gli edifici e sono state rilevate informazioni da utilizzare nella fase di completamento della rilevazione sul campo (cfr. paragrafo 1.6). Inoltre, sempre nei Comuni di maggiori dimensioni, al fine di ridurre il carico statistico complessivo sulle famiglie e produrre un risparmio sui costi della rilevazione, è stata adottata la cosiddetta strategia *short form/long form*. Le variabili socio-economiche sono state rilevate su base campionaria (solo un campione di famiglie ha ricevuto il questionario completo o *long form*, mentre la maggioranza delle famiglie ha ricevuto il questionario ridotto o *short form*) (cfr. paragrafo 1.5). Infine, l'aggiornamento delle Lac alla data di riferimento del censimento è stato effettuato con modalità diverse a seconda dell'ampiezza demografica (cfr. paragrafo 2.1.3.).

Le numerose innovazioni di metodo e di processo che caratterizzavano la nuova strategia censuaria sono state testate nell'ambito dell'indagine pilota (condotta a ottobre 2009). Data la complessità della nuova strategia, gli obiettivi dell'indagine sono stati variamente combinati all'interno di diversi schemi di rilevazione, ciascuno dei quali è stato testato su un certo numero di Comuni rappresentativi delle diverse realtà territoriali e demografiche. Ciò ha consentito di limitare l'impatto sui singoli Comuni coinvolti, rendendo possibile al tempo stesso la stima dei parametri necessari per la messa a punto della strategia censuaria e la valutazione dell'impatto complessivo delle innovazioni proposte (Cassata e Tamburrano, 2011). Tra queste, una delle più importanti era costituita dal sistema di restituzione multicanale, che prevedeva la possibilità per le famiglie di compilare il questionario elettronico sul web, o di compilare il questionario cartaceo e restituirlo presso gli Uffici postali, presso i Centri comunali di raccolta (Ccr) o al rilevatore.

1.2 Il censimento assistito da lista

L'uso delle fonti amministrative a supporto della rilevazione sul campo ha rappresentato la principale innovazione di metodo del Censimento 2011. Si è trattato infatti di un censimento assistito da liste: le Liste anagrafiche comunali (Lac), utilizzate per la spedizione dei questionari e le liste ausiliarie, utilizzate per individuare unità di rilevazione non presenti nelle liste anagrafiche ma facenti parte del campo di osservazione del censimento ovvero dimoranti abitualmente sul territorio comunale.

La prima fase del censimento assistito da lista consiste proprio nella creazione degli archivi di base necessari allo svolgimento della rilevazione: le Lac che guidano la rilevazione (cfr. paragrafo 2.1) e le liste ausiliarie (cfr. paragrafi 2.2 e 2.3) necessarie al completamento della rilevazione sul campo, effettuato a cura degli Ucc attraverso l'invio sul campo dei rilevatori (cfr. paragrafo 4.4).

L'acquisizione ed elaborazione delle Lac deve essere svolta con congruo anticipo rispetto alla data di riferimento del censimento, in quanto gli archivi comunali devono essere resi omogenei per la loro utilizzazione ai fini dell'individuazione delle unità di rilevazione e dell'invio postale dei questionari.

In particolare, le 8.092 Lac che hanno costituito le liste di partenza del Censimento 2011 sono state acquisite con riferimento al 31 dicembre 2010 ed elaborate dall'Istat a partire dalla seconda metà di gennaio 2011 (cfr. paragrafo 2.1.2.). Oltre alle operazioni di standardizzazione delle liste e normalizzazione degli indirizzi, è stato necessario associare a ciascuna famiglia delle Lac le informazioni necessarie alla predisposizione del piano di stampa e spedizione dei questionari e alla tracciatura di ciascun questionario nel corso delle diverse fasi del processo di rilevazione:

- il codice questionario, associato univocamente all'intestatario della scheda di famiglia e all'indirizzo di iscrizione in anagrafe, necessario per tracciare il questionario nelle diverse fasi del processo tramite Sgr (Istat, 2016);
- il codice di accesso ai servizi telematici, da utilizzare per la compilazione del questionario *online*;

¹⁹ I comuni con almeno 20 mila abitanti o capoluogo di provincia alla data del 1° gennaio 2008.

²⁰ La Rnc è una rilevazione condotta dall'Istat sui 509 comuni capoluogo di provincia e/o con più di 20 mila abitanti, come attività propedeutica al 15° Censimento generale della popolazione e delle abitazioni. Gli obiettivi della rilevazione erano: realizzare un archivio di numeri civici validato da ciascun comune e geocodificati alle sezioni di censimento; permettere la costituzione di una lista comunale volta a facilitare le operazioni di recupero dell'eventuale sotto-copertura delle anagrafi; raccogliere informazioni relative al numero di edifici e alle loro caratteristiche strutturali.

- il tipo di questionario (questionario da 3 componenti/da 6 componenti; questionario in versione completa/in versione ridotta²¹ e corrispondente lettera informativa²²), determinato sulla base del numero dei componenti della famiglia risultante in Lac e del piano di campionamento predisposto dall'Istat per ciascun Comune;
- la presenza di componenti stranieri, al fine dell'inserimento nel plico del modello di ausilio plurilingue²³.

Inoltre, al fine di rilevare le famiglie iscritte in anagrafe dopo il 31 dicembre 2010 e prima della data di riferimento del censimento, è stato necessario aggiornare le Lac. L'aggiornamento è avvenuto con modalità diverse a seconda dell'ampiezza demografica dei Comuni: i Comuni di maggiori dimensioni (circa 2.400) hanno inviato nuovamente i dati della propria anagrafe, e l'Istat ha ricavato, attraverso il confronto tra i due archivi, le variazioni intervenute tra le due date di riferimento; mentre i Comuni più piccoli hanno inserito direttamente le variazioni anagrafiche mediante un'apposita funzione del Sistema di gestione (cfr. paragrafo 2.1.3).

Come già detto, l'aggiornamento delle Lac alla data di riferimento del censimento consentiva di "recuperare" quei rispondenti che avevano ufficialmente trasferito la propria dimora abituale nel Comune dopo il congelamento della Lac ma prima della data di riferimento del censimento. Le liste anagrafiche non erano però gli unici archivi necessari allo svolgimento della rilevazione. Infatti, se l'esistenza di una lista di unità di rilevazione presentava indubbi vantaggi rispetto alla rilevazione "porta a porta", riducendo gli oneri della rilevazione sul campo e facilitando le operazioni di consegna dei questionari, al fine di garantire l'eshaustività della conta censuaria era necessario prevedere dei meccanismi di controllo degli errori di "copertura" delle liste anagrafiche, affette da errori di sovra-copertura (mancate cancellazioni di persone non più abitualmente dimoranti nel Comune) e di sotto-copertura (mancate iscrizioni di persone dimoranti abitualmente nel Comune).

Il meccanismo di correzione degli errori di sovra-copertura era in un certo senso implicito nella rilevazione assistita da lista: le persone non più dimoranti abitualmente nel Comune presumibilmente non sarebbero state rilevate e quindi non sarebbero state conteggiate. Per correggere la sotto-copertura era invece necessario il ricorso a fonti terze che fornissero segnali utili ai fini dell'individuazione di unità (famiglie o convivenze) non presenti nelle Lac ma dimoranti abitualmente nel Comune. Queste informazioni ausiliarie sono state utilizzate nell'ambito delle operazioni di completamento della rilevazione sul campo, effettuate, appunto ai fini del recupero della sotto-copertura anagrafica, con modalità che si differenziavano a seconda della dimensione demografica del Comune, come verrà illustrato in dettaglio nel paragrafo 4.4.

In particolare, per tutti i Comuni sono state utilizzate le informazioni confluite nella Lista integrativa da fonti ausiliarie (Lifa), costituita dall'Istat incrociando con le Lac i dati di due archivi: l'Anagrafe tributaria (AT) per i cittadini italiani e comunitari e i Permessi di soggiorno (PS) per i cittadini stranieri non comunitari. Più precisamente, per ciascun Comune la Lifa ha incluso tutti gli individui che fossero:

- presenti nella Lifa del Comune e nella Lac di un altro Comune;
- presenti nella Lifa del Comune e non presenti in Lac (in nessun Comune);
- presenti nella Lifa del Comune (lo stesso in cui l'individuo risultava presente in Lac) a un indirizzo diverso da quello di iscrizione in anagrafe (in questo caso l'informazione non è stata utilizzata per il recupero della sotto-copertura ma è stata resa disponibile all'Ucc per il confronto censimento-anagrafe).

Inoltre, per i Comuni di maggiori dimensioni, confrontando il numero di unità immobiliari potenzialmente abitative rilevato con la Rnc con il numero di interni occupati da unità presenti nella lista anagrafica (abitazioni occupate da famiglie iscritte in anagrafe), è stato possibile individuare ulteriori interni abitativi da verificare sul campo (cfr. paragrafo 2.3).

Inizialmente, dunque, Sgr è stato alimentato con i dati di oltre 25 milioni di famiglie provenienti dalle Liste anagrafiche comunali al 31 dicembre 2010 (i dati utilizzati per la stampa personalizzata e la spedizione postale dei questionari), e di quasi 9 milioni di individui presenti nella Lifa (individui potenzialmente presenti sul territorio ma non iscritti in anagrafe). Successivamente, i dati caricati in Sgr sono stati aggiornati sulla base

²¹ Si fa riferimento alla summenzionata strategia *short form/long form*, descritta in maggior dettaglio nel paragrafo 1.5.

²² La lettera di accompagnamento era stata anch'essa predisposta in due versioni (lettera di accompagnamento alla versione completa del questionario e lettera di accompagnamento alla versione ridotta) in quanto nella lettera di accompagnamento alla versione completa del questionario si faceva riferimento ai quesiti sensibili contenuti nella sezione "Difficoltà nella vita quotidiana".

²³ Cfr. paragrafo 4.2.

delle Lac riferite alla data del Censimento che, fornendo dati sui movimenti in ingresso e in uscita dal Comune dopo il congelamento della Lac (cioè dopo l'invio all'Istat della Lac), hanno consentito di individuare sia famiglie non presenti nella Lac che avevano trasferito la propria dimora abituale nel Comune dopo il congelamento della Lac ma prima della data di riferimento del censimento (da "recuperare" attraverso la consegna del questionario a cura dell'Ucc – cfr. paragrafo 4.2), sia famiglie che non rientravano più nel campo di osservazione perché avevano trasferito la propria dimora abituale fuori dal comune (e dunque unità della lista di partenza che non rappresentavano mancate risposte da recuperare – cfr. paragrafo 4.4). Inoltre, l'aggiornamento delle Lac consentiva agli Ucc di svolgere correttamente le operazioni di confronto censimento-anagrafe (cioè di confrontare i dati rilevati con i dati delle Lac riferiti alla stessa data).

1.3 Uso di nuove tecniche di rilevazione

Diversamente dai censimenti precedenti, in cui il rilevatore batteva la sezione di censimento a tappeto "alla cieca" (cioè senza l'ausilio di liste precompilate delle unità che rientrano nel campo di osservazione) per consegnare e ritirare i questionari cartacei, il censimento del 2011 si è avvalso delle Lac come lista di unità di rilevazione. Ciò ha reso possibile l'adozione della tecnica del *mail-out* (spedizione postale) per la consegna del questionario alle famiglie.

Inoltre, per consentire la massima flessibilità ai rispondenti e conseguire un elevato tasso di risposta "spontanea" (percentuale di questionari restituiti tramite modalità alternative al rilevatore), è stata adottata una tecnica di restituzione multicanale che comprendeva una pluralità di canali per la restituzione del questionario cartaceo, oltre alla possibilità di compilare il questionario *online*. Per far fronte ad esigenze diversificate e garantire al contempo una riduzione del ricorso ai rilevatori, si decise infatti di associare l'opzione internet a soluzioni, anch'esse innovative per il contesto italiano, ovvero la possibilità di restituire il questionario presso gli Uffici postali (diffusi in maniera capillare sul territorio italiano) o presso Centri di raccolta appositamente istituiti sul territorio comunale.

Come già accennato, la restituzione multicanale, seppur complessa da gestire, appariva necessaria date le difficoltà legate alla strategia tradizionale e la necessità di ridurre il numero dei rilevatori sul campo, garantendo al contempo una maggiore qualità e tempestività della rilevazione. Anche dall'analisi del contesto internazionale emergeva con chiarezza l'orientamento verso la compresenza di una molteplicità di tecniche.

Per quanto riguarda l'adozione dell'opzione internet, era una scelta in un certo qual modo obbligata, seppur non ovvia. D'altra parte, date le percentuali di utilizzo delle tecnologie informatiche da parte delle famiglie ancora relativamente basse (seppur in crescita), la compilazione via internet veniva utilizzata nell'ambito di sistemi *multi-mode*, associata (in via sperimentale su sottogruppi della popolazione o come possibilità offerta a tutti i potenziali rispondenti) al questionario cartaceo, distribuito con uno qualsiasi dei metodi tradizionali (*hand-delivery*, *list-leave*, *mail-out*, *mail back*, eccetera) (Unece, 2006).

Oltre agli ovvi vantaggi associati all'utilizzo del questionario elettronico (possibilità di editing interattivo e dunque migliore qualità dei dati rilevati), si evidenziava come la compilazione via internet potesse facilitare la rilevazione di una parte delle sottopopolazioni difficili da rilevare (cosiddetti *hard to count*), quali i giovani adulti. Tuttavia, l'opportunità di adottare la tecnica Cawi doveva essere valutata sulla base della proporzione della popolazione in grado di accedere a internet da casa, della proporzione di utenti che si avvalevano di sistemi di connessione a banda larga e della proporzione di utenti di internet per altri scopi (*online banking*, *e-commerce*, eccetera); a riguardo, e i dati sulla diffusione e l'utilizzo delle nuove tecnologie dell'informazione e della comunicazione da parte delle famiglie non erano molto incoraggianti²⁴. Inoltre, l'analisi delle esperienze degli altri paesi evidenziava come l'utilizzo del Cawi comportasse inizialmente un aumento dei costi in quanto, non potendo prevedere in anticipo chi avrebbe usato il web, si richiedeva comunque di consegnare un questionario cartaceo a tutte le unità di rilevazione²⁵.

²⁴ Nel 2006, infatti, l'Italia era al quindicesimo posto tra i paesi dell'Unione Europea per percentuale di famiglie con almeno un componente tra i 16 e i 64 anni che possedevano un accesso a internet da casa (appena il 40 per cento rispetto alla media europea del 52 per cento). Inoltre, si evidenziava una situazione molto disomogenea sul territorio e forti divari di tipo generazionale, culturale ed economico. Cfr. Istat (2005), Istat (2006) e Romano (2017, paragrafo 2.1.1).

²⁵ L'utilizzo di internet consentiva un'effettiva riduzione dei costi necessari per la stampa e l'acquisizione dei questionari cartacei e per la rilevazione sul campo (pagamento dei rilevatori e/o spedizione postale dei questionari) solo nel caso di un censimento *totally paperless* (che, nella maggior parte dei casi, rappresentava ancora solo un modello a cui tendere). Ad esempio, in Canada, uno dei primi paesi a utilizzare il Cawi nell'ambito del

Nonostante i bassi tassi di utilizzo prevedibili, la complessità che caratterizzava la progettazione del software²⁶ e la verosimile lievitazione dei costi che avrebbe comportato, l'opzione internet appariva in qualche modo come una "scelta necessaria"²⁷. L'Istat riteneva infatti che un'operazione massiccia e "visibile" come il censimento non potesse prescindere dalle aspettative di miglioramento/semplificazione dei rapporti tra cittadini e istituzioni. Veniva insomma intesa come un "servizio" da fornire alla popolazione, indipendentemente dal contributo che si stimava potesse offrire alla copertura e alla qualità della rilevazione censuaria (Istat, 2012a).

Nel censimento 2011 la tecnica Cawi è stata quindi utilizzata all'interno di un *concurrent design*, come uno dei possibili canali di restituzione, al fine di innalzare il tasso di "restituzione spontanea" (Aragona e Zindato, 2016). In particolare, scegliendo la modalità che più soddisfaceva le proprie esigenze, le famiglie potevano optare per la compilazione del questionario *online* oppure compilare il questionario cartaceo e riconsegnarlo presso uno dei 14 mila Uffici postali sul territorio italiano o presso uno dei Centri comunali di raccolta (Ccr) appositamente istituiti sul territorio comunale. Il questionario poteva essere riconsegnato anche presso un Ufficio postale di un Comune diverso da quello di residenza²⁸ o presso un Ccr di una zona della città diversa da quella di residenza (ad esempio, al Ccr più vicino al luogo di lavoro o alla scuola dei figli).

Un ulteriore canale di restituzione era rappresentato dai rilevatori, che entravano in azione nella seconda fase della rilevazione e la cui funzione è stata fondamentale (come descritto in dettaglio nel capitolo 4) nonostante la drastica riduzione di numero (65 mila rilevatori contro i 100 mila del 2001) consentita dalle innovazioni di metodo e di tecnica, per effettuare la consegna dei questionari non spediti (questionari che per vari motivi non era stato possibile "postalizzare") e il recupero mirato delle mancate risposte (i questionari non restituiti spontaneamente dalle famiglie). Il canale "rilevatore" era dunque un canale sequenziale rispetto agli altri canali (nella fase di "restituzione spontanea" non era previsto il ricorso ai rilevatori, ma le famiglie avevano la possibilità di recarsi presso i Ccr per ricevere supporto alla compilazione), ma anche concomitante (limitatamente alla fase di recupero delle mancate risposte), in quanto, proprio per garantire ai rispondenti la massima flessibilità e ridurre il più possibile il ricorso ai rilevatori, i canali di restituzione spontanea restavano attivi fino al termine della rilevazione.

Tutti i questionari cartacei (quelli consegnati dai rispondenti al rilevatore o portati direttamente nei Centri comunali di raccolta o restituiti presso gli Uffici postali), indipendentemente dal canale prescelto dalla famiglia, venivano "restituiti" agli Uffici comunali di censimento (Ucc), ai quali erano richieste notevoli capacità organizzativo-logistiche per tenere conto della pluralità di canali disponibili. In generale, se la restituzione multicanale consentiva al rispondente la massima flessibilità, attraverso la scelta non solo della modalità di compilazione/restituzione più confacente alle proprie esigenze ma anche del punto di restituzione (Ufficio postale o Ccr) più compatibile con i propri spostamenti, richiedeva all'Ucc di predisporre le strategie organizzative necessarie a gestire la flessibilità della restituzione sul territorio. In particolare: per i questionari cartacei restituiti presso gli Uffici postali, l'indirizzo di recapito finale era quello stabilito dall'Ucc e prestampato sui questionari (uno o più indirizzi, a seconda della dimensione e dell'organizzazione dell'Ucc); per i questionari restituiti presso i Ccr, l'Ucc doveva organizzarsi per effettuare le operazioni di revisione indipendentemente dalla sezione di censimento di residenza del rispondente (che, come detto, poteva scegliere di restituire il questionario presso uno qualsiasi dei Ccr attivi sul territorio comunale). In alternativa, doveva provvedere a trasferire i questionari dal Ccr presso il quale erano stati restituiti a quello territorialmente "competente".

Se, nell'impianto complessivo, il ricorso ai rilevatori sul campo veniva ridotto, adottando il *mail-out* in sostituzione della tradizionale tecnica di *hand-delivery*, il ruolo degli Ucc rimaneva dunque di primaria

Censimento della popolazione e dove la compilazione via internet aveva riscosso il maggior successo (18,6 per cento delle famiglie nel 2006), si progettava di continuare ad utilizzare un sistema di rilevazione multicanale, concentrando al contempo le risorse sull'identificazione a priori dei potenziali rispondenti *online* (Abbatini *et al.*, 2007).

²⁶ Le caratteristiche della rilevazione censuaria (alti numeri in gioco; necessità di un'infrastruttura capace di sostenere possibili situazioni di picco, di garantire tempi di risposta rapidi per non rendere l'esperienza *online* frustrante per l'utente, e di assicurare elevati standard di sicurezza) rendevano la progettazione del software per la compilazione del questionario *online* particolarmente complessa (Tinini e Virgillito, 2013).

²⁷ Cfr. Romano (2017), paragrafo 2.1.1.

²⁸ Durante le fasi preparatorie del censimento, era stato chiesto ai comuni di indicare (ai fini della stampa sui questionari) l'indirizzo presso il quale sarebbero state effettuate le operazioni di revisione dei questionari cartacei a cura degli operatori comunali e registrazione dell'avvenuta consegna nel sistema di monitoraggio. Su ciascun questionario era quindi prestampato (oltre all'indirizzo del destinatario, necessario per la spedizione) l'indirizzo presso il quale dovevano essere recapitati da Poste italiane i questionari riconsegnati presso gli Uffici postali (Istat, 2011).

importanza, per coordinare i rilevatori ai fini del completamento della rilevazione sul campo (consegna dei questionari non spediti, recupero delle mancate risposte, recupero della sotto-copertura), organizzare le operazioni di revisione dei questionari cartacei, effettuare il confronto censimento-anagrafe e gestire i Centri comunali di raccolta, che costituivano essi stessi un canale di rilevazione. Spettava dunque al responsabile dell'Ucc individuare il numero di operatori necessari a garantire il rispetto dei tempi previsti dal calendario delle operazioni e valutare quali compiti assegnare a ciascuno degli addetti alla rilevazione.

Ai Ccr competevano due funzioni fondamentali:

- fornire chiarimenti e assistenza alla compilazione e restituzione del questionario ai rispondenti che sceglievano di recarsi presso un Centro comunale;
- garantire lo svolgimento delle attività necessarie alla ricezione dei questionari cartacei consegnati dai rispondenti presso i Ccr invece che presso gli Uffici postali.

Un'altra funzione fondamentale dei Centri comunali di raccolta era la "promozione" della compilazione *online* del questionario, attraverso l'installazione di postazioni a disposizione dei rispondenti e la presenza di operatori formati all'utilizzo del questionario elettronico. Data la maggior qualità dei dati e tempestività dell'acquisizione garantita dalla compilazione del questionario elettronico, l'Istat ha incentivato gli Ucc a promuovere presso le famiglie l'uso della tecnica Cawi prevedendo che il compenso per i questionari compilati *online* variasse in base alla percentuale complessiva di Cawi raggiunta nel Comune²⁹.

Tutte le attività volte a facilitare e promuovere la compilazione del questionario elettronico contribuivano a ridurre il carico di lavoro dell'Ucc. Se per i questionari cartacei era richiesto agli Ucc di registrarne l'arrivo, effettuarne la revisione (e, nel caso di questionari incompleti, ricontattare la famiglia) e compilare i modelli riepilogativi (necessari all'Istat per la diffusione dei dati provvisori e al Comune per il confronto censimento-anagrafe), per i questionari web tutte queste operazioni avvenivano in automatico, contestualmente alla compilazione. La compilazione web era quindi "vantaggiosa" per gli Ucc, in quanto comportava una notevole riduzione del lavoro di *back office* e, grazie all'aggiornamento immediato del sistema di monitoraggio, consentiva una gestione più efficiente degli interventi sul campo.

Dal punto di vista dei risultati, la scelta di adottare tecniche di rilevazione innovative è stata premiata dall'elevatissimo tasso di risposta spontanea (87,7 per cento)³⁰, consentendo al tempo stesso una significativa riduzione sia del lavoro di *front office* (riduzione del numero di rilevatori) che del lavoro di *back office* dei Comuni. Allo stesso modo, la scelta di adottare l'opzione Cawi ha facilitato il rispondente attraverso la compilazione "guidata" del questionario e la generazione automatica dei fogli aggiuntivi e, pur non giustificata dalle previsioni basate sulla diffusione di internet, si è rivelata più che opportuna alla luce della risposta sorprendentemente positiva dei rispondenti (quasi una famiglia su tre ha scelto il web per compilare il proprio questionario, con oltre un milione di questionari restituiti nella prima settimana di esercizio)³¹. Inoltre, la riduzione del numero di pagine da acquisire ha consentito una certa riduzione dei costi, anche se, come detto, una consistente riduzione dei costi totali si sarebbe potuta ottenere solo evitando di inviare il questionario cartaceo a tutte le unità di rilevazione.

1.4 Una strategia di rilevazione modulare

Dall'esame dei dati del sistema di monitoraggio predisposto dall'Istat in occasione del 14° Censimento generale della popolazione (Mtcens) emergeva con chiarezza come l'impatto organizzativo delle operazioni censuarie fosse legato fortemente alla dimensione demografica dei Comuni; se le maggiori difficoltà organizzative venivano sperimentate dai Comuni di maggiori dimensioni, i Comuni piccoli mostravano

²⁹ Il sistema di pagamenti era differenziato per canale di restituzione: i questionari consegnati presso i Ccr o recuperati dai rilevatori erano pagati 6 euro, mentre per i questionari compilati *online* veniva corrisposta al comune una quota di 4 euro (ma pari a 5,50 euro se la percentuale di Cawi nel comune fosse risultata superiore al 25 per cento). In alcuni contesti territoriali la funzione di incentivazione alla compilazione web svolta dagli Ucc ha contribuito in maniera decisiva all'innalzamento del tasso di restituzione web. Cfr. Romano (2017), paragrafo 2.4.

³⁰ In particolare, il 33,4 per cento dei questionari sono stati compilati via web, il 31,7 per cento delle famiglie ha preferito recarsi presso i Centri comunali di raccolta e il 22,6 per cento ha scelto di riconsegnare il questionario compilato presso un Ufficio postale. Solo nel 12,3 per cento dei casi è stato necessario l'intervento dei rilevatori.

³¹ Sorprendente è stata anche la partecipazione nelle primissime ore di apertura del sito, quando sono stati registrati fino a 500 mila accessi in contemporanea, tali da provocare rallentamenti e rendere necessario il potenziamento dell'infrastruttura tecnologica dimensionata sulla base di aspettative di un utilizzo più ridotto (Tinini e Virgillito, 2012).

problemi di tenuta dal punto di vista economico. Nell'ambito della progettazione del censimento 2011, si riteneva quindi vitale adottare metodi e tecniche in parte differenziati per classe di dimensione demografica dei Comuni. In particolare, il disegno censuario adottato si fondava sulla distinzione dei Comuni in quattro classi. Nel Prospetto 1.1 sono riportati il numero dei Comuni e la percentuale di popolazione residente al 1° gennaio 2007 per ciascuna delle classi definite.

Prospetto 1.1 – Numero di Comuni e popolazione al 1° gennaio 2007 per classe di ampiezza demografica.

Tipo	Ampiezza demografica	Numero	Popolazione (%)
A	Comuni con almeno 50.000 abitanti o comuni capoluogo di provincia	165	35,8
B	Comuni non capoluogo di provincia con 20.000 - 49.999 abitanti	339	17,0
C1	Comuni non capoluogo di provincia con 5.000 - 19.999 abitanti	1.859	29,6
C2	Comuni con < 5.000 abitanti	5.738	17,6
Totale		8.101	100,00

Fonte: Direzione centrale dei Censimenti generali (DCCG), Istat.

Nel Prospetto 1.2 sono invece riportati i principali moduli della nuova strategia censuaria per tipo di Comune. Come descritto nei paragrafi precedenti, la nuova strategia era contraddistinta, oltre che dall'uso delle liste anagrafiche per guidare la rilevazione, da un sistema di restituzione multicanale e dall'adozione di nuovi strumenti territoriali.

Prospetto 1.2 – Principali moduli della strategia censuaria del Censimento 2011.

Moduli	Tipo di comune			
	A	B	C1	C2
Aggiornamento delle basi territoriali	X	X	X	X
Costituzione delle aree di censimento	X	X		
Rilevazione dei numeri civici (Rnc) e Censimento degli edifici	X	X		
Uso delle liste anagrafiche per guidare la rilevazione	X	X	X	X
Acquisizione Lac aggiornata alla data del censimento	X	X	X	
Acquisizione variazioni anagrafiche alla data di riferimento del censimento			X	X
Rilevazione campionaria delle informazioni socio-economiche	X	X		
Recupero della sotto-copertura	X	X	X	X
Censimento degli edifici contestuale alla rilevazione di famiglie e abitazioni	X ^(*)	X ^(*)	X	X
Confronto censimento-anagrafe contestuale alla rilevazione sul campo e revisione	X	X	X	X

(*) Solo nelle sezioni di tipo "nucleo abitato" e "case sparse".

Fonte: Direzione centrale dei Censimenti generali (DCCG), Istat.

Come si evince dal Prospetto 1.2, si è prevista l'applicazione di alcuni moduli indipendentemente dal tipo di Comune. Si tratta innanzitutto del modulo relativo all'aggiornamento delle basi territoriali (propedeutico a qualsiasi censimento) e dei moduli costitutivi del censimento da lista ovvero:

- l'uso delle liste anagrafiche per guidare la rilevazione abbinato all'uso di nuove tecniche di distribuzione e di raccolta dei questionari (*mail-out* e restituzione multicanale);
- il recupero della sotto-copertura della lista, ovvero l'individuazione e la rilevazione delle unità di rilevazione che, pur dimorando abitualmente nel Comune, non sono presenti nella lista anagrafica. Questa fase veniva svolta con modalità diverse a seconda dell'ampiezza demografica del Comune (cfr. paragrafo 4.4);
- il confronto censimento-anagrafe contestuale alla rilevazione e la conseguente revisione delle anagrafi effettuata in coordinamento con le operazioni di rilevazione sul campo.

Altri moduli erano invece funzionali all'adozione di ulteriori innovazioni che riguardavano solo i Comuni di maggiori dimensioni (Comuni di tipo A e B). In questi ultimi, era prevista innanzitutto l'effettuazione della

Rilevazione dei numeri civici³², finalizzata alla costituzione di archivi comunali di numeri civici geocodificati alla sezione di censimento, che sostituivano i tradizionali itinerari di sezione, consentendo di migliorare l'accuratezza della rilevazione. Per ciascun indirizzo (numero civico) veniva inoltre rilevato il numero di unità immobiliari potenzialmente abitative, informazione che, in abbinamento con le Lac, consentiva di indirizzare in modo mirato l'intervento dei rilevatori nella fase di recupero della sotto-copertura (che, come anticipato, veniva effettuata in maniera diversa a seconda del tipo di Comune, proprio in funzione della disponibilità o meno di questo dato). Contestualmente alla rilevazione dei numeri civici veniva effettuato il censimento degli edifici (che invece nei Comuni di tipo C1 e C2 veniva effettuato nella fase di completamento della rilevazione sul campo attraverso la ricognizione completa della sezione da parte del rilevatore (cfr. paragrafo 4.4).

Altri moduli adottati solo per i Comuni di tipo A e B erano quelli relativi alla costituzione delle *aree di censimento di centro abitato* (Ace)³³ – domini di stima e di diffusione – e all'adozione di una strategia campionaria per la rilevazione delle variabili socio-economiche. In particolare, al fine di ridurre il carico statistico complessivo sulle famiglie e produrre un risparmio sui costi della rilevazione, nei centri abitati dei Comuni di maggiori dimensioni demografiche³⁴ la maggioranza delle famiglie (circa i due terzi) ha ricevuto il questionario in forma ridotta (*short form*), mentre il questionario completo (*long form*), contenente quesiti aggiuntivi rispetto alla versione breve, è stato compilato solo da un campione di famiglie. Nelle altre località abitate dei Comuni di maggiori dimensioni e negli altri Comuni (ovvero laddove le dimensioni demografiche non consentivano l'utilizzo di una tecnica campionaria), tutte le famiglie compilavano il questionario completo (cfr. paragrafo 1.5). Riepilogando, le aree di censimento costituivano aree di output per la diffusione dei dati censuari, intermedie tra i centri abitati e le sezioni di censimento (definite come insiemi di sezioni di censimento contigue della dimensione di circa 15 mila abitanti), funzionali alla necessità di ridurre il carico statistico sui rispondenti garantendo al contempo la produzione di dati rappresentativi a livello sub-comunale.

Infine, l'aggiornamento delle Lac, necessario per rilevare le famiglie iscritte in anagrafe dopo il 31 dicembre 2010 e prima della data di riferimento del censimento (9 ottobre 2011), è stato effettuato con modalità diverse a seconda dell'ampiezza demografica del Comune. In particolare, per i Comuni di tipo A e B era previsto un secondo invio della Lac (aggiornata alla data di riferimento del censimento), mentre per i Comuni di tipo C2 era previsto l'inserimento in Sgr (cfr. paragrafo 1.7) delle variazioni anagrafiche intercorse tra il 1° gennaio e l'8 ottobre. Infine, i Comuni di tipo C1 potevano scegliere tra le due modalità di aggiornamento (cfr. Prospetto 1.2).

Le principali innovazioni che distinguevano la strategia di rilevazione adottata nei Comuni di maggiori dimensioni (Comuni di tipo A e B) erano quindi:

- l'uso di archivi di numeri civici geocodificati alla sezione di censimento;
- la costituzione delle aree di censimento;
- l'adozione di una strategia campionaria per la produzione di stime di incroci relativi alle variabili socio-economiche, rappresentative a livello di area di censimento.

In sintesi, la nuova strategia era costituita dai seguenti moduli:

- aggiornamento delle basi territoriali;
- rilevazione dei numeri civici (Rnc);
- costituzione delle aree di censimento;
- uso delle liste anagrafiche per guidare la rilevazione;
- strategia *short form/long form*;
- invio della Lac aggiornata alla data di riferimento del censimento;
- recupero delle mancate risposte;
- recupero mirato della sotto-copertura;
- confronto censimento anagrafe e revisione delle anagrafi.

³² Più precisamente, il campo di osservazione della Rnc era costituito dai "centri abitati" (erano quindi esclusi "nuclei abitati" e "case sparse") dei Comuni con almeno 20 mila abitanti o capoluogo di provincia alla data del 1° gennaio 2008.

³³ Cfr. paragrafo 3.3.2.

³⁴ Le località di "centro abitato" dei Comuni con popolazione residente uguale o superiore a 20 mila abitanti o capoluogo di provincia al 1° gennaio 2008, nelle quali è stato possibile individuare le "aree di censimento", ovvero unità territoriali intermedie fra le sezioni di censimento e i limiti comunali, ottenute come raggruppamenti di sezioni di censimento contigue, omogenee rispetto ad alcuni parametri statistici, appartenenti ai centri capoluogo e idonee per il campionamento delle famiglie presenti in Lac.

Nei Comuni di minori dimensioni (Comuni di tipo C1 e C2) la nuova strategia era invece costituita dai seguenti moduli:

- aggiornamento delle basi territoriali;
- uso delle liste anagrafiche per guidare la rilevazione;
- inserimento delle variazioni anagrafiche in Sgr (in alternativa all’invio della Lac aggiornata per i Comuni di tipo C1);
- recupero delle mancate risposte;
- ricognizione completa della sezione per il recupero della sotto-copertura e contestuale rilevazione degli edifici;
- confronto censimento anagrafe e revisione delle anagrafi.

Inoltre, nell’ambito delle attività volte ad ottimizzare il grado di copertura della rilevazione dei cittadini stranieri³⁵, l’Istat ha condotto alcune analisi di approfondimento per i Comuni con popolazione superiore ai 150 mila abitanti (nei quali si concentrava il 26,1 per cento dei cittadini stranieri residenti in Italia) e per i Comuni con popolazione superiore ai 100 mila abitanti con un’incidenza degli stranieri iscritti in anagrafe superiore alla media nazionale (pari al 7 per cento). I risultati di queste analisi sono stati forniti agli Ucc per facilitare l’organizzazione della rilevazione dei cittadini stranieri e le operazioni di recupero della sotto-copertura³⁶.

Infine, è opportuno evidenziare che, oltre che nella differenziazione relativa ai moduli della strategia censuaria, la variabile dimensione demografica dei Comuni è stata presa in considerazione nella definizione del calendario della rilevazione. In particolare, per tenere conto della maggiore durata delle operazioni nei Comuni di maggiori dimensioni, la data di chiusura delle operazioni sul campo, la data per il completamento delle operazioni di confronto censimento-anagrafe e l’invio delle risultanze all’Istat, e la data per la chiusura dei pacchi dei questionari cartacei per il ritiro da parte della ditta di registrazione erano differenziate per classe di ampiezza demografica del Comune. Più precisamente: i Comuni di tipo C1 e C2 dovevano completare le operazioni sul campo entro il 31 dicembre 2011, i Comuni con popolazione compresa tra 20 mila e 150 mila abitanti entro il 31 gennaio 2012 e i Comuni con oltre 150 mila abitanti entro il 29 febbraio 2012; allo stesso modo, i Comuni C1 e C2 dovevano completare le operazioni di confronto censimento-anagrafe e di chiusura dei pacchi entro il 10 febbraio 2011, mentre i Comuni con popolazione compresa tra 20 mila e 150 mila abitanti entro il 10 marzo 2012 e i Comuni più grandi entro il 10 aprile 2012. E ancora, era prevista una frequenza di recapito presso gli Ucc dei questionari cartacei restituiti dalle famiglie presso gli Up diversa a seconda che i Comuni avessero più o meno di 150 mila abitanti (cfr. paragrafo 4.3).

1.5 Strategia *short form/long form*

Per ridurre il fastidio statistico sui rispondenti, in special modo di quelli residenti nei Comuni più grandi, e per massimizzare il ritorno spontaneo dei questionari compilati, riducendo così le operazioni di sollecito da parte dei rilevatori, insieme al questionario completo contenente le tradizionali variabili di censimento, è stato disegnato anche un questionario con un ridotto numero di quesiti. Inoltre, per mantenere inalterato il contenuto

³⁵ I cittadini stranieri e gli apolidi, dimoranti abitualmente o temporaneamente presenti in Italia alla data del Censimento, venivano rilevati con le stesse modalità dei cittadini italiani. Tuttavia, al fine di ottimizzare il grado di copertura della rilevazione dei cittadini stranieri (che costituivano un collettivo a più elevato rischio di sotto-copertura) e la qualità dei dati rilevati, erano state progettate una serie di azioni mirate per agevolare il compito degli Ucc: predisposizione del facsimile del Foglio di famiglia in diciassette lingue per facilitare la compilazione del questionario da parte dei cittadini stranieri; servizio di *contact centre* telefonico ed e-mail con operatori in grado di gestire le richieste di aiuto in inglese, francese, spagnolo, rumeno, arabo, cinese, singalese e ucraino; invito agli Ucc ad affidare la rilevazione a rilevatori esperti e valutare i requisiti per la selezione degli stessi sulla base delle caratteristiche linguistiche e culturali della popolazione da rilevare e facendo ricorso, laddove necessario, a mediatori culturali; invito a predisporre azioni di coinvolgimento delle associazioni di stranieri e degli enti “no profit” per sensibilizzare i cittadini stranieri sul censimento, fornire assistenza nel reclutamento dei rilevatori e fornire indicazioni utili ai fini dell’individuazione di aree o centri di aggregazione della popolazione straniera; campagna pubblicitaria mirata in sei lingue (inglese, francese, spagnolo, arabo, cinese e ucraino) e azioni specifiche di promozione e di *partnership* con le associazioni delle diverse comunità di stranieri, con le comunità religiose e con le associazioni di volontariato. Cfr. Istat (2011), paragrafo 4.4.

³⁶ In particolare, sono stati forniti i seguenti materiali: matrice che raggruppa le sezioni di censimento per livello di incidenza e di densità (alta/media/bassa) della popolazione straniera complessiva (presente in Lac o nella Lifa); matrice che raggruppa le sezioni di censimento per livello di incidenza e di densità della popolazione straniera e numero di cittadini stranieri in possesso dei requisiti per l’iscrizione ma non ancora iscritti in anagrafe; dati sulla distribuzione per cittadinanza (sulla base dei quali pianificare la distribuzione dei facsimile del questionario in lingua straniera); elenco e cartogrammi raffiguranti le sezioni ad alta incidenza ed elevata densità di presenza straniera.

informativo e rispettare i vincoli internazionali (Unece, 2006; Unsd, 2007) in merito alle variabili censuarie, è stato deciso di rilevare un più ampio insieme di variabili su campioni di famiglie. Si è quindi stabilito di adottare una strategia basata sull'utilizzo contemporaneo di un questionario in forma breve (*short form*), contenente pochi quesiti relativi alle abitazioni, alle caratteristiche familiari e ad alcune variabili socio-demografiche degli individui, e di un questionario in forma estesa (*long form*), contenente tutte le variabili tradizionalmente osservate in occasione del censimento.

Nei Comuni capoluogo di Provincia e in tutti i Comuni di almeno 20 mila abitanti (486 in totale) è stato somministrato il questionario *long form* a un campione rappresentativo di famiglie, mentre lo *short form* è stato sottoposto alle rimanenti famiglie non incluse nel campione. In tutti gli altri Comuni si è invece proceduto alla rilevazione censuaria unicamente con il questionario *long form* inviato a tutte le famiglie. Con tale strategia, i dati riferiti alle variabili presenti su entrambi i tipi di questionario sono stati determinati tramite un conteggio esaustivo, mentre, esclusivamente per i Comuni sottoposti a campionamento, tutti gli altri dati riferiti alle variabili presenti solo nel *long form* e il loro incrocio con le variabili esaustive sono state desunte tramite stima campionaria.

La proposta metodologica praticata per i Comuni capoluogo di Provincia e i Comuni superiori a 20 mila abitanti ha previsto l'impiego di un disegno di campionamento casuale semplice da lista per la selezione di campioni di famiglie e per la produzione di stime accurate riferibili ad opportune aree di output (le aree di censimento di centro abitato) aventi le seguenti caratteristiche: aggregazione di sezioni di censimento di tipo "centro" contigue con popolazione mediamente intorno alle 15 mila unità. Sono state escluse dal disegno delle aree di censimento le aree extra-urbane e le zone rurali.

È importante sottolineare che, sia nei Comuni non campionati, sia nelle aree escluse dal disegno delle aree di censimento nei Comuni a campione, la rilevazione censuaria si è avvalsa solo del questionario *long form* che è stato consegnato a tutte le famiglie presenti in lista; in questi ambiti tutte le informazioni censuarie erano ottenibili tramite un conteggio esaustivo.

Questa strategia ha comportato, per i Comuni di maggiore dimensione, una riduzione della mole dei dati da acquisire ed elaborare così da permettere maggiori controlli sui dati raccolti a vantaggio della diminuzione degli errori di misura.

1.6 Uso di nuovi strumenti territoriali: la Rilevazione dei numeri civici (Rnc)

Le già richiamate innovazioni di processo introdotte con la tornata censuaria del 2011, ovvero il passaggio da un censimento tradizionale con rilevazione esaustiva della sezione da parte dei rilevatori comunali a un censimento assistito da lista in cui il rilevatore è intervenuto sul campo soltanto per recuperare i questionari non restituiti dalle famiglie e per controllare situazioni di potenziale sotto-copertura, hanno richiesto strategie *ad hoc* per tenere sotto controllo alcuni aspetti più deboli della rilevazione. In particolare, si è dovuto prestare attenzione alle unità di rilevazione per le quali non si disponeva di una lista, soprattutto le abitazioni non occupate da residenti (mentre le abitazioni occupate da residenti sono state di fatto rilevate insieme alle famiglie iscritte in anagrafe).

Inoltre, con il censimento 2011 si è deciso, per la prima volta, di raccogliere e memorizzare digitalmente anche l'informazione relativa all'indirizzo di ciascuna unità. Attraverso un'attività piuttosto complessa, cominciata prima e proseguita durante e dopo le fasi di lavoro sul campo, tutti gli indirizzi rilevati sono stati controllati e geocodificati, assegnando in questo modo la corretta sezione di censimento alle unità rilevate.

Le informazioni relative agli indirizzi e al patrimonio edilizio collegato agli stessi sono state acquisite in due momenti diversi: nei centri abitati dei Comuni capoluogo di Provincia e nei Comuni sopra i 20 mila abitanti i dati sono stati raccolti durante la Rilevazione dei numeri civici (Rnc), svoltasi tra la fine del 2010 e l'inizio dell'anno successivo. Nella parte rimanente del territorio (i Comuni sotto i 20 mila abitanti e le sezioni non di centro abitato dei Comuni più grandi) la rilevazione ha avuto luogo contestualmente alle altre operazioni censuarie e in particolare durante la fase di completamento della rilevazione sul campo³⁷.

La Rilevazione dei numeri civici (Rnc) aveva tre obiettivi principali:

³⁷ Come già detto, il campo di osservazione della Rnc era costituito dai centri abitati dei Comuni con almeno 20 mila abitanti o capoluogo di provincia alla data del 1° gennaio 2008. Erano quindi escluse le località di nucleo abitato e case sparse, nelle quali il censimento degli edifici (e il recupero della sotto-copertura) sono stati effettuati con le stesse modalità dei comuni non capoluogo di provincia con meno di 20 mila abitanti. Cfr. paragrafo 4.4.

- realizzare, prima dell'inizio del Censimento, un archivio di numeri civici validato da ciascun Comune e geocodificato alle sezioni di censimento aggiornate al 2010;
- disporre delle informazioni necessarie per realizzare una lista comunale di dati volta a facilitare, durante la rilevazione censuaria del 2011, le operazioni di recupero della eventuale sotto-copertura della Lac;
- raccogliere le informazioni relative al numero degli edifici e alle loro caratteristiche strutturali, anticipando una parte consistente del lavoro dei Comuni per il censimento degli edifici.

Ai fini del secondo degli obiettivi sopra illustrati, per ciascun numero civico (accesso all'area di circolazione) classificato come residenziale si è calcolata la differenza tra il numero di interni abitativi e il numero di famiglie in Lac risultanti allo stesso indirizzo; in caso di differenza positiva, l'informazione relativa è stata resa disponibile ai rilevatori (visualizzata in Sgr), come segnale di potenziale sotto-copertura anagrafica o di presenza di abitazioni non occupate da rilevare.

Per la Rnc i Comuni avevano a disposizione le ortoimmagini del territorio integrate con le carte tecniche catastali geocodificate alle sezioni di censimento e le liste di numeri civici precompilate in formato digitale, con i numeri civici ordinati per sezione di censimento e arco di strada. Durante la rilevazione sul campo ciascun rilevatore doveva percorrere le strade della propria sezione di censimento orientandosi con l'aiuto della mappa cartografica. Per ciascun arco di strada il rilevatore doveva controllare tutti i numeri civici riportati nel modello cartaceo precompilato di rilevazione dei numeri civici (Mod. Istat RNC.1) e correggere eventuali errori riscontrati, cancellando i civici non trovati o integrando, quando necessario, con indirizzi non presenti nel modello. Per ciascun numero civico, inoltre, il rilevatore doveva registrare il corrispondente codice di edificio, individuandolo sulla mappa cartografica fornita dall'Istat. Inoltre, per ciascun edificio il rilevatore riportava sul modello non precompilato di rilevazione degli edifici (Mod. Istat EDI.1) il codice di edificio e le principali caratteristiche, tra cui il numero di unità immobiliari, distinte in abitative e non abitative (il numero di unità immobiliari abitative veniva poi confrontato con il numero di famiglie residenti per ricavare segnali di potenziale sotto-copertura).

I dati della Rnc, raccolti sul campo dai rilevatori comunali mediante il modello cartaceo, sono stati poi registrati in acquisizione controllata a cura dell'Ucc attraverso un portale dedicato alla rilevazione (portale Rnc) e sono stati trasmessi all'Istat attraverso un'apposita funzione del portale. L'operazione di invio, a cura del responsabile tecnico comunale, certificava la qualità dei dati contenuti nei file prodotti. Tutti i dati trasmessi dai Comuni sono stati comunque sottoposti dall'Istat a procedure di controllo e validazione e successivamente sono stati elaborati e caricati in Sgr (cfr. paragrafo 1.7) ai fini del loro utilizzo per la rilevazione della popolazione e delle abitazioni.

Il censimento degli edifici nelle porzioni di territorio non interessate dalla Rnc (i Comuni dove non era stata effettuata la Rnc e le località di nucleo abitato e case sparse dei Comuni dove era stata effettuata) si è svolto attraverso l'ausilio di modelli di rilevazione cartacei (Mod. Istat CP.ED), compilati dai rilevatori comunali al momento della rilevazione esaustiva della sezione (cfr. paragrafo 4.4). Il codice identificativo e l'indirizzo dell'edificio sono stati registrati in formato digitale in Sgr, mentre la parte restante è stata acquisita tramite lettura ottica.

1.7 Il Sistema di gestione e monitoraggio della rilevazione (Sgr)

Le innovazioni metodologiche introdotte con il censimento del 2011 hanno consentito di raggiungere un equilibrio soddisfacente tra costi e benefici ma hanno aumentato la complessità della rilevazione e il rischio di errori (dovuti all'uso dei registri per guidare la rilevazione, o a disallineamenti temporali, o alle nuove tecniche di raccolta dati, che consentivano una maggiore flessibilità ma determinavano anche una maggiore possibilità di duplicazioni nella rilevazione), rendendo necessario l'utilizzo di un sistema informativo che consentisse di gestire un flusso di lavoro che coinvolgeva più attori, monitorare le diverse fasi del processo e integrare diverse fonti di dati.

Più in generale, l'esigenza di adottare una metodologia innovativa, in grado di garantire la qualità dei dati rilevati, migliorare l'efficienza del processo e la tempestività di diffusione dei dati, facile da gestire da parte della rete di rilevazione e monitorato in tutte le sue fasi, ha portato l'Istat verso un censimento incentrato sulle tecnologie web. Questo ha richiesto lo sviluppo di un sistema informativo web altamente integrato che ha

supportato tutte le diverse fasi del processo di rilevazione (Benassi *et al.*, 2014). Tale sistema, che ha costituito esso stesso una delle principali innovazioni della nuova strategia, era composto da tre applicazioni web:

- a) un Sistema di gestione e monitoraggio (Sgr) accessibile a tutti gli operatori censuari (Istat, 2016);
- b) il questionario *online* – Qpop (Istat, 2016);
- c) un sistema di documentazione *online*, comprendente diversi tipi di contenuti come wiki, Faq, manuali, circolari, eccetera.

Il Sistema di gestione della rilevazione (Sgr) ha supportato la rete di rilevazione nell'esecuzione di tutte le operazioni previste dal Piano generale di Censimento, rappresentando di fatto la spina dorsale dell'intero processo di rilevazione, grazie alla possibilità di gestire un flusso di lavoro che coinvolgeva più attori, monitorare le diverse fasi del processo e integrare diverse fonti di dati. L'applicazione Qpop ha guidato gli intervistati nella corretta compilazione del proprio questionario attraverso regole di coerenza e verifica degli errori. Inoltre, Qpop ha presentato al rispondente solo il sotto-insieme corretto di domande a cui rispondere, in modo che la compilazione *online* risultasse più semplice, rapida e meno soggetta a errori rispetto al questionario cartaceo (Tinolini e Virgillito, 2012). Infine, un ruolo non trascurabile è stato svolto dal sistema di documentazione *online*, che ha costituito un ausilio fondamentale per i diversi operatori della rete di rilevazione, che lo hanno utilizzato per accedere alla documentazione sul censimento, quali manuali e riferimenti normativi, ma anche per essere costantemente aggiornati accedendo a contenuti quali wiki, Faq e comunicazioni a vario titolo. Va sottolineata l'importanza per un'organizzazione così complessa di uno spazio per la condivisione di informazioni standardizzate, la cui esistenza ha contribuito a ridurre i problemi dovuti alla presenza di asimmetrie informative.

Per quanto riguarda il Sistema di gestione e monitoraggio, l'adozione di un sistema di restituzione multicanale ha richiesto lo sviluppo di un sistema informatico web che consentisse di gestire in contemporanea le diverse opzioni per la restituzione del questionario e di supportare in tempo reale il lavoro dei rilevatori e degli operatori comunali di *back office*, consentendo di tenere traccia costantemente dello stato di ciascun questionario all'interno del processo di rilevazione. In particolare, la coesistenza di diverse modalità di restituzione, ovvero di informazioni provenienti da fonti diverse (questionario *online*, sistema di monitoraggio degli Uffici postali, Ccr), ha richiesto lo sviluppo di un sistema *online* costantemente aggiornato che consentisse alla rete di rilevazione di seguire nel tempo lo stato di ogni questionario. Infatti, la gestione efficiente di un sistema di restituzione multicanale richiede che il sistema sia aggiornato tempestivamente sui questionari che man mano vengono restituiti tramite i diversi canali, al fine di evitare l'invio dei rilevatori presso famiglie che avessero già restituito il questionario

Dal punto di vista informatico, Sgr è stato concepito come un insieme di funzionalità, ognuna relativa ad una fase del processo e personalizzata in base al profilo utente che vi accedeva. A supporto di tutte le attività di censimento sono state sviluppate circa ottanta funzioni, raggruppate in menù e sottomenù. Il sistema è stato personalizzato dinamicamente in base al profilo dell'utente loggato: sia i menù che i sottomenù sono stati personalizzati permettendo solo la visualizzazione delle funzioni che l'operatore era autorizzato ad utilizzare (Istat, 2016).

Una delle principali complessità della configurazione di un sistema così importante e *business critical* era costituita dalla necessità di integrare più fonti di dati e di garantire la disponibilità dell'intero sistema 24 ore su 24 (in particolare nelle ore di punta) e 7 giorni su 7. Questi obiettivi potevano essere raggiunti solo progettando e implementando un'architettura flessibile, che si basasse su sistemi bilanciati e replicabili e utilizzasse un'infrastruttura *middleware* consolidata in cui i componenti potessero essere collegati facilmente e senza massicce rielaborazioni del codice. Pertanto Sgr si è basato su un sistema modulare, flessibile e scalabile. Sebbene più di venti sviluppatori siano stati coinvolti nella sua realizzazione, la progettazione *plug-and-play* ha consentito una forte collaborazione e un rapido sviluppo di tutte le funzionalità (Istat, 2016).

Dal punto di vista informativo, Sgr è stato progettato come un sistema di *workflow* distribuito, nel quale ogni operatore lavorava seguendo un percorso standardizzato, in modo da garantire in tempo reale un flusso bidirezionale di informazioni tra responsabili della rilevazione e coordinatori da una parte e rilevatori sul campo dall'altra. Un altro obiettivo del Sistema era garantire visibilità sul processo di rilevazione all'intera rete, tenendo traccia dei cambiamenti nel tempo dello stato di ogni singolo questionario e consentendo di intervenire tempestivamente in caso di problemi o ritardi (Benassi *et al.*, 2014).

I rilevatori registravano direttamente nel sistema tutte le informazioni relative all'esito della rilevazione, dai tentativi di contatto dei rispondenti all'eventuale recupero del questionario, tramite un pannello di controllo

denominato “Diario di sezione”, che conteneva l’elenco ordinato per indirizzo di tutti i questionari della sezione, identificati dal codice questionario assegnato dall’Istat, dal tipo di questionario (S - versione ridotta; L - versione completa), dal tipo di unità di rilevazione (F - famiglia; C - convivenza), e dal cognome e nome dell’Intestatario del Foglio di famiglia o dalla denominazione della convivenza. Lo stato del questionario nelle diverse fasi della rilevazione veniva visualizzato nella colonna “Stato questionario”.

Inizialmente popolato con i dati delle Liste anagrafiche comunali e delle liste ausiliarie (Lifa e Rnc), il Diario di sezione veniva aggiornato in tempo quasi reale dal sistema di acquisizione dei dati per i questionari web mentre doveva essere aggiornato dagli operatori comunali per i questionari cartacei. Era quindi possibile seguire l’evoluzione di ogni singolo questionario, dalla consegna postale o tramite rilevatore, fino alla sua compilazione via internet o al suo arrivo presso l’Ufficio comunale di censimento, alla revisione, alla compilazione dei modelli riepilogativi e alla conclusione delle operazioni di confronto censimento-anagrafe. Attraverso i rapporti riassuntivi, aggiornati in tempo reale, era possibile inoltre monitorare l’andamento complessivo della rilevazione sul territorio.

Questa modalità operativa ha prodotto benefici in termini di tempestività, costi e capacità di controllo centralizzato della rete censuaria. La tracciabilità di qualsiasi operazione effettuata sul sistema ha permesso all’Istat, così come agli Uffici regionali e provinciali di censimento, di avere costantemente un quadro completo relativo all’andamento delle operazioni e allo stato di avanzamento delle varie attività presso ciascun Ufficio comunale di censimento. Nel contempo, ha reso gli operatori della rete comunale indipendenti nella gestione delle operazioni loro affidate. Infatti, il sistema è stato progettato per automatizzare il lavoro di *back office* e per garantire flessibilità all’organizzazione del lavoro sul campo all’interno di ogni Ucc. I Comuni hanno creato autonomamente la propria rete di rilevazione, assegnando un profilo di sistema a ciascun utente (sono stati inseriti i dati anagrafici di quasi 85 mila operatori, suddivisi tra rilevatori, coordinatori comunali e altri operatori degli Uffici comunali di censimento), stabilendo le relazioni gerarchiche all’interno della rete di rilevazione comunale (ad esempio tra i coordinatori e i rilevatori) e assegnando le sezioni di censimento ai rilevatori.

Il sistema, accessibile *online* a tutti gli operatori censuari, ha permesso di seguire lo stato di ogni singolo questionario in tempo quasi reale, consentendo così il recupero mirato dei questionari mancanti. La disponibilità di informazioni costantemente aggiornate sullo stato di ogni singolo questionario ha consentito di indirizzare i rilevatori solo presso le famiglie alle quali il questionario era stato inviato ma non ancora restituito. Inoltre, le informazioni delle liste ausiliarie, integrate e caricate nel sistema, sono state utilizzate dai rilevatori per il recupero mirato e sistematico della sotto-copertura anagrafica, contestualmente al recupero delle mancate risposte.

Infine, il sistema prevedeva anche un’importante funzione da utilizzare per l’effettuazione del confronto censimento-anagrafe ovvero del controllo incrociato *online* tra dati censuari e anagrafi e per la produzione del relativo bilancio. Inserendo in Sgr i dati identificativi delle persone censite, e confrontandoli con le liste anagrafiche comunali aggiornate all’8 ottobre 2011, gli operatori comunali hanno effettuato il confronto censimento-anagrafe in tempo reale. Ciò ha consentito quindi la tempestiva diffusione dei risultati di conteggio del censimento. Infatti, i primi e importantissimi dati definitivi (ovvero i dati sulla popolazione legale dei Comuni per sesso, età e cittadinanza) sono stati diffusi prima della fine dell’acquisizione dei dati dei questionari cartacei, proprio sulla base dei dati inseriti in Sgr.

CAPITOLO 2

PREDISPOSIZIONE DELLE LISTE (LAC, LIFA, RNC)³⁸

2.1 Produzione e uso delle Liste anagrafiche comunali (Lac)

Per la prima volta in Italia, nel 2011 sono stati utilizzati i dati della popolazione residente registrata nelle anagrafi di ciascun Comune per guidare la rilevazione censuaria. Le informazioni di base su individui, famiglie e convivenze sono state acquisite grazie ad una nuova rilevazione: la Rilevazione delle Liste anagrafiche comunali (Lac). Dopo alcune sperimentazioni iniziate nel 2008, nel 2011 è stata avviata la nuova rilevazione, nell'ambito della quale tutti i Comuni italiani sono stati chiamati a trasmettere file standardizzati contenenti il set ottimale di variabili su individui, famiglie e convivenze.

I dati raccolti sono stati trattati e successivamente caricati nel Sistema di gestione della rilevazione censuaria in modo da: inviare i questionari agli intestatari dei fogli di famiglia, tracciare le attività dei rispondenti e degli operatori censuari e consentire ai Comuni di effettuare il confronto censimento/anagrafe contestualmente alle attività censuarie; il processo, così costruito, ha permesso di determinare l'ammontare e la struttura della popolazione legale in tempi di gran lunga più rapidi rispetto ai censimenti precedenti oltre, ovviamente, a semplificare attività che in passato avrebbero richiesto un carico di lavoro di gran lunga superiore per tutti gli attori coinvolti.

La rilevazione delle Lac ha consentito di acquisire un patrimonio informativo vasto e utile per la costruzione di registri integrati di famiglie e individui e per l'estrazione centralizzata, presso l'Istat, delle unità campione di tutte le rilevazioni che coinvolgono la popolazione. Inoltre, il software implementato per la trasmissione/acquisizione dei dati e per il monitoraggio e il supporto all'attività dei Comuni, mette a disposizione dei Comuni i risultati delle analisi sulla qualità dei dati; con tale funzione viene data ai Comuni sia la possibilità di visualizzare i risultati di sintesi delle regole di compatibilità, sia la possibilità di scaricare le informazioni sui singoli individui che presentano errori, anomalie e dati mancanti. In questo modo si è innescato un processo virtuoso volto a garantire sempre più elevati standard di qualità e, in futuro, l'utilizzabilità degli archivi amministrativi locali per calibrare in maniera efficiente ed efficace il censimento permanente che si è realizzato a partire dal 2018.

Nel corso del periodo post-censuario, dai controlli quantitativi e qualitativi a cui sono stati sottoposti i dati di fonte anagrafica e dalle risultanze censuarie, è emerso che il divario tra dato amministrativo locale e dato censuario, talvolta anche di non modesta entità, è da imputare ai seguenti motivi:

- alla diversa natura delle due fonti - l'anagrafe, infatti, è un registro di popolazione che risente fortemente di tempi e modalità delle procedure amministrative basate sulle dichiarazioni rese dai cittadini mentre il censimento è un'operazione statistica condotta in un periodo di tempo relativamente breve da diversi attori sul territorio con l'obiettivo di determinare l'ammontare della popolazione residente di ciascun Comune;
- alle dimensioni della qualità (accuratezza, consistenza, completezza e non ridondanza) ritenute necessarie da ciascuna amministrazione comunale per la tenuta dei dati del proprio registro anagrafico;
- all'impegno degli operatori censuari e dei rispondenti di ciascun comune nel garantire l'aderenza alla realtà del dato rilevato.

Focalizzando l'attenzione sull'analisi delle dimensioni della qualità dei registri anagrafici comunali è emerso che questi possono essere influenzati da numerosi aspetti quali:

- la dimensione demografica dei Comuni;
- la localizzazione territoriale delle anagrafi;

³⁸ Il Capitolo 2 è stato curato da Donatella Zindato. Il paragrafo 2.1 è stato redatto da Anna Pezone. La stesura del paragrafo 2.2 è stata curata da Luca Mancini. Ha contribuito al paragrafo 2.2.1 Federico Benassi. Il processo di creazione delle Lifa per i cittadini stranieri non comunitari, introdotto all'inizio del paragrafo 2.2.1 e descritto nel dettaglio nei paragrafi 2.2.2 e 2.2.3 è frutto del lavoro congiunto con Marco Fortini e Luigi Marcone. Si ringraziano Francesco Borrelli, Stefania Di Domenico e Alessandra Ronconi per aver collaborato alle operazioni di de-duplicazione e *record linkage* descritte nei paragrafi 2.2.2 e 2.2.3. Il paragrafo 2.3 è stato redatto da Maura Giacommo.

- i controlli e i vincoli di qualità imposti dai software utilizzati per l'acquisizione e la gestione delle pratiche anagrafiche rese ai cittadini.

Per il Censimento permanente i dati delle anagrafi comunali rappresenteranno una irrinunciabile fonte informativa sulla quale eseguire il conteggio di popolazione in termini di individui, famiglie e convivenze, pertanto per i controlli eseguiti su questi dati devono sempre essere garantiti i più elevati standard di qualità.

2.1.1 Le sperimentazioni

Per poter acquisire i dati in formato elettronico forniti da 8.094 unità di rilevazione diverse (i Comuni italiani esistenti al 31 dicembre 2010), per elaborare i dati e predisporre il piano di stampa e spedizione dei questionari agli intestatari dei Fogli di famiglia, era necessario ottenere informazioni dettagliate sulla qualità, esaustività e accuratezza delle informazioni anagrafiche nonché verificare le dotazioni tecnico-informatiche di tutte le anagrafi italiane e la capacità di questi ad attenersi a specifici standard tecnici predefiniti dall'Istat.

Per rispondere a queste esigenze conoscitive, l'Istituto ha intrapreso le seguenti attività:

- 1) nell'ambito della Direttiva 2008 TRAD14, Prodotto TRAD14.1.3 relativo alla "Progettazione di soluzioni tecniche e organizzative *ad hoc* per la massimizzazione della copertura degli stranieri nel censimento del 2011" è stata condotta una sperimentazione per prendere visione dei dati anagrafici di 7 Comuni con riferimento ad una partizione degli individui residenti (nati all'estero e stranieri nati in Italia);
- 2) la rilevazione pilota 2009, condotta su 31 Comuni al fine di sperimentare l'introduzione di un protocollo standardizzato per l'invio/ricezione dei dati (formato e nome dei file, tracciato record e modalità di trasmissione/acquisizione dati);
- 3) nell'ambito della Direttiva 2009 TRAD03, Fase 2 "Preparazione del 15° Censimento generale della popolazione con particolare riferimento ai problemi dell'acquisizione per via elettronica di liste anagrafiche comunali (Lac)", ha condotto la Rilevazione sperimentale Lac 2010 per determinare in via definitiva il tracciato record, la classificazione di alcune variabili e l'implementazione dell'applicazione web per la trasmissione dei dati.

2.1.1.1 Dal primo approccio alla rilevazione pilota nel 2009

Il primo contatto con gli uffici di anagrafe dei Comuni è avvenuto nel 2007 quando, nell'ambito della progettazione del Censimento 2011, è stato ipotizzato di ricorrere agli archivi in generale e alle anagrafi in particolare, per migliorare la copertura della rilevazione dei cittadini stranieri, una delle popolazioni più a rischio di sfuggire al censimento. Per tale motivo si è dato avvio ad una sperimentazione che ha coinvolto 7 Comuni italiani (Roma, Firenze, Bologna, Padova, Salerno, Monopoli e Capri). La convenzione stipulata tra l'Istat e i Comuni era finalizzata all'acquisizione dei dati anagrafici riferiti alla popolazione straniera residente e presente (cittadini stranieri e italiani nati all'estero) per realizzare un sistema informativo integrato utilizzando più fonti di dati, al fine di minimizzare il rischio di sotto-copertura di questa sottopopolazione.

Ai Comuni è stato richiesto di fornire i dati sulla popolazione straniera residente e sugli italiani nati all'estero inviando all'Istat, su protocollo protetto, le informazioni in un file di tipo testo che doveva rispettare un tracciato record definito per formato, lunghezza e riclassificazione di alcune variabili.

Dall'acquisizione è emerso che un solo Comune ha rispettato contemporaneamente il formato, il contenuto e il tracciato record richiesti. Per i restanti 6 Comuni è stato necessario intervenire a livello informatico per ripristinare, ove possibile, gli standard richiesti; sono stati inviati, infatti, file in diversi formati (.csv, .xls, .doc, eccetera); in alcuni casi la sequenza delle variabili non è stata rispettata; in molti casi il formato delle date (nascita e iscrizione in anagrafe) era così compromesso da renderne impossibile il ripristino.

Un ulteriore elemento che è emerso dall'acquisizione dei dati dei 7 Comuni è stata la mancanza di uniformità di classificazione per le variabili "relazione di parentela" e "stato civile"; ogni Comune, infatti, utilizzava codici diversi, e con un numero di modalità (e quindi un dettaglio) fortemente variabile.

Le informazioni acquisite con questa prima sperimentazione hanno gettato le basi per definire nuovi standard che potessero agevolare le operazioni di trasmissione/acquisizione. Contemporaneamente, cominciava a prendere forma l'idea di utilizzare i dati anagrafici non solo per sottopopolazioni critiche ma, per l'insieme della popolazione e la Direzione centrale dei Censimenti generali (DCCG) dell'Istat si impegnava nella ricerca di soluzioni appropriate per renderne possibile l'acquisizione.

Tenendo presente le difficoltà incontrate nella gestione dei dati della sperimentazione sugli stranieri, si è cominciato a progettare un nuovo standard per la gestione dei dati anagrafici dei 31 Comuni coinvolti nella rilevazione pilota del 2009; i Comuni del campione per la rilevazione sono stati: Abbiategrasso, Alba, Alghero, Appiano sulla strada del vino, Aprilia, Belmonte Mezzagno, Brindisi, Challand-Saint-Anselme, Città di Castello, Genova, Latina, Marano Principato, Messina, Napoli, Nimis, Novara, Palermo, Parma, Passignano sul Trasimeno, Perugia, Porto Sant'Elpidio, Prato, Roma, Rovereto, Rovigo, Scandicci, Tolentino, Trezzano sul Naviglio, Tricarico e Villaricca.

Con la prima circolare inviata dalla DCCG (7 maggio 2009) ai Comuni è stato richiesto di effettuare le seguenti operazioni:

1. nominare e comunicare i responsabili tecnici della rilevazione;
2. trasmettere all'Istat entro il 30 maggio 2009 le Lac per consentire:
 - la standardizzazione della variabile indirizzo per la postalizzazione dei questionari;
 - l'estrazione del campione di famiglie a cui somministrare il *long form*;
 - la predisposizione del piano di spedizione;
 - la stampa dei questionari.

La Lac di ciascun Comune, riferita al 31 dicembre 2008, doveva contenere l'insieme di individui residenti sia in famiglia che in convivenza. Anche se le convivenze e gli individui in esse residenti non erano oggetto della rilevazione pilota, si è comunque ritenuto opportuno inserire informazioni ad esse legate per simulare nel modo più completo possibile le operazioni di estrazione dei dati per il censimento del 2011.

Questo tipo di approccio ha consentito di acquisire un ulteriore patrimonio informativo rispetto alle unità contenute nelle Lac e alle modalità con cui le anagrafi registrano le informazioni relative ai propri residenti. Inoltre, con la prima circolare sono state fornite ai Comuni tutte le specifiche tecniche necessarie per la predisposizione e l'invio del file.

