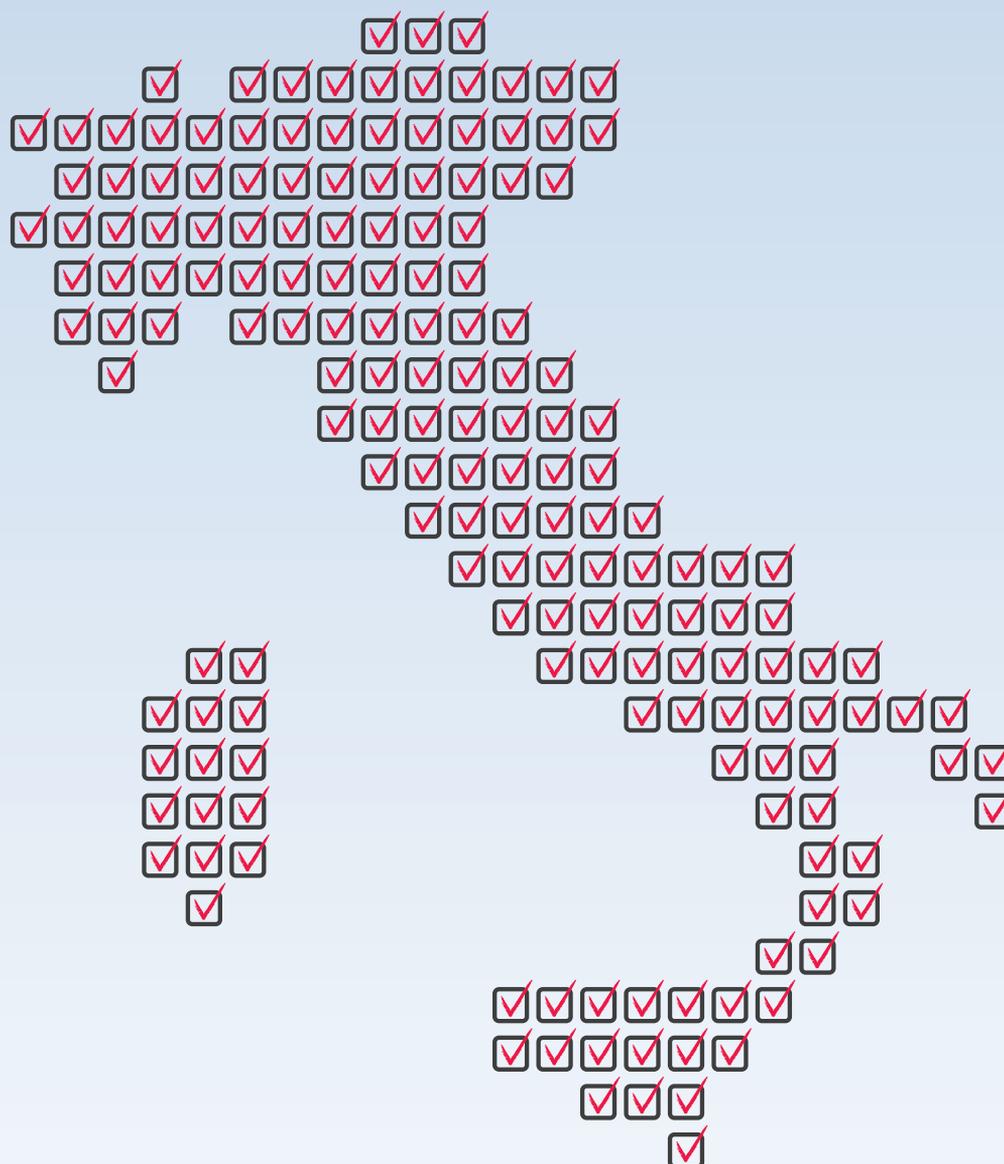


Atti del 15° Censimento generale della popolazione e delle abitazioni



6 La valutazione della qualità

Atti a cura di: Giuseppe Sindoni

Fascicolo a cura di: Patrizia Grossi e Matteo Mazziotta

Atti del 15° Censimento generale della popolazione e delle abitazioni
6 - LA VALUTAZIONE DELLA QUALITÀ

© 2016

Istituto nazionale di statistica
Via Cesare Balbo, 16 - Roma

ISBN 978-88-458-1885-1 (elettronico)

ISBN 978-88-458-1884-4 (stampa)

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

	Pagina
Presentazione dell'opera	7
Il 15° Censimento generale della popolazione e delle abitazioni	9
Introduzione	13
PARTE PRIMA	
L'ORGANIZZAZIONE DELLA RILEVAZIONE	
Capitolo 1 - Quadro normativo e finanziario	19
1.1 Cenni storici ed obiettivi	19
1.2 Quadro normativo	23
1.3 Aspetti finanziari e amministrativi	26
1.4 Gestione dei servizi in outsourcing	27
Capitolo 2 - Organizzazione e compiti della rete	29
2.1 Organizzazione della rilevazione	29
2.2 Ruolo e compiti dell'Istat	39
2.3 Ruolo e compiti della rete e operazioni sul campo	41
2.4 Formazione	46
2.5 Monitoraggio del materiale della rilevazione	50
2.6 Monitoraggio della rilevazione sul campo	53
Capitolo 3 - La rilevazione sul campo: metodi e strumenti	59
3.1 Unità di rilevazione	59
3.2 Questionario	60
3.3 Manuale	61
3.4 Strumenti ed operazioni di ausilio alla rilevazione sul campo	63
Capitolo 4 - Applicazioni geografiche per la rilevazione	67
4.1 Ripartizione della popolazione censita al 2001 nelle sezioni del 2011	67
4.2 Abbinamento delle sezioni 2001-2011 per la preparazione cartografica a supporto della PES	74
4.3 Strumenti cartografici e basi di dati geografiche a supporto della rilevazione	80
4.4 Elaborati cartografici	84
4.5 Itinerari di sezione a supporto della rilevazione	87
Capitolo 5 - Il sistema informatico a supporto della rilevazione (SGRPES)	93
5.1 Introduzione al sistema	93
5.2 Profili applicativi e visibilità territoriale	96
5.3 Flusso di lavorazione del questionario	99
5.4 Lista e modelli riepilogativi	100
5.5 Monitoraggio del lavoro sul campo	102
5.6 Rapporti riassuntivi	104
5.7 Test	106
5.8 Utilizzo del sistema	109

PARTE SECONDA

GLI ASPETTI METODOLOGICI E I RISULTATI

Capitolo 1 - Disegno di campionamento, processo di stima e attendibilità delle stime prodotte	115
1.1 Obiettivi dell'indagine e parametri di interesse	115
1.2 Popolazione obiettivo e domini di stima	115
1.3 Errori di copertura	116
1.4 Disegno di campionamento	118
1.4.1 Caratteristiche generali	118
1.4.2 Scelta delle variabili di stratificazione	119
1.4.3 Determinazione della numerosità campionaria e l'allocazione del campione tra gli strati di primo stadio	120
1.4.4 Formazione del campione di primo stadio	121
1.4.5 Valutazione degli errori relativi attesi	124
1.4.6 Formazione del campione di secondo stadio	125
1.5 Procedura di stima	127
1.5.1 Caratteristiche generali	127
1.5.2 Determinazione dei pesi campionari	128
1.5.3 Modello di stima dell'errore di sovra-copertura	129
1.5.4 Dual-System Estimator corretto per sovra-copertura	133
1.5.5 Regression Estimator	138
1.5.6 Sample Balance Adjustment	139
1.5.7 Determinazione delle stime del totale "vero" (incognito) della popolazione residente a livello regionale e nazionale	140
1.6 HtC Index	141
1.7 Attendibilità delle stime dei tassi di sotto e sovra-copertura	145
Capitolo 2 - Fasi post-rilevazione	147
2.1 Obiettivi	147
2.2 Costruzione della base dati 'PES'	147
2.3 Costruzione dell'archivio 'CENS'	151
2.4 Controllo e correzione dei microdati	151
2.4.1 Principali variabili	152
2.4.2 Normalizzazione degli indirizzi	153
2.4.3 Corretta localizzazione degli individui	153
2.5 Analisi dell'eleggibilità	154
2.6 I numeri della rilevazione	156
Capitolo 3 - Le procedure di abbinamento	159
3.1 Quadro generale e sintesi della strategia	159
3.2 Attività preliminari	161
3.2.1 Riconciliazione degli schemi	161
3.2.2 Deduplicazione	162
3.3 Metodi di linkage	162
3.3.1 Record linkage deterministico	163
3.3.2 Record linkage probabilistico	163
3.3.3 Revisione manuale dei casi ambigui e la ricerca manuale dei non abbinati	165
3.4 RELAIS: uno strumento per l'abbinamento	166
3.5 Strategia di abbinamento adottata	167
3.6 Confronto con gli archivi amministrativi e lista pre-censuaria	170
3.7 I risultati del linkage	172

Capitolo 4 - I risultati della PES	175
4.1 Livello nazionale	175
4.2 Livello regionale	177
4.3 Variabilità delle stime	178
4.4 La PES in Italia, i confronti nel tempo	178
Conclusioni	181
Riferimenti bibliografici	183
Riferimenti legislativi	186
Appendice 2.A	187

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. *I contenuti informativi della rilevazione, la validazione e diffusione 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 *I contenuti informativi della rilevazione, la validazione e diffusione 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.

¹ <http://www.istat.it/it/archivio/174654>

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 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

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

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.

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 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 (8 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 on line. La cifra assoluta di questionari compilati on line 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

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.

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

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 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 fascicolo 4 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 on line 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

⁹ <http://www.istat.it/it/censimento-popolazione/censimento-popolazione-2011>

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

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. I fascicoli 2 e 5 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 fascicolo 6 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 sottocopertura a livello nazionale pari all'1,45 per cento della popolazione censita e un tasso di sovracopertura 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 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

L'Istat, nel corso del 2012, ha svolto la "Rilevazione campionaria areale di qualità della copertura del 15° Censimento generale della popolazione e delle abitazioni" (secondo la terminologia anglosassone, *Post Enumeration Survey*, PES 2011).