I dati in formato elettronico dovevano essere inviati alla DCCG attraverso un protocollo protetto e secondo degli standard predefiniti relativi a:

- nome, struttura e formato del file di dati (.csv con separatore “;”);
- caratteristiche, formato e lunghezza delle variabili;
- nome, struttura e formato del file di decodifica per le variabili “relazione di parentela” e “stato civile” se archiviate con codici numerici o stringhe alfabetiche.

Il tracciato record prevedeva 27 variabili ognuna delle quali necessaria a rispondere ad esigenze diverse: l'individuazione delle unità di rilevazione e di analisi (variabili familiari e variabili individuali); la gestione della stampa e dell'invio dei questionari di rilevazione; la verifica della qualità dei dati forniti e il controllo informatizzato dei file pervenuti alla DCCG.

Trattandosi di una sperimentazione che coinvolgeva un limitato numero di Comuni, si è preferito ricorrere ad una procedura, già presente in Istat, per la trasmissione di file contenenti dati personali e sensibili (<https://indata.istat.it/datistat>). La procedura di trasmissione, poco *user-friendly*, ha richiesto un supporto ai Comuni nelle operazioni di invio dei file di dati anagrafici e degli eventuali file di decodifica.

L'applicazione simultanea di tutti gli standard è stata accertata solo per 8 su 31 (25,8 per cento) Comuni; questo dato ha destato non poche preoccupazioni nell'ottica di gestione degli 8.094 Comuni del Censimento 2011; infatti, la proiezione dei Comuni esatti sul totale dei Comuni avrebbe garantito solo 2.089 Comuni con dati formalmente corretti.

Dall'analisi della qualità dei dati inviati, inoltre, sono emerse numerose criticità. Più nel dettaglio, verificando per ciascun individuo (record), sia la completezza dei dati sia la congruenza inter e intra variabili è emerso che, rispetto alle 27 variabili richieste, gli individui con almeno un errore o dato mancante era pari al 17,52 per cento. La distribuzione dei casi è però molto concentrata sui Comuni di Palermo, Parma, Roma e Napoli, che coprono l'87 per cento degli errori riscontrati. Inoltre, il peso percentuale degli errori e dati mancanti è molto variabile tra i vari Comuni; si va da un tasso massimo del 100 per cento nella lista di Alghero, Marano Principato, Palermo e Parma ad un tasso nullo di Belmonte Mezzagno, e Tolentino: Comunque il 50 per cento dei Comuni considerati presenta tassi inferiori all'1 per cento.

I circa 130 mila dati mancanti della variabile “codice fiscale”, concentrati nei Comuni di Roma e Napoli, hanno destato non poche preoccupazioni in quanto l'assenza di questa variabile avrebbe inciso sulla possibilità di compilare ed inviare i questionari con il canale web.

Un altro elemento di riflessione che è emerso riguarda la variabilità, in termini di numero e tipo di modalità, delle variabili “relazione di parentela” e “stato civile” per ogni Comune; in particolare, esaminando i file di decodifica della variabile “relazione di parentela”, sono state riscontrate delle modalità anomale (“deceduto”; “ex-intestatario/a ora deceduto/a”; “ex-intestatario scheda e deceduto”); queste modalità, pur presenti per un numero poco rilevante di casi, segnalavano che oltre al problema della “mancata cancellazione per morte”, le anagrafi disponevano di specifiche modalità per archiviare (e non cancellare) persone decedute.

Si è rilevato, di contro, che le variabili strettamente necessarie all’invio postale dei questionari (cognome, nome, codice famiglia) sono risultate presenti nella quasi totalità degli individui. Questo è un segnale molto confortante per l’uso censuario delle liste anagrafiche, sebbene anche su queste variabili siano stati riscontrati casi relativamente limitati di errori e incongruenze fra variabili.

Infine, sono emerse discrepanze dal confronto fra le numerosità di individui e famiglie presenti nelle Lac al 31 dicembre 2008 e i dati Istat del Bilancio demografico alla stessa data; il divario fra i due tipi di dati, sia per il numero di famiglie che per quello degli individui, è risultato particolarmente accentuato nei Comuni di dimensione medio-grande (Roma, Napoli, Palermo e Latina) dove la revisione anagrafica era meno accurata.

In conclusione, il patrimonio informativo acquisito con i dati delle 31 Lac della rilevazione pilota 2009 è stato certamente rilevante; le modalità di contatto con i Comuni e l’analisi dei dati hanno fatto emergere criticità e punti di forza portando a valutazioni strategiche in ottica censuaria. Sulla base dei risultati ottenuti, si è ritenuto opportuno procedere con una sperimentazione il cui unico obiettivo fosse focalizzato su tutti gli aspetti del processo di trasmissione e acquisizione delle Lac.

2.1.1.2 La rilevazione sperimentale delle Lac nel 2010

La Rilevazione sperimentale del 2010 è stata realizzata, nelle more dell’adozione del Piano generale di Censimento, sotto l’egida dell’Art. 50 del D.L. n. 78 (convertito con modificazioni nella L. 122/2010) del maggio 2010 dove al comma 2 si definisce che *“l’Istat provvede alle iniziative necessarie e urgenti preordinate ad effettuare [...] la predisposizione di liste pre-censuarie di famiglie e convivenze desunte dagli archivi di anagrafi comunali attraverso apposite circolari. Con apposite circolari e nel rispetto della riservatezza, l’Istat stabilisce la tipologia e il formato dei dati individuali nominativi dell’anagrafe della popolazione residente, utili per le operazioni censuarie, che i Comuni devono fornire all’Istat”*.

Si è così potuto progettare e realizzare la sperimentazione che era estremamente importante per la fase iniziale del Censimento della popolazione del 2011.

Gli obiettivi che ci si proponeva di raggiungere con la Rilevazione sperimentale sulle Liste anagrafiche comunali del 2010 erano legati primariamente alla conoscenza anticipata che l’Istat avrebbe ottenuto rispetto alle differenti realtà comunali, all’acquisizione della maggiore quantità possibile di informazioni sui criteri di archiviazione e codifica dei dati comunali e alla verifica della capacità di adeguamento a standard predefiniti nell’ottica di predisporre anticipatamente le procedure necessarie a minimizzare le criticità del processo di acquisizione, analisi e validazione dei dati.

In fase di progettazione si è ritenuto importante consentire al vasto numero di Comuni campione di prendere dimestichezza con le procedure di predisposizione e invio delle liste, consentendo loro di individuare anche le eventuali problematiche interne rispetto a due diversi tipi di errore:

- *errori formali* cioè l’insieme degli errori relativi alla violazione degli standard tecnici richiesti;
- *errori sostanziali* l’insieme delle anomalie presenti nei dati anagrafici e relative a errori di registrazione e, in alcuni casi, al mancato rispetto del regolamento anagrafico.

Per il primo tipo di errore la soluzione scelta fin dall’inizio è stata quella di utilizzare un portale web che acquisisse solo i file formalmente corretti. Per il secondo tipo di errore, invece, è stata attivata sia una campagna di sensibilizzazione degli Uffici di anagrafe con il supporto delle principali Associazioni di Comuni (Anusca, Usci, eccetera), sia il coinvolgimento delle *software house* che si occupavano della gestione degli aspetti tecnico-informatici della gran parte dei Comuni italiani.

2.1.1.3 La scelta dei Comuni campione

L’obiettivo che ci si poneva era di osservare il maggior numero possibile di casi problematici, al fine di valutarne l’incidenza sul totale dei Comuni e di definire le risorse – umane, finanziarie e di tempo – da destinare alla loro soluzione.

In questo senso il campione doveva garantire, oltre ad una determinata accuratezza delle stime dell'incidenza dei diversi problemi riscontrabili, anche una bassa probabilità che problemi con bassa incidenza non venissero mai riscontrati.

La stratificazione del disegno ha tenuto conto della ripartizione geografica, della classe di ampiezza demografica e di una classificazione dei Comuni in base all'importanza delle *software house* affidatarie del servizio di gestione informatica degli archivi anagrafici. Dall'estrazione del campione sono stati inoltre esclusi i Comuni della rilevazione pilota del 2009 e quei Comuni che, notoriamente virtuosi, adempiono sempre con regolarità e correttezza alle richieste di dati da parte dell'Istat; allo stesso modo sono stati inclusi i Comuni notoriamente più problematici o per i quali si era a conoscenza di peculiarità sulle quali era opportuno indagare più approfonditamente. Si è arrivati così alla determinazione di un campione di 1.850 Comuni³⁹.

Per armonizzare e gestire al meglio i diversi aspetti di un'operazione complessa come l'acquisizione delle oltre 8 mila Lac per il Censimento è stata predisposta l'applicazione web Starlac – *Sistema di trasmissione, acquisizione e reportistica delle Lac* (Istat, 2016) per gestire su protocollo protetto tutti gli aspetti del processo di acquisizione dei dati, dalla gestione degli aspetti organizzativi⁴⁰ al monitoraggio dello stato di avanzamento delle attività di invio dei dati e al supporto ai Comuni nelle varie fasi delle operazioni⁴¹, fino al caricamento del file comunale di dati, formalmente corretto, su un Data Base Oracle.

In particolare, Starlac ha garantito l'acquisizione controllata dei dati, fornendo supporto ai Comuni per la trasmissione di dati formalmente esatti⁴² e consentendo all'Istat di restituire informazioni dettagliate sulla qualità dei dati anagrafici inviati.

Fondamentale ai fini dell'individuazione e della gestione delle relative criticità relative al software Starlac e alla predisposizione del file Lac è stato il contatto diretto (telefonico e per email) con i Responsabili Comunali e Referenti tecnici dei Comuni e con i referenti informatici delle *software house* coinvolte.

2.1.2 La prima acquisizione delle Lac per il Censimento 2011

Per poter gestire l'acquisizione dei dati anagrafici di tutti i Comuni italiani e per garantire il trattamento dei dati necessario per l'invio postale dei questionari agli intestatari dei Fogli di famiglia, è stato indispensabile anticipare a gennaio 2011 l'acquisizione dei dati.

Con le Circolari n° 1 del Censimento generale della popolazione e delle abitazioni – Acquisizione delle liste anagrafiche comunali – riportanti, la prima, protocollo n. 7966 e indirizzata al Responsabile dei Servizi demografici del Comune e la seconda, protocollo n. 7967, indirizzata al Responsabile dell'Ufficio di statistica del Comune, entrambe con data 20 dicembre 2010, si è dato avvio alle attività censuarie come previsto dal Programma statistico nazionale 2011-2013 e dall'Art. 50 della legge n. 122 del 30 luglio 2010.

A partire dall'11 gennaio 2011, quindi, l'Istat ha iniziato ad acquisire i dati delle Lac degli 8.094 Comuni italiani, riferite al 31 dicembre 2010. Il termine stabilito per l'invio dei dati è stato fissato al 15 febbraio 2011; ovviamente, trattandosi della prima occasione di acquisizione totale dei dati, attività cruciale ai fini delle operazioni successive, in alcuni casi i dati sono stati acquisiti anche oltre la data stabilita⁴³.

³⁹ Dati gli obiettivi, la numerosità complessiva del campione doveva attestarsi intorno ai 1.500-2.000 comuni. Tale dimensione campionaria è giustificata secondo l'approssimazione di Chernoff, per la quale il campione di numerosità n più bassa tale da assicurare una probabilità $Pr(|p_{est,n}-p| < \epsilon) > (1-\delta)$ per $p \in [a,b]$ è determinato tramite la relazione $n > \ln(2/\delta)/2\epsilon^2$. Usando questa formula si ricava che, per un campione di numerosità $n=2.000$ e per $p \in [0,1]$, eventi con incidenza del 3 per cento hanno una probabilità di almeno il 95 per cento di essere presente nel campione. Secondo la medesima formula, eventi la cui probabilità di occorrenza è pari al 6 per cento hanno probabilità superiore al 95 per cento di essere stimati da valori compresi tra 0,3 e 0,9. È inoltre opportuno osservare che l'approssimazione di Chernoff è nota in letteratura per fornire stime protettive e quindi ci si aspetta che, nella pratica, i risultati campionari siano più accurati delle attese. Se consideriamo la probabilità con la quale eventi che pure occorrono tra gli 8.101 comuni (conteggiate al 31 dicembre 2006) possano rimanere non osservati in un campione di 2 mila di loro, occorre osservare che la più piccola probabilità con la quale un evento può verificarsi tra gli 8.101 comuni italiani corrisponde a $p=1,23 \cdot 10^{-4}$ (cioè 1/8.101). A questo proposito, essendo pari a $P(x=0|n,p)=(1-p)^n$ la probabilità che tale evento non sia mai osservato nel nostro campione è piuttosto elevata e pari al 78 per cento. Tuttavia, sotto l'ipotesi che eventi di tale rarità si verificano in modo tra loro indipendente, risulta inferiore al 5 per cento la probabilità che ne rimangano ignoti esattamente 13 tra loro diversi in un campione di 2 mila comuni.

⁴⁰ Ad esempio, la comunicazione all'Istat dei dati relativi al responsabile comunale della rilevazione.

⁴¹ Ad esempio, tramite la richiesta di riportare la codifica di alcune variabili (quali "relazione di parentela" e "stato civile") alla codifica standard stabilita dall'Istat.

⁴² I dati inviati dai Comuni sono stati considerati acquisiti solo se superavano tutte le soglie di accettabilità stabilite per l'insieme dei controlli volti ad individuare errori, anomalie e dati mancanti inter e intra individui (controlli relativi alla validità del formato e nome del file, validità del formato delle variabili, compatibilità fra variabili, eccetera).

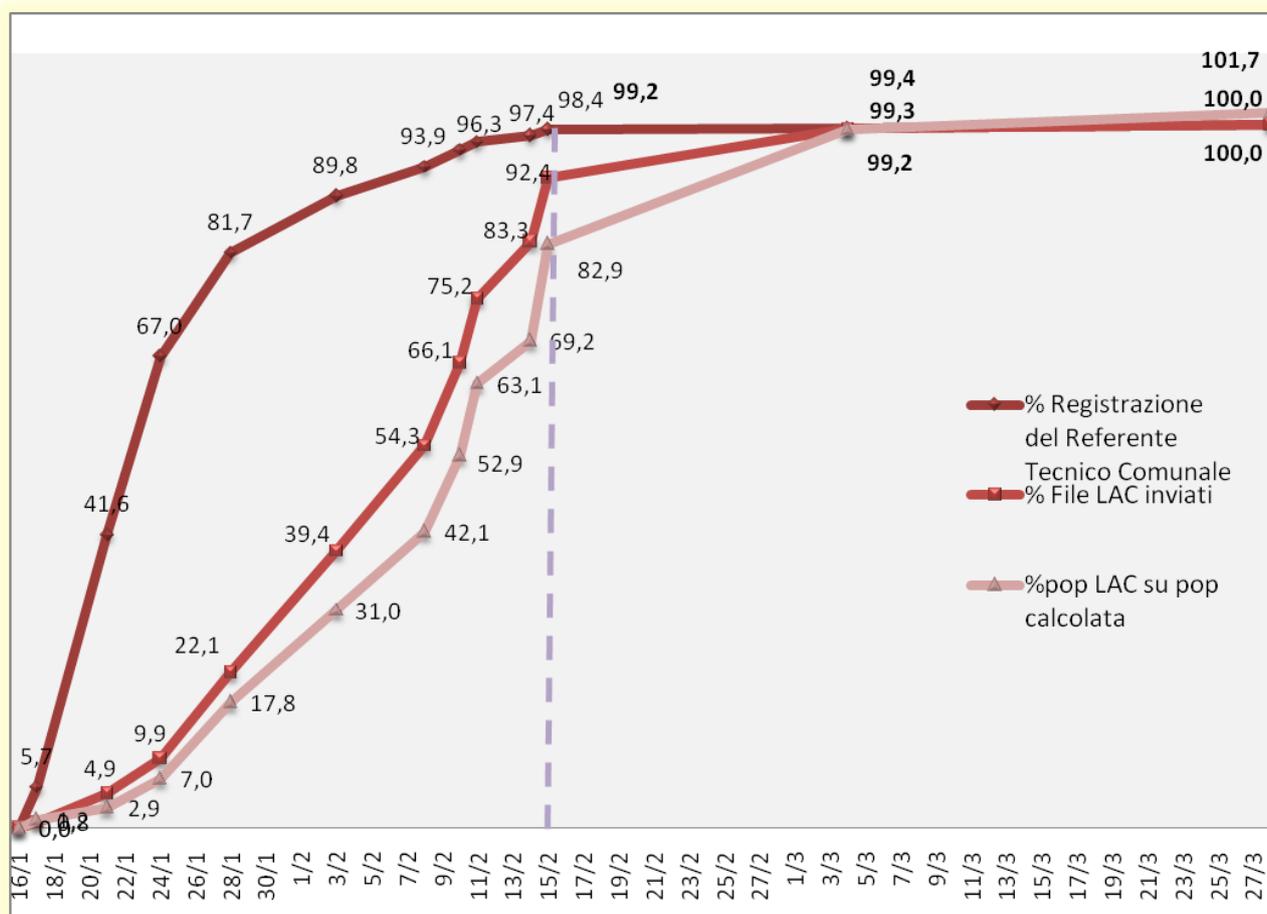
⁴³ Solo per il comune di Ponti (AL), malgrado i numerosi contatti con il Responsabile Comunale e i numerosi solleciti, non è stata acquisita la Lac.

Utilizzando l'applicazione Starlac, migliorata e semplificata in seguito all'esperienza acquisita soprattutto con la rilevazione sperimentale del 2010, è stato richiesto ai Comuni di inviare in formato elettronico i dati relativi a individui, famiglie e convivenze iscritti nell'anagrafe comunale.

Il set di variabili richieste è rimasto inalterato rispetto a quelle della rilevazione sperimentale; l'unica modifica ha riguardato la riclassificazione delle variabili "relazione di parentela" (in particolare si è visto che i dati a disposizione dei Comuni non riuscivano a discriminare il "nipote discendente" dal "nipote collaterale" pertanto si è scelto di inserire la sola modalità "09" = "Nipote senza ulteriori specifiche") e lo "stato civile" (la modalità "separato legalmente" è stata eliminata poiché non comunicata alle anagrafi).

Di seguito è riportato il cronogramma della prima acquisizione delle Lac per il 15° Censimento della popolazione e delle abitazioni (Figura 2.1).

Figura 2.1 – Cronogramma dell'acquisizione delle Lac riferite al 31 dicembre 2010.



Fonte: Direzione centrale dei Censimenti generali (DCCG), Istat.

Dal grafico si può notare che alla data fissata per la chiusura delle attività di invio (15 febbraio 2011):

- il 98,4 per cento dei Referenti tecnici Comunali si erano registrati sull'applicazione Starlac;
- il 94,4 per cento dei file Lac erano stati inviati correttamente;
- l'82,9 per cento degli individui residenti, rispetto alla popolazione Istat-Posas, erano stati acquisiti.

Grazie al coinvolgimento del maggior numero possibile di *software house* nella rilevazione sperimentale del 2010 si è constatato che:

- la quasi totalità dei Comuni ha inviato correttamente i dati entro marzo (oltre questa data solo 69 Comuni hanno inviato la Lac ad aprile e gli ultimi 11 a maggio 2011);

- solo per 77 Comuni è stato necessario chiedere un nuovo invio della Lac perché emergevano delle anomalie dal confronto con il dato Posas sulla popolazione;
- il 93,2 per cento dei Comuni ha effettuato un solo invio dei dati, il 4,3 per cento ha effettuato 2 invii e solo il 2,5 per cento ha effettuato 3 o più invii; questo aspetto in particolare ha fatto capire che le *software house* avevano implementato prontamente applicazioni per la generazione dei file elettronici così come richiesto dall'Istat.

Al fine di monitorare e migliorare la qualità dei dati acquisiti, sono state implementate in Starlac diverse regole per il conteggio degli errori, anomalie e dati mancanti, sia a livello di singoli individui che fra individui appartenenti alla stessa famiglia. Per ognuno di questi controlli è stata stabilita una soglia di accettabilità, calcolata rapportando i casi anomali alla rispettiva popolazione di riferimento (totale individui, totale famiglie o totale convivenze); solo il superamento contemporaneo di tutte le regole di controllo consentiva ai Comuni di completare la procedura di invio e ricevere l'email di conferma dell'invio.

Inoltre, disponendo per ciascun Comune di un'analisi della qualità del dato inviato, è stato scelto di restituire l'informazione ai Responsabili comunali e ai Referenti tecnici comunali direttamente sull'applicazione Starlac. Alla voce Report della schermata Home dell'applicazione Starlac, è stato reso disponibile un "Prospetto di Dettaglio" in cui, per ogni regola, veniva indicato il numero di individui che la violavano. Cliccando sulla regola di errore, anomalia o mancante era inoltre possibile visualizzare gli individui coinvolti con tutte le informazioni fornite dallo stesso Comune. Questa funzione ha consentito ai Comuni di migliorare la qualità dei dati presenti nelle anagrafi, innescando un processo virtuoso nel tempo.

Sebbene i filtri per il controllo dei dati presenti in Starlac fossero numerosi e accurati, si sono resi necessari ulteriori controlli alla fine dell'acquisizione per evitare di appesantire l'utilizzabilità dell'applicazione web; così, l'insieme di controlli di confronto con il dato Posas, di verifica degli indirizzi presenti e la correzione delle famiglie senza intestatario o con più di un intestatario e altri controlli di qualità sono stati inseriti nella procedura Traslac di trattamento dei dati post-acquisizione. Una volta eseguito il controllo e la correzione dei dati acquisiti, questi dovevano essere trattati per conseguire due diverse finalità:

1. l'invio postale dei questionari personalizzati con le credenziali per la compilazione web;
2. il caricamento dei dati in Sgr.

Ai fini della spedizione dei questionari è stato necessario:

- individuare le **unità eleggibili** - famiglie o convivenze per le quali erano presenti e congruenti il tipo di residenza (1 = famiglia, 2 = convivenza) e il codice corrispondente;
- individuare gli intestatari dei Fogli di Famiglia con codice fiscale valido; il codice fiscale, non stampato sul questionario, era la *user ID* per l'accesso alla **compilazione web**, associata alla *password* di primo accesso (generata in modo tale che fosse unica) stampata sul questionario;
- calcolare il **numero dei componenti** per famiglia, così da consentire la selezione del formato del questionario da spedire (da 3 o da 6 componenti);
- normalizzare e geocodificare gli indirizzi per la **postalizzazione dei questionari** e per determinare la numerosità della popolazione all'interno delle sezioni di censimento da sottoporre a campionamento, operazione necessaria per la selezione del campione di famiglie a cui inviare il questionario nella sua forma estesa (*long form*);
- attribuire (a valle dell'estrazione del campione) il **tipo di questionario** da inviare (*short o long*).

La seconda attività, prettamente di tipo informatico, ha gestito il caricamento dei dati acquisiti in Sgr.

2.1.3. La seconda acquisizione delle Lac e il calcolo delle variazioni anagrafiche

Al fine di consentire ai Comuni di effettuare più agevolmente il confronto censimento anagrafe, si è ritenuto opportuno aggiornare i dati presenti in Sgr, riferiti a gennaio 2011, sulla base delle variazioni anagrafiche intercorse fra il 1° gennaio e l'8 ottobre 2011.

Dopo i necessari confronti con i rappresentanti dei Comuni e le principali associazioni (Anusca, Anci, Usci, eccetera), si è deciso di seguire una strategia differenziata tra i Comuni, in base all'ampiezza demografica (Prospetto 2.1), prevedendo due modalità per l'acquisizione delle variazioni:

1. lo scarico di una nuova Lac riferita all'8 ottobre 2011, sulla base della quale l'Istat ha ricavato e caricato in Sgr le variazioni anagrafiche;
2. l'inserimento manuale delle variazioni anagrafiche direttamente in Sgr.

Prospetto 2.1 – Modalità di aggiornamento delle Lac.

Ampiezza demografica	Modalità di aggiornamento	Opzioni
Più di 20.000 ab.	Invio della Lac	Obbligatorio
Fra 5.000 e 20.000 ab.	Invio della Lac o inserimento manuale delle variazioni anagrafiche	Scelta facoltativa
Meno di 5.000 ab.	Inserimento manuale delle variazioni anagrafiche	Obbligatorio

Fonte: Direzione centrale dei Censimenti generali (DCCG), Istat.

Si riporta che, nell'insieme dei Comuni di medie e grandi dimensioni (2.437 Comuni, per un totale di circa 49 milioni di individui), oltre il 90 per cento ha inviato nuovamente la Lac, confermando che il processo di estrazione e trasmissione della Lac tramite l'applicazione Starlac è risultato agevole per gli organi comunali. Il restante 10 per cento di tale insieme, costituito dai Comuni di più piccola dimensione e con un numero di variazioni anagrafiche molto limitato, ha proceduto all'inserimento manuale delle variazioni anagrafiche. Di seguito è riportato il cronogramma della seconda acquisizione dei dati Lac (Figura 2.2).

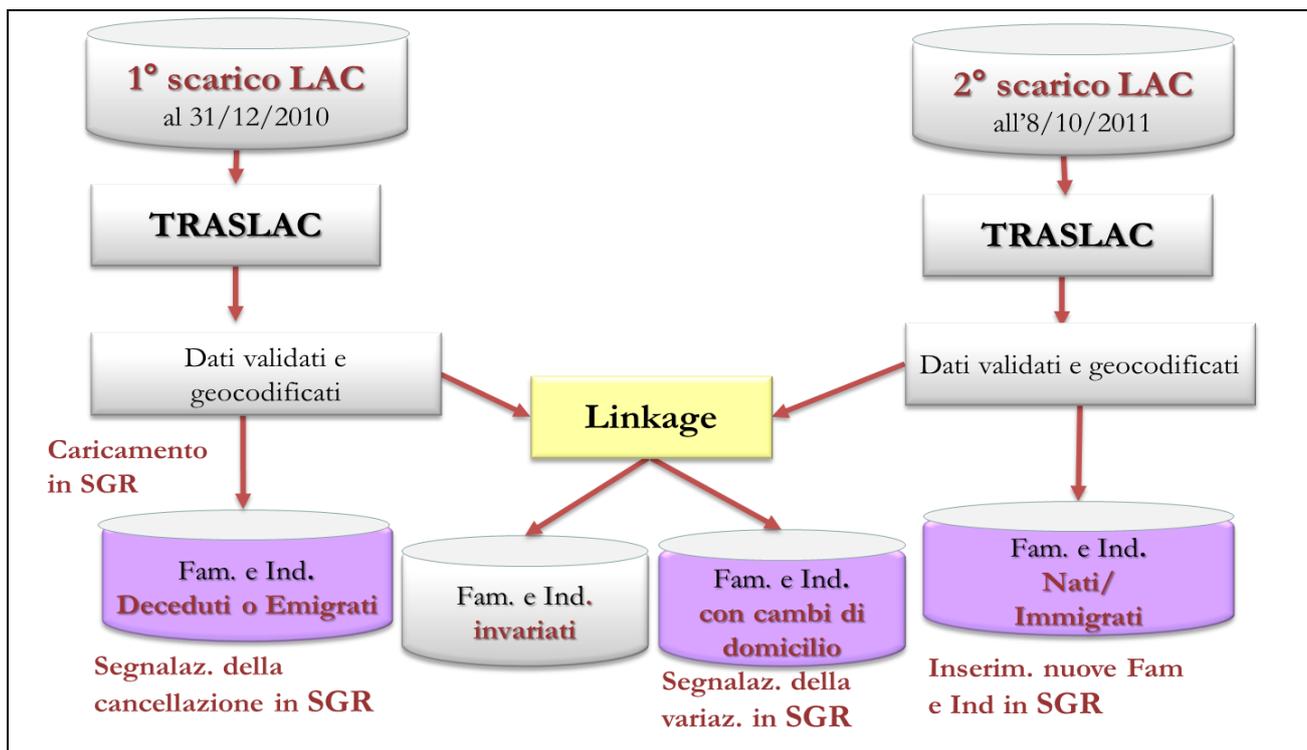
Figura 2.2 – Cronogramma dell'acquisizione delle Lac riferite all'8 ottobre 2010.



Fonte: Direzione centrale dei Censimenti generali (DCCG), Istat.

A differenza della prima acquisizione, è stato previsto un intervallo di sole tre settimane, a partire dal 10 ottobre 2011, per l'invio della Lac aggiornata, per consentire all'Istat di effettuare tutte le attività di controllo e di confronto dei dati sia a livello individuale che familiare prima del caricamento degli aggiornamenti in Sgr (Figura 2.3).

In particolare, man mano che i dati venivano inviati, l'Istat li ha elaborati per blocchi di Comuni. Dopo i controlli operati in Starlac, i dati sono stati ulteriormente verificati con la procedura di controllo e correzione e, prima di procedere al confronto tra le due Lac, sono stati sottoposti al processo di normalizzazione e geocodifica degli indirizzi.

Figura 2.3 – Schema di lavoro per la determinazione delle variazioni anagrafiche fra il 1° gennaio e l'8 ottobre.


Fonte: Direzione centrale dei Censimenti generali (DCCG), Istat.

Le variazioni anagrafiche complessivamente riscontrate al termine delle attività di aggiornamento dei dati in Sgr sono riportate nella Tavola 2.1.

Tavola 2.1 – Variazioni anagrafiche fra il 31 dicembre 2010 e l'8 ottobre 2011.

Tipo di Variazione	Calcolate		Inserite manualmente	
	Famiglie (migliaia)	Individui (migliaia)	Famiglie (migliaia)	Individui (migliaia)
Nuove famiglie	663	1.110	138	242
Famiglie con almeno un componente aggiunto	356	856	194	263
Famiglie con almeno un componente in meno	948	1.200	188	227
Famiglie che hanno cambiato domicilio	233	380	33	81
Famiglie cancellate	487	744	124	193
Cambi di intestatario	140	264	65	126

Fonte: Direzione centrale dei Censimenti generali (DCCG), Istat.

2.2 Produzione e uso della Lista integrativa da fonti ausiliarie (Lifa)

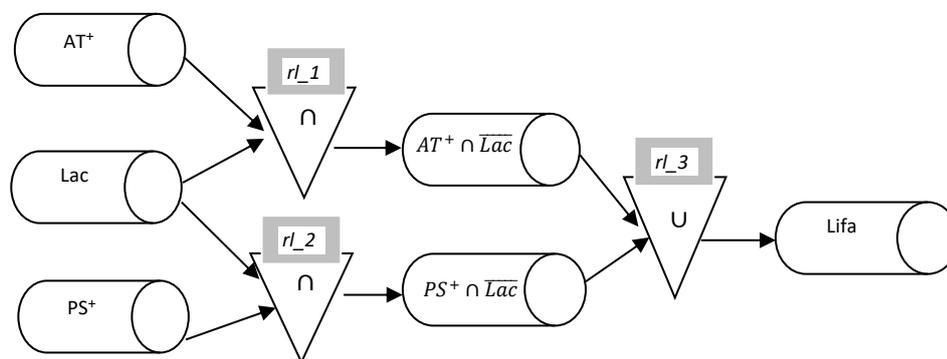
Com'è noto, pur in presenza di Lac tempestivamente aggiornate a seguito di trasferimenti di residenza da o verso un altro Comune (italiano o estero), la popolazione abitualmente dimorante raramente coincide con quella anagraficamente residente nel territorio. Può accadere, infatti, che un numero più o meno consistente di individui e/o famiglie viva stabilmente in un Comune senza averne acquisito la residenza anagrafica. Per converso, la Lac di un Comune può contenere un numero non trascurabile di record relativi a individui e/o famiglie che vivono stabilmente o abitualmente in un altro Comune (italiano o estero).

Le ragioni della mancata sovrapposizione tra le due popolazioni possono essere molteplici. Ad esempio, la precarietà occupazionale diffusa non favorisce il trasferimento della residenza anagrafica presso il Comune dove ci si stabilisce per motivi di lavoro. Per i cittadini stranieri, inoltre, l'iter burocratico di "iscrizione nel" o "cancellazione dal" registro anagrafico tende a scoraggiare ulteriormente il richiedente. Ma anche per i cittadini italiani con un lavoro stabile esiste l'incentivo a mantenere o trasferire la residenza anagrafica in un Comune pur vivendo stabilmente o abitualmente altrove per usufruire, ad esempio, di vantaggi di natura fiscale.

In previsione dei potenziali errori di copertura da cui un censimento assistito da lista avrebbe potuto essere affetto a causa di liste di individui obsolete e/o incomplete, l'Istat in occasione del censimento 2011 ha deciso di affiancare alle Lac liste integrative desunte da altre fonti amministrative. Le Liste integrative da fonti ausiliarie (Lifa) sono state popolate, per ciascun Comune italiano, con record individuali contenenti segnali di presenza alternativi o aggiuntivi rispetto a quelli esistenti nelle liste anagrafiche. Tali segnali sono stati forniti ai rilevatori come ausilio nella fase di ritorno sul campo, al fine di censire nuovi individui o famiglie "non coperti" dalla prima fase della rilevazione guidata esclusivamente dalle Lac.

La Lifa nasce dall'integrazione di due archivi: l'Anagrafe tributaria (AT) per i cittadini italiani e comunitari e i Permessi di soggiorno (PS) per i cittadini stranieri non comunitari, secondo lo schema riportato nella Figura 2.4. Gli archivi AT^{+44} e PS^{+45} , ottenuti dopo aver ripulito, validato e de-duplicato i record contenuti nelle fonti originarie, sono stati abbinati ciascuno in modo indipendente con le Lac. L'abbinamento tra AT^{+} e Lac e tra PS^{+} e Lac sono indicati nella Figura 2.4 rispettivamente con rl_1 e rl_2 . L'abbinamento tra PS^{+} e Lac (rl_2)⁴⁶ e la creazione dell'archivio Lifa (rl_3) dall'integrazione tra i record PS^{+} e AT^{+} non abbinati, in quanto portatori di segnali di presenza sul territorio che non trovano corrispondenza nelle rispettive Lac, sono descritti nel paragrafo 2.2.3.

Figura 2.4 – Schema sintetico di produzione della Lifa.



Fonte: Direzione centrale dei Censimenti generali (DCCG), Istat.

2.2.1 L'archivio dei Permessi di Soggiorno (PS)

L'archivio dei permessi di soggiorno utilizzato per la produzione delle Lifa è stato ottenuto grazie all'utilizzo congiunto di due fonti di tipo amministrativo entrambe facenti capo al Ministero dell'Interno: la banca dati del Dipartimento per la pubblica sicurezza e a quella del Dipartimento per le libertà civili⁴⁷.

⁴⁴ L'Anagrafe tributaria arricchita (AT^{+}) è stata ottenuta da un confronto di AT con altri archivi minori tra cui l'Anagrafe degli studenti, il Casellario dei pensionati, l'Archivio dei lavoratori dipendenti (mod. 770), l'Archivio dei lavoratori autonomi (mod. 770), il Registro statistico delle imprese attive. Si veda Mancini *et al.* (2011) per uno studio preliminare sull'efficacia delle Lifa a fini censuari basato su una simulazione su dati reali provenienti dall'indagine pilota del 2009.

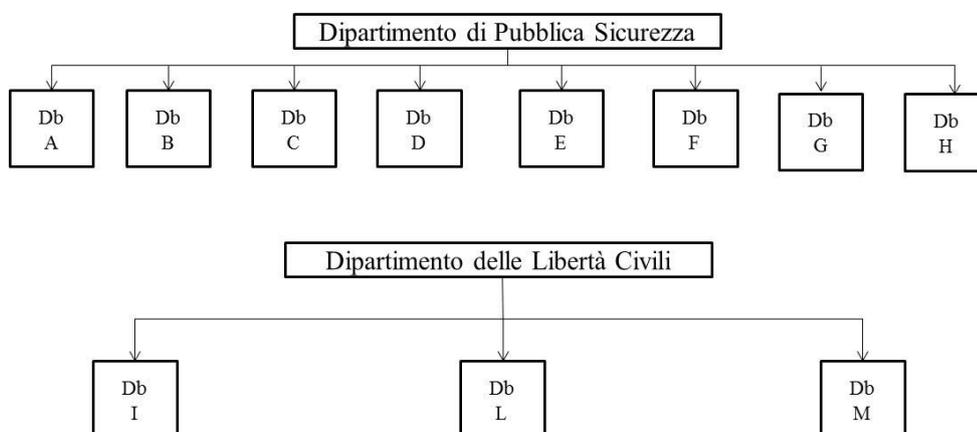
⁴⁵ L'archivio PS^{+} è stato ottenuto dopo le operazioni di de-duplicazione descritte nel paragrafo 2.2.2.

⁴⁶ A differenza dell'operazione rl_1 tra AT^{+} e Lac che ha riguardato tutti i record contenuti nelle liste anagrafiche comunali, l'operazione rl_2 è circoscritta ai soli record Lac relativi ai cittadini stranieri non comunitari.

⁴⁷ I dati contenuti nelle due fonti si riferiscono alla data di estrazione: novembre 2010.

All'interno di ciascuna di queste fonti le informazioni individuali sono archiviate in diversi data base origine a seconda della natura e della fase in cui si trova il processo relativo a ciascun individuo. Uno schema della struttura complessiva delle due fonti è rappresentata nella Figura 2.5.

Figura 2.5 – Struttura originale dell'archivio dei permessi di soggiorno. Fonti e data base origine.



Fonte: Fortini et al. (2011).

Dal successivo Prospetto 2.2 si evince che ciascuno degli undici data base origine ha una propria numerosità nonché un proprio contenuto informativo. Tuttavia, come illustrato in dettaglio nel paragrafo 2.2.2, tra i diversi data base origine erano molto frequenti aree di sovrapposizione. Ad uno stadio iniziale l'archivio dei permessi di soggiorno consisteva in un insieme di data base origine parzialmente sovrapposti; ciò in ragione sia di cause fisiologiche (uno stesso record è presente in due data base origine diversi ma con contenuto informativo simile, ad esempio H ed L) ma anche perché ottenere l'acquisizione di un permesso di soggiorno è spesso un processo lungo che porta inevitabilmente il richiedente a transitare per una serie di status successivi.

Prospetto 2.2 – Contenuto informativo, numerosità e fonte (origine) di ciascun data base costituente l'archivio dei Permessi di soggiorno (PS).

Data base	Contenuto informativo	Numerosità	Fonte ⁽¹⁾
A	Nuove richieste di PS in stato di pre-trattamento ⁽²⁾	185.000	Dps
B	PS validi	2.600.000	Dps
C	Richieste di PS ancora in esame	500.000	Dps
D	Richieste rifiutate	9.500	Dps
E	PS scaduti	755.000	Dps
F	PS di minori registrati PS dei genitori	39.000	Dps
G	PS per lavoro rilasciati per la prima volta	57.000	Dps
H	PS rilasciati per motivi di ricongiungimento familiare	235.000	Dps
I	Richieste di regolarizzazione	250.000	Dlc
L	Richieste PS per ricongiungimenti familiari	131.000	Dlc
M	PS per motivi di lavoro: estensione temporale/richiesta di rinnovo	122.000	Dlc

⁽¹⁾ Dps = Dipartimento di pubblica sicurezza; Dlc = Dipartimento per le libertà civili.

⁽²⁾ Questa è una prima fase di *screening* in cui si verifica la completezza delle informazioni fornite.

Fonte: Direzione centrale dei Censimenti generali (DCCG), Istat.

Per la costruzione dell'archivio unico dei permessi di soggiorno, ottenuto come unione dei data base origine, è stato necessario effettuare alcune operazioni di pulizia e pre-processamento dei dati grezzi. In queste operazioni, particolare attenzione è stata dedicata all'individuazione, al conteggio e, laddove possibile, al trattamento dei record con codice fiscale mancante o incompleto. Nel Prospetto 2.3 sono riportate alcune informazioni sulla numerosità assoluta e relativa dei record distinti per data base di origine con codice fiscale mancante o incompleto e sprovvisti di indirizzo stradale. Per il reperimento delle informazioni mancanti o

incomplete sono state utilizzate diverse strategie tra cui l'incrocio con archivi di tipo tributario e amministrativo e la donazione da record duplicato, che verrà illustrata nel paragrafo seguente.

È utile ricordare che il lavoro svolto sull'archivio PS non è servito solo per la costruzione delle Lifa ma anche per la predisposizione di pacchetti informativi ad uso dei grandi Comuni (≥ 150 mila residenti), contenenti informazioni sugli stranieri a livello micro territoriale (sezioni di censimento) utili ad una migliore rilevazione sia della componente stabile (stranieri iscritti in anagrafe) sia della componente semi stabile (stranieri con permesso di soggiorno ma non iscritti in anagrafe) della popolazione straniera (Gallo *et al.*, 2014).

Prospetto 2.3 – Record con codice fiscale (CF) mancante o incompleto. Valori assoluti e percentuali per ciascun data base origine.

Dipartimento di pubblica sicurezza	Numerosità	CF mancante o incompleto ^(a)	Percentuale di record con CF mancante o incompleto
Nuove richieste di PS in stato di pre-trattamento	185.000	27.893	15,1
PS validi	2.600.000	278.528	10,7
Richieste di PS ancora in esame	500.000	95.179	19,0
Richieste rifiutate	9.500	2.193	23,1
PS scaduti	755.000	137.565	18,2
PS di minori registrati sul PS dei genitori	39.000	3.174	8,1
PS per lavoro rilasciati per la prima volta	57.000	10.064	17,6
PS rilasciati per motivi di ricongiungimento familiare	235.000	6	0,0
Dipartimento delle libertà civili			
Richieste di regolarizzazione	250.000	61.502	24,6
Richieste di PS per ricongiungimenti familiari	131.000	0	0,0
PS per motivi di lavoro: estensione temporale/riciesta di rinnovo	122.000	0	0,0
Totale complessivo	4.969.000	616.104	12,4
<i>di cui con indirizzo</i>	<i>4.918.895</i>	<i>612.930</i>	<i>12,5</i>

^(a) Incompleti sono tutti quei casi in cui il codice fiscale è parziale o rappresentato da un codice numerico.

Fonte: Direzione centrale dei Censimenti generali (DCCG), Istat.

2.2.2 L'identificazione ed eliminazione dei record PS duplicati

L'esistenza di più record riconducibili ad un medesimo individuo nell'insieme delle banche dati PS è in parte dovuta al fatto che due dipartimenti (Dps – Dipartimento di pubblica sicurezza e Dlc – Dipartimento per le libertà civili) dello stesso dicastero registrano e mantengono, per finalità diverse, le stesse informazioni sulla presenza della popolazione straniera in Italia come avviene, ad esempio, per le richieste di permesso di soggiorno per ricongiungimento familiare. Tuttavia, la maggior parte delle duplicazioni riguarda individui che transitano attraverso due o più archivi durante l'arco temporale necessario per ottenere un permesso di soggiorno definitivo: ad esempio (Figura 2.5), un richiedente può emergere da uno stato di clandestinità (Db_I), rimanere sotto scrutinio per un determinato periodo (Db_C), ottenere un primo permesso di lavoro di breve durata (Db_G) fino a conseguire un titolo a soggiornare progressivamente più permanente (Db_B). In altre parole, l'ipotetico individuo in questione è presente quattro volte nell'archivio consolidato.

La presenza di duplicati nelle banche dati PS complica le fasi d'integrazione tra archivi schematizzate nella Figura 2.4. Per tale ragione, l'individuazione dei record duplicati e la loro successiva de-duplicazione è un'operazione preliminare necessaria per la creazione di un archivio ausiliario a supporto della rilevazione censuaria come le Lifa.

Sebbene il codice fiscale sia senza dubbio la chiave più potente e affidabile per identificare univocamente un individuo all'interno di un sistema di archivi amministrativi, è altrettanto noto che per i cittadini stranieri il codice fiscale non offre le stesse garanzie di efficacia e affidabilità. Come evidenziato nel paragrafo 2.2.1, si riscontrano all'interno delle banche dati PS numerosi casi di record con codice fiscale errato, incompleto o del tutto mancante. Di conseguenza, non è possibile abbinare in modo esatto, ovvero senza gradi di incertezza, due o più record relativi allo stesso individuo sulla base di una chiave identificativa univoca. Diventa, pertanto, necessario non solo utilizzare congiuntamente altre chiavi più complete seppure meno discriminanti, ma anche accettare la possibilità che, in presenza di chiavi d'aggancio comunque affette da errore, una certa percentuale di *link* trovati non si riferisca in realtà allo stesso individuo e, di contro, che esista un certo numero di record

che, sebbene riconducibili allo stesso individuo, non vengono riconosciuti come tali (falsi *link* positivi e negativi). Un operatore addestrato è in grado di giudicare se un *link* sia vero o falso attraverso un rapido esame delle concordanze e delle discordanze sulle chiavi d'aggancio utilizzate. Tuttavia, all'aumentare del numero di record da confrontare è necessario automatizzare tale scrutinio definendo delle opportune regole di decisione. L'utilizzo di regole probabilistiche che tengono conto sia della qualità delle chiavi disponibili sia del loro potere discriminante ha il pregio della parsimonia e della riduzione del controllo manuale rispetto a regole di tipo deterministico.

La ricerca dei record PS duplicati, così come quella della successiva integrazione tra gli archivi (paragrafo 2.2.3), ha utilizzato un abbinamento di tipo probabilistico basato sull'approccio formalizzato da Fellegi e Sunter (1969) e implementato con l'ausilio del software Relais 2.2 (Fortini *et al.*, 2011). Per circoscrivere il confronto tra i record per individuare quelli tra loro sufficientemente simili, si è ricorso sia a criteri di ordinamento sul nome e cognome, sia a tecniche di bloccaggio sulla nazionalità degli individui. Le chiavi di abbinamento considerate sono state l'anno, il mese e il giorno di nascita, il nominativo e il codice fiscale. In virtù della possibilità di abbinamenti multipli i *link* trovati sono stati rappresentati sotto forma di grappoli utilizzando la teoria dei grafi.

Successivamente, per decidere quale record di ciascun grappolo (individuo) mantenere nell'archivio de-duplicato si è scelto di privilegiare quello con data di movimentazione più recente e, a parità di data, quello proveniente dalla banca dati ritenuta più affidabile sulla base dell'analisi preliminare discussa nel corso del presente paragrafo. Nei casi in cui il codice fiscale del record sopravvivente era incompleto o mancante oppure l'indirizzo non era geocodificabile ad una sezione di censimento si è ricorso, ove disponibile, ad un "donatore" scelto tra i record eliminati con data di movimentazione più recente.

I risultati sono descritti nel Prospetto 2.4; oltre al numero complessivo di record duplicati (ultima riga) sono riportati i risultati relativi ad alcune nazionalità. In particolare, la colonna (1) riporta il numero di record presenti nelle banche dati PS⁴⁸, (2) mostra il numero di duplicati individuati dal *record linkage*, (3) riporta il valore aggiunto della strategia probabilistica rispetto ad un abbinamento esatto su codice fiscale, (4) e (5) mostrano rispettivamente i record eliminati applicando le regole gerarchiche di de-duplicazione, mentre (6) riporta il numero di record sopravviventi che formano l'archivio PS⁺. Infine, le ultime due colonne mostrano rispettivamente la percentuale residua di record de-duplicati senza codice fiscale completo (7) e/o senza indirizzo geocodificato (8) dopo l'imputazione da donatore. I record eliminati sulla base delle regole sono oltre 530 mila e dunque PS⁺ contiene oltre 3,32 milioni di record rispetto ai circa 3,86 milioni di partenza.

Prospetto 2.4 – I risultati sull'identificazione ed eliminazione dei duplicati.

Paese	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Albania	441.311	81.543	3,6	41.263	201	399.847	5,4	8,4
Bangladesh	111.948	37.712	4,8	20.096	62	91.790	10,1	5,5
Cina	255.772	57.413	10,2	29.102	75	226.595	6,4	6,8
Ecuador	87.298	22.004	11,8	11.147	38	76.113	4,1	8,3
India	162.480	68.510	17,6	35.140	75	127.265	11,4	10,7
Macedonia	70.314	15.147	15,8	7.640	37	62.637	7,8	11,0
Tunisia	107.824	24.818	21,3	12.493	106	95.225	19,0	11,3
Ucraina	269.966	113.663	8,2	57.767	106	212.093	8,2	8,9
.....
Totale	3.859.869	1.043.317	15,0	529.225	2.825	3.327.819	12,5	8,5

Fonte: Direzione centrale dei Censimenti generali (DCCG), Istat.

2.2.3 L'integrazione dei record di stranieri con permesso di soggiorno con le Lac e la costruzione delle Lifa

L'obiettivo della Lifa era costituire una lista ausiliaria di individui (corredata ovviamente di indirizzi validi dove poterli rintracciare) per i quali esistessero segnali di presenza sul territorio alternativi e/o complementari alla residenza anagrafica.

⁴⁸ L'archivio PS sottoposto a de-duplicazione è quello ottenuto a seguito della fase di pulizia e pre-processamento delle banche dati origine discusso nel paragrafo 2.2.1. Tra i record esclusi ci sono, ad esempio, tutti quelli provenienti dalla banca dati E dei permessi scaduti.

Poiché solitamente l'indirizzo indicato da un individuo all'atto d'iscrizione in un generico archivio amministrativo coincide con la sua residenza anagrafica, occorre integrare la lista ausiliaria con le Lac per eliminare dalla prima tutti quei record (la maggior parte) presenti in entrambe allo stesso indirizzo. In altre parole, occorre mantenere nella lista ausiliaria soltanto i record relativi alle seguenti tipologie di individui:

- iscritti in anagrafe nello stesso Comune ma ad un indirizzo diverso (Lifa_1);
- non iscritti in alcuna anagrafe nazionale (Lifa_3);
- iscritti all'anagrafe di un altro Comune (Lifa_2).

Mentre i record inseriti nelle sotto-liste Lifa_2 e Lifa_3 erano uno strumento di supporto alla rilevazione nella fase di recupero sul campo della sotto-copertura delle Lac (cfr. paragrafo 4.4), la sotto-lista Lifa_1 è stata utilizzata per visualizzare in Sgr indirizzi alternativi a quello di iscrizione anagrafica, informazione utile all'Ucc in fase di confronto censimento-anagrafe.

L'integrazione tra PS⁺ e Lac (indicata con *rl_2* nella Figura 2.4) è articolata in due fasi: una prima fase deterministica basata su una semplice operazione di *merge* su codice fiscale tra i record per i quali tale chiave risultava completa; una seconda fase probabilistica tra i record (residui) non abbinati nella fase precedente⁴⁹, utilizzando chiavi di aggancio alternative al codice fiscale (tra cui nome, cognome e data di nascita dell'individuo).

Dato l'elevato numero di record coinvolti in questa seconda fase, è stato necessario per l'abbinamento procedere alla ricerca per "blocchi", dove i blocchi erano rappresentati generalmente dalle singole Regioni e, per le aree più densamente popolate come quelle metropolitane, dalle singole Province o addirittura dal singolo Comune. Gli esiti dell'operazione *rl_2* sono riportati nella Tavola 2.2.

Oltre la metà dei record PS⁺ inseriti nelle Lifa si riferisce ad individui non presenti in alcuna Lac (Lifa_3). Ciò conferma come una quota rilevante della popolazione straniera in possesso di un regolare permesso di soggiorno e dunque avente, potenzialmente, la propria dimora abituale in un Comune italiano, non lasci traccia di sé nei registri anagrafici.

Tavola 2.2 – L'esito del record linkage tra PS+ e Lac e l'integrazione con AT+.

Tipo lista	rl_2			rl_3				
	Fase (1)	Fase (2)	PS ⁺ ∩ LAC	In entrambi	Lifa	Di cui da:	PS	AT
Lifa_1	336.252	50.923	387.175	2.386	1.445.730		386.706	1.059.024
Lifa_2	255.439	23.416	278.855	1.179	739.794		278.837	460.957
Lifa_3	-	772.470	772.470	2.437	1.363.978		772.470	591.508
Totale Lifa	591.691	846.809	1.438.500	6.013	3.549.502		1.438.013	2.111.489

Fonte: Direzione centrale dei Censimenti generali (DCCG), Istat.

La Tavola 2.2 riporta, inoltre, la numerosità e la composizione delle liste Lifa definitive ottenute come integrazione finale (*rl_3*) degli archivi intermedi prodotti rispettivamente nelle operazioni di *record linkage* di AT⁺ (*rl_1*) e PS⁺ (*rl_2*) con le Lac. La sovrapposizione tra i due archivi (circostrita a poco più di 6 mila record) è stata eliminata privilegiando la fonte il cui record presentasse un indirizzo geocodificato ad una sezione di censimento o, a parità di qualità della geocodifica, mantenendo il record AT⁺, qualora la data di variazione del domicilio fiscale fosse più recente rispetto a quella di movimentazione/richiesta del permesso di soggiorno, o il record PS⁺ nel caso contrario.

Se si esclude la sotto-lista Lifa_1 che, come detto, è stata utilizzata per finalità diverse dal recupero della sotto-copertura delle Lac, l'archivio Lifa così come originariamente concepito contiene poco più di 2,1 milioni di record la cui provenienza è equamente ripartita tra PS (per i record relativi ai cittadini stranieri non comunitari) e AT (per quelli riguardanti i cittadini italiani o gli stranieri comunitari).

⁴⁹ Nella fase (2) si è deciso di coinvolgere, oltre ai record PS⁺ esclusi dalla fase (1) in quanto sprovvisti di codice fiscale completo, anche i record PS⁺ con codice fiscale completo non abbinati nella fase (1) con l'intento di recuperare eventuali *link* genuini (falsi negativi) sfuggiti a causa della presenza di possibili errori sul codice fiscale.

2.2.4 L'utilizzo e il rendimento della Lifa nella fase di recupero della sotto-copertura delle Lac durante il Censimento 2011

La Lifa è stata concepita come uno strumento di ausilio alla rilevazione per il recupero mirato degli individui non anagraficamente residenti ma con segnali amministrativi di presenza sul territorio. Una valutazione complessiva dell'efficacia delle Lifa ai fini del recupero di popolazione eleggibile ma assente dai registri anagrafici è resa difficile da una serie di fattori, primo fra tutti il parziale o mancato utilizzo delle stesse da parte della rete di rilevazione. In base ai dati provenienti dal Sistema di gestione della rilevazione, poco più del 3 per cento della sotto-copertura delle Lac è stata individuata grazie alle Lifa.

Gli Uffici comunali di censimento hanno spesso segnalato all'Istat liste ausiliarie ridondanti o non utili, cioè contenenti individui che erano già stati censiti o che non erano presenti all'indirizzo Lifa. Ciò è dovuto in parte ad errori di mancato abbinamento tra i record degli archivi ausiliari e le Lac, e in parte a segnali di presenza già obsoleti rispetto alla data del censimento. A causa di questi inconvenienti è ragionevole pensare che, dopo le prime verifiche sul campo, la rete di rilevazione abbia sottoutilizzato le Lifa, considerandole fonte di aggravio per le operazioni sul campo. Tuttavia, è comunque possibile valutarne l'efficacia a prescindere dal loro effettivo utilizzo da parte dei rilevatori.

Ad esempio, è interessante conoscere la percentuale di individui non iscritti nelle Lac ma presenti nelle Lifa del Comune allo stesso indirizzo presso il quale sono stati censiti. Si tratta di un indicatore di rendimento potenziale (o teorico) delle liste ausiliarie, qualora fossero state utilizzate come canale prioritario di recupero della sotto-copertura anagrafica.

La Tavola 2.3 presenta i risultati di tale valutazione circoscritta ai cittadini stranieri extra-comunitari. Questa sottopopolazione, notoriamente tra le più elusive da censire, rappresenta il bacino "naturale" su cui verificare l'efficacia dei segnali provenienti dall'archivio dei Permessi di soggiorno a fini censuari. La colonna (1) fa riferimento alla sotto-copertura delle Lac ufficialmente riconosciuta in sede di validazione della popolazione legale 2011, mentre la (2) rettifica al ribasso tale conteggio a seguito di verifiche post-censuarie.

Tavola 2.3 – Efficacia potenziale ed effettiva della Lifa nel recupero della sotto-copertura delle Lac.

Censiti all'indirizzo Lifa e non iscritti in Lac ^(a)		
Tipo lista	(1)	(2)
Lifa_1	822	245
Lifa_2	1.725	1.608
Lifa_3	16.718	15.526
Totale recuperi potenziali da Lifa	19.267	17.379
Totale recuperi Lifa effettivi	4.543	4.157
Totale censiti non iscritti	139.731	119.545
Indicatore di efficacia effettiva (%)	3,2	3,5
Indicatore di efficacia potenziale (%)	13,8	14,5

^(a) Ci sono complessivamente 1.346.628 segnali di presenza relativi a cittadini stranieri non comunitari in Lifa PS⁺.

Fonte: Direzione centrale dei Censimenti generali (DCCG), Istat.

I risultati confermano che poco più del 3 per cento degli stranieri non comunitari censiti sono stati effettivamente rilevati con l'ausilio dei segnali Lifa. Tuttavia, i risultati sono meno deludenti quando si considerano i tassi di recupero potenziale, cioè la percentuale di individui che si sarebbero potuti censire all'indirizzo Lifa qualora questo canale di recupero fosse stato effettivamente utilizzato. Sia che si consideri la sotto-copertura ufficiale (1) che quella confermata a seguito della revisione anagrafica (2), il rendimento delle Lifa sale al 14 per cento ovvero ad un livello quasi cinque volte superiore a quello calcolato in base al loro effettivo utilizzo durante il censimento. Questi risultati appaiono più in linea con quelli ottenuti dalla sperimentazione effettuate su quattro Comuni (Abbiategrasso, Genova, Prato e Scandicci) in occasione dell'indagine pilota dell'ottobre 2009, quando furono testate le potenzialità di un prototipo di Lifa basato su archivi aggiornati a giugno 2009⁵⁰.

⁵⁰ I risultati della sperimentazione 2009 si riferiscono a tutti i censiti non iscritti (cittadini italiani, stranieri comunitari e stranieri extracomunitari) e dunque non sono direttamente confrontabili con quelli del 2011 (Mancini *et al.*, 2011).

Il forte divario tra rendimento effettivo e potenziale non implica, tuttavia, che le Lifa debbano essere giudicate in base al secondo. Infatti tale divario è da imputare, almeno in parte, ai quei problemi di ridondanza cui si è fatto cenno all’inizio del paragrafo, che ne hanno certamente intaccato l’efficacia rendendole impopolari presso gli operatori della rete di rilevazione chiamati a utilizzarle. Anche alla luce della sperimentazione del 2009, l’utilizzo di archivi amministrativi aggiornati appare imperativo per il futuro sviluppo di questo strumento a fini censuari.

2.3. Produzione e uso della lista degli interni potenzialmente abitativi

Durante il Censimento 2011, sono state messe a disposizione dei rilevatori informazioni aggiuntive provenienti da liste ausiliarie di individui (Lifa) e da liste di interni potenzialmente abitativi. L’obiettivo di fornire questi dati è stato quello di coadiuvare il rilevatore nella ricerca delle famiglie da intervistare al fine di agevolare il recupero mirato della sotto-copertura delle Lac. In questo paragrafo l’attenzione viene posta sulla gestione dell’informazione relativa agli interni potenzialmente abitativi e in particolare verrà descritto il processo di popolamento del Base dati e il flusso di lavorazione previsto per l’utilizzo della Rilevazione sui numeri civici (Rnc) ai fini del recupero della sotto-copertura.

La Rnc ha permesso ai Comuni di effettuare, attraverso l’utilizzo di un sistema web, la revisione dello stradario e della numerazione civica, e di rilevare le caratteristiche degli edifici, dato fondamentale per reperire le unità potenzialmente abitative.

Questo lavoro preliminare è stato richiesto solo ai Comuni più grandi, ossia quelli con almeno 20 mila abitanti, e ai Comuni capoluogo di Provincia; solo in questi Comuni, dunque, è stato possibile effettuare il recupero mirato della sotto-copertura (cfr. paragrafo 4.4). L’aggiornamento degli stradari ha garantito una lista di indirizzi aggiornata e controllata direttamente sul territorio e ha consentito di avere, per ciascun indirizzo (ovvero per ciascun numero civico presente sul territorio), il dato sul numero di interni abitativi. Questa lista di indirizzi, normalizzata e pulita, è stata successivamente linkata con i dati presenti nelle Liste anagrafiche comunali. Per effettuare tale confronto è stato tuttavia necessario effettuare anche la normalizzazione degli indirizzi delle anagrafi. Il software utilizzato a questo scopo (Egon)⁵¹ ha permesso di scomporre l’indirizzo in sezione di censimento, Denominazione urbanistica generica (Dug), toponimo, civico ed esponente, e ha fornito indicatori di valutazione della qualità del processo di normalizzazione. Gli indirizzi per i quali non è stato possibile effettuare la normalizzazione e quelli per i quali la qualità della geocodifica (assegnazione della sezione di censimento) era insufficiente non sono stati sottoposti al *linkage* con gli indirizzi provenienti da Rnc.

Ai fini del *linkage* tra i due archivi, si è deciso di utilizzare le informazioni relative a sezione di censimento, Dug, toponimo e civico, e di tralasciare l’esponente in quanto, grazie ad una serie di simulazioni condotte in fase di analisi preliminare, avrebbe condotto alla sovrastima della sotto-copertura delle Lac, a causa della frequente gestione non corretta dell’esponente nelle liste anagrafiche. Il *linkage* è stato di tipo deterministico: si è cercata l’esatta uguaglianza tra tutti e quattro i campi nelle due fonti. Il risultato ha evidenziato le seguenti possibilità:

- a) civico presente sia nella tabella Rnc che nelle Lac con numero di famiglie corrispondente con il numero di interni potenzialmente abitativi;
- b) civico presente sia nella tabella Rnc che nella Lac con numero di famiglie maggiore rispetto al numero di interni potenzialmente abitativi;
- c) civico presente sia nella tabella Rnc che nelle Lac con numero di famiglie della Lac minore rispetto al numero di interni potenzialmente abitativi;
- d) civico presente nella tabella Rnc ma non trovato nella Lac.

I primi due casi identificavano situazioni non rilevanti ai fini del recupero della sotto-copertura, identificando rispettivamente indirizzi in cui tutti gli interni abitativi erano occupati da famiglie residenti (caso a) o, nel caso di un numero di famiglie residenti maggiore degli interni ad uso abitativo (caso b), indirizzi ai

⁵¹ Egon è un software, installabile in locale, oppure disponibile direttamente *online* via web, progettato per garantire la puntuale “normalizzazione” di indirizzi errati o incompleti. La normalizzazione degli indirizzi consiste nella verifica, validazione e correzione delle inesattezze e delle incoerenze riscontrate in un indirizzo o in un insieme di indirizzi, al fine di ridurre ogni singolo recapito alla sua forma universalmente intellegibile, rendendolo così rintracciabile dai più diversi sistemi informativi mondiali.

quali erano presenti famiglie coabitanti (o, in qualche caso, evidenziando un errato conteggio degli interni potenzialmente abitativi durante la rilevazione). I casi *c* e *d* identificavano invece un possibile problema di sotto-copertura, in quanto gli alloggi (interni) potenzialmente abitativi non trovavano un corrispettivo nel numero di famiglie risultante in anagrafe, che era inferiore (caso *c*) o pari a zero (caso *d*)⁵². In questi due casi si è proceduto quindi a popolare Sgr con un numero di righe equivalente alla differenza tra il numero di interni abitativi rilevato alla Rnc e il numero di famiglie risultante in Lac allo stesso indirizzo. Più in dettaglio, nel caso non fosse stato trovato l'intero civico nella Lac, tutte le unità potenzialmente abitative conteggiate alla Rnc sono state identificate come segnali di potenziale sotto-copertura. In particolare, il Diario di sezione – il principale strumento di lavoro del rilevatore (interfaccia web presente come funzionalità di Sgr) – conteneva una lista di indirizzi ordinabile e filtrabile, corrispondente alle famiglie della Lac (a cui era stato spedito o doveva essere consegnato il questionario), agli individui presenti nella Lifa e non in Lac (cfr. paragrafo 2.2.3) e agli interni potenzialmente abitativi. Ogni riga del Diario mostrava inoltre lo stato di lavorazione della riga e, laddove esistenti (ovvero per i record da Lac e da Lifa), i dati anagrafici (dell'intestatario del Foglio di famiglia per le famiglie da Lac o dell'individuo per i record della Lifa).

Le righe aggiunte nel Diario sulla base della Rnc erano identificate dallo stato “sotto-copertura da Rnc”, che era uno stato iniziale. Per queste righe, come per le altre (ad eccezione dei questionari compilati *online* per i quali lo stato veniva aggiornato automaticamente dal sistema), gli operatori della rete di rilevazione (rilevatori o operatori di *back office*) dovevano aggiornare lo stato in base all'esito della rilevazione e alla fase del processo. In particolare, per poter procedere alle operazioni di chiusura della sezione di censimento, il rilevatore doveva “risolvere” ogni riga, cioè procedere alla compilazione del questionario, nel caso in cui avesse identificato una famiglia non presente nella lista iniziale, o chiudere la riga con uno stato di lavorazione finale coerente con il *workflow* di riferimento. Di seguito i possibili stati di chiusura previsti dal sistema, in caso non fosse stata identificata la famiglia: famiglia non trovata all'indirizzo, abitazione non occupata o la voce generica “Altro”, per tutti i casi particolari.

Da un punto di vista tecnico le procedure di *linkage* e di popolamento del Diario sono state sviluppate in linguaggio PL/SQL. Poiché queste operazioni sono state effettuate durante la fase di esercizio, è stato necessario schedare dei “*job a tempo*” per non rallentare il funzionamento di Sgr e quindi le operazioni sul campo. Inoltre, grazie ad un sistema di tracciatura e di controllo delle operazioni svolte, è stato possibile monitorare quanto veniva caricato nell'applicazione, consentendo il ripristino dei dati in caso di situazioni non previste o di difficile gestione. Complessivamente, sono stati immessi in Sgr più di 9 milioni di segnali di sotto-copertura, la maggior parte dei quali relativi a record provenienti dalle Lifa.

⁵² È opportuno ricordare che le Lac, essendo liste di famiglie iscritte nell'anagrafe del comune, contenevano solo indirizzi con almeno una famiglia residente.

CAPITOLO 3

DISEGNO DI CAMPIONAMENTO E SELEZIONE DEL CAMPIONE⁵³

3.1 Introduzione

La decisione di impiegare una strategia campionaria tramite *short/long form* al 15° Censimento generale della popolazione e delle abitazioni è stata avallata dai risultati di uno studio preliminare avviato con lo scopo di valutare quale disegno di campionamento fosse possibile praticare con maggiore favore nel contesto censuario, sia per le ricadute sulle operazioni in campo che per l'impatto sull'affidabilità dei risultati finali. Una volta definita la strategia, l'operazione successiva è stata la predisposizione del data base necessario alla fase di estrazione, da ciascuno dei Comuni coinvolti dal campionamento, dei campioni rappresentativi di famiglie.

Il capitolo descrive l'ambito di riferimento, i dati, le problematiche e alcune principali soluzioni praticate per la determinazione delle liste di famiglie residenti iscritte nelle anagrafi comunali e geocodificate nei domini dei Comuni sottoposti a campionamento, ai fini della formazione dei campioni per il Censimento della popolazione e delle abitazioni del 2011.

Dopo la presentazione del disegno di campionamento adottato (paragrafo 3.2), vengono rappresentati i Comuni interessati dalla strategia campionaria, i domini di campionamento e le unità dichiarate eleggibili per la rilevazione tramite modelli *short* e *long* (paragrafo 3.3). Nel paragrafo 3.4 viene illustrata la fase di realizzazione delle liste di campionamento inclusa l'operazione di normalizzazione degli indirizzi, utile alla geocodifica delle famiglie eleggibili nelle sezioni di censimento. Il paragrafo 3.5 riporta l'elenco delle operazioni eseguite fino all'estrazione dei campioni di famiglie per ciascuno dei Comuni campione. In seguito sono illustrati i risultati delle singole operazioni e i principali aspetti di qualità (paragrafo 3.6). Il capitolo si conclude (paragrafo 3.7) con alcune considerazioni di sintesi sull'intero processo seguito, con l'evidenza delle criticità incontrate.

3.2 La strategia di campionamento tramite *short/long form*

Il 15° Censimento generale della popolazione e delle abitazioni ha visto, per la prima volta in Italia, l'impiego di una strategia campionaria basata sull'utilizzo contemporaneo di un questionario in forma breve (*short form*) e un questionario in forma estesa (*long form*). Con il modello breve sono state osservate le caratteristiche familiari e demografiche, le principali variabili socio-economiche degli individui, e alcune caratteristiche delle abitazioni; con il modello completo sono stati rilevati i dati per tutte le variabili tradizionalmente osservate in occasione del censimento.

L'adozione delle tecniche di campionamento rientra nel complesso delle innovazioni introdotte per il censimento del 2011 al fine di diminuire il fastidio statistico sui rispondenti, in special modo di quelli residenti nei Comuni più grandi. Quindi, per massimizzare il ritorno spontaneo dei questionari compilati, riducendo così le operazioni di sollecito e di recupero da parte dei rilevatori, è stato definito un questionario breve con un ridotto numero di domande; mentre, per garantire la continuità del contenuto informativo, rispetto al passato, e rispettare i vincoli internazionali (Unece, 2006; Unsd, 2007) in merito alle variabili di censimento, è stato deciso di rilevare l'insieme completo di variabili su campioni di famiglie opportunamente selezionati.

La strategia campionaria ha interessato i Comuni capoluogo di Provincia e quelli non-capoluogo con almeno 20 mila abitanti. In tali Comuni, il questionario *long* è stato somministrato a campioni rappresentativi di famiglie mentre quello *short* alle rimanenti famiglie non incluse nel campione. In tutti i Comuni non coinvolti dal campionamento la rilevazione è avvenuta unicamente tramite il questionario in versione *long* per

⁵³ Il Capitolo 3 è a cura di Giancarlo Carbonetti. I paragrafi 3.1, 3.2, 3.6, 3.7 e i paragrafi 3.4.1, 3.4.4, 3.4.5 sono stati redatti da Giancarlo Carbonetti; i paragrafi 3.3.1, 3.3.3 sono stati redatti da Epifania Fiorello; il paragrafo 3.3.2 è stato redatto da Stefania Di Domenico; i paragrafi 3.4.2, 3.4.3 sono stati redatti da Sara Giavante; i paragrafi 3.5.1, 3.5.2, 3.5.3 sono stati redatti da Francesco Borrelli; i paragrafi 3.4.6, 3.5.4, 3.5.5 sono stati redatti da Luana De Felici. Le tavole contenute nell'Allegato A sono state curate da Giancarlo Carbonetti.

tutte le famiglie. Con tale approccio i dati relativi alle variabili osservate con il modello *short* derivano da un conteggio esaustivo (per qualunque livello territoriale) mentre, per i Comuni sottoposti a campionamento e per i livelli territoriali superiori (Provincia, Regione, ripartizione geografica), le informazioni rilevate esclusivamente con il modello *long* (le variabili di tipo socio-economico e il loro incrocio con le variabili demografiche) sono ottenute con procedimenti di stima campionaria (cfr. Capitolo 5).

Per la scelta della tecnica campionaria (disegno e stimatore) per il censimento del 2011 è stata effettuata una valutazione metodologica e un'articolata sperimentazione (su dati del censimento del 2001) di varie strategie campionarie, per valutare l'impatto del campionamento sia sulle operazioni in campo che sull'accuratezza dei risultati finali (Borrelli *et al.*, 2011; Borrelli *et al.*, 2012). A conclusione della suddetta fase, per l'estrazione dei campioni di famiglie si è stabilito di adottare un disegno casuale semplice da Lista anagrafica comunale (Lac), con un tasso di campionamento (di progettazione⁵⁴) pari al 33 per cento, quantità che permette di produrre stime con la desiderata accuratezza, fino ai livelli comunale e sub-comunale. Per la fase di calcolo delle stime si è deciso di impiegare lo stimatore di ponderazione vincolata per le proprietà statistiche che tale approccio garantisce (cfr. paragrafo 5.2).

La selezione dei campioni di famiglie è avvenuta su domini di campionamento sub-comunali, le aree di censimento di centro abitato (Ace), opportunamente disegnati (cfr. paragrafo 3.3.2) nei Comuni interessati dalla strategia campionaria tramite *short* e *long form*. Tali aree si configurano sia come *domini pianificati* dal disegno di campionamento per la formazione dei campioni, che come *aree di output*, offrendo in tal modo la possibilità di diffondere dati riferiti agli incroci di tipo *long* anche ad un livello di dettaglio territoriale più fine rispetto a quello comunale.

3.3 I Comuni e le unità eleggibili alla rilevazione campionaria

3.3.1 I Comuni rilevati a campione tramite *short/long form*

La strategia campionaria è stata disegnata per la popolazione italiana dei seguenti Comuni:

- Comuni capoluogo di Provincia;
- Comuni non capoluogo di Provincia con almeno 20 mila abitanti.

In base al dato della popolazione ufficiale Istat riferito al 31 dicembre 2008, sono stati dichiarati eleggibili al campionamento 509 Comuni. In seguito agli eventi sismici che hanno colpito la Regione dell'Abruzzo (aprile 2009), il Comune dell'Aquila è stato escluso dalla rilevazione campionaria. Sono stati esclusi dall'operazione campionaria ulteriori 22 Comuni per i quali non era stato possibile, per motivi diversi, disegnare le aree di censimento (Ace); a riguardo, una delle principali condizioni richieste per il disegno delle Ace era che il centro principale contenesse almeno 10 mila persone. La lista definitiva dei Comuni sottoposti a campionamento si è così ridotta a 486 (cfr. Allegato A).

Si fa presente che la Provincia di Ogliastra (Sardegna) è stata l'unica Provincia esclusa dalla strategia campionaria del censimento del 2011, in quanto nessuno dei suoi due Comuni capoluogo di Provincia (Tortolì e Lanusei) aveva i requisiti dimensionali minimi per il disegno delle aree di censimento (cfr. paragrafo 3.3.2).

Nella Tavola 3.1 è riportata la distribuzione dei Comuni a campione per classi demografiche. Si evidenzia che oltre il 90 per cento dei Comuni a campione ha dimensione tra 20 mila e 100 mila abitanti; nello specifico, il 69,14 per cento è tra 20 mila e 50 mila abitanti, il 21,19 per cento è tra 50 mila e 100 mila abitanti. Solo il 2,47 per cento dei Comuni campionati è superiore ai 250 mila abitanti. Due Comuni della Regione Sardegna, pur avendo popolazione inferiore a 20 mila abitanti, sono stati ritenuti eleggibili al campionamento in quanto Comuni capoluogo di Provincia: Tempio Pausania, della Provincia di Tempio-Olbia; Villacidro, della Provincia di Medio Campidano.

⁵⁴ Il tasso del 33 per cento è applicato alla lista delle famiglie eleggibili al campionamento che costituiscono un sottoinsieme delle famiglie iscritte in anagrafe.

Tavola 3.1 – Distribuzione dei 486 Comuni sottoposti a strategia di campionamento tramite *short/long form* al censimento del 2011, per classe demografica (popolazione residente al 31 dicembre 2008).

Classe di ampiezza demografica	Comuni	%
fino a 20.000	2 ^(*)	0,41
20.000 - 50.000	336	69,14
50.000 - 100.000	103	21,19
100.000 - 150.000	18	3,70
150.000 - 250.000	15	3,09
oltre 250.000	12	2,47
Totale	486	100,00

^(*) I due Comuni con meno di 20 mila abitanti sono: Tempio Pausania (prov. di Olbia-Tempio) - Villacidro (prov. di Medio Campidano).

Fonte: Direzione centrale dei Censimenti generali (DCCG), Istat.

3.3.2 I domini di campionamento: le aree di censimento

Uno degli strumenti più utilizzati per la realizzazione dei censimenti sono le Basi territoriali⁵⁵ (Bt), predisposte dall'Istat con il supporto dei Comuni italiani, per favorire le operazioni connesse alla rilevazione censuaria.

Gli oggetti delle Bt sono, in ordine gerarchico decrescente: i limiti amministrativi (regionali, provinciali e comunali); le località abitate (distinte in: “centri”; “nuclei abitati”; “case sparse”); le località produttive; i nuclei speciali; le aree speciali; le sezioni di censimento.

In occasione dei Censimenti generali del 2011 sono state introdotte le aree di censimento (Ace)⁵⁶, definite come aggregazioni di sezioni di censimento di tipo centro⁵⁷ contigue, di dimensione media intorno ai 15 mila abitanti (residenti). Le Ace determinano una suddivisione territoriale dei centri capoluogo⁵⁸ dei Comuni per i quali sono state costruite (i 486 Comuni sottoposti a campionamento) e rappresentano i “domini di campionamento” per i quali è stata pianificata l'estrazione dei campioni di famiglie iscritte nelle Lac per la rilevazione censuaria tramite *short/long form*. L'output di definizione delle Ace riporta l'elenco delle sezioni di censimento di tipo “centro” riferite alle Basi territoriali del 2011 appartenenti a ciascuna Ace.

A supporto delle successive operazioni di definizione del campione, è stata predisposta un data base contenente, per ciascun Comune eleggibile al campionamento, le seguenti informazioni: denominazione; codice identificativo Istat; popolazione ufficiale al 31 dicembre 2008; elenco delle aree di censimento disegnate e relativo ammontare di popolazione censita nel 2001. Al fine di effettuare ulteriori valutazioni, è stata calcolata anche la stima della popolazione delle Ace al 2010 facendo riferimento alla popolazione residente in famiglia al 31 dicembre 2010 (dati demografici Istat, giugno 2011). I dati, aggregati per Regione sono riportati nella Tavola 3.2.

⁵⁵ Le Basi territoriali sono predisposte (e periodicamente aggiornate) dall'Istat nella fase di preparazione dei censimenti generali per tener conto delle modifiche intervenute nel territorio comunale durante gli anni successivi all'ultimo censimento. Si tratta di un data base geografico realizzato con il supporto dei comuni a cui spetta il compito di occuparsi degli adempimenti topografici e di aggiornare il piano topografico; ciò è anche una naturale conseguenza della specifica conoscenza del territorio di propria competenza.

⁵⁶ Le “aree di censimento di centro abitato” (Bianchi *et al.*, 2007; Bianchi *et al.*, 2010) sono aree sub-comunali disegnate dall'Istat per gli obiettivi del 15° Censimento generale della popolazione e delle abitazioni; sono state definite per i comuni oltre i 20 mila abitanti interessati dalla strategia campionaria tramite *short/long form* impiegata nel censimento del 2011. Tali aree sono aggregazioni di sezioni di censimento di tipo “centro” con il vincolo della contiguità e secondo criteri di omogeneità statistica. Le aree di censimento sono state disegnate in modo tale da avere una dimensione media di circa 15 mila unità e da rispettare i limiti geografici delle suddivisioni (meno fini) predefinite dai comuni a scopi amministrativi o funzionali.

⁵⁷ Le sezioni di censimento sono le unità territoriali più piccole per la suddivisione del territorio comunale. Il loro perimetro è definito dall'Istat, in accordo con i comuni, rispettando limiti fisici (strade, ferrovie) e geografici (corsi di fiumi) facilmente individuabili sul territorio. Ad un livello più aggregato ci sono le “località abitate” distinte in “centri”, “nuclei” e “case sparse”. Le sezioni di censimento di tipo “centro” sono quelle di pertinenza delle località abitate “centro”.

⁵⁸ Il centro capoluogo di un comune corrisponde al centro abitato dove è localizzata la casa comunale e, non necessariamente, coincide con il centro più importante dal punto di vista demografico. I limiti geografici dei centri abitati principali impiegati per il disegno delle Ace si riferiscono alle Basi territoriali pre-censuarie del 2010.

Tavola 3.2 – Distribuzione per Regione del numero di Comuni eleggibili al campionamento, del numero di Ace, della popolazione legale 2001 nelle Ace e della popolazione “stimata” al 31 dicembre 2010 nelle Ace.

Regione	Comuni	Ace	Pop. 2001 nelle Ace	Pop. 2010 nelle Ace
Lombardia	66	249	3.768.026	4.005.244
Campania	63	224	3.364.884	3.407.993
Sicilia	56	194	3.000.832	3.050.756
Puglia	49	156	2.348.102	2.392.025
Lazio	37	210	3.443.698	3.810.458
Toscana	33	115	1.644.855	1.749.481
Piemonte	32	123	1.910.058	2.010.762
Veneto	32	94	1.428.675	1.506.637
Emilia-Romagna	31	119	1.786.510	1.952.678
Marche	16	37	536.807	567.772
Sardegna	16	46	624.624	646.541
Abruzzo	12	27	413.597	449.018
Calabria	10	33	524.430	531.773
Liguria	10	61	937.518	957.497
Umbria	7	21	321.615	355.979
Friuli-Venezia Giulia	5	24	400.580	405.951
Trentino-Alto Adige-Südtirol	5	18	242.305	269.174
Molise	3	6	86.599	90.058
Basilicata	2	8	110.315	112.540
Valle d'Aosta	1	2	32.534	33.551
Totale	486	1.767	26.926.564	28.305.888

Fonte: Direzione centrale dei Censimenti generali (DCCG), Istat.

Nei 486 Comuni a campione sono state disegnate 1.767 Ace, la maggior parte delle quali (1.033) si concentra nelle Regioni più popolose: Lombardia, Lazio, Campania, Sicilia, e Puglia. Nella Tavola 3.3 è riportato il numero di aree di censimento dei Comuni più grandi (superiori a 250 mila abitanti); si può notare che Roma è il Comune per il quale è stato disegnato il numero maggiore di Ace (146).

Tavola 3.3 – Distribuzione dei Comuni sopra i 250 mila abitanti, per numero di Ace (Pop. residente al 31-12-2008).

Comuni > 250.000 ab.	Ace
Roma	146
Milano	85
Napoli	68
Torino	56
Palermo	43
Genova	39
Bologna	23
Firenze	21
Bari	22
Catania	19
Venezia	14
Verona	14
Totale	550

Fonte: Direzione centrale dei Censimenti generali (DCCG), Istat.

Ai fini del campionamento, sono state ritenute eleggibili le aree di censimento il cui totale di popolazione residente in famiglia iscritta in anagrafe e geocodificata a una delle sezioni di censimento dell’Ace risulta non inferiore alla soglia di 5 mila unità; tale soglia, definita a seguito di valutazioni statistico-inferenziali, è stata posta come limite minimo per estrarre campioni utili a produrre stime significative su tali domini (Borrelli *et al.*, 2011). Nelle Ace aventi un numero di unità inferiore a tale soglia la rilevazione è stata condotta assegnando il modello *long* a tutte le famiglie residenti.

3.3.3 Le unità eleggibili al campionamento

Le informazioni contenute nelle Lac e il disegno definitivo delle Ace hanno permesso la definizione dei campioni di famiglie per ciascuno dei 486 Comuni coinvolti nella strategia campionaria, previa distinzione tra “unità eleggibili” e “unità non eleggibili” al campionamento.

Sono state considerate eleggibili tutte le famiglie residenti presenti nelle Lac e geocodificate alle Ace con dimensione superiore a 5 mila abitanti. Al 33 per cento di queste famiglie, estratte tramite uno schema di campionamento casuale semplice senza reimmissione, è stato assegnato il questionario *long* (*long* campione). Sono stati altresì assegnati i questionari *long* a tutte le famiglie geocodificate in aree di censimento con un numero di componenti superiore a 6; queste sono quindi state considerate famiglie auto-rappresentative, cioè comprese con certezza nel campione (*long* auto-rappresentativi). Alle famiglie non estratte per il campione e non considerate auto-rappresentative è stato consegnato il questionario *short* così come previsto dalla strategia censuaria.

Nei 486 Comuni considerati per il campionamento, sono definite non eleggibili le seguenti tipologie di famiglie:

- famiglie presenti in Lac e geocodificate ad aree di censimento con dimensione inferiore a 5 mila abitanti (Ace non eleggibili);
- famiglie geocodificate in sezioni non appartenenti ad aree di censimento (sezioni non di tipo “centro” o sezioni di nuova costituzione non considerate nel disegno delle Ace);
- famiglie non geocodificate (a causa di indirizzi errati o mancanti in Lac);
- individui senza fissa dimora;
- nomadi;
- cittadini italiani iscritti all’Aire (Anagrafe degli italiani residenti all’estero).

Le informazioni censuarie delle suddette unità non-eleggibili sono rilevate mediante il questionario in forma completa (*long*).

3.4 Le liste di campionamento

3.4.1 Premessa

Una volta definiti lo schema di estrazione dei campioni, l’insieme delle unità eleggibili al campionamento e i domini da cui estrarre i campioni di famiglie a cui inviare il questionario *long* per gli scopi inferenziali della strategia censuaria, l’operazione successiva è stata la formazione delle Liste di campionamento. La disponibilità delle informazioni sulle famiglie residenti iscritte nelle Lac, richieste dall’Istat a tutti i Comuni italiani per l’operazione censuaria, ha consentito la predisposizione delle liste di famiglie eleggibili al campionamento. Per tale scopo, è stata cruciale l’operazione di geocodifica delle famiglie presenti in Lac, tramite l’indirizzo di residenza, alle sezioni di censimento e, quindi, alle Ace di riferimento. L’obiettivo era quello di definire, per ciascun Comune campione, la Lista di campionamento delle famiglie eleggibili geocodificate alle proprie aree di censimento così da poter formare i campioni secondo il disegno campionario fissato (estrazione casuale semplice senza re-immissione con tasso di campionamento pari al 33 per cento).

In questo paragrafo si illustrano le operazioni svolte per la definizione delle Liste di campionamento, le criticità riscontrate nella fase di geocodifica e le possibili conseguenze da mancata o da errata geocodifica.

3.4.2 La geocodifica delle famiglie presenti nelle Liste anagrafiche comunali

Il complesso di informazioni statistiche e territoriali impiegato per la realizzazione della strategia campionaria nel 15° Censimento della popolazione e delle abitazioni è rappresentato da:

- le Liste anagrafiche comunali (Lac), elenchi delle famiglie residenti fornite all’Istat da tutti i Comuni italiani, secondo predefinite specifiche tecniche (“campi” e “formati”) (cfr. paragrafo 2.1);
- le Basi territoriali (Bt), strumenti territoriali (mappe digitali) definite dall’Istat per la predisposizione dei piani topografici utili alla realizzazione dei censimenti generali;
- le aree di censimento di centro abitato (Ace), suddivisioni territoriali dei Comuni più grandi (indicati nel paragrafo 3.3.1) che permettono di identificare domini in cui affiancare una strategia campionaria alla rilevazione censuaria tradizionale.

Il processo di trasmissione delle Lac all'Istat ha coinvolto tutti i Comuni italiani e ha previsto una modalità di scambio circolare di informazioni in cui l'Istat ha avuto i seguenti compiti: definire gli standard di riferimento a cui dovevano attenersi i Comuni; verificare i livelli di accuratezza⁵⁹ delle Lac; richiedere, in caso di qualità non accettabile, l'invio di una nuova Lac.

Lo scarico delle Liste anagrafiche comunali (Lac) è riferito al 31 dicembre 2010; la struttura dei dati è quella della matrice *individui x variabili* e include tutti gli individui residenti nel territorio del Comune ad esclusione degli iscritti nell'Anagrafe degli italiani residenti all'estero (Aire).

Le Basi territoriali sono state utilizzate per la tornata censuaria del 2011 per geocodificare strade e numeri civici alle sezioni di censimento, e per costruire le aree di censimento. Il passaggio dalle Basi territoriali del 2001 alle Basi territoriali del 2011 ha presupposto la revisione e la convalida della proposta Istat da parte del Comune.

Il disegno delle aree di censimento è stata un'attività innovativa rispetto al passato (cfr. paragrafo 3.3.2). L'Istat ha inviato al Comune la proposta di disegno delle aree di censimento; l'ufficio tecnico del Comune ha effettuato le dovute verifiche e apportato le eventuali modifiche. Una volta concordate le proposte di modifica ed eseguito i controlli di coerenza⁶⁰, si è pervenuti alla versione definitiva delle aree di censimento.

Le informazioni derivanti dall'uso delle Lac e dall'individuazione delle Ace ha permesso di distinguere tra unità eleggibili al campionamento e unità da escludere dalla strategia campionaria.

Le operazioni di normalizzazione⁶¹ e di geocodifica (attribuzione del codice di sezione di censimento riferito alle Basi territoriali del 2001) degli indirizzi presenti in Lac si sono rivelate azioni cruciali per la formazione delle liste di campionamento e la formazione dei campioni di famiglie.

Tutti i file delle Lac ricevuti dai Comuni sono stati sottoposti al processo di normalizzazione degli indirizzi⁶². Si sono stabilite tre soglie percentuali di normalizzazione (rispetto al totale di indirizzi presenti in Lac) e conseguentemente sono stati individuati tre gruppi di Lac con differente qualità di normalizzazione⁶³:

- percentuale di indirizzi non normalizzati inferiore al 5 per cento;
- percentuale di indirizzi non normalizzati tra il 5 e il 10 per cento;
- percentuale di indirizzi non normalizzati superiore al 10 per cento.

Le Lac appartenenti ai primi due gruppi sono state giudicate idonee per essere sottoposte alle procedure di campionamento, previo controllo del campo "Nsez" che riporta il "codice di sezione di censimento". Quelle del terzo gruppo hanno presentato maggiori criticità e hanno richiesto interventi *ad hoc*; per queste, gli indirizzi ritenuti errati sono stati sottoposti ad un intervento manuale e ad un ulteriore procedimento di normalizzazione per portare la percentuale di indirizzi scartati sotto il 10 per cento, a prescindere dal contenuto del campo "Nsez".

Il campo "Nsez" all'interno della Lac riportava il codice di sezione di censimento solo se questo era presente nell'archivio anagrafico del Comune; tale codice poteva essere relativo alle Basi territoriali del 2001 o alle Basi territoriali del 2011. Non tutti i Comuni avevano valorizzato tale campo e non sempre veniva compilato in modo corretto. Il codice riportato, ove presente, è stato comunque utile per il geocodifica della famiglia nella sezione e quindi nella relativa Ace.

Si fa presente che il codice di sezione era assegnato alla famiglia e non all'indirizzo; questo significa che nel caso di cambio di dimora di una famiglia nello stesso Comune, il codice di sezione di censimento non veniva aggiornato in modo automatico, risultando così erroneo rispetto al nuovo indirizzo.

⁵⁹ Le verifiche di accuratezza delle Lac hanno principalmente riguardato: identificabilità dell'intestatario di famiglia; normalizzazione dell'indirizzo dell'intestatario; copertura.

⁶⁰ Le verifiche di coerenza del disegno delle Ace hanno riguardato: copertura (inclusione solo delle sezioni di tipo "centro"); perimetro (rispetto dei limiti geografici sub-comunali); popolazione (ampiezza demografica coerente con le specifiche progettuali - circa 15 mila abitanti).

⁶¹ Il processo di normalizzazione degli indirizzi consiste in una serie di controlli, del data base di riferimento, utili a correggere gli errori e completare le informazioni mancanti (per esempio, recupero dei "codici di avviamento postale"). Tale processo, una volta completato, fa sì che gli indirizzi utilizzati siano corretti per le successive operazioni. Per le attività del censimento del 2011 la normalizzazione degli indirizzi, avvenuta mediante l'impiego del software Egon (cfr. nota 51), ha consentito di verificare la correttezza degli indirizzi presenti nelle Lac con un elevato margine di completezza.

⁶² La procedura automatizzata di normalizzazione degli indirizzi usata ha richiesto l'assegnazione, ad ogni indirizzo processato, di un valore numerico da 0 a 3, attinente l'accuratezza dell'esito del processo. Sono stati giudicati normalizzati gli indirizzi ai quali è stato assegnato il codice 0 o 1. Non sono stati ritenuti normalizzati gli indirizzi con codice 2 o 3; per questi si è dovuto procedere con procedimenti di aggiustamento manuali.

⁶³ I livelli di qualità sono influenzati dall'esito della normalizzazione: un indirizzo è scartato proprio quando il processo di normalizzazione relativo al corrispondente record non è andato a buon fine.

A seconda della presenza del codice di sezione nel campo “Nsez” si è proceduto secondo una delle seguenti modalità:

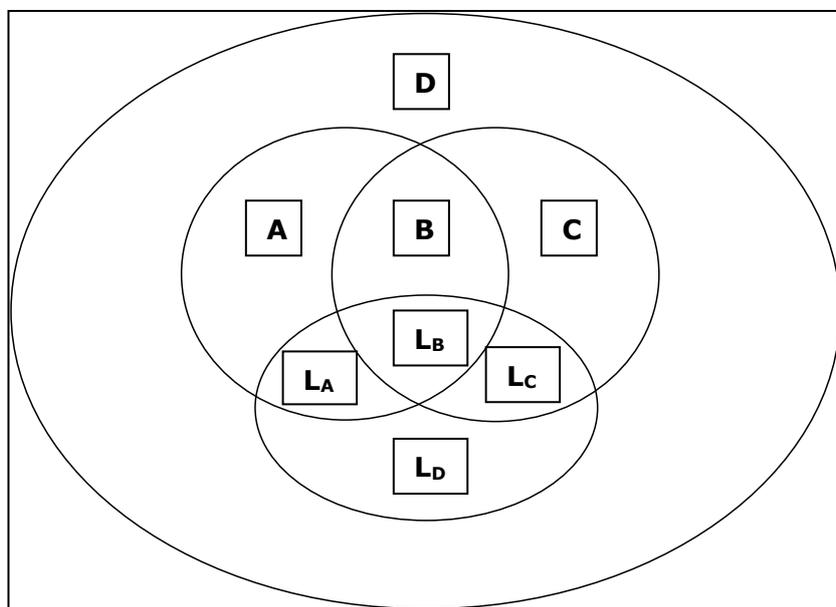
1. *campo “Nsez” vuoto* – il record corrispondente è stato geocodificato alla sezione di censimento mediante le informazioni derivanti dal processo di normalizzazione degli indirizzi; nello specifico, l’indirizzo normalizzato risulta geocodificato alle Basi territoriali del 2001.
2. *campo “Nsez” pieno* – si è proceduto a verificare l’affidabilità dell’informazione fornita dal Comune, intesa come coerenza del codice di sezione di censimento rispetto all’elenco delle sezioni di censimento relative alle Basi territoriali del 2001 (BT01) o del 2011 (BT11). A riguardo si sono riscontrate due situazioni:
 - 2.1. contenuto di “Nsez” inaffidabile (codice incoerente con le BT Istat) – ciò ha determinato l’impiego del procedimento precedente, ovvero la geocodifica alla sezione di censimento mediante le informazioni derivanti dal processo di normalizzazione degli indirizzi, in modo tale da ottenere un indirizzo normalizzato geocodificato al 2001;
 - 2.2. contenuto di “Nsez” affidabile (coerenza con una delle due BT di riferimento) – non è stato necessario ricorrere ad ulteriori azioni per la geocodifica.

3.4.3 Le Lac “ibride”

Per Lac “ibrida” si intende la Lac di un Comune per la quale, a seguito delle valutazioni sulla conformità dei codici presenti nel campo “Nsez” e dell’eventuale operazione di normalizzazione, non si riesce ad attribuire, in modo univoco, i codici di sezione di censimento ad una tra le due Basi territoriali di riferimento (BT01 - BT11). Tra i 486 Comuni a campione, questo evento è stato osservato nelle Lac di 134 Comuni.

Nella Figura 3.1 è rappresentata la casistica possibile riferita alle Lac ibride.

Figura 3.1 – Possibili gruppi derivanti dall’incrocio tra le liste dei codici di sezione di censimento delle Basi territoriali Istat (BT01 e BT11) e l’insieme dei codici di sezioni contenuti nelle Liste anagrafiche comunali (Lac).



Fonte: Direzione centrale dei Censimenti generali (DCCG), Istat.

Con A, B e C si rappresentano le liste dei codici di sezione di censimento delle Basi territoriali Istat (BT01, BT11). Gli insiemi dei codici di sezione presenti nelle Lac e riconducibili alle suddette liste sono indicati con: LA, LB, LC; con LD è indicato l’insieme di codici presenti in Lac ma inesistenti nelle Basi territoriali. In dettaglio:

- A – Lista delle sezioni di censimento delle Basi territoriali del 2001 (BT01);
- B – Lista delle sezioni di censimento dell’intersezione tra BT01 e BT11;
- C – Lista delle sezioni di censimento delle Basi territoriali del 2011 (BT11);
- L_A – Insieme delle sezioni di censimento della Lac presenti solo in BT01;
- L_B – Insieme delle sezioni di censimento della Lac presenti in BT01 e in BT11;
- L_C – Insieme delle sezioni di censimento della Lac presenti solo in BT011;
- L_D – Insieme delle sezioni di censimento della Lac non presenti né in BT01, né in BT11.

L’insieme dei codici di sezione di censimento presenti nella Lac possono essere ricondotti ad una sola delle tipologie di insiemi L_A, L_B, L_C, L_D, oppure a due o più tra tali insiemi. Per esempio, la Lac di un generico Comune può contenere solo i codici della lista BT01 (quindi vale la tipologia di gruppo L_A), oppure può contenere sia codici della BT01 che codici inesistenti nelle Bt prese in considerazione (quindi vale anche la tipologia L_D).

Per risolvere le criticità delle Lac ibride sono stati considerati i seguenti dati comunali:

- anno dichiarato dal Comune al quale sono riferiti i codici delle sezioni di censimento;
- numero di sezioni di censimento delle Basi territoriali del 2001;
- numero di sezioni di censimento delle Basi territoriali del 2011;
- numero di sezioni di censimento valide presenti in Lac;
- popolazione delle sezioni di censimento valide in Lac;
- popolazione delle sezioni di censimento dell’insieme L_D;
- percentuale di popolazione dell’insieme L_D sulla popolazione delle Sezioni valide in Lac;
- numero di intestatari di famiglia delle sezioni di censimento dell’insieme L_D.

L’impiego di tali informazioni ha permesso, tramite specifiche analisi (con il supporto cartografico) e assegnazione “manuale” dei codici di sezione, di attribuire una delle due Basi territoriali di riferimento alle Lac ibride. Nel caso di quattro Comuni si sono evidenziate situazioni con una criticità tale da richiedere azioni *ad hoc*, con una procedura più articolata e l’impiego delle informazioni raccolte con la Rilevazione dei numeri civici⁶⁴ (Rnc).

3.4.4 Mancata geocodifica ed errata geocodifica

L’estrazione di famiglie, nei Comuni a campione, e la successiva assegnazione dei modelli *long* (alle famiglie estratte e a quelle auto-rappresentative) e dei modelli *short* (alle famiglie eleggibili non estratte) presuppongono un’elevata correttezza dell’operazione di geocodifica delle famiglie alle sezioni di censimento e, di conseguenza, alle aree di censimento intese come domini di campionamento.

Infatti, la disponibilità di Lac con alti livelli di qualità in termini di completezza e correttezza dei dati presenti (in particolare, l’indirizzo di residenza: via e numero civico) favorirebbe l’esatta geocodifica di tutte le famiglie residenti nel territorio italiano alle corrispondenti sezioni di censimento; in tal modo, quelle che risultano geocodificate nelle aree di censimento sarebbero tutte eleggibili al campionamento mentre quelle fuori dalle suddette aree sarebbero automaticamente escluse dal campionamento.

Quanto più ci si allontana da questa situazione ottimale, tanto più si potrebbero avere ricadute negative sull’operazione campionaria e sulla successiva fase di stima delle variabili oggetto di interesse per il censimento. La trasmissione di Lac “imperfette” per la presenza di errori o l’assenza di dati nel campo relativo all’indirizzo, oltre a richiedere onerosi interventi *ad hoc* per correggere o ripristinare, ove possibile, il dato mancante, evidenzia delle criticità che potrebbero avere conseguenze sulla qualità finale delle informazioni censuarie raccolte.

La mancata geocodifica alle sezioni di censimento di famiglie residenti eleggibili alla strategia campionaria comporta l’esclusione di queste unità dalle procedure di campionamento, con una possibile riduzione dell’accuratezza delle stime finali che sarà tanto maggiore quanto più sarà ridotta la lista di famiglie eleggibili dell’area di censimento. Nel caso limite di una riduzione della lista di campionamento sotto la soglia minima di 5 mila unità (residenti nell’Ace), si modifica lo stato di eleggibilità dell’area di censimento; questo, però, non ha riflessi negativi sulla bontà del dato finale in quanto comporta l’esclusione di tale area dal campionamento, l’assegnazione del modello *long* a tutte le famiglie in essa geocodificate e quindi la raccolta

⁶⁴ Cfr. nota 20.

esaustiva delle informazioni censuarie.

Una errata geocodifica delle famiglie residenti alle sezioni di censimento può avere ricadute ex-post alla rilevazione. Infatti, una volta evidenziati gli errori di geocodifica, si potrebbero osservare variazioni alla dimensione dell'area di censimento, e quindi alla frazione sondata reale. Inoltre, si potrebbe aver assegnato il modello *short* a famiglie che invece avrebbero dovuto ricevere il modello *long*, generando in tal modo “mancate risposte parziali”, con conseguente perdita di informazioni.

La verifica dell'impatto di una errata operazione di geocodifica viene effettuata a posteriori, ad operazioni censuarie concluse, ma è possibile individuare a priori almeno tre casi ciascuno dei quali comporta una diversa distorsione dei conteggi di censimento (Prospetto 3.1).

Prospetto 3.1 – Conseguenze di geocodifica errata.

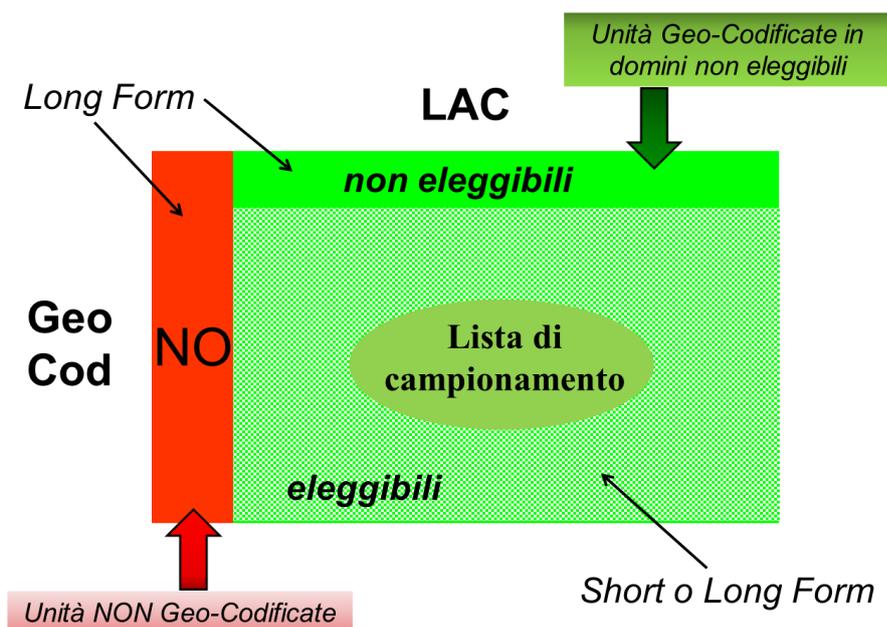
Errata geocodifica	Conseguenza
Geocodifica di una famiglia eleggibile al campionamento in un'area di censimento diversa da quella a cui realmente appartiene.	L'eleggibilità della famiglia viene mantenuta: cambiano le dimensioni delle popolazioni delle aree di censimento coinvolte (errata e reale) e le rispettive dimensioni campionarie.
Geocodifica di una famiglia eleggibile al campionamento in un dominio territoriale non sottoposto alle procedure di campionamento.	La famiglia diventa non eleggibile al campionamento: si modifica l'ammontare della popolazione dell'area di censimento coinvolta (assegnata per errore) e la corrispondente dimensione campionaria.
Geocodifica di una famiglia non eleggibile al campionamento in un'area di censimento.	La famiglia non eleggibile diventa eleggibile al campionamento; si modifica la dimensione della popolazione dell'area di censimento coinvolta. Se la famiglia viene campionata si modifica la dimensione campionaria dell'area di censimento interessata; se la famiglia non viene estratta per il campione, l'errata assegnazione del modello <i>short</i> al posto del <i>long</i> implica delle “mancate risposte parziali”.

Fonte: Direzione centrale dei Censimenti generali (DCCG), Istat.

3.4.5 Integrazione tra i dati delle Liste anagrafiche comunali e le aree di censimento

L'operazione di integrazione tramite il codice di sezione di censimento impiegato come *linkage key* (chiave di aggancio) tra l'elenco delle famiglie presenti in Lac e il disegno delle Ace (l'elenco delle sezioni di censimento della BT11 appartenenti a ciascuna area di censimento) ha portato a costruire le liste di campionamento dei 486 Comuni coinvolti dalla strategia campionaria.

L'operazione di geocodifica delle famiglie presenti in Lac ha prodotto due insiemi (Figura 3.2): l'insieme delle famiglie geocodificate (riquadri in “verde”) e l'insieme delle famiglie non geocodificate (riquadro in “rosso”). Le famiglie che sono state geocodificate nei domini eleggibili al campionamento costituiscono la lista di riferimento per la formazione dei campioni; da questo insieme è stato estratto il campione con frazione sondata pari al 33 per cento. Alle unità estratte è stato assegnato il modello *long* e alle famiglie non estratte quello *short*. Invece, tutte le famiglie geocodificate nei domini non eleggibili hanno ricevuto il modello *long* e, ai fini inferenziali, sono state considerate auto-rappresentative (con peso finale di riporto all'universo pari a “1”). Allo stesso modo sono state considerate tutte le famiglie non geocodificate che hanno ricevuto il modello *long*.

Figura 3.2 – Schema dell'integrazione tra i dati contenuti nelle Lac e il disegno delle Ace.

Fonte: Direzione centrale dei Censimenti generali (DCCG), Istat.

3.4.6 Lo sdoppiamento delle sezioni di censimento

Come illustrato nel paragrafo 3.4.5, per l'integrazione tra i dati contenuti nelle Lac e i riferimenti territoriali delle Ace è stata utilizzata come chiave di aggancio il codice di sezione di censimento.

In particolare, per 206 Comuni dei 486 sottoposti a campionamento è stato usato il codice di sezione di censimento relativo alle Basi territoriali disegnate per il censimento del 2011 (BT11); per i restanti 280 Comuni è stato usato il codice di sezione di censimento relativo alle Basi territoriali predisposte per il censimento del 2001 (BT01). Fra questi ultimi, 35 Comuni presentavano alcune sezioni di censimento 2001 che, per il disegno delle nuove Basi territoriali, sono state divise in due o più sezioni 2011 ricadenti, a volte, in aree di censimento differenti, e impedendo così l'assegnazione delle relative famiglie ad una sezione precisa e quindi ad una Ace. In questi casi, non essendoci disponibilità di altre informazioni per geocodificare la famiglia all'Ace corretta, è stato deciso di assegnare tutte le famiglie di tali sezioni, ritenute "critiche", ad una sola sezione della BT11 (scelta a caso tra quelle di "nuova creazione") e di conseguenza all'area di censimento ad essa associata.

Successive analisi hanno evidenziato che i casi in questione, numericamente esigui, non hanno modificato lo stato di eleggibilità delle Ace e hanno cambiato la frazione di campionamento reale in misura trascurabile.

3.5 Le procedure di estrazione dei campioni e la validazione dei risultati

3.5.1 Caricamento dei dati di input e di output

Per quanto riguarda gli aspetti operativi sulla formazione dei campioni di famiglie, per ciascuno dei 486 Comuni coinvolti nella strategia campionaria, si è proceduto con le seguenti fasi:

- caricamento dei dati di input (Lac e Ace);
- predisposizione della lista di campionamento (geocodifica delle unità in Lac alle Ace);
- estrazione dei campioni di famiglie;
- assegnazione della tipologia di questionario (*short o long*) alle famiglie della Lac;
- caricamento dei dati di output.

La prima operazione ha previsto il caricamento, sui sistemi informatici, dei dati utili alla fase di integrazione (cfr. paragrafo 3.4.5) per la predisposizione della lista di campionamento: l'elenco delle famiglie

presenti in Lac e geocodificate alle Ace, complete dei codici identificativi (dell'intestatario), dell'indirizzo e del codice di sezione di censimento riportato alla BT11 (tramite la tabella di transcodifica da BT01 a BT11); l'elenco dei codici delle sezioni di censimento delle Basi territoriali del 2011 appartenenti alle Ace.

Si è quindi proceduto prima con l'estrazione da ogni Ace, del campione di famiglie a cui fornire il questionario completo (*long*) e, successivamente, all'assegnazione a ciascuna famiglia presente nella Lac, della versione del questionario (*short* o *long*). Nell'ultima fase si è proceduto al caricamento dei dati di output relativi al risultato dell'attività di estrazione dei campioni.

3.5.2 Preparazione della lista di campionamento

La creazione della lista di campionamento ha previsto due operazioni preliminari:

- 1) esclusione delle unità non eleggibili al campionamento: le “famiglie in convivenza” e le “persone senza fissa dimora”⁶⁵;
- 2) associazione, tramite il codice di sezione di censimento, di ogni famiglia residente geocodificata alla propria area di censimento.

Le famiglie non geocodificate o geocodificate in domini (sezioni di censimento o Ace) non campionabili sono state escluse dalla strategia campionaria ed è stato assegnato loro il questionario di tipo *long*.

3.5.3 Estrazione dei campioni di famiglie

Dalla lista di campionamento di ciascun Comune campione sono stati estratti campioni di famiglie residenti secondo uno schema casuale semplice senza ripetizione con un tasso di campionamento pari al 33 per cento. Così facendo, è stato estratto un campione rappresentativo di famiglie residenti per ogni area di censimento del Comune, previa verifica dell'eleggibilità dell'area stessa.

L'estrazione delle unità campione è avvenuta mediante l'utilizzo di specifiche procedure scritte in linguaggio Sas che hanno previsto dapprima l'assegnazione di un valore casuale alle unità di campionamento e successivamente la selezione della parte riferita al primo terzo della lista stessa, dopo averla ordinata in senso decrescente rispetto al valore casuale assegnato.

3.5.4 Controlli effettuati

Dopo la conclusione delle operazioni di campionamento, i risultati sono stati sottoposti ad alcuni controlli di coerenza per la validazione finale. Di seguito alcuni dei controlli effettuati tramite specifiche procedure scritte in linguaggio Sas:

- i. numero di *long* assegnati + numero di *short* assegnati = numero famiglie presenti in Lac;
- ii. se l'area di censimento è non eleggibile, tutte le famiglie geocodificate in tale area devono avere assegnato il modello di tipo *long*;
- iii. se ad una famiglia è stato assegnato il modello di tipo *short*, l'Ace deve essere eleggibile e la famiglia non deve appartenere al campione.

3.5.5 Output prodotto

Dopo aver effettuato, per ognuno dei 486 Comuni campionati, l'estrazione dei campioni di famiglie e i relativi controlli di coerenza per la validazione, si è proceduto a caricare sul *data warehouse* del censimento la tabella di riassunto avente un numero di record pari al numero di famiglie presenti in Lac e contenente le seguenti informazioni:

- codice Istat identificativo della Provincia;
- codice Istat identificativo del Comune;
- codice di Ace;
- codice identificativo della famiglia;
- codice identificativo dell'intestatario di famiglia;
- stato di eleggibilità della Ace (che assume i seguenti valori: “1” se la famiglia ricade in un'Ace campionabile; “-1” se la famiglia ricade in un'Ace non campionabile per dimensione inferiore alla soglia di eleggibilità; “0” se la famiglia ricade in una sezione di censimento esclusa dal disegno delle Ace);
- stato di appartenenza della famiglia al campione (“1” alle famiglie estratte per il campione e alle famiglie

⁶⁵ Cfr. paragrafo 4.5.2.

- auto rappresentative; “0” per tutte le altre);
- tipologia di assegnato alla generica famiglia (“L” per *long*; “S” per *short*).

3.6 Risultati

In questo paragrafo sono riportate tre tavole che descrivono alcuni risultati delle operazioni di campionamento e di assegnazione dei questionari *short* e *long* alle famiglie presenti nelle Lac dei 486 Comuni a campione (i dati su famiglie e individui sono conteggiati sulle Lac riferite al 31 dicembre 2010).

Nella Tavola 3.4 è possibile osservare che il modello *long* è stato assegnato al 41,84 per cento delle famiglie residenti nei Comuni a campione, corrispondente al 43,00 per cento del totale degli individui.

Rispetto all’insieme delle famiglie residenti, le famiglie eleggibili al campionamento hanno rappresentato l’87,60 per cento del totale (Tavola 3.5). In particolare, al 33,20 per cento di queste, estratte per i campioni, è stato spedito il modello *long*, mentre al 66,39 per cento delle famiglie residenti non campionate è stato inviato il modello *short*; l’assegnazione dei modelli si è completata con l’invio del *long* alle famiglie con più di 6 componenti considerate nel campione auto-rappresentative (lo 0,41 per cento del totale delle famiglie residenti in Lac).

Tavola 3.4 – Distribuzione del numero di modelli *short* e *long* assegnati alle famiglie residenti (presenti nelle Lac riferite al 31 dicembre 2010) nei 486 Comuni a campione.

Questionari assegnati alle unità presenti nelle Lac dei 486 Comuni a campione	Long assegnati alle unità eleggibili campionate e alle unità non eleggibili	Short assegnati alle unità eleggibili non campionate	Totale
Famiglie residenti	5.673.461 41,84%	7.886.207 58,16%	13.559.668 100,00%
Individui	13.578.949 43,00%	18.002.539 57,00%	31.581.488 100,00%

Fonte: Direzione centrale dei Censimenti generali (DCCG), Istat.

Tavola 3.5 – Distribuzione del numero di modelli *short* e *long* assegnati alle famiglie residenti (presenti nelle Lac riferite al 31 dicembre 2010) eleggibili al campionamento dei 486 Comuni a campione, per tipologia di assegnazione.

Unità eleggibili al campionamento	Long assegnati alle famiglie eleggibili campionate	Short assegnati alle famiglie eleggibili non campionate	Long assegnati alle famiglie eleggibili con più di 6 componenti (auto-rappresentative)	Totale delle unità eleggibili	Totale delle unità in Lac
Famiglie residenti	3.944.074 33,20%	7.886.207 66,39%	47.967 0,41%	11.878.248 100,00%	13.559.668 100,00%
Individui	9.001.964 32,89%	18.002.539 65,78%	364.992 1,33%	27.369.495 100,00%	31.581.488 100,00%
	29,09%	58,16%	0,35%	87,60%	100,00%
	28,50%	57,00%	1,16%	86,66%	100,00%

Fonte: Direzione centrale dei Censimenti generali (DCCG), Istat.

Alle famiglie residenti nei 486 Comuni campione ritenute non eleggibili al campionamento, il 12,40 per cento del totale, è stato assegnato il modello *long*. Tra queste (Tavola 3.6), lo 0,53 per cento sono state dichiarate non eleggibili in quanto geocodificate in Ace non campionabili⁶⁶; le famiglie escluse a posteriori

⁶⁶ La soglia di eleggibilità al campionamento delle Ace è stata posta, per motivi di accuratezza delle stime finali, pari a 5 mila individui (cfr. paragrafo 3.3.2). A seguito dell’operazione di geocodifica delle famiglie residenti nelle Ace, in 6 delle 1.767 Ace disegnate per i 486 comuni a campione il numero di individui delle famiglie residenti della lista di campionamento è risultato inferiore a tale soglia: 4 nel comune di Messina; 1 nel comune di Trento; 1 nel comune di Bari. Di conseguenza, tali Ace sono state escluse dal campionamento, portando così a 1.761 il numero di Ace in cui si è seguita la strategia campionaria tramite *short/long form*.

(ad esempio: gli “irreperibili”⁶⁷) ammontano all’1,14 per cento; la quota di famiglie non geocodificate per vari motivi, pari al 29,28 per cento di quelle non eleggibili (il 3,63 per cento di tutte le famiglie presenti nelle Lac), è indicativa sia della qualità delle informazioni contenute nelle Lac che di difetti strutturali propri del processo di geocodifica (il software di normalizzazione; le basi dati degli indirizzi impiegate per la geocodifica). Inoltre, il 69,05 per cento delle famiglie residenti non eleggibili è stato escluso dal campionamento in quanto geocodificate in sezioni di censimento non incluse nel disegno delle aree di censimento.

In conclusione, dai dati presentati, riferiti ai 486 Comuni a campione, si può osservare che, del totale di questionari *long* assegnati (5.673.461), il 70,36 per cento è stato associato alle famiglie residenti campionate (3.944.074 + 47.967), mentre il restante 29,64 per cento alle famiglie escluse (1.681.420) dalla strategia di campionamento.

Tavola 3.6 – Distribuzione del numero di modelli *long* assegnati alle unità non eleggibili al campionamento nei 486 Comuni a campione (presenti nelle Lac riferite al 31 dicembre 2010).

Unità non eleggibili al campionamento	Long assegnati alle unità geocodificate in Ace non eleggibili	Long assegnati alle unità geocodificate in sezioni non eleggibili	Long assegnati alle unità non geocodificate	Long assegnati alle unità escluse "a posteriori"	Unità non eleggibili totali	Unità totali in Lac
	8.873	1.160.994	492.333	19.220	1.681.420	13.559.668
Famiglie residenti	0,53%	69,05%	29,28%	1,14%	100,00%	
	0,07%	8,56%	3,63%	0,14%	12,40%	100,00%
	22.221	2.890.529	1.274.583	24.660	4.211.993	31.581.488
Individui	0,53%	68,63%	30,26%	0,59%	100,00%	
	0,07%	9,15%	4,04%	0,08%	13,34%	100,00%

Fonte: Direzione centrale dei Censimenti generali (DCCG), Istat.

3.7 Considerazioni conclusive

L’adozione delle tecniche di campionamento è stata una delle principali innovazioni introdotte nel Censimento della popolazione e delle abitazioni del 2011 con lo scopo di contenere i costi, ridurre il fastidio statistico sui rispondenti, offrire una migliore accuratezza dei risultati e garantire una maggiore tempestività per il rilascio dei dati definitivi. La sua realizzazione ha richiesto la formazione di campioni rappresentativi di famiglie residenti, riferiti ad opportuni domini di campionamento, su cui rilevare l’intero insieme di variabili tradizionalmente osservate in occasione del censimento.

Tale operazione ha interessato i Comuni capoluogo di Provincia e i Comuni non capoluogo con almeno 20 mila abitanti e si è articolata nelle seguenti principali attività: disegno dei domini di campionamento da cui estrarre i campioni rappresentativi di famiglie residenti; geocodifica delle unità alle sezioni di censimento e ai domini di campionamento; assegnazione dello stato di eleggibilità delle unità di rilevazione; formazione delle liste di campionamento; estrazione dei campioni per l’attribuzione dei modelli *short* e *long*.

Per l’identificazione esatta dei domini di campionamento, le aree di censimento di centro abitato (Ace), si è fatto ricorso alla lista delle sezioni di censimento delle Basi territoriali predisposte per il 15° Censimento della popolazione e delle abitazioni.

Per la fase di geocodifica delle famiglie presenti nelle liste anagrafiche comunali (Lac) alle sezioni di censimento è stato considerato, come chiave di aggancio, l’indirizzo di residenza della famiglia. Questa operazione, per alcuni Comuni, ha presentato delle complessità a causa della qualità non elevata delle informazioni territoriali presenti in Lac. L’impiego di opportuni algoritmi di elaborazione e alcuni successivi interventi manuali hanno, comunque, permesso di associare alle famiglie residenti, per percentuali molto alte di casi presenti in Lac, un indirizzo valido e, di conseguenza, la sezione di censimento corrispondente e l’Ace di appartenenza.

Per stabilire l’eleggibilità delle famiglie al campionamento sono stati seguiti prefissati criteri con

⁶⁷ Gli “irreperibili” al censimento del 2011 sono coloro che erano iscritti in anagrafe (riferita al 31 dicembre 2010) e non sono stati trovati nella rilevazione censuaria, anche dopo diversi tentativi di contatto. In fase di revisione anagrafica, se questi risultavano irreperibili anche agli accertamenti anagrafici, allora si era nei termini di legge per operare la cancellazione dall’anagrafe per irreperibilità al censimento.

l'obiettivo di dare la corretta rappresentatività al campione finale. In questa fase si è proceduto anche a caratterizzare l'insieme delle famiglie auto-rappresentative (aventi caratteristiche tali da dover essere rilevate con il questionario completo), e l'insieme di unità da escludere dalla pratica campionaria per vari motivi.

Il completamento delle suddette attività ha portato alla determinazione delle liste di campionamento come sottoinsieme delle famiglie presenti in Lac geocodificate alle diverse Ace. Da tali liste sono stati estratti i campioni di famiglie a cui è stato inviato il questionario *long* al fine di inferire su tutta la popolazione le informazioni raccolte; alle famiglie non estratte per il campione è stato assegnato il modello *short*. Infine, alle unità non eleggibili per il campionamento e a tutte le famiglie dei Comuni più piccoli non coinvolti dalla strategia campionaria è stato inviato il modello in forma completa.

Le maggiori criticità si sono riscontrate nella fase di geocodifica delle famiglie delle Lac alle rispettive aree di censimento a causa della qualità delle fonti impiegate e per problemi di disallineamento delle basi di dati territoriali impiegate. Per alcuni Comuni si sono registrate delle difficoltà tecniche principalmente indotte dalla presenza di errori o assenza degli indirizzi presenti in Lac, dalla non piena completezza degli archivi di numeri civici disponibili e adoperati per la geocodifica, del contemporaneo impiego degli elenchi di sezioni di censimento (riferiti alle Basi territoriali del 2001 e a quelle del 2011).

Seppur in presenza di tali complicazioni, l'operazione di geocodifica si è completata con risultati pienamente soddisfacenti e utili alla formazione dei campioni di famiglie per il Censimento della popolazione e delle abitazioni del 2011. Sono state rimandate alle operazioni post-censuarie la soluzione delle problematiche connesse alla mancata o errata geocodifica. Per esempio, l'assegnazione del modello *short* alle famiglie di sezioni fuori dal disegno delle Ace (per un errore di indirizzo o di geocodifica) ha prodotto "mancate risposte" in quanto si sarebbe dovuto inviare il modello *long*; di conseguenza si è dovuto procedere con opportune procedure per l'imputazione dei dati mancanti. Invece, l'assegnazione di modelli *long* alle famiglie di sezioni appartenenti ad Ace diverse ha avuto conseguenze trascurabili sull'ampiezza del campione osservato (riduzione della numerosità dell'Ace originaria) e sull'accuratezza delle stime riferite al dominio erroneamente assegnato.

Le operazioni descritte in questo capitolo hanno rappresentato un primo passo per la creazione delle liste di campionamento impiegate dall'Istat per il disegno di campioni areali e da lista a supporto dell'attuale Censimento permanente della popolazione e delle abitazioni. Negli ultimi anni, infatti, l'Istat ha fatto un enorme investimento metodologico e informatico per migliorare sia la qualità delle informazioni territoriali presenti negli archivi amministrativi, sia l'accuratezza dell'operazione di georiferimento delle unità statistiche (individui, famiglie, abitazioni, edifici) al territorio. Questa attività ha portato, secondo un processo in continua evoluzione, alla realizzazione di una infrastruttura statistica di fondamentale importanza per il disegno di campioni di famiglie da lista, geocodificate a prefissati domini di campionamento, utili per le attuali e le future indagini Istat. In tal modo sarà possibile progettare campioni che possano tendere, come situazione ottimale, alla piena coincidenza tra campione pianificato e campione osservato per la garanzia di esaustività (sono tutte campionate) ed esclusività (sono campionate solo quelle) delle unità presenti nel dominio di indagine.

CAPITOLO 4

LE TECNICHE DI RILEVAZIONE⁶⁸

4.1 Introduzione

In questo capitolo verranno esaminate in dettaglio le tecniche di rilevazione utilizzate nelle diverse fasi delle operazioni sul campo e le strategie progettate *ad hoc* per la rilevazione di alcune sottopopolazioni difficili da rilevare o le cui caratteristiche richiedevano l'adozione di tecniche differenziate (la popolazione nelle convivenze anagrafiche, le persone senza tetto, le minoranze linguistiche e la popolazione residente nei Comuni del “cratere sismico” relativo al terremoto avvenuto in Abruzzo nell'aprile del 2009).

Per comodità di esposizione, nel corso del volume si fa riferimento a una suddivisione che comprende quattro fasi delle operazioni: creazione degli archivi di base, consegna, restituzione spontanea multicanale e completamento della rilevazione sul campo. Si tratta tuttavia di una sequenza logica più che temporale. Infatti, ad esclusione della fase relativa alla creazione degli archivi di base, le altre fasi della rilevazione (consegna dei questionari, restituzione spontanea multicanale e completamento della rilevazione sul campo) erano parzialmente sovrapposte:

- la consegna dei questionari non spediti poteva essere effettuata in contemporanea con la spedizione dei questionari o nelle fasi successive, secondo diverse valutazioni da parte del responsabile dell'Ucc;
- la restituzione spontanea cominciava il 9 ottobre 2011, quando le operazioni di spedizione dei questionari su una parte del territorio non erano ancora concluse, e proseguiva anche durante la fase di completamento della rilevazione del campo; allo stesso modo, l'Ucc doveva garantire il funzionamento dei Ccr per l'assistenza ai rispondenti e la ricezione dei questionari cartacei per tutta la durata della rilevazione, a partire dalla data di riferimento del censimento e fino alla data di chiusura delle operazioni censuarie;
- alcune operazioni concettualmente relative alla fase di completamento della rilevazione sul campo potevano essere effettuate contestualmente al recupero delle mancate risposte (ad esempio il recupero della sotto-copertura), ma potevano anche essere anticipate, se il responsabile dell'Ucc lo riteneva funzionale all'organizzazione della rilevazione sul territorio comunale. Inoltre, l'analisi dell'andamento dei questionari recuperati dai rilevatori ha mostrato che molti Ucc hanno scelto di inviarli sul campo prima dell'avvio ufficiale della fase di recupero delle mancate risposte⁶⁹.

4.2 La consegna dei questionari

Come detto, la prima fase del censimento assistito da lista consiste nella creazione degli archivi di base, a partire dall'acquisizione delle Lac utilizzate come lista delle unità di rilevazione. Come anticipato, al fine di ridurre il numero di rilevatori e, quindi, l'impatto sugli Ucc delle operazioni sul campo, la consegna dei questionari cartacei alle famiglie iscritte nelle anagrafi comunali è stata attuata tramite la tecnica del *mail-out* (spedizione postale).

La spedizione è stata effettuata da Poste italiane nell'arco di sei settimane (12 settembre - 22 ottobre) a cavallo della data di riferimento del Censimento⁷⁰, tramite circa 19 mila portalettere, sulla base degli elenchi

⁶⁸ Il Capitolo 4 è stato curato da Donatella Zindato, che ha redatto anche i paragrafi 4.1, 4.2, 4.3, 4.4 e 4.5.1. I paragrafi 4.5.2 e 4.5.4 sono stati redatti da Lorenzo Cassata. Il paragrafo 4.5.3 è stato redatto da Tiziana Tamburrano.

⁶⁹ Cfr. Romano (2017), paragrafo 2.4.

⁷⁰ La notevole durata delle operazioni di consegna, determinata dalla grande mole di questionari da consegnare (circa 25 milioni), comportava delle criticità. Per una parte delle famiglie la data di consegna avrebbe potuto precedere significativamente la data di riferimento del Censimento, aumentando il rischio che provassero a compilare il questionario prima della data di riferimento. Gli Ucc erano quindi stati invitati a chiarire ai cittadini che il questionario doveva essere compilato con riferimento al 9 ottobre, e che quindi poteva essere compilato *online*/restituito solo a partire da quella data. Inoltre, aumentava il rischio che il questionario venisse smarrito o che le famiglie si rivolgessero ai Ccr per richiedere un questionario sostitutivo perché altre famiglie allo stesso indirizzo avevano già ricevuto il questionario (i questionari avrebbero potuto essere recapitati allo stesso indirizzo in momenti diversi). Si richiedeva quindi di monitorare attentamente l'impiego dei questionari di scorta, rivolgendo tempestivamente all'Istat le richieste di ulteriori scorte.

comunali rielaborati dall'Istat e contenenti tutte le informazioni necessarie alla spedizione e al tracciamento del questionario: nominativo e indirizzo dell'intestatario della scheda di famiglia in anagrafe, codice questionario associato univocamente all'intestatario, *password* per la compilazione del questionario elettronico, tipo di questionario assegnato alla famiglia sulla base del numero di componenti e del piano di campionamento, presenza di componenti stranieri (per l'inserimento del modello di ausilio plurilingue).

Oltre al questionario, personalizzato con i dati del destinatario e con i relativi codici, il plico inviato alle famiglie conteneva:

- la lettera informativa sul Censimento a firma del Presidente dell'Istat;
- la guida alla compilazione;
- la busta da utilizzare per la restituzione del questionario.

Nei plichi destinati alle famiglie con almeno un componente straniero veniva inserito anche un foglio prestampato plurilingue, che conteneva un messaggio, tradotto in dieci lingue (albanese, francese, inglese, polacco, portoghese, rumeno, serbo, sloveno, spagnolo e tedesco), con il quale si comunicava ai cittadini stranieri che il questionario doveva essere compilato in lingua italiana e che la traduzione della lettera informativa, del facsimile del questionario e della guida alla compilazione erano disponibili sul sito internet del Censimento e presso i Centri comunali di raccolta, ai quali era possibile rivolgersi anche per chiedere assistenza alla compilazione. Inoltre, si specificava, in italiano, che i facsimile del questionario e le guide alla compilazione in arabo, bengali, bulgaro, cinese, macedone, russo, singalese, ucraino e urdu erano anch'essi scaricabili da internet e disponibili presso i Ccr⁷¹.

Qualche giorno prima dell'inizio della distribuzione dei questionari a cura del portalettere veniva lasciata, al portiere o nella cassetta delle lettere, un'informativa che riportava la descrizione dell'iniziativa istituzionale del Censimento, i tempi della rilevazione, il periodo di consegna dei plichi di rilevazione e le modalità di gestione di un secondo tentativo di consegna.

L'Istat comunicava all'Ucc la data di inizio e di conclusione delle operazioni di consegna dei questionari sul territorio del Comune, mentre l'esito individuale della spedizione veniva monitorato dall'Ucc attraverso Sgr, alimentato a cura dell'Istat con le informazioni ricevute dal vettore postale. Il questionario era classificato come "consegnato" nel caso di consegna diretta al destinatario, ad altro componente della famiglia o al portiere, o se imbucato all'indirizzo riportato sul questionario, nel caso in cui il nominativo sulla buca delle lettere corrispondesse esattamente a quello riportato sul questionario.

In caso di fallita consegna, il questionario non consegnato veniva recapitato da Poste italiane all'indirizzo prestampato sulla prima pagina del questionario (ovvero l'indirizzo stabilito dall'Ucc per la "restituzione" dei questionari consegnati dalle famiglie presso gli Uffici postali o non recapitati) e l'informazione veniva visualizzata in Sgr con lo stato "fallita consegna da vettore". In questo caso, infatti, come nel caso dei questionari esclusi a priori dal piano di stampa e spedizione (casi particolari, convivenze, indirizzi non riconosciuti dal normalizzatore), la consegna veniva effettuata a cura dell'Ucc tramite i rilevatori.

In particolare, erano escluse dal piano di spedizione, pur rientrando nel campo di osservazione della rilevazione, le convivenze anagrafiche (per le quali erano state stabilite modalità di rilevazione *ad hoc*⁷²) e, ovviamente (dati i tempi di acquisizione delle Lac), le famiglie iscritte in anagrafe dopo il 31 dicembre 2010 (e prima della data di riferimento del Censimento), visualizzate in Sgr con lo stato "variazione anagrafica - nuova fam/conv" a partire dal 15 novembre 2011 (a seguito dell'aggiornamento delle Lac alla data del censimento).

Altre categorie di unità di rilevazione escluse dalla spedizione, e che quindi dovevano essere rilevate a cura dell'Ucc, che provvedeva alla consegna tramite rilevatore (nel Diario di sezione tali record erano contrassegnati con stati *ad hoc*), erano:

⁷¹ Per facilitare la compilazione del questionario da parte dei cittadini stranieri, l'Istat aveva predisposto un facsimile del Foglio di famiglia (nelle due versioni, completa e ridotta) in diciassette lingue: albanese, arabo, bengali, bulgaro, cinese, francese, inglese, macedone, polacco, portoghese, rumeno, russo, serbo, singalese, spagnolo, ucraino e urdu. A questi si aggiungono i facsimili in tedesco e sloveno, predisposti grazie alla disponibilità delle traduzioni effettuate a uso dei cittadini italiani di lingua tedesca o slovena residenti nei territori tutelati delle province di Gorizia, Trieste e Udine e nella Provincia Autonoma di Bolzano, dove il questionario poteva essere compilato rispettivamente in tedesco o in sloveno, oltre che in italiano (cfr. paragrafo 4.5.3). Inoltre, ha predisposto la traduzione delle relative guide alla compilazione e delle due versioni (corrispondenti alle due tipologie, completa e ridotta, del questionario) della lettera informativa sul Censimento a firma del Presidente dell'Istat. I cittadini stranieri venivano informati della disponibilità di questi materiali attraverso un modello plurilingue inserito nei plichi inviati alle famiglie con almeno un componente straniero.

⁷² Cfr. paragrafo 4.5.1.

- le famiglie che in Lac risultavano avere più di 6 componenti: in questo caso il plico contenente il questionario personalizzato (al quale abbinare in fase di consegna i Fogli/Questionari aggiuntivi necessari, a seconda del numero di componenti) veniva inviato all'Ucc (nel Diario di sezione questi record erano contrassegnati con lo stato "da consegnare");
- le famiglie il cui indirizzo non avesse i requisiti per la normalizzazione: anche in questo caso il plico contenente il questionario personalizzato veniva inviato all'Ucc (nel Diario di sezione anche questi record erano contrassegnati con lo stato "da consegnare");
- i record che in Lac non erano chiaramente individuabili come famiglie o convivenze: in questo caso venivano visualizzati nel Diario di sezione con lo stato "consegna specializzata" e l'Ucc provvedeva alla rilevazione utilizzando i questionari di scorta (Foglio di famiglia o Foglio di convivenza, sulla base di quanto accertato dal rilevatore sul campo);
- le persone senza fissa dimora iscritte in anagrafe: anche questi record venivano visualizzati nel Diario di sezione con lo stato "consegna specializzata" e l'Ucc provvedeva alla rilevazione utilizzando i questionari di scorta⁷³.

Sulla base dell'organizzazione interna predisposta da ciascun Ucc in relazione alla gestione dei Ccr e alla suddivisione dei compiti tra personale di *back office* e rilevatori, ciascun responsabile valutava autonomamente il momento più opportuno per inviare i rilevatori sul campo per effettuare la consegna alle unità in Lac dei questionari non spediti (o non consegnati): se durante la fase di completamento della rilevazione, quando i rilevatori venivano inviati sul campo per il recupero delle mancate risposte (cfr. paragrafo 4.4.1), o durante le fasi precedenti.

Infine, erano escluse dalla spedizione anche le unità di rilevazione (famiglie, convivenze) dimoranti abitualmente nel territorio comunale ma non ancora iscritte in anagrafe (la cosiddetta sotto-copertura anagrafica). Anche in questo caso, spettava all'Ucc valutare autonomamente il momento più opportuno per inviare i rilevatori sul campo ai fini del recupero della sotto-copertura (cfr. paragrafo 4.4.2): se contestualmente al recupero delle mancate risposte o durante le fasi precedenti.

4.3 La restituzione multicanale

La restituzione multicanale dei questionari compilati cominciava a partire dalla mezzanotte tra l'8 e il 9 ottobre 2011 (data di riferimento del Censimento). Era prevista una prima fase (della durata di circa sei settimane) di esclusiva restituzione spontanea⁷⁴, in cui le famiglie potevano scegliere una tra le seguenti alternative:

1. compilare la versione elettronica del questionario;
2. compilare il questionario cartaceo e consegnarlo in busta chiusa presso un Ufficio postale (Up);
3. compilare il questionario cartaceo e consegnarlo presso uno dei Centri comunali di raccolta istituiti dall'Ucc.

A partire dal 21 novembre, si aggiungeva a questi canali un quarto canale di "restituzione" cioè il recupero delle mancate risposte da parte dei rilevatori. È opportuno però ricordare che, al fine di garantire la massima flessibilità ai rispondenti, era stato stabilito di mantenere i canali di restituzione spontanea attivi fino al termine della rilevazione. In molti casi, dunque, il passaggio del rilevatore fungeva da sollecito per la compilazione autonoma del questionario da parte della famiglia.

Si illustrano in dettaglio le tre modalità di restituzione multicanale sopra elencate:

⁷³ Sui questionari di scorta il codice questionario era prestampato ma, a differenza dei questionari personalizzati, non era abbinato a priori a una specifica famiglia e a un indirizzo. Al momento della consegna del questionario di scorta alla famiglia (in sostituzione di un questionario smarrito o deteriorato; per una famiglia iscritta in anagrafe dopo il 31 dicembre 2010; per la sotto-copertura anagrafica), il rilevatore doveva riportare sulla stampa del Diario di sezione il codice questionario, che andava successivamente associato alla famiglia in Sgr. Inoltre, per consentire la restituzione dei questionari di scorta (non prestampati con l'indirizzo di recapito del questionario) presso gli Uffici postali, al momento della consegna il rilevatore doveva apporre sull'ultima pagina del modello (nel riquadro in basso a sinistra) l'etichetta adesiva con l'indirizzo di recapito del questionario. Si ricorda infine che i questionari di scorta erano tutti in forma completa e da 6 componenti (Mod. Istat CP.1_6P_SC).

⁷⁴ Per semplicità, nel prosieguo si definisce fase di restituzione spontanea il periodo che andava dal 9 ottobre al 21 novembre, cioè la fase in cui erano attivi solo i canali di restituzione spontanea (prima dell'intervento dei rilevatori per il recupero delle mancate risposte).

1) Per accedere alla compilazione web, la famiglia doveva utilizzare la *password* stampata sul frontespizio del questionario ricevuto e il codice fiscale dell'intestatario della scheda di famiglia in anagrafe (*user ID*)⁷⁵. In caso di smarrimento del questionario ricevuto per posta, l'intestatario del Foglio di famiglia poteva chiedere il ripristino della *password* recandosi presso l'Ucc. A procedura conclusa, il sistema rilasciava una ricevuta di avvenuta compilazione. Inoltre, la compilazione del questionario veniva registrata in Sgr in tempo quasi reale. L'Ucc poteva visualizzare (ma non modificare) il questionario e quindi procedere alle operazioni di confronto censimento-anagrafe.