Scopo della rilevazione è di determinare gli indicatori di qualità e di copertura del 15° Censimento generale della popolazione e delle abitazioni richiesti dal regolamento (CE) n. 763/08 del Parlamento Europeo e del Consiglio dell'8 luglio 2008 e stabiliti dal Regolamento (UE) n. 1151/2010 della Commissione Europea dell'8 dicembre 2010.

Con la PES 2011 l'Istat è alla sua quarta indagine di copertura. La prima si è avuta in occasione del Censimento della popolazione e delle abitazioni del 1981, proseguendo poi, con la cadenza decennale dei censimenti generali, rispettivamente nel 1991 e 2001.

La PES viene condotta da numerosi stati dell'area OCSE al termine delle operazioni censuarie. Negli ultimi anni, le tecniche di rilevazione della PES e le metodologie statistiche per la stima dei parametri incogniti sono stati oggetto di un fervente dibattito scientifico e molti documenti sono disponibili in letteratura.

La PES del 15° Censimento della popolazione italiano ha il principale obiettivo di stimare il numero di individui abitualmente dimoranti nel periodo di riferimento (9 Ottobre 2011), il tasso di copertura e il tasso di sovra-copertura DEL Censimento.

La PES si basa su un disegno campionario areale a due stadi stratificati in cui: il primo stadio è un campione di 252 comuni, stratificati per Regione e cinque classi di ampiezza demografica; il secondo stadio è composto da 2.500 sezioni di Censimento stratificate per ampiezza demografica della sezione stessa.

L'indagine è disegnata per garantire l'indipendenza tra le due rilevazioni (Censimento e PES) in modo da consentire l'applicazione di modelli di stima dei parametri basati sul classico "*dual system*". L'indagine è stata condotta da aprile a luglio 2012 ed ha coinvolto circa 1.200 rilevatori selezionati tra i migliori del Censimento (a ciascuno di loro è stata assegnata una sezione differente rispetto a quella del Censimento). La PES ha enumerato circa 329 mila unità eleggibili; una accurata fase di *record linkage* e l'applicazione di una complessa metodologia di stima (derivata dall'ONS e modificata dai ricercatori dell'Istat), ha prodotto i risultati finali presentati nella parte seconda del presente volume.

La PES, oltre ad essere la più grande indagine censuaria di qualità mai condotta in Italia, è stata realizzata attraverso un sistema informatico in grado di garantire l'indipendenza statistica con il Censimento, oltre al monitoraggio in tempo reale delle operazioni sul territorio.

Vista la rilevanza e la complessità della rilevazione, il presente documento illustra l'iter normativo percorso, nonché le principali soluzioni organizzative, tecniche e metodologiche. Il lavoro è stato realizzato in maniera unitaria, essendo evidenti le interconnessioni delle fasi e dei processi produttivi, ma ciascun paragrafo è autonomo ed è stato realizzato per affrontare in maniera esauriente il tema trattato.

I cinque capitoli della parte prima riguardano

1. *Quadro normativo e finanziario;*
2. *Organizzazione e compiti della rete;*
3. *La rilevazione sul campo: metodi e strumenti;*

¹¹ L'Introduzione è a cura di Patrizia Grossi e Matteo Mazziotta.

4. *Applicazioni geografiche per la rilevazione;*
5. *Il sistema informatico SGRPES.*

Il capitolo primo, *Quadro normativo e finanziario*, contiene cenni storici sulle rilevazioni di copertura dei censimenti a giustificazione delle scelte tecniche e metodologiche adottate, nonché le norme internazionali e nazionali sulla base delle quali si è progettata e realizzata questa rilevazione. Le competenze e le conoscenze delle persone che hanno collaborato a questa fase del processo sono di tipo gestionale-amministrativo.

Il capitolo secondo, *Organizzazione e compiti della rete*, descrive le linee guida adottate nell'organizzazione della rilevazione, la composizione e le caratteristiche della rete di rilevazione, nonché le modalità di definizione ed erogazione della formazione. Vengono altresì descritte le caratteristiche organizzative del sistema di monitoraggio della rilevazione. Le competenze e le conoscenze delle persone che hanno collaborato a questa fase del processo sono di tipo gestionale-organizzativo.

Il capitolo terzo, *La rilevazione sul campo: metodi e strumenti*, descrive le metodologie adottate e gli strumenti utilizzati per garantire un adeguato funzionamento della rilevazione sul territorio. Le competenze e le conoscenze delle persone che hanno collaborato a questa fase del processo sono di tipo tecnico-statistico.

Il capitolo quarto, *Applicazioni geografiche per la rilevazione*, descrive gli strumenti informativi geografici di ausilio all'estrazione delle unità statistiche da inserire nel campione, nonché l'utilità della Rilevazione dei numeri civici (RNC), degli strumenti cartografici e delle basi di dati geografiche a supporto della rilevazione. Le competenze e le conoscenze delle persone che hanno collaborato a questa fase del processo sono di tipo geografico-territoriale.

Il capitolo quinto, *Il sistema informatico SGRPES*, descrive l'architettura e le soluzioni informatiche adottate per coadiuvare gli operatori nelle diverse fasi del processo, garantendo un continuo monitoraggio della rilevazione per singolo ente territoriale coinvolto, con velocità e trasparenza. Le competenze e le conoscenze delle persone che hanno collaborato a questa fase del processo sono di tipo informatico.

I quattro capitoli della parte seconda riguardano:

1. *Il disegno di campionamento, il processo di stima e l'attendibilità delle stime prodotte;*
2. *Le fasi post-rilevazione;*
3. *Le procedure di abbinamento;*
4. *I risultati della PES.*

Il capitolo primo, *Il disegno di campionamento, il processo di stima e l'attendibilità delle stime prodotte*, contiene la sfida chiave della PES 2011 ovvero la ricerca di una metodologia che fosse in grado di produrre una stima attendibile del livello di sovra-copertura del Censimento e di come includere tale stima nel modello *dual system*. Le competenze e le conoscenze delle persone che hanno collaborato a questa fase del processo sono di tipo metodologico.

Il capitolo secondo, *Le fasi post-rilevazione*, descrive i controlli ed aggiustamenti si sono resi necessari al fine di disporre di valori validi per ciascun individuo, eliminando gli errori e le incongruenze. Nello specifico i controlli sono stati finalizzati al soddisfacimento delle condizioni necessarie per il corretto prosieguo del lavoro. Le competenze e le conoscenze delle persone che hanno collaborato a questa fase del processo sono di tipo metodologico.

Il capitolo terzo, *Le procedure di abbinamento*, descrive le metodologie adottate e gli strumenti utilizzati per assicurare la massima correttezza degli abbinamenti. Sono stati adottati modelli di record linkage probabilistici, iterati considerando differenti metriche per il confronto delle variabili comuni nei vari passi della procedura. Le competenze e le conoscenze delle persone che hanno collaborato a questa fase del processo sono di tipo informatico-metodologico.

Il capitolo quarto, *I risultati della PES*, illustra i risultati dell'indagine per livello territoriale, ossia dapprima il nazionale e poi con dettaglio regionale. Oltre ai risultati riferiti a questi domini sono riportate le stime della popolazione e i tassi di sovra e sottocopertura riferiti ad alcune variabili sociodemografiche quali l'età e la cittadinanza. In ultimo sono riportati gli errori campionari sotto forma di coefficiente di variazione per la stima della popolazione 'vera' e di intervallo di confidenza per i tassi di sottocopertura e sovracopertura. Le competenze e le conoscenze delle persone che hanno collaborato a questa fase del processo sono di tipo statistico-metodologico.

PARTE PRIMA

L'ORGANIZZAZIONE DELLA RILEVAZIONE

CAPITOLO 1

Quadro normativo e finanziario¹²

La Rilevazione campionaria areale di controllo della copertura e della qualità del 15° Censimento generale della popolazione e delle abitazioni (PES2011) viene condotta, al termine delle operazioni censuarie, per stimare l'incidenza degli errori di copertura del Censimento della popolazione, la cui presenza conduce a errori nel conteggio della popolazione residente.

La popolazione obiettivo della PES2011 è costituita dagli individui e dalle famiglie residenti in abitazioni situate sul territorio nazionale; la principale stima prodotta è rappresentata dall'errore di sottocopertura, cioè dalla distorsione per difetto del reale ammontare dei conteggi di popolazione a causa della pratica impossibilità di enumerare tutti gli individui. L'obiettivo è stimare il tasso di copertura del Censimento, relativamente a un campione di 2.488 sezioni di Censimento, che insiste su 255 comuni.

La data di riferimento dell'indagine è, analogamente al Censimento, quella del 9 ottobre 2011. Tuttavia, per intervenire sulla stessa popolazione, evitando al contempo di interferire con le normali operazioni censuarie, si è scelto di condurre la rilevazione in un momento immediatamente successivo alle operazioni censuarie sul campo e, nello specifico, dall'11 aprile al 30 giugno 2012. Dal punto di vista della tecnica di rilevazione, si è scelto di ricorrere alla modalità tradizionale del Censimento, per una enumerazione esaustiva delle famiglie e degli individui nelle sezioni di interesse, senza utilizzo di liste di partenza. Pertanto, i Rilevatori¹³ hanno consegnato il questionario di indagine alle famiglie e, successivamente, hanno provveduto al ritiro dello stesso, a compilazione avvenuta a cura dei rispondenti.

Il questionario della PES è stato progettato in coerenza con il formato del questionario breve (short form) proposto in occasione del Censimento.

1.1 Cenni storici ed obiettivi

A partire dal 1960, ogni tornata censuaria è stata "guidata" dalle Raccomandazioni delle Nazioni Unite in un'ottica di sincronizzazione delle operazioni di raccolta/produzione dei dati censuari e armonizzazione dei concetti, finalizzata alla comparabilità internazionale delle misure. Seppur nella forma del *Gentlemen's Agreement*, accordo informale, non vincolante per i singoli Stati, le Nazioni Unite forniscono indicazioni sui requisiti fondamentali di un Censimento, dalle principali variabili da rilevare alle definizioni e classificazioni da adottare, come le tavole da produrre (Statistical Commission & Economic Commission For Europe, 2006).

Nel 2000 le Raccomandazioni si focalizzavano sui contenuti informativi, con rare indicazioni metodologiche e appena un accenno all'uso degli archivi amministrativi circa l'integrazione o la sostituzione del questionario per la rilevazione sul campo. Al contrario, le Raccomandazioni del 2010 relative ai censimenti dedicavano ampio spazio agli aspetti metodologici e alle nuove tecnologie, ipotizzavano i possibili usi degli archivi amministrativi a fini censuari ed esaminavano

¹² I paragrafi sono a cura di:

- 1.1 Cenni storici ed obiettivi: Anna Maria Cecchini
- 1.2 Quadro normativo: Eleonora Ciocca
- 1.3 Aspetti finanziari e amministrativi: Sara Orecchini
- 1.4 Gestione dei servizi in outsourcing: Fabio Falcone

¹³ Per assicurare l'indipendenza tra le due occasioni di rilevazione, i Rilevatori impiegati dall'indagine di copertura, scelti tra quelli più esperti, verranno inviati in sezioni differenti da quelle del Censimento.

alcune tecniche campionarie e di valutazione della qualità dei dati, passando in rassegna vantaggi e svantaggi delle diverse strategie (United Nations Economic Commission for Europe & Statistical Office of the European Communities (EUROSTAT), 2006).

Nell'ambito dell'esecuzione dei censimenti, durante la rilevazione sul campo, possono verificarsi due tipi di errori statistici: da un lato, errori nei dati censiti, dall'altro, la rilevazione può non essere esaustiva con conseguente perdita di unità. Ecco, quindi, che diventa prioritario appurare e misurare la qualità dei dati censuari attraverso un'indagine di misura e un'indagine di copertura. Oggi, le indagini post-censuarie finalizzate a tale scopo, nonché alla misurazione dell'errore non campionario, sono previste da numerose fonti normative nazionali¹⁴ e internazionali¹⁵.

Con la PES 2011 l'Istat è alla sua quarta indagine di copertura. La prima si è avuta in occasione del Censimento della popolazione e delle abitazioni del 1981, proseguendo poi, con la cadenza decennale dei censimenti generali, rispettivamente nel 1991 e nel 2001. I risultati della prima indagine di controllo, effettuata nel mese di dicembre 1981 "allo scopo di valutare la qualità dei dati del Censimento della Popolazione", furono presentati due anni dopo in occasione del Convegno della Società Italiana di Statistica (SIS) di Trieste. Più precisamente, l'indagine è stata concepita per individuare i caratteri più soggetti ad "errore" e per quantificare tale "errore", oltre che per conoscere, relativamente ai caratteri rilevati, la distribuzione dell'errore sulle modalità dei caratteri stessi.

Seguirono due Commissioni il cui lavoro ebbe lo scopo di valutare la qualità dei censimenti, mentre nell'ambito del Servizio Studi dell'Istat nel 1984 veniva costituito il Progetto sul controllo della qualità dei dati, i cui risultati sarebbero stati pubblicati nel 1986 in un documento intitolato *Indagine interna sui controlli di qualità utilizzati nelle diverse indagini dell'ISTAT* (Masselli, 1986).

Nel novembre 1986, presso l'Università degli Studi di Padova, durante il Convegno di presentazione dei risultati del gruppo di ricerca Biggeri-Trivellato sull'*Attendibilità e tempestività delle stime di contabilità nazionale*, fu ospitata anche la relazione sul primo *error profile* (oggi meglio denominato profilo di qualità) di un'indagine statistica in Italia, dal titolo, *Una ricognizione dell'error profile dell'indagine ISTAT sul prodotto lordo*. La novità del documento consisteva nel fatto che, per la prima volta, un processo era esaminato dall'interno.

Nel tempo, l'approccio basato sulla qualità dei dati ha guadagnato via via il consenso del mondo accademico, grazie a diversi rapporti di ricerca, pubblicazioni e dibattiti scientifici sull'adeguamento di metodi e strumenti di calcolo già in uso a livello internazionale; in particolare, l'obiettivo era ricostituire la struttura della popolazione interna. Storicamente, tali metodi furono impiegati per la prima volta nell'Indagine sulla salute del 1983 e nell'Indagine sulle forze di lavoro degli ultimi anni Ottanta. Riguardo alle definizioni e classificazioni di mancata risposta, coerentemente alla letteratura internazionale, anche l'Istat si è attestato sulla tripartizione mancato contatto, rifiuto a rispondere, impossibilità a rispondere¹⁶.

¹⁴ Tra le tante: Programma statistico nazionale 2011-2013, aggiornamento 2012-2013; Piano Generale di Censimento, art.50 della Legge 30 luglio 2010 n. 122.

¹⁵ Fonti normative: Regolamento Ue n. 1151/2010 della Commissione Europea dell'8 dicembre 2010 – "Attuazione del regolamento (CE) n. 763/2008 del Parlamento europeo e del Consiglio relativo ai censimenti della popolazione e delle abitazioni per quanto riguarda le modalità e la struttura delle relazioni sulla qualità e il formato tecnico per la trasmissione dei dati".

¹⁶ Quintano C. e Castellano R. (2000), Strategie di trattamento delle mancate risposte ..., in Quintano C. (a cura di), *Scritti di Statistica economica 7*, Quaderni di discussione, ISM-IUN, n. 19, Napoli.

Nell'intento di individuare le unità sfuggite al Censimento, nel 1981 si abbinarono le unità rilevate con quelle censite, attraverso il nominativo della persona di riferimento del Foglio di famiglia (fornendo così tassi percentuali di copertura solo per l'universo delle famiglie), nell'ambito di un campione stratificato per ripartizione geografica (Nord, Centro, Mezzogiorno e isole), estratto in due stadi (primo stadio rappresentato da 66 comuni (Quintano, 2000) con popolazione maggiore ai 10 mila abitanti, secondo stadio rappresentato dalle sezioni di Censimento). Per massimizzare l'apporto informativo, al campione di sezioni originario, in fase di analisi, furono sostituite sia le sezioni risultate vuote, sia quelle con un numero non significativo di famiglie. L'indagine riguardò 120 mila famiglie e 15 mila abitazioni non occupate, riscontrando un tasso di copertura delle famiglie pari al 96,45 per cento.

Il campo di osservazione coincideva con quello del Censimento, ovvero famiglie, convivenze e abitazioni non occupate, così come la rete di Rilevatori era composta anche da quelli "dimostratisi particolarmente validi durante il Censimento della popolazione". "Accorgimento" adottato in seguito, in occasione dell'Indagine di copertura (Idc) del 2001, come ai giorni nostri, in occasione della PES 2011, laddove però (sia nel 2001 sia nel 2011) si è scelto, per il principio di indipendenza delle due indagini, di non assegnare ai Rilevatori lo stesso itinerario di sezione assegnato durante il Censimento.

Per l'indagine di copertura del 1991 vennero estratti 85 comuni campione, comprensivi di tutti i comuni con dimensione demografica maggiore di 500 mila abitanti, ritenuti autorappresentativi, e gli altri estratti a seguito di una stratificazione per ripartizione geografica e ampiezza demografica, e 638 sezioni di Censimento, selezionate con probabilità di inclusione costante. La rilevazione riguardò 52.844 famiglie e 11.157 abitazioni non occupate. Per raccordare Censimento e indagine di copertura a livello di singolo record si usò una duplice modalità: *record linkage* e dichiarazioni delle famiglie. Nella fase di *record linkage* si utilizzarono identificativi che si riferivano alle famiglie e non agli individui e ciò non risultò abbastanza discriminante; inoltre, non si fece riferimento al nominativo della persona intestataria del Foglio di famiglia, ma alle seguenti chiavi: codice provinciale, codice comunale, codice di sezione di Censimento, via, numero civico, piano ed interno e domanda diretta presso le famiglie a conferma dell'avvenuta adesione al Censimento. Il grado di copertura riferito alle famiglie italiane risultò pari al 99,3 per cento (Istat, 2009a). Successive analisi fanno supporre che il grado di copertura del 1991 debba essere considerato come un limite superiore in quanto deriva dalla semplice risposta degli intervistati alla richiesta di specificare se erano stati censiti o meno e non su di un rigoroso *record linkage* con i dati censuari. I risultati dello studio suggeriscono una lettura al ribasso: da 99,3 per cento a 98,2 per cento (Fortini, Gallo, Paluzzi, Reale, & Silvestrini, 2007).

Nel 2001, l'Indagine sul grado di copertura del 14° Censimento generale della popolazione (Idc) si estese a 98 comuni (di cui 14 autorappresentativi) e a 1.154 sezioni campione che vennero estratte secondo una stratificazione per tipologia di località abitata (centri, nuclei, case sparse/località produttive). Furono rilevate oltre 68 mila famiglie. Per eludere i problemi di *record linkage* del 1991, per la prima volta si acquisirono i nominativi e gli indirizzi degli individui rilevati al Censimento e all'Indagine di copertura che permisero di riferire le stime anche agli individui e non solo ai nuclei familiari. Tra le novità principali che caratterizzarono l'indagine si annoverano: la stima dell'errore di sottocopertura attraverso il modello statistico cattura-ricattura e, a livello sperimentale, la stima dell'errore di sovracopertura.

Per stimare la quantità di individui sfuggiti alla rilevazione, cioè la stima dell'errore di sottocopertura, si utilizzò il metodo cattura-ricattura¹⁷ basato sull'appaiamento di due o più liste, tutte affette da possibile sottonotifica, con lo scopo di contare i soggetti che, inclusi in una di queste, non lo sono in una, o in più d'una, delle altre. Questo metodo, denominato Dual-System Enumeration (DSE), permette di stimare l'ammontare ignoto di una popolazione in occasione del Censimento attraverso un primo conteggio esaustivo (cattura), la ripetizione della conta su un campione di sezioni di Censimento e una successiva integrazione finale tra i due insiemi di dati per calcolare coloro che sono enumerati sia in occasione del Censimento, sia in occasione della successiva indagine post-censuaria (Wolter, 1986).