2) Il questionario poteva essere restituito presso tutti gli sportelli dedicati ai Servizi Postali in funzione presso i circa 14 mila Uffici postali (Up) presenti sul territorio nazionale⁷⁶, anche in un Comune diverso da quello di residenza (in quanto il sistema di tracciatura di Poste italiane era unico e quindi raccoglieva le informazioni provenienti da ogni Up sul territorio nazionale). Al momento della consegna del questionario, l'operatore postale leggeva il codice a barre del questionario e rilasciava la ricevuta dell'avvenuta restituzione⁷⁷. Le consegne multiple venivano gestite singolarmente per consentire il rilascio della singola ricevuta che riportava il timbro da sistema, il codice questionario, la data e l'ora dell'accettazione. Una volta emessa la ricevuta di avvenuta consegna, il questionario risultava restituito dal rispondente nel sistema di tracciatura di Poste italiane ma non automaticamente in Sgr. Infatti, l'informazione sull'avvenuta restituzione presso l'Ufficio postale non veniva trasmessa all'Istat in tempo reale⁷⁸ e solo dopo l'aggiornamento di Sgr (che avveniva periodicamente sulla base dei file trasmessi da Poste italiane) veniva visualizzata dall'Ucc con lo stato "restituito all'Ufficio postale".

Dagli Uffici postali i questionari cartacei venivano recapitati periodicamente all'indirizzo stabilito dall'Ucc⁷⁹. In particolare, era previsto che i Comuni fino a 150 mila abitanti ricevessero i colli dei questionari due volte a settimana mentre i Comuni con oltre 150 mila abitanti quattro volte la settimana.

3) I Centri comunali di raccolta dovevano essere operativi tra il 9 ottobre 2011 (data di riferimento) e la data di chiusura delle operazioni⁸⁰. Ad essi competevano principalmente due funzioni:

- garantire lo svolgimento delle attività necessarie alla ricezione dei questionari cartacei consegnati direttamente dai rispondenti;
- fornire chiarimenti e assistenza alla compilazione e restituzione del questionario.

Altra funzione fondamentale che poteva essere svolta presso i Centri comunali di raccolta era quella di incentivare i rispondenti alla compilazione del questionario elettronico, attraverso la predisposizione di una o più postazioni web a disposizione dei rispondenti che non disponessero di un computer e/o di una connessione a internet, e la presenza presso il Centro di operatori formati all'utilizzo del questionario elettronico, così da poter fornire assistenza ai rispondenti che non avessero familiarità con l'uso del PC e/o di internet.

In caso di consegna del questionario presso i Ccr, l'addetto alla ricezione, dopo aver verificato, tramite l'apposita funzione di Sgr, che il rispondente fosse residente nel Comune⁸¹, doveva effettuare la revisione del questionario in presenza del rispondente e assisterlo per le eventuali correzioni, così da evitare l'invio sul campo dei rilevatori per integrare/correggere i questionari. Nel caso in cui non fosse possibile effettuare la

⁷⁵ La compilazione web non poteva essere effettuata se nella Lac non era presente il codice fiscale dell'intestatario: in questo caso sul frontespizio del questionario non veniva stampata la *password* (si è trattato di una quota residuale pari a circa 89 mila famiglie).

⁷⁶ I 14 mila Uffici postali coprivano oltre il 96 per cento dei Comuni. 10 mila Up erano aperti 5 giorni a settimana per almeno 5 ore al giorno mentre i restanti 4 mila erano aperti con orari differenti, da 1 a 4 giorni la settimana. Il contratto con Poste italiane prevedeva che circa 22 mila operatori, debitamente e costantemente formati, fossero impegnati nelle operazioni di restituzione. Prevedeva inoltre, se necessario, di modificare gli orari di apertura per consentire la restituzione dei questionari e che, nei periodi di picco delle operazioni di restituzione, fossero attivati Uffici mobili (fino a un massimo di 20 su tutto il territorio nazionale) a supporto degli Up con maggior afflusso di rispondenti.

⁷⁷ Nel caso in cui l'errato posizionamento del questionario compilato nella busta di restituzione non consentiva la lettura del codice questionario, l'operatore consegnava al rispondente gli strumenti necessari (forbici e nastro adesivo) per poter riaprire la busta, posizionare correttamente il questionario e richiudere la busta.

⁷⁸ Il sistema di tracciatura di Poste italiane non comunicava direttamente con Sgr. Ciò ha comportato notevoli ritardi nell'aggiornamento del sistema per quanto riguarda i questionari cartacei restituiti presso gli Up, rendendo difficoltoso il monitoraggio delle operazioni e complicando le operazioni di recupero delle mancate risposte (che in molti casi si basavano su informazioni non aggiornate).

⁷⁹ Ovvero: per i questionari personalizzati, l'indirizzo stampato sull'ultima pagina del questionario; per i questionari di scorta, l'indirizzo riportato sulle etichette adesive stampate a cura dell'Ucc. Le etichette adesive venivano consegnate agli Ucc insieme ai questionari di scorta e dovevano essere prestampate a cura dell'Ucc prima della consegna ai rilevatori, con l'indirizzo a cui il questionario doveva essere recapitato nel caso in cui la famiglia avesse optato per la restituzione presso gli Uffici postali.

⁸⁰ Si ricorda che la data di chiusura delle operazioni sul campo era differenziata per classe di ampiezza demografica del Comune.

⁸¹ Nel caso in cui il rispondente non fosse risultato residente nello specifico Comune, doveva essere invitato a portare il questionario presso un Ufficio postale o presso uno dei Centri comunali di raccolta istituiti presso il proprio Comune di residenza.

revisione sul momento, era necessario verificare la presenza sul questionario di un recapito (numero di telefono e/o indirizzo di posta elettronica) da utilizzare per l'eventuale necessità di ricontattare la famiglia a seguito della revisione. L'addetto doveva quindi rilasciare al rispondente la ricevuta dell'avvenuta consegna, registrare la restituzione in Sgr e biffare la casella corrispondente (a seconda che il questionario fosse stato revisionato o meno) sul questionario.

Queste operazioni erano fondamentali per monitorare in tempo reale l'andamento della rilevazione e gestire efficacemente la successiva fase di rilevazione sul campo, in quanto solo la tempestiva registrazione dell'arrivo dei questionari cartacei consentiva all'Ucc di evitare l'invio dei rilevatori presso famiglie che avevano già provveduto alla restituzione del questionario.

4.3.1 Il monitoraggio della restituzione multicanale tramite Sgr

Dopo il primo popolamento di Sgr, il sistema veniva alimentato da parte dell'Istat con i dati sui flussi dei questionari compilati via web e dei questionari consegnati dalle famiglie presso gli Uffici postali. Doveva invece essere alimentato dall'Ucc per le consegne dei questionari effettuate tramite rilevatore, per i questionari restituiti dai rispondenti presso i Ccr o ai rilevatori o restituiti presso gli Up e poi recapitati all'Ucc, per le famiglie non presenti in Lac rilevate nella fase di recupero della sotto-copertura.

In particolare, presso i Ccr o altra sede designata dal responsabile, l'Ucc doveva provvedere alla registrazione in Sgr dell'arrivo dei questionari cartacei (oltre che alla revisione degli stessi e alla compilazione dei modelli riepilogativi). In proposito, è opportuno ricordare i questionari cartacei che confluivano presso l'Ucc:

- i questionari consegnati dai rispondenti presso i Centri comunali di raccolta;
- i questionari consegnati dai rispondenti presso gli Uffici postali e successivamente recapitati all'indirizzo stabilito dall'Ucc, prestampato sull'ultima pagina del questionario;
- i questionari recuperati sul campo dai rilevatori.

Il monitoraggio della rilevazione attraverso Sgr consentiva di ridurre il numero di interventi sul campo dei rilevatori. Sulla base delle informazioni inserite in Sgr era possibile visualizzare e stampare i rapporti riassuntivi relativi ai questionari non ancora restituiti, e ai questionari via via acquisiti. L'Ucc poteva inoltre controllare la situazione relativa a ciascuna famiglia della sezione attraverso il Diario di Sezione, ottenendo ad esempio, l'elenco ordinato per indirizzo delle famiglie che non avevano ancora restituito il questionario, da sollecitare telefonicamente (nel caso in cui il Comune disponesse di un recapito telefonico) o presso cui inviare i rilevatori, per verificarne l'effettiva dimora abituale all'indirizzo della Lac ed eventualmente procedere all'intervista o a sollecitare la compilazione/restituzione del questionario tramite altri canali.

4.4 Il completamento della rilevazione sul campo

A partire dal 21 novembre 2011, e fino alla chiusura delle operazioni sul campo, gli Ucc dovevano curare il completamento della rilevazione sul campo. In particolare, attraverso l'invio dei rilevatori sul campo, dovevano coordinare:

- il recupero delle mancate risposte, cioè dei questionari non restituiti spontaneamente dalle famiglie (cfr. paragrafo 4.4.1);
- il recupero della sotto-copertura anagrafica, ovvero la rilevazione delle persone dimoranti abitualmente nel Comune ma non iscritte in anagrafe, la rilevazione delle abitazioni non occupate e la rilevazione degli edifici nei Comuni e nelle località abitate dove non era stata effettuata la Rnc (cfr. paragrafo 4.4.2).

Inoltre, in questa fase veniva effettuata anche la consegna dei questionari non spediti/non consegnati (cfr. paragrafo 4.2) e la rilevazione delle convivenze (cfr. paragrafo 4.5.1), qualora il responsabile dell'Ucc avesse scelto di concentrare in questa fase lo svolgimento di tutte le operazioni sul campo.

In parallelo, l'Ucc doveva continuare a garantire lo svolgimento delle attività avviate nella fase di restituzione spontanea: monitorare l'esito della spedizione dei questionari e della restituzione multicanale; gestire i Ccr per fornire assistenza ai rispondenti; garantire la ricezione, la revisione e la registrazione in Sgr dei questionari cartacei; effettuare il confronto censimento-anagrafe.

Nel corso dello svolgimento delle attività sul campo, il rilevatore doveva tornare frequentemente presso la sede designata dal responsabile dell'Ucc al fine di consentire il tempestivo aggiornamento di Sgr. In

particolare, secondo quanto concordato con il responsabile/coordinatore, il rilevatore doveva tornare periodicamente presso l'Ucc per:

- consegnare i questionari recuperati sul campo;
- consegnare la versione cartacea del Diario sulla quale erano registrati gli esiti relativi a ciascun questionario da recuperare/consegnare/revisionare, al fine di consentire l'aggiornamento del Diario di sezione in Sgr e ottenere una nuova versione stampata del Diario aggiornata sulla base del lavoro svolto sul campo e del lavoro di *back office* effettuato a cura dell'Ucc;
- acquisire i materiali necessari alla prosecuzione della rilevazione (questionari da fornire in sostituzione di questionari smarriti, Fogli/Questionari aggiuntivi, etichette, eccetera).

L'aggiornamento costante e tempestivo del Diario, secondo quanto stabilito in autonomia dal responsabile dell'Ucc, poteva essere curato dallo stesso rilevatore o da un altro operatore designato, ma era essenziale al fine di gestire efficacemente la rilevazione sul campo e l'insieme delle operazioni che competevano all'Ucc.

4.4.1 Il recupero delle mancate risposte

Al fine del recupero delle mancate risposte, il rilevatore percorreva l'intera sezione di censimento in base alle informazioni riportate nel Diario di sezione, aggiornato dalle informazioni inserite in Sgr, e si recava presso le famiglie che risultavano non avere restituito il questionario. Oltre al Diario di sezione, portava con sé un congruo numero di questionari di scorta, di fogli aggiuntivi, di etichette adesive e di ricevute da compilare e consegnare all'atto della ricezione dei questionari, eventuali questionari compilati da revisionare e, nel caso in cui la consegna dei questionari non postalizzati non si fosse svolta nelle fasi precedenti, anche i questionari da recapitare a cura dell'Ucc.

Al fine di consentire al rilevatore di effettuare il recupero mirato delle mancate risposte ed evitare di recarsi presso famiglie che avevano già restituito il questionario, nel Diario di sezione erano visualizzate le informazioni relative a tutte le unità di rilevazione della sezione (comprese, quindi, le informazioni relative alle famiglie da non contattare). Inoltre, al fine di individuare eventuali unità non presenti in Lac, nel Diario di sezione comparivano, ordinate per indirizzo, anche righe (una per ogni individuo) relative alla potenziale sotto-copertura (cfr. paragrafo 4.4.2).

Ad esempio, in corrispondenza dello stesso indirizzo, il rilevatore avrebbe potuto trovare una riga corrispondente a una famiglia presente in Lac che non aveva restituito il questionario (identificata dallo stato "mancata risposta"), una riga corrispondente a una famiglia presente in Lac che aveva compilato il questionario elettronico (identificata dallo stato "restituito web"), una riga corrispondente a una famiglia con più di 6 componenti a cui consegnare il questionario (identificata dallo stato "consegna specializzata"), una riga corrispondente a una convivenza (identificata dallo stato "convivenza in convenzione") e tre righe corrispondenti a unità potenzialmente abitative, presso le quali verificare la presenza di eventuali famiglie non incluse nella Lac (cfr. paragrafo 4.4.2).

Ai fini del recupero delle mancate risposte, il rilevatore doveva, quindi, individuare nel Diario di sezione gli alloggi corrispondenti alle righe del Diario con stato "mancata risposta" o "salvataggio parziale via web". Nel caso dei questionari con stato "mancata risposta", se la famiglia era presente nell'alloggio il rilevatore offriva la propria collaborazione per la compilazione immediata del questionario o, nel caso in cui non fosse possibile, sollecitava la compilazione/restituzione del questionario. Se la famiglia non era presente nell'alloggio, prendeva nota ed effettuava ulteriori tentativi di contatto, fino alla compilazione/restituzione del questionario. Nel caso in cui l'abitazione fosse risultata non occupata o utilizzata a scopo non abitativo (ufficio, studio professionale, eccetera), registrava l'informazione nel Diario di sezione (e successivamente in Sgr). Nel caso in cui l'alloggio fosse occupato da un'altra famiglia/convivenza o da persone non dimoranti abitualmente presenti alla data di riferimento del censimento, registrava l'informazione nel Diario di sezione e consegnava un questionario non personalizzato alla nuova unità da rilevare.

Nel caso di questionari parzialmente compilati via web, il rilevatore doveva sollecitare la famiglia a completare la compilazione web, eventualmente offrendo la propria collaborazione, o a effettuare la compilazione del questionario cartaceo. Nel caso in cui fosse stato smarrito il questionario cartaceo, il rilevatore doveva invitare il rispondente a recarsi presso il Ccr per il ripristino della *password* o consegnare un questionario di scorta.

4.4.2 Il recupero della sotto-copertura

Obiettivo del recupero della sotto-copertura era l'individuazione di unità (famiglie o convivenze) dimoranti abitualmente nel Comune ma non presenti in Lac e la rilevazione esaustiva delle abitazioni non occupate. A tal fine, il rilevatore doveva percorrere la sezione verificando, per ciascun segnale di potenziale sotto-copertura presente nel Diario, se fossero presenti:

- famiglie e convivenze dimoranti abitualmente nel Comune ma non presenti in Lac;
- abitazioni non occupate.

Come anticipato, le operazioni connesse al recupero della sotto-copertura si differenziavano a seconda che il Comune avesse o meno effettuato la Rilevazione dei numeri civici. Infatti, la presenza nel Diario di sezione delle informazioni derivanti dalla Rnc (in aggiunta a quelle derivanti dalla Lifa) forniva al rilevatore indicazioni puntuali, consentendo il recupero mirato della sotto-copertura. In assenza di queste informazioni (Comuni non Rnc e sezioni di nucleo abitato e case sparse dei Comuni Rnc), il rilevatore doveva invece effettuare la ricognizione esaustiva del territorio della sezione al fine di integrare le informazioni presenti nel Diario (segnali di potenziale sotto-copertura derivanti dalla Lifa), in quanto in questo caso non era stato possibile stimare gli interni potenzialmente abitativi non corrispondenti a famiglie in Lac (cfr. paragrafo 2.3).

4.4.2.1 Il recupero della sotto-copertura nei Comuni Rnc

Ai fini del recupero della sotto-copertura, i Comuni Rnc⁸² non costituivano un insieme omogeneo poiché non tutta la loro superficie era stata interessata dalla Rnc. Le informazioni che, nel Diario di sezione, guidavano il rilevatore per il recupero della sotto-copertura, erano quindi differenziate a seconda del tipo di sezione: nelle sezioni di centro abitato erano presenti sia informazioni da Lifa che da Rnc, mentre nelle sezioni di nucleo abitato e case sparse erano presenti solo informazioni da Lifa.

Più precisamente, per le sezioni dove era stata completata la Rnc, nel Diario di sezione comparivano due tipi di segnali di potenziale sotto-copertura:

- righe con stato questionario “sotto-copertura da Lifa”, con nome, cognome e indirizzo del potenziale intestatario (corrispondenti a individui presenti nella Lifa del Comune e nella Lac di un altro Comune o presenti nella Lifa del Comune e non presenti in alcuna Lac);
- righe corrispondenti alle unità potenzialmente abitative stimate sulla base della Rnc, con il solo indirizzo valorizzato (corrispondenti alla differenza tra il numero di interni ad uso abitativo rilevato dalla Rnc e il numero di famiglie iscritte in anagrafe a quell'indirizzo).

Per ciascuna delle righe relative alla potenziale sotto-copertura (da Lifa o da Rnc), il rilevatore doveva quindi verificare se esistevano famiglie/convivenze da rilevare. In particolare, dopo aver individuato l'alloggio corrispondente al segnale di potenziale sotto-copertura anagrafica, se nell'alloggio era effettivamente dimorante abitualmente una famiglia/convivenza, il rilevatore consegnava il questionario offrendo la propria collaborazione per la compilazione immediata o, nel caso in cui non fosse possibile, informava i rispondenti sulle modalità di restituzione del questionario. Nel caso in cui l'abitazione fosse non occupata o fosse utilizzata a scopo non abitativo (ufficio, studio professionale, eccetera), registrava l'informazione nel Diario di sezione (e successivamente in Sgr). Se l'alloggio era vuoto al passaggio del rilevatore ma presumibilmente occupato, prendeva nota ed effettuava ulteriori tentativi di contatto, fino alla consegna del questionario o alla classificazione dell'alloggio come non occupato.

Infine, se il numero di alloggi (non corrispondenti a famiglie in Lac) individuati presso l'indirizzo era inferiore rispetto alle righe di potenziale sotto-copertura presenti nel Diario di sezione, le righe eccedenti dovevano essere chiuse con l'esito “sotto-copertura non trovata”. Lo stesso stato di “sotto-copertura non trovata” doveva essere utilizzato nel caso in cui a una stessa famiglia corrispondessero diverse righe di “sotto-copertura da Lifa”⁸³. In questo caso, il questionario consegnato doveva essere associato alla riga corrispondente alla persona di riferimento della famiglia, e le altre righe dovevano essere chiuse con lo stato “sotto-copertura non trovata”. Se, al contrario, a un indirizzo presente nel Diario venivano individuati più interni abitativi

⁸² Come detto, il campo di osservazione della Rnc era costituito dalle sezioni di centro abitato dei comuni con almeno 20 mila abitanti o capoluogo di provincia alla data del 1° gennaio 2008.

⁸³ Poiché gli archivi amministrativi che costituivano la Lifa erano archivi individuali, nella Lifa avrebbero potuto essere incluse persone appartenenti alla stessa famiglia. Quando nel Diario erano presenti più righe di “sotto-copertura da Lifa” allo stesso indirizzo, il rilevatore doveva verificare che si trattasse di individui appartenenti alla stessa famiglia e consegnare un unico questionario.

rispetto alla potenziale sotto-copertura stimata sulla base di Lifa e Rnc (e dunque rispetto alle righe presenti nel Diario), il rilevatore doveva aggiungere le righe corrispondenti nel Diario di sezione con il relativo stato (“questionario consegnato” o “abitazione non occupata”).

Per le sezioni di nucleo abitato e case sparse, dove non era stata effettuata la Rnc, il rilevatore doveva invece procedere con le stesse modalità previste per il recupero della sotto-copertura nei Comuni non Rnc (cfr. paragrafo 4.4.2.2).

4.4.2.2 Il recupero della sotto-copertura nei Comuni non Rnc

Obiettivo del recupero della sotto-copertura nei Comuni non Rnc (e nelle sezioni di nucleo abitato e case sparse dei Comuni Rnc) era l’individuazione di unità (famiglie o convivenze) dimoranti abitualmente nel Comune, ma non presenti in Lac e la rilevazione esaustiva delle abitazioni non occupate. Gli strumenti a disposizione del rilevatore erano:

- l’Itinerario di sezione, cioè l’elenco puntuale degli indirizzi (toponimi e numeri civici) che rientrano nei confini della sezione di censimento;
- il Diario di sezione, contenente l’elenco ordinato per indirizzo di tutte le unità di rilevazione della sezione, inclusi (“sotto-copertura da Lifa”).

Ai fini del recupero della sotto-copertura, nel Diario comparivano solo le righe corrispondenti ai segnali di potenziale sotto-copertura stimati sulla base della Lifa (“sotto-copertura da Lifa”). Portando con sé i modelli necessari per la rilevazione (oltre all’Itinerario di sezione e alla versione stampata del Diario), il rilevatore doveva effettuare la ricognizione esaustiva della sezione, percorrendo tutte le vie presenti nell’Itinerario di sezione, a partire dal numero civico più basso fino al più alto, ed esplorando tutti gli accessi esterni individuati lungo il percorso, ai fini di:

- verificare l’esistenza di famiglie o convivenze per ciascuna delle righe relative alla “sotto-copertura da Lifa” (come nelle sezioni di centro abitato);
- individuare eventuale ulteriore sotto-copertura (oltre a quella della Lifa) e rilevare tutte le abitazioni non occupate.

In particolare, il rilevatore doveva esplorare tutti gli accessi esterni individuati lungo il percorso e controllare tutti gli interni presenti presso ciascuno di essi, individuando gli interni corrispondenti alle righe di “sotto-copertura da Lifa” presenti nel Diario di sezione e aggiungendo le righe corrispondenti agli ulteriori interni abitativi (eccedenti rispetto alla potenziale sotto-copertura stimata sulla base della Lifa). Presso ciascun alloggio doveva rilevare le famiglie/convivenze ivi dimoranti abitualmente (o classificare gli alloggi come non occupati o utilizzati a scopo non abitativo). Allo stesso modo, il rilevatore doveva esplorare e aggiungere al Diario di sezione eventuali indirizzi ricadenti nella sezione ma non presenti nel Diario. Se invece il numero di alloggi (non corrispondenti a famiglie in Lac) individuati presso un determinato indirizzo era inferiore rispetto alle righe di potenziale sotto-copertura presenti nel Diario di sezione, le righe eccedenti dovevano essere chiuse con l’esito “sotto-copertura non trovata”.

4.5 Casi particolari della rilevazione della popolazione

4.5.1 La rilevazione delle convivenze⁸⁴

Ai fini del censimento si intende per *convivenza* un insieme di persone che, senza essere legate da vincoli di matrimonio, parentela, affinità e simili, conducono vita in comune per differenti motivi⁸⁵.

Come anticipato, le convivenze erano escluse dal piano di stampa e spedizione dei questionari, ma le modalità di rilevazione delle stesse si differenziavano in base al tipo di convivenza:

- convivenze in convenzione (convivenze dei Ministeri della difesa, della Giustizia e dell’internò);
- convivenze non in convenzione.

⁸⁴ Cfr. Istat (2011), paragrafo 6.1.

⁸⁵ Più precisamente, le convivenze si distinguono in convivenze anagrafiche e convivenze non anagrafiche. La convivenza anagrafica è un insieme di persone normalmente coabitanti per motivi religiosi, di cura, di assistenza, militari, di pena e simili, aventi dimora abituale nello stesso Comune (art. 5 comma 1 del Regolamento Anagrafico - D.P.R. 30 maggio 1989, n. 223). La convivenza non anagrafica è una struttura residenziale collettiva che, in genere, ospita solo persone non dimoranti abitualmente (alberghi, alcuni ospedali, eccetera).

Le convivenze in convenzione ricevevano le credenziali per l'accesso alla compilazione web, mentre le convivenze non in convenzione dovevano essere censite a cura dell'Ucc tramite consegna del questionario cartaceo. In entrambi i casi, il questionario doveva essere compilato dal responsabile della convivenza o da un'altra persona incaricata.

4.5.1.1 *La rilevazione delle convivenze in convenzione*

Al fine di facilitare l'attività di rilevazione da parte degli Ucc e di migliorare la qualità delle informazioni rilevate, l'Istat si è avvalso della collaborazione delle Amministrazioni centrali dello Stato per la rilevazione di alcune categorie di convivenze.

Sulla base di accordi specifici stipulati con il Ministero della difesa, con il Ministero della giustizia e con il Ministero dell'interno, le convivenze in convenzione comprendevano:

- le convivenze militari (caserme dell'Esercito, della Marina, dell'Aeronautica, dei Carabinieri e gli ospedali e carceri militari);
- le convivenze afferenti al Ministero della giustizia (istituti penitenziari, inclusi gli istituti per minorenni e gli ospedali psichiatrici giudiziari);
- le convivenze afferenti al Ministero dell'interno (caserme della Polizia di Stato e dei Vigili del fuoco, Centri di soccorso e di prima accoglienza, Centri identificazione ed espulsione e Centri assistenza richiedenti asilo).

Il censimento delle convivenze militari è avvenuto esclusivamente tramite compilazione del questionario elettronico, a cura del Ministero della difesa, che monitorava direttamente la compilazione dei questionari, provvedendo ad effettuare i solleciti sulla base dei report forniti dall'Istat. Le convivenze militari erano visibili in Sgr ma, a differenza delle altre unità di rilevazione, lo stato di lavorazione dei questionari non poteva essere modificato dall'Ucc.

Per le convivenze afferenti al Ministero della giustizia e al Ministero dell'interno, invece, l'Istat ha fornito ai responsabili delle convivenze le credenziali per la compilazione del questionario elettronico (tramite i rispettivi Ministeri), mentre la rilevazione è stata monitorata e completata a cura degli Ucc. In particolare, a partire dal 21 novembre 2011 (ovvero al termine della fase di restituzione spontanea), i rilevatori dovevano procedere alla consegna e al ritiro del questionario cartaceo presso quelle convivenze per quali il questionario non risultasse "restituito web".

Ai fini della consegna del questionario cartaceo, era necessario innanzitutto stabilire il tipo di convivenza (con almeno una persona dimorante abitualmente o solo con persone non dimoranti abitualmente)⁸⁶ e conteggiare il numero di persone dimoranti abitualmente, per consegnare uno o più questionari (ciascun questionario conteneva sette fogli individuali) e/o fogli individuali aggiuntivi.

Inoltre, poiché si trattava di questionari non personalizzati, prima della consegna il rilevatore doveva compilare il riquadro sul frontespizio del questionario con tutti i dati identificativi della convivenza e del responsabile, e alcuni riquadri sul retro con le informazioni relative al tipo e al numero di questionari consegnati.

Infine, offriva la propria collaborazione per la compilazione del questionario o, nel caso in cui non fosse possibile compilare il questionario sul momento, concordava un appuntamento per il ritiro del questionario compilato, aggiornando sul Diario di sezione (e successivamente in Sgr) lo stato della riga corrispondente alla convivenza e riportando il codice del questionario consegnato.

4.5.1.2 *La rilevazione delle convivenze non in convenzione*

La rilevazione delle convivenze non in convenzione era condotta tramite i rilevatori, che provvedevano alla consegna e al ritiro del questionario cartaceo sulla base del Diario di sezione. La consegna dei questionari poteva essere effettuata in parallelo alla spedizione dei questionari alle famiglie, o nelle fasi successive, secondo valutazioni da parte del responsabile dell'Ucc.

Come per la consegna del questionario cartaceo alle convivenze in convenzione, il rilevatore doveva determinare il tipo di convivenza (se con almeno una persona dimorante abitualmente o solo con persone non dimoranti abitualmente) e conteggiare il numero di persone dimoranti abitualmente, in base al quale stabilire

⁸⁶ Erano stati predisposti due tipi di questionari - Mod. Istat CP.2 o Mod. Istat CP.2NAD - perché le persone non dimoranti abitualmente venivano solo conteggiate cioè non compilavano il questionario individuale (cfr. Istat, 2017b, Appendice).

quali e quanti questionari/fogli individuali aggiuntivi consegnare. Contestualmente doveva aggiornare la riga corrispondente alla convivenza nel Diario di sezione, riportando la data concordata per il ritiro del questionario nel caso in cui non fosse possibile compilare il questionario sul momento.

Le stesse operazioni dovevano essere effettuate nel caso in cui, nel corso delle operazioni di recupero della sotto-copertura, venisse individuata una convivenza non presente in Lac. In questo caso, dopo aver riportato sul questionario le informazioni propedeutiche alla consegna, il rilevatore doveva associare il codice del questionario consegnato alla riga di sotto-copertura presente nel Diario di sezione o inserire una riga ex novo.

4.5.2 La rilevazione delle persone senza tetto e delle persone senza fissa dimora⁸⁷

La rilevazione delle persone senza tetto e delle persone senza fissa dimora⁸⁸, nell'ambito del censimento della popolazione, richiede l'adozione di un approccio dedicato in quanto si tratta di una popolazione ad elevato rischio di sotto-copertura, cosiddetta difficile da rilevare (*hard to count*). Al censimento del 2011 sono state adottate strategie differenziate a partire dalle informazioni presenti in Lac⁸⁹. In particolare, tra le persone senza tetto/senza fissa dimora presenti in Lac era possibile distinguere due casi:

1. le persone senza fissa dimora iscritte a un indirizzo fittizio (ad esempio "via della Casa comunale"), indicato dall'Ucc in fase di trasmissione all'Istat della Lac⁹⁰ (indirizzo fittizio appositamente istituito dal Comune, come previsto dal Regolamento anagrafico per l'iscrizione in anagrafe di persone senza fissa dimora) o iscritte all'indirizzo del Comune/Municipio (categoria assimilabile a quella degli iscritti a un indirizzo fittizio - in questo caso il Comune aveva utilizzato l'indirizzo del Comune/Municipio come indirizzo convenzionale invece di istituire un indirizzo fittizio);
2. persone iscritte in anagrafe all'indirizzo corrispondente a strutture di accoglienza presso le quali era possibile eleggere residenza (indirizzi individuati preventivamente dall'Istat attraverso il confronto tra gli indirizzi delle Lac e gli indirizzi raccolti nell'ambito dell'indagine Istat sui servizi per le persone senza dimora).

Nel primo caso, gli iscritti a un indirizzo convenzionale (fittizio o corrispondente all'indirizzo del Comune/Municipio), ovviamente esclusi dalla spedizione, venivano censiti a cura dell'Ucc, nel caso in cui fossero reperibili sul territorio comunale alla data del censimento o nel caso di persone comunque domiciliate nel Comune (per esempio, i giostrai).

Nel secondo caso, in presenza di indirizzi realmente esistenti ma corrispondenti ad associazioni o strutture di accoglienza, trattandosi di indirizzi reali e non fittizi, in linea di principio era possibile che fossero associati a edifici residenziali aventi al proprio interno non solo la sede dell'associazione ma anche persone residenti in abitazione. Per ridurre il carico di lavoro degli Ucc in fase di recupero delle mancate risposte (nel caso in cui l'associazione non fosse riuscita a consegnare i questionari agli interessati), si è scelto di spedire i questionari personalizzati nel caso di indirizzi con un numero di iscritti in anagrafe inferiore a 200, comunicandolo agli Ucc, affinché provvedessero a informare i responsabili delle strutture. Per gli indirizzi presso i quali risultavano iscritte più di 200 famiglie, sono invece state adottate strategie differenziate, sulla base di quanto concordato con i singoli Ucc: in alcuni casi i questionari personalizzati sono stati inviati direttamente all'indirizzo (se, ad esempio, si trattava della sede dei Servizi sociali del Comune); in altri casi sono stati inviati all'Ucc, che ha provveduto a contattare il responsabile della struttura per coinvolgerlo nella rilevazione; in altri ancora l'Ucc ha inviato i rilevatori muniti di questionari di scorta presso la sede dell'associazione, previo contatto con la stessa⁹¹.

⁸⁷ Cfr. Istat, 2011, paragrafo 6.2.

⁸⁸ Si definisce persona senza fissa dimora chi non abbia in alcun Comune quella dimora abituale che è elemento necessario per l'accertamento della residenza (girovaghi, artisti di imprese-spettacoli itineranti, commercianti e artigiani ambulanti, eccetera). Per persone senza tetto si intendono le persone senza fissa dimora che non hanno alcun domicilio (né abitazione, né altro tipo di alloggio - persone che vivono in strada, sotto i ponti, eccetera).

⁸⁹ Il regolamento anagrafico prevede per le persone senza fissa dimora il diritto all'iscrizione in anagrafe. Ai sensi dell'art. 2 della vigente legge anagrafica (legge n. 1228 del 24 dicembre 1954), la persona senza fissa dimora "si considera residente nel Comune ove ha il domicilio e, in mancanza di questo, nel Comune di nascita". Ai fini del censimento, per le persone senza fissa dimora si considera Comune di dimora abituale quello di iscrizione anagrafica.

⁹⁰ Gli iscritti in Lac a indirizzi fittizi dichiarati tali all'atto della trasmissione all'Istat delle Lac al 1° gennaio 2011 erano 40.473 (di cui 20.421 effettivamente censiti come persone senza tetto).

⁹¹ In particolare, sono stati presi accordi con gli Ucc di Roma, Milano, Livorno, Genova e Firenze.

Inoltre, per garantire l'esaustività del conteggio, era necessario rilevare le persone senza tetto presenti sull'intero territorio comunale. Per evitare duplicazioni, la rilevazione doveva avvenire in modo simultaneo in tutto il Comune. A tal fine, gli Ucc erano stati invitati ad effettuare una ricognizione preventiva del territorio comunale, per individuare aree e specifici luoghi (stazioni ferroviarie, parchi pubblici, eccetera) nelle quali era più diffusa la presenza di persone senza tetto, presso le quali inviare i rilevatori alla data di riferimento del censimento (9 ottobre 2011), muniti dei questionari da compilare con le notizie fornite dagli interessati.

4.5.3 La rilevazione delle minoranze linguistiche

I territori in cui risiedono cittadini appartenenti alle minoranze linguistiche riconosciute e di interesse ai fini del censimento sono la Provincia di Trento, la Provincia di Bolzano e la Regione Friuli-Venezia Giulia (Province di Trieste, Gorizia e Udine)⁹². Per questi territori è stato necessario prevedere strumenti e procedure atte a garantire la tutela dei diritti previsti dalle norme in vigore e, per quanto riguarda le Province autonome di Trento e di Bolzano, ad armonizzare, regolare e coordinare le attività relative al 15° Censimento generale della popolazione e delle abitazioni con le attività relative alle rilevazioni sulla consistenza e la dislocazione territoriale di alcuni gruppi linguistici⁹³.

A tal fine, l'Istat ha predisposto specifiche intese con il Servizio di statistica della Provincia autonoma di Trento e l'Istituto provinciale di statistica della Provincia autonoma di Bolzano, di seguito denominato Astat, per il censimento nei rispettivi territori, e ha regolamentato con la Circolare n. 8 dell'8 settembre 2011, prot. n. 6858, le attività censuarie nelle Province della regione Friuli-Venezia Giulia interessate dalla tutela del diritto all'uso della lingua slovena.

4.5.3.1 La rilevazione censuaria nella Provincia autonoma di Trento

Nell'ambito della strategia censuaria del 2011, la rilevazione nella Provincia di Trento si è svolta con alcune specificità di processo. Le rilevazioni sono state due:

- il 15° Censimento generale della popolazione e delle abitazioni, per il quale l'organo responsabile è l'Istat;
- la "Rilevazione sulla consistenza e la dislocazione territoriale degli appartenenti alle popolazioni di lingua ladina, mòchena e cimbra", la cui responsabilità è affidata alla Provincia autonoma di Trento. Si tratta di apposite dichiarazioni rese dagli interessati, senza alcun obbligo di risposta. Per i cittadini di età inferiore ai 14 anni, la dichiarazione è resa dai genitori o da coloro che ne esercitano la patria potestà.

La rilevazione censuaria si è svolta come in tutti i Comuni del territorio nazionale ad eccezione dei seguenti 11 Comuni: Campitello di Fassa, Canazei, Mazzin, Moena, Pozza di Fassa, Soraga e Vigo di Fassa (località ladine); Fierozzo, Frassilongo e Palù del Fersina (località mòchene); Luserna (località cimbra). Questi Comuni rappresentano le località su cui insistono le comunità da tutelare, e quindi, come previsto dalla normativa vigente (nello specifico il D.lgs. 16 dicembre 1993, n. 592), si è garantito anche l'uso della lingua della minoranza locale.

Negli 11 Comuni i rilevatori hanno consegnato alle famiglie:

- ✓ il questionario in forma completa (*long form*) del 15° Censimento generale della popolazione;
- ✓ il facsimile del questionario nella lingua della minoranza, come *ausilio* alla compilazione;
- ✓ il modello di Rilevazione sulla consistenza e la dislocazione territoriale degli appartenenti alle popolazioni di lingua ladina, mòchena e cimbra in lingua italiana e nella lingua della minoranza locale, completo di lettera informativa alle famiglie.

In tutti gli altri Comuni della Provincia di Trento, le famiglie hanno ricevuto tramite posta:

⁹² La tutela delle minoranze linguistiche è garantita dall'articolo XXX della Costituzione (*dello Stato italiano concorre alla tutela delle minoranze linguistiche e riconosce loro il diritto di usare la propria lingua nei rapporti sociali e amministrativi*) e normata dalla legge 15 dicembre 1999 n. 482 "Norme a tutela delle minoranze linguistiche storiche" (e relativo regolamento approvato con D.P.R. 2 maggio 2001 n. 345) e legge 23 febbraio 2001 n. 38 "Norme a favore della minoranza slovena del Friuli-Venezia Giulia". Il presupposto per l'operatività di tali norme è definito dalla *delimitazione territoriale* nella quale insiste la comunità da tutelare.

⁹³ Specifiche disposizioni normative prevedono che, *in concomitanza* con il Censimento della popolazione, vengano acquisite informazioni sulla consistenza e la dislocazione territoriale di alcuni gruppi linguistici. In particolare:

- il D.lgs. 16 dicembre 1993 n. 592 stabilisce di effettuare la rilevazione relativa agli appartenenti alle minoranze di lingua ladina, mòchena e cimbra residenti nella provincia di Trento
- il D.P.R. 26 luglio 1976 n. 752 stabilisce di effettuare la rilevazione relativa ai gruppi linguistici italiano, tedesco e ladino nella provincia di Bolzano.

- ✓ il questionario del 15° Censimento generale della popolazione;
- ✓ la lettera informativa sul censimento a firma del Presidente dell'Istat;
- ✓ la guida alla compilazione;
- ✓ la busta da utilizzare per la restituzione del questionario;
- ✓ il modello di Rilevazione sulla consistenza e la dislocazione territoriale degli appartenenti alle popolazioni di lingua ladina, mòchena e cimbra in lingua italiana completo di lettera informativa alle famiglie.

La spedizione del plico alle famiglie è stata effettuata a cura dell'Istat sulla base delle Liste anagrafiche comunali (Lac) riferite al 31 dicembre 2010.

La restituzione/compilazione del questionario ha seguito i canali prestabiliti (web, Ccr, Uffici postali, rilevatore), mentre la restituzione del modello della rilevazione sulle minoranze linguistiche è stata effettuata tramite consegna al Centro comunale di raccolta competente per territorio o tramite consegna al rilevatore comunale.

Come previsto dall'intesa stipulata con il Servizio di statistica della Provincia autonoma di Trento, l'Istat ha provveduto a tutte le attività previste dal Piano generale di Censimento per la rilevazione censuaria (spedizione alle famiglie presenti in Lac di tutto il materiale in italiano: questionari di censimento, guide alla compilazione e lettere informative). Inoltre, ha stampato e spedito i modelli della rilevazione sulle minoranze linguistiche, ad eccezione degli 11 Comuni sopra elencati (nei quali ha spedito agli Ucc solo i materiali del censimento, mentre i modelli della rilevazione sulle minoranze linguistiche sono stati spediti a cura del Servizio di Statistica della Provincia di Trento).

L'Istat ha inoltre personalizzato il Sistema di gestione della rilevazione (Sgr), inserendo un modulo per la documentazione delle fasi di consegna e restituzione del modello di rilevazione di appartenenza linguistica precaricando la rispettiva lista delle unità eleggibili.

Il Servizio di statistica della Provincia di Trento si è occupato di specifiche attività nell'ambito delle operazioni censuarie e, in particolare, della traduzione e della stampa in ladino, mocheno e cimbro del questionario di censimento e della lettera informativa, predisposti in italiano dall'Istat. Nell'ambito della rilevazione sulle minoranze linguistiche, ha inoltre predisposto e stampato i modelli di rilevazione e l'informativa in lingua italiana e in ladino, mocheno e cimbro, e inviato i materiali per la rilevazione agli Ucc degli undici Comuni sopra indicati. Infine, si è occupato dell'acquisizione, validazione e diffusione dei dati contenuti nelle "dichiarazioni di appartenenza" alle popolazioni di lingua ladina, mocheno e cimbro.

Il servizio di assistenza alle famiglie per il 15° Censimento generale della popolazione è stato garantito mediante il numero verde telefonico (*contact centre*) valido per tutto il territorio nazionale, mentre tutte le informazioni riguardanti la "Rilevazione sulla consistenza e la dislocazione territoriale degli appartenenti alle popolazioni di lingua ladina, mocheno e cimbra" sono state fornite dal numero verde attivato dalla Provincia autonoma di Trento.

Il monitoraggio sull'andamento delle operazioni censuarie, la segnalazione all'Istat dei casi critici e il monitoraggio delle operazioni relative alla "Rilevazione sulla consistenza e la dislocazione territoriale degli appartenenti alle popolazioni di lingua ladina, mocheno e cimbra" sono state effettuate dalla Commissione tecnica regionale, nominata dall'amministrazione provinciale.

4.5.3.2 La rilevazione censuaria nella Provincia autonoma di Bolzano

In linea con le principali innovazioni previste dal Piano generale di Censimento per la rilevazione censuaria su tutto il territorio nazionale, il censimento nella Provincia di Bolzano è stato effettuato a cura dell'Astat e degli Ucc, secondo specifiche modalità. Anche su questo territorio le rilevazioni sono state due:

- il 15° Censimento generale della popolazione e delle abitazioni, per il quale l'organo responsabile è l'Istat;
- la "Rilevazione di appartenenza linguistica o aggregazione a gruppo linguistico", per la quale l'organo responsabile è la Provincia autonoma di Bolzano. Ogni cittadino italiano non interdetto e residente nella Provincia di Bolzano, di 14 anni o più, è tenuto a rendere, in forma anonima, una dichiarazione individuale di appartenenza o aggregazione ad uno dei tre gruppi linguistici (italiano, tedesco e ladino). La dichiarazione è resa in busta chiusa di colore bianco, anonima e recante l'indicazione del Comune di residenza. Per i cittadini di età inferiore ai 14 anni, la dichiarazione è resa dai genitori o da coloro che

ne esercitano la patria potestà, in busta chiusa di colore rosa, anch'essa anonima e recante l'indicazione del Comune di residenza.

In tutti i Comuni della Provincia di Bolzano le famiglie hanno ricevuto tramite posta:

- ✓ la lettera informativa del 15° Censimento generale della popolazione a firma del Presidente dell'Istat e dell'Astat, contenente anche le credenziali per l'accesso al questionario elettronico;
- ✓ la lettera informativa sulla "Rilevazione di appartenenza linguistica o aggregazione a gruppo linguistico" a firma dell'Astat.

La spedizione delle due lettere alle famiglie è stata effettuata a cura dell'Astat sulla base delle Liste anagrafiche comunali (Lac) aggiornate al 10 settembre 2011. La lista delle unità eleggibili alla rilevazione di appartenenza linguistica, anch'essa riferita al 10 settembre, è stata inviata all'Istat contemporaneamente all'invio della Lac.

A partire dal 22 ottobre, i rilevatori si sono recati presso le famiglie e:

- ✓ hanno consegnato il modello "Dichiarazione di appartenenza o aggregazione a gruppo linguistico" per ciascun individuo presente nella lista delle unità eleggibili per il censimento linguistico;
- ✓ se la famiglia non aveva ancora compilato il questionario del censimento della popolazione hanno offerto supporto alla compilazione *online* tramite l'uso del proprio computer portatile. Il questionario cartaceo è stato consegnato solo nel caso di richiesta esplicita o nel caso di mancata copertura della rete internet. Nel caso in cui il questionario di censimento fosse stato compilato in forma cartacea, la sua restituzione è stata effettuata tramite consegna al rilevatore, oppure mediante consegna ad un Centro comunale di raccolta (Ccr) o presso le sedi dell'Astat.

Inoltre è stato compito dei rilevatori effettuare:

- ✓ la rilevazione degli edifici;
- ✓ il recupero delle mancate risposte presso le unità di rilevazione del censimento della popolazione non eleggibili alla rilevazione sull'appartenenza ai gruppi linguistici (cittadini stranieri abitualmente dimoranti nella Provincia di Bolzano);
- ✓ il recupero della sotto-copertura della Lac;
- ✓ la rilevazione delle abitazioni non occupate.

Poiché Il D.P.R. 574/1988 stabilisce che, sul territorio della Provincia di Bolzano, nei rapporti fra la pubblica amministrazione e i cittadini, ogni cittadino ha diritto ad usare la propria lingua madre, e in particolare l'art. 4 del citato D.P.R. prevede l'uso congiunto delle due lingue, italiana e tedesca, negli atti trasmessi alla generalità dei cittadini, il materiale della rilevazione e tutte le informazioni sono state predisposte nel rispetto del bilinguismo.

L'Istat ha personalizzato il Sistema di gestione della rilevazione (Sgr) inserendo un modulo per la documentazione delle fasi di consegna e restituzione della dichiarazione di appartenenza e caricando la rispettiva lista delle unità eleggibili. Ha sviluppato il questionario elettronico di censimento in italiano e tedesco e predisposto il manuale della rilevazione per gli Ucc e il manuale di Sgr in lingua italiana. Infine, ha fornito all'Astat l'applicazione per la generazione dei codici di questionario e delle *password* di accesso al questionario *online*.

L'Astat, invece, si è occupato della traduzione (in tedesco e ladino) e della stampa (in italiano e tedesco) dei questionari, delle guide, delle lettere informative (inclusa la lettera informativa dedicata alla rilevazione di appartenenza) e di tutto il restante materiale predisposto dall'Istat in italiano per la rilevazione censuaria, inclusa la stampa del manuale di rilevazione per gli Ucc in versione italiana e tedesca. Inoltre, ha predisposto la stampa delle etichette contenenti le informazioni dell'intestatario del foglio di famiglia e del responsabile di convivenza, da apporre sui questionari cartacei eventualmente utilizzati dai rilevatori. L'acquisizione dei dati registrati sui questionari cartacei è stata effettuata a cura dell'Astat mediante l'applicazione web (questionario elettronico) predisposta dall'Istat. Infine, la spedizione agli Ucc dei pacchi contenenti tutto il materiale per la rilevazione censuaria è stata effettuata dall'Astat.

Il servizio di assistenza alle famiglie tramite il numero verde telefonico (*contact centre*) è stato fornito dall'Astat sia per il censimento della popolazione che per la "Rilevazione di appartenenza linguistica o aggregazione a gruppo linguistico".

L'Astat ha svolto inoltre compiti specifici per le operazioni attinenti la rilevazione sull'appartenenza linguistica. In particolare, si è occupata della formazione della rete di rilevazione a livello provinciale, e di predisporre, stampare e inviare agli Ucc i modelli della rilevazione di appartenenza linguistica e le specifiche

istruzioni. Infine, ha curato l'acquisizione e la validazione dei dati contenuti nelle Dichiarazioni di appartenenza linguistica.

Ai sensi dell'art. 18 del D.P.R. 752/1976, l'Istat ha provveduto ad indicare nelle proprie pubblicazioni ufficiali i dati relativi alla consistenza proporzionale nella Provincia dei tre gruppi linguistici (italiano, tedesco e ladino), risultanti dalle dichiarazioni di appartenenza linguistica.

Il monitoraggio sull'andamento delle operazioni censuarie, la segnalazione all'Istat dei casi critici e il monitoraggio delle operazioni relative alla rilevazione sull'appartenenza ai gruppi linguistici sono stati effettuati dalla Commissione tecnica regionale, nominata dall'amministrazione provinciale.

4.5.3.3 La rilevazione censuaria nella Regione Friuli-Venezia Giulia

Ai cittadini italiani appartenenti alla minoranza linguistica slovena è riconosciuto il diritto all'uso della lingua slovena nei territori delle Province di Trieste, Gorizia e Udine. A tal fine sono state previste specifiche modalità di conduzione del 15° Censimento generale della popolazione e delle abitazioni (Circolare n. 8 dell'8 settembre 2011, prot. n. 6858).

La modalità di esercizio dei diritti di cui all'Art. 8 della legge n. 38 del 23 febbraio 2001, si esprime diversamente a seconda dei territori:

- ✓ nel territorio del Comune di Cividale del Friuli e nelle zone centrali dei Comuni di Trieste e Gorizia, le modalità di rilevazione sono state uguali a quelle previste per tutti gli altri Comuni italiani, ma nei Centri comunali di raccolta (Ccr) erano presenti operatori di censimento bilingue. Erano inoltre disponibili questionari di censimento e altra documentazione in lingua slovena come *ausilio* alla compilazione del questionario in lingua italiana, per chi ne avesse fatto richiesta;
- ✓ nei Comuni o zone sub-comunali nei quali si applicano le misure di tutela della minoranza slovena (L. 38/2001, D.P.R. 12 settembre 2007), le modalità di rilevazione erano differenti da quelle stabilite. In particolare, i rilevatori, guidati dall'Ucc e con l'ausilio delle informazioni precaricate nel Sistema di gestione della rilevazione (Sgr), consegnavano alle famiglie i plichi contenenti tutta la documentazione di censimento (questionario, guida alla compilazione e lettera informativa) sia in lingua italiana che in lingua slovena, al fine di garantire ad ogni famiglia la possibilità di compilare il questionario di censimento nella lingua preferita.

In particolare, le famiglie presenti nella Lista anagrafica comunale al 31 dicembre 2010 ricevevano dai rilevatori:

- ✓ plichi in lingua italiana, personalizzati con nome e indirizzo dell'intestatario della scheda di famiglia e associati univocamente al codice del questionario prestampato; quindi la famiglia poteva compilare il questionario via web utilizzando la propria *password*;
- ✓ plichi in lingua slovena, se richiesti. In questo caso il rilevatore effettuava la personalizzazione del questionario, consegnando anche il plico in italiano per dare la possibilità alla famiglia di compilare via web il questionario (la *password* era stampata solo sul questionario in italiano). Il rilevatore per personalizzare il questionario sloveno apponeva sull'ultima pagina un'etichetta prestampata con il codice questionario corrispondente al codice stampato sul questionario in italiano associato alla famiglia.

Diversamente, le famiglie non presenti nella Lista anagrafica comunale al 31 dicembre 2010 ricevevano:

- ✓ i plichi in lingua italiana non personalizzati con nome e indirizzo dell'intestatario della scheda di famiglia (il codice questionario prestampato non era associato a priori alla famiglia). Al momento della consegna, il rilevatore doveva annotare il codice di questionario sul Diario di sezione in forma cartacea e registrarlo appena possibile in Sgr;
- ✓ i plichi in lingua slovena (se richiesti). Il rilevatore annotava il codice questionario sul Diario di sezione in forma cartacea e successivamente procedeva alla registrazione in Sgr.

La restituzione avveniva secondo le modalità previste per l'insieme dei Comuni italiani.

I questionari cartacei compilati dalle famiglie in lingua slovena venivano poi registrati dall'Ufficio comunale di censimento (Ucc) competente per territorio, tramite l'applicativo web (questionario elettronico)⁹⁴.

⁹⁴ Le particolari modalità di conduzione del censimento nei comuni e nelle zone sub-comunali di cui all'allegato 1 della Circolare sono state adottate al fine di corrispondere a quanto previsto dall'art. 8 della legge n. 38 del 23 febbraio 2001, e non per eseguire un censimento della minoranza di lingua slovena. Di conseguenza, l'Ucc non poteva comunicare o diffondere dati, né individuali né in forma aggregata, relativi ai questionari compilati in lingua slovena.

L'Istat ha inoltre predisposto il questionario elettronico in lingua slovena, che era sempre possibile compilare indipendentemente dalla zona di residenza.

Per la rilevazione delle convivenze non era prevista la traduzione dei materiali in lingua slovena, ma l'Ucc inviava presso la convivenza un rilevatore bilingue.

Per il Comune di Gorizia, è stato inoltre attivato un numero telefonico dedicato con operatori plurilingue che fornivano assistenza alla compilazione in italiano, sloveno e friulano.

4.5.4 La rilevazione nei Comuni del “cratere sismico”

Nell'ambito delle attività di progettazione del censimento 2011, a seguito dei gravi danni provocati dal terremoto che ha colpito il territorio abruzzese il 6 aprile 2009, l'Istat decise cautelativamente di non includere nessun Comune abruzzese tra i Comuni campione della rilevazione pilota del Censimento della popolazione e delle abitazioni svolta a ottobre 2009.

Inoltre, i 57 Comuni del “cratere sismico” (ovvero i Comuni maggiormente colpiti dal sisma)⁹⁵ furono esclusi all'atto della pubblicazione della proposta delle basi territoriali dei Comuni abruzzesi, in considerazione del fatto che le immagini del territorio disponibili erano precedenti all'evento sismico, e che molti Comuni erano in fase di riorganizzazione, impegnati nell'emergenza, e in alcuni casi privati, a causa del terremoto, di documentazione importante, quale quella cartacea del censimento del 2001, o di strumentazione informatica adeguata (Cassata e Lipizzi, 2012).

Per gli stessi motivi, il Comune dell'Aquila fu esonerato dall'esecuzione della Rilevazione dei numeri civici (Rnc), che ha riguardato, tra il novembre 2010 e il marzo del 2011, i Comuni italiani con popolazione superiore ai 20 mila abitanti (cfr. paragrafo 1.6) (Istat, 2012b).

Infine, nell'ambito delle attività di preparazione del censimento nei Comuni del “cratere sismico”, date le difficoltà derivanti dallo stato di inagibilità di numerosi edifici privati e dalla conseguente mobilità forzata della popolazione sul territorio (in parte rilocata in campi, alberghi, case in affitto o presso amici/parenti, in parte migrata in altri Comuni o, in alcuni casi, in altre Regioni italiane o all'estero)⁹⁶, si ritenne necessario un approfondimento per predisporre procedure di rilevazione *ad hoc*.

A partire da settembre 2010, la Direzione centrale dei Censimenti generali (DCCG) dell'Istat ha quindi attivato contatti diretti con i Comuni del “cratere sismico” e con le altre istituzioni locali, anche attraverso l'Ufficio territoriale dell'Abruzzo con sede a Pescara. In particolare, venne avviato un confronto con il Comune dell'Aquila e con il suo ufficio tecnico, volto a trovare soluzioni in vista della rilevazione censuaria dell'ottobre 2011, cercando di conciliare per quanto possibile le novità di processo del 15° Censimento della popolazione e delle abitazioni con la difficile situazione del territorio a causa del sisma.

Sulla base delle proposte elaborate nell'ambito del “tavolo tecnico” appositamente costituito, che accoglievano parzialmente le esigenze sollevate dalle istituzioni locali, la Circolare Istat numero 7 del 20 luglio del 2011 ha stabilito specifiche modalità di conduzione del Censimento nei Comuni del “cratere sismico” abruzzese, prevedendo inoltre incentivi finanziari e prolungando i tempi per il completamento delle operazioni di censimento, in considerazione della particolare difficoltà delle operazioni di rilevazione sul campo.

Per quanto riguarda le modalità specifiche di realizzazione del censimento, la Circolare Istat n. 7 (Istat, 2017a) prevedeva, per i 57 Comuni del “cratere sismico” abruzzese, la possibilità di censire come dimoranti abitualmente nel proprio Comune anche le persone che, alla data del censimento, dimoravano in un altro Comune italiano perché costrette dall'emergenza legata al terremoto⁹⁷.

⁹⁵ I comuni facenti parte del cosiddetto “cratere sismico” erano 57, appartenenti alle province dell'Aquila (42 comuni), di Teramo (8 comuni) e di Pescara (7 comuni). Cfr. Decreti del Commissario delegato per la Ricostruzione (ai sensi del Decreto del Presidente del Consiglio dei Ministri del 6 aprile 2009) n. 3 del 16 aprile 2009 e n. 11 del 17 luglio 2009.

(<https://leg16.camera.it/561?appro=61#:~:text=Con%20il%20decreto%20n.,Adriano%20e%20Penna%20Sant'Andrea>).

⁹⁶ Nelle prime settimane gli sfollati ammontavano ad oltre 60 mila persone. A gennaio 2010 risultavano alloggiate in alberghi 6.461 persone coinvolte nel sisma dell'aprile 2009 (di queste, meno della metà dimoravano nel territorio della provincia dell'Aquila) mentre 1.191 persone erano alloggiate presso le strutture della Guardia di Finanza e la caserma Campomizzi. Alla stessa data, 12.059 residenti dell'Aquila risultavano dimoranti negli alloggi del Progetto C.A.S.E. (Complessi Antisismici Sostenibili ed Ecocompatibili) e 2.336 in Moduli Abitativi Provvisori (MAP). Cfr. Protezione Civile, Rapporto sintetico dell'attività della DICOMAC (Direzione di comando e controllo - organo di coordinamento nazionale delle strutture di Protezione civile nelle aree colpite da un evento disastroso; viene attivato dal Dipartimento della protezione civile in seguito alla dichiarazione dello “stato di emergenza”) per l'avvio della fase di affiancamento alla struttura del Commissario Delegato – presidente della Regione Abruzzo, 29 gennaio 2010.

⁹⁷ Se i comuni avessero cancellato dalla propria anagrafe le famiglie non più dimoranti abitualmente sul territorio perché sfollate a causa del terremoto, queste avrebbero perso il diritto al contributo per l'autonoma sistemazione perché sprovviste del requisito di iscrizione anagrafica nel comune erogatore dei servizi. Si tratta dell'Ordinanza del Presidente del Consiglio dei Ministri del 9 aprile 2009 (articolo 11) e successive modifiche e

Per quanto riguarda la consegna dei questionari, si è scelto di effettuarla tramite spedizione postale, come per il resto dei Comuni italiani, pur sapendo che un elevato numero degli stessi non sarebbe stato recapitato, ma sulla base del lavoro preliminare effettuato con i Comuni e con le altre istituzioni, i residenti nei 57 Comuni erano stati suddivisi in tre categorie:

- a) gli individui la cui dimora abituale dopo il terremoto (e alla data del censimento) coincideva con il luogo di residenza, evidentemente agibile;
- b) gli individui temporaneamente dimoranti, in sistemazioni varie, a un indirizzo diverso da quello di iscrizione anagrafica ma all'interno dello stesso Comune;
- c) gli individui costretti a una dimora temporanea fuori dal Comune di iscrizione anagrafica.

All'interno del Sistema di gestione e monitoraggio della rilevazione (Sgr) i residenti della categoria *a*) erano collocati all'indirizzo (e sezione di censimento) di iscrizione anagrafica, i residenti della categoria *b*) erano collocati (e quindi censiti) nelle sezioni del Comune corrispondenti all'indirizzo "alternativo" e i residenti della categoria *c*) erano collocati in sezioni fittizie. Questi ultimi, pur essendo rilevati in un Comune diverso da quello di residenza, venivano poi attribuiti al Comune di iscrizione anagrafica. Infatti, come detto, in considerazione dell'evento eccezionale, era possibile, in deroga al regolamento anagrafico, essere considerati dimoranti abitualmente nel Comune di iscrizione anagrafica, indipendentemente dal luogo effettivo di dimora abituale alla data del censimento.

Infine, sono state predisposte ulteriori specifiche modalità di rilevazione per il solo Comune dell'Aquila. Innanzitutto, come già accennato, il Comune, pur rientrando nel campo di osservazione della Rilevazione dei numeri civici (propedeutica al Censimento 2011 per tutti i Comuni con almeno 20 mila abitanti), era stato esonerato dalla Rnc condotta nel 2010, e quindi ha effettuato il censimento degli edifici durante il censimento del 2011 come i Comuni più piccoli. Questa decisione era stata presa sia per ridurre il carico di lavoro dell'Ufficio di censimento sia per disporre di dati il più possibile aggiornati, data l'estrema variabilità nel tempo della situazione sul territorio.

Inoltre, data la difficoltà di estrarre un campione effettivamente rappresentativo della situazione territoriale, il Comune dell'Aquila è stato escluso dalla strategia campionaria e, di conseguenza, a tutte le famiglie del Comune è stato assegnato il questionario in forma completa (*long form*), diversamente dagli altri Comuni con almeno 20 mila abitanti sottoposti a campionamento tramite *short/long form*.

Allo scopo di ridurre le difficoltà di recapito, dovute alla mutata situazione sul territorio rispetto a quella fotografata dalle Liste anagrafiche comunali, è stata utilizzata una lista anagrafica "arricchita", fornita all'Istat dal Comune dell'Aquila, comprendente gli indirizzi di dimora temporanea delle famiglie che, in seguito al terremoto, erano state costrette a cambiare alloggio. La lista di indirizzi, estratta dalla *Banca dati dell'emergenza*, è stata elaborata dai tecnici del Comune dell'Aquila, grazie all'incrocio dei dati del censimento sull'agibilità degli edifici effettuato dalla Protezione civile⁹⁸, con quelli di vari archivi amministrativi legati all'emergenza⁹⁹. È stata quindi questa lista di indirizzi a costituire la base per la

integrazioni: "Art. 11. **1.** Il Commissario delegato, anche avvalendosi dei Sindaci, è autorizzato ad assegnare ai nuclei familiari la cui abitazione principale, abituale e continuativa sita nei comuni di cui all'articolo 1 sia stata distrutta in tutto o in parte, ovvero sia stata sgomberata in esecuzione di provvedimenti delle competenti autorità, adottati a seguito degli eccezionali eventi sismici di cui in premessa, un contributo per l'autonoma sistemazione fino ad un massimo di 400,00 euro mensili, e, comunque, nel limite di 100,00 euro per ogni componente del nucleo familiare abitualmente e stabilmente residente nell'abitazione; ove si tratti di un nucleo familiare composto da una sola unità, il contributo medesimo è stabilito in 200,00 euro. Qualora nel nucleo familiare siano presenti persone di età superiore a 65 anni, portatori di handicap, ovvero disabili con una percentuale di invalidità non inferiore al 67 per cento, è concesso un contributo aggiuntivo di 100,00 euro mensili per ognuno dei soggetti sopra indicati. **2.** Il Commissario delegato, anche avvalendosi dei Sindaci, è autorizzato, laddove non è possibile l'autonoma sistemazione dei nuclei familiari, a disporre per il reperimento di una sistemazione alloggiativa alternativa. **3.** I benefici economici di cui al comma 1 sono concessi a decorrere dalla data di sgombero dell'immobile e fino al 31 dicembre 2009, salvo che non si siano realizzate le condizioni per il rientro nell'abitazione ovvero si sia provveduto ad altra sistemazione avente carattere di stabilità".

⁹⁸ Il Dipartimento della protezione civile, incaricato della gestione del post-terremoto in base all'Ordinanza 3753/2009, ha condotto attività di sopralluogo in provincia dell'Aquila nelle settimane successive al 6 aprile 2009, allo scopo di verificare lo stato di agibilità degli edifici. Al 25 luglio 2009 (cfr. *Protezione civile, Rapporto del 26 luglio 2009*) risultavano compilate 64.187 schede di edifici (di cui 59.606 relative a edifici ad uso privato e le restanti relative ad edifici per utilizzo pubblico o commerciale). Gli edifici sono stati classificati in 6 categorie, in base al grado dei danni subiti. Circa il 50 per cento degli edifici risultava di categoria A (agibile), la restante metà era suddivisa in edifici "temporaneamente inagibili, ma agibili con provvedimenti di pronto intervento" (tipo B), "parzialmente inagibili" (tipo C), "temporaneamente inagibili, da approfondire" (tipo D), "inagibili" (tipo E) e "inagibili per rischio esterno" (tipo F). I dati della classificazione di agibilità sono stati trasmessi ai comuni del cratere, che li hanno utilizzati per l'erogazione di finanziamenti e altri interventi mirati.

⁹⁹ Questi contengono dati su: intestatari di richieste di agibilità parziale e relativi familiari; persone ospitate in albergo; intestatari di pratiche B o C e familiari; richiedenti del Contributo di autonoma sistemazione; assegnatari di appartamento facente parte del fondo immobiliare; assegnatari del Progetto C.A.S.E. (Complessi antisismici sostenibili ed ecocompatibili) e dei Moduli abitativi provvisori; assegnatari di appartamenti nel villaggio di Onna; intestatari di pratiche per costruzione di manufatti temporanei.

spedizione dei questionari di censimento per i residenti nel Comune dell'Aquila. In particolare, nei casi di intestatari di pratiche per edifici “agibili”, “temporaneamente inagibili, ma agibili con provvedimenti di pronto intervento” o “parzialmente inagibili”¹⁰⁰, il questionario è stato inviato all'indirizzo di residenza anagrafica, mentre in tutti gli altri casi il modello è stato spedito all'indirizzo alternativo presente nella Banca dati¹⁰¹ e caricato in Sgr come informazione di ausilio al monitoraggio della rilevazione per l'Ucc.

¹⁰⁰ Ovvero intestatari di pratiche A, B o C secondo la classificazione degli edifici del censimento della Protezione civile (cfr. Istat, 2017a).

¹⁰¹ Nel primo indirizzario inviato all'Istat a fine ottobre 2010 erano presenti informazioni sull'indirizzo o sul recapito telefonico per il 90 per cento delle famiglie. Il *linkage* con altre fonti effettuato nel 2011 ha ulteriormente migliorato la qualità e la quantità dei dati presenti; nella lista finale, solo per il 4,5 per cento degli individui non erano presenti informazioni sull'eventuale indirizzo alternativo di spedizione.

CAPITOLO 5

I PESI DI RIPORTO ALL'UNIVERSO PER LA STIMA DELLE FREQUENZE¹⁰²

5.1 Introduzione

Il censimento, oltre a determinare la popolazione legale e la sua struttura demografica, ha l'obiettivo di produrre le frequenze assolute per celle di incrocio (semplici o complesse) definite dalle modalità di classificazione delle variabili rilevate, per i differenti universi oggetto di rilevazione: i dati socio-economici degli individui; le tipologie familiari; le caratteristiche strutturali delle abitazioni; le informazioni sugli edifici.

La decisione di adottare una strategia campionaria basata su *short/long form* al 15° Censimento generale della popolazione e delle abitazioni ha necessariamente richiesto, per una parte dei dati previsti dal Piano di diffusione italiano e dal Piano di diffusione europeo, il passaggio dalla semplice operazione di conteggio ad una, più complessa, di stima.

La struttura dei due tipi di questionario impiegati per la raccolta dei dati, uno in forma ridotta (*short*) e uno in forma estesa (*long*), porta a distinguere come incroci di “*tipo short*” quelli riferibili alle variabili in comune ai due questionari, da quelli di “*tipo long*” che considerano almeno una variabile rilevata a campione con il modello completo. Per i primi si procede al conteggio esaustivo delle unità, per i secondi, unicamente nei domini soggetti a rilevazione campionaria (cfr. paragrafo 3.3), si deve seguire una procedura di stima.

La produzione delle stime del censimento del 2011 ha previsto l'impiego dello stimatore di ponderazione vincolata (o calibrazione). Pur in presenza di complicazioni di natura computazionale, la scelta metodologica ha offerto garanzie in termini di accuratezza delle stime finali e di coerenza dei valori stimati rispetto ai dati derivanti da conteggio, e la certezza di poter completare l'intera operazione nel rispetto dei tempi prefissati per il rilascio dei dati.

Dopo una breve illustrazione di alcuni aspetti metodologici dello stimatore di ponderazione vincolata (paragrafo 5.2), vengono presentate le analisi e i test sperimentali, condotti su dati del censimento del 2001, per differenti approcci di calcolo dei pesi calibrati (paragrafo 5.3). Per la strategia di calcolo scelta, viene illustrato il sistema di vincoli implementato nella procedura e i test condotti al fine di rendere praticabile l'approccio prescelto, nelle modalità e nei tempi dovuti (paragrafo 5.4), per ciascuno degli universi censiti: popolazione, famiglie, alloggi ed edifici. Successivamente (paragrafo 5.5) viene descritta la fase operativa di calcolo dei pesi finali di riporto all'universo. Infine (paragrafo 5.6), il capitolo si chiude con delle considerazioni di sintesi sul processo di riponderazione impiegato, con l'evidenza di criticità e vantaggi.

5.2 Lo stimatore di ponderazione vincolata

5.2.1 Aspetti metodologici

L'impiego di una strategia campionaria in una rilevazione statistica comporta l'adozione di particolari tecniche per la stima dei parametri nella popolazione di interesse.

Dato un campione probabilistico, ad ogni unità è associata una “probabilità di inclusione” calcolabile in base al disegno di campionamento adottato per la selezione del campione. Ogni stimatore fa uso del peso associato; nel caso dello stimatore più semplice (stimatore di espansione o di Horvitz-Thompson) il peso corrisponde all'inverso della probabilità di inclusione dell'unità.

Lo stimatore di Horvitz-Thompson gode della proprietà della correttezza ma può non essere il più efficiente, quello cioè che comporta il minimo possibile della variabilità campionaria. L'obiettivo di produrre stime maggiormente accurate spinge verso l'impiego di stimatori alternativi che sfruttano informazione ausiliaria, disponibile per l'intera popolazione e correlata a quella oggetto di interesse per le stime. Un esempio, in tal senso, è rappresentato dagli stimatori di regressione generalizzata che sfruttano informazione ausiliaria,

¹⁰² Il Capitolo 5 è a cura di Giancarlo Carbonetti. I paragrafi 5.1, 5.6 e i paragrafi 5.2.1, 5.2.2, 5.2.3, 5.5.6 sono stati redatti da Giancarlo Carbonetti; il paragrafo 5.4 e i paragrafi 5.2.4, 5.3.1 sono stati redatti da Sara Giavante; i paragrafi 5.3.2, 5.5.4 sono stati redatti da Luana De Felici; i paragrafi 5.5.1, 5.5.2, 5.5.3 sono stati redatti da Epifania Fiorello; il paragrafo 5.5.5 è stato redatto da Ferdinando Ruggiero.

tramite un modello che lega la variabile oggetto di stima (variabile dipendente) a una o più variabili ausiliarie (variabili indipendenti).

Lo stimatore di regressione generalizzata implica un processo di “riponderazione” delle unità campionarie: il peso iniziale, l’inverso della probabilità di inclusione (peso base o da disegno) viene moltiplicato per un fattore di correzione determinato tramite una procedura iterativa, tale da assicurare da una parte la minimizzazione della distanza euclidea tra peso iniziale e peso finale, e dall’altra la coincidenza tra i totali noti nella popolazione e i totali stessi ottenuti mediante la stima calcolata tramite i valori campionari. Quest’ultimo risultato è estremamente importante ai fini della coerenza delle stime (specialmente se provenienti da indagini diverse).

5.2.2 La riponderazione

Si consideri una popolazione finita P costituita da N unità che vengono indicate con $j = \{1, \dots, k, \dots, N\}$ e sia s un campione casuale di n elementi ($n < N$) che si possono indicare con $s = \{1, \dots, k, \dots, n\}$, estratto da P secondo un disegno di campionamento che assegna al generico campione s la probabilità $p(s)$ di essere estratto.

Per la generica unità j di P , si indichi:

π_j , la probabilità di inclusione di j nel campione s ;

$x_j = (x_{1j}, \dots, x_{mj}, \dots, x_{Mj})$ il valore assunto dal vettore $X = (X_1, \dots, X_m, \dots, X_M)$ di M variabili ausiliarie.

Si supponga di voler stimare, sulla popolazione P il totale T_y della variabile y dato dalla seguente espressione:

$$T_y = \sum_{j \in P} y_j \quad (5.1)$$

dove sono disponibili le seguenti informazioni:

- ✓ per ciascuna unità del campione s le $M+1$ osservazioni $(y_j, x_{1j}, \dots, x_{mj}, \dots, x_{Mj})$;
- ✓ i valori dei totali delle M variabili ausiliarie su tutta la popolazione P .

In base alla sopraindicata notazione, uno stimatore del totale T_y appartenente alla classe degli *stimatori di ponderazione vincolata* può essere espresso nel seguente modo:

$$\hat{T}_y = \sum_{j \in s} d_j \delta_j y_j = \sum_{j \in s} w_j y_j \quad (5.2)$$

in cui:

$d_j = 1/\pi_j$ rappresenta il peso diretto (o da disegno) associato alla j -ma unità campionaria di s ;

$w_j = d_j \delta_j$ indica il peso finale di riporto all’universo assegnato a tale unità;

δ_j indica il correttore del peso diretto.

L’espressione (5.2) è simile a quella dello stimatore di Horvitz-Thompson (Horvitz e Thompson, 1952) che, come noto, è espresso come somma pesata (con i pesi diretti) dei dati campionari. Si fa però presente che nello stimatore di Horvitz-Thompson i pesi diretti dipendono unicamente dalle unità estratte nel campione, mentre nella (5.2) i pesi finali dipendono, per costruzione, dai totali noti delle variabili ausiliarie, dai valori assunti dalle variabili ausiliarie nel campione rilevato e dalla variabilità della variabile oggetto di interesse.

La determinazione dei pesi finali w_j è ottenuta risolvendo un problema di minimo vincolato in cui la funzione obiettivo è minimizzare la somma delle “distanze” tra i pesi diretti e i pesi finali:

$$\min\{\sum_{j \in s} \text{dist}(d_j; w_j)\} \quad (5.3)$$

e, come vincoli, che i pesi finali siano tali da produrre, con i dati del campione s , stime per i totali delle variabili ausiliarie pari ai valori noti sulla popolazione P :

$$\sum_{j \in s} w_j x_{mj} = T_{x_m}, \quad \forall m \quad (5.4).$$

L'obiettivo è, quindi, quello di individuare un insieme di pesi finali w_j che consenta di rispettare il sistema di vincoli (5.4) e che contemporaneamente modifichi il meno possibile, sulla base della funzione di distanza scelta ed eventualmente su limiti di variazione predefiniti per i pesi finali, l'insieme dei pesi diretti d_j .

Lo stimatore di regressione generalizzata appartiene alla classe degli "stimatori di calibrazione". Ogni possibile stimatore di calibrazione è individuato dalla specifica funzione di distanza tra peso iniziale e peso finale utilizzata nel processo di riponderazione: le distanze euclidea, logaritmica, logaritmica troncata, lineare troncata, eccetera definiscono altrettanti stimatori di calibrazione, la cui scelta è legata alle esigenze dell'indagine e alle proprietà stesse degli stimatori.

Per il calcolo dei pesi di calibrazione è stato impiegato il software *Genesees v3.0*, sviluppato in Istat (Pagliuca, 2005) proprio con l'obiettivo di calcolare i pesi tenendo conto di totali noti della popolazione oggetto di studio (informazione ausiliaria) e garantire la coincidenza tra questi e le corrispondenti stime campionarie.

5.2.3 Alcune proprietà dello stimatore di ponderazione vincolata

Lo stimatore di ponderazione vincolata è asintoticamente corretto; quindi, per campioni sufficientemente grandi lo stimatore produce stime approssimativamente corrette. Inoltre, in moltissimi casi l'impiego dello stimatore calibrato porta ad una riduzione della variabilità campionaria rispetto allo stimatore diretto (basato sui "pesi base").

La proprietà più importante riguarda la *coerenza* delle stime prodotte. L'uso dello stimatore di ponderazione vincolata garantisce, per costruzione, la coerenza dei totali noti riferiti alle variabili ausiliarie con quelli ottenibili come conteggio su tutta la popolazione di interesse.

Nel contesto censuario, i totali noti considerati si riferiscono ad alcuni degli incroci di *tipo short* osservati in modo esaustivo su tutta la popolazione e riferibili a modalità delle variabili contenute nei modelli *short* e *long*. Una volta definito l'insieme dei totali noti, i pesi calibrati garantiscono la coerenza dei valori delle frequenze assolute riferite alle variabili *short* (per gli incroci che includono almeno una variabile campionata) con quelli ottenibili come conteggio su tutta la popolazione (negli incroci relativi solo a variabili *short*).

5.2.4 Criteri per il disegno di un sistema di vincoli in ambito censuario

La strategia campionaria impiegata per il 15° Censimento generale della popolazione e delle abitazioni ha previsto l'adozione di due modelli di rilevazione: uno in forma breve (*short*) per la raccolta dei dati di natura demografica e delle principali variabili socio-economiche; uno in forma completa (*long*) per la raccolta di tutte le variabili tradizionalmente osservate al censimento. L'uso di due diversi questionari ha implicato la necessità di definire un opportuno sistema di vincoli sul quale impostare il calcolo dei pesi di riporto all'universo.

Con riferimento al Piano di diffusione italiano dei risultati del censimento del 2011, le tavole multidimensionali (cubi di diffusione) disegnate si riferiscono ai seguenti oggetti:

- ✓ popolazione;
- ✓ famiglie;
- ✓ alloggi ed edifici.

Il complesso delle tavole di diffusione sono state raggruppate in differenti aree tematiche:

- 1) numero di famiglie e di convivenze, numero di componenti;
- 2) variabili anagrafiche, presenza e dimora, popolazione in famiglia e popolazione in convivenza, tipo di famiglia, tipo di nucleo e tipo di convivenza;
- 3) alloggi ed edifici e famiglie;
- 4) istruzione e formazione;
- 5) condizione professionale e lavoro;
- 6) pendolarismo;
- 7) *disability*.

Per ogni oggetto e tematica il Piano di diffusione del 2011 stabilisce il dettaglio territoriale di riferimento per la pubblicazione dei dati. I livelli territoriali coinvolti sono: nazione, ripartizione geografica, regione, provincia/grande comune, comune, località abitata.

Il disegno del sistema dei vincoli prevede, in una prima fase, che per ogni oggetto, per ogni livello territoriale e per ogni tavola, le variabili siano classificate in "esaustive", se sono state rilevate sia tramite il modello in forma breve (*short*) che tramite il modello in forma completa (*long*), e "a stima", se sono state

rilevate (nei domini campionati¹⁰³) esclusivamente mediante il modello *long*. Esistono poi casi particolari di aggregati che sono oggetto di stima in quanto derivanti da variabili osservate con il modello *long* (esempio: l'aggregato "Occupati dipendenti" della tematica "condizione professionale e lavoro").

Individuate e classificate le variabili e gli aggregati, si costituiscono due gruppi di tavole: il primo contiene quelle sulle quali la strategia di campionamento non ha impatto, ovvero le tavole che riportano solo dati delle variabili esaustive; il secondo gruppo contiene invece le tavole cosiddette "a stima" perché incrociano almeno una variabile, o una modalità, o un aggregato oggetto di stima.

Il disegno del sistema di vincoli per il calcolo dei pesi finali si riferisce proprio al secondo gruppo di tavole. Infatti, per popolare le celle di queste tavole bisogna tenere conto del contenuto degli incroci tra le modalità delle variabili coinvolte. Le frequenze assolute degli incroci delle variabili esaustive, prodotte tramite conteggio, rappresentano i vincoli che devono essere soddisfatti dal processo di riponderazione; così facendo, i pesi finali permetteranno di popolare le restanti celle e produrre stime che siano coerenti, nella stessa tavola, con quanto già prodotto tramite conteggio.

Si consideri a titolo esemplificativo la tavola "Occupati per sesso (3), Attività lavorativa svolta (11) e Classe di età da 15 anni in poi (14)", in cui viene indicato in parentesi, accanto ad ogni variabile di incrocio, il numero di modalità rilevate per quella variabile, incluso di totali e sub-totali. Il primo passo è l'individuazione del numero "netto" di modalità di classificazione riferite ad ogni variabile, in modo tale che nessuna modalità possa essere ricavata da combinazioni delle restanti modalità. Nel caso in esame, poiché per ogni variabile è stato incluso anche il valore totale (marginale), la tavola diventa: "Occupati per sesso (2), Attività lavorativa svolta (10) e Classe di età da 15 anni in poi (13)", in cui ci sono $2 \times 13 \times 10 = 260$ possibili incroci.

I valori presenti nelle celle derivanti dalla combinazione delle modalità delle variabili "Sesso" e "Classe di età da 15 anni in poi", essendo esaustive, devono essere riprodotti con le stime delle frequenze relative agli incroci con le modalità della variabile "Attività lavorativa svolta". Il numero totale di vincoli da soddisfare (il numero di celle coinvolte) è ottenuto moltiplicando tra loro il numero delle modalità di classificazione delle variabili esaustive presenti nella tavola; per la tavola in esame il numero di vincoli è pari a 26, cioè le due modalità della variabile "Sesso" moltiplicate per le 13 modalità della variabile "Classe di età da 15 anni in poi". I valori delle 26 "celle vincolo" (dette anche *celle di calibrazione*) devono essere garantiti per popolare in modo coerente le celle relative all'incrocio delle 10 modalità della variabile "Attività lavorativa svolta" con le modalità delle variabili "Sesso" e "Classe di età da 15 anni in poi" (in totale 260 incroci).

Riassumendo, per la tavola "Occupati per Sesso (2), Attività lavorativa svolta (10) e Classe di età da 15 anni in poi (13)", l'insieme dei vincoli da considerare per la riponderazione è dato dai 26 valori delle celle (*di calibrazione*) riferite agli "Occupati" incrociati per il "Sesso" e per la "Classe di età" (*distribuzioni marginali di riponderazione*).

Reiterando tale procedimento per ciascuna delle tavole del Piano di diffusione si ottengono le celle di calibrazione e le relative frequenze assolute che devono essere riprodotte in fase di stima per la definizione del sistema di vincoli da implementare nel procedimento di calcolo dei pesi finali. La determinazione dei totali noti indotti da ciascuna tavola conduce ad un primo sistema di vincoli complessivo; tale sistema si dice *non ottimizzato* in quanto presenta vincoli ridondanti anche su diversa scala geografica. Il sistema non ottimizzato viene sottoposto ad un processo di ottimizzazione "orizzontale" e "verticale" per poter definire l'*insieme minimo di vincoli ottimizzato*.

L'ottimizzazione "orizzontale" consiste nell'eliminazione di ridondanze di vincoli indotti da più tavole con riferimento allo stesso livello territoriale. Operando ad un dato livello territoriale, una volta individuate tutte le tavole con oggetto e variabili esaustive in comune, si ritengono garantite dalla tavola più complessa (in termini di modalità e/o variabili) tutte le tavole in essa contenute, a prescindere da quali siano le variabili oggetto di stima.

Ad esempio, la tavola provinciale relativa a "Popolazione residente in famiglia in abitazione per Disponibilità di servizi (11) e Numero di occupanti (7)" è garantita dalla tavola provinciale riferita alla "Popolazione residente in famiglia in abitazione per Numero di occupanti (7), Numero di stanze (7) e Titolo

¹⁰³ Occorre distinguere tra "domini campionati" e "domini non campionati". I domini campionati sono quelli in cui la raccolta dei dati è avvenuta tramite *short/long form*: se ciò ha riguardato l'intero dominio allora si dice "interamente campionato" (generalmente lo sono i comuni con più di 20 mila abitanti coinvolti dal campionamento); se l'operazione campionaria non ha interessato l'intero territorio del dominio allora questo si dice "parzialmente campionato" (lo sono tutte le province e le regioni in quanto hanno al loro interno comuni oltre 20 mila abitanti campionati e comuni di minori dimensioni non campionati).

di godimento (4)”, in cui l’insieme dei vincoli in comune è dato dalle celle relative alle modalità della variabile esaustiva “Popolazione residente in famiglia in abitazione per Numero di occupanti (7)”.

Come descritto in precedenza, i livelli territoriali sono articolati in maniera gerarchica a partire dal livello minimo delle località abitate fino al livello di ripartizione geografica e di nazione. Per realizzare il processo di ottimizzazione “verticale”, una volta individuate tutte le tavole con oggetto e variabili esaustive in comune, ma appartenenti a livelli territoriali diversi, si stabilisce che l’insieme dei vincoli individuato per il livello territoriale inferiore valga anche per tutti i livelli superiori.

Per esempio, la tavola a livello di ripartizione geografica “Occupati per sesso (3), Attività lavorativa svolta (11) e Posizione nella professione (8)” è garantita dalla tavola provinciale “Occupati per sesso (3), Grado di istruzione (8) e Tipologia attività lavorativa (pieno-parziale) (3)” perché i vincoli richiesti, cioè i valori della distribuzione degli “Occupati per sesso”, sono contenuti nei vincoli definiti a livello provinciale dalla variabile “Sesso” in comune alle due tavole. Infatti, ipotizzando di effettuare la riponderazione su domini di calibrazione che considerano l’intera ripartizione geografica, i vincoli provinciali garantiscono i vincoli in comune riferiti alla ripartizione (per somma dei valori delle Province appartenenti alla stessa ripartizione).

5.3 Analisi preliminari per la definizione della strategia di calcolo dei pesi calibrati

5.3.1 Possibili approcci al disegno del sistema di vincoli

Il Piano di diffusione dei risultati del 14° Censimento generale della popolazione e delle abitazioni del 2001 ha rappresentato, durante la fase di studio sperimentale, la base di riferimento per la costruzione del sistema di vincoli che avrebbe successivamente interessato il Piano di diffusione relativo ai risultati del 15° Censimento generale della popolazione e delle abitazioni del 2011.

Per arrivare al sistema di vincoli definitivo bisognava non solo attendere la versione finale del Piano di diffusione 2011, ma anche tenere conto delle problematiche computazionali, sia in termini di software che di hardware, nonché considerare i risultati dei diversi test condotti sulla procedura di riponderazione.

Il processo ha richiesto la collaborazione di diverse professionalità: metodologiche, per le soluzioni statistiche e matematiche impiegabili nella tecnica di riponderazione; tematiche, per intervenire sul disegno di alcune tavole più complesse e semplificare il sistema di vincoli; informatiche, per il dimensionamento delle risorse software e hardware necessarie.

In questo ambito si illustrano le fasi di sperimentazione e le ipotesi sottostanti i tre principali approcci seguiti, per ciascuno dei quali si è proceduto a determinare: i vincoli richiesti per soddisfare la coerenza dei dati; la praticabilità in termini di processo; l’impatto delle procedure informatiche; la disponibilità di opportuna strumentazione hardware; i tempi di elaborazione dei dati per il rilascio dei risultati.

Nei paragrafi seguenti si illustrano le sperimentazioni e le considerazioni che ne sono seguite, che hanno portato a scartare alcune delle possibilità convergendo sull’opzione di compromesso più promettente dal punto di vista del rapporto tra costi e benefici.

5.3.1.1 Approccio A - Sistema di pesi unico

Inizialmente, con riferimento alla strategia campionaria definita per il censimento del 2011, si era deciso di considerare come massimo dettaglio territoriale per la diffusione dei dati, il dominio sub-comunale dell’area di censimento di centro abitato (Ace)¹⁰⁴. Le Ace sono state disegnate per i Comuni oltre i 20 mila abitanti e si configurano come domini pianificati per la selezione di campioni rappresentativi di famiglie, e come domini di output (cfr. paragrafo 3.3.2).

Il primo tentativo di disegno del sistema di vincoli ha previsto la creazione di due sistemi che potessero garantire tutte le tavole del Piano di diffusione del 2001: un sistema per l’oggetto “popolazione” e un sistema congiunto per gli oggetti “famiglie” e “alloggi ed edifici”, individuando come livelli territoriali l’area di censimento, la provincia/grande comune, la regione e la ripartizione (Prospetti 5.1 e 5.2). Il processo di ottimizzazione “orizzontale” e “verticale” ha generato un numero consistente di vincoli (Prospetto 5.3).

¹⁰⁴ Cfr. nota 56.

Prospetto 5.1 - Approccio A: sistema di vincoli complessivo ottimizzato (oggetto “popolazione”).

Descrizione	Numero di vincoli
AREA DI CENSIMENTO	
Occupati X Sesso X Classe di età 15 anni e più (4) X Attività economica (3)	8
PROVINCIA / GRANDE COMUNE	
Popolazione residente di 15 anni e più X Sesso X Età per singolo anno da 15 a 19 anni e classe di età 15 anni e più (15) X Iscrizione ad un corso regolare di studi (2)	30
Popolazione residente di 6 anni e più X Sesso X Classe di età 6 anni e più (15) X Grado di istruzione (9)	26
Popolazione residente fino a 14 anni X Sesso X Età per singolo anno da 0 a 14 anni (15)	30
Popolazione residente in famiglia in abitazione X Numero di occupanti (6) X Titolo di godimento (3) X Numero stanze (6)	18
Occupati X Sesso X Classe di età 15 anni e più (13) X Attività lavorativa svolta (10)	18
Occupati X Sesso X Stato civile (6) X Attività lavorativa svolta (10)	12
Occupati X Sesso X Grado di istruzione (7) X Tipologia attività lavorativa (pieno-parziale) (2)	14
Popolazione residente che si è recata il mercoledì precedente la data del censimento al luogo abituale di studio X Sesso X Luogo di destinazione (9) X Mezzo utilizzato (9) X Tempo impiegato in minuti (5)	18
Popolazione residente di 15 anni e più che si è recata il mercoledì precedente la data del censimento al luogo abituale di lavoro X Sesso X Luogo di destinazione (9) X Mezzo utilizzato (9) X Tempo impiegato in minuti (5)	18
Popolazione straniera residente nata all'estero X Sesso X Area geografica di cittadinanza (6) X Anno di trasferimento in Italia (5)	12
Popolazione straniera residente occupata X Sesso X Classe di età 15 anni e più (9) X Attività economica (3) X Posizione nella professione (2)	18
REGIONE	
Popolazione femminile residente in famiglia di 15 anni e più X Classe di età 15 anni e più (9) X Numero di figli (4) X Condizione e posizione nella professione (10)	36
Popolazione residente di 14 anni e più X Sesso X Classe di età 14 anni e più (12) X Grado di istruzione (7) X Iscrizione ad un corso regolare di studi (2)	168
Popolazione residente di 15 anni e più X Sesso X Condizione (8) X Iscrizione ad un corso regolare di studi (2)	16
Popolazione residente di 6 anni e più X Sesso X Classe di età 6 anni e più (15) X Titolo di studio (273)	130
Popolazione residente in famiglia in abitazione X Numero di famiglie (2) X Figura giuridica del proprietario (7)	2
Popolazione residente in famiglia in abitazione X Numero di componenti (6) X Tipo di servizi (12)	6
Popolazione residente in cerca di prima occupazione X Sesso X Classe di età 15 anni e più (6) X Titolo di studio (77)	12
Popolazione residente che si è recata il mercoledì precedente la data del censimento al luogo abituale di studio o di lavoro X Sesso X Classe di età (9) X Mezzo utilizzato (9)	18
Popolazione straniera residente nata all'estero X Sesso X Area geografica e paese di cittadinanza (96) X Anno di trasferimento in Italia (5)	180
RIPARTIZIONE	
Occupati X Sesso X Classe di età (9) X Grado di istruzione (6) X Settore di attività economica (6)	60
Occupati X Sesso X Titolo di studio (9) X Posizione nella professione (5) X Settore di attività economica (17)	18
Popolazione straniera residente occupata X Sesso X Area geografica e paese di cittadinanza (96) X Attività economica (4)	192

Fonte: Direzione centrale dei Censimenti generali (DCCG), Istat.

Prospetto 5.2 - Approccio A: sistema di vincoli complessivo ottimizzato (oggetto “famiglie” e “alloggi ed edifici”).

Descrizione	Numero di vincoli
AREA DI CENSIMENTO	
Numero abitazioni X Tipo di occupazione e servizi (7)	3
Stanze delle abitazioni in edifici ad uso abitativo X Epoca di costruzione (7)	7
Abitazione occupate da persone residenti X Disponibilità di servizi (11)	1
PROVINCIA / GRANDE COMUNE	
Famiglie in abitazione X Titolo di godimento (3) X Numero di stanze (6)	3
Abitazioni occupate da persone residenti X Classe di superficie (9) X Disponibilità e numero di cucine (9)	9
Abitazioni occupate da persone residenti in edifici ad uso abitativo X Epoca di costruzione (7) X Titolo di godimento (3) X Figura giuridica del proprietario (7)	21
Abitazioni occupate da persone residenti in edifici ad uso abitativo X Numero di abitazioni (6) X Servizi installati (6)	6
Abitazioni occupate da persone residenti X Titolo di godimento (3) X Numero di occupanti (6) X Numero di stanze (6)	18
REGIONE	
Famiglie in abitazione X Numero di famiglie (2) X Figura giuridica del proprietario (7)	2
Famiglie in abitazione X Numero di componenti (6) X Tipo di servizi (12)	6
Famiglie unipersonali non in coabitazione in abitazione X Classe di età decennale (8) X Numero di stanze (6)	8
Abitazioni occupate da persone residenti X Numero di famiglie (2) X Numero di stanze (6)	2

Fonte: Direzione centrale dei Censimenti generali (DCCG), Istat.

Prospetto 5.3 - Approccio A: numero delle tavole soggette a stima, numero delle tavole ottimizzate e rispettivi vincoli per oggetto e livello territoriale.

Oggetto	Livello territoriale	Numero di tavole a stima ⁽¹⁾	Numero di vincoli indotti	Numero di tavole a stima ⁽¹⁾ ottimizzate	Numero di vincoli indotti
Popolazione	Area di censimento	1	8	1	8
	Provincia/Grande Comune	25	401	11	214
	Regione	19	1.078	9	568
	Ripartizione	8	356	3	270
Famiglie e Alloggi ed edifici	Area di censimento	6	14	3	11
	Provincia/Grande Comune	22	135	5	57
	Regione	8	23	4	18
	Ripartizione	0	0	0	0

⁽¹⁾ Ipotesi di Piano di diffusione dei risultati del Censimento della popolazione e delle abitazioni del 2001 confermato per il 2011.

Fonte: Direzione centrale dei Censimenti generali (DCCG), Istat.

In conclusione, l’approccio A del “sistema di pesi unico” ha condotto al disegno due sistemi di vincoli per i tre oggetti. Però, dovendo necessariamente procedere con la ponderazione su domini per ripartizione geografica, si sarebbe generato un così elevato numero di vincoli e di dati da rendere impraticabile tale soluzione per motivi di complessità computazionale (cfr. paragrafo 5.3.2).

5.3.1.2 Approccio B - Sistema di pesi a blocchi

Per superare la complessità operativa del processo di riponderazione riscontrata con l’approccio A si è pensato di disegnare, per uno stesso oggetto, diversi sistemi di vincoli ciascuno dei quali riferito a uno specifico insieme di tavole e quindi di variabili da stimare (Carbonetti e Righi, 2011). L’idea era quella di suddividere le tavole a stima in insiemi, detti “blocchi”, disgiunti (ogni tavola appartiene ad uno e un solo blocco) e indipendenti rispetto alle variabili da stimare. In tal modo, il generico sistema di pesi finale sarebbe stato impiegato per stimare l’insieme delle variabili, e solo quelle, del blocco di riferimento. Quindi, a partire dal raggruppamento delle tavole a stima per aree tematiche e livelli territoriali e applicando i processi di ottimizzazione descritti nel paragrafo 5.2.4, si ottiene il sistema di vincoli ottimizzato su più blocchi riportato nel Prospetto 5.4.

Prospetto 5.4 - Approccio B: sistema di vincoli a blocchi per oggetto, livello territoriale e vincoli indotti.

Oggetto	Blocco	Variabili esaustive	Livello territoriale	Variabili da stimare	Numero di vincoli
Popolazione	1	Occupati X Sesso X Classe di età 15 anni e più (4)	Ace	Attività economica (3)	8
Popolazione	1	Occupati X Sesso X Classe di età 15 anni e più (13)	Provincia	Attività lavorativa (10)	26
Popolazione	1	Occupati X Sesso X Stato civile (6)	Provincia	Attività lavorativa (10)	12
Popolazione	1	Occupati X Sesso X Grado di istruzione (7)	Provincia	Tipologia attività lavorativa (pieno-parziale) (2)	14
Popolazione	1	Pendolari per studio X Sesso X Luogo di destinazione (5)	Provincia	Mezzo utilizzato (9) X Tempo impiegato in minuti (5)	10
Popolazione	1	Pendolari per lavoro X Sesso X Luogo di destinazione (5)	Provincia	Mezzo utilizzato (9) X Tempo impiegato in minuti (5)	10
Popolazione	1	Occupati stranieri X Sesso X Classe di età 15 anni e più (9)	Provincia	Attività economica (3) X Posizione nella professione (2)	18
Popolazione	1	Popolazione di 6 anni e più con diploma scuola secondaria superiore X Sesso X Classe di età 6 anni e più (13)	Provincia	Grado di istruzione (9)	26
Popolazione	1	Popolazione in cerca di prima occupazione X Sesso X Classe di età 15 anni e più (6) X Titolo di studio (5)	Regione	Titolo di studio (77)	60
Popolazione	1	Occupati (femmine) residenti in famiglia di 15 anni e più X Classe di età 15 anni e più (9) X Numero di figli (4)	Regione	Condizione e posizione nella professione (10)	36
Popolazione	1	Popolazione di 6 anni e più X Sesso X Classe di età 6 anni e più (13) X Titolo di studio (5)	Regione	Titolo di studio (273)	130
Popolazione	1	Pendolari per studio o lavoro X Sesso X Classe di età (9)	Regione	Mezzo utilizzato (9)	18
Popolazione	1	Occupati X Sesso X Classe di età 15 anni e più (9) X Grado di istruzione (6)	Ripartizione	Settore di attività economica (6)	108
Popolazione	1	Occupati X Sesso X Titolo di studio (9)	Ripartizione	Posizione nella professione (5) X Settore di attività economica (17)	18
Popolazione	1	Occupati stranieri X Sesso X Area geografica e paese di cittadinanza (90)	Ripartizione	Attività economica (4)	180
Popolazione	2	Stranieri nati all'estero X Sesso X Area geografica di residenza (6)	Provincia	Anno di trasferimento in Italia (5)	12
Popolazione	2	Stranieri nati all'estero X Sesso X Area geografica e paese di cittadinanza (90)	Regione	Anno di trasferimento in Italia (5)	180
Popolazione	3	Popolazione di 15 anni e più X Sesso X Età per singolo anno da 15 a 19 anni e classi di età 15 anni e più (15)	Provincia	Iscrizione ad un corso regolare di studi (2)	30
Popolazione	3	Popolazione fino a 14 anni X Sesso X Età singolo anno da 0 a 14 anni (15)	Provincia	Frequenza asilo nido/scuola materna/prima elementare (2) X Iscrizione ad un corso regolare di studi (2)	30
Popolazione	4	Popolazione in abitazione X Numero occupanti (6) X Titolo di godimento (3)	Provincia	Numero di stanze (6)	18
Popolazione	4	Popolazione in abitazione X Numero di famiglie (2)	Regione	Figura giuridica del proprietario (7)	2
Popolazione	4	Popolazione in abitazione X Numero di componenti (6)	Regione	Tipo di servizi (12)	6
Popolazione	5	Popolazione di 14 anni e più X Sesso X Classe di età 14 anni e più (12) X Grado di istruzione (7)	Regione	Iscrizione ad un corso regolare di studi (2)	158
Popolazione	5	Popolazione di 15 anni e più X Sesso X Condizione nella professione (8)	Regione	Iscrizione ad un corso regolare di studi (2)	16

Fonte: Direzione centrale dei Censimenti generali (DCCG), Istat.

Prospetto 5.4 - Approccio B: sistema di vincoli a blocchi per oggetto, livello territoriale e vincoli indotti (segue).

Oggetto	Blocco	Variabili esaustive	Livello territoriale	Variabili da stimare	Numero di vincoli
Famiglie e Alloggi ed edifici	6	Numero di abitazioni X Tipo di occupazione e servizi (3)	Ace	Tipo di occupazione e servizi (4)	3
Famiglie e Alloggi ed edifici	6	Numero di abitazioni in edifici ad uso abitativo X Epoca di costruzione (7)	Ace	Stanze delle abitazioni in edifici ad uso abitativo (1)	7
Famiglie e Alloggi ed edifici	6	Numero di abitazioni occupate da residenti	Ace	Disponibilità di servizi (11)	1
Famiglie e Alloggi ed edifici	6	Numero di famiglie in abitazione X Titolo di godimento (3)	Provincia	Numero di stanze (6)	3
Famiglie e Alloggi ed edifici	6	Numero di abitazioni occupate da residenti X Classi di superficie (mq) (9)	Provincia	Disponibilità e numero di cucine (9)	9
Famiglie e Alloggi ed edifici	6	Numero di abitazioni occupate da residenti in edifici ad uso abitativo X Epoca di costruzione (7) X Titolo di godimento (3)	Provincia	Figura giuridica del proprietario (7)	21
Famiglie e Alloggi ed edifici	6	Numero di abitazioni occupate da residenti in edifici ad uso abitativo X Numero di abitazioni (6)	Provincia	Servizi installati (6)	6
Famiglie e Alloggi ed edifici	6	Numero di abitazioni occupate da residenti X Titolo di godimento (3) X Numero di occupanti (6)	Provincia	Numero di stanze (6)	18
Famiglie e Alloggi ed edifici	6	Numero di famiglie in abitazione X Numero di famiglie (2)	Regione	Figura giuridica del proprietario (7)	2
Famiglie e Alloggi ed edifici	6	Numero di famiglie in abitazione X Numero di componenti (6)	Regione	Tipo di servizi (12)	6
Famiglie e Alloggi ed edifici	6	Famiglie unipersonali non in coabitazione in abitazione X Classe di età decennale (8)	Regione	Numero di stanze (6)	8
Famiglie e Alloggi ed edifici	6	Numero di abitazioni occupate da residenti X Numero di famiglie (2)	Regione	Numero di stanze (6)	2

Fonte: Direzione centrale dei Censimenti generali (DCCG), Istat.

Prospetto 5.5 - Approccio B: numero di tavole per oggetto e blocco.

Oggetto	Blocco	Numero di tavole
Popolazione	1	39
Popolazione	2	3
Popolazione	3	5
Popolazione	4	4
Popolazione	5	2
Famiglie e Alloggi ed edifici	6	36
Totale		89

Fonte: Direzione centrale dei Censimenti generali (DCCG), Istat.

Sulla base del Piano di diffusione del 2001 sono stati individuati sei blocchi (Prospetto 5.5), cinque per l'oggetto "popolazione" e uno unico per gli oggetti "famiglie" e "alloggi ed edifici". Tale approccio porta a determinare sei sistemi di pesi finali. La soluzione del sistema multiplo di pesi, pur risolvendo la criticità computazionale incontrata con il precedente approccio, non è stata presa in considerazione in quanto giudicata incompatibile con le architetture software impiegate per la messa in esercizio dei *data warehouse* di produzione e di diffusione dei risultati finali (cfr. paragrafo 5.3.2).

5.3.1.3 Approccio C - Sistema di pesi primario e secondario

Prima di rilasciare la versione definitiva del Piano di diffusione nazionale dei dati finali del 15° Censimento, sono state prodotte differenti versioni che hanno, di volta in volta, recepito:

- gli avanzamenti proposti dai tematici sulla strategia di diffusione dei risultati;
- i suggerimenti giunti da alcuni gruppi di lavoro Istat¹⁰⁵ preposti a definire le attività di produzione dei dati censuari e le linee operative per lo sviluppo dei sistemi informatici di supporto alla diffusione;
- le indicazioni derivanti dai risultati delle sperimentazioni effettuate sui dati del censimento 2001 per valutare la praticabilità della procedura di riponderazione.

Dal disegno delle tavole di diffusione nella loro versione definitiva e dall'esito dei test sperimentali condotti per i vari approcci ipotizzati, la strategia di riponderazione si è orientata verso uno schema composto da tre sistemi di vincoli, uno per ciascuno degli "oggetti" del Piano di diffusione (cfr. paragrafo 5.2.4). I relativi sistemi definitivi di vincoli (esposti nel successivo paragrafo 5.4) sono stati disegnati sul Piano di diffusione nazionale dei dati censuari definitivo rilasciato a gennaio 2014 (versione 2.0).

I limiti computazionali e le scelte dei sistemi informatici per la diffusione dei risultati hanno reso impraticabili gli approcci A e B descritti in precedenza (paragrafi 5.3.1.1 e 5.3.1.2). La soluzione ritenuta più facilmente praticabile è stata quella di determinare un sistema di pesi primario per ognuno degli oggetti del Piano di diffusione, uno per "popolazione", uno per "famiglie" e uno per "alloggi ed edifici", in modo tale da garantire la coerenza dei dati per la quasi totalità delle tavole statistiche di diffusione.

Poiché il raggiungimento di tale obiettivo richiedeva una diminuzione del numero dei vincoli indotti dalle tavole a stima, è stato necessario apportare alcune modifiche al Piano di diffusione. La modifica più rilevante ha riguardato, per alcune tavole statistiche, la riduzione del dettaglio territoriale eliminando il livello Ace e proponendo una diffusione dei dati per domini territoriali più aggregati. Un'altra importante modifica è stata la riduzione del dettaglio classificatorio di alcune tavole tramite l'adozione, per alcune variabili di incrocio, di classificazioni con un numero ridotto di modalità. Quest'ultima soluzione ha permesso di risolvere in anticipo eventuali problemi di natura computazionale e di non convergenza dell'algoritmo di riponderazione.

In alternativa, per la produzione di tre tavole particolarmente complesse che non potevano essere garantite dai tre sistemi di pesi primari e perciò escluse dal disegno dei vincoli, è stato calcolato un sistema di pesi secondario, detto anche "*ad hoc*" (cfr. paragrafo 5.4.1).

5.3.2 Test sulle procedure di calibrazione

Le sperimentazioni finalizzate alla scelta del sistema di calibrazione più opportuno sono state condotte sull'oggetto "popolazione" utilizzando il data base del censimento del 2001 da cui sono stati estratti campioni casuali di famiglie secondo la strategia di campionamento adottata per il censimento del 2011. Si è deciso di preferire l'universo della popolazione residente in famiglia rispetto agli altri due ("famiglie"; "alloggi ed edifici") perché, comportando un sistema di vincoli molto più articolato e una grande mole di dati, avrebbe permesso di studiare le soluzioni software e hardware necessarie per implementare le procedure di riponderazione nella situazione di maggiore complessità. Tali soluzioni si sarebbero così potute impiegare senza rischi anche per gli altri "oggetti" per cui si prevedeva una minore difficoltà computazionale.

Il primo disegno di sistema di vincoli, descritto nel paragrafo 5.3.1.1 (Approccio A), ha rappresentato una situazione "estrema" studiata solo da un punto di vista teorico. L'adozione di un sistema unico di vincoli che potesse garantire la coerenza di tutte le tavole previste dal Piano di diffusione del 2011 a tutti i livelli territoriali, dall'area di censimento fino alla ripartizione geografica, era impossibile da sperimentare per motivi legati alla complessità computazionale (troppi vincoli causati dai troppi domini di stima); sono state quindi condotte sperimentazioni sulle altre strategie proposte per il calcolo dei pesi di riporto all'universo.

¹⁰⁵ I gruppi di lavoro fanno riferimento ai seguenti due progetti:

- Progetto per lo sviluppo e la gestione del sistema di diffusione dei dati censuari (gruppi di lavoro DIDAC costituito con deliberazione Istat n. 5/DIQR del 28.10.2011, e DIDAC2 costituito con deliberazione Istat n. 34/DGEN del 27.02.2013);
- Progetto per la produzione dei dati definitivi del 15° Censimento generale della popolazione e delle abitazioni (gruppo di Lavoro PRODAC costituito con deliberazione Istat n. 13/DICA del 15.04.2013).

5.3.2.1 Riponderazione multipla e procedura bottom-up

Per risolvere le criticità computazionali indotte dal sistema unico è stato seguito l'approccio che ha portato all'individuazione di 6 differenti sistemi di pesi (cfr. paragrafo 5.3.1.2). I sistemi di pesi fanno riferimento a specifici insiemi disgiunti di tavole oggetto di stima: una tavola appartiene ad uno e un solo insieme e, di conseguenza, le frequenze assolute oggetto di stima riferite alle celle di massimo incrocio di una data tavola sono determinate da uno e un solo sistema di pesi di riporto all'universo. La possibilità di impiegare differenti sistemi di pesi aveva l'obiettivo di garantire la perfetta coerenza tra dati stimati e dati esaustivi per tutte le tavole del Piano di diffusione.

In fase sperimentale, non era possibile procedere sui domini di calibrazione riferiti alle ripartizioni geografiche in quanto il relativo campione di famiglie (unione dei campioni di tutti i comuni superiori ai 20 mila abitanti coinvolti dalla strategia campionaria) comportava un numero di record individuali troppo elevato per essere gestibile in fase di elaborazione dati. Quindi, per ridurre le dimensioni del data base di riferimento, le sperimentazioni sono state eseguite su domini regionali; a tale scopo i vincoli disegnati per la ripartizione geografica sono stati applicati già al livello regionale in modo da essere automaticamente soddisfatti anche a livello più aggregato.

I risultati delle prime prove effettuate su alcune Regioni, scelte per diversa dimensione demografica, hanno mostrato la convergenza della procedura di calibrazione per tutti i blocchi disegnati ad eccezione del blocco 1.

Il sistema di vincoli riferito al blocco 1 prevedeva 8 vincoli a livello di area di censimento, 116 a livello di provincia/grande comune, 244 a livello di Regione e 306 a livello di ripartizione (riportati a livello regionale). L'elevato numero di vincoli¹⁰⁶ che si andava a configurare ha fatto emergere problemi computazionali simili a quelli riscontrati per il "sistema unico", per quasi tutte le Regioni.

Al fine di risolvere le criticità emerse per il blocco 1, si è proceduto con un approccio di tipo *bottom-up*: nella prima fase si calcolano i pesi di area per tutte le aree di censimento della Regione in esame; nella seconda fase si determinano i pesi per ciascuna delle Province contenute nella Regione e, nel caso di una Provincia che includa un grande Comune, si calcolano contemporaneamente a quella Provincia anche i pesi riferiti a quel grande Comune; infine, nella terza fase, si calcolano i pesi finali relativi al contesto regionale. Ad ogni livello territoriale il sistema di vincoli da implementare nella riponderazione doveva includere (per garantire la coerenza delle stime finali), come ulteriori dati di *benchmark*, le frequenze assolute (stimate) relative alle tavole a stima dei livelli territoriali inferiori, definite tramite i pesi calcolati a quel livello territoriale. Il risultato finale ha condotto alla definizione, per il blocco 1, di tre sistemi di pesi: uno per le tavole di diffusione disegnate per le aree di censimento, uno per quelle provinciali e di grande Comune, uno per quelle regionali.

La strategia della riponderazione per diversi blocchi (Approccio B) praticata con la procedura di calcolo *bottom-up*, porta all'impiego di otto sistemi di pesi finali: tre per il complesso blocco 1 e uno per ciascuno degli altri cinque blocchi. I risultati consentono di stimare tutte le tavole statistiche stabilite dal Piano di diffusione; per ogni nuova tavola non prevista, l'uso di uno dei sistemi di pesi definiti richiede la verifica della compatibilità dei relativi vincoli indotti con il corrispondente sistema implementato nella procedura. In alternativa si dovrà necessariamente procedere con una calibrazione *ad hoc*.

Però, per il calcolo dei pesi finali per il censimento del 2011 l'approccio appena esposto non è stato adottato perché la presenza di più sistemi di pesi è stato ritenuto incompatibile con le architetture software implementate per lo sviluppo e la gestione del *data warehouse* di diffusione. Per risolvere questo problema si è deciso di impiegare un solo sistema di pesi per ciascuno degli universi oggetto della rilevazione censuaria ("popolazione", "famiglie", "alloggi ed edifici") contemporaneamente ad una revisione del Piano di diffusione al fine di modificare le tavole più complesse (cfr. paragrafo 5.3.1.3).

¹⁰⁶ La numerosità del sistema dei vincoli per i domini regionali viene calcolata moltiplicando il numero di vincoli di ciascun dominio territoriale per il numero di domini esistenti nella prefissata regione. Per esempio, per la regione Liguria, che ha 61 aree di censimento, 4 province e 1 grande comune, il numero totale di vincoli è pari $(61 \times 8) + (116 \times 4) + (116 \times 1) + (244 \times 1) + (306 \times 1) = 1.618$.

5.3.2.2 Riponderazione unica

Con l'approccio della riponderazione unica il sistema di vincoli non viene più diviso in blocchi ma, con riferimento al dominio regionale, se ne disegna uno unico che considera tutti i vincoli indotti dalle tavole a stima tranne quelli riferiti all'area di censimento che, come già esposto, causano problemi computazionali per l'elevato numero di vincoli associato. Per esempio, nei 66 Comuni a campione della Regione Lombardia (cfr. Tavola 3.2) sono state disegnate 249 aree di censimento. Poiché il numero di vincoli riferiti all'area di censimento è pari ad 8, il numero totale di vincoli riferiti "solo" a tale livello territoriale per la Lombardia è pari a 1.992 (249x8); aggiungendo poi i vincoli per provincia/grande comune e quelli per Regione, il numero complessivo diventa così elevato da rendere impraticabile la procedura di riponderazione.

Per questo è stato deciso di escludere dal Piano di diffusione del 2011 tutte le tavole a stima riferibili al livello territoriale dell'area di censimento, rimandando ad una successiva fase di pubblicazione dei dati ufficiali la possibilità di produrre tavole che incrociano variabili *long* per le aree di censimento dei Comuni a campione tramite l'impiego di "pesi *ad hoc*". Così facendo, nella procedura di riponderazione per il calcolo del sistema di pesi unico, non è stato considerato il dettaglio territoriale dell'area di censimento.

Le sperimentazioni sono state condotte, con riferimento al data base del censimento del 2001, sull'oggetto "popolazione" e per un sistema di vincoli implementato a partire dal livello territoriale di provincia/grande comune. Il disegno del sistema di vincoli ha dapprima previsto l'inclusione di tutti i vincoli dati dalle tavole a stima e, nei passi successivi, l'esclusione di quelli che causavano problemi alla convergenza dell'algoritmo o generavano incompatibilità con altri vincoli. Durante la fase di test sono emerse delle problematiche per alcune tavole.

- Nelle tavole sul pendolarismo [P3.66], [P3.68], e [P3.69] la modalità relativa al luogo di destinazione "di cui Comune capoluogo" ha creato problemi di incoerenza all'interno dello stesso dominio territoriale. Per esempio, si supponga che il vincolo al livello provinciale "numero di pendolari maschi che si recano per motivi di studio in altra Provincia della stessa Regione" presenti un totale noto pari a 9 e un totale campionario pari a 2, mentre il vincolo "pendolari maschi che si recano per motivi di studio in altra Provincia della stessa Regione: di cui Comune capoluogo" presenti un totale noto pari a 6 e un totale campionario pari a 2; ciò comporta che per entrambi i vincoli sono state estratte le stesse unità campionarie anche se il totale riferito al secondo vincolo è un sub-totale. Per risolvere tale incompatibilità si è deciso di ridurre, nel sistema di vincoli, il dettaglio classificatorio della variabile "luogo di destinazione" eliminando la sotto-modalità "di cui Comune capoluogo".

- La tavola [P2.33] relativa alla "Popolazione straniera residente nata all'estero per Sesso (3), Anno di trasferimento (6), Area geografica e Paese di cittadinanza (127)" ha causato la mancata convergenza dell'algoritmo di calibrazione. Per risolvere il problema è stato deciso di ridurre il numero di variabili di incrocio eliminando la variabile a stima "Anno di trasferimento (6)"; così facendo suddetta la tavola non è più "a stima" perché include solo variabili esaustive.

- La tavola regionale [2.18] relativa alla "Popolazione residente di 6 anni e più che ha conseguito il titolo di studio più elevato all'estero per Sesso (3), Grado di istruzione (8) e Area geografica di cittadinanza (19)" entrava in conflitto con la tavola provinciale [P3.78] riferita alla "Popolazione straniera residente nata all'estero per Sesso (3), Anno di trasferimento (6) e Area geografica di cittadinanza (continenti) (7)". Infatti, pur avendo differente oggetto, le unità campionarie relative alle modalità "Oceania" e "Apolide" per la variabile "Area geografica di cittadinanza" erano in numero esiguo e coincidenti in entrambe le tavole. È stato così deciso di escludere la tavola [P2.18] dal disegno del sistema di vincoli.

Dopo aver risolto le criticità delle tavole sopra illustrate, si è giunti al disegno definitivo del sistema di vincoli in seguito implementato per il calcolo dei pesi finali associabili all'oggetto "popolazione" (cfr. paragrafo 5.4.1).

Sono state escluse dal disegno del suddetto sistema di vincoli le seguenti tavole:

- tavola a livello di ripartizione geografica [P1.9] relativa alla "Popolazione straniera residente occupata, per Sesso (3), Area geografica di cittadinanza (21), Attività economica (4)";
- tavola a livello di ripartizione geografica [P1.10] relativa alla "Popolazione straniera residente occupata, per Sesso (3), Area geografica e Paese di cittadinanza (127), Attività economica (4)";
- tavola a livello di provincia/grande comune [P3.43] relativa alla "Popolazione residente di 6 anni e più per Sesso (3), Età in classi quinquennali dai 6 anni in poi (16), Grado di istruzione (19)".

Per la produzione di queste tavole si è proceduto all'elaborazione di un sistema di “*pesi ad hoc*”.

L'approccio basato sulla riponderazione unica è stato adottato per calcolare i pesi di riporto all'universo anche per gli oggetti “famiglie” e “alloggi ed edifici”.

I tre sistemi di vincoli sono stati sperimentati su domini regionali di diversa dimensione da cui è emerso che sulle Regioni molto piccole come la Valle d'Aosta, data la ridotta consistenza numerica delle frequenze assolute riferite ai vincoli di calibrazione, il processo di convergenza dell'algoritmo di riponderazione ha successivamente presentato incompatibilità con la tecnica della donazione implementata nel processo di calcolo per superare alcune criticità di tipo computazionale (cfr. paragrafo 5.5.2).

5.4 Il disegno definitivo del sistema dei vincoli

5.4.1 Oggetto “popolazione”

Nel Prospetto 5.6 è presentato l'insieme delle tavole a stima per l'oggetto “popolazione”. Per ciascuna delle 57 tavole iniziali sono state individuate le marginali esaustive da garantire con la calibrazione; tali marginali sono state sottoposte al processo di ottimizzazione “orizzontale” e “verticale” (cfr. paragrafo 5.2.4), arrivando così a determinare il sistema finale di totali noti ottimizzato (Prospetto 5.7) che comprende 184 vincoli a livello di provincia/grande comune¹⁰⁷, 618 vincoli a livello regionale e 20 vincoli a livello di ripartizione (Prospetto 5.8).

Come illustrato nel paragrafo 5.3.2.2, per le tre tavole riferite all'oggetto “popolazione” [P1.9], [P1.10], e [P3.43], di particolare interesse per la diffusione, è stato deciso di calcolare un sistema di “*pesi ad hoc*” per evitare che l'elevato numero di vincoli indotti dalle marginali esaustive di queste tavole potesse creare problemi di tipo computazionale. Nel Prospetto 5.9 è descritto il sistema ottimizzato di vincoli implementato nella procedura di riponderazione per il calcolo del sistema di “*pesi ad hoc*” (chiamato anche sistema di pesi secondario). Si fa presente che, poiché le tavole [P1.9] e [P1.10] hanno un dettaglio informativo che si differenzia solo per le variabili “Area geografica di cittadinanza (21)” nella [P1.9] e “Area geografica e Paese di cittadinanza (127)” nella [P1.10], i vincoli indotti dalla [P1.9] sono un sotto-insieme di quelli indotti dalla [P1.10]. Il processo di ottimizzazione tende quindi ad eliminare i vincoli ridondanti (cfr. paragrafo 5.2.4) che sono proprio quelli indotti dalla tavola [P1.9].

Per facilitare la lettura del Prospetto 5.6 si fa presente che nella prima colonna è riportato il codice della tavola seguito dall'oggetto di interesse (seconda colonna) e dal livello territoriale di riferimento (terza colonna). La quarta colonna riporta le variabili componenti la tavola, mentre nella quinta e sesta colonna sono riportati rispettivamente gli incroci che derivano dal conteggio esaustivo e quelli che sono determinati tramite stima. Nella settima colonna è riportato il numero di vincoli riferiti alla tavola considerata e coincidente con il relativo numero di incroci di tipo esaustivo.

La struttura descritta per il Prospetto 5.6 si ripete in modo analogo per i successivi Prospetti 5.7 - 5.9 - 5.10 - 5.11 - 5.13 - 5.14 che illustrano il dettaglio di ulteriori sistemi di vincoli.

¹⁰⁷ Al censimento del 2011 tutti i comuni con più di 150 mila abitanti sono stati considerati “Grandi Comuni”. In totale sono 25 e sono, in ordine di ampiezza demografica, i seguenti: Roma, Milano, Napoli, Torino, Palermo, Genova, Bologna, Firenze, Bari, Catania, Venezia, Verona, Messina, Padova, Trieste, Taranto, Brescia, Prato, Reggio di Calabria, Modena, Parma, Perugia, Reggio nell'Emilia, Livorno, Ravenna. Per tali comuni il Piano di diffusione nazionale dei risultati censuari ha previsto la pubblicazione delle stesse tavole statistiche definite per il livello territoriale provinciale.

Prospetto 5.6 - Sistema di vincoli non ottimizzato per l'oggetto "popolazione". Tavole soggette a stima e rispettivi vincoli indotti per livello territoriale (Piano di diffusione 2011 ver. 2.0 gennaio 2014).

Tavola	Oggetto	Livello territoriale	Incrocio	Marginali esaustive (al netto di totali e sub-totali)	Marginali a stima (al netto di totali e sub-totali)	Numero di vincoli
P3.30	Popolazione residente in famiglia in abitazione	Provincia	Disponibilità di servizi (11) X Numero di occupanti (7)	Numero di occupanti (6)	Disponibilità di servizi (11)	6
P3.31	Popolazione residente in famiglia in abitazione	Provincia	Numero di occupanti (7) X Numero di stanze (7) X Titolo di godimento (4)	Numero di occupanti (6) X Titolo di godimento (3)	Numero di stanze (6)	18
P3.40	Popolazione residente con meno di 6 anni	Provincia	Sesso (3) X Età per singolo anno da 0 a 5 anni (9) X Frequenza asilo nido/scuola materna/prima elementare (3)	Sesso (2) X Età per singolo anno da 0 a 5 anni (6)	Frequenza asilo nido/scuola materna/prima elementare (2)	12
P3.41	Popolazione residente con meno di 6 anni	Provincia	Sesso (3) X Età per singolo anno da 0 a 5 anni (9) X Frequenza asilo nido/scuola materna/prima elementare (7)	Sesso (2) X Età per singolo anno da 0 a 5 anni (6)	Frequenza asilo nido/scuola materna/prima elementare (5)	12
P3.42	Popolazione residente di 15 anni e più	Provincia	Sesso (3) X Età per singolo anno da 15 a 19 anni e classi di età 15 anni e più (17) X Iscrizione ad un corso regolare di studi (3)	Sesso (2) X Età per singolo anno da 15 a 19 anni e classi di età 15 anni e più (15)	Iscrizione ad un corso regolare di studi (2)	30
P3.43	Popolazione residente di 6 anni e più	Provincia	Sesso (3) X Classe di età quinquennale 6 anni e più (16) X Grado di istruzione (19)	Sesso (2) X Classe di età quinquennale 6 anni e più (15) X Grado di istruzione (2 modalità: Diploma 2,3 anni o Diploma 4,5 anni)	Grado di istruzione (7)	60
P3.46	Popolazione residente fino a 16 anni che frequenta l'asilo nido, micro-asilo, baby-parking, eccetera o la scuola dell'infanzia o che è iscritta ad un corso regolare di studi	Provincia	Sesso (3) X Età per singolo anno da 0 a 16 anni (23)	Sesso (2) X Età per singolo anno da 0 a 16 anni (17)		34
P3.47	Popolazione residente da 6 a 16 anni	Provincia	Sesso (3) X Età per singolo anno da 6 a 16 anni (15) X Iscrizione ad un corso regolare di studi (3)	Sesso (2) X Età per singolo anno da 6 a 16 anni (11)	Iscrizione ad un corso regolare di studi (2)	22
P3.48	Tasso di frequenza/iscrizione	Provincia	Sesso (3) X Età per singolo anno da 0 a 16 anni (23)	Sesso (2) X Età per singolo anno da 0 a 16 anni (17)		34
P3.49	Occupati	Provincia	Sesso (3) X Attività economica (4) X Grado di istruzione (7) X Posizione nella professione (7)	Sesso (2) X Grado di istruzione (6)	Attività economica (3) X Posizione nella professione (6)	12
P3.50	Occupati	Provincia	Sesso (3) X Attività lavorativa svolta (11) X Classe di età 15 anni e più (14)	Sesso (2) X Classe di età 15 anni e più (13)	Attività lavorativa svolta (10)	26
P3.51	Occupati	Provincia	Sesso (3) X Attività lavorativa svolta (11) X Ore di lavoro effettuate (7)	Sesso (2)	Attività lavorativa svolta (10) X Ore di lavoro effettuate (6)	2
P3.52	Occupati	Provincia	Sesso (3) X Stato civile (7) X Attività lavorativa svolta (11)	Sesso (2) X Stato civile (6)	Attività lavorativa svolta (10)	12
P3.53	Occupati	Provincia	Sesso (3) X Classe di età 15 anni e più (12) X Ore di lavoro effettuate (7)	Sesso (2) X Classe di età 15 anni e più (11)	Ore di lavoro effettuate (6)	22

Fonte: Direzione centrale dei Censimenti generali (DCCG), Istat.

Prospetto 5.6 - Sistema di vincoli non ottimizzato per l'oggetto "popolazione". Tavole soggette a stima e rispettivi vincoli indotti per livello territoriale (Piano di diffusione 2011, versione 2.0 - gennaio 2014) (segue).

Tavola	Oggetto	Livello territoriale	Incrocio	Marginali esaustive (al netto di totali e sub-totali)	Marginali a stima (al netto di totali e sub-totali)	Numero di vincoli
P3.54	Occupati	Provincia	Sesso (3) X Classe di età 15 anni e più (12) X Posizione nella professione (4) X Sezioni di attività economica (18)	Sesso (2) X Classe di età 15 anni e più (11)	Posizione nella professione (3) X Sezioni di attività economica (17)	22
P3.55	Occupati	Provincia	Sesso (3) X Classe di età 15 anni e più (14) X Posizione nella professione (8)	Sesso (2) X Classe di età 15 anni e più (13)	Posizione nella professione (7)	26
P3.56	Occupati	Provincia	Sesso (3) X Grado di istruzione (8) X Tipologia attività lavorativa (pieno-parziale) (3)	Sesso (2) X Grado di istruzione (7)	Tipologia attività lavorativa (pieno-parziale) (2)	14
P3.57	Occupati	Provincia	Sesso (3) X Ore di lavoro effettuate (7) X Posizione nella professione (8)	Sesso (2)	Ore di lavoro effettuate (6) X Posizione nella professione (7)	2
P3.58	Occupati	Provincia	Sesso (3) X Posizione nella professione (7) X Settori di attività economica (22)	Sesso (2)	Posizione nella professione (6) X Settori di attività economica (21)	2
P3.59	Occupati	Provincia	Sesso (3) X Stato civile (7) X Posizione nella professione (7) X Sezioni di attività economica (18)	Sesso (2) X Stato civile (6)	Posizione nella professione (6) X Sezioni di attività economica (17)	12
P3.60	Occupati dipendenti	Provincia	Sesso (3) X Classe di età 15 anni e più (12) X Rapporto di lavoro (Tempo indeterminato-determinato) (3)	Sesso (2) X Classe di età 15 anni e più (11)	Rapporto di lavoro (Tempo indeterminato-determinato) (2)	22
P3.66	Occupati che si recano al luogo abituale di lavoro	Provincia	Attività economica (4) X Luogo di destinazione (12) X Mezzo utilizzato (10) X Tempo impiegato in minuti (6)	Luogo di destinazione (5)	Attività economica (3) X Mezzo utilizzato (9) X Tempo impiegato in minuti (5)	5
P3.68	Popolazione residente che si reca al luogo abituale di studio	Provincia	Sesso (3) X Luogo di destinazione (12) X Mezzo utilizzato (10) X Tempo impiegato in minuti (6)	Sesso (2) X Luogo di destinazione (5)	Mezzo utilizzato (9) X Tempo impiegato in minuti (5)	10
P3.69	Occupati che si recano al luogo abituale di lavoro	Provincia	Sesso (3) X Luogo di destinazione (12) X Mezzo utilizzato (10) X Tempo impiegato in minuti (6)	Sesso (2) X Luogo di destinazione (5)	Mezzo utilizzato (9) X Tempo impiegato in minuti (5)	10
P3.78	Popolazione straniera residente nata all'estero	Provincia	Sesso (3) X Anno di trasferimento (6) X Area geografica di cittadinanza (continenti) (7)	Sesso (2) X Area geografica di cittadinanza (continenti) (6)	Anno di trasferimento (5)	12
P3.80	Popolazione straniera residente occupata	Provincia	Sesso (3) X Classe di età 15 anni e più (10) X Attività economica (4) X Posizione nella professione (4)	Sesso (2) X Classe di età 15 anni e più (9)	Attività economica (3) X Posizione nella professione (3)	18
P3.1	Occupati	Regione/Provincia	Sesso (3) X Attività economica (4) X Classe di età 15 anni e più (5)	Sesso (2) X Classe di età 15 anni e più (4)	Attività economica (3)	8
P2.7	Popolazione residente in famiglia in abitazione	Regione	Figura giuridica del proprietario (8) X Numero di famiglie (3)	Numero di famiglie (2)	Figura giuridica del proprietario (7)	2
P2.8	Popolazione residente in famiglia in abitazione	Regione	Numero di componenti (7) X Tipo di servizi (7)	Numero di componenti (6)	Tipo di servizi (7)	6

Fonte: Direzione centrale dei Censimenti generali (DCCG), Istat.

Prospetto 5.6 - Sistema di vincoli non ottimizzato per l'oggetto "popolazione". Tavole soggette a stima e rispettivi vincoli indotti per livello territoriale (Piano di diffusione 2011, versione 2.0 - gennaio 2014) (segue).

Tavola	Oggetto	Livello territoriale	Incrocio	Marginali esaustive (al netto di totali e sub-totali)	Marginali a stima (al netto di totali e sub-totali)	Numero di vincoli
P2.9	Popolazione residente di 14 anni e più	Regione	Sesso (3) X Classe di età 14 anni e più (13) X Frequenza corsi di formazione (3) X Grado di istruzione (8)	Sesso (2) X Classe di età 14 anni e più (12) X Grado di Istruz. (4) (modalità 5,6,7 sono garantite dalla P2.16)	Frequenza corsi di formazione (2)	96
P2.11	Popolazione residente di 15 anni e più	Regione	Sesso (3) X Frequenza corsi di formazione (3) X Condizione (9)	Sesso (2) X Condizione (6)	Frequenza corsi di formazione (2)	12
P2.12	Popolazione residente di 6 anni e più	Regione	Sesso (3) X Classe di età quinquennale 6 anni e più (16) X Grado di istruzione (19)	Sesso (2) X Classe di età quinq. 6 anni e più (15) X Grado di Istruzione 5,6	Grado di istruzione (6)	30
P2.13	Popolazione residente che ha conseguito almeno il diploma di scuola secondaria superiore	Regione	Sesso (3) X Corso di formazione professionale reg./prov. di durata pari o sup. a 6 mesi, a cui si accede con il diploma di scuola secondaria superiore (3)	Sesso (2) X Grado di istruzione 5-6-7 di Grado di istruzione (8)	Corso di formazione professionale reg./prov. di durata pari o sup. 6 mesi, a cui si accede con diploma di scuola secondaria superiore (2)	16
P2.14	Popolazione residente che ha conseguito almeno la licenza di scuola media (o avviamento professionale)	Regione	Sesso (3) X Corso di formazione professionale regionale/provinciale della durata pari o superiore ai 24 mesi, a cui si accede con la licenza media (4)	Sesso (2) Grado di istruzione 4-5-6-7 di Grado di istruzione (8)	Corso di formazione professionale regionale/provinciale della durata pari o sup. a 24 mesi, a cui si accede con licenza media (3)	2
P2.15	Popolazione residente di 6 anni e più	Regione	Sesso (3) X Classe di età quinquennale 6 anni e più (16) X Titolo di studio (X)	Sesso (2) X Classe di età quinquennale 6 anni e più (15)	Titolo di studio (X)	30
P2.16	Popolazione residente di 6 anni e più	Regione	Sesso (3) X Classe di età quinquennale 6 anni e più (16) X Titolo di studio (160)	Sesso (2) X Classe di età quinquennale 6 anni e più (15) X Grado di istruzione (8) (modalità Short: da 5 a 12)	Titolo di studio (160)	240
P2.18	Popolazione residente di 6 anni e più che ha conseguito il titolo di studio più elevato all'estero	Regione	Sesso (3) X Grado di istruzione (8) X Area geografica di cittadinanza (21)	Sesso (2) X Grado di istruzione (5) X Area geografica di cittadinanza (15)		150
P2.19	Occupati	Regione	Sesso (3) X Attività economica (4) X Età per singolo anno da 15 anni e più (74) X Posiz. nella Prof. (7)	Sesso (2) X Età per singolo anno da 15 anni e più (61)	Attività economica (3) X Posizione nella professione (6)	122
P2.20	Occupati	Regione	Sesso (3) X Attività lavorativa svolta (11) X Grado di Istruz. (7)	Sesso (2) X Grado di istruzione (6)	Attività lavorativa svolta (10)	12
P2.21	Occupati	Regione	Sesso (3) X Attività lavorativa svolta (11) X Tipologia attività lavorativa (pieno-parziale) (3)	Sesso (2)	Attività lavorativa svolta (10) X Tipologia attività lavorativa (pieno-parziale) (2)	2
P2.22	Occupati	Regione	Sesso (3) X Classe di età 15 anni e più (12) X Tipologia attività lavorativa (pieno-parziale) (3)	Sesso (2) X Classe di età 15 anni e più (11)	Tipologia attività lavorativa (pieno-parziale) (2)	22
P2.23	Occupati	Regione	Sesso (3) X Età per singolo anno da 15 anni e più (74) X Sez. di Att. Economica (18)	Sesso (2) X Età per singolo anno da 15 anni e più (61)	Sezioni di attività economica (17)	122
P2.24	Occupati dipendenti	Regione	Sesso (3) X Grado di Istruz. (7) X Rapporto di lavoro (Tempo Indet.-determinato) (3)	Sesso (2) X Grado di istruzione (6)	Rapporto di lavoro (Tempo indeterminato-determinato) (2)	12

Fonte: Direzione centrale dei Censimenti generali (DCCG), Istat.

Prospetto 5.6 - Sistema di vincoli non ottimizzato per l'oggetto "popolazione". Tavole soggette a stima e rispettivi vincoli indotti per livello territoriale (Piano di diffusione 2011, versione 2.0 - gennaio 2014) (segue).

Tavola	Oggetto	Livello territoriale	Incroccio	Marginali esaustive (al netto di totali e sub-totali)	Marginali a stima (al netto di totali e sub-totali)	Numero di vincoli
P2.25	Popolazione residente in cerca di prima occupazione	Regione	Sesso (3) X Titolo di studio (160) X Classe di età 15 anni e più (7)	Sesso (2) X Classe di età 15 anni e più (6) X Grado di istruzione (8) (modalità dello Short: da 5 a 12)	Titolo di studio (160)	96
P2.27	Popolazione residente che si reca al luogo abituale di studio o di lavoro	Regione	Sesso (3) X Classe di età (10) X Mezzo utilizzato (10)	Sesso (2) X Classe di età (9)	Mezzo utilizzato (9)	18
P2.28	Popolazione residente che si reca al luogo abituale di studio o di lavoro	Regione	Sesso (3) X Classe di età (10) X Tempo impiegato in minuti (6)	Sesso (2) X Classe di età (9)	Tempo impiegato in minuti (5)	18
P2.32	Popolazione straniera residente nata all'estero	Regione	Sesso (3) X Anno di trasferimento (6) X Area geografica di cittadinanza (21)	Sesso (2) X Area geografica di cittadinanza (12)	Anno di trasferimento (5)	24
P2.33	Popolazione straniera residente nata all'estero	Regione	Sesso (3) X Anno di trasferimento (6) X Area geografica e paese di cittadinanza (97)	Sesso (2) X Area geografica e paese di cittadinanza (97)	Anno di trasferimento (5)	194
P2.1	Popolazione che è stata residente all'estero	Ripartizione/Regione	Sesso (3) X Area di residenza precedente (7) X Cittadinanza (3)	Sesso (2) X Cittadinanza (2)	Area di residenza precedente (6)	4
P2.2	Popolazione che è stata residente all'estero	Ripartizione/Regione	Sesso (3) X Anno di trasferimento (3)	Sesso (2)	Anno di trasferimento (2)	2
P2.3	Popolazione che è stata residente all'estero	Ripartizione/Regione	Sesso (3) X Classe di età (7)	Sesso (2) X Classe di età (6)		12
P1.1	Occupati	Ripartizione	Sesso (3) X Attività lavorativa svolta (11) X Posizione nella professione (8)	Sesso (2)	Attività lavorativa svolta (10) X Posizione nella professione (7)	2
P1.2	Occupati	Ripartizione	Sesso (3) X Attività lavorativa svolta (11) X Settori di attività economica (22)	Sesso (2)	Attività lavorativa svolta (10) X Settori di attività economica (21)	2
P1.3	Occupati	Ripartizione	Sesso (3) X Classe di età 15 anni e più (10) X Grado di istruzione (7) X Sezioni di attività economica (7)	Sesso (2) X Classe di età 15 anni e più (9) X Grado di istruzione (6)	Sezioni di attività economica (6)	108
P1.4	Occupati	Ripartizione	Sesso (3) X Ore di lavoro effettuate (7) X Posizione nella professione (4) X Settori di attività economica (22)	Sesso (2)	Ore di lavoro effettuate (6) X Posizione nella professione (3) X Settori di attività economica (21)	2
P1.5	Occupati	Ripartizione	Sesso (3) X Posizione nella professione (7) X Sezioni di attività economica (18) X Titolo di studio (160)	Sesso (2) X Grado di istruzione (8) (otto modalità dello Short: da 5 a 12)	Posizione nella professione (6) X Sezioni di attività economica (17) X Titolo di studio (160)	16
P1.6	Occupati	Ripartizione	Sesso (3) X Posizione nella professione (7) X Sezioni di attività economica (18) X Grado di istruzione S (11)	Sesso (2) X Grado di istruzione S (6)	Posizione nella professione (6) X Sezioni di attività economica (17)	12
P1.9	Popolazione straniera residente occupata	Ripartizione	Sesso (3) X Area geografica di cittadinanza (21) X Attività economica (4)	Sesso (2) X Area geografica di cittadinanza (21)	Attività economica (3)	22
P1.10	Popolazione straniera residente occupata	Ripartizione	Sesso (3) X Area geografica e paese di cittadinanza (127) X Attività economica (4)	Sesso (2) X Area geografica e paese di cittadinanza (127)	Attività economica (3)	194

Fonte: Direzione centrale dei Censimenti generali (DCCG), Istat.