Vale la pena sottolineare che, affinché il metodo *cattura-ricattura* sia applicabile senza distorsioni nelle stime, occorre che vi siano: indipendenza statistica tra le due rilevazioni (Censimento e copertura), la medesima probabilità di cattura tra tutti gli individui rilevati in una stessa occasione e una corretta (senza errore) identificazione del numero di "catture" di un individuo. La data di riferimento dell'indagine fu, analogamente al 14° Censimento generale della popolazione, quella del 21 ottobre 2001. Tuttavia, proprio per rispettare l'indipendenza statistica e intervenire sulla stessa popolazione di riferimento, evitando al contempo di interferire con le normali operazioni censuarie, si scelse di procedere all'Idc, nelle sezioni campione, in un momento immediatamente successivo alla conclusione del Censimento.

A differenza del Censimento dell'81 e del '91, nel 2001 non fu eseguita un'indagine di qualità *ad hoc*, per non gravare l'Istituto di costi aggiuntivi. Si calcolò piuttosto la variabilità causata dall'errore di misurazione a partire dall'utilizzo dei dati riguardanti gli individui censiti, insieme a quelli relativi ai record abbinati in occasione dell'Idc. Questo rendeva necessaria la confrontabilità tra i questionari. Il questionario della copertura fu quindi progettato nel formato, nel testo e nella grafica similmente a quello del Censimento, con l'unica variante della scelta del colore, così come in occasione della PES2011.

La PES 2011 è finalizzata alla stima dell'errore non campionario, ovvero di quell'errore che non dipende dal campionamento, ma che può manifestarsi in qualsiasi fase dell'indagine (raccolta dei dati, trattamento, etc.)¹⁸. Le occasioni di cattura sono due (Censimento ed indagine post censuaria), di cui la seconda a carattere campionario (Wolter, 1986).

Dall'indagine sul grado di copertura del 1981 alla PES 2011 la struttura del questionario è profondamente cambiata¹⁹ (Figura 1.1).

Oggi, come in passato, si garantisce l'indipendenza tra le due rilevazioni (Censimento e Copertura) mediante l'assegnazione al medesimo Rilevatore di sezioni diverse.

Un altro elemento di continuità è la registrazione manuale al posto della lettura ottica, adottata nel 2001, dovuta ad esigenze di contenimento di tempi e costi, quindi a motivi pratici più che a una scelta metodologica vera e propria.

¹⁷ Il metodo cattura-ricattura, introdotto per la stima della dimensione ignota di popolazione di origine animale è stato successivamente applicato anche alle popolazioni umane, ambito in cui il metodo è anche noto come dual system.

¹⁸ "La misurazione della qualità nelle indagini statistiche: concetti base", Il sistema di controllo della qualità di un'indagine statistica, Istat, M. D'Orazio.

¹⁹ Con riferimento all'indagine di copertura del 1981 si veda mod. ISTAT/CP/1/C; per la PES 2011 si veda Mod.Istat.PES.1.

Figura 1.1 - Indagine sul grado di copertura del 1981 – Questionario - Mod. Istat/CP/1/C



ISTITUTO CENTRALE DI STATISTICA
12° CENSIMENTO GENERALE DELLA POPOLAZIONE
25 ottobre 1981

Mod. ISTAT/CP/1/C

MUNICIPIO DI CODICE	SEZIONE DI CENSIMENTO
Comune	N.
Provincia	Foglio N.

INDAGINE SUL GRADO DI COPERTURA

N° INDIZIO	indirizzo	abitazione o convivenza	cognome e nome del capofamiglia - specie e denominazione della convivenza	la famiglia era abitualmente disabitata all'attuale indirizzo il giorno 25/10/81	la famiglia o la convivenza è stata censita all'attuale indirizzo il giorno		N° componenti abitualmente disabitati allo attuale indirizzo il giorno	osservazioni
					25/10/81	25/10/81		
1	2	3	4	5	6	7	8	9
		<input type="checkbox"/> conviv. <input type="checkbox"/> ab.occ. <input type="checkbox"/> ab.non occ. 1 2 3	SI NO	SI NO		
		<input type="checkbox"/> conviv. <input type="checkbox"/> ab.occ. <input type="checkbox"/> ab.non occ. 1 2 3	SI NO	SI NO		
		<input type="checkbox"/> conviv. <input type="checkbox"/> ab.occ. <input type="checkbox"/> ab.non occ. 1 2 3	SI NO	SI NO		

1.2 Quadro normativo

Il Regolamento (CE) n. 763/2008 del Parlamento europeo e del Consiglio del 9 luglio 2008 relativo ai censimenti della popolazione e delle abitazioni la cui attuazione, per quanto riguarda le specifiche tecniche delle variabili e delle loro classificazioni, viene disposta dal successivo Regolamento (CE) n. 1201/2009 della Commissione del 30 novembre 2009, ha come oggetto norme comuni per la fornitura decennale di dati esaurienti sulla popolazione e sulle abitazioni. Esso pone, tra i suoi obiettivi principali (al fine di disporre di dati sufficientemente affidabili e dettagliati per quanto riguarda le definizioni e il programma dei dati e dei metadati) la raccolta e la compilazione di statistiche comunitarie comparabili ed esaurienti sulla popolazione e sulle abitazioni. Allo stesso tempo il citato Regolamento presuppone che questi obiettivi non possano essere adeguatamente realizzati dagli Stati membri per la mancanza di parametri statistici e di requisiti di qualità comuni, nonché per l'insufficiente trasparenza metodologica. Per tale motivo, pertanto, la maniera più efficace di realizzazione degli obiettivi risulta essere, a livello comunitario, la creazione di un quadro statistico comune. In virtù del principio di sussidiarietà, sancito all'art 5 del trattato istitutivo della Comunità europea, si prevede, dunque, che la Comunità possa intervenire nei limiti di quanto risulti necessario per il conseguimento di tali obiettivi, sempre in ottemperanza al principio di proporzionalità enunciato nello stesso articolo. Ai fini del Regolamento n. 763/2008 gli Stati membri sono dunque tenuti a presentare alla Commissione (Eurostat) una relazione sulla qualità dei dati forniti in occasione dell'esecuzione del Censimento. A tali dati si applicano le seguenti dimensioni per la valutazione della qualità:

- ▶ pertinenza: il grado in cui le statistiche rispondono alle esigenze attuali e potenziali degli utenti;
- ▶ accuratezza: la vicinanza tra le stime e i valori reali non noti;

- ▶ tempestività e puntualità: l'intervallo di tempo che intercorre fra il periodo di riferimento e la disponibilità dei risultati;
- ▶ accessibilità e chiarezza: le condizioni e le modalità con cui gli utenti possono ottenere, utilizzare e interpretare i dati;
- ▶ comparabilità: la misurazione dell'impatto delle differenze tra i concetti di statistica applicata e gli strumenti e le procedure di misurazione, quando le statistiche si comparano per aree geografiche, ambiti settoriali o periodi di tempo;
- ▶ coerenza: la possibilità di combinare i dati in modo attendibile secondo modalità differenti e per usi diversi.

Il Regolamento (UE) n. 1151/2010 della Commissione dell'8 dicembre 2010, in attuazione di quanto stabilito dal Regolamento (CE) n. 763/2008, al fine di valutare la qualità dei dati trasmessi dagli Stati membri alla Commissione (Eurostat), stabilisce le modalità e la struttura delle relazioni che gli Stati membri sono tenuti a presentare sulla qualità dei dati dei Censimenti della popolazione e delle abitazioni. A tal fine, gli Stati membri sono tenuti ad effettuare una valutazione della copertura dei propri censimenti della popolazione e delle abitazioni per l'anno di riferimento 2011. L'articolo 3 del suddetto Regolamento detta i tempi entro i quali gli Stati membri sono tenuti a fornire le informazioni richieste e inoltre specifica i dati e i metadati relativi alla qualità da rilevare. Al fine di rispondere adeguatamente alle esigenze informative imposte dai regolamenti europei è stata inserita, come nuovo progetto nel Programma Statistico Nazionale 2011-2013, aggiornamento 2012-2013, la Rilevazione campionaria di controllo della copertura e qualità del 15° Censimento generale della popolazione e delle abitazioni (IST-02410).

Tale rilevazione è prevista altresì dal Piano Generale di Censimento (PGC)²⁰, adottato con deliberazione Istat n. 6/PRES del 18 febbraio 2011 e comunicato in Gazzetta Ufficiale l'8 marzo 2011, serie generale n. 55, nel rispetto di quanto stabilito dall'art. 50, comma 1, del Decreto legge 31 maggio 2010, n. 78, convertito con modificazioni in Legge 30 luglio 2010, n. 122 e recante "Misure urgenti in materia di stabilizzazione finanziaria e di competitività economica". In particolare, il PGC prevede che gli organi di Censimento siano tenuti ad adempiere, tra le altre attività, anche all'esecuzione della Rilevazione campionaria areale di controllo della copertura del Censimento (PES 2011).

Come il 15° Censimento generale della popolazione e delle abitazioni, anche la Rilevazione campionaria areale di controllo della copertura del Censimento è inserita nel PSN 2011-2013, aggiornamento 2012-2013, tra le rilevazioni da svolgere a cura dell'Istituto Nazionale di Statistica per le quali è previsto l'obbligo di risposta.

L'obbligo di fornire dati statistici è previsto dall' art. 7 del Decreto Legislativo 6 settembre 1989, n. 322, che detta le "Norme sul Sistema statistico nazionale e sulla riorganizzazione dell'Istituto Nazionale di Statistica". Tale norma fa obbligo a tutte le amministrazioni, enti e organismi pubblici, nonché ai soggetti privati di fornire tutti i dati che vengano loro richiesti per le rilevazioni previste dal Programma Statistico Nazionale ed espressamente indicate con delibera del Consiglio dei Ministri, che nel caso della Rilevazione PES è il DPR 26 settembre 2012 (G.U. n. 267 del 15 novembre 2012). L'obbligo di risposta si sostanzia nella compilazione, in modo completo e veritiero, del questionario predisposto dall'Istat. Si riscontra quindi violazione esclusivamente nelle ipotesi in cui il soggetto obbligato non fornisca le informazioni richieste o le fornisca scientemente errate o incomplete.