Prospetto 5.7 - Sistema di vincoli ottimizzato per l'oggetto "popolazione" (Piano di diffusione 2011, versione 2.0 - gennaio 2014).

Tavola	Oggetto	Livello territoriale	Incrocio	Marginali esaustive (al netto di totali e sub-totali)	Marginali a stima (al netto di totali e sub-totali)	Numero di vincoli
P3.NEW ^(*)	Popolazione residente	Provincia		Sesso (2) X Età per singolo anno da 0 a 19 anni e classe di età quinquennale da 20 anni e più (32)		64
P3.31	Popolazione residente in famiglia in abitazione	Provincia	Numero di occupanti (7) X Numero di stanze (7) X Titolo di godimento (4)	Numero di occupanti (6) X Titolo di godimento (3)	Numero di stanze (6)	18
P3.50	Occupati	Provincia	Sesso (3) X Attività lavorativa svolta (11) X Classe di età 15 anni e più (14)	Sesso (2) X Classe di età da 15 anni e più (13)	Attività lavorativa svolta (10)	26
P3.52	Occupati	Provincia	Sesso (3) X Stato civile (7) X Attività lavorativa svolta (11)	Sesso (2) X Stato civile (6)	Attività lavorativa svolta (10)	12
P3.56	Occupati	Provincia	Sesso (3) X Grado di istruzione (8) X Tipologia attività lavorativa (pieno-parziale) (3)	Sesso (2) X Grado di istruzione (7)	Tipologia attività lavorativa (pieno-parziale) (2)	14
P3.68	Popolazione residente che si reca al luogo abituale di studio	Provincia	Sesso (3) X Luogo di destinazione (12) X Mezzo utilizzato (10) X Tempo impiegato in minuti (6)	Sesso (2) X Luogo di destinazione (5)	Mezzo utilizzato (9) X Tempo impiegato in minuti (5)	10
P3.69	Occupati che si recano al luogo abituale di lavoro	Provincia	Sesso (3) X Luogo di destinazione (12) X Mezzo utilizzato (10) X Tempo impiegato in minuti (6)	Sesso (2) X Luogo di destinazione (5)	Mezzo utilizzato (9) X Tempo impiegato in minuti (5)	10
P3.78	Popolazione straniera residente nata all'estero	Provincia	Sesso (3) X Anno di trasferimento (6) X Area geo. di cittadinanza - continenti (7)	Sesso (2) X Area geografica di cittadinanza (continenti) (6)	Anno di trasferimento (5)	12
P3.80	Popolazione straniera residente occupata	Provincia	Sesso (3) X Classe di età 15 anni e più (10) X Attività economica (4) X Posizione nella professione (4)	Sesso (2) X Classe di età 15 anni e più	Attività economica (3) X Posizione nella professione (3)	18
P2.7	Popolazione residente in famiglia in abitazione	Regione	Figura giuridica del proprietario (8) X Numero di famiglie (3)	Numero di famiglie (2)	Figura giuridica del proprietario (7)	2
P2.8	Popolazione residente in famiglia in abitazione	Regione	Numero di componenti (7) X Tipo di servizi (7)	Numero di componenti (6)	Tipo di servizi (7)	6
P2.9	Popolazione residente di 14 anni e più	Regione	Sesso (3) X Classe di età 14 anni e più (13) X Frequenza corsi di formazione (3) X Grado di istruzione (8)	Sesso (2) X Classe di età 14 anni e più (12) X Grado di istruzione (4) (le modalità 5 6 7 sono garantite dalla P2.16)	Frequenza corsi di formazione (2)	96
P2.11	Popolazione residente di 15 anni e più	Regione	Sesso (3) X Frequenza corsi di formazione (3) X Condizione (9)	Sesso (2) X Condizione (6)	Frequenza corsi di formazione (2)	12
P2.14	Popolazione residente che ha conseguito almeno la licenza di scuola media (o avviamento professionale)	Regione	Sesso (3) X Corso di formazione professionale regionale/provinciale della durata pari o superiore ai 24 mesi, a cui si accede con la licenza media (4)	Sesso (2) Grado di istruzione 4-5-6-7 di Grado di istruzione (8)	Corso di formazione professionale regionale/provinciale della durata pari o superiore ai 24 mesi, a cui si accede con la licenza media (3)	2
P2.16	Popolazione residente di 6 anni e più	Regione	Sesso (3) X Classe di età quinquennale da 6 anni e più (16) X Titolo di studio (160)	Sesso (2) X Classe di età quinquennale da 6 anni e più (15) X Grado di istruzione (8) (modalità dello Short: da 5 a 12)	Titolo di studio (160)	240

^(*) Tavola fittizia. La variabile "età" viene costruita in sede di definizione vincoli per inglobare modalità relative a tavole diverse.

Fonte: Direzione centrale dei Censimenti generali (DCCG), Istat.

Prospetto 5.7 - Sistema di vincoli ottimizzato per l'oggetto "popolazione" (Piano di diffusione 2011, versione 2.0 - gennaio 2014) (segue).

Tavola	Oggetto	Livello territoriale	Incrocio	Marginali esaustive (al netto di totali e sub-totali)	Marginali a stima (al netto di totali e sub-totali)	Numero di vincoli
P2.19	Occupati	Regione	Sesso (3) X Attività economica (4) X Et� per singolo anno da 15 anni e pi� (74) X Posizione nella professione (7)	Sesso (2) X Et� per singolo anno da 15 anni e pi� (61)	Attivit� economica (3) X Posizione nella professione (6)	122
P2.25	Popolazione residente in cerca di prima occupazione	Regione	Sesso (3) X Titolo di studio (160) X Classe di et� da 15 anni e pi� (7)	Sesso (2) X Classe di et� da 15 anni e pi� (6) X Grado di istruzione (8) (otto modalit� dello Short: da 5 a 12)	Titolo di studio (160)	96
P2.27	Popolazione residente che si reca al luogo abituale di studio o di lavoro	Regione	Sesso (3) X Classe di et� (10) X Mezzo utilizzato (10)	Sesso (2) X Classe di et� (9)	Mezzo utilizzato (9)	18
P2.32	Popolazione straniera residente nata all'estero	Regione	Sesso (3) X Anno di trasferimento (6) X Area geografica di cittadinanza (21)	Sesso (2) X Area geografica di cittadinanza (12)	Anno di trasferimento (5)	24
P2.1	Popolazione che � stata residente all'estero	Ripartizione/Regione	Sesso (3) X Area di residenza precedente (7) X Cittadinanza (3)	Sesso (2) X Cittadinanza (2)	Area di residenza precedente (6)	4
P1.5	Occupati	Ripartizione	Sesso (3) X Posizione nella professione (7) X Sezioni di attivit� economica (18) X Titolo di studio (160)	Sesso (2) X Grado di istruzione (8) (otto modalit� dello Short: da 5 a 12)	Posizione nella professione (6) X Sezioni di attivit� economica (17) X Titolo di studio (160)	16

Fonte: Direzione centrale dei Censimenti generali (DCCG), Istat.

Prospetto 5.8 - Sistema definitivo di vincoli per l'oggetto "popolazione": numero di tavole per livello territoriale e vincoli indotti (Piano di diffusione 2011, versione 2.0 - gennaio 2014).

Livello Territoriale	Numero di tavole a stima	Numero di vincoli indotti	Numero di tavole a stima ottimizzate	Numero di vincoli indotti
Provincia/Grande Comune	25	397	8 ^(**)	184 ^(**)
Regione	22	1234	10	618
Ripartizione	10	354	2	20

(**) Viene conteggiata anche una tavola fittizia con i rispettivi vincoli. Tale tavola viene costruita in sede di definizione dei vincoli per includere le modalit  relative a tavole diverse.

Fonte: Direzione centrale dei Censimenti generali (DCCG), Istat.

Prospetto 5.9 - Sistema definitivo di vincoli ottimizzato per l'oggetto "popolazione": tavole con sistema di "pesi ad hoc" per livello territoriale e vincoli indotti (Piano di diffusione 2011 ver. 2.0 gennaio 2014).

Tavola	Oggetto	Livello territoriale	Incrocio	Marginali esaustive (al netto di totali e sub-totali)	Marginali a stima (al netto di totali e sub-totali)	Numero di vincoli
P3.43	Popolazione residente di 6 anni e pi�	Provincia	Sesso (3) X Classe di et� quinquennale da 6 anni e pi� (16) X Grado di istruzione (19)	Sesso (2) X Classe di et� quinquennale da 6 anni e pi� (15) X Grado di istruzione (2 modalit�: Diploma 2,3 anni; Diploma 4,5 anni)	Grado di istruzione (7)	60
P1.10	Popolazione straniera residente occupata	Ripartizione	Sesso (3) X Area geografica e paese di cittadinanza (127) X Attivit� economica (4)	Sesso (2) X Area geografica e paese di cittadinanza (119)	Attivit� economica (3)	238

Fonte: Direzione centrale dei Censimenti generali (DCCG), Istat.

5.4.2 Oggetto “famiglie”

Le tavole sottoposte a stima dell’oggetto “famiglie”, in base al previsto Piano di diffusione nazionale, ammontavano a 14 (Prospetto 5.10); dopo il processo di ottimizzazione “orizzontale e verticale” (Prospetto 5.11), si è giunti al disegno finale del sistema composto da 20 vincoli: 9 per il livello territoriale di provincia/grande comune; 11 per il livello di Regione (Prospetto 5.12).

Prospetto 5.10 - Sistema di vincoli non ottimizzato per l’oggetto “famiglie”: tavole soggette a stima e rispettivi vincoli indotti per livello territoriale (Piano di diffusione 2011, versione 2.0 - gennaio 2014).

Tavola	Oggetto	Livello territoriale	Incroccio	Marginali esaustive (al netto di totali e sub-totali)	Marginali a stima (al netto di totali e sub-totali)	Numero di vincoli
F3.19	Famiglie in abitazione	Provincia	Disponibilità di servizi (11)	Tipo alloggio=1	Disponibilità di servizi (11)	1
F3.20	Famiglie in abitazione	Provincia	Numero di stanze (7) X Titolo di godimento (4)	Titolo di godimento (3) x Tipo alloggio=1	Numero di stanze (6)	3
F3.36	Famiglie in abitazione	Provincia	Automobile (4) X Posto auto (4)	Tipo alloggio=1	Automobile (3) X Posto auto (3)	1
F3.37	Famiglie in abitazione	Provincia	Posto auto (4) X Titolo di godimento (4)	Titolo di godimento (3) x Tipo alloggio=1	Posto auto (3)	3
F3.38	Famiglie in abitazione	Provincia	Servizi posseduti dalla famiglia (5) X Titolo di godimento (4)	Titolo di godimento (3) x Tipo alloggio=1	Servizi posseduti dalla famiglia (5)	3
F3.39	Famiglie in abitazione	Provincia	Telefono fisso (3) X Telefono cellulare (3)	Tipo alloggio=1	Telefono fisso (2) X Telefono cellulare (2)	1
F3.41	Famiglie in abitazione	Provincia	Numero di componenti (7) X Componenti in famiglia con cellulare (4)	Numero di componenti (6) x Tipo alloggio=1	Componenti in famiglia con cellulare (3)	6
F3.44	Famiglie in abitazione	Provincia	Connessione a internet (7)	Tipo alloggio=1	Connessione a internet (6)	1
F2.3	Famiglie in abitazione	Regione	Figura giuridica del proprietario (8) X Numero di famiglie (3)	Numero di famiglie (2) x Tipo alloggio=1	Figura giuridica del proprietario (7)	2
F2.4	Famiglie in abitazione	Regione	Numero di componenti (7) X Tipo di servizi (7)	Numero di componenti (6) x Tipo alloggio=1	Tipo di servizi (7)	6
F2.7	Famiglie unipersonali non in coabitazione in abitazione	Regione	Classe di età decennale (9) X Numero di stanze (7)	Classe di età decennale (8) x Tipo alloggio=1	Numero di stanze (6)	8
F2.10	Famiglie in altri tipi di alloggio	Regione	Automobile (4)	Tipo alloggio=2 or Tipo alloggio=3	Automobile (3)	1
F2.11	Famiglie in altri tipi di alloggio	Regione	Componenti in famiglia con telefono cellulare (4)	Tipo alloggio=2 or Tipo alloggio=3	Componenti in famiglia con telefono cellulare (3)	1
F2.12	Famiglie in altri tipi di alloggio	Regione	Connessione a internet (7)	Tipo alloggio=2 or Tipo alloggio=3	Connessione a internet (6)	1

Fonte: Direzione centrale dei Censimenti generali (DCCG), Istat.

Prospetto 5.11 - Sistema definitivo di vincoli ottimizzato per l'oggetto "famiglie" (Piano di diffusione 2011, versione 2.0 - gennaio 2014).

Tavola	Oggetto	Livello territoriale	Incrocio	Marginali esaustive (al netto di totali e sub-totali)	Marginali a stima (al netto di totali e sub-totali)	Numero di vincoli
F3.20	Famiglie in abitazione	Provincia	Numero di stanze (7) X Titolo di godimento (4)	Titolo di godimento (3) x Tipo alloggio=1	Numero di stanze (6)	3
F3.41	Famiglie in abitazione	Provincia	Numero di componenti (7) X Componenti in famiglia con cellulare (4)	Numero di componenti (6) x Tipo alloggio=1	Componenti in famiglia con cellulare (3)	6
F2.3	Famiglie in abitazione	Regione	Figura giuridica del proprietario (8) X Numero di famiglie (3)	Numero di famiglie (2) x Tipo alloggio=1	Figura giuridica del proprietario (7)	2
F2.7	Famiglie unipersonali non in coabitazione in abitazione	Regione	Classe di età decennale (9) X Numero di stanze (7)	Classe di età decennale (8) x Tipo alloggio=1	Numero di stanze (6)	8
F2.10	Famiglie in altri tipi di alloggio	Regione	Automobile (4)	Tipo alloggio=2 or Tipo alloggio=3	Automobile (3)	1

Fonte: Direzione centrale dei Censimenti generali (DCCG), Istat.

Prospetto 5.12 - Sistema definitivo dei vincoli per l'oggetto "famiglie": numero di tavole per livello territoriale e vincoli indotti (Piano di diffusione 2011, versione 2.0 - gennaio 2014).

Oggetto	Livello Territoriale	Numero di tavole a stima	Numero di vincoli indotti	Numero di tavole a stima ottimizzate	Numero di vincoli indotti
Famiglie	Provincia/Grande Comune	8	19	2	9
	Regione	6	19	3	11

Fonte: Direzione centrale dei Censimenti generali (DCCG), Istat.

5.4.3 Oggetto "alloggi ed edifici"

Le tavole dell'oggetto "alloggi ed edifici" che incrociano variabili da stimare, in base al Piano di diffusione nazionale, erano 34 (Prospetto 5.13); queste tavole, dopo il processo di ottimizzazione "orizzontale e verticale" (Prospetto 5.14), hanno condotto ad un sistema finale di 24 vincoli: 2 per il livello comunale; 20 per il livello di provincia/grande comune; 2 per il livello di Regione (Prospetto 5.15).

Prospetto 5.13 - Sistema di vincoli non ottimizzato per l'oggetto "alloggi ed edifici": tavole soggette a stima e rispettivi vincoli indotti per livello territoriale (Piano di diffusione 2011, versione 2.0 - gennaio 2014).

Tavola	Oggetto	Livello territoriale	Incrocio	Marginali esaustive (al netto di totali e sub-totali)	Marginali a stima (al netto di totali e sub-totali)	Numero di vincoli
A4.3	Abitazioni occupate da persone residenti	Comune	Disponibilità di servizi (11)	Numero abitazioni occupate per Comune	Disponibilità di servizi (11)	1
A4.4	Abitazioni occupate da persone residenti	Comune	Numero di gabinetti (3)	Numero abitazioni occupate per Comune	Numero di gabinetti (2)	1
A4.5	Abitazioni occupate da persone residenti	Comune	Numero di impianti doccia e vasche da bagno (3)	Numero abitazioni occupate per Comune	Numero di impianti doccia e vasche da bagno (2)	1
A4.6	Abitazioni occupate da persone residenti	Comune	Numero di stanze (7)	Numero abitazioni occupate per Comune	Numero di stanze (6)	1
A4.17	Numero di occupanti per stanza in abitazioni occupate da persone residenti	Comune	Numero di occupanti per stanza in abitazioni occupate da persone residenti	Numero di occupanti per Comune	Numero di occupanti per stanza in abitazioni occupate da persone residenti	1
A3.5	Abitazioni occupate da persone residenti	Provincia	Classi di superficie (mq) (10) X Disponibilità e numero di cucine (10)	Classi di superficie (mq) (9)	Disponibilità e numero di cucine (9)	9
A3.6	Abitazioni occupate da persone residenti	Provincia	Classi di superficie (mq) (10) X Numero di gabinetti (3)	Classi di superficie (mq) (9)	Numero di gabinetti (2)	9
A3.7	Abitazioni occupate da persone residenti	Provincia	Classi di sup. (mq) (10) X N. di impianti doccia e vasche da bagno (3)	Classi di superficie (mq) (9)	Numero di impianti doccia e vasche da bagno (2)	9
A3.8	Abitazioni occupate da persone residenti	Provincia	Classi di superficie (mq) (10) X Numero di stanze (7)	Classi di superficie (mq) (9)	Numero di stanze (6)	9
A3.9	Abitazioni occupate da persone residenti	Provincia	Numero di stanze (7) X Numero di stanze ad uso professionale (6)	Numero abitazioni occupate per Comune	Numero di stanze (6) X Numero di stanze ad uso professionale (4)	1
A3.11	Abitazioni occupate da persone residenti	Provincia	Figura giuridica del proprietario (8)	Numero abitazioni occupate per Comune	Figura giuridica del proprietario (7)	1
A3.12	Abitazioni occupate da persone residenti	Provincia	Numero di occupanti (7) X Numero di stanze (7)	Numero di occupanti (6)	Numero di stanze (6)	6
A3.13	Abitazioni occupate da persone residenti	Provincia	Servizi presenti nell'abitazione (8)	Numero abitazioni occupate per Comune	Servizi presenti nell'abitazione (8)	1
A3.14	Abitazioni occupate da persone residenti con impianto di riscaldamento	Provincia	Tipo di combustibile o energia che alimenta l'impianto di riscaldamento (8)	Numero abitazioni occupate per Comune	Tipo di combustibile o energia che alimenta l'impianto di riscaldamento (7)	1
A3.15	Abitazioni occupate da persone residenti con impianto di riscaldamento	Provincia	Tipo di impianto di riscaldamento (5)	Numero abitazioni occupate per Comune	Tipo di impianto di riscaldamento (4)	1
A3.18	Abitazioni occupate da persone residenti in edifici residenziali	Provincia	Servizi installati (6)	Edifici residenziali	Servizi installati (6)	1
A3.36	Numero di occupanti per stanza in abitazioni occupate da persone residenti	Provincia	Anno di Censimento (5)	Numero abitazioni occupate per Comune		1
A3.37	Numero di occupanti per stanza in abitazioni occupate da persone residenti in affitto	Provincia	Anno di Censimento (5)	Numero di occupanti X Titolo di godimento (2 affitto)		1

Fonte: Direzione centrale dei Censimenti generali (DCCG), Istat.

Prospetto 5.13 - Sistema di vincoli non ottimizzato per l'oggetto "alloggi ed edifici": tavole soggette a stima e rispettivi vincoli indotti per livello territoriale (Piano di diffusione 2011, versione 2.0 - gennaio 2014) (segue).

Tavola	Oggetto	Livello territoriale	Incrocio	Marginali esaustive (al netto di totali e sub-totali)	Marginali a stima (al netto di totali e sub-totali)	Numero di vincoli
A3.38	Numero di occupanti per stanza in abitazioni occupate da persone residenti in proprietà	Provincia	Anno di Censimento (5)	Numero di occupanti X Titolo di godimento (1 proprietà)		1
A3.39	Numero di stanze per abitazione occupata da persone residenti	Provincia	Anno di Censimento (5)	Numero abitazioni occupate per Comune		1
A3.40	Numero di stanze per abitazione occupata da persone residenti in affitto	Provincia	Anno di Censimento (5)	Titolo di godimento (2 affitto)		1
A3.41	Numero di stanze per abitazione occupata da persone residenti in proprietà	Provincia	Anno di Censimento (5)	Titolo di godimento (1 proprietà)		1
A3.49	Stanze ad uso professionale in abitazioni occupate da persone residenti	Provincia	Numero di stanze (7)	Numero abitazioni occupate per Comune	Numero di stanze (6)	1
A3.50	Stanze in abitazioni occupate da persone residenti	Provincia	Anno di Censimento (5)	Numero abitazioni occupate per Comune		1
A3.51	Stanze in abitazioni occupate da persone residenti	Provincia	Classi di superficie (mq) (10) X Numero di stanze (7)	Classi di superficie (mq) (9)	Numero di stanze (6)	9
A3.52	Stanze in abitazioni occupate da persone residenti	Provincia	Numero di stanze (7) X Numero di stanze ad uso professionale (6)	Numero abitazioni occupate per Comune	Numero di stanze (6) X Numero di stanze ad uso professionale (4)	1
A3.54	Stanze in abitazioni occupate da persone residenti in edifici residenziali	Provincia	Numero di stanze (7)	Numero abitazioni occupate per Comune	Numero di stanze (6)	1
A2.1	Abitazioni occupate da persone residenti	Regione	Figura giuridica del proprietario (8) X Numero di stanze (7)	Numero abitazioni occupate per Comune	Figura giuridica del proprietario (7) X Numero di stanze (6)	1
A2.2	Abitazioni occupate da persone residenti	Regione	Numero di famiglie (3) X Numero di stanze (7)	Numero di famiglie (2)	Numero di stanze (7)	2
A2.3	Abitazioni occupate da persone residenti	Regione	Numero di stanze (7) X Tipo di servizi (7)	Numero abitazioni occupate per Comune	Numero di stanze (6) X Tipo di servizi (7)	1
A2.4	Abitazioni occupate da persone residenti con impianto di riscaldamento	Regione	Disponibilità di acqua calda e impianto di riscaldamento (6) X Tipo di combustibile o energia che alimenta l'impianto di riscaldamento (8)	Numero abitazioni occupate per Comune	Disponibilità di acqua calda e impianto di riscaldamento (6) X Tipo di combustibile o energia che alimenta l'impianto di riscaldamento (7)	1
A2.5	Abitazioni occupate da persone residenti con impianto di riscaldamento	Regione	Tipo di combustibile o energia che alimenta l'impianto di riscaldamento (8) X Tipo di impianto di riscaldamento (5)	Numero abitazioni occupate per Comune	Tipo di combustibile o energia che alimenta l'impianto di riscaldamento (7) X Tipo di impianto di riscaldamento (4)	1
A2.8	Stanze in abitazioni occupate da persone residenti	Regione	Figura giuridica del proprietario (8) X Numero di famiglie (3) X Numero di stanze (7)	Numero di famiglie (2)	Figura giuridica del proprietario (7) X Numero di stanze (6)	2
A2.9	Stanze in abitazioni occupate da persone residenti	Regione	Numero di stanze (7) X Tipo di servizi (7)	Numero abitazioni occupate per Comune	Numero di stanze (6) X Tipo di servizi (7)	1

Fonte: Direzione centrale dei Censimenti generali (DCCG), Istat.

Prospetto 5.14 - Sistema definitivo di vincoli ottimizzato per l'oggetto "alloggi ed edifici" (Piano di diffusione 2011, versione 2.0 - gennaio 2014).

Tavola	Oggetto	Livello territoriale	Incrocio	Marginali esaustive (al netto di totali e sub-totali)	Marginali a stima (al netto di totali e sub-totali)	Numero di vincoli
A4.3	Abitazioni occupate da persone residenti	Comune	Disponibilità di servizi (11)	Numero abitazioni occupate per Comune	Disponibilità di servizi (11)	1
A4.17	Numero di occupanti per stanza in abitazioni occupate da persone residenti	Comune	Numero di occupanti per stanza in abitazioni occupate da persone residenti	Numero di occupanti per Comune	Numero di occupanti per stanza in abitazioni occupate da persone residenti	1
A3.5	Abitazioni occupate da persone residenti	Provincia	Classi di superficie (mq) (10) X Disponibilità e numero di cucine (10)	Classi di superficie (mq) (9)	Disponibilità e numero di cucine (9)	9
A3.12	Abitazioni occupate da persone residenti	Provincia	Numero di occupanti (7) X Numero di stanze (7)	Numero di occupanti (6)	Numero di stanze (6)	6
A3.18	Abitazioni occupate da persone residenti in edifici residenziali	Provincia	Servizi installati (6)	Edifici residenziali	Servizi installati (6)	1
A3.37	Numero di occupanti per stanza in abitazioni occupate da persone residenti in affitto	Provincia	Anno di Censimento (5)	Numero di occupanti X Titolo di godimento (2 affitto)		1
A3.38	Numero di occupanti per stanza in abitazioni occupate da persone residenti in proprietà	Provincia	Anno di Censimento (5)	Numero di occupanti X Titolo di godimento (1 proprietà)		1
A3.40	Numero di stanze per abitazione occupata da persone residenti in affitto	Provincia	Anno di Censimento (5)	Titolo di godimento (2 affitto)		1
A3.41	Numero di stanze per abitazione occupata da persone residenti in proprietà	Provincia	Anno di Censimento (5)	Titolo di godimento (1 proprietà)		1
A2.2	Abitazioni occupate da persone residenti	Regione	Numero di famiglie (3) X Numero di stanze (7)	Numero di famiglie (2)	Numero di stanze (7)	2

Fonte: Direzione centrale dei Censimenti generali (DCCG), Istat.

Prospetto 5.15 - Sistema definitivo di vincoli per l'oggetto "alloggi ed edifici": numero di tavole per livello territoriale e vincoli indotti (Piano di diffusione 2011, versione 2.0 - gennaio 2014).

Oggetto	Livello territoriale	Numero di tavole a stima	Numero di vincoli indotti	Numero di tavole a stima ottimizzate	Numero di vincoli indotti
Alloggi ed edifici	Comune	5	5	2	2
	Provincia/Grande Comune	22	67	7	20
	Regione	7	9	1	2

Fonte: Direzione centrale dei Censimenti generali (DCCG), Istat.

5.5 Il calcolo dei pesi di riporto all'universo

5.5.1 Introduzione

La scelta di determinare un sistema di pesi di riporto all'universo distinto per ciascuno dei tre universi oggetto di rilevazione al censimento ("popolazione", "famiglie", "alloggi ed edifici"), associato ad una revisione delle tavole di diffusione previste per il 2011, ha richiesto dapprima il disegno del sistema di vincoli di calibrazione e poi l'impiego di una opportuna procedura matematica implementata nel software *Genesees v3.0*, sviluppato in Istat per facilitare i calcoli richiesti dalla riponderazione (Pagliuca, 2005).

La struttura di vincoli è stata indotta dalle tavole a stima stabilite per la diffusione nazionale dei risultati del Censimento della popolazione e delle abitazioni del 2011. Il sistema di pesi finali è stato ottenuto come soluzione di un problema di ottimo vincolato che minimizza la differenza tra i pesi iniziali (quelli da disegno) e i pesi finali, sotto l'insieme di vincoli precedentemente individuati.

L'impiego dei pesi finali consente da un lato di migliorare la rappresentatività del campione e dall'altro di aumentare l'accuratezza delle stime finali ottenute tramite con lo stimatore di ponderazione vincolata. Infatti, il principio su cui è basato ogni metodo di stima diretta (cfr. paragrafo 5.2) è che il peso di riporto all'universo esprima il peso che assume ogni unità statistica campionata su cui è stato osservato il dato, rispetto alle unità della popolazione che non sono incluse nel campione; quindi il peso attribuito ad ogni unità campionaria indica il numero di unità della popolazione rappresentate dall'unità medesima.

5.5.2 La preparazione dei dati di input

La procedura di riponderazione ha richiesto, in via preliminare, per ciascun dominio di calibrazione¹⁰⁸, la definizione di due *dataset di input*: "totali noti" e "dati campionari". Il primo riporta, per i domini riferiti al sistema di vincoli, i totali delle variabili *short* impiegate nella procedura di calibrazione; il secondo contiene i dati riferiti alle stesse variabili osservate sulle unità campionarie.

Il *dataset* dei totali noti è stato ottenuto calcolando le frequenze assolute delle variabili ausiliarie su tutte le unità del dominio. Tali frequenze si riferiscono alle celle di calibrazione indotte dal sistema di vincoli richiesto dalle tavole a stima.

Per la costruzione del *dataset* dei dati campionari (record di osservazioni rilevate tramite il modello *long*) è stato necessario selezionare, per ogni dominio di calibrazione:

- 1) i Comuni sottoposti a campionamento;
- 2) le aree di censimento di centro abitato (Ace) campionate dei Comuni a campione;
- 3) le famiglie, all'interno delle Ace, selezionate per il campione a cui è stato assegnato il modello *long* per la rilevazione censuaria.

Si ricorda che il modello *long* è stato assegnato, a fini "non campionari" a tutte le famiglie non eleggibili al campionamento appartenenti ai seguenti domini: i Comuni non campionati; le sezioni di censimento delle Ace non campionate; le sezioni di censimento escluse dal disegno delle Ace dei Comuni a campione.

Inoltre, nei domini campionati, a seconda dell'universo di riferimento, non tutte le unità sono considerate eleggibili per il calcolo dei pesi finali. La variabile discriminante è il "Tipo di alloggio"¹⁰⁹. Si considerano eleggibili, per l'universo "popolazione", gli individui di tutte le famiglie, indipendentemente dalla tipologia di alloggio; per l'universo delle "famiglie" le unità eleggibili sono solo quelle residenti in "Abitazione", "Altro tipo di alloggio" o "Alloggio presso sede diplomatica o consolare"; mentre per "alloggi ed edifici" si considera solo la prima modalità "Abitazione". Le unità degli universi "famiglie" e "alloggi ed edifici", ritenute non eleggibili (a cui si attribuisce il questionario di tipo *long*), sono considerate "auto-rappresentative" per cui, nel sistema di pesi finale, si assegna un peso pari ad "1".

La definizione del *dataset* campionario finale da sottoporre alla procedura di riponderazione in *Genesees* ha richiesto ulteriori operazioni finalizzate a valutare la presenza e, nel caso, a risolvere alcune criticità riscontrate nella fase di sperimentazione.

¹⁰⁸ I "domini di calibrazione" sono i domini territoriali di riferimento per l'insieme di vincoli e per il data base campionario utile al calcolo dei pesi finali; questi ultimi, una volta prodotti, saranno riferibili solo alle unità campionarie di quel dominio.

¹⁰⁹ La variabile "Tipo di alloggio" (codificata con TIPALL) è rilevata, al censimento del 2011, tramite la domanda 1.1 della SEZIONE I – Notizie su famiglia e alloggio e unità eleggibili. Le modalità possibili di classificazione sono quattro: 1 – "Abitazione"; 2 – "Altro tipo di alloggio"; 3 – "Alloggio presso sede diplomatica o consolare"; 4 – "Struttura residenziale collettiva".

La prima ha riguardato la possibile presenza di zeri campionari non strutturali¹¹⁰ in alcuni domini di calibrazione. Tale criticità è stata risolta con il ricorso alla “tecnica della donazione”. A tale scopo sono stati creati opportuni serbatoi di “donatori” costituiti da unità non appartenenti al campione rilevate con il modello *short*. In presenza di uno zero campionario per un certo vincolo (incrocio) e un dato dominio, la procedura di donazione prevedeva l'estrazione casuale, dal serbatoio dei donatori, di un prefissato numero di unità¹¹¹ appartenenti allo stesso dominio e con le caratteristiche *short* riferibili all'incrocio per il quale si è osservato lo zero campionario. Tali unità, a cui saranno in seguito imputati¹¹² i valori mancanti delle variabili proprie del modello *long*, vengono di fatto integrate nel campione al fine di rappresentare i casi presenti nell'universo ma non incontrati nella rilevazione.

Si supponga, per esempio, per l'incrocio-vincolo “femmine occupate con età 16 anni” della generica Provincia K di aver osservato il totale pari a 12, dai dati *short* e *long*, e pari a 0 sul campione (zero campionario) dai dati *long*. Allora, per popolare tale incrocio campionario “vuoto” si procede dapprima ad estrarre, dal serbatoio delle famiglie della Provincia K rilevate tramite il modello *short* in cui c'è la presenza di “femmine occupate con età 16 anni”, 4 unità (il 33 per cento di 12) come se fossero state precedentemente estratte per il campione e, successivamente, a imputare i dati delle variabili *long* mancanti (trasformazione da *short* a *long*).

Un'altra criticità è derivata dalle incoerenze nel sistema di vincoli per la presenza, per uno stesso incrocio-vincolo, di totali identici riferiti a domini geograficamente innestati. Un esempio è dato dal caso di coincidenza tra il totale campionario di Provincia e quello del grande Comune ad essa appartenente, a fronte di una differenza (ovviamente a vantaggio della Provincia) tra i corrispondenti totali noti. Per risolvere questa incoerenza tra i vincoli che pone problemi alla convergenza dell'algoritmo per la determinazione della soluzione ottima, si è fatto ancora ricorso alla tecnica della donazione. Nella fase di esercizio si è proceduto includendo nel campione una sola unità estratta casualmente dal serbatoio dei donatori, con la caratteristica riferita all'incrocio che generava l'incoerenza, selezionata all'interno della Provincia, ma non appartenente al grande Comune.

Infine, la tecnica della donazione è stata impiegata anche nel caso in cui la procedura di riponderazione non giungeva a convergenza per la presenza di totali campionari troppo esigui rispetto ai corrispondenti totali noti (per esempio: totale campionario pari a 1 e totale noto pari a 20).

Una volta risolte le problematiche descritte sopra si è proceduto prima al calcolo del peso base per ciascuna unità campionaria (inverso della probabilità di inclusione dell'unità nel campione) e poi alla definizione dei *dataset* dei “Totali noti” e dei “Dati campionari”, comprensivi delle unità selezionate per la donazione, da sottoporre al software *Genesees* per il calcolo dei pesi finali di calibrazione.

Il risultato della riponderazione ha fornito i pesi di riporto all'universo finali; questi sono dati dal prodotto tra i pesi base e dei fattori correttivi calcolati in modo da soddisfare la condizione di uguaglianza tra i valori noti sulla popolazione, riferiti al sistema di vincoli e le corrispondenti stime campionarie ottenute con i pesi calibrati. I fattori correttivi sono quindi dati dalla soluzione di un problema di minimo vincolato in cui la funzione da minimizzare è la distanza tra i pesi base e i pesi finali (cfr. paragrafo 5.2). La funzione di distanza utilizzata nel processo di calibrazione è stata la “logaritmica troncata”, sulla base della quale si individua l'insieme dei pesi finali che da un lato soddisfa il sistema di vincoli e dall'altro modifica il meno possibile l'insieme dei pesi diretti.

Al fine di avere garanzie sulla convergenza dell'algoritmo di riponderazione, nella procedura in *Genesees* è stato ampliato l'intervallo di variazione del peso finale tramite la modifica dei valori minimo e massimo dei moltiplicatori preimpostati per default¹¹³. Per l'universo “popolazione”, nonostante la modifica fatta al software, la convergenza dell'algoritmo di ponderazione non è stata raggiunta per la Provincia di Trento e per la Regione Basilicata. Su tali domini si è proceduto con la “distanza logaritmica” che non pone un limite superiore al valore dei pesi finali.

¹¹⁰ Gli “zeri campionari” sono valori pari a zero riscontrati sul campione per alcun incroci delle modalità delle variabili osservate. Questi si dicono “strutturali” se corrispondono a incroci di cui non si sono registrati casi nell'universo (a volte perché impossibile: per esempio, “persona di 0-9 anni in possesso del diploma di laurea”); si dicono “non strutturali” se fanno riferimento incroci di cui si è osservato almeno un caso nell'universo. Quindi, gli “zeri campionari non strutturali” sono riferiti a casi esistenti, generalmente con bassa frequenza, nella popolazione ma non rilevati nel campione.

¹¹¹ Il numero di unità da estrarre era pari al prodotto tra il totale noto riferito al vincolo e la frazione di campionamento (0,33).

¹¹² L'imputazione è avvenuta tramite le stesse tecniche impiegate nella fase di “controllo e correzione” dei dati del Censimento 2011.

¹¹³ La procedura implementata in *Genesees* v.3.0 prevede, per i moltiplicatori, un campo di variazione iniziale pari a 0,5-1,5. Dai risultati delle sperimentazioni si è verificato che la convergenza dell'algoritmo è quasi sempre garantita con un ampliamento del *range* a 0,0005-5.

In alcuni casi si è fatto ricorso anche alla “distanza euclidea”, ma solo a fini diagnostici per valutare l’eventuale presenza di incompatibilità nel sistema di vincoli e quindi poter procedere con azioni preliminari in fase di definizione dei *dataset* di input.

5.5.3 Domini di riponderazione

5.5.3.1 Oggetto “popolazione”

Come descritto nel paragrafo 5.4.1, il sistema definitivo di vincoli per l’oggetto “popolazione” ha previsto 184 vincoli a livello di provincia/grande comune, 618 a livello regionale e 20 a livello di ripartizione geografica (Prospetto 5.8). Al fine di operare la riponderazione su scala regionale, riportando i vincoli di ripartizione a livello di Regione, il numero totale di vincoli definiti a tale livello territoriale è salito a 638.

Il Prospetto 5.16 riporta il numero di vincoli definiti per ogni dominio regionale; in tre Regioni (Lombardia, Emilia-Romagna, Toscana) poiché il numero di vincoli superava il limite massimo impostato¹¹⁴ pari a 3 mila, si è proceduto ad una opportuna suddivisione in due blocchi di Province.

La ripartizione in due domini sub-regionali è stata effettuata cercando di determinare un numero equivalente di record individuali nei due sotto-domini. La Regione Lombardia è stata ripartita in due blocchi: il primo ha previsto solo la Provincia di Milano; il secondo ha considerato tutte le altre Province. L’Emilia-Romagna è stata suddivisa in: Emilia-Romagna_1, comprendente le Province di Piacenza, Bologna, Ferrara, Rimini; Emilia-Romagna_2, contenente le Province di Parma, Reggio Emilia, Modena, Forlì, Ravenna. La Regione Toscana è stata ripartita in: Toscana_1 comprendente le Province di Pistoia, Firenze, Arezzo, Siena, Grosseto; Toscana_2 contenente le Province di Lucca, Livorno, Pisa, Prato, Massa-Carrara.

Il calcolo dei pesi per la Regione Trentino-Alto Adige è stato effettuato separatamente per le Province autonome di Bolzano e Trento in quanto il Piano di diffusione prevedeva per tali Province le stesse tavole disegnate per i domini regionali.

Prospetto 5.16 – Numero di vincoli dei domini di riponderazione per il calcolo dei pesi di riporto all’universo finali riferiti all’oggetto “popolazione”.

Dominio regionale/sub-regionale	Numero di province	Numero di grandi comuni	Numero di vincoli
Piemonte	8	1	2.294
Valle d’Aosta	1	0	822
Prov. di Milano	1	1	1.006
Lombardia (esclusa la Prov. di Milano)	11	1	2.846
Prov. Aut. di Bolzano	1	0	822
Prov. Aut. di Trento	1	0	822
Veneto	7	3	2.478
Friuli-Venezia Giulia	4	1	1.558
Liguria	4	1	1.558
Emilia-Romagna_1	4	1	1.558
Emilia-Romagna_2	5	4	2.294
Toscana_1	5	1	1.742
Toscana_2	5	2	1.926
Umbria	2	1	1.190
Marche	5	0	1.558
Lazio	5	1	1.742
Abruzzo	4	0	1.374
Molise	2	0	1.006
Campania	5	1	1.742
Puglia	6	2	2.110
Basilicata	2	0	1.006
Calabria	5	1	1.742
Sicilia	9	3	2.846
Sardegna	7	0	1.926

Fonte: Direzione centrale dei Censimenti generali (DCCG), Istat.

¹¹⁴ Gli esperti informatici hanno implementato le procedure in modo da portare fino a 3 mila il numero massimo di vincoli gestibili dal software *Genesees* v.3.0; oltre non era possibile per limiti computazionali.

Come esposto nel paragrafo 5.4.1, sono stati elaborati dei “*pesi ad hoc*” per le tavole [P1.9], [P1.10], e [P3.43]. A riguardo, il sistema dei vincoli ha previsto 60 incroci a livello di provincia/grande comune e 238 a livello di ripartizione geografica (Prospetto 5.17). Questo è stato l’unico caso in cui è stato possibile impiegare la procedura di riponderazione su domini di ripartizione in quanto il numero di vincoli rientrava nei limiti previsti dal software *Genesees v3.0* (Pagliuca, 2005).

Prospetto 5.17 – Numero di vincoli dei domini di riponderazione per il calcolo dei pesi di riporto all’universo *ad hoc* riferiti all’oggetto “popolazione”.

Dominio di ripartizione	Numero di province	Numero di grandi comuni	Numero di vincoli
Nord-Ovest	25	4	1.978
Nord-Est	22	9	2.098
Centro	22	5	1.858
Sud	20	4	1.678
Isole	16	3	1.378

Fonte: Direzione centrale dei Censimenti generali (DCCG), Istat.

5.5.3.2 Oggetto “famiglie”

Come descritto nel paragrafo 5.4.2, il sistema definitivo dei vincoli per l’oggetto “famiglie” ha previsto 9 vincoli a livello di provincia/grande comune e 11 vincoli a livello di Regione. In questo caso, dato l’esiguo numero di vincoli, è stato possibile operare la riponderazione su domini coincidenti con le Regioni, tranne che per le Province autonome di Trento e Bolzano trattate separatamente. Il Prospetto 5.18 riporta il numero dei vincoli definito per ogni dominio di riponderazione.

Prospetto 5.18 – Numero di vincoli dei domini di riponderazione per il calcolo dei pesi di riporto all’universo finali riferiti all’oggetto “famiglie”.

Dominio regionale/sub-regionale	Numero di province	Numero di grandi comuni	Numero di vincoli
Piemonte	8	1	92
Valle d’Aosta	1	0	20
Lombardia	12	2	137
Prov. Aut. di Bolzano	1	0	20
Prov. Aut. di Trento	1	0	20
Veneto	7	3	101
Friuli-Venezia Giulia	4	1	56
Liguria	4	1	56
Emilia-Romagna	9	5	137
Toscana	10	3	128
Umbria	2	1	38
Marche	5	0	56
Lazio	5	1	65
Abruzzo	4	0	47
Molise	2	0	29
Campania	5	1	65
Puglia	6	2	83
Basilicata	2	0	29
Calabria	5	1	65
Sicilia	9	3	119
Sardegna	7	0	74

Fonte: Direzione centrale dei Censimenti generali (DCCG), Istat.

5.5.3.3 Oggetto “alloggi ed edifici”

Il disegno definitivo del sistema di vincoli per l’oggetto “alloggi ed edifici” (cfr. paragrafo 5.4.3) ha richiesto l’inclusione di 2 vincoli a livello di Comune (solo per i Comuni campionati), 20 a livello di provincia/grande comune e 2 a livello di Regione. Anche in questo caso, dato il ridotto numero di vincoli, è stato possibile effettuare il calcolo dei pesi finali su domini coincidenti con le Regioni, sempre con l’eccezione delle province autonome di Trento e Bolzano lavorate separatamente. Nel Prospetto 5.19 viene riportato il numero di vincoli disegnato per ogni dominio di riponderazione.

Prospetto 5.19 – Numero di vincoli dei domini di riponderazione per il calcolo dei pesi di riporto all’universo finali riferiti all’oggetto “alloggi ed edifici”.

Dominio regionale/sub-regionale	Numero di province	Numero di grandi comuni	Numero di comuni a campione	Numero di vincoli
Piemonte	8	1	32	246
Valle d’Aosta	1	0	1	24
Lombardia	12	2	66	414
Prov. Aut. di Bolzano	1	0	3	28
Prov. Aut. di Trento	1	0	2	26
Veneto	7	3	32	266
Friuli-Venezia Giulia	4	1	5	112
Liguria	4	1	10	122
Emilia-Romagna	9	5	31	344
Toscana	10	3	33	328
Umbria	2	1	7	76
Marche	5	0	16	134
Lazio*	5	1	37	196
Abruzzo	4	0	12	106
Molise	2	0	3	48
Campania	5	1	63	248
Puglia	6	2	49	260
Basilicata	2	0	2	46
Calabria	5	1	10	142
Sicilia	9	3	56	354
Sardegna	7	0	16	174

Fonte: Direzione centrale dei Censimenti generali (DCCG), Istat.

5.5.4 La coerenza delle stime calibrate

I pesi finali ottenuti dalla procedura di riponderazione per calibrazione hanno garantito, in modo largamente diffuso sulle tavole a stima, la coerenza tra gli incroci riferiti alle variabili *short*, calcolate come somma dei pesi finali, e i valori relativi agli stessi incroci derivanti però da un conteggio esaustivo. In rari casi si è verificata un’incoerenza (generalmente nell’ordine di 1 unità) tra i suddetti valori, essenzialmente dovuta ad una mancata convergenza dell’algoritmo di riponderazione.

Per l’oggetto “popolazione” sono state osservate, per alcuni domini, delle differenze (sempre pari a +/- 1) tra il dato stimato (somma dei pesi finali) e il dato ottenuto come conteggio dei record.

I Prospetti 5.20 (per i pesi finali) e 5.21 (per i “*pesi ad hoc*”) descrivono gli incroci, riferiti all’oggetto “popolazione”, per i quali sono state osservate delle incoerenze. Prendendo ad esempio il primo riportato nel Prospetto 5.20: per la Regione Valle d’Aosta, il totale stimato relativo all’incrocio *short* “femmine, in cerca di prima occupazione, di età 15-19 anni, con diploma di scuola secondaria superiore 2-3 anni” differisce di 1 unità (in meno) rispetto al totale dello stesso incrocio ottenuto come conteggio esaustivo sul dominio preso in considerazione.

Per l’oggetto “famiglie” la coerenza dei dati è stata verificata per tutte le tavole a stima. Invece, per l’oggetto “alloggi ed edifici” sono state riscontrate delle incoerenze (nell’ordine di +/- 1 e, in un caso, pari a 2), così come riportato nel Prospetto 5.22.

Prospetto 5.20 – Incoerenze riscontrate sull’oggetto “popolazione” con il sistema di pesi finali.

Regione	Provincia	Descrizione del vincolo	Totale noto	Totale stimato	Incoerenza	Dominio
Valle d'Aosta		totale femmine in cerca di prima occupazione 15-19 anni con diploma di scuola secondaria superiore 2-3 anni	3	2	-1	Regione
Piemonte	Verbano-Cusio-Ossola	totale maschi che si recano al luogo abituale di studio luogo di destinazione: estero	2	3	1	Provincia
	Verbano-Cusio-Ossola	totale femmine che si recano al luogo abituale di studio luogo di destinazione: estero	3	2	-1	Provincia
Lombardia		totale maschi in cerca di prima occupazione 15-19 anni con laurea triennale	1	2	1	Regione
Umbria		totale femmine 20-24 anni con AFAM II livello	3	2	-1	Regione
		totale femmine 65-69 anni con AFAM I livello	1	2	1	Regione
		totale femmine 70-74 anni con AFAM I livello	6	7	1	Regione
		totale femmine in cerca di prima occupazione 25-29 anni con AFAM I livello	1	0	-1	Regione
Marche		totale maschi 65-69 anni con AFAM I livello	1	2	1	Regione

Fonte: Direzione centrale dei Censimenti generali (DCCG), Istat.

Prospetto 5.21 – Incoerenze riscontrate sull’oggetto “popolazione” con il sistema dei “pesi ad hoc”.

Ripartizione	Descrizione del vincolo	Totale noto	Totale stimato	Incoerenza	Dominio
Sud	totale femmine straniere residenti occupate paese di cittadinanza: CIPRO	1	2	1	Ripartizione
Sud	totale maschi stranieri residenti occupati paese di cittadinanza: MALTA	2	3	1	Ripartizione
Sud	totale femmine straniere residenti occupate paese di cittadinanza: MALTA	3	2	-1	Ripartizione

Fonte: Direzione centrale dei Censimenti generali (DCCG), Istat.

Prospetto 5.22 – Incoerenze riscontrate sull’oggetto “alloggi ed edifici” con il sistema di pesi finali.

Regione	Provincia	Comune	Descrizione del vincolo	Totale noto	Totale stimato	Incoerenza	Dominio
Abruzzo	Chieti		Numero di abitazioni occupate in edifici residenziali	52.612	52.613	1	provincia
Calabria	Catanzaro	Catanzaro	Numero di abitazioni occupate	29.371	29.370	-1	comune
	Catanzaro	Lamezia Terme	Numero di abitazioni occupate	18.009	18.008	-1	comune
	Catanzaro		Numero di abitazioni occupate in edifici residenziali	47.376	47.378	2	provincia
Friuli-Venezia Giulia	Pordenone		Numero di abitazioni occupate in edifici residenziali	20.608	20.609	1	provincia
Sicilia	Trapani		Numero di abitazioni occupate in edifici residenziali	100.787	100.788	1	provincia
Toscana	Firenze	Firenze	Numero di abitazioni occupate in edifici residenziali	146.751	146.752	1	grande comune

Fonte: Direzione centrale dei Censimenti generali (DCCG), Istat.

5.5.5 Le soluzioni informatiche

Al fine di rendere efficiente la gestione delle operazioni è stato utilizzato un approccio “a processo” suddiviso in tre fasi separate e consequenziali:

1. scarico dati;
2. preparazione *dataset*;
3. esecuzione della procedura di riponderazione in *Genesees*.

La fase di “scarico dati”, che aveva come obiettivo l'estrapolazione dei dati utili a preparare i *dataset* di input (cfr. paragrafo 5.5.2) con formato compatibile con il software *Genesees v3.0*, è stata eseguita attraverso una procedura PL/SQL.

La fase di “preparazione *dataset*” consisteva nell'estrazione delle famiglie campione e delle unità eleggibili alla riponderazione; la creazione dei *dataset* è avvenuta attraverso uno *script* eseguito in ambiente Sas e ha fornito come output i “*dataset*” stessi.

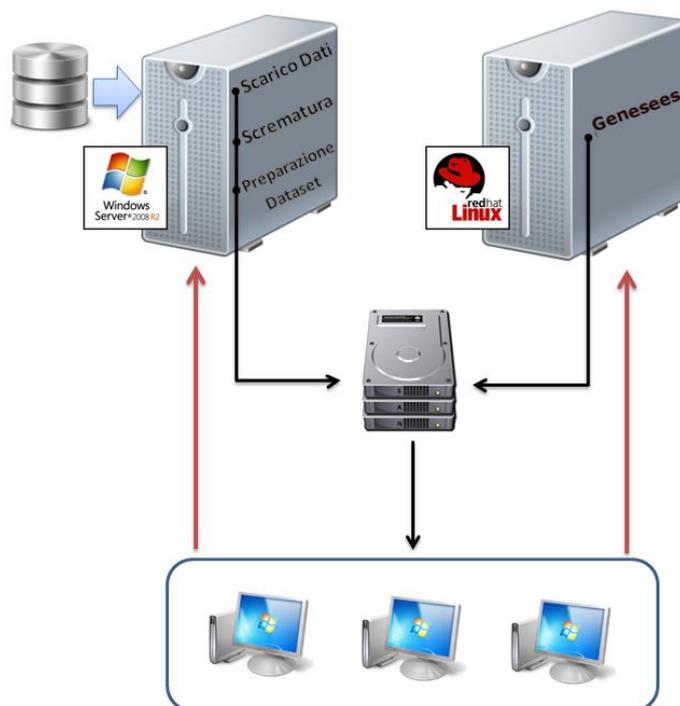
Nella fase 3 si è proceduto al calcolo dei pesi finali con il software *Genesees* e con l'impiego di procedure Sas.

Dopo aver monitorato l'impiego delle risorse hardware durante la sperimentazione delle varie fasi di lavoro, per l'esecuzione dell'intero processo si è deciso di seguire la seguente organizzazione tecnologica: la procedura PL/SQL per la fase di “scarico dati” e di “preparazione data set” è stata eseguita su un server Windows e i *dataset* estratti sono stati memorizzati su una SAN accessibile sia dai PC locali, sia dal server Linux che ospitava il software *Genesees v3.0*.

L'accesso diretto ai *dataset* ha permesso di eseguire più istanze di processamento al software *Genesees*, ognuna su un *dataset* differente riducendo, in tal modo, i tempi di esecuzione dell'intero processo.

Con una tale organizzazione del processo, è stato ridotto l'*overhead* (le risorse accessorie, richieste in sovrappiù rispetto a quelle strettamente necessarie) dovuto ai molteplici spostamenti di dati tra i vari server; inoltre, l'intera attività elaborativa è stata trasferita ai server così da liberare le risorse dei PC locali per le operazioni di routine. La Figura 5.1 illustra lo schema ottimizzato del processo informatico praticato.

Figura 5.1 – Processo informatico seguito per l'operazione di calcolo dei pesi di riporto all'universo.



Fonte: Direzione centrale dei Censimenti generali (DCCG), Istat.

5.5.6 I risultati del calcolo dei pesi finali

In questo paragrafo sono presentate alcune tabelle riepilogative (Prospetti 5.23 – 5.24 – 5.25) riferite alle distribuzioni dei pesi base (da disegno) e dei pesi finali (da calibrazione) calcolati per ciascuno degli universi presi in considerazione: “popolazione”, “famiglie”, “alloggi ed edifici”.

Le tabelle riportano, per entrambe le distribuzioni di pesi, i valori minimo e massimo, la mediana, i quartili e i principali percentili (il 1°, il 5°, il 10°, il 90°, il 95°, il 99°). Sulla base di tali valori è possibile fare valutazioni su: campo di variazione (*range*), concentrazione dei valori intorno alla mediana, consistenza numerica dei casi sulle code (più significativi per valutare l’effetto della calibrazione). Infatti, laddove i pesi finali rimangono il più vicino possibile ai pesi base la calibrazione risulta ottimale con vantaggi sull’efficienza delle stime (minor variabilità campionaria).

Si nota subito, come risultato comune, che il *range* dei valori dei pesi finali, calcolati sull’intero dominio di calibrazione, si amplia moltissimo rispetto a quello del peso base calcolato invece a livello di area di censimento (in quanto questo era il dominio di campionamento pianificato).

Prendendo ad esempio le distribuzioni dei pesi calcolati per l’oggetto “popolazione” (Prospetto 5.23) con riferimento al dominio “Piemonte”, il *range* dei pesi base passa da [2,6620; 3,1118] all’intervallo molto più ampio [0,0005; 22,3865] per i pesi finali. Con riferimento ai quartili, il 50 per cento dei valori intorno alla mediana passa dall’intervallo [2,8273; 2,9312] per la distribuzione dei pesi base a quello poco più ampio [2,6418; 3,0451] per la distribuzione dei pesi finali. Si fa inoltre notare, come dato particolare, che per i valori sulla coda di destra, l’1 per cento delle famiglie coinvolte nella riponderazione distribuiscono pesi finali ai rispettivi individui di appartenenza con valori compresi nell’intervallo [4,5698; 22,3865] in quanto, per costruzione, la riponderazione tende a sovra-rappresentare unità sotto-campionate riferite a sotto-classi (incroci di variabili/modalità) con frequenze piccole.

Prospetto 5.23 – Percentili delle distribuzioni dei pesi base e dei pesi finali per l’oggetto “popolazione”.

Dominio	Numero di unità	Pesi	Valori percentili										
			Min	p1	p5	p10	Q1	Me	Q3	p90	p95	p99	Max
Valle d'Aosta	5.018	Base	2,8495	2,8495	2,8495	2,8495	2,8495	2,9172	2,9172	2,9172	2,9172	2,9172	2,9172
		Finale	0,0005	0,9998	1,5599	1,9040	2,4113	2,8358	3,1964	3,8891	4,4757	6,1396	15,4937
Piemonte	296.309	Base	2,6620	2,6620	2,6823	2,7341	2,8273	2,8871	2,9312	2,9744	2,9960	3,0785	3,1118
		Finale	0,0005	1,6402	2,1643	2,3729	2,6418	2,8535	3,0451	3,3291	3,6336	4,5698	22,3865
Prov. Milano	321.382	Base	2,1304	2,4521	2,7521	2,8131	2,9069	2,9542	2,9934	3,0295	3,0631	3,0875	3,1079
		Finale	0,1861	1,7689	2,2157	2,4187	2,7056	2,9074	3,0724	3,3750	3,7196	4,7799	18,1818
Lombardia (senza Prov. Milano)	246.566	Base	2,4759	2,4759	2,7655	2,8151	2,8653	2,9325	3,0101	3,1302	3,1798	3,3405	3,3405
		Finale	0,2807	1,6678	2,2059	2,4211	2,6958	2,9191	3,1402	3,4443	3,7420	4,5777	15,7423
Veneto	221.020	Base	2,2760	2,2760	2,4631	2,5563	2,6480	2,8399	2,8981	2,9782	3,0582	3,1136	3,2774
		Finale	0,3518	1,6148	2,0850	2,2756	2,5324	2,7831	3,0040	3,2847	3,5632	4,3911	12,4748
Prov. Aut. Trento	17.385	Base	1,0545	1,0545	1,0545	1,0545	2,9517	3,0388	3,0503	3,1667	3,1667	3,1667	3,1667
		Finale	0,0005	0,8400	1,0406	1,2768	2,5395	2,9609	3,2348	3,5856	3,8586	5,1203	16,2310
Prov. Aut. Bolzano	23.144	Base	2,8032	2,8032	2,8032	2,8032	2,8493	2,8828	2,9139	2,9167	2,9708	2,9708	2,9708
		Finale	0,0005	1,4525	2,0329	2,2605	2,5520	2,8208	3,1211	3,5341	3,9457	5,0824	14,0000
Friuli	62.392	Base	2,8786	2,8786	2,8922	2,9032	2,9321	2,9599	2,9737	3,0073	3,0868	3,2048	3,2048
		Finale	0,3063	1,7198	2,3254	2,5357	2,7583	2,9301	3,1040	3,4215	3,7479	4,5385	11,1368
Liguria	145.405	Base	2,3994	2,3994	2,7046	2,7575	2,8561	2,9627	3,0274	3,1192	3,1888	3,2461	3,3799
		Finale	0,1957	1,5676	2,1503	2,3897	2,6850	2,9174	3,1415	3,4505	3,8051	4,9562	26,6844
E. Romagna 1	131.788	Base	2,5592	2,5592	2,7887	2,8174	2,9128	2,9867	3,0385	3,1699	3,2006	3,3246	3,3246
		Finale	0,3940	1,6994	2,2616	2,4827	2,7547	2,9667	3,1705	3,4431	3,7344	4,6353	20,9240
E. Romagna 2	150.269	Base	2,4287	2,4287	2,5436	2,6582	2,7658	2,8651	2,9568	3,0348	3,1361	3,4113	3,4113
		Finale	0,0004	1,4579	2,0071	2,2550	2,5939	2,8670	3,0902	3,3845	3,6484	4,4822	15,4012
Toscana 1	126.589	Base	2,5925	2,5925	2,7375	2,7649	2,8325	2,9570	3,0801	3,1879	3,2485	3,4779	3,4779
		Finale	0,3691	1,6425	2,1795	2,4149	2,7180	2,9524	3,1799	3,4691	3,7844	4,8728	21,2518
Toscana 2	132.132	Base	2,2223	2,2223	2,3320	2,4239	2,6837	2,7943	2,8934	2,9856	3,0342	3,1779	3,1779
		Finale	0,3999	1,5076	2,0032	2,2106	2,4942	2,7556	2,9863	3,2681	3,5632	4,5268	18,2282
Umbria	49.258	Base	2,5043	2,5043	2,5043	2,6753	2,7701	2,8165	2,8844	2,9019	2,9809	2,9809	2,9809
		Finale	0,0005	1,4850	2,0167	2,2587	2,5540	2,8097	3,0109	3,2901	3,5694	4,4662	10,9975
Marche	77.975	Base	2,5823	2,5823	2,6745	2,7576	2,8627	2,9550	2,9992	3,0697	3,1334	3,3115	3,3115
		Finale	0,2539	1,5005	2,0813	2,3365	2,6533	2,9107	3,1549	3,4884	3,8034	4,8349	18,7957
Lazio	527.669	Base	1,6019	1,6019	2,1909	2,5387	2,7937	2,8982	2,9706	3,0322	3,0675	3,1451	3,1948
		Finale	0,3269	1,4669	1,8988	2,1954	2,5677	2,8367	3,0440	3,2948	3,5801	4,6323	19,8458
Campania	418.763	Base	1,0450	1,0450	2,1242	2,2872	2,4780	2,6972	2,8272	2,9216	2,9850	3,0320	3,1784
		Finale	0,1633	1,0330	1,6768	1,9214	2,2603	2,6019	2,9150	3,2666	3,6066	4,8352	19,3130
Abruzzo	61.612	Base	2,0621	2,0621	2,0621	2,4058	2,7314	2,8131	2,9030	2,9385	2,9475	2,9583	2,9583
		Finale	0,2037	1,4410	1,8958	2,1067	2,4627	2,7663	3,0030	3,2680	3,5183	4,4308	11,9713
Molise	12.807	Base	2,5296	2,5296	2,5296	2,5296	2,5303	2,5624	2,6733	2,8276	2,8276	2,8276	2,8276
		Finale	0,0004	1,0403	1,6353	1,8709	2,2356	2,5651	2,8932	3,3391	3,7605	5,2084	11,8570
Puglia	302.328	Base	1,0533	1,9450	2,2457	2,2977	2,6275	2,8276	2,9362	3,0328	3,1207	3,1795	3,2178
		Finale	0,2218	1,3935	1,9025	2,1135	2,4360	2,7461	3,0174	3,3038	3,5833	4,7567	24,5138
Basilicata	17.215	Base	1,7758	1,7758	1,7758	1,7758	1,8556	2,4959	2,6850	2,8277	2,8277	2,8277	2,8277
		Finale	0,0003	1,0329	1,4094	1,5879	1,8762	2,3090	2,7343	3,1240	3,4155	4,3275	11,9225
Calabria	74.533	Base	1,7308	1,7308	1,8369	1,9085	2,3010	2,4982	2,7241	2,8918	2,9498	3,0974	3,0974
		Finale	0,3435	1,2895	1,6253	1,7739	2,0790	2,4241	2,7816	3,1713	3,4766	4,4051	17,8905
Sicilia	396.975	Base	1,0424	1,0851	2,0423	2,2879	2,5144	2,7176	2,8819	2,9906	3,0231	3,1333	3,4305
		Finale	0,0989	1,0449	1,6446	1,9628	2,3470	2,6655	2,9482	3,2292	3,4729	4,4356	20,3929
Sardegna	87.742	Base	2,3686	2,3686	2,6284	2,6587	2,7591	2,8184	2,8835	2,9457	2,9696	3,1324	3,1324
		Finale	0,1548	1,5188	2,0747	2,2939	2,5587	2,7697	2,9715	3,2676	3,6176	4,7258	18,0854

Fonte: Direzione centrale dei Censimenti generali (DCCG), Istat.

Prospetto 5.24 – Percentili delle distribuzioni dei pesi base e dei pesi finali per l’oggetto “famiglie”.

Dominio	Numero di unità	Pesi	Valori percentili										
			Min	p1	p5	p10	Q1	Me	Q3	p90	p95	p99	Max
Valle d'Aosta	4.942	Base	2,8919	2,8919	2,8919	2,8919	2,8919	2,9557	2,9557	2,9557	2,9557	2,9557	2,9557
		Finale	2,1450	2,5675	2,6532	2,7249	2,8766	2,9401	3,0027	3,0595	3,1269	3,1681	8,6667
Piemonte	295.886	Base	2,6630	2,6630	2,6817	2,7355	2,8273	2,8861	2,9318	2,9752	2,9956	3,0776	3,1141
		Finale	1,7155	2,2878	2,5894	2,6684	2,7850	2,8977	2,9801	3,0434	3,0730	3,1361	6,0077
Lombardia	566.901	Base	2,1297	2,4601	2,7653	2,8126	2,8801	2,9438	2,9969	3,0603	3,1056	3,2999	3,3434
		Finale	1,4811	2,1598	2,5671	2,7341	2,8611	2,9601	3,0410	3,1101	3,1601	3,3515	6,9617
Veneto	220.636	Base	2,2761	2,2761	2,4633	2,5547	2,6463	2,8402	2,8966	2,9781	3,0570	3,1101	3,2795
		Finale	1,6293	2,1294	2,4111	2,5099	2,6626	2,8320	2,9470	3,0443	3,1079	3,2172	4,6850
Prov. Aut. Trento	17.331	Base	1,0545	1,0545	1,0545	1,0545	2,9604	3,0420	3,0629	3,1794	3,1794	3,1794	3,1794
		Finale	0,7042	1,0315	1,0658	1,1116	2,9259	3,0353	3,1245	3,2008	3,2598	3,3515	5,0378
Prov. Aut. Bolzano	22.870	Base	2,8325	2,8325	2,8325	2,8325	2,8668	2,8930	2,9323	2,9463	2,9801	2,9801	2,9801
		Finale	1,8982	1,9651	2,3210	2,5582	2,7970	2,9483	3,0841	3,1372	3,1719	3,2316	3,6137
Friuli	62.230	Base	2,8853	2,8853	2,9052	2,9067	2,9370	2,9597	2,9778	3,0106	3,0855	3,2075	3,2075
		Finale	1,7483	2,1745	2,7884	2,8242	2,9080	2,9781	3,0422	3,1062	3,1539	3,3020	4,6237
Liguria	145.233	Base	2,4003	2,4003	2,7109	2,7442	2,8580	2,9669	3,0290	3,1193	3,1883	3,2479	3,3810
		Finale	1,7856	2,3072	2,5886	2,6886	2,8370	2,9668	3,0746	3,1756	3,2428	3,3784	8,4298
E. Romagna	281.537	Base	2,4310	2,4598	2,6473	2,6967	2,8179	2,9278	3,0111	3,1402	3,1795	3,3260	3,4110
		Finale	1,4575	2,1611	2,5012	2,6143	2,7840	2,9501	3,0683	3,1713	3,2532	3,3949	6,6071
Toscana	258.144	Base	2,2227	2,2227	2,4259	2,5923	2,7767	2,8658	3,0048	3,1349	3,1926	3,3642	3,4896
		Finale	1,6864	2,0943	2,3841	2,5337	2,7304	2,8960	3,0229	3,1804	3,2560	3,4309	4,9859
Umbria	49.123	Base	2,5027	2,5027	2,5027	2,6861	2,7741	2,8280	2,8894	2,9045	2,9829	2,9829	2,9829
		Finale	1,6368	1,9561	2,4370	2,5277	2,7074	2,8575	2,9655	3,0213	3,0526	3,1221	4,2297
Marche	77.828	Base	2,5804	2,5804	2,6755	2,7613	2,8649	2,9568	2,9996	3,0759	3,1352	3,3176	3,3176
		Finale	1,6557	2,1740	2,5839	2,6584	2,8130	2,9592	3,0745	3,1656	3,2593	3,3527	6,8540
Lazio	530.843	Base	1,6014	1,6014	2,0888	2,4778	2,7763	2,8963	2,9690	3,0318	3,0675	3,1498	3,1941
		Finale	1,1623	1,5509	1,9729	2,3750	2,6954	2,9084	3,0117	3,0886	3,1335	3,2275	4,0545
Campania	418.521	Base	1,0450	1,0450	2,1242	2,2866	2,4780	2,6968	2,8272	2,9223	2,9875	3,0320	3,1784
		Finale	0,8226	1,0960	2,0119	2,1608	2,4230	2,6862	2,8902	3,0220	3,0823	3,1799	5,0258
Abruzzo	61.469	Base	2,0625	2,0625	2,0625	2,4124	2,7307	2,8151	2,9023	2,9467	2,9470	2,9587	2,9587
		Finale	1,5840	1,9616	2,0780	2,3722	2,6696	2,8030	2,9410	3,0162	3,0482	3,0879	6,4207
Molise	12.750	Base	2,5345	2,5345	2,5345	2,5345	2,5400	2,5725	2,6911	2,8414	2,8414	2,8414	2,8414
		Finale	1,6793	1,8826	2,2852	2,3520	2,5273	2,6530	2,7498	2,8588	2,9404	3,0014	4,8464
Puglia	302.123	Base	1,0533	1,9453	2,2442	2,2971	2,6287	2,8268	2,9355	3,0328	3,1192	3,1779	3,2171
		Finale	0,8690	1,9349	2,1887	2,3000	2,6122	2,8186	2,9672	3,0848	3,1472	3,2605	6,1060
Basilicata	17.184	Base	1,7779	1,7779	1,7779	1,7779	1,8595	2,4993	2,6866	2,8354	2,8354	2,8354	2,8354
		Finale	1,1810	1,5226	1,7237	1,7740	1,9037	2,4543	2,7049	2,8583	2,8891	2,9780	2,9780
Calabria	74.456	Base	1,7311	1,7311	1,8369	1,9090	2,3006	2,4977	2,7236	2,8936	2,9483	3,1005	3,1005
		Finale	1,4946	1,6617	1,8036	1,9176	2,2521	2,4731	2,7407	2,9803	3,1107	3,2797	3,3647
Sicilia	396.613	Base	1,0421	1,0851	2,0424	2,2870	2,5151	2,7169	2,8808	2,9906	3,0273	3,1348	3,4285
		Finale	0,7922	1,0818	1,9443	2,2094	2,4919	2,7251	2,9030	3,0365	3,1061	3,2672	5,5434
Sardegna	87.599	Base	2,3712	2,3712	2,6266	2,6590	2,7634	2,8215	2,8821	2,9482	2,9705	3,1603	3,1603
		Finale	1,6554	2,0188	2,4819	2,5742	2,7057	2,8369	2,9290	2,9988	3,0516	3,2455	3,3413

Fonte: Direzione centrale dei Censimenti generali (DCCG), Istat.

Prospetto 5.25 – Percentili delle distribuzioni dei pesi base e dei pesi finali per l’oggetto “alloggi e edifici”.

Dominio	Numero di unità	Pesi	Valori percentili										
			Min	p1	p5	p10	Q1	Me	Q3	p90	p95	p99	Max
Valle d'Aosta	4.914	Base	2,8796	2,8796	2,8796	2,8796	2,8796	2,9388	2,9388	2,9388	2,9388	2,9388	2,9388
		Finale	0,8876	1,6652	2,7206	2,7670	2,8173	2,8787	2,9985	3,1546	3,3244	3,7496	4,8938
Piemonte	293.824	Base	2,6522	2,6522	2,6738	2,7204	2,8146	2,8721	2,9166	2,9645	2,9871	3,0743	3,1097
		Finale	0,1434	1,6968	2,5825	2,6591	2,7637	2,8728	2,9728	3,0692	3,1582	3,3785	5,9209
Lombardia	561.469	Base	2,1191	2,4537	2,7475	2,7918	2,8631	2,9265	2,9792	3,0395	3,0642	3,2754	3,3371
		Finale	0,0119	1,6438	2,5214	2,7263	2,8561	2,9574	3,0399	3,1286	3,1938	3,3574	6,5754
Veneto	217.212	Base	2,2705	2,2705	2,4498	2,5367	2,6237	2,8022	2,8644	2,9540	3,0231	3,0696	3,2473
		Finale	0,0658	1,5760	2,3064	2,4643	2,6317	2,8228	2,9442	3,0575	3,1389	3,3903	14,1724
Prov. Aut. Trento	17.179	Base	1,0544	1,0544	1,0544	1,0544	2,9471	3,0099	3,0405	3,1496	3,1496	3,1496	3,1496
		Finale	0,5729	1,0348	1,0714	1,1462	2,9000	3,0182	3,1304	3,2249	3,2860	3,5493	5,6003
Prov. Aut. Bolzano	22.676	Base	2,8090	2,8090	2,8090	2,8090	2,8447	2,8631	2,8978	2,9163	2,9636	2,9636	2,9636
		Finale	0,1674	1,4772	2,5043	2,6241	2,7565	2,9387	3,0678	3,1456	3,1851	3,2601	3,3617
Friuli	61.439	Base	2,8546	2,8546	2,8810	2,8842	2,9099	2,9341	2,9501	2,9902	3,0424	3,1883	3,1883
		Finale	0,2204	1,6561	2,7152	2,8141	2,8906	2,9760	3,0487	3,1210	3,1808	3,3702	5,7727
Liguria	143.169	Base	2,3928	2,3928	2,6825	2,7420	2,8342	2,9206	2,9941	3,0884	3,1416	3,2237	3,3339
		Finale	0,0486	1,5986	2,4415	2,6685	2,8162	2,9532	3,0746	3,2059	3,3049	3,5418	6,1232
E. Romagna	276.398	Base	2,4071	2,4467	2,6182	2,6762	2,7899	2,9003	2,9686	3,0846	3,1530	3,2549	3,3866
		Finale	0,1142	1,5623	2,3197	2,5597	2,7648	2,9356	3,0657	3,2081	3,3165	3,5485	8,2140
Toscana	254.188	Base	2,2194	2,2194	2,4229	2,5700	2,7458	2,8420	2,9702	3,0897	3,1263	3,2881	3,4512
		Finale	0,0737	1,6211	2,1974	2,4337	2,7032	2,8878	3,0227	3,1934	3,2712	3,4873	6,5304
Umbria	48.399	Base	2,4748	2,4748	2,4748	2,6212	2,7374	2,7926	2,8558	2,8645	2,9626	2,9626	2,9626
		Finale	0,0524	1,5536	2,2148	2,4503	2,6816	2,8515	2,9526	3,0272	3,0723	3,2165	16,4400
Marche	76.856	Base	2,5684	2,5684	2,6622	2,7393	2,8410	2,9134	2,9733	3,0413	3,0877	3,2875	3,2875
		Finale	0,1027	1,6203	2,4934	2,6268	2,7981	2,9502	3,0590	3,2108	3,3084	3,5366	14,4340
Lazio	523.505	Base	1,5986	1,5986	2,0766	2,4731	2,7589	2,8757	2,9504	3,0102	3,0487	3,1326	3,1884
		Finale	0,0661	1,5750	1,7628	2,2806	2,7270	2,8846	2,9953	3,0988	3,1708	3,3345	12,8963
Campania	415.672	Base	1,0465	1,0465	2,1217	2,2734	2,4719	2,6876	2,8114	2,9078	2,9698	3,0257	3,1728
		Finale	0,0551	1,0875	1,9506	2,1462	2,4124	2,6777	2,8746	3,0322	3,1158	3,3390	10,0000
Abruzzo	61.032	Base	2,0553	2,0553	2,0553	2,3996	2,7122	2,7903	2,8840	2,9181	2,9235	2,9249	2,9249
		Finale	0,1274	1,6413	2,0569	2,2841	2,6418	2,7995	2,9295	3,0258	3,0693	3,1842	10,5475
Molise	12.673	Base	2,5111	2,5111	2,5111	2,5111	2,5255	2,5585	2,6890	2,8441	2,8441	2,8441	2,8441
		Finale	0,7617	1,6879	2,2731	2,3625	2,5018	2,6285	2,7649	2,8801	2,9854	3,0933	3,9630
Puglia	300.523	Base	1,0528	1,9352	2,2394	2,2913	2,6152	2,8151	2,9228	3,0164	3,1056	3,1628	3,1992
		Finale	0,1046	1,6293	2,1140	2,2912	2,5965	2,8110	2,9707	3,0985	3,1781	3,3228	5,9383
Basilicata	16.886	Base	1,7720	1,7720	1,7720	1,7720	1,8517	2,4914	2,6879	2,8187	2,8187	2,8187	2,8187
		Finale	0,6060	1,5220	1,6758	1,7512	1,8925	2,4589	2,7055	2,8680	2,9747	3,1840	9,5000
Calabria	73.769	Base	1,7259	1,7259	1,8328	1,8999	2,2898	2,4896	2,7007	2,8813	2,9264	3,0870	3,0870
		Finale	0,3913	1,5783	1,7789	1,8892	2,2217	2,4677	2,7580	2,9856	3,0757	3,2971	13,0178
Sicilia	394.118	Base	1,0427	1,0862	2,0096	2,2612	2,5050	2,7066	2,8613	2,9676	3,0178	3,1264	3,4187
		Finale	0,0179	1,0791	1,8073	2,1726	2,4753	2,7110	2,9056	3,0538	3,1403	3,3537	5,8157
Sardegna	86.544	Base	2,3635	2,3635	2,6062	2,6383	2,7252	2,7926	2,8585	2,9188	2,9542	3,1348	3,1348
		Finale	0,0959	1,6494	2,3410	2,5286	2,6974	2,8311	2,9268	3,0094	3,0638	3,3241	4,8179

Fonte: Direzione centrale dei Censimenti generali (DCCG), Istat.

5.6 Considerazioni conclusive

Il calcolo dei pesi di riporto all'universo è stata una operazione tanto necessaria quanto complessa, sia in fase di valutazione e progettazione che in fase di realizzazione. L'impiego della calibrazione è stato dettato dall'esigenza di produrre, per gli incroci relativi alle variabili rilevate solo su campioni di famiglie, stime coerenti, accurate e tempestive.

La coerenza dei dati finali, è intesa in senso "orizzontale" (per incroci riferiti a due o più tavole per lo stesso dominio territoriale), in senso "verticale" (per incroci riferiti a due o più tavole, per lo stesso dominio territoriale, aventi però un diverso dettaglio classificatorio), in senso "territoriale" (per incroci riferiti a una stessa tavola riferita però a domini geografici relativi a diversa scala). Il soddisfacimento di questa proprietà aumenta le possibilità di impiego dei dati finali.

L'accuratezza dei dati stimati, rispetto ai valori marginali riferiti ad incroci esaustivi (dati derivanti da un conteggio) è legata alla soluzione ottima del problema di minimo vincolato che si trova alla base del procedimento di calibrazione. La convergenza dell'algoritmo computazionale impiegato ha permesso il raggiungimento dell'obiettivo evitando così problemi di disallineamento tra i totali (marginali e complessivo) nell'ambito di una stessa tavola. Piccole differenze si osservano a causa dell'arrotondamento delle stime in quanto queste sono il risultato della somma, non intera, di pesi finali (valori non interi).

La tempestività dell'operazione di calcolo dei pesi è stata, infine, dettata dalla necessità di rispettare i tempi di consegna previsti degli output finali; a riguardo è stato definito un piano di elaborazione dei dati che, sfruttando le disponibilità sw/hw in dotazione ha permesso di completare il lavoro nei tempi prestabiliti.

Data la complessità, sia in termini di numero di variabili di incrocio che di dettaglio classificatorio, degli ipercubi (tavole multidimensionali) di diffusione richiesti da Eurostat, il processo di calcolo dei pesi di riporto all'universo è stato "calato" sulle tavole proposte per la diffusione nazionale. In particolare, tali tavole sono state disegnate (per gli oggetti "popolazione", "famiglie" e "alloggi ed edifici") in modo da rendere praticabile il conseguente sistema di vincoli impiegato nella procedura di riponderazione. Si pone in evidenza il lavoro di interazione tra gli esperti tematici (per il disegno delle tavole) e gli esperti metodologi (per la definizione del sistema di vincoli) al fine di raggiungere un giusto compromesso tra quanto si voleva proporre come output informativo e quanto era possibile produrre, in termini di stime calibrate, con le proprietà di coerenza, accuratezza e tempestività. Per le tavole non rientranti nel sistema di vincoli finale sono stati successivamente prodotti dei "pesi ad hoc" al fine di renderle lo stesso disponibili per l'utenza finale.

Per gli ipercubi richiesti dal Piano di diffusione europeo è stato impiegato il sistema di pesi definito per il piano nazionale; i dati così prodotti hanno però presentato disallineamenti in termini di coerenza e accuratezza, proprietà non richieste esplicitamente da Eurostat e per le quali sono state fatte le dovute misurazioni tramite indicatori di qualità in seguito riportati nel *Quality Report* prodotto e inviato a Eurostat.

Al fine di decidere quale strategia adottare per il calcolo dei pesi finali, sono state sperimentate diverse possibilità, dall'impiego di un peso unico alla possibilità di adottare sistemi indipendenti di pesi. Dati però i limiti computazionali e le scelte legate all'architettura informatica per il disegno dei *data warehouse* di produzione e di diffusione, è stato ritenuto praticabile l'impiego di un peso finale unico basato su un sistema di vincoli che fosse in grado di garantire l'intero Piano di diffusione nazionale progettato.

Quindi, una volta decisa la metodologia per la determinazione dei pesi finali, sono stati eseguiti specifici e approfonditi test su dati relativi al censimento del 2001 per valutare i limiti computazionali massimi e per dimensionare le risorse informatiche (hw e sw) necessarie a compiere le operazioni di calcolo, così da ottenere i risultati nei tempi dovuti e con le proprietà attese. Le sperimentazioni hanno inoltre permesso di studiare specifiche soluzioni per risolvere i casi di "non convergenza" dell'algoritmo di calibrazione dovuti, per esempio, a incompatibilità tra i vincoli (presenza di domini innestati – grande Comune dentro una Provincia – con totali noti identici) o zeri campionari (a fronte di totali noti diversi da zero). A riguardo, sono state studiate e messe in esercizio specifiche procedure che hanno preliminarmente risolto le suddette situazioni permettendo così il raggiungimento della convergenza.

La fase sperimentale ha inoltre permesso di potenziare la "macchina computazionale" messa in campo per il calcolo dei pesi finali. Infatti, con il contributo degli esperti informatici dell'Istituto sia sul versante Sas che per la procedura di riponderazione in *Geneses*, il software impiegato nell'operazione è stato adattato in modo da permettere l'impiego fino a 3 mila vincoli in un solo dominio di calibrazione (in fase di esercizio il numero massimo di vincoli è stato quello della Regione Lombardia – senza la Provincia di Milano – per l'oggetto

“popolazione” in cui si sono impiegati ben 2.846 vincoli). Lo sforzo compiuto per adattare le procedure computazionali e la predisposizione delle macchine (PC e server) per il calcolo dei pesi finali del Censimento della popolazione e delle abitazioni del 2011 ha rappresentato un grande risultato mai realizzato in altre indagini campionarie in Istat.

La conclusione dell’attività di calcolo dei pesi di riporto all’universo finali è stata realizzata nei tempi previsti e i risultati, a meno di piccolissime differenze, ha garantito la coerenza e l’accuratezza dei valori delle frequenze assolute riportate nelle tavole di diffusione presenti sul sito istituzionale dell’Istat.

I pesi finali prodotti presentano valori non negativi e mostrano un campo di variazione non troppo ampio; ciò è stato dovuto sia alla distanza impiegata nel procedimento (la “logaritmica troncata”) che all’ampia disponibilità di informazione campionaria raccolta sul campo (il disegno campionario prevedeva un tasso di campionamento pari al 33 per cento).

L’ipotesi di adozione di una frazione sondata più piccola avrebbe richiesto attente valutazioni in quanto, oltre a far aumentare la variabilità campionaria (la varianza delle stime finali) può far crescere il rischio di osservare “zeri campionari” (mancanza di osservazioni nel campione rispetto a fenomeni presenti nella popolazione). Questo rischio è più alto per tavole che prevedono incroci molto fini (elevato numero di celle rispetto alle unità del dominio considerato).

Uno degli aspetti maggiormente critici dell’intera attività della riponderazione è stato il disegno del sistema dei vincoli implementati nella procedura. Infatti, l’adozione nella strategia campionaria del censimento del 2011 di un questionario *short* non troppo ridotto (come inizialmente disegnato) rispetto al questionario *long*, se da un lato ha offerto una maggiore disponibilità di dati esaustivi, dall’altro ha determinato un insieme cospicuo di informazioni di *benchmark* tale da creare un eccesso di “totali noti” per la procedura (le informazioni raccolte in modo completo sulla popolazione, sulle famiglie e sugli alloggi) complicando così il disegno dei vincoli di calibrazione. L’impiego, in alternativa, di un modello *short* con meno domande avrebbe permesso di disegnare un sistema di vincoli con pochi totali noti anche per domini territoriali più piccoli, al limite fino a quello dell’area di censimento, così come è stato sperimentato per la progettazione della strategia campionaria tramite *short/long form* (Borrelli *et al.*, 2011).

L’adozione di un peso finale unico, congiuntamente alla revisione delle tavole del Piano di diffusione nazionale, ha portato vantaggi alla progettazione e alla gestione del *data warehouse* dei risultati finali. Inoltre, il peso finale unico offre la possibilità di modificare ex-post al Piano di diffusione nazionale con l’inserimento di una qualunque nuova tavola che incrocia almeno una variabile *long*, purché i relativi vincoli indotti dalle distribuzioni marginali riferite solo alle variabili *short* siano contenuti nel sistema di vincoli impiegato per il calcolo dei pesi finali.

CAPITOLO 6

ERRORI DI CAMPIONAMENTO¹¹⁵

6.1 Introduzione

A ciascuna stima campionaria ottenuta a partire da un campione di unità viene associato l'errore commesso nell'effettuare tale stima. Tale errore consiste nella differenza tra il valore vero da stimare e la stima ottenuta. Generalmente maggiore è la dimensione del campione, minore è l'errore commesso con la stima.

Poiché le tavole "a stima" (cfr. paragrafo 5.2.4) sono costruite a partire dalle stime delle variabili *long* rilevate in alcuni domini solamente a livello campionario, è necessario pubblicare l'errore campionario atteso da associare alla stima riferita a ciascuna delle celle di ciascuna tavola pubblicata. La pubblicazione di una stima dell'errore per ciascuna cella prevista dagli ipercubi di diffusione risulterebbe però di difficile lettura. Pertanto si ricorre alla costruzione di opportuni modelli di rappresentazione sintetica degli errori campionari che li mettono in relazione con la corrispondente stima della frequenza osservata. I parametri di questi modelli vengono stimati a partire dalle stime e dai relativi errori di un insieme di tavole rappresentative. Pertanto in fase di pubblicazione vengono diffusi non i singoli errori ma i modelli sintetici degli errori. Attraverso questi modelli l'utente potrà calcolare in maniera autonoma l'errore associato alle sole stime di proprio interesse.

Nel seguito del capitolo verranno descritti gli errori campionari (paragrafo 6.2) e la costruzione dei modelli per la loro rappresentazione sintetica (paragrafo 6.3). Nell'Allegato B, alla fine del volume, sono riportati in dettaglio i modelli di errore costruiti, applicabili a tutti gli ipercubi censuari pubblicati.

6.2 Gli errori campionari

Ai fini dell'illustrazione dei passi seguiti per il calcolo degli errori campionari, si indica con:

C , censimento;

S , campionamento;

d , indice del generico dominio territoriale di riferimento delle stime;

i , indice di individuo;

h , indice di strato;

Y , generica variabile oggetto di indagine;

Y_{hi} , valore di Y osservato sull'individuo i dello strato h ;

N_h , numero totale di individui dello strato h ;

n_h , ampiezza del campione di individui selezionati dallo strato h ;

H_d , numero totale di strati nel generico dominio territoriale d .

Si ipotizzi di voler stimare, con riferimento ad un generico dominio d , il totale della generica variabile Y oggetto di indagine. Questo totale sarà determinato dalla somma del totale nelle aree dove i valori sono stati censiti (C) su tutte le unità della popolazione e il totale nelle aree dove tale variabile è stata rilevata a livello campionario (S).

$$Y_d = Y_{C,d} + Y_{S,d} \quad (6.1).$$

La stima del totale $Y_{S,d}$, relativo alle aree campionate, è data da:

$$\hat{Y}_{S,d} = \sum_{h=1}^{H_d} \hat{Y}_h \quad (6.2)$$

¹¹⁵ Il Capitolo 6 è a cura di Emanuela Scavalli. I paragrafi 6.1, 6.2 e 6.3.1 sono stati redatti da Emanuela Scavalli. Il paragrafo 6.3.2 è stato redatto da Francesco Borrelli; il paragrafo 6.3.3 è stato redatto da Stefania Di Domenico. Le elaborazioni e le Tavole contenute nell'Allegato B sono state curate da Francesco Borrelli e Stefania Di Domenico.

essendo

$$\hat{Y}_h = \sum_{i=1}^{n_h} W_{hi} Y_{hi} \quad (6.3)$$

in cui W_{hi} è il peso finale da attribuire agli individui i campionati dal generico strato h .

Pertanto la stima del totale (6.1) è pari a:

$$\hat{Y}_d = Y_{C,d} + \hat{Y}_{S,d} \quad (6.4).$$

Le principali statistiche di interesse per valutare la variabilità campionaria delle stime prodotte dall'indagine sono l'errore di campionamento assoluto e l'errore di campionamento relativo.

Indicando con $\hat{V}ar(\hat{Y}_d)$ la stima della varianza di \hat{Y}_d , la stima dell'errore di campionamento assoluto di \hat{Y}_d si può ottenere mediante la seguente espressione:

$$\hat{\sigma}_d(\hat{Y}_d) = \sqrt{\hat{V}ar(\hat{Y}_d)} \quad (6.5a).$$

In questo caso la parte di varianza riguarderà soltanto la componente campionaria (quella censuaria è nulla in quanto, per definizione, non è soggetta a errore di campionamento), per cui:

$$\hat{\sigma}(\hat{Y}_d) = \sqrt{\hat{V}ar(\hat{Y}_{S,d})} \quad (6.5b).$$

La stima dell'errore di campionamento relativo di \hat{Y}_d è invece definita dall'espressione:

$$\hat{\varepsilon}(\hat{Y}_d) = \hat{\sigma}(\hat{Y}_d)/\hat{Y}_d \quad (6.6).$$

Come descritto nel paragrafo 5.2, 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 preso in considerazione non è funzione lineare dei dati campionari, per la stima della varianza $\hat{V}ar(\hat{Y}_{S,d})$ si è utilizzato il metodo proposto da Woodruff (1971); in base a tale metodo, che si basa sul procedimento di linearizzazione in serie di Taylor, è possibile ricavare la varianza di ogni stimatore non lineare (funzione regolare di totali) calcolando la varianza dell'espressione linearizzata ottenuta.

In particolare, per la definizione dell'espressione linearizzata dello stimatore ci si è riferiti allo stimatore di regressione generalizzata, sfruttando la convergenza asintotica di tutti gli stimatori di ponderazione vincolata a tale stimatore, poiché nel caso di stimatori di ponderazione vincolata che utilizzino funzioni di 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 è data, quindi, da:

$$\hat{Y}_{S,d} \cong \hat{Z}_d = \sum_{h=1}^{H_d} \hat{Z}_{d,h} \quad (6.7a)$$

Essendo, con riferimento al dominio d

$$\hat{Z}_{d,h} = \sum_{i=1}^{n_{d,h}} W_{d,hi} Z_{d,hi} \quad (6.7b)$$

dove, indipendentemente dal dominio d , Z_{hi} è la variabile linearizzata espressa come $Z_{hi} = Y_{hi} - X'_{hi}\beta$, e $X_{hi} = (X_{hi,1}, \dots, X_{hi,k}, \dots, X_{hi,K})'$ il vettore contenente i valori delle K variabili ausiliarie, osservati per il generico individuo i (dello strato h) e $\hat{\beta}$ il vettore (stimato) dei coefficienti di regressione del modello lineare che lega la variabile di interesse Y alle K variabili ausiliarie X . In base alla (6.7a), si ha, quindi, che la stima della varianza della stima $\hat{Y}_{S,d}$ è ottenuta mediante la seguente relazione:

$$\hat{V}ar(\hat{Y}_{S,d}) \cong \hat{V}ar(\hat{Z}_d) = \sum_{h=1}^{H_d} \hat{V}ar(\hat{Z}_{d,h}) \quad (6.8).$$

Dalla (6.8) risulta che la stima della varianza di $\hat{Y}_{S,d}$ è calcolata come somma della stima delle varianze delle stime $\hat{Z}_{a,h}$ nei singoli strati h appartenenti al dominio d .

Gli errori campionari espressi dalla (6.5b) e dalla (6.6) consentono di valutare il grado di accuratezza 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 nell'ipotesi di una sua distribuzione (asintoticamente) Normale; in questo modo l'intervallo viene espresso come segue:

$$\{\hat{Y}_d - k_p \hat{\sigma}(\hat{Y}_d) \leq Y_d \leq \hat{Y}_d + k_p \hat{\sigma}(\hat{Y}_d)\} \quad (6.9).$$

Nella (6.9) il valore di k_p dipende dal valore fissato per la probabilità p ; ad esempio, per $p = 0,95$ si ha $k = 1,96$.

6.3 Definizione dei modelli di rappresentazione sintetica degli errori campionari

La determinazione per ogni stima \hat{Y}_d del suo corrispondente errore di campionamento relativo $\hat{\varepsilon}(\hat{Y}_d)$ non è stato possibile sia per limiti di tempo che per costi di elaborazione; inoltre, le tavole della pubblicazione sarebbero risultate appesantite e di non facile consultazione da parte degli utenti. A questo si aggiunge che non si sarebbero comunque resi disponibili gli errori delle stime non pubblicate.

Per le ragioni sopra esposte, solitamente, associato ai dati di output, si offre una presentazione sintetica degli errori relativi basata sul metodo dei modelli regressivi. Questo metodo richiede la determinazione di una funzione matematica che mette in relazione ciascuna stima con il proprio errore relativo.

La procedura standard per la pubblicazione dei modelli di errore di tutte le indagini adottata dall'Istat prevede l'adozione del seguente modello:

$$\log(\hat{\varepsilon}^2(\hat{Y}_d)) = a + b \cdot \log(\hat{Y}_d) \quad (6.10)$$

dove i parametri a e b vengono stimati utilizzando il metodo dei minimi quadrati.

Nella Tavola 6.1 sono riportati i valori dei coefficienti a e b e dell'indice di determinazione R^2 del modello utilizzato per l'interpolazione degli errori campionari di stime riferite agli individui, per il totale Italia e per ripartizione geografica (nell'Allegato B sono riportati anche quelli per Regione e per Provincia relativi a individui, famiglie e alloggi).

Tavola 6.1 - Valori dei coefficienti a e b e dell'indice di determinazione R^2 del modello per l'interpolazione degli errori campionari delle stime riferite agli individui per ripartizione geografica.

Dominio di stima	a	b	R^2
Ripartizione geografica			
Nord-Ovest	0,09134	-1,04247	0,99000
Nord-Est	-0,14753	-1,04148	0,98670
Centro	0,14635	-1,02727	0,99340
Sud	0,13867	-1,03565	0,99630
Isole	0,12966	-1,02864	0,99580
Italia	0,11122	-1,03491	0,99410

Fonte: Direzione centrale dei Censimenti generali (DCCG), Istat.

Sulla base delle informazioni contenute nella Tavola 6.1, è possibile calcolare la stima dell'errore di campionamento relativo di una determinata stima di frequenza assoluta \hat{Y}_d mediante la formula:

$$\hat{\varepsilon}(\hat{Y}_d) = \sqrt{\exp[a + b \log(\hat{Y}_d)]} \quad (6.11)$$

che si ricava direttamente dalla (6.10).

Per esempio, se la stima \hat{Y}_d si riferisce ad una frequenza osservata su una tavola censuaria pubblicata sugli individui e riferita al dominio ripartizione Nord-occidentale, l'errore relativo corrispondente si ottiene sostituendo nella (6.11), oltre al valore della stima, i valori dei relativi parametri a e b riportati nella riga riferita alla ripartizione Nord-ovest della Tavola 6.1 ($a = 0,09134$; $b = -1,04247$).

La Tavola 6.2 consente di rendere più agevole il calcolo degli errori campionari. Essa ha la seguente struttura: in fiancata sono elencati valori crescenti di stima; le colonne successive contengono gli errori di campionamento relativo, per ciascun dominio territoriale di interesse, già calcolati mediante l'espressione (6.11), corrispondenti alle stime delle frequenze assolute riportate in fiancata.

Tavola 6.2 - Valori interpolati degli errori campionari relativi percentuali delle stime riferite agli individui per ripartizione geografica.

Stima	Ripartizione geografica					Italia
	Nord-Ovest	Nord-Est	Centro	Sud	Isole	
25	0,1960	0,1740	0,2060	0,2020	0,2040	0,2000
50	0,1360	0,1210	0,1440	0,1410	0,1430	0,1400
100	0,0950	0,0840	0,1010	0,0990	0,1000	0,0980
250	0,0590	0,0520	0,0630	0,0610	0,0620	0,0610
500	0,0410	0,0370	0,0440	0,0430	0,0440	0,0420
1.000	0,0290	0,0250	0,0310	0,0300	0,0310	0,0300
5.000	0,0120	0,0110	0,0140	0,0130	0,0130	0,0130
10.000	0,0090	0,0080	0,0090	0,0090	0,0090	0,0090
100.000	0,0030	0,0020	0,0030	0,0030	0,0030	0,0030

Fonte: Direzione centrale dei Censimenti generali (DCCG), Istat.

Le informazioni contenute in tali prospetti permettono di calcolare l'errore relativo di una generica stima di frequenza mediante due procedimenti che risultano di facile applicazione, anche se conducono a risultati meno precisi di quelli ottenibili mediante l'espressione (6.11). Il primo metodo consiste nell'individuare, nella prima colonna del prospetto, il livello di stima che più si avvicina alla stima di interesse e nel considerare come errore relativo il valore che si trova sulla stessa riga, nella colonna corrispondente al dominio territoriale di riferimento.

Con il secondo metodo, poiché al crescere del valore della stima diminuisce l'errore relativo, l'errore campionario della stima \hat{Y}_d si ricava direttamente dalla seguente espressione:

$$\hat{\varepsilon}(\hat{Y}_d) = \hat{\varepsilon}(\hat{Y}_d^{k-1}) - \frac{(\hat{Y}_d - \hat{Y}_d^{k-1})}{(\hat{Y}_d^k - \hat{Y}_d^{k-1})} [\hat{\varepsilon}(\hat{Y}_d^{k-1}) - \hat{\varepsilon}(\hat{Y}_d^k)] \quad (6.12)$$

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 , mentre $\hat{\varepsilon}(\hat{Y}_d^{k-1})$ e $\hat{\varepsilon}(\hat{Y}_d^k)$ sono i corrispondenti errori relativi ricavati nella colonna del dominio territoriale di interesse.

6.3.1 La strategia di costruzione dei modelli di errore

Per ciascuna popolazione di riferimento (individui, famiglie e abitazioni) e per ciascun dominio territoriale (ripartizione geografica, Regione e Provincia) sono state individuate alcune tavole pubblicate ritenute più rappresentative e, a partire dalle stime e dagli errori campionari riferiti a tali tavole, sono stati calcolati i diversi modelli di errore. Il processo si è articolato nelle seguenti fasi:

1. individuazione per ciascun dominio (Regione e Provincia) delle tavole maggiormente rappresentative;
2. per le sopra individuate tavole, calcolo delle stime e delle varianze sui dati campionari utilizzando il software *Genesees v.3.0*¹¹⁶;

¹¹⁶ Maggiori dettagli sul software impiegato sono consultabili sul sito istituzionale dell'Istat al seguente link:

[https://www.istat.it/it/metodi-e-strumenti/metodi-e-strumenti-it-per-la-produzione-statistica/elaborazione/strumenti-di-elaborazione/regenesees-\(r-evolved-generalised-software-for-sampling-estimates-and-errors-in-surveys\)](https://www.istat.it/it/metodi-e-strumenti/metodi-e-strumenti-it-per-la-produzione-statistica/elaborazione/strumenti-di-elaborazione/regenesees-(r-evolved-generalised-software-for-sampling-estimates-and-errors-in-surveys))

3. calcolo delle stesse tavole anche sui Comuni non a campione: questo si è reso necessario in quanto l'errore relativo è dato dalla radice quadrata della varianza della stima sulla stima totale che in questo caso è pari alla parte stimata a campione più la parte rilevata per intero sui Comuni non a campione;
4. calcolo delle stime totali (6.2);
5. calcolo degli errori relativi riportando la radice quadrata della varianza (della parte a campione) sulle stime totali;
6. costruzione (tramite il software *Genesees* e con specifici programmi Sas) dei modelli di errore che tengano conto delle stime e degli errori calcolati secondo le (6.6) e (6.8);
7. verifica tramite l' R^2 dell'attendibilità dei modelli costruiti.

Per svolgere le suddette operazioni sono stati necessari, in ciascuno dei domini in cui sono state effettuate le sperimentazioni:

- i dati campionari con i relativi pesi finali calcolati con il processo di riponderazione;
- il sistema di variabili utilizzate per definire i vincoli di calibrazione;
- i dati dei Comuni non campionati;
- il software *Genesees v3.0* (funzione modelli di errore) e i programmi Sas per la preparazione dei dati.

I modelli di errore per ripartizione geografica sono stati invece calcolati a partire dalle varianze calcolate a livello regionale. Infatti, la mole di dati da elaborare non consentiva al software il calcolo diretto delle varianze campionarie a livello di ripartizione, per problemi di tipo computazionale. Pertanto, le varianze sono state ottenute come somma delle varianze stimate per le Regioni appartenenti alla stessa ripartizione, nell'ipotesi di indipendenza delle varianze tra le Regioni. La procedura seguita nella costruzione dei modelli è stata successivamente replicata per la costruzione dei modelli di errore a livello regionale e provinciale.

6.3.2 Alcuni risultati

I modelli degli errori di campionamento individuati hanno tutti mostrato una bontà di adattamento molto elevata (R^2 vicino a 1) indicando così da farli ritenere altamente rappresentativi.

Gli errori calcolati sono risultati molto bassi sia per l'elevata frazione di campionamento (oltre il 33 per cento), sia per l'alto numero di Comuni che hanno rilevato le variabili censuarie in maniera esaustiva (il 94 per cento dei Comuni italiani con il 54 per cento di popolazione residente). La Tavola 6.3 riporta agli errori calcolati al livello provinciale.

Tavola 6.3 – Statistiche sugli errori campionari provinciali relativi a diversi livelli di stima (individui).

Stima	Media	Minimo	Massimo
25	0,1670	0,1100	0,2350
50	0,1150	0,0720	0,1660
100	0,0790	0,0470	0,1170
250	0,0490	0,0270	0,0740
500	0,0330	0,0180	0,0520
1.000	0,0230	0,0110	0,0360
5.000	0,0100	0,0040	0,0160
10.000	0,0070	0,0030	0,0110
100.000	0,0020	0,0010	0,0040

Fonte: Direzione centrale dei Censimenti generali (DCCG), Istat.

Dall'analisi dei valori riportati nella Tavola 6.3, si evidenzia un comportamento dell'errore diversificato tra le Province a causa della diversa tipologia di Comuni presenti al loro interno in termini di peso demografico, che ha comportato numerosità campionarie molto differenti.

Le Province caratterizzate da valori minimi di errore relativo sono¹¹⁷ Avellino, Lecco, Medio Campidano, Sondrio, Vibo Valentia e Oristano. In larga misura, tale comportamento può essere spiegato dallo scarso ammontare di popolazione sottoposto a campionamento. Le Province che presentano, invece, valori più alti

¹¹⁷ Si fa presente che la provincia di Ogliastra, poiché interessata da una rilevazione totalmente esaustiva, non è stata considerata ai fini del calcolo del valore minimo.

sono Milano, Roma, Napoli, Trieste, Barletta-Andria-Trani e Genova, poiché caratterizzate da una massiccia presenza di Comuni più grandi e, quindi, con una maggiore quota di popolazione sottoposta a campionamento.

6.3.3 Il calcolo degli errori di campionamento per il *Quality report* di Eurostat

Oltre alla costruzione del modello di errore richiesto dal piano di diffusione italiano, è stato necessario fornire alcune misure di errore campionario a Eurostat per la pubblicazione del *Quality report*¹¹⁸. Eurostat non richiede un modello per la rappresentazione sintetica dell'errore simile a quello appena presentato per il piano di diffusione italiano, bensì la stima puntuale dell'errore per ogni singola cella di quattro specifici ipercubi, nei quali sono presenti sia variabili rilevate esaustivamente, sia variabili oggetto di stima.

Gli indicatori di qualità sono stati calcolati con riferimento a quattro ipercubi di diffusione europea relativi rispettivamente a “*occupation*”, “*industry*”, “*status in employment*”, e “*year of arrival in the country*”. Partendo dai microdati con i relativi pesi, è stata ricostruita la componente campionaria e la componente non campionaria. In seguito, la componente campionaria è stata processata con il software *Genesees v.3.0* per calcolare le varianze di ogni singola cella e di ogni distribuzione marginale relativa ad ogni ipercubo. Le varianze così stimate sono state rapportate al totale (la parte campionaria più la parte non campionaria) di ogni singola cella ai fini del calcolo dell'errore relativo.

¹¹⁸ Cfr. “Regulation 1151/2010 - Implementation of Regulation 763/2008”, documento di attuazione del regolamento sui censimenti della popolazione e delle abitazioni, relativo alle modalità e struttura del Quality Report, e al formato tecnico per la trasmissione dei metadati.

Conclusioni

Il secondo fascicolo della raccolta dedicata al 15° *Censimento generale della popolazione e delle abitazioni* illustra “I metodi campionari e le tecniche di rilevazione”. Il dettaglio con cui sono documentate le tecniche di rilevazione adottate nel quadro della strategia censuaria del Censimento 2011, il disegno di campionamento e i processi di calcolo delle stime e degli errori campionari restituisce la complessità del processo censuario e illustra le tante innovazioni, di metodo e di tecniche, che hanno caratterizzato il 15° Censimento generale della popolazione e delle abitazioni.

A partire dall’analisi delle criticità della strategia tradizionale, e dallo studio delle esperienze estere, è stata progettata una strategia di censimento fortemente innovativa, basata sull’uso dei dati delle anagrafi comunali per guidare la rilevazione e sull’adozione di soluzioni differenziate in base all’ampiezza demografica dei Comuni. L’uso delle fonti amministrative a supporto della rilevazione sul campo ha rappresentato la principale innovazione di metodo del Censimento 2011.

Sotto questo aspetto, il Censimento del 2011 rappresenta una tappa fondamentale nel passaggio dal Censimento tradizionale al Censimento permanente, che l’Istat ha avviato nel 2018. Si è trattato infatti di un censimento assistito da liste: le Liste anagrafiche comunali (Lac), utilizzate per la spedizione dei questionari, e le liste ausiliarie, utilizzate per individuare le unità di rilevazione non presenti nelle liste anagrafiche ma facenti parte del campo di osservazione del censimento ovvero dimoranti abitualmente sul territorio comunale. In particolare, per garantire l’esaustività della rilevazione era necessario il ricorso a fonti terze, che fornissero segnali utili per correggere la sotto-copertura delle liste anagrafiche. A tal fine, per tutti i Comuni sono state utilizzate le informazioni della Lista integrativa da fonti ausiliarie (Lifa), costituita dall’Istat incrociando le Lac con i dati dell’Anagrafe tributaria e dell’archivio dei Permessi di soggiorno. Inoltre, nei Comuni di maggiori dimensioni, è stata effettuata la Rilevazione dei numeri civici (Rnc), che ha consentito di individuare ulteriori interni abitativi da verificare sul campo ai fini del recupero della sotto-copertura anagrafica.

Gli obiettivi principali della nuova strategia censuaria erano la riduzione del carico di lavoro degli Uffici comunali di censimento, la massimizzazione del tasso di risposta spontanea (anch’essa mirata alla riduzione del lavoro di *front office* della rete di rilevazione comunale) attraverso la flessibilità garantita ai rispondenti dalla diversificazione delle tecniche di compilazione/restituzione del questionario, e la riduzione del carico statistico sui rispondenti. Questi obiettivi sono stati raggiunti attraverso una strategia modulare e flessibile, i cui caposaldi erano l’utilizzo delle Liste anagrafiche comunali (Lac) come liste di unità di rilevazione, la spedizione postale dei questionari (*mail-out*), l’adozione di un sistema di restituzione multicanale.

Al fine di ridurre il carico statistico complessivo sulle famiglie e produrre un risparmio sui costi della rilevazione, è stata inoltre adottata una strategia di tipo *short form/long form*, per la rilevazione su base campionaria delle variabili socio-economiche; un campione di famiglie ha ricevuto il questionario in forma ridotta, mentre il resto delle famiglie ha compilato il questionario in forma completa.

Un’altra innovazione fondamentale è stata l’adozione di un sistema informatizzato di gestione e monitoraggio della rilevazione (Sgr), che ha rappresentato la spina dorsale dell’intero processo di rilevazione, consentendo di gestire in contemporanea le diverse opzioni per la restituzione del questionario e la tracciabilità in tempo reale di tutte le operazioni sul territorio.

Insieme agli altri fascicoli degli *Atti del 15° Censimento generale della popolazione e delle abitazioni*, il volume fornisce un contributo prezioso alla conoscenza del processo di progettazione e realizzazione del Censimento 2011, utile per gli utilizzatori dei dati censuari e per la comprensione dell’evoluzione del censimento della popolazione in Italia.

RIFERIMENTI BIBLIOGRAFICI

- Abbatini, D., L. Cassata, F. Martire, A. Reale, G. Ruocco, e D. Zindato. 2007. “La progettazione dei Censimenti Generali 2010-2011. 2 - Analisi comparativa di esperienze censuarie estere e valutazione di applicabilità di metodi e tecniche ai censimenti italiani”. *Documenti Istat*, N. 9/2007. Roma, Italia: Istat. http://www.istat.it/it/files/2011/04/2007_9.pdf.
- Aragona, B., and D. Zindato. 2016. “Counting People in the Data Revolution Era: Challenges and Opportunities for Population Censuses”. *International Review of Sociology*, Volume 26, Issue 3: 367-385.
- Benassi, F., M. Bruno, M. Giacommo, M. Silipo, G. Vaste, and D. Zindato. 2014. “Managing census complexity through highly integrated web systems”. *Rivista di statistica ufficiale/Review of official statistics*, N. 3/2014: 43-59. Roma, Italia: Istat. <https://www.istat.it/it/archivio/152014>.
- Bianchi, G., F. Di Pedè, A. Reale, e S. Talice. 2010. “Aree di censimento, nuove suddivisioni sub-comunali per la raccolta campionaria di informazioni aggiuntive durante il prossimo censimento della popolazione: applicazioni nella regione Marche”. Relazione presentata alla XXXI Conferenza Italiana di Scienze Regionali, *Identità, Qualità e Competitività Territoriale, Sviluppo economico e coesione nei Territori alpini*. Aosta, Italia, 20-22 settembre 2010.
- Bianchi, G., A. Reale, F. Di Pedè, S. Talice, E. Patrino, e I. Ronchi. 2007. “Processi per la definizione di aree di censimento sub comunali da utilizzare per il Censimento della popolazione e delle abitazioni del 2011”. Relazione presentata alla XXVIII Conferenza Italiana di Scienze Regionali, *Lo sviluppo territoriale nell'Unione europea. Obiettivi, strategie, politiche*. Bolzano, Italia, 26-28 settembre 2007.
- Borrelli, F., G. Carbonetti, S. Dardanelli, L. De Felici, E. Fiorello, M. Marrone, e M. Verrascina. 2012. “La progettazione dei censimenti generali 2010-2011: misure di accuratezza di tavole di diffusione per livelli territoriali e dettagli informativi”. *Istat working papers*, N. 15/2012. Roma, Italia: Istat. <https://www.istat.it/it/archivio/77982>.
- Borrelli, F., G. Carbonetti, L. De Felici, E. Fiorello, e M. Marrone. 2011. “La progettazione dei censimenti generali 2010-2011: disegni campionari e stima di errori di campionamento”. *Istat working papers*, N. 2/2011. Roma, Italia: Istat. <https://www.istat.it/it/archivio/30080>.
- Carbonetti, G., and P. Righi. 2011. “Estimation Procedures for the Long Form Frequencies of the Italian Population Census”. In *Survey Research Methods and Applications. Proceedings of the Second ITACOMS Conference - Second ITALIAN Conference on Survey Methodology*. Pisa, Italia: Plus - Università di Pisa.
- Cassata, L., e F. Lipizzi. 2012. “L'uso dei GIS nell'aggiornamento delle basi territoriali per i censimenti 2011. Un approfondimento sull'Abruzzo”. *Semestrare di Studi e Ricerca di Geografia*, XXIV, Fascicolo 1, gennaio-giugno 2012: 151-167. https://rosa.uniroma1.it/rosa03/semestrare_di_geografia/article/view/15202/14667.
- Cassata, L., and M.T. Tamburrano. 2011. “The 15th Population Census Pilot Survey: how the register driven census changes the enumerators' role”. In Migani, S., and M. Costa (Eds.). “Statistics in the 150 years from Italian Unification. SIS 2011 Statistical Conference, Bologna, 8 - 10 June 2011”. *Book of short papers, Quaderni di Dipartimento, Serie Ricerche 2011*, N. 10. Bologna, Italy: Università di Bologna, Dipartimento di Scienze Statistiche “Paolo Fortunati”. http://amsacta.unibo.it/3202/1/Quaderni_2011_10_SIS2011_BookofShortPaper.pdf.
- Fellegi, I.P., and A.B. Sunter. 1969. “A Theory for Record Linkage”. *Journal of the American Statistical Association*, Volume 64, Issue 328: 1183-1210.
- Fortini, M., G. Gallo, F. Benassi, L. Mancini, and L. Marcone. 2011. “Counting in Non-Italian Residents: The Use of Permits to Stay Archives in the Next Population Census”. In *Survey Research Methods and*

- Applications. Proceedings of the Second ITACOMS Conference - Second ITALian CONference on Survey Methodology.* Pisa, Italia: Plus - Università di Pisa.
- Fortini, M., G. Gallo, E. Paluzzi, A. Reale, e A. Silvestrini. 2007. “La progettazione dei Censimenti Generali 2010-2011. 3 - Criticità di processo e di prodotto nel 14° Censimento generale della popolazione e delle abitazioni: aspetti rilevanti per la progettazione del 15° Censimento”. *Documenti Istat*, N. 10/2007. Roma, Italia: Istat. http://www.istat.it/it/files/2011/04/2007_10.pdf.
- Gallo, G., E. Paluzzi, F. Benassi, and R. Ferrara. 2014. “The 2011 Italian experience towards *supported-Census* for measuring migration”. *Working Paper N.7.* United Nations Economic Commission for Europe - UNECE. Conference of European Statisticians - CES, *Work Session on Migration Statistics.* Chisinau, Republic of Moldova, 10-12 September 2014.
- Horvitz, D.G., and D.J. Thompson. 1952. “A Generalization of Sampling Without Replacement From a Finite Universe”. *Journal of the American Statistical Association*, Volume 47, N. 260: 663-685.
- Istituto Nazionale di Statistica - Istat. 2017a. “1 - Le norme, l’organizzazione e la progettazione finanziaria”. *Atti del 15° Censimento generale della popolazione e delle abitazioni*, Fascicolo 1. Roma, Italia: Istat. <https://www.istat.it/it/files/2016/08/Atti-Censimento-popolazione-Fascicolo-1.pdf>.
- Istituto Nazionale di Statistica - Istat. 2017b. “3 - I contenuti informativi della rilevazione, la validazione e diffusione dei dati”. *Atti del 15° Censimento generale della popolazione e delle abitazioni*, Fascicolo 3. Roma, Italia: Istat. https://www.istat.it/storage/Atti-15-Censimento-generale/Fascicolo-2/AttiPOP_Fascicolo%203.pdf.
- Istituto Nazionale di Statistica - Istat. 2016. “5 - I sistemi informatici a supporto delle operazioni censuarie”. *Atti del 15° Censimento generale della popolazione e delle abitazioni*, Fascicolo 5. Roma, Italia: Istat. https://www.istat.it/it/files/2016/08/AttiPOP_FASCICOLO-5-web.pdf.
- Istituto Nazionale di Statistica - Istat. 2012a. “L’innovazione tecnologica del Censimento: un vantaggio per i cittadini e per gli enti locali”. *Comunicato stampa.* Roma, Italia: Istat. <https://www.istat.it/it/files/2012/04/innovazione-tecnologica.pdf>.
- Istituto Nazionale di Statistica - Istat. 2012b. “Il Censimento degli edifici e l’archivio dei numeri civici”. *Comunicato stampa.* Roma, Italia: Istat. <http://www.istat.it/it/files/2012/04/edifici.pdf>.
- Istituto Nazionale di Statistica - Istat. 2011. *15° Censimento Generale della popolazione e delle abitazioni 2011. Manuale della rilevazione.* Roma, Italia: Istat.
- Istituto Nazionale di Statistica - Istat. 2006. “Le tecnologie dell’informazione e della comunicazione: disponibilità nelle famiglie e utilizzo degli individui. Anno 2006”. *Statistiche in breve.* Roma, Italia: Istat. http://www3.istat.it/salastampa/comunicati/non_calendario/20061218_01/testointegrale.pdf.
- Istituto Nazionale di Statistica - Istat. 2005. “Le tecnologie dell’informazione e della comunicazione: disponibilità nelle famiglie e utilizzo degli individui. Anno 2005”. *Statistiche in breve.* Roma, Italia: Istat.
- Mancini, L., M. Fortini, L. Marcone, F. Borrelli, and A. Ronconi. 2011. “Assessing the Effectiveness of Administrative Registers in Reducing Under-coverage Errors in a Population Census: Evidence from the 2009 Italian Census Pilot Study”. In Migani, S., and M. Costa (Eds.). “Statistics in the 150 years from Italian Unification. SIS 2011 Statistical Conference, Bologna, 8 - 10 June 2011”. *Book of short papers, Quaderni di Dipartimento*, Serie Ricerche 2011, N. 10. Bologna, Italy: Università di Bologna, Dipartimento di Scienze Statistiche “Paolo Fortunati”.
- Pagliuca, D. 2005. “Genesees V.3.0. Funzione Riponderazione. Manuale utente e aspetti metodologici”. *Tecniche e Strumenti*, N. 2/2005. Roma, Italia: Istat. https://www.istat.it/it/files/2013/12/MANU_RIPONDERAZIONE.pdf.
- Romano, M.C. (a cura di). 2017. “L’utilizzo della tecnica CAWI nelle indagini su individui e famiglie”. *Lecture Statistiche - Metodi.* Roma, Italia: Istat. <https://www.istat.it/it/archivio/203729>.

- Tininini, L., and A. Virgillito. 2013. "The Design of the Online Questionnaire of the Italian Population Census". In *Proceedings of the Conferences on New Techniques and Technologies for Statistics - NTTS*. Brussels, Belgium, 5-7 March 2013.
- Tininini, L., and A. Virgillito. 2012. "The Web-based Data Collection in the Italian Population and Housing Census". Paper presented at the *Meeting on the Management of Statistical Information Systems (MSIS 2012)*. Washington, DC, U.S., 21-23 May 2012.
http://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.50/2012/18_Italy.pdf.
- United Nations, Department of Economic and Social Affairs, Statistics Division. 2007. "Principles and Recommendations for Population and Housing Censuses. Revision 2". *Statistical papers*, Series M, N. 67/Rev.2. New York, NY, U.S., and Geneva, Switzerland: United Nations Publication.
https://unstats.un.org/unsd/demographic-social/Standards-and-Methods/files/Principles_and_Recommendations/Population-and-Housing-Censuses/Series_M67Rev2-E.pdf.
- United Nations Economic Commission for Europe - UNECE. 2014. "Measuring Population and Housing. Practices of UNECE countries in the 2010 round of censuses". *ECE/CES/34*. New York, NY, U.S., and Geneva, Switzerland: United Nations. <https://unece.org/measuring-population-and-housing>.
- United Nations Economic Commission for Europe - UNECE, and Conference of European Statisticians - CES. 2015. "Recommendations for the 2020 Censuses of Population and Housing". *ECE/CES/41* New York, NY, U.S., and Geneva, Switzerland: United Nations.
https://unece.org/DAM/stats/publications/2015/ECECES41_EN.pdf.
- United Nations Economic Commission for Europe – UNECE, and Conference of European Statisticians - CES. 2006. "Recommendations for the 2010 Censuses of Population and Housing. Prepared in cooperation with the Statistical Office of the European Communities (Eurostat)". *ECE/CES/STAT/NONE/2006/4*. New York, NY, U.S., and Geneva, Switzerland: United Nations.
https://unece.org/fileadmin/DAM/stats/publications/CES_2010_Census_Recommendations_English.pdf.
- Woodruff, R.S. 1971. "A Simple Method for Approximating the Variance of a Complicated Estimate". *Journal of the American Statistical Association*, Volume 66, Issue 334: 411-414.

ALLEGATO A
ELENCO DEI COMUNI SOTTOPOSTI A CAMPIONAMENTO AL 15° CENSIMENTO
GENERALE DELLA POPOLAZIONE E DELLE ABITAZIONI

Allegato A – Elenco dei Comuni sottoposti a campionamento al 15° Censimento generale della popolazione e delle abitazioni.

Regione	Provincia	Comune	Cod. Istat
Piemonte	Torino	Carmagnola	001059
Piemonte	Torino	Chieri	001078
Piemonte	Torino	Chivasso	001082
Piemonte	Torino	Collegno	001090
Piemonte	Torino	Grugliasco	001120
Piemonte	Torino	Ivrea	001125
Piemonte	Torino	Moncalieri	001156
Piemonte	Torino	Nichelino	001164
Piemonte	Torino	Orbassano	001171
Piemonte	Torino	Pinerolo	001191
Piemonte	Torino	Rivoli	001219
Piemonte	Torino	Settimo Torinese	001265
Piemonte	Torino	Torino	001272
Piemonte	Torino	Venaria Reale	001292
Piemonte	Vercelli	Vercelli	002158
Piemonte	Novara	Borgomanero	003024
Piemonte	Novara	Novara	003106
Piemonte	Cuneo	Alba	004003
Piemonte	Cuneo	Bra	004029
Piemonte	Cuneo	Cuneo	004078
Piemonte	Cuneo	Fossano	004089
Piemonte	Cuneo	Mondovì	004130
Piemonte	Cuneo	Savigliano	004215
Piemonte	Asti	Asti	005005
Piemonte	Alessandria	Acqui Terme	006001
Piemonte	Alessandria	Alessandria	006003
Piemonte	Alessandria	Casale Monferrato	006039
Piemonte	Alessandria	Novi Ligure	006114
Piemonte	Alessandria	Tortona	006174
Piemonte	Alessandria	Valenza	006177
Piemonte	Biella	Biella	096004
Piemonte	Verbania	Verbania	103072
Valle d'Aosta	Aosta	Aosta	007003
Lombardia	Varese	Busto Arsizio	012026
Lombardia	Varese	Cassano Magnago	012040
Lombardia	Varese	Gallarate	012070
Lombardia	Varese	Saronno	012119
Lombardia	Varese	Varese	012133
Lombardia	Como	Cantù	013041
Lombardia	Como	Como	013075
Lombardia	Como	Mariano Comense	013143
Lombardia	Sondrio	Sondrio	014061
Lombardia	Milano	Abbiategrosso	015002
Lombardia	Milano	Bollate	015027
Lombardia	Milano	Bresso	015032
Lombardia	Milano	Buccinasco	015036
Lombardia	Milano	Cernusco sul Naviglio	015070
Lombardia	Milano	Cesano Boscone	015074
Lombardia	Milano	Cinisello Balsamo	015077

Fonte: Direzione centrale dei Censimenti generali (DCCG), Istat.

Allegato A – Elenco dei 486 Comuni sottoposti a campionamento al 15° Censimento generale della popolazione e delle abitazioni del 2011 (segue).

Regione	Provincia	Comune	Cod. Istat
Lombardia	Milano	Cologno Monzese	015081
Lombardia	Milano	Cornaredo	015087
Lombardia	Milano	Corsico	015093
Lombardia	Milano	Garbagnate Milanese	015105
Lombardia	Milano	Lainate	015116
Lombardia	Milano	Legnano	015118
Lombardia	Milano	Magenta	015130
Lombardia	Milano	Milano	015146
Lombardia	Milano	Novate Milanese	015157
Lombardia	Milano	Paderno Dugnano	015166
Lombardia	Milano	Parabiago	015168
Lombardia	Milano	Peschiera Borromeo	015171
Lombardia	Milano	Pioltello	015175
Lombardia	Milano	Rho	015182
Lombardia	Milano	Rozzano	015189
Lombardia	Milano	San Donato Milanese	015192
Lombardia	Milano	San Giuliano Milanese	015195
Lombardia	Milano	Segrate	015205
Lombardia	Milano	Senago	015206
Lombardia	Milano	Sesto San Giovanni	015209
Lombardia	Bergamo	Bergamo	016024
Lombardia	Bergamo	Dalmine	016091
Lombardia	Bergamo	Seriate	016198
Lombardia	Bergamo	Treviglio	016219
Lombardia	Brescia	Brescia	017029
Lombardia	Brescia	Desenzano del Garda	017067
Lombardia	Brescia	Lumezzane	017096
Lombardia	Brescia	Montichiari	017113
Lombardia	Pavia	Pavia	018110
Lombardia	Pavia	Vigevano	018177
Lombardia	Pavia	Voghera	018182
Lombardia	Cremona	Crema	019035
Lombardia	Cremona	Cremona	019036
Lombardia	Mantova	Castiglione delle Stiviere	020017
Lombardia	Mantova	Mantova	020030
Lombardia	Lecco	Lecco	097042
Lombardia	Lodi	Lodi	098031
Lombardia	Monza e della Brianza	Brugherio	108012
Lombardia	Monza e della Brianza	Cesano Maderno	108019
Lombardia	Monza e della Brianza	Desio	108023
Lombardia	Monza e della Brianza	Giussano	108024
Lombardia	Monza e della Brianza	Limbiate	108027
Lombardia	Monza e della Brianza	Lissone	108028
Lombardia	Monza e della Brianza	Meda	108030
Lombardia	Monza e della Brianza	Monza	108033
Lombardia	Monza e della Brianza	Muggiò	108034
Lombardia	Monza e della Brianza	Nova Milanese	108035
Lombardia	Monza e della Brianza	Seregno	108039
Lombardia	Monza e della Brianza	Seveso	108040

Fonte: Direzione centrale dei Censimenti generali (DCCG), Istat.

Allegato A – Elenco dei 486 Comuni sottoposti a campionamento al 15° Censimento generale della popolazione e delle abitazioni del 2011 (segue).

Regione	Provincia	Comune	Cod. Istat
Lombardia	Monza e della Brianza	Vimercate	108050
Trentino-Alto Adige-Südtirol	Provincia autonoma di Bolzano	Bolzano - Bozen	021008
Trentino-Alto Adige-Südtirol	Provincia autonoma di Bolzano	Bressanone/Brixen	021011
Trentino-Alto Adige-Südtirol	Provincia autonoma di Bolzano	Merano - Meran	021051
Trentino-Alto Adige-Südtirol	Provincia autonoma di Trento	Rovereto	022161
Trentino-Alto Adige-Südtirol	Provincia autonoma di Trento	Trento	022205
Veneto	Verona	Legnago	023044
Veneto	Verona	San Giovanni Lupatoto	023071
Veneto	Verona	Verona	023091
Veneto	Verona	Villafranca di Verona	023096
Veneto	Vicenza	Arzignano	024008
Veneto	Vicenza	Bassano del Grappa	024012
Veneto	Vicenza	Montebelluna	024061
Veneto	Vicenza	Schio	024100
Veneto	Vicenza	Thiene	024105
Veneto	Vicenza	Valdagno	024111
Veneto	Vicenza	Vicenza	024116
Veneto	Belluno	Belluno	025006
Veneto	Belluno	Feltre	025021
Veneto	Treviso	Castelfranco Veneto	026012
Veneto	Treviso	Conegliano	026021
Veneto	Treviso	Mogliano Veneto	026043
Veneto	Treviso	Montebelluna	026046
Veneto	Treviso	Treviso	026086
Veneto	Treviso	Vittorio Veneto	026092
Veneto	Venezia	Chioggia	027008
Veneto	Venezia	Mira	027023
Veneto	Venezia	Mirano	027024
Veneto	Venezia	Portogruaro	027029
Veneto	Venezia	San Donà di Piave	027033
Veneto	Venezia	Spinea	027038
Veneto	Venezia	Venezia	027042
Veneto	Padova	Albignasego	028003
Veneto	Padova	Padova	028060
Veneto	Padova	Selvazzano Dentro	028086
Veneto	Padova	Vigonza	028100
Veneto	Rovigo	Adria	029001
Veneto	Rovigo	Rovigo	029041
Friuli-Venezia Giulia	Udine	Udine	030129
Friuli-Venezia Giulia	Gorizia	Gorizia	031007
Friuli-Venezia Giulia	Gorizia	Monfalcone	031012
Friuli-Venezia Giulia	Trieste	Trieste	032006
Friuli-Venezia Giulia	Pordenone	Pordenone	093033
Liguria	Imperia	Imperia	008031
Liguria	Imperia	Sanremo	008055
Liguria	Imperia	Ventimiglia	008065
Liguria	Savona	Albenga	009002
Liguria	Savona	Savona	009056
Liguria	Genova	Chiavari	010015

Fonte: Direzione centrale dei Censimenti generali (DCCG), Istat.

Allegato A – Elenco dei 486 Comuni sottoposti a campionamento al 15° Censimento generale della popolazione e delle abitazioni del 2011 (segue).

Regione	Provincia	Comune	Cod. Istat
Liguria	Genova	Genova	010025
Liguria	Genova	Rapallo	010046
Liguria	La Spezia	La Spezia	011015
Liguria	La Spezia	Sarzana	011027
Emilia-Romagna	Piacenza	Piacenza	033032
Emilia-Romagna	Parma	Fidenza	034014
Emilia-Romagna	Parma	Parma	034027
Emilia-Romagna	Reggio nell'Emilia	Correggio	035020
Emilia-Romagna	Reggio nell'Emilia	Reggio nell'Emilia	035033
Emilia-Romagna	Reggio nell'Emilia	Scandiano	035040
Emilia-Romagna	Modena	Carpi	036005
Emilia-Romagna	Modena	Castelfranco Emilia	036006
Emilia-Romagna	Modena	Formigine	036015
Emilia-Romagna	Modena	Mirandola	036022
Emilia-Romagna	Modena	Modena	036023
Emilia-Romagna	Modena	Sassuolo	036040
Emilia-Romagna	Modena	Vignola	036046
Emilia-Romagna	Bologna	Bologna	037006
Emilia-Romagna	Bologna	Casalecchio di Reno	037011
Emilia-Romagna	Bologna	Castel San Pietro Terme	037020
Emilia-Romagna	Bologna	Imola	037032
Emilia-Romagna	Bologna	San Giovanni in Persiceto	037053
Emilia-Romagna	Bologna	San Lazzaro di Savena	037054
Emilia-Romagna	Ferrara	Cento	038004
Emilia-Romagna	Ferrara	Comacchio	038006
Emilia-Romagna	Ferrara	Ferrara	038008
Emilia-Romagna	Ravenna	Cervia	039007
Emilia-Romagna	Ravenna	Faenza	039010
Emilia-Romagna	Ravenna	Lugo	039012
Emilia-Romagna	Ravenna	Ravenna	039014
Emilia-Romagna	Forlì-Cesena	Cesena	040007
Emilia-Romagna	Forlì-Cesena	Cesenatico	040008
Emilia-Romagna	Forlì-Cesena	Forlì	040012
Emilia-Romagna	Rimini	Riccione	099013
Emilia-Romagna	Rimini	Rimini	099014
Toscana	Massa - Carrara	Carrara	045003
Toscana	Massa - Carrara	Massa	045010
Toscana	Lucca	Camaione	046005
Toscana	Lucca	Capannori	046007
Toscana	Lucca	Lucca	046017
Toscana	Lucca	Pietrasanta	046024
Toscana	Lucca	Viareggio	046033
Toscana	Pistoia	Monsummano Terme	047009
Toscana	Pistoia	Montecatini-Terre	047011
Toscana	Pistoia	Pistoia	047014
Toscana	Firenze	Campi Bisenzio	048006
Toscana	Firenze	Empoli	048014
Toscana	Firenze	Firenze	048017
Toscana	Firenze	Fucecchio	048019

Fonte: Direzione centrale dei Censimenti generali (DCCG), Istat.

Allegato A – Elenco dei 486 Comuni sottoposti a campionamento al 15° Censimento generale della popolazione e delle abitazioni del 2011 (segue).

Regione	Provincia	Comune	Cod. Istat
Toscana	Firenze	Pontassieve	048033
Toscana	Firenze	Scandicci	048041
Toscana	Firenze	Sesto Fiorentino	048043
Toscana	Livorno	Cecina	049007
Toscana	Livorno	Livorno	049009
Toscana	Livorno	Piombino	049012
Toscana	Livorno	Rosignano Marittimo	049017
Toscana	Pisa	Cascina	050008
Toscana	Pisa	Pisa	050026
Toscana	Pisa	Pontedera	050029
Toscana	Pisa	San Miniato	050032
Toscana	Arezzo	Arezzo	051002
Toscana	Arezzo	Montevarchi	051026
Toscana	Siena	Colle di Val d'Elsa	052012
Toscana	Siena	Poggibonsi	052022
Toscana	Siena	Siena	052032
Toscana	Grosseto	Follonica	053009
Toscana	Grosseto	Grosseto	053011
Toscana	Prato	Prato	100005
Umbria	Perugia	Bastia Umbra	054002
Umbria	Perugia	Città di Castello	054013
Umbria	Perugia	Foligno	054018
Umbria	Perugia	Gubbio	054024
Umbria	Perugia	Perugia	054039
Umbria	Perugia	Spoletto	054051
Umbria	Terni	Terni	055032
Marche	Pesaro e Urbino	Fano	041013
Marche	Pesaro e Urbino	Pesaro	041044
Marche	Ancona	Ancona	042002
Marche	Ancona	Fabriano	042017
Marche	Ancona	Falconara Marittima	042018
Marche	Ancona	Jesi	042021
Marche	Ancona	Osimo	042034
Marche	Ancona	Senigallia	042045
Marche	Macerata	Civitanova Marche	043013
Marche	Macerata	Macerata	043023
Marche	Macerata	Recanati	043044
Marche	Macerata	Tolentino	043053
Marche	Ascoli Piceno	Ascoli Piceno	044007
Marche	Ascoli Piceno	San Benedetto del Tronto	044066
Marche	Fermo	Fermo	109006
Marche	Fermo	Porto Sant'Elpidio	109034
Lazio	Viterbo	Viterbo	056059
Lazio	Rieti	Rieti	057059
Lazio	Roma	Albano Laziale	058003
Lazio	Roma	Anzio	058007
Lazio	Roma	Cerveteri	058029
Lazio	Roma	Civitavecchia	058032
Lazio	Roma	Colleferro	058034

Fonte: Direzione centrale dei Censimenti generali (DCCG), Istat.

Allegato A – Elenco dei 486 Comuni sottoposti a campionamento al 15° Censimento generale della popolazione e delle abitazioni del 2011 (segue).

Regione	Provincia	Comune	Cod. Istat
Lazio	Roma	Frascati	058039
Lazio	Roma	Genzano di Roma	058043
Lazio	Roma	Grottaferrata	058046
Lazio	Roma	Guidonia Montecelio	058047
Lazio	Roma	Marino	058057
Lazio	Roma	Monterotondo	058065
Lazio	Roma	Nettuno	058072
Lazio	Roma	Pomezia	058079
Lazio	Roma	Roma	058091
Lazio	Roma	Tivoli	058104
Lazio	Roma	Velletri	058111
Lazio	Roma	Ladispoli	058116
Lazio	Roma	Ardea	058117
Lazio	Roma	Ciampino	058118
Lazio	Roma	Fiomicino	058120
Lazio	Roma	Fonte Nuova	058122
Lazio	Latina	Aprilia	059001
Lazio	Latina	Cisterna di Latina	059005
Lazio	Latina	Fondi	059007
Lazio	Latina	Formia	059008
Lazio	Latina	Gaeta	059009
Lazio	Latina	Latina	059011
Lazio	Latina	Sezze	059028
Lazio	Latina	Terracina	059032
Lazio	Frosinone	Alatri	060003
Lazio	Frosinone	Cassino	060019
Lazio	Frosinone	Ceccano	060024
Lazio	Frosinone	Ferentino	060033
Lazio	Frosinone	Frosinone	060038
Lazio	Frosinone	Sora	060074
Abruzzo	L'Aquila	Avezzano	066006
Abruzzo	L'Aquila	Sulmona	066098
Abruzzo	Teramo	Giulianova	067025
Abruzzo	Teramo	Roseto degli Abruzzi	067037
Abruzzo	Teramo	Teramo	067041
Abruzzo	Pescara	Montesilvano	068024
Abruzzo	Pescara	Pescara	068028
Abruzzo	Chieti	Chieti	069022
Abruzzo	Chieti	FrancaVilla al Mare	069035
Abruzzo	Chieti	Lanciano	069046
Abruzzo	Chieti	Ortona	069058
Abruzzo	Chieti	Vasto	069099
Molise	Campobasso	Campobasso	070006
Molise	Campobasso	Termoli	070078
Molise	Isernia	Isernia	094023
Campania	Caserta	Aversa	061005
Campania	Caserta	Casal di Principe	061019
Campania	Caserta	Caserta	061022
Campania	Caserta	Castel Volturno	061027

Fonte: Direzione centrale dei Censimenti generali (DCCG), Istat.

Allegato A – Elenco dei 486 Comuni sottoposti a campionamento al 15° Censimento generale della popolazione e delle abitazioni del 2011 (segue).

Regione	Provincia	Comune	Cod. Istat
Campania	Caserta	Maddaloni	061048
Campania	Caserta	Marcianise	061049
Campania	Caserta	Mondragone	061052
Campania	Caserta	Orta di Atella	061053
Campania	Caserta	San Nicola la Strada	061078
Campania	Caserta	Santa Maria Capua Vetere	061083
Campania	Benevento	Benevento	062008
Campania	Napoli	Acerra	063001
Campania	Napoli	Afragola	063002
Campania	Napoli	Arzano	063005
Campania	Napoli	Bacoli	063006
Campania	Napoli	Boscoreale	063008
Campania	Napoli	Caivano	063011
Campania	Napoli	Cardito	063016
Campania	Napoli	Casalnuovo di Napoli	063017
Campania	Napoli	Casoria	063023
Campania	Napoli	Castellammare di Stabia	063024
Campania	Napoli	Frattamaggiore	063032
Campania	Napoli	Giugliano in Campania	063034
Campania	Napoli	Gragnano	063035
Campania	Napoli	Marano di Napoli	063041
Campania	Napoli	Marigliano	063043
Campania	Napoli	Melito di Napoli	063045
Campania	Napoli	Mugnano di Napoli	063048
Campania	Napoli	Napoli	063049
Campania	Napoli	Nola	063050
Campania	Napoli	Ottaviano	063051
Campania	Napoli	Poggioreale	063055
Campania	Napoli	Pomigliano d'Arco	063057
Campania	Napoli	Pompei	063058
Campania	Napoli	Portici	063059
Campania	Napoli	Pozzuoli	063060
Campania	Napoli	Qualiano	063062
Campania	Napoli	Quarto	063063
Campania	Napoli	Ercolano	063064
Campania	Napoli	San Giorgio a Cremano	063067
Campania	Napoli	San Giuseppe Vesuviano	063068
Campania	Napoli	Sant'Anastasia	063072
Campania	Napoli	Sant'Antimo	063073
Campania	Napoli	Somma Vesuviana	063079
Campania	Napoli	Torre Annunziata	063083
Campania	Napoli	Torre del Greco	063084
Campania	Napoli	Vico Equense	063086
Campania	Napoli	Villaricca	063087
Campania	Napoli	Volla	063089
Campania	Avellino	Avellino	064008
Campania	Salerno	Agropoli	065002
Campania	Salerno	Angri	065007
Campania	Salerno	Battipaglia	065014

Fonte: Direzione centrale dei Censimenti generali (DCCG), Istat.