²⁰ Il Piano generale di Censimento (PGC) è un atto programmatico di natura generale che detta le linee di indirizzo in merito alla pianificazione, l'organizzazione e l'esecuzione delle operazioni relative al 15° Censimento Generale della Popolazione e delle Abitazioni, mentre la disciplina degli aspetti di dettaglio e operativi del Censimento è rinviata ad ulteriori atti e circolari.

Le attività realizzate in esecuzione della PES che comportino l'effettuazione di qualsiasi operazione concernente il trattamento di dati e di informazioni relative a persone fisiche²¹ identificate o identificabili anche indirettamente, devono essere svolte nel rispetto di quanto stabilito dal Decreto Legislativo 30 giugno 2003, n. 196 (Codice in materia di protezione dei dati personali) come modificato dal D.L. 6 dicembre 2011 n. 201, convertito con modificazioni in L. 22 dicembre 2011 n. 214 (Disposizioni urgenti per la crescita, l'equità e il consolidamento dei conti pubblici) e del Codice di deontologia e di buona condotta per i trattamenti di dati personali a scopi statistici e di ricerca scientifica effettuati nell'ambito del Sistema statistico nazionale (All. A3 del D.Lgs. n. 196/2003) e successive modificazioni e integrazioni.

Ai sensi dell'art. 28 del D.Lgs n. 196/2003, titolare del trattamento dei dati personali è l'Istat. Responsabili del trattamento dei dati personali, ai sensi dell'art. 29 del citato D.Lgs, sono, per le operazioni di rispettiva competenza, il Direttore del Dipartimento per i censimenti e gli archivi amministrativi e statistici dell'Istat, i Responsabili degli Uffici Regionali di Censimento (URC), i Responsabili degli Uffici Provinciali di Censimento (UPC) e i Responsabili degli Uffici Comunali di Censimento in forma singola o associata (UCC).

I dati raccolti o comunque utilizzati nel corso della rilevazione sono coperti da segreto d'ufficio e da segreto statistico, ai sensi degli artt. 8 e 9 del D.Lgs. n. 322/89, e successive modificazioni e integrazioni. La loro diffusione o comunicazione a soggetti non autorizzati è, pertanto, severamente vietata. Ai sensi dell'art. 9 del D.Lgs. n. 322/89, il segreto statistico garantisce che i dati individuali acquisiti per la produzione statistica ufficiale non vengano utilizzati per altri fini o comunicati a terzi e siano comunque diffusi in forma aggregata in modo che non si possa risalire al soggetto al quale si riferiscono.

A livello comunitario il segreto statistico è disciplinato dal Regolamento (CE) n. 223/2009 del Parlamento europeo e del Consiglio dell'11 marzo 2009 il quale prevede che, al fine di assicurare che i dati riservati siano usati esclusivamente a fini statistici e di prevenirne la divulgazione illecita, è necessario adottare una serie di norme e misure. In particolare, ai sensi dell'art. 20 comma 2, i dati riservati ottenuti esclusivamente per la produzione di statistiche europee sono utilizzati dagli Istituti Nazionali di Statistica (INS) e dalle altre autorità nazionali, nonché dalla Commissione (Eurostat) esclusivamente per fini statistici, a meno che l'unità statistica non abbia in maniera non equivoca autorizzato l'impiego per altri scopi. In caso di risultati statistici che rendano possibile l'identificazione dell'unità statistica, questi possono essere diffusi dagli Istituti Nazionali di Statistica e dalle altre autorità nazionali, nonché dalla Commissione (Eurostat) solo in alcuni casi specificatamente indicati dall'art. 20 comma 3, il quale prevede che i risultati statistici suscettibili di rendere possibile l'identificazione di un'unità statistica possono essere diffusi dagli INS e dalle altre autorità nazionali e dalla Commissione (Eurostat) nei seguenti casi eccezionali:

- ▶ qualora condizioni e modalità specifiche siano fissate da un atto del Parlamento europeo e del Consiglio, adottato a norma dell'articolo 251 del trattato, e i risultati della statistica siano modificati in modo tale che la loro diffusione non ne comprometta la riservatezza ogni qualvolta l'unità statistica lo richieda;
- ▶ qualora l'unità statistica abbia inequivocabilmente autorizzato la divulgazione dei dati.

Inoltre, nell'ambito delle rispettive sfere di competenza, gli INS, le altre autorità nazionali e la Commissione (Eurostat) sono tenuti ad adottare tutte le misure regolamentari, amministrative, tecniche e organizzative necessarie a garantire la protezione fisica e logica di dati riservati.

²¹ Per operazioni concernenti il trattamento di dati e di informazioni relative a persone fisiche si intende: la raccolta, la registrazione, l'organizzazione, la conservazione, la consultazione, l'elaborazione, la modificazione, la selezione, l'estrazione, il raffronto, l'utilizzo, l'interconnessione, il blocco (conservazione con sospensione temporanea di ogni trattamento), la comunicazione, la diffusione, la cancellazione e la distruzione di dati.

1.3 Aspetti finanziari e amministrativi

Il Piano generale di Censimento (PGC) al paragrafo 3.3.1 stabilisce che per la rilevazione campionaria areale di controllo della copertura del Censimento (PES) è previsto un contributo totale massimo pari a € 5.000.000,00. Esso sarà ripartito solo tra i Comuni che formeranno il campione areale di controllo della qualità e della copertura della rilevazione censuaria, secondo criteri di commisurazione al numero delle sezioni di Censimento da sottoporre a controllo e del numero di questionari rilevati.

In particolare è stata prevista una spesa complessiva per i comuni che partecipano alla rilevazione di € 2.666.387,50 sulla base del numero di sezioni di Censimento da sottoporre a controllo e della stima del numero dei questionari rilevati. La Circolare n. 17 del 23 gennaio 2012 prot. n. 1990, con oggetto "Rilevazione campionaria areale di controllo della copertura e della qualità del 15° Censimento generale della popolazione e delle abitazioni", al paragrafo 5 prevede la corresponsione di un contributo ripartito tra i 255 Comuni che formano il campione della rilevazione secondo criteri di commisurazione al numero di sezioni di Censimento da sottoporre a controllo e al numero di unità rilevate.

In linea con la suddetta Circolare, è stato corrisposto un contributo fisso pari a €150,00 per ogni sezione di Censimento da sottoporre a controllo e un contributo variabile pari a € 15,00 per ogni questionario compilato, revisionato e trasmesso all'ISTAT (Prospetto 1.1).

Prospetto 1.1 - Contributi ai Comuni

Contributo fisso	€ 150,00 a sezione di Censimento da sottoporre a controllo
Contributo variabile	€ 15,00 a questionario trasmesso all'Istat

I contributi di cui sopra sono stati erogati a ciascun Comune in due soluzioni (Prospetto 1.2):

- ▶ l'anticipo, pari alla somma dell'intero contributo fisso e al 20 per cento del contributo variabile, calcolato in base al numero presunto di unità da rilevare, di complessivi € 832.755,19. Questa somma si riferisce ad un primo gruppo di 248 Comuni per i quali è stata disposta la liquidazione di complessivi € 814.277,73 nel mese di aprile 2012 e un secondo gruppo di 5 comuni per i quali è stata disposta la liquidazione di complessivi € 18.477,46 nel mese di maggio 2012;
- ▶ il saldo del contributo variabile, pari al residuale 80 per cento, calcolato sulla base del numero di unità effettivamente rilevate a conclusione della rilevazione, per complessivi € 1.611.441,30 nel mese di agosto 2012.

Prospetto 1.2 - Importo liquidato

Totale anticipo (*)	Saldo (80% variabile)	Totale contributo liquidato
€ 832.755,19	€ 1.611.441,30	€ 2.444.196,49

(*) contributo fisso + 20% contributo variabile.

In totale i comuni che hanno partecipato alla rilevazione PES sono risultati 252 sui 255 previsti, poiché i Comuni di Veggiano (PD) e Formello (RM) non hanno potuto parteciparvi dall'inizio, mentre, il comune di Tivoli (RM) ha restituito all'ISTAT l'anticipo riscosso di complessivi € 3.486,49 avendo ad indagine iniziata deciso di non prendere parte alla rilevazione.

Il totale complessivo del contributo corrisposto ai 252 comuni è risultato perciò inferiore di € 225.677,50 rispetto alla spesa prevista inizialmente di € 2.666.387,50 (Prospetto 1.3).

Prospetto 1.3 - Dimostrazione del risultato economico del contributo ai Comuni

Importo previsto	Importo liquidato (*)	Differenza
€ 2.666.387,50	€ 2.440.710,00	€ 225.677,50

(*) Al netto del contributo del Comune di Tivoli

1.4 Gestione dei servizi in outsourcing

L'esecuzione dell'indagine PES ha richiesto, per quanto riguarda l'allestimento del materiale della rilevazione, di predisporre alcune attività di servizi. In particolare, prima dell'inizio della rilevazione sul campo si sono rese necessarie attività di pre stampa, stampa, allestimento, confezionamento, trasporto e consegna di colli relative al materiale di rilevazione e a conclusione dell'indagine si sono dovute operare attività di ritiro delle scatole contenenti i questionari compilati e il relativo trasporto.

L'Istat, relativamente al 15° Censimento generale della popolazione e delle abitazioni 2011, aveva stipulato il "Contratto di appalto per i servizi di stampa, trasporto, acquisizione di dati e immagini mediante lettura ottica e registrazione tradizionale relativi ai questionari e all'altro materiale del 15° Censimento generale della popolazione e delle abitazioni 2011"(rep. n. 188/2011), a seguito dell'espletamento di una procedura pubblica con aggiudicazione a favore del Raggruppamento Temporaneo di Imprese (RTI) con mandataria Postel S.p.A..

Le attività di servizi relative alla PES hanno richiesto, da parte dell'Istat, una gestione contrattuale specifica disciplinata dall'art. 57 del D.Lgs. 12 aprile 2006, n.163, Codice dei Contratti Pubblici, che dà la possibilità alle stazioni appaltanti di aggiudicare contratti pubblici mediante procedura negoziata senza previa pubblicazione di un bando di gara nelle ipotesi indicate, dandone conto con adeguata motivazione nella delibera o determina a contrarre. Tra le varie ipotesi indicate, il comma 5 lettera a) dispone che la procedura è consentita per i lavori o i servizi complementari, non compresi nel progetto iniziale né nel contratto iniziale che, a seguito di una circostanza imprevista, sono divenuti necessari all'esecuzione dell'opera o del servizio oggetto del progetto o del contratto iniziale, purché aggiudicati all'operatore economico che presta tale servizio o esegue tale opera, nel rispetto delle seguenti condizioni:

- ▶ tali lavori o servizi complementari non possono essere separati, sotto il profilo tecnico o economico, dal contratto iniziale, senza recare gravi inconvenienti alla stazione appaltante, ovvero, pur essendo separabili dall'esecuzione del contratto iniziale, sono strettamente necessari al suo perfezionamento;
- ▶ il valore complessivo stimato dei contratti aggiudicati per lavori o servizi complementari non superi il cinquanta per cento dell'importo del contratto iniziale.

L'Istat, quindi, in virtù del dettato dell'art 57 del Codice dei Contratti Pubblici, facendo ricorso allo strumento "dell'estensione per servizi complementari" ha disposto l'affidamento diretto a favore del predetto Raggruppamento Temporaneo di Imprese (RTI) con mandataria Postel S.p.A., già appaltatore del "Contratto di appalto per i servizi di stampa, trasporto, acquisizione di dati e immagini mediante lettura ottica e registrazione tradizionale relativi ai questionari e all'altro materiale del 15° Censimento generale della popolazione e delle abitazioni 2011", avente ad oggetto il medesimo tipo di prestazione necessaria per la PES.

Questo accorgimento, ammesso dal legislatore comunitario, ha consentito all'Istat di anteporre l'esigenza dell'economia procedimentale della PES in termini di opportunità tecnica, temporale e organizzativa rispetto al sistema di garanzie a tutela della concorrenza; infatti, le attività eseguite per la PES rispettavano il concetto di complementarietà così come definito dal legislatore ovvero quello di "lavori suppletivi" e non di "lavori extracontrattuali" (cioè aventi una propria individualità distinta rispetto a quella originaria).

L'affidamento a Postel S.p.A. dei servizi relativi alla PES è da intendersi, quindi, integrativo di quello del 15° Censimento generale della popolazione e delle abitazioni e per nulla correlato con il contratto originario (n.188/2011), sotto il profilo dei rapporti quantitativi ed economici, nonostante la corrispondenza soggettiva (i contraenti) ed oggettiva (il tipo di prestazione) delle due fattispecie negoziali. In questa tipologia di estensione, non solo non viene imposta una proporzione limite, come la regola del "sesto-quinto"²², ma non è previsto il mantenimento, nel nuovo contratto, delle originarie condizioni esecutive pattuite nell'affidamento principale. Infatti le parti (Istat e Postel) hanno avuto modo di rinegoziare le modalità di svolgimento delle attività esecutive del servizio in estensione (PES), senza essere vincolate a quanto pattuito con riferimento alla fornitura originaria. Dal punto di vista operativo, l'Istat ha preliminarmente richiesto a Postel S.p.A di quotare il servizio per le attività PES suindicate e successivamente ha operato il confronto tra i prezzi offerti con quelli relativi agli analoghi servizi previsti nel contratto di appalto per i servizi relativi al 15° Censimento generale della popolazione e delle abitazioni 2011 (n.188/2011), al fine di esprimere una congruità tecnica/economica per il nuovo affidamento. Il confronto ha tenuto conto sia della relazione inversa tra prezzo e quantità (il prezzo unitario aumenta se la quantità diminuisce, e viceversa), sia del valore del servizio PES rispetto all'importo del contratto n.188/2011.

²² La regola del quinto d'obbligo o "sesto-quinto" si riferisce al caso in cui la stazione appaltante disponga un aumento o una diminuzione delle opere da eseguirsi, in questo caso l'appaltatore ha l'obbligo di eseguire, senza alcuna indennità, le variazioni che non superino il quinto dell'importo complessivo dell'appalto.

CAPITOLO 2

Organizzazione e compiti della rete²³

2.1 Organizzazione della rilevazione

L'organizzazione della Rilevazione campionaria areale di controllo della copertura e della qualità del 15° Censimento generale della popolazione e delle abitazioni (PES) ha riguardato la gestione e il monitoraggio di risorse umane e materiali tenendo presente i vincoli normativi e finanziari. Ogni attività del processo organizzativo è stata programmata, organizzata (individuando i soggetti e il tipo di responsabilità per la cura della sua realizzazione), guidata e infine controllata.

Scopo della rilevazione è di determinare gli indicatori di qualità e di copertura del 15° Censimento generale della popolazione e delle abitazioni stabiliti dal Regolamento (UE) n. 1151/2010 della Commissione Europea dell'8-12-2010.

In questo capitolo si affrontano le tematiche organizzative relative: al ruolo e ai compiti dell'Istat e della rete di rilevazione, alle operazioni sul campo, alla formazione, al monitoraggio del materiale di rilevazione (allestimento, trasmissione, confezionamento e ritiro) e al monitoraggio della rilevazione.

Per l'esecuzione della rilevazione l'Istat si è avvalsa della collaborazione degli Uffici di Censimento costituiti ai sensi del Piano Generale di Censimento (PGC) a vari livelli territoriali e, a livello comunale, degli Uffici Comunali di Censimento (UCC) dei comuni appartenenti al primo stadio del campione areale di controllo, secondo le modalità stabilite dalla circolare n. 3 del 3 marzo 2011, prot. n.1896²⁴, tenendo conto anche degli UCC costituiti in forma associata tra più comuni. L'associazione di comuni è stata incentivata al fine di una più efficiente organizzazione delle attività e, nelle ipotesi in cui la costituzione di UCC in ciascun singolo comune avrebbe potuto determinare diseconomie di scala. Per tale ragione, dei 255 comuni facenti parte del campione di primo stadio, 14 appartengono a 12 UCC in forma associata, mentre 241 comuni hanno costituito UCC in forma singola.

Dei 14 comuni appartenenti ad UCC in forma associata, sette sono situati in Emilia-Romagna (dei quali tre appartenenti allo stesso UCC), tre nella regione Toscana, due in Valle d'Aosta, uno in Piemonte e uno in Puglia.

Tutte le regioni, attraverso i loro comuni, sono rappresentate nel campione. Il foglio di famiglia, la guida alla compilazione e la lettera a firma del Presidente dell'Istat sono stati tradotti nella lingua slovena e tedesca per gli UCC di alcuni comuni della regione Friuli-Venezia Giulia e della provincia di Bolzano, in cui sono presenti zone soggette alla tutela della minoranza linguistica.

²³ I paragrafi sono a cura di:

- 2.1 Organizzazione della rilevazione: Patrizia Perini
- 2.2 Ruolo e compiti dell'Istat: Loredana De Gaetano
- 2.3 Ruolo e compiti della rete e operazioni sul campo: Valentina Fusco
- 2.4 Formazione: Gerarda Grippo
- 2.5 Monitoraggio del materiale della rilevazione: Claudia Naccarato
- 2.6 Monitoraggio della rilevazione sul campo: Gianluca Grauso

²⁴ La Circolare n. 3 del 3 marzo 2011, prot. n. 1896 in esecuzione delle disposizioni citate nel PGC, fornisce indicazioni in ordine a compiti, modalità e tempi di costituzione degli UCC, nonché a modalità e tempi di corresponsione dei contributi loro spettanti.

I comuni di Veggiano, Formello e Tivoli non hanno partecipato alla rilevazione, quindi i comuni campione, nel complesso, sono stati 252 e gli Uffici Comunali di Censimento 250. Il campione di secondo stadio, composto dalle sezioni di Censimento definitive, è stato comunicato a ogni Responsabile di UCC circa un mese prima dell'inizio della rilevazione; precedentemente, attraverso la Circolare n.17 del 23 gennaio 2012 prot. n.1990²⁵, era stata comunicata una stima del campione.

Nel complesso, quindi, sono state interessate 2.488 sezioni di Censimento, corrispondenti a una stima di circa 152 mila intestatari di fogli di famiglia rilevati al 31 dicembre 2010 (Tavola 2.1).

Tavola 2.1 - Campione stimato e campione effettivo per comune (valori assoluti)

Comune	Cod. Reg.	Campione stimato (a)		Campione effettivo (b)	
		N° famiglie	N° sezioni	N° famiglie	N° sezioni
1 Bene Vagienna	1	656	22	699	22
2 Pinasca	1	658	18	733	18
3 Garbagna Novarese	1	559	10	589	10
4 Bannio Anzino	1	218	8	240	8
5 Crescentino	1	472	8	512	8
6 Dronero	1	496	10	531	10
7 Villanova d'Asti	1	483	6	526	6
8 Domodossola	1	495	15	546	15
9 Galliate	1	529	3	569	3
10 Ovada	1	490	5	548	6
11 Alessandria	1	595	16	666	16
12 Nichelino	1	602	2	681	2
13 Casale Monferrato	1	630	7	700	7
14 Fossano	1	583	7	621	7
15 Torino	1	1.981	20	2.169	20
16 Novara	1	229	5	247	5
17 Quart	2	770	20	898	20
18 Challand-Saint-Victor	2	252	17	288	17
19 Aosta	2	892	17	997	16
20 Monzambano	3	603	12	607	13
21 Cassago Brianza	3	605	7	615	7
22 Castione della Presolana	3	602	14	584	14
23 Castelbelforte	3	603	11	612	11
24 Angolo Terme	3	601	11	609	11
25 Entratico	3	600	8	499	9
26 San Cipriano Po	3	210	9	230	9
27 Castelleone	3	487	6	482	6
28 Magnago	3	491	5	547	5

segue

²⁵ Circolare n. 17 del 23 gennaio 2012 prot. n° 1990 con oggetto: Rilevazione campionaria areale di controllo della copertura e della qualità del 15° Censimento generale della popolazione e delle abitazioni.