Allegato A – Elenco dei 486 Comuni sottoposti a campionamento al 15° Censimento generale della popolazione e delle abitazioni del 2011 (segue).

Regione	Provincia	Comune	Cod. Istat
Campania	Salerno	Cava de' Tirreni	065037
Campania	Salerno	Eboli	065050
Campania	Salerno	Mercato San Severino	065067
Campania	Salerno	Nocera Inferiore	065078
Campania	Salerno	Nocera Superiore	065079
Campania	Salerno	Pagani	065088
Campania	Salerno	Pontecagnano Faiano	065099
Campania	Salerno	Salerno	065116
Campania	Salerno	Sarno	065135
Campania	Salerno	Scafati	065137
Puglia	Foggia	Cerignola	071020
Puglia	Foggia	Foggia	071024
Puglia	Foggia	Lucera	071028
Puglia	Foggia	Manfredonia	071029
Puglia	Foggia	San Giovanni Rotondo	071046
Puglia	Foggia	San Severo	071051
Puglia	Bari	Acquaviva delle Fonti	072001
Puglia	Bari	Altamura	072004
Puglia	Bari	Bari	072006
Puglia	Bari	Bitonto	072011
Puglia	Bari	Conversano	072019
Puglia	Bari	Corato	072020
Puglia	Bari	Gioia del Colle	072021
Puglia	Bari	Giovinazzo	072022
Puglia	Bari	Gravina in Puglia	072023
Puglia	Bari	Modugno	072027
Puglia	Bari	Mola di Bari	072028
Puglia	Bari	Molfetta	072029
Puglia	Bari	Monopoli	072030
Puglia	Bari	Noicattaro	072032
Puglia	Bari	Palo del Colle	072033
Puglia	Bari	Putignano	072036
Puglia	Bari	Ruvo di Puglia	072038
Puglia	Bari	Santeramo in Colle	072041
Puglia	Bari	Terlizzi	072043
Puglia	Bari	Triggiano	072046
Puglia	Taranto	Ginosa	073007
Puglia	Taranto	Grottaglie	073008
Puglia	Taranto	Manduria	073012
Puglia	Taranto	Martina Franca	073013
Puglia	Taranto	Massafra	073015
Puglia	Taranto	Taranto	073027
Puglia	Brindisi	Brindisi	074001
Puglia	Brindisi	Ceglie Messapica	074003
Puglia	Brindisi	Fasano	074007
Puglia	Brindisi	Francavilla Fontana	074008
Puglia	Brindisi	Mesagne	074010
Puglia	Brindisi	Ostuni	074012
Puglia	Lecce	Casarano	075016

Fonte: Direzione centrale dei Censimenti generali (DCCG), Istat.

Allegato A – Elenco dei 486 Comuni sottoposti a campionamento al 15° Censimento generale della popolazione e delle abitazioni del 2011 (segue).

Regione	Provincia	Comune	Cod. Istat
Puglia	Lecce	Copertino	075022
Puglia	Lecce	Galatina	075029
Puglia	Lecce	Gallipoli	075031
Puglia	Lecce	Lecce	075035
Puglia	Lecce	Nardò	075052
Puglia	Barletta-Andria-Trani	Andria	110001
Puglia	Barletta-Andria-Trani	Barletta	110002
Puglia	Barletta-Andria-Trani	Bisceglie	110003
Puglia	Barletta-Andria-Trani	Canosa di Puglia	110004
Puglia	Barletta-Andria-Trani	Trani	110009
Basilicata	Potenza	Potenza	076063
Basilicata	Matera	Matera	077014
Calabria	Cosenza	Acri	078003
Calabria	Cosenza	Castrovillari	078033
Calabria	Cosenza	Cosenza	078045
Calabria	Cosenza	Rende	078102
Calabria	Cosenza	Rossano	078108
Calabria	Catanzaro	Catanzaro	079023
Calabria	Catanzaro	Lamezia Terme	079160
Calabria	Reggio di Calabria	Reggio di Calabria	080063
Calabria	Crotone	Crotone	101010
Calabria	Vibo Valentia	Vibo Valentia	102047
Sicilia	Trapani	Alcamo	081001
Sicilia	Trapani	Castelvetrano	081006
Sicilia	Trapani	Erice	081008
Sicilia	Trapani	Marsala	081011
Sicilia	Trapani	Mazara del Vallo	081012
Sicilia	Trapani	Trapani	081021
Sicilia	Palermo	Bagheria	082006
Sicilia	Palermo	Carini	082021
Sicilia	Palermo	Misilmeri	082048
Sicilia	Palermo	Monreale	082049
Sicilia	Palermo	Palermo	082053
Sicilia	Palermo	Partinico	082054
Sicilia	Palermo	Termini Imerese	082070
Sicilia	Palermo	Villabate	082079
Sicilia	Messina	Barcellona Pozzo di Gotto	083005
Sicilia	Messina	Messina	083048
Sicilia	Messina	Milazzo	083049
Sicilia	Agrigento	Agrigento	084001
Sicilia	Agrigento	Canicatti	084011
Sicilia	Agrigento	Favara	084017
Sicilia	Agrigento	Licata	084021
Sicilia	Agrigento	Palma di Montechiaro	084027
Sicilia	Agrigento	Sciacca	084041
Sicilia	Caltanissetta	Caltanissetta	085004
Sicilia	Caltanissetta	Gela	085007
Sicilia	Caltanissetta	Niscemi	085013
Sicilia	Caltanissetta	San Cataldo	085016

Fonte: Direzione centrale dei Censimenti generali (DCCG), Istat.

Allegato A – Elenco dei 486 Comuni sottoposti a campionamento al 15° Censimento generale della popolazione e delle abitazioni del 2011 (segue).

Regione	Provincia	Comune	Cod. Istat
Sicilia	Enna	Enna	086009
Sicilia	Enna	Piazza Armerina	086014
Sicilia	Catania	Aci Catena	087003
Sicilia	Catania	Acireale	087004
Sicilia	Catania	Adrano	087006
Sicilia	Catania	Belpasso	087007
Sicilia	Catania	Biancavilla	087008
Sicilia	Catania	Caltagirone	087011
Sicilia	Catania	Catania	087015
Sicilia	Catania	Giarre	087017
Sicilia	Catania	Gravina di Catania	087019
Sicilia	Catania	Mascalucia	087024
Sicilia	Catania	Misterbianco	087029
Sicilia	Catania	Paternò	087033
Sicilia	Catania	San Giovanni la Punta	087041
Sicilia	Catania	Tremestieri Etneo	087051
Sicilia	Ragusa	Comiso	088003
Sicilia	Ragusa	Modica	088006
Sicilia	Ragusa	Ragusa	088009
Sicilia	Ragusa	Scicli	088011
Sicilia	Ragusa	Vittoria	088012
Sicilia	Siracusa	Augusta	089001
Sicilia	Siracusa	Avola	089002
Sicilia	Siracusa	Floridia	089009
Sicilia	Siracusa	Lentini	089011
Sicilia	Siracusa	Noto	089013
Sicilia	Siracusa	Pachino	089014
Sicilia	Siracusa	Rosolini	089016
Sicilia	Siracusa	Siracusa	089017
Sardegna	Sassari	Alghero	090003
Sardegna	Sassari	Porto Torres	090058
Sardegna	Sassari	Sassari	090064
Sardegna	Nuoro	Nuoro	091051
Sardegna	Cagliari	Assemini	092003
Sardegna	Cagliari	Cagliari	092009
Sardegna	Cagliari	Capoterra	092011
Sardegna	Cagliari	Quartu Sant'Elena	092051
Sardegna	Cagliari	Selargius	092068
Sardegna	Cagliari	Monsezzato	092109
Sardegna	Oristano	Oristano	095038
Sardegna	Olbia-Tempio	Olbia	104017
Sardegna	Olbia-Tempio	Tempio Pausania	104025
Sardegna	Medio Campidano	Villacidro	106025
Sardegna	Carbonia-Iglesias	Carbonia	107003
Sardegna	Carbonia-Iglesias	Iglesias	107009

Fonte: Direzione centrale dei Censimenti generali (DCCG), Istat.

ALLEGATO B
MODELLI SINTETICI DI RAPPRESENTAZIONE DEGLI ERRORI CAMPIONARI

Tavola B.1 – Valori dei coefficienti a e b e dell'indice di determinazione R² del modello per l'interpolazione degli errori campionari delle stime riferite alle famiglie per ripartizione geografica.

Dominio di stima	a	b	R ²
Ripartizione geografica			
Nord-Ovest	0,17873	-1,09500	0,9909
Nord-Est	-0,12843	-1,07690	0,9895
Centro	0,22889	-1,08483	0,9973
Sud	0,46718	-1,11172	0,9914
Isole	0,60241	-1,11665	0,9955
Italia	0,43174	-1,09844	0,9938

Fonte: Direzione centrale dei Censimenti generali (DCCG), Istat.

Tavola B.2 – Valori interpolati degli errori campionari relativi percentuali delle stime riferite alle famiglie per ripartizione geografica.

Stima	Ripartizione geografica					Italia
	Nord-Ovest	Nord-Est	Centro	Sud	Isole	
25	0,188	0,166	0,196	0,211	0,224	0,212
50	0,128	0,114	0,134	0,144	0,152	0,145
100	0,088	0,079	0,092	0,098	0,103	0,099
250	0,053	0,048	0,056	0,059	0,062	0,060
500	0,036	0,033	0,039	0,040	0,042	0,041
1.000	0,025	0,023	0,026	0,027	0,029	0,028
5.000	0,010	0,010	0,011	0,011	0,012	0,012
10.000	0,007	0,007	0,008	0,008	0,008	0,008
100.000	0,002	0,002	0,002	0,002	0,002	0,002

Fonte: Direzione centrale dei Censimenti generali (DCCG), Istat.

Tavola B.3 – Valori dei coefficienti a e b e dell'indice di determinazione R² del modello per l'interpolazione degli errori campionari delle stime riferite agli alloggi per ripartizione geografica.

Dominio di stima	a	b	R ²
Ripartizione geografica			
Nord-Ovest	0,35829	-1,09483	0,9935
Nord-Est	0,22405	-1,09388	0,9902
Centro	0,21568	-1,06794	0,9961
Sud	0,21861	-1,07796	0,9930
Isole	0,29788	-1,07821	0,9980
Italia	0,4175	-1,0845	0,9956

Fonte: Direzione centrale dei Censimenti generali (DCCG), Istat.

Tavola B.4 – Valori interpolati degli errori campionari relativi percentuali delle stime riferite agli alloggi per ripartizione geografica.

Stima	Ripartizione geografica					Italia
	Nord-Ovest	Nord-Est	Centro	Sud	Isole	
25	0,205	0,192	0,200	0,197	0,205	0,215
50	0,141	0,132	0,138	0,135	0,141	0,148
100	0,096	0,090	0,095	0,093	0,097	0,101
250	0,058	0,055	0,058	0,057	0,059	0,062
500	0,040	0,037	0,040	0,039	0,041	0,042
1.000	0,027	0,026	0,028	0,027	0,028	0,029
5.000	0,011	0,011	0,012	0,011	0,012	0,012
10.000	0,008	0,007	0,008	0,008	0,008	0,008
100.000	0,002	0,002	0,002	0,002	0,002	0,002

Fonte: Direzione centrale dei Censimenti generali (DCCG), Istat.

Tavola B.5 – Parametri del modello, indice di adattamento e valori interpolati degli errori campionari relativi percentuali delle stime riferite agli individui per regione.

Regione	Parametri		R ²	Frequenze								
	Alfa	Beta		25	50	100	250	500	1.000	5.000	10.000	100.000
Piemonte	-0,13841	-1,02375	0,986	0,180	0,126	0,088	0,055	0,039	0,027	0,012	0,008	0,003
Valle d'Aosta	-0,60491	-1,04823	0,959	0,137	0,095	0,066	0,041	0,028	0,020	0,009	0,006	0,002
Lombardia	0,10379	-1,05036	0,989	0,194	0,135	0,094	0,058	0,040	0,028	0,012	0,008	0,002
Bolzano/Bozen	-0,32250	-1,05627	0,957	0,155	0,108	0,075	0,046	0,032	0,022	0,009	0,007	0,002
Trento	-0,69672	-1,04777	0,959	0,131	0,091	0,063	0,039	0,027	0,019	0,008	0,006	0,002
Veneto	-0,20507	-1,05262	0,984	0,166	0,115	0,080	0,049	0,034	0,024	0,010	0,007	0,002
Friuli-Venezia Giulia	-0,37858	-1,03198	0,979	0,157	0,110	0,077	0,048	0,034	0,023	0,010	0,007	0,002
Liguria	0,07535	-1,01463	0,994	0,203	0,143	0,100	0,063	0,044	0,031	0,014	0,010	0,003
Emilia-Romagna	-0,03909	-1,03463	0,989	0,185	0,130	0,091	0,056	0,039	0,028	0,012	0,008	0,003
Toscana	-0,00957	-1,02492	0,994	0,191	0,134	0,094	0,059	0,041	0,029	0,013	0,009	0,003
Umbria	-0,29360	-1,02342	0,984	0,166	0,117	0,082	0,051	0,036	0,025	0,011	0,008	0,002
Marche	-0,23915	-1,03293	0,983	0,168	0,118	0,082	0,051	0,036	0,025	0,011	0,008	0,002
Lazio	0,30397	-1,02486	0,995	0,224	0,157	0,110	0,069	0,048	0,034	0,015	0,010	0,003
Abruzzo	-0,17041	-1,05052	0,986	0,169	0,118	0,082	0,051	0,035	0,024	0,010	0,007	0,002
Molise	-0,45344	-1,04649	0,967	0,148	0,103	0,072	0,044	0,031	0,021	0,009	0,006	0,002
Campania	0,14601	-1,02372	0,996	0,207	0,145	0,102	0,064	0,045	0,031	0,014	0,010	0,003
Puglia	0,25403	-1,03174	0,998	0,216	0,151	0,106	0,066	0,046	0,032	0,014	0,010	0,003
Basilicata	-0,45181	-1,08452	0,968	0,139	0,096	0,066	0,040	0,027	0,019	0,008	0,005	0,002
Calabria	-0,26664	-1,06372	0,984	0,158	0,109	0,076	0,046	0,032	0,022	0,009	0,007	0,002
Sicilia	0,16840	-1,02322	0,997	0,210	0,147	0,103	0,065	0,045	0,032	0,014	0,010	0,003
Sardegna	0,00037	-1,05402	0,986	0,183	0,127	0,088	0,054	0,038	0,026	0,011	0,008	0,002

Fonte: Direzione centrale dei Censimenti generali (DCCG), Istat.

Tavola B.6 – Parametri del modello, indice di adattamento e valori interpolati degli errori campionari relativi percentuali delle stime riferite alle famiglie per regione.

Regione	Parametri		R ²	Frequenze								
	Alfa	Beta		25	50	100	250	500	1.000	5.000	10.000	100.000
Piemonte	0,10860	-1,09811	0,987	0,180	0,123	0,084	0,051	0,035	0,024	0,010	0,007	0,002
Valle d'Aosta	-0,91800	-1,04814	0,962	0,117	0,081	0,057	0,035	0,024	0,017	0,007	0,005	0,002
Lombardia	0,07930	-1,09655	0,991	0,178	0,122	0,083	0,050	0,034	0,024	0,010	0,007	0,002
Bolzano/Bozen	-0,28882	-1,09580	0,964	0,148	0,101	0,069	0,042	0,029	0,020	0,008	0,006	0,002
Trento	-1,12422	-1,04031	0,971	0,107	0,074	0,052	0,032	0,022	0,016	0,007	0,005	0,001
Veneto	-0,37983	-1,07302	0,987	0,147	0,101	0,070	0,043	0,029	0,020	0,009	0,006	0,002
Friuli-Venezia Giulia	-0,19461	-1,08768	0,980	0,158	0,108	0,074	0,045	0,031	0,021	0,009	0,006	0,002
Liguria	0,25289	-1,09218	0,995	0,196	0,134	0,092	0,056	0,038	0,026	0,011	0,007	0,002
Emilia-Romagna	0,03859	-1,08313	0,995	0,178	0,123	0,084	0,051	0,035	0,024	0,010	0,007	0,002
Toscana	0,26135	-1,09403	0,998	0,196	0,134	0,092	0,056	0,038	0,026	0,011	0,007	0,002
Umbria	-0,49493	-1,05366	0,993	0,143	0,099	0,069	0,043	0,030	0,021	0,009	0,006	0,002
Marche	-0,18408	-1,07900	0,994	0,161	0,111	0,076	0,046	0,032	0,022	0,009	0,006	0,002
Lazio	0,26236	-1,08792	0,997	0,198	0,136	0,093	0,057	0,039	0,027	0,011	0,008	0,002
Abruzzo	-0,25180	-1,08476	0,988	0,154	0,106	0,073	0,044	0,030	0,021	0,009	0,006	0,002
Molise	-0,46502	-1,08953	0,977	0,137	0,094	0,064	0,039	0,027	0,018	0,008	0,005	0,001
Campania	0,42212	-1,11059	0,991	0,207	0,141	0,096	0,058	0,039	0,027	0,011	0,007	0,002
Puglia	0,54255	-1,11073	0,997	0,220	0,149	0,102	0,061	0,042	0,028	0,012	0,008	0,002
Basilicata	-1,08361	-1,06463	0,969	0,105	0,072	0,050	0,031	0,021	0,015	0,006	0,004	0,001
Calabria	-0,20650	-1,11335	0,981	0,150	0,102	0,069	0,042	0,028	0,019	0,008	0,005	0,001
Sicilia	0,61875	-1,11571	0,996	0,226	0,154	0,104	0,063	0,043	0,029	0,012	0,008	0,002
Sardegna	0,08729	-1,10297	0,992	0,177	0,121	0,082	0,050	0,034	0,023	0,010	0,007	0,002

Fonte: Direzione centrale dei Censimenti generali (DCCG), Istat.

Tavola B.7 – Parametri del modello, indice di adattamento e valori interpolati degli errori campionari relativi percentuali delle stime riferite agli alloggi per regione.

Regione	Parametri		R ²	Frequenze								
	Alfa	Beta		25	50	100	250	500	1.000	5.000	10.000	100.000
Piemonte	0,17573	-1,08844	0,987	0,189	0,130	0,089	0,054	0,037	0,025	0,011	0,007	0,002
Valle d'Aosta	-0,87237	-1,03531	0,965	0,122	0,085	0,060	0,037	0,026	0,018	0,008	0,005	0,002
Lombardia	0,27365	-1,09872	0,992	0,196	0,134	0,091	0,055	0,038	0,026	0,011	0,007	0,002
Bolzano/Bozen	-0,54219	-1,05376	0,948	0,140	0,097	0,067	0,042	0,029	0,020	0,009	0,006	0,002
Trento	-0,68330	-1,07003	0,977	0,127	0,088	0,060	0,037	0,026	0,018	0,007	0,005	0,002
Veneto	0,05193	-1,09891	0,990	0,175	0,120	0,082	0,049	0,034	0,023	0,010	0,007	0,002
Friuli-Venezia Giulia	0,11685	-1,10637	0,978	0,179	0,122	0,083	0,050	0,034	0,023	0,010	0,006	0,002
Liguria	0,34359	-1,08101	0,998	0,208	0,143	0,099	0,060	0,041	0,028	0,012	0,008	0,002
Emilia-Romagna	0,30560	-1,09320	0,994	0,201	0,137	0,094	0,057	0,039	0,027	0,011	0,008	0,002
Toscana	0,24781	-1,07763	0,996	0,200	0,138	0,095	0,058	0,040	0,027	0,012	0,008	0,002
Umbria	-0,31101	-1,05507	0,986	0,157	0,109	0,075	0,047	0,032	0,022	0,010	0,007	0,002
Marche	0,05209	-1,08706	0,992	0,178	0,122	0,084	0,051	0,035	0,024	0,010	0,007	0,002
Lazio	0,21277	-1,06556	0,998	0,200	0,138	0,096	0,059	0,041	0,028	0,012	0,008	0,002
Abruzzo	-0,40688	-1,05570	0,993	0,149	0,103	0,072	0,044	0,031	0,021	0,009	0,006	0,002
Molise	-0,78269	-1,04063	0,966	0,127	0,088	0,062	0,038	0,027	0,019	0,008	0,006	0,002
Campania	0,48033	-1,09939	0,997	0,217	0,148	0,101	0,061	0,042	0,029	0,012	0,008	0,002
Puglia	0,49738	-1,09335	0,997	0,221	0,151	0,103	0,063	0,043	0,029	0,012	0,008	0,002
Basilicata	-1,68252	-0,98367	0,949	0,089	0,063	0,045	0,029	0,020	0,014	0,007	0,005	0,001
Calabria	-0,59161	-1,06512	0,985	0,134	0,093	0,064	0,039	0,027	0,019	0,008	0,006	0,002
Sicilia	0,38109	-1,08151	0,998	0,212	0,146	0,100	0,061	0,042	0,029	0,012	0,008	0,002
Sardegna	-0,45335	-1,03712	0,986	0,150	0,105	0,073	0,046	0,032	0,022	0,010	0,007	0,002

Fonte: Direzione centrale dei Censimenti generali (DCCG), Istat.

Tavola B.8 – Parametri del modello, indice di adattamento e valori interpolati degli errori campionari relativi percentuali delle stime riferite agli individui per Provincia.

Provincia	Parametri		R ²	Frequenze								
	Alfa	Beta		25	50	100	250	500	1.000	5.000	10.000	100.000
Agrigento	-0,01454	-1,04740	0,976	0,184	0,128	0,089	0,055	0,038	0,027	0,011	0,008	0,002
Alessandria	-0,13337	-1,04334	0,966	0,174	0,122	0,085	0,052	0,037	0,025	0,011	0,008	0,002
Ancona	0,00689	-1,11559	0,973	0,187	0,131	0,091	0,056	0,039	0,027	0,012	0,008	0,002
Aosta	-0,22331	-1,13643	0,923	0,144	0,097	0,065	0,039	0,026	0,018	0,007	0,005	0,001
Arezzo	-0,33429	-1,07866	0,952	0,149	0,103	0,071	0,043	0,030	0,020	0,009	0,006	0,002
Ascoli	-0,06093	-1,06970	0,946	0,173	0,120	0,083	0,051	0,035	0,024	0,010	0,007	0,002
Asti	-0,21788	-1,10857	0,930	0,151	0,103	0,070	0,042	0,029	0,019	0,008	0,005	0,002
Avellino	-0,47520	-1,22196	0,892	0,110	0,072	0,047	0,027	0,018	0,012	0,004	0,003	0,001
Bari	0,23837	-1,00917	0,992	0,222	0,156	0,110	0,069	0,049	0,035	0,015	0,011	0,003
Barletta - Andria - Trani	0,36021	-1,01069	0,990	0,235	0,166	0,117	0,074	0,052	0,036	0,016	0,011	0,004
Belluno	-0,35501	-1,14433	0,923	0,133	0,089	0,060	0,036	0,024	0,016	0,006	0,004	0,001
Benevento	-0,23507	-1,18240	0,909	0,133	0,088	0,058	0,034	0,023	0,015	0,006	0,004	0,001
Bergamo	-0,32656	-1,12996	0,944	0,138	0,093	0,063	0,038	0,025	0,017	0,007	0,005	0,001
Biella	-0,14518	-1,14331	0,937	0,148	0,099	0,067	0,040	0,027	0,018	0,007	0,005	0,001
Bologna	-0,03019	-1,02770	0,975	0,188	0,132	0,092	0,058	0,040	0,028	0,012	0,009	0,003
Bolzano	-0,09639	-1,10331	0,932	0,161	0,110	0,075	0,045	0,031	0,021	0,009	0,006	0,002
Brescia	-0,50215	-1,08463	0,943	0,136	0,093	0,064	0,039	0,027	0,018	0,008	0,005	0,002
Brindisi	0,00149	-1,03581	0,976	0,189	0,132	0,092	0,057	0,040	0,028	0,012	0,008	0,003
Cagliari	0,14113	-1,04145	0,975	0,201	0,140	0,098	0,061	0,042	0,029	0,013	0,009	0,003
Caltanissetta	0,22855	-1,03444	0,981	0,212	0,148	0,104	0,064	0,045	0,031	0,014	0,010	0,003
Campobasso	-0,08024	-1,11165	0,932	0,161	0,109	0,074	0,045	0,030	0,021	0,008	0,006	0,002
Carbonia-Iglesias	0,00720	-1,04257	0,938	0,167	0,113	0,077	0,046	0,031	0,021	0,009	0,006	0,002
Caserta	-0,06957	-1,05965	0,975	0,175	0,122	0,084	0,052	0,036	0,025	0,011	0,007	0,002
Catania	0,09412	-1,01226	0,986	0,206	0,145	0,102	0,064	0,045	0,032	0,014	0,010	0,003
Catanzaro	-0,05536	-1,07859	0,956	0,171	0,118	0,081	0,050	0,034	0,023	0,010	0,007	0,002
Chieti	-0,05931	-1,08130	0,955	0,170	0,117	0,081	0,049	0,034	0,023	0,010	0,007	0,002
Como	-0,49436	-1,06291	0,954	0,141	0,098	0,068	0,042	0,029	0,020	0,008	0,006	0,002
Cosenza	-0,36877	-1,11635	0,947	0,138	0,094	0,064	0,038	0,026	0,018	0,007	0,005	0,001
Cremona	-0,15655	-1,12780	0,933	0,151	0,102	0,069	0,041	0,028	0,019	0,008	0,005	0,001
Crotone	0,04864	-1,16861	0,899	0,156	0,104	0,069	0,041	0,027	0,018	0,007	0,005	0,001
Cuneo	-0,31369	-1,10292	0,940	0,145	0,099	0,067	0,041	0,028	0,019	0,008	0,005	0,001
Enna	-0,05148	-1,16833	0,934	0,149	0,099	0,066	0,039	0,026	0,017	0,007	0,004	0,001
Fermo	-0,25045	-1,13395	0,931	0,142	0,096	0,065	0,039	0,026	0,018	0,007	0,005	0,001
Ferrara	-0,14343	-1,08553	0,932	0,162	0,111	0,076	0,046	0,032	0,022	0,009	0,006	0,002
Firenze	-0,02197	-1,02013	0,979	0,192	0,134	0,094	0,059	0,042	0,029	0,013	0,009	0,003
Foggia	0,12517	-1,03602	0,972	0,201	0,140	0,098	0,061	0,043	0,030	0,013	0,009	0,003
Forlì-Cesena	0,01514	-1,04311	0,973	0,188	0,131	0,091	0,057	0,039	0,027	0,012	0,008	0,002
Frosinone	-0,40814	-1,07048	0,949	0,146	0,100	0,069	0,042	0,029	0,020	0,009	0,006	0,002
Genova	0,22344	-1,00420	0,990	0,222	0,157	0,111	0,070	0,049	0,035	0,016	0,011	0,003
Gorizia	0,02592	-1,06980	0,956	0,181	0,125	0,086	0,053	0,036	0,025	0,011	0,007	0,002
Grosseto	-0,14699	-1,07697	0,932	0,164	0,113	0,078	0,048	0,033	0,023	0,009	0,007	0,002
Imperia	0,02380	-1,04798	0,970	0,187	0,130	0,091	0,056	0,039	0,027	0,012	0,008	0,002
Isernia	-0,18645	-1,19975	0,875	0,132	0,087	0,058	0,033	0,022	0,014	0,006	0,004	0,001
La Spezia	-0,01202	-1,03517	0,970	0,188	0,131	0,092	0,057	0,040	0,028	0,012	0,008	0,003
L'Aquila	-0,43864	-1,12738	0,939	0,131	0,089	0,060	0,036	0,024	0,016	0,007	0,004	0,001
Latina	-0,05709	-1,03545	0,969	0,184	0,128	0,090	0,056	0,039	0,027	0,012	0,008	0,003
Lecce	-0,29233	-1,09356	0,959	0,149	0,102	0,070	0,042	0,029	0,020	0,008	0,006	0,002
Lecco	-0,69523	-1,13601	0,924	0,113	0,077	0,052	0,031	0,021	0,014	0,006	0,004	0,001
Livorno	0,16798	-1,02111	0,980	0,210	0,148	0,104	0,065	0,046	0,032	0,014	0,010	0,003
Lodi	-0,32797	-1,16454	0,919	0,130	0,087	0,058	0,034	0,023	0,015	0,006	0,004	0,001
Lucca	0,05470	-1,02293	0,978	0,198	0,139	0,097	0,061	0,043	0,030	0,013	0,009	0,003
Macerata	-0,29902	-1,08158	0,943	0,151	0,104	0,071	0,043	0,030	0,021	0,009	0,006	0,002
Mantova	-0,51859	-1,15890	0,917	0,119	0,080	0,054	0,031	0,021	0,014	0,006	0,004	0,001
Massa-Carrara	0,12938	-1,01529	0,976	0,208	0,146	0,103	0,065	0,045	0,032	0,014	0,010	0,003
Matera	-0,08417	-1,13551	0,914	0,154	0,104	0,070	0,042	0,028	0,019	0,008	0,005	0,001

Fonte: Direzione centrale dei Censimenti generali (DCCG), Istat.

Tavola B.8 (segue).

Provincia	Parametri		R ²	Frequenze									
	Alfa	Beta		25	50	100	250	500	1.000	5.000	10.000	100.000	
Medio Campidano	-0,31963	-1,23388	0,912	0,117	0,076	0,050	0,028	0,018	0,012	0,004	0,003	0,001	
Messina	-0,14334	-1,06349	0,958	0,168	0,116	0,080	0,049	0,034	0,024	0,010	0,007	0,002	
Milano	0,28690	-1,01809	0,994	0,224	0,158	0,111	0,069	0,049	0,034	0,015	0,011	0,003	
Modena	-0,06315	-1,03109	0,973	0,184	0,129	0,090	0,056	0,039	0,028	0,012	0,008	0,003	
Monza e della Brianza	0,07552	-1,01905	0,989	0,201	0,141	0,099	0,062	0,044	0,031	0,014	0,010	0,003	
Napoli	0,27575	-1,00623	0,995	0,227	0,160	0,113	0,071	0,050	0,036	0,016	0,011	0,004	
Novara	-0,21683	-1,07567	0,958	0,159	0,109	0,075	0,046	0,032	0,022	0,009	0,006	0,002	
Nuoro	-0,03827	-1,18163	0,904	0,146	0,097	0,065	0,038	0,025	0,017	0,006	0,004	0,001	
Ogliastra	-	-	-	-	-	-	-	-	-	-	-	-	
Olbia-Tempio	-0,18527	-1,06695	0,948	0,164	0,113	0,078	0,048	0,033	0,023	0,010	0,007	0,002	
Oriстано	-0,00698	-1,24918	0,907	0,133	0,087	0,056	0,032	0,021	0,013	0,005	0,003	0,001	
Padova	-0,22514	-1,09492	0,944	0,153	0,105	0,072	0,043	0,030	0,020	0,008	0,006	0,002	
Palermo	0,24928	-1,02058	0,988	0,219	0,154	0,108	0,068	0,048	0,033	0,015	0,010	0,003	
Parma	-0,02361	-1,07320	0,961	0,176	0,121	0,083	0,051	0,035	0,024	0,010	0,007	0,002	
Pavia	-0,34135	-1,07101	0,947	0,150	0,104	0,072	0,044	0,030	0,021	0,009	0,006	0,002	
Perugia	-0,25514	-1,04506	0,959	0,164	0,114	0,079	0,049	0,034	0,024	0,010	0,007	0,002	
Pesaro e Urbino	-0,16483	-1,08404	0,946	0,161	0,110	0,076	0,046	0,032	0,022	0,009	0,006	0,002	
Pescara	0,12062	-1,05250	0,964	0,195	0,136	0,094	0,058	0,040	0,028	0,012	0,008	0,002	
Piacenza	-0,17945	-1,08122	0,944	0,160	0,110	0,076	0,046	0,032	0,022	0,009	0,006	0,002	
Pisa	-0,23222	-1,04839	0,954	0,165	0,115	0,080	0,049	0,034	0,024	0,010	0,007	0,002	
Pistoia	-0,14500	-1,06107	0,965	0,169	0,117	0,081	0,050	0,034	0,024	0,010	0,007	0,002	
Pordenone	-0,39608	-1,15755	0,924	0,127	0,085	0,057	0,034	0,022	0,015	0,006	0,004	0,001	
Potenza	-0,45548	-1,16744	0,902	0,122	0,081	0,054	0,032	0,021	0,014	0,006	0,004	0,001	
Prato	0,28830	-1,01914	0,988	0,224	0,157	0,111	0,069	0,049	0,034	0,015	0,011	0,003	
Ragusa	0,15739	-1,03521	0,981	0,204	0,143	0,100	0,062	0,043	0,030	0,013	0,009	0,003	
Ravenna	-0,05143	-1,05663	0,960	0,178	0,123	0,086	0,053	0,037	0,025	0,011	0,008	0,002	
Reggio di Calabria	-0,12242	-1,10861	0,938	0,158	0,108	0,073	0,044	0,030	0,020	0,008	0,006	0,002	
Reggio nell'Emilia	-0,28579	-1,06349	0,954	0,157	0,108	0,075	0,046	0,032	0,022	0,009	0,006	0,002	
Rieti	-0,28494	-1,11954	0,914	0,143	0,097	0,066	0,039	0,027	0,018	0,007	0,005	0,001	
Rimini	-0,04554	-1,04131	0,966	0,183	0,128	0,089	0,055	0,038	0,027	0,012	0,008	0,002	
Roma	0,26649	-1,00509	0,995	0,227	0,160	0,113	0,071	0,050	0,036	0,016	0,011	0,004	
Rovigo	-0,19519	-1,15960	0,916	0,140	0,094	0,063	0,037	0,025	0,017	0,007	0,004	0,001	
Salerno	-0,11307	-1,03600	0,970	0,178	0,125	0,087	0,054	0,038	0,026	0,011	0,008	0,002	
Sassari	0,09290	-1,06678	0,964	0,188	0,130	0,090	0,055	0,038	0,026	0,011	0,008	0,002	
Savona	-0,36334	-1,08105	0,952	0,146	0,101	0,069	0,042	0,029	0,020	0,008	0,006	0,002	
Siena	-0,30287	-1,08734	0,935	0,149	0,102	0,070	0,043	0,029	0,020	0,008	0,006	0,002	
Siracusa	0,22371	-1,02836	0,987	0,214	0,150	0,105	0,065	0,046	0,032	0,014	0,010	0,003	
Sondrio	-0,17656	-1,27076	0,917	0,118	0,076	0,049	0,027	0,018	0,011	0,004	0,003	0,001	
Taranto	0,13158	-1,02942	0,982	0,204	0,143	0,100	0,062	0,044	0,031	0,013	0,009	0,003	
Teramo	-0,30614	-1,11133	0,935	0,143	0,098	0,066	0,040	0,027	0,018	0,008	0,005	0,001	
Terni	-0,18565	-1,03511	0,935	0,172	0,120	0,084	0,052	0,037	0,026	0,011	0,008	0,002	
Torino	0,06728	-1,01503	0,986	0,202	0,142	0,100	0,063	0,044	0,031	0,014	0,010	0,003	
Trapani	0,14830	-1,02018	0,984	0,209	0,146	0,103	0,064	0,045	0,032	0,014	0,010	0,003	
Trento	-0,41128	-1,10259	0,930	0,138	0,094	0,064	0,039	0,026	0,018	0,007	0,005	0,001	
Treviso	-0,34059	-1,09133	0,953	0,146	0,100	0,068	0,041	0,028	0,019	0,008	0,006	0,002	
Trieste	0,31267	-1,00631	0,991	0,231	0,163	0,115	0,073	0,051	0,036	0,016	0,011	0,004	
Udine	-0,40956	-1,12468	0,926	0,133	0,090	0,061	0,037	0,025	0,017	0,007	0,005	0,001	
Varese	-0,33525	-1,04944	0,969	0,156	0,109	0,075	0,047	0,032	0,023	0,010	0,007	0,002	
Venezia	0,01125	-1,04644	0,974	0,187	0,130	0,090	0,056	0,039	0,027	0,012	0,008	0,002	
Verbania	-0,34960	-1,15506	0,938	0,131	0,088	0,059	0,035	0,023	0,016	0,006	0,004	0,001	
Vercelli	-0,12685	-1,13468	0,939	0,151	0,102	0,069	0,041	0,028	0,019	0,007	0,005	0,001	
Verona	-0,19007	-1,07909	0,951	0,160	0,110	0,076	0,046	0,032	0,022	0,009	0,006	0,002	
Vibo Valentia	-0,11617	-1,27683	0,911	0,121	0,078	0,050	0,028	0,018	0,011	0,004	0,003	0,001	
Vicenza	-0,26021	-1,06504	0,956	0,158	0,109	0,076	0,046	0,032	0,022	0,009	0,007	0,002	
Viterbo	-0,60393	-1,14362	0,911	0,117	0,079	0,053	0,031	0,021	0,014	0,006	0,004	0,001	

Fonte: Direzione centrale dei Censimenti generali (DCCG), Istat.

Tavola B.9 – Parametri del modello, indice di adattamento e valori interpolati degli errori campionari relativi percentuali delle stime riferite alle famiglie per Provincia.

Provincia	Parametri		R ²	Frequenze								
	Alfa	Beta		25	50	100	250	500	1.000	5.000	10.000	100.000
Agrigento	0,12995	-1,11277	0,997	0,178	0,121	0,082	0,049	0,034	0,023	0,009	0,006	0,002
Alessandria	0,15974	-1,11995	0,986	0,179	0,121	0,082	0,049	0,033	0,023	0,009	0,006	0,002
Ancona	-0,07484	-1,08038	0,993	0,169	0,116	0,080	0,049	0,034	0,023	0,010	0,007	0,002
Aosta	-0,94118	-1,04609	0,961	0,116	0,081	0,056	0,035	0,024	0,017	0,007	0,005	0,002
Arezzo	-0,41220	-1,09263	0,993	0,140	0,096	0,066	0,040	0,027	0,019	0,008	0,005	0,002
Ascoli	0,06087	-1,10628	0,994	0,174	0,118	0,081	0,049	0,033	0,023	0,009	0,006	0,002
Asti	-0,36683	-1,10068	0,969	0,142	0,097	0,066	0,040	0,027	0,019	0,008	0,005	0,001
Avellino	-2,23779	-1,01157	0,908	0,064	0,045	0,032	0,020	0,014	0,010	0,004	0,003	0,001
Bari	0,58676	-1,11052	0,999	0,224	0,153	0,104	0,063	0,043	0,029	0,012	0,008	0,002
Barletta - Andria - Trani	0,70162	-1,11287	0,998	0,237	0,161	0,110	0,066	0,045	0,030	0,012	0,008	0,002
Belluno	-0,77606	-1,09187	0,976	0,117	0,080	0,055	0,033	0,023	0,016	0,006	0,004	0,001
Benevento	-0,65589	-1,12953	0,928	0,117	0,079	0,053	0,032	0,022	0,015	0,006	0,004	0,001
Bergamo	-0,91006	-1,08237	0,986	0,111	0,076	0,052	0,032	0,022	0,015	0,006	0,004	0,001
Biella	-0,32976	-1,11589	0,987	0,141	0,096	0,065	0,039	0,026	0,018	0,007	0,005	0,001
Bologna	-0,02021	-1,08240	0,993	0,173	0,119	0,082	0,050	0,034	0,024	0,010	0,007	0,002
Bolzano	-0,31329	-1,09380	0,963	0,147	0,101	0,069	0,042	0,029	0,020	0,008	0,006	0,002
Brescia	-0,41060	-1,11938	0,989	0,134	0,091	0,062	0,037	0,025	0,017	0,007	0,005	0,001
Brindisi	0,25499	-1,10922	0,993	0,191	0,130	0,088	0,053	0,036	0,025	0,010	0,007	0,002
Cagliari	0,35063	-1,11262	0,995	0,199	0,135	0,092	0,055	0,038	0,026	0,010	0,007	0,002
Caltanissetta	0,59600	-1,12957	0,997	0,219	0,148	0,100	0,060	0,040	0,027	0,011	0,007	0,002
Campobasso	-0,40434	-1,09112	0,976	0,141	0,097	0,066	0,040	0,028	0,019	0,008	0,005	0,002
Carbonia-Iglesias	-0,52112	-1,05690	0,985	0,141	0,098	0,068	0,042	0,029	0,020	0,009	0,006	0,002
Caserta	0,10874	-1,12475	0,989	0,173	0,117	0,079	0,047	0,032	0,022	0,009	0,006	0,002
Catania	0,52960	-1,11985	0,997	0,215	0,146	0,099	0,059	0,040	0,027	0,011	0,008	0,002
Catanzaro	0,09139	-1,12966	0,987	0,170	0,115	0,078	0,046	0,031	0,021	0,009	0,006	0,002
Chieti	-0,45621	-1,06794	0,981	0,143	0,099	0,068	0,042	0,029	0,020	0,008	0,006	0,002
Como	-0,72574	-1,07218	0,989	0,124	0,085	0,059	0,036	0,025	0,017	0,007	0,005	0,001
Cosenza	-0,44550	-1,12884	0,976	0,130	0,088	0,059	0,035	0,024	0,016	0,007	0,004	0,001
Cremona	-0,43080	-1,09424	0,983	0,139	0,095	0,065	0,039	0,027	0,018	0,008	0,005	0,001
Crotone	-0,26358	-1,11769	0,974	0,145	0,098	0,067	0,040	0,027	0,018	0,008	0,005	0,001
Cuneo	-0,73376	-1,07580	0,984	0,123	0,084	0,058	0,036	0,024	0,017	0,007	0,005	0,001
Enna	-0,61961	-1,09451	0,983	0,126	0,086	0,059	0,036	0,024	0,017	0,007	0,005	0,001
Fermo	-0,48476	-1,10016	0,992	0,134	0,091	0,062	0,038	0,026	0,018	0,007	0,005	0,001
Ferrara	-0,27931	-1,09052	0,987	0,150	0,103	0,071	0,043	0,029	0,020	0,008	0,006	0,002
Firenze	0,07161	-1,08914	0,993	0,180	0,123	0,084	0,051	0,035	0,024	0,010	0,007	0,002
Foggia	0,25962	-1,10469	0,994	0,192	0,131	0,089	0,054	0,037	0,025	0,010	0,007	0,002
Forlì-Cesena	-0,06947	-1,07603	0,997	0,171	0,118	0,081	0,050	0,034	0,023	0,010	0,007	0,002
Frosinone	-0,11685	-1,13373	0,985	0,152	0,103	0,069	0,041	0,028	0,019	0,008	0,005	0,001
Genova	0,30813	-1,08573	0,998	0,203	0,140	0,096	0,058	0,040	0,027	0,011	0,008	0,002
Gorizia	0,38100	-1,13575	0,995	0,194	0,131	0,089	0,053	0,035	0,024	0,010	0,006	0,002
Grosseto	-0,40832	-1,07241	0,988	0,145	0,100	0,069	0,042	0,029	0,020	0,008	0,006	0,002
Imperia	0,26181	-1,11513	0,998	0,189	0,129	0,087	0,052	0,036	0,024	0,010	0,007	0,002
Isernia	-0,61925	-1,12237	0,961	0,121	0,082	0,055	0,033	0,022	0,015	0,006	0,004	0,001
La Spezia	0,04577	-1,09562	0,989	0,175	0,120	0,082	0,050	0,034	0,023	0,010	0,007	0,002
L'Aquila	-1,15382	-1,06234	0,949	0,102	0,070	0,049	0,030	0,021	0,014	0,006	0,004	0,001
Latina	-0,19481	-1,07158	0,993	0,162	0,112	0,077	0,047	0,032	0,022	0,009	0,007	0,002
Lecce	-0,22982	-1,12419	0,990	0,146	0,099	0,067	0,040	0,027	0,018	0,007	0,005	0,001
Lecco	-0,96151	-1,09607	0,980	0,106	0,072	0,050	0,030	0,021	0,014	0,006	0,004	0,001
Livorno	0,21326	-1,08411	0,994	0,194	0,133	0,092	0,056	0,038	0,026	0,011	0,008	0,002
Lodi	-0,26887	-1,14626	0,985	0,138	0,093	0,062	0,037	0,025	0,017	0,007	0,004	0,001
Lucca	0,21676	-1,09563	0,999	0,191	0,131	0,089	0,054	0,037	0,025	0,010	0,007	0,002
Macerata	-0,53326	-1,07576	0,994	0,136	0,093	0,064	0,039	0,027	0,019	0,008	0,005	0,002
Mantova	-1,86969	-0,99640	0,942	0,079	0,056	0,040	0,025	0,018	0,013	0,006	0,004	0,001
Massa-Carrara	0,37674	-1,10365	0,998	0,204	0,139	0,095	0,057	0,039	0,027	0,011	0,007	0,002
Matera	-0,64228	-1,08748	0,960	0,126	0,086	0,059	0,036	0,025	0,017	0,007	0,005	0,001

Fonte: Direzione centrale dei Censimenti generali (DCCG), Istat.

Tavola B.9 (segue).

Provincia	Parametri		R ²	Frequenze									
	Alfa	Beta		25	50	100	250	500	1.000	5.000	10.000	100.000	
Medio Campidano	-1,17848	-1,08341	0,983	0,097	0,067	0,046	0,028	0,019	0,013	0,005	0,004	0,001	
Messina	-0,18389	-1,10326	0,988	0,154	0,105	0,072	0,043	0,030	0,020	0,008	0,006	0,002	
Milano	0,54171	-1,10976	0,998	0,220	0,150	0,102	0,061	0,042	0,028	0,012	0,008	0,002	
Modena	-0,12487	-1,07600	0,997	0,166	0,115	0,079	0,048	0,033	0,023	0,010	0,007	0,002	
Monza e della Brianza	0,41208	-1,11362	0,999	0,205	0,139	0,095	0,057	0,039	0,026	0,011	0,007	0,002	
Napoli	0,51904	-1,10114	0,998	0,220	0,150	0,103	0,062	0,042	0,029	0,012	0,008	0,002	
Novara	-0,51493	-1,07651	0,987	0,137	0,094	0,065	0,040	0,027	0,019	0,008	0,005	0,002	
Nuoro	-0,78752	-1,07889	0,974	0,119	0,082	0,056	0,034	0,024	0,016	0,007	0,005	0,001	
Ogliastra	-	-	-	-	-	-	-	-	-	-	-	-	
Olbia-Tempio	-0,39683	-1,07858	0,987	0,145	0,099	0,068	0,042	0,029	0,020	0,008	0,006	0,002	
Oriстано	-0,68273	-1,12461	0,975	0,116	0,079	0,053	0,032	0,022	0,015	0,006	0,004	0,001	
Padova	-0,57626	-1,07647	0,984	0,133	0,091	0,063	0,038	0,026	0,018	0,008	0,005	0,002	
Palermo	0,45766	-1,10805	0,995	0,211	0,144	0,098	0,059	0,040	0,027	0,011	0,008	0,002	
Parma	-0,13505	-1,09269	0,990	0,161	0,110	0,076	0,046	0,031	0,021	0,009	0,006	0,002	
Pavia	-0,40307	-1,08857	0,990	0,142	0,097	0,067	0,040	0,028	0,019	0,008	0,005	0,002	
Perugia	-0,65130	-1,04689	0,993	0,134	0,093	0,065	0,040	0,028	0,019	0,008	0,006	0,002	
Pesaro e Urbino	-0,45593	-1,07443	0,991	0,141	0,097	0,067	0,041	0,028	0,019	0,008	0,006	0,002	
Pescara	0,06834	-1,08458	0,994	0,181	0,124	0,085	0,052	0,036	0,024	0,010	0,007	0,002	
Piacenza	-0,48785	-1,07314	0,984	0,139	0,096	0,066	0,040	0,028	0,019	0,008	0,006	0,002	
Pisa	-0,09002	-1,09969	0,994	0,163	0,111	0,076	0,046	0,031	0,021	0,009	0,006	0,002	
Pistoia	-0,15321	-1,09301	0,995	0,159	0,109	0,075	0,045	0,031	0,021	0,009	0,006	0,002	
Pordenone	-0,97748	-1,08010	0,984	0,108	0,074	0,051	0,031	0,021	0,015	0,006	0,004	0,001	
Potenza	-1,14606	-1,08728	0,958	0,098	0,067	0,046	0,028	0,019	0,013	0,005	0,004	0,001	
Prato	0,44440	-1,09587	0,998	0,214	0,146	0,100	0,061	0,041	0,028	0,012	0,008	0,002	
Ragusa	0,36891	-1,10874	0,996	0,202	0,137	0,094	0,056	0,038	0,026	0,011	0,007	0,002	
Ravenna	-0,11470	-1,08566	0,994	0,165	0,113	0,078	0,047	0,032	0,022	0,009	0,006	0,002	
Reggio di Calabria	-0,53624	-1,08412	0,969	0,134	0,092	0,063	0,038	0,026	0,018	0,008	0,005	0,001	
Reggio nell'Emilia	-0,51373	-1,07287	0,988	0,138	0,095	0,065	0,040	0,028	0,019	0,008	0,006	0,002	
Rieti	-0,30769	-1,12815	0,957	0,140	0,094	0,064	0,038	0,026	0,017	0,007	0,005	0,001	
Rimini	0,03102	-1,08957	0,998	0,176	0,121	0,083	0,050	0,034	0,024	0,010	0,007	0,002	
Roma	0,42656	-1,09305	0,999	0,213	0,146	0,100	0,061	0,041	0,028	0,012	0,008	0,002	
Rovigo	-0,59014	-1,09461	0,982	0,128	0,087	0,060	0,036	0,025	0,017	0,007	0,005	0,001	
Salerno	0,03732	-1,10630	0,992	0,172	0,117	0,080	0,048	0,033	0,022	0,009	0,006	0,002	
Sassari	0,03151	-1,09583	0,989	0,174	0,119	0,081	0,049	0,034	0,023	0,010	0,007	0,002	
Savona	-0,49681	-1,10101	0,976	0,133	0,091	0,062	0,037	0,025	0,017	0,007	0,005	0,001	
Siena	-0,93336	-1,02978	0,987	0,120	0,084	0,059	0,037	0,026	0,018	0,008	0,005	0,002	
Siracusa	0,45211	-1,11061	0,996	0,210	0,143	0,097	0,058	0,040	0,027	0,011	0,008	0,002	
Sondrio	-1,47198	-1,06620	0,917	0,086	0,060	0,041	0,025	0,017	0,012	0,005	0,004	0,001	
Taranto	0,34051	-1,10746	0,996	0,199	0,136	0,093	0,056	0,038	0,026	0,011	0,007	0,002	
Teramo	-0,71723	-1,07502	0,985	0,124	0,085	0,059	0,036	0,025	0,017	0,007	0,005	0,001	
Terni	-0,12782	-1,08397	0,992	0,164	0,113	0,077	0,047	0,032	0,022	0,009	0,006	0,002	
Torino	0,19625	-1,09672	0,989	0,189	0,129	0,088	0,053	0,037	0,025	0,010	0,007	0,002	
Trapani	0,41646	-1,10831	0,998	0,207	0,141	0,096	0,058	0,039	0,027	0,011	0,007	0,002	
Trento	-1,15854	-1,03754	0,969	0,105	0,074	0,051	0,032	0,022	0,016	0,007	0,005	0,001	
Treviso	-0,62977	-1,08173	0,989	0,128	0,088	0,060	0,037	0,025	0,017	0,007	0,005	0,001	
Trieste	0,48331	-1,09109	0,997	0,220	0,151	0,103	0,063	0,043	0,029	0,012	0,008	0,002	
Udine	-0,46040	-1,12152	0,980	0,131	0,089	0,060	0,036	0,024	0,017	0,007	0,005	0,001	
Varese	-0,19436	-1,10826	0,995	0,152	0,104	0,071	0,043	0,029	0,020	0,008	0,006	0,002	
Venezia	-0,14172	-1,07768	0,990	0,164	0,113	0,078	0,048	0,033	0,023	0,009	0,007	0,002	
Verbania	-0,63284	-1,11407	0,988	0,121	0,082	0,056	0,034	0,023	0,016	0,006	0,004	0,001	
Vercelli	-0,71525	-1,06954	0,980	0,125	0,086	0,060	0,037	0,025	0,017	0,007	0,005	0,001	
Verona	-0,77994	-1,04538	0,982	0,126	0,088	0,061	0,038	0,026	0,018	0,008	0,005	0,002	
Vibo Valentia	-0,78858	-1,15152	0,945	0,106	0,071	0,048	0,028	0,019	0,013	0,005	0,003	0,001	
Vicenza	-0,38051	-1,08511	0,990	0,144	0,099	0,068	0,041	0,028	0,019	0,008	0,006	0,002	
Viterbo	-0,74005	-1,11908	0,979	0,114	0,077	0,053	0,031	0,021	0,014	0,006	0,004	0,001	

Fonte: Direzione centrale dei Censimenti generali (DCCG), Istat.

Tavola B.10 – Parametri del modello, indice di adattamento e valori interpolati degli errori campionari relativi percentuali delle stime riferite agli alloggi per Provincia.

Provincia	Parametri		R ²	Frequenze								
	Alfa	Beta		25	50	100	250	500	1.000	5.000	10.000	100.000
Agrigento	-0,02455	-1,08056	0,993	0,174	0,119	0,082	0,050	0,034	0,024	0,010	0,007	0,002
Alessandria	0,06514	-1,09437	0,978	0,178	0,121	0,083	0,050	0,034	0,024	0,010	0,007	0,002
Ancona	0,21971	-1,08347	0,992	0,192	0,131	0,090	0,054	0,037	0,025	0,011	0,007	0,002
Aosta	-0,72228	-1,05497	0,967	0,128	0,089	0,061	0,038	0,026	0,018	0,008	0,005	0,002
Arezzo	-0,03673	-1,12249	0,991	0,161	0,109	0,074	0,044	0,030	0,020	0,008	0,006	0,002
Ascoli	0,17882	-1,10391	0,992	0,185	0,126	0,086	0,052	0,035	0,024	0,010	0,007	0,002
Asti	-0,21629	-1,10615	0,954	0,151	0,103	0,070	0,042	0,029	0,020	0,008	0,006	0,002
Avellino	-1,39901	-1,08855	0,937	0,086	0,059	0,041	0,025	0,017	0,012	0,005	0,003	0,001
Bari	0,55323	-1,08957	0,998	0,228	0,157	0,107	0,065	0,045	0,031	0,013	0,009	0,002
Barletta - Andria - Trani	0,65427	-1,09142	0,998	0,239	0,164	0,112	0,068	0,047	0,032	0,013	0,009	0,003
Belluno	-0,92084	-1,05960	0,973	0,115	0,079	0,055	0,034	0,023	0,016	0,007	0,005	0,001
Benevento	-0,84643	-1,08804	0,938	0,114	0,078	0,053	0,032	0,022	0,015	0,006	0,004	0,001
Bergamo	-0,62118	-1,09842	0,986	0,125	0,086	0,058	0,035	0,024	0,016	0,007	0,005	0,001
Biella	-0,26714	-1,10850	0,978	0,147	0,100	0,068	0,041	0,028	0,019	0,008	0,005	0,001
Bologna	0,27199	-1,09554	0,993	0,196	0,134	0,092	0,056	0,038	0,026	0,011	0,007	0,002
Bolzano	-0,34367	-1,07652	0,955	0,149	0,103	0,071	0,043	0,030	0,020	0,009	0,006	0,002
Brescia	-0,51572	-1,09718	0,987	0,132	0,090	0,062	0,037	0,026	0,017	0,007	0,005	0,001
Brindisi	0,46301	-1,11715	0,992	0,209	0,142	0,096	0,058	0,039	0,027	0,011	0,007	0,002
Cagliari	0,22029	-1,09512	0,993	0,195	0,134	0,092	0,056	0,039	0,026	0,011	0,008	0,002
Caltanissetta	0,46492	-1,09982	0,996	0,215	0,147	0,100	0,061	0,041	0,028	0,012	0,008	0,002
Campobasso	-0,42472	-1,07394	0,971	0,144	0,099	0,068	0,042	0,029	0,020	0,008	0,006	0,002
Carbonia-Iglesias	-0,54626	-1,03733	0,976	0,143	0,100	0,070	0,043	0,030	0,021	0,009	0,006	0,002
Caserta	-0,21925	-1,07549	0,995	0,159	0,109	0,075	0,046	0,032	0,022	0,009	0,006	0,002
Catania	0,46722	-1,07782	0,997	0,216	0,148	0,101	0,061	0,042	0,029	0,012	0,008	0,002
Catanzaro	-0,17566	-1,08422	0,991	0,160	0,110	0,075	0,046	0,032	0,022	0,009	0,006	0,002
Chieti	-0,11630	-1,08960	0,988	0,163	0,112	0,077	0,047	0,032	0,022	0,009	0,006	0,002
Como	-0,39341	-1,09251	0,991	0,142	0,097	0,066	0,040	0,028	0,019	0,008	0,005	0,002
Cosenza	-0,90180	-1,06744	0,983	0,114	0,079	0,055	0,033	0,023	0,016	0,007	0,005	0,001
Cremona	-0,27612	-1,09642	0,981	0,149	0,102	0,070	0,042	0,029	0,020	0,008	0,006	0,002
Crotone	-0,14439	-1,11972	0,970	0,153	0,104	0,071	0,042	0,029	0,019	0,008	0,005	0,001
Cuneo	-0,64527	-1,07118	0,968	0,129	0,089	0,061	0,038	0,026	0,018	0,008	0,005	0,002
Enna	-0,74827	-1,06430	0,978	0,124	0,086	0,059	0,036	0,025	0,017	0,007	0,005	0,002
Fermo	-0,34584	-1,10264	0,982	0,143	0,097	0,066	0,040	0,027	0,019	0,008	0,005	0,001
Ferrara	-0,06330	-1,09673	0,985	0,166	0,113	0,078	0,047	0,032	0,022	0,009	0,006	0,002
Firenze	0,26800	-1,09084	0,993	0,198	0,135	0,093	0,056	0,039	0,026	0,011	0,008	0,002
Foggia	0,15873	-1,07705	0,994	0,191	0,132	0,091	0,055	0,038	0,026	0,011	0,008	0,002
Forlì-Cesena	0,12656	-1,08013	0,994	0,187	0,129	0,089	0,054	0,037	0,026	0,011	0,007	0,002
Frosinone	-0,31058	-1,10154	0,991	0,145	0,099	0,068	0,041	0,028	0,019	0,008	0,005	0,002
Genova	0,53820	-1,08730	0,999	0,227	0,156	0,107	0,065	0,045	0,031	0,013	0,009	0,003
Gorizia	0,22244	-1,10593	0,989	0,188	0,128	0,088	0,053	0,036	0,025	0,010	0,007	0,002
Grosseto	-0,48992	-1,04277	0,981	0,146	0,102	0,071	0,044	0,031	0,021	0,009	0,006	0,002
Imperia	0,22754	-1,09387	0,995	0,193	0,132	0,090	0,055	0,037	0,026	0,011	0,007	0,002
Isernia	-0,77216	-1,08834	0,966	0,118	0,081	0,055	0,034	0,023	0,016	0,007	0,005	0,001
La Spezia	-0,04947	-1,06335	0,993	0,176	0,122	0,084	0,052	0,036	0,025	0,011	0,007	0,002
L'Aquila	-1,17487	-1,04760	0,963	0,103	0,072	0,050	0,031	0,021	0,015	0,006	0,004	0,001
Latina	-0,04717	-1,07009	0,991	0,175	0,120	0,083	0,051	0,035	0,024	0,010	0,007	0,002
Lecce	-0,13504	-1,12202	0,989	0,154	0,104	0,071	0,042	0,029	0,019	0,008	0,005	0,001
Lecco	-0,54108	-1,13504	0,982	0,123	0,083	0,056	0,033	0,022	0,015	0,006	0,004	0,001
Livorno	0,07809	-1,04698	0,988	0,193	0,134	0,093	0,058	0,040	0,028	0,012	0,008	0,003
Lodi	-0,17775	-1,14892	0,983	0,144	0,097	0,065	0,038	0,026	0,017	0,007	0,005	0,001
Lucca	0,46730	-1,09767	0,997	0,212	0,145	0,099	0,059	0,040	0,028	0,011	0,008	0,002
Macerata	-0,19931	-1,09491	0,991	0,155	0,106	0,073	0,044	0,030	0,021	0,009	0,006	0,002
Mantova	-0,40024	-1,14482	0,970	0,130	0,087	0,059	0,035	0,023	0,016	0,006	0,004	0,001
Massa-Carrara	0,42148	-1,09282	0,995	0,213	0,146	0,100	0,060	0,041	0,028	0,012	0,008	0,002
Matera	-0,81621	-1,04969	0,945	0,123	0,085	0,059	0,037	0,025	0,018	0,008	0,005	0,002

Fonte: Direzione centrale dei Censimenti generali (DCCG), Istat.

Tavola B.10 (segue).

Provincia	Parametri		R ²	Frequenze									
	Alfa	Beta		25	50	100	250	500	1.000	5.000	10.000	100.000	
Medio Campidano	-0,64766	-1,14195	0,972	0,115	0,077	0,052	0,031	0,021	0,014	0,006	0,004	0,001	
Messina	-0,47464	-1,05930	0,977	0,143	0,099	0,069	0,042	0,029	0,020	0,009	0,006	0,002	
Milano	0,70073	-1,09917	0,999	0,239	0,163	0,111	0,067	0,046	0,031	0,013	0,009	0,002	
Modena	0,11503	-1,08406	0,995	0,185	0,127	0,087	0,053	0,036	0,025	0,010	0,007	0,002	
Monza e della Brianza	0,37897	-1,09466	0,998	0,208	0,142	0,097	0,059	0,040	0,028	0,011	0,008	0,002	
Napoli	0,70119	-1,10675	0,999	0,242	0,165	0,113	0,068	0,047	0,032	0,013	0,009	0,003	
Novara	-0,41694	-1,06641	0,988	0,146	0,101	0,070	0,043	0,030	0,020	0,009	0,006	0,002	
Nuoro	-0,58201	-1,08703	0,975	0,130	0,089	0,061	0,037	0,026	0,018	0,007	0,005	0,001	
Ogliastra	-	-	-	-	-	-	-	-	-	-	-	-	
Olbia-Tempio	-0,38625	-1,06015	0,983	0,150	0,104	0,072	0,044	0,031	0,021	0,009	0,006	0,002	
Oristano	-0,69926	-1,11090	0,954	0,118	0,080	0,055	0,033	0,022	0,015	0,006	0,004	0,001	
Padova	-0,19599	-1,10174	0,983	0,154	0,105	0,072	0,043	0,030	0,020	0,008	0,006	0,002	
Palermo	0,43105	-1,08853	0,998	0,215	0,148	0,101	0,061	0,042	0,029	0,012	0,008	0,002	
Parma	-0,04952	-1,08385	0,986	0,170	0,117	0,080	0,049	0,034	0,023	0,010	0,007	0,002	
Pavia	-0,01914	-1,11653	0,982	0,164	0,112	0,076	0,045	0,031	0,021	0,009	0,006	0,002	
Perugia	-0,31539	-1,06266	0,990	0,154	0,107	0,074	0,045	0,031	0,022	0,009	0,006	0,002	
Pesaro e Urbino	-0,71115	-1,02717	0,974	0,134	0,094	0,066	0,041	0,029	0,020	0,009	0,006	0,002	
Pescara	0,25771	-1,08790	0,994	0,197	0,135	0,093	0,056	0,039	0,027	0,011	0,008	0,002	
Piacenza	-0,25503	-1,08118	0,973	0,154	0,106	0,073	0,044	0,031	0,021	0,009	0,006	0,002	
Pisa	0,16578	-1,11118	0,993	0,182	0,124	0,084	0,051	0,034	0,023	0,010	0,007	0,002	
Pistoia	-0,11369	-1,08294	0,990	0,165	0,114	0,078	0,048	0,033	0,022	0,009	0,006	0,002	
Pordenone	-0,67820	-1,10092	0,970	0,121	0,083	0,056	0,034	0,023	0,016	0,007	0,004	0,001	
Potenza	-1,49533	-1,03147	0,948	0,090	0,063	0,044	0,027	0,019	0,013	0,006	0,004	0,001	
Prato	0,27896	-1,05686	0,989	0,210	0,145	0,101	0,062	0,043	0,030	0,013	0,009	0,003	
Ragusa	0,42371	-1,09894	0,997	0,211	0,144	0,098	0,059	0,041	0,028	0,011	0,008	0,002	
Ravenna	0,05318	-1,08475	0,991	0,179	0,123	0,084	0,051	0,035	0,024	0,010	0,007	0,002	
Reggio di Calabria	-0,48616	-1,07788	0,981	0,138	0,095	0,066	0,040	0,028	0,019	0,008	0,005	0,002	
Reggio nell'Emilia	-0,31385	-1,08044	0,969	0,150	0,103	0,071	0,043	0,030	0,020	0,009	0,006	0,002	
Rieti	-0,92339	-1,03941	0,954	0,118	0,083	0,058	0,036	0,025	0,017	0,008	0,005	0,002	
Rimini	0,14836	-1,08893	0,996	0,187	0,128	0,088	0,053	0,037	0,025	0,010	0,007	0,002	
Roma	0,46740	-1,10766	0,999	0,223	0,153	0,106	0,064	0,044	0,031	0,013	0,009	0,003	
Rovigo	-0,49677	-1,09158	0,975	0,135	0,092	0,063	0,038	0,026	0,018	0,007	0,005	0,001	
Salerno	0,02237	-1,08894	0,994	0,175	0,120	0,082	0,050	0,034	0,024	0,010	0,007	0,002	
Sassari	0,03320	-1,08249	0,984	0,178	0,122	0,084	0,051	0,035	0,024	0,010	0,007	0,002	
Savona	-0,44220	-1,08484	0,986	0,140	0,096	0,066	0,040	0,028	0,019	0,008	0,005	0,002	
Siena	-0,17850	-1,09922	0,990	0,156	0,107	0,073	0,044	0,030	0,021	0,008	0,006	0,002	
Siracusa	0,49035	-1,09792	0,998	0,218	0,149	0,102	0,062	0,042	0,029	0,012	0,008	0,002	
Sondrio	-1,02146	-1,11106	0,959	0,100	0,068	0,046	0,028	0,019	0,013	0,005	0,004	0,001	
Taranto	0,50069	-1,10723	0,996	0,216	0,147	0,100	0,060	0,041	0,028	0,012	0,008	0,002	
Teramo	-0,59289	-1,07353	0,982	0,132	0,091	0,063	0,038	0,026	0,018	0,008	0,005	0,002	
Terni	-0,11957	-1,06906	0,982	0,169	0,116	0,080	0,049	0,034	0,023	0,010	0,007	0,002	
Torino	0,42827	-1,09885	0,993	0,211	0,144	0,099	0,060	0,041	0,028	0,011	0,008	0,002	
Trapani	0,35973	-1,08544	0,996	0,209	0,143	0,098	0,060	0,041	0,028	0,012	0,008	0,002	
Trento	-0,45927	-1,09515	0,978	0,136	0,093	0,064	0,039	0,026	0,018	0,007	0,005	0,001	
Treviso	-0,50092	-1,08382	0,985	0,136	0,093	0,064	0,039	0,027	0,018	0,008	0,005	0,002	
Trieste	0,48411	-1,07238	0,995	0,227	0,156	0,108	0,066	0,045	0,031	0,013	0,009	0,003	
Udine	-0,51047	-1,10300	0,976	0,131	0,090	0,061	0,037	0,025	0,017	0,007	0,005	0,001	
Varese	-0,33845	-1,07721	0,992	0,149	0,103	0,071	0,043	0,030	0,020	0,009	0,006	0,002	
Venezia	0,28911	-1,10705	0,993	0,195	0,133	0,090	0,054	0,037	0,025	0,010	0,007	0,002	
Verbania	-0,45001	-1,12400	0,982	0,131	0,089	0,060	0,036	0,024	0,016	0,007	0,005	0,001	
Vercelli	-0,18106	-1,11603	0,975	0,152	0,103	0,070	0,042	0,028	0,019	0,008	0,005	0,001	
Verona	-0,33049	-1,07493	0,984	0,150	0,104	0,071	0,044	0,030	0,021	0,009	0,006	0,002	
Vibo Valentia	-1,05168	-1,11030	0,938	0,099	0,067	0,046	0,028	0,019	0,013	0,005	0,004	0,001	
Vicenza	-0,21831	-1,08659	0,985	0,156	0,107	0,073	0,045	0,031	0,021	0,009	0,006	0,002	
Viterbo	-1,24918	-1,04582	0,966	0,099	0,069	0,048	0,030	0,021	0,014	0,006	0,004	0,001	

Fonte: Direzione centrale dei Censimenti generali (DCCG), Istat.