Tavola 2.1 - Campione stimato e campione effettivo per comune (valori assoluti)

Comune	Cod. Reg.	Campione stimato (a)		Campione effettivo (b)	
		N° famiglie	N° sezioni	N° famiglie	N° sezioni
29 Sovico	3	488	3	491	3
30 Chiavenna	3	486	3	488	3
31 Pegognaga	3	481	7	472	7
32 Casteggio	3	485	7	546	7
33 Siziano	3	480	4	541	4
34 Figino Serenza	3	480	4	470	4
35 Cassano d'Adda	3	534	6	600	6
36 Erba	3	519	3	531	3
37 Merate	3	535	4	517	4
38 Alzano Lombardo	3	518	3	511	3
39 Calcinato	3	520	5	527	5
40 Nave	3	520	3	516	3
41 Gerenzano	3	539	6	556	6
42 Cinisello Balsamo	3	758	11	862	11
43 Pavia	3	783	20	844	19
44 Cologno Monzese	3	743	11	860	12
45 Cesano Maderno	3	746	12	738	11
46 Garbagnate Milanese	3	720	3	805	3
47 Novate Milanese	3	757	14	859	15
48 Milano	3	2.808	31	3.129	31
49 Brescia	3	411	9	415	10
50 Monza	3	260	4	268	4
51 Bergamo	3	253	3	244	3
52 Veggiano	5	493	8	494	8
53 San Martino di Venezze	5	495	6	509	7
54 Bergantino	5	488	8	500	8
55 Cencenighe Agordino	5	493	14	549	14
56 Maserada sul Piave	5	563	7	536	7
57 Limena	5	553	13	529	13
58 Noventa di Piave	5	548	7	548	7
59 Longare	5	552	4	540	4
60 Abano Terme	5	518	4	496	4
61 Rubano	5	502	4	510	4
62 Roncade	5	521	5	507	5
63 Caldogno	5	514	4	514	4
64 Camisano Vicentino	5	524	5	524	5
65 Bassano del Grappa	5	774	26	768	26
66 Spinea	5	740	4	763	4
67 Legnago	5	755	31	759	31

segue

Tavola 2.1 - Campione stimato e campione effettivo per comune (valori assoluti)

Comune	Cod. Reg.	Campione stimato (a)		Campione effettivo (b)		
		N° famiglie	N° sezioni	N° famiglie	N° sezioni	
68	Venezia	5	643	22	650	22
69	Verona	5	626	11	612	11
70	Padova	5	508	10	494	10
71	Vicenza	5	275	6	270	6
72	San Giorgio della Richinvelda	6	489	4	482	4
73	Ronchis	6	487	13	517	14
74	Brugnera	6	485	6	485	6
75	Basiliano	6	470	5	513	5
76	Azzano Decimo	6	403	3	396	3
77	Ronchi dei Legionari	6	445	3	501	3
78	Udine	6	365	8	392	8
79	Gorizia	6	497	11	551	11
80	Trieste	6	798	9	998	8
81	Millesimo	7	886	9	1.089	9
82	Vado Ligure	7	402	4	475	4
83	Diano Marina	7	383	8	437	8
84	Loano	7	366	4	444	4
85	Alassio	7	372	4	432	4
86	La Spezia	7	635	19	701	19
87	Chiavari	7	512	5	603	5
88	Genova	7	1.565	22	1.840	22
89	Castello di Serravalle	8	689	13	820	13
90	Albareto	8	697	39	757	39
81	Millesimo	7	886	9	1.089	9
82	Vado Ligure	7	402	4	475	4
83	Diano Marina	7	383	8	437	8
84	Loano	7	366	4	444	4
85	Alassio	7	372	4	432	4
86	La Spezia	7	635	19	701	19
87	Chiavari	7	512	5	603	5
88	Genova	7	1.565	22	1.840	22
89	Castello di Serravalle	8	689	13	820	13
90	Albareto	8	697	39	757	39
91	Argelato	8	573	6	659	6
92	San Giorgio di Piano	8	591	5	700	5
93	Bentivoglio	8	585	14	688	15
94	Maranello	8	491	4	488	4
95	Bondeno	8	497	9	530	9
96	Portomaggiore	8	506	7	566	7

segue

Tavola 2.1 - Campione stimato e campione effettivo per comune (valori assoluti)

Comune	Cod. Reg.	Campione stimato (a)		Campione effettivo (b)	
		N° famiglie	N° sezioni	N° famiglie	N° sezioni
97 Castelnovo ne' Monti	8	520	12	533	13
98 Cesena	8	673	20	673	20
99 Lugo	8	706	9	762	9
100 Santarcangelo di Romagna	8	658	6	675	6
101 Bologna	8	719	11	836	10
102 Parma	8	353	6	390	6
103 Modena	8	350	10	364	10
104 Reggio nell'Emilia	8	622	14	635	14
105 Rimini	8	528	17	526	16
106 Piacenza	8	419	10	450	10
107 Licciana Nardi	9	591	15	635	16
108 Castelfranco di Sopra	9	592	11	551	11
109 Pieve a Nievole	9	478	6	474	6
110 Castiglione della Pescaia	9	477	12	526	12
111 Castelnuovo di Garfagnana	9	474	10	483	10
112 Agliana	9	567	16	585	16
113 Certaldo	9	574	10	625	10
114 Montignoso	9	578	7	630	7
115 Pistoia	9	512	10	518	10
116 Carrara	9	550	6	571	7
117 Capannori	9	565	8	580	8
118 Campi Bisenzio	9	565	11	608	12
119 Pietrasanta	9	538	7	545	7
120 Firenze	9	855	18	913	12
121 Prato	9	433	8	405	8
122 Livorno	9	371	6	394	6
123 Arezzo	9	231	4	226	4
124 Porano	10	630	7	682	7
125 Torgiano	10	542	7	526	7
126 Marsciano	10	620	8	628	8
127 Spoleto	10	502	14	494	14
128 Corciano	10	528	5	510	5
129 Perugia	10	612	7	623	7
130 Terni	10	413	6	432	6
131 San Costanzo	11	530	7	526	7
132 Macerata Feltria	11	535	16	537	16
133 Fossombrone	11	434	5	429	5
134 Gabicce Mare	11	438	3	435	3
135 Urbino	11	463	8	468	8

segue

Tavola 2.1 - Campione stimato e campione effettivo per comune (valori assoluti)

Comune	Cod. Reg.	Campione stimato (a)		Campione effettivo (b)	
		N° famiglie	N° sezioni	N° famiglie	N° sezioni
136 Cingoli	11	417	8	389	8
137 Pesaro	11	540	14	521	14
138 Jesi	11	560	11	561	11
139 Falconara Marittima	11	513	15	521	14
140 Ancona	11	601	10	600	10
141 Vitorchiano	12	800	7	805	7
142 Picinisco	12	523	35	502	34
143 Montalto di Castro	12	506	10	526	10
144 Toscana	12	493	9	497	9
145 Genazzano	12	511	5	510	6
146 Santa Marinella	12	512	5	525	5
147 Vetralla	12	553	5	561	5
148 Formello	12	542	4	526	4
149 Tivoli	12	798	6	812	7
150 Anzio	12	770	22	764	23
151 Fondi	12	746	3	708	3
152 Sora	12	687	6	643	6
153 Roma	12	4.574	54	4.675	54
154 Latina	12	198	4	190	4
155 Controguerra	13	564	9	605	10
156 Castellafiume	13	468	7	474	7
157 Bucchianico	13	764	10	738	10
158 San Salvo	13	400	2	375	2
159 Atessa	13	368	7	366	7
160 Chieti	13	664	19	645	19
161 Avezzano	13	687	20	696	20
162 Pescara	13	625	9	604	10
163 Pozzilli	14	551	9	710	9
164 Lucito	14	318	12	309	12
165 Larino	14	518	7	509	7
166 Venafro	14	340	2	337	2
167 Isernia	14	835	19	844	20
168 San Giovanni a Piro	15	633	16	578	16
169 Capriglia Irpina	15	638	9	582	9
170 Casalbore	15	638	12	601	12
171 Carinola	15	609	8	523	8
172 Grazzanise	15	612	5	522	5
173 Cancellò ed Arnone	15	620	4	553	4
174 Grumo Nevano	15	542	8	462	8

segue

Tavola 2.1 - Campione stimato e campione effettivo per comune (valori assoluti)

Comune	Cod. Reg.	Campione stimato (a)		Campione effettivo (b)	
		N° famiglie	N° sezioni	N° famiglie	N° sezioni
175 Castel San Giorgio	15	544	3	507	3
176 Sala Consilina	15	539	6	472	6
177 Sant'Agata de' Goti	15	510	12	480	13
178 Pozzuoli	15	566	4	475	4
179 Marano di Napoli	15	624	3	521	3
180 Aversa	15	630	2	533	2
181 Eboli	15	574	17	522	17
182 Angri	15	597	2	524	2
183 San Giuseppe Vesuviano	15	588	2	499	2
184 Ottaviano	15	625	3	525	3
185 Napoli	15	2.053	22	1.691	22
186 Salerno	15	550	10	503	11
187 Specchia	16	536	4	487	4
188 Patù	16	535	6	373	6
189 Parabita	16	691	7	643	7
190 Corsano	16	701	3	667	3
191 San Vito dei Normanni	16	498	3	466	3
192 Maglie	16	497	4	458	4
193 Leverano	16	511	2	479	2
194 Taviano	16	499	2	434	2
195 Cerignola	16	859	9	767	9
196 Corato	16	780	33	702	33
197 Francavilla Fontana	16	809	10	751	10
198 Conversano	16	808	19	729	20
199 Bari	16	770	9	710	9
200 Taranto	16	461	6	423	6
201 Andria	16	608	16	499	15
202 Marsico Nuovo	17	492	19	471	19
203 Montemilone	17	486	18	454	18
204 Senise	17	745	6	736	6
205 Melfi	17	375	4	356	4
206 Bernalda	17	418	5	384	5
207 Matera	17	759	13	681	13
208 San Calogero	18	546	4	457	4
209 Platania	18	541	14	509	14
210 Amato	18	359	9	342	9
211 Pizzo	18	525	4	473	4
212 Motta San Giovanni	18	530	7	484	7
213 Taurianova	18	506	4	465	4

segue

Tavola 2.1 - Campione stimato e campione effettivo per comune (valori assoluti)

Comune	Cod. Reg.	Campione stimato (a)		Campione effettivo (b)	
		N° famiglie	N° sezioni	N° famiglie	N° sezioni
214 Amantea	18	495	3	463	3
215 Lamezia Terme	18	623	4	608	4
216 Crotonè	18	729	11	668	11
217 Reggio di Calabria	18	791	23	732	23
218 Vallelunga Pratameno	19	782	9	715	9
219 Fiumedinisi	19	644	8	663	8
220 Palazzolo Acreide	19	553	6	523	6
221 Valverde	19	561	5	538	5
222 Cammarata	19	575	6	522	6
223 Palagonia	19	468	3	468	3
224 Leonforte	19	471	5	450	6
225 Terrasini	19	445	3	425	3
226 Capaci	19	460	3	429	3
227 Trapani	19	878	37	825	37
228 Paternò	19	871	30	844	30
229 Carini	19	897	4	838	4
230 Gravina di Catania	19	881	7	841	7
231 Palermo	19	1.329	14	1.262	14
232 Catania	19	595	12	590	12
233 Messina	19	491	8	512	8
234 Siracusa	19	251	3	243	3
235 Orgosolo	20	702	20	674	20
236 Chiaramonti	20	705	26	737	26
237 Decimomannu	20	451	3	456	3
238 Pula	20	427	10	437	10
239 Sant'Antioco	20	787	12	778	12
240 Nuoro	20	590	17	593	18
241 Oristano	20	610	17	583	17
242 Cagliari	20	521	11	504	10
243 Sassari	20	435	8	436	8
244 San Candido/Innichen	41	511	19	510	19
245 Nalles/Nals	41	511	15	532	16
246 Sarentino/Sarntal	41	540	19	522	19
247 Laives/Leifers	41	466	6	455	6
248 Merano/Meran	41	468	8	467	8
249 Bolzano/Bozen	41	635	4	610	4
250 Givo	42	584	18	596	18
251 Faver	42	345	8	355	8
252 Ala	42	490	15	494	15

segue

Tavola 2.1 - Campione stimato e campione effettivo per comune (valori assoluti)

Comune	Cod. Reg.	Campione stimato (a)		Campione effettivo (b)	
		N° famiglie	N° sezioni	N° famiglie	N° sezioni
253 Arco	42	373	5	373	5
254 Rovereto	42	465	11	477	11
255 Trento	42	671	8	701	8
TOTALE		152.779	2.498	153.844	2.507
TOTALE (esclusi i caduti)		150.946	2.480	152.012	2.488

(a) All. B – Circolare n.17 dati riferiti al Censimento 2001.

(b) Dati riferiti alle famiglie da Liste Anagrafiche Comunali riferita al 31 dicembre 2010.

Per quanto riguarda la comunicazione degli atti procedurali alla rete di rilevazione, sono stati inviati solo due atti ufficiali: una circolare e un’informativa. Tutte le altre specifiche procedurali, relative alla conduzione della rilevazione, indirizzate principalmente ai Responsabili degli UCC, sono state comunicate attraverso una casella di posta elettronica dedicata alla rilevazione²⁶. Le operazioni di comunicazione sia degli atti ufficiali sia delle mail meno formali, hanno tenuto presente un principio fondamentale per la buona riuscita dell’indagine: la diffusione capillare delle informazioni verso tutti gli attori interessati. Infatti, oltre ad essere indirizzati per conoscenza a tutti i responsabili della rete, sono stati pubblicati avvisi sul loro contenuto sulle “Bacheche” del SISTAN e/o sul Sistema di Gestione della Rilevazione (SGRPES).

La Circolare n. 17²⁷, indirizzata ai Responsabili degli Uffici di Censimento ai vari livelli territoriali, al Responsabile dell’Ufficio di statistica del Ministero dell’Interno e al Ministero dell’Interno, dopo una premessa normativa sulle motivazioni per le quali l’Istat effettua la rilevazione, ha trattato le seguenti tematiche:

- ▶ tempi, metodi e tecniche della rilevazione;
- ▶ organizzazione della rilevazione;
- ▶ modalità di reclutamento, di impiego e assicurazione dei Rilevatori;
- ▶ contributo agli UCC per lo svolgimento della rilevazione;
- ▶ protezione dei dati personali e segreto statistico.

L’Informativa²⁸ n.1, inviata a tutti i Responsabili degli UCC e per conoscenza ai Responsabili URC e al Ministero dell’Interno, aveva la finalità di fornire agli UCC le istruzioni utili alla conclusione della rilevazione PES. L’informativa ha trattato i seguenti punti:

- ▶ termine dell’indagine PES;
- ▶ operazioni preliminari alla fase di ritiro dei pacchi a cura dell’UCC;
- ▶ preparazione e confezionamento in scatole dei questionari per la restituzione;
- ▶ confezionamento dei questionari in scatole;
- ▶ distinte di spedizione e documento di trasporto;
- ▶ questionari e altro materiale non utilizzato;
- ▶ questionari in lingua slovena e tedesca.

²⁶ Il processo di comunicazione con la rete di rilevazione non si è esaurito con tali atti, infatti, attraverso l’indirizzo di posta elettronica dedicato, le Bacheche del SISTAN e il numero telefonico dedicato, lo scambio di comunicazioni ha interessato tutto l’arco temporale relativo all’indagine.

²⁷ Vedi nota n.20.

²⁸ Informativa n.1 del 25 giugno 2012.

- Come già previsto dal PGC per il Censimento della Popolazione e delle abitazioni, anche la rilevazione PES è stata gestita con l'ausilio di un Sistema di Gestione della Rilevazione (SGRPES) realizzato attraverso la rete Internet e volto a facilitare il monitoraggio di tutte le fasi del processo produttivo ai vari livelli di operatività della rete di rilevazione, compresa l'acquisizione dei questionari. SGRPES ha consentito di seguire le fasi della rilevazione prima, durante e dopo la raccolta dei dati. Le scadenze che hanno subito variazioni durante la rilevazione non hanno inficiato il suo buon esito, grazie all'organizzazione della rilevazione di tipo dinamico, orientata alla gestione dell'imprevisto.

Prospetto 2.1 - Calendario delle attività della PES 2011

Attività	Scadenze	
	previste (a)	effettive
Definizione degli UCC coinvolti e prosecuzione delle attività di Rilevatori ed eventuali coordinatori	Entro gennaio 2012	Entro gennaio 2012
Richiesta di prosecuzione delle attività dei Responsabili Istat territoriali (RIT) presso gli UURR	Entro gennaio 2012	Entro gennaio 2012
Produzione materiale censuario (questionario, lettera informativa, guida, manuale) e relative traduzioni	Entro gennaio 2012	Entro gennaio 2012
Stampa materiale censuario (questionario e manuale)	Entro febbraio 2012	Entro marzo 2012
Invio ai responsabili degli UURR del manuale in versione pdf	Entro febbraio 2012	Entro febbraio 2012
Rilascio SGRPES - Sistema di monitoraggio della rilevazione	Entro febbraio 2012	4 aprile 2012
Formazione ai responsabili degli URC	1-15 marzo 2012	8-9 marzo 2012
Consegna materiale censuario agli UCC, comprensivo di scorte	Entro marzo 2012	Dal 2 all'11 aprile 2012
Contributo agli UCC - versamento acconto	Entro marzo 2012	Aprile – maggio 2012
Inizio indagine (data di riferimento 9 ottobre 2011)	11 aprile 2012	11 aprile 2012
Raccolta dati	11 aprile - 31 maggio 2012	11 aprile - 30 giugno 2012
Revisione e controllo dei questionari da parte degli UCC	Entro 15 giugno 2012	Entro il 30 giugno 2012 (b)
Confezionamento pacchi/plichi	Entro 30 giugno 2012	Dal 30 giugno al 15 luglio
Ritiro pacchi c/o UCC	Entro 15 luglio 2012	Dal 16 luglio al 3 agosto
Contributo agli UCC - versamento saldo	Entro 15 luglio 2012	12 settembre 2012
Registrazione a cura della ditta aggiudicataria	Da luglio a ottobre 2012	Da ottobre a dicembre 2012
Analisi dei primi risultati	Entro dicembre 2013	Da gennaio a settembre 2013
Record linkage con dati censuari	Dal gennaio a dicembre 2013	Da settembre a dicembre 2013
Rilascio dati a EUROSTAT	Entro maggio 2014	Entro maggio 2014
Diffusione dei dati definitivi in Italia	Entro dicembre 2014	Entro dicembre 2014

(a) All. A - Circolare n.17.

(b) Alcuni comuni hanno protratto la fase di revisione e controllo dei questionari fino al 15 luglio 2012; solo cinque comuni hanno terminato entro il 27 luglio 2012.

2.2 Ruolo e compiti dell'Istat

In attuazione del Regolamento (CE) n. 763/08 del Parlamento europeo e del Consiglio del 9 luglio 2008 e del Regolamento (UE) n. 1151/2010 della Commissione dell'8 dicembre 2010 l'Istat è obbligato a svolgere la "Rilevazione campionaria di controllo della copertura e della qualità del 15° Censimento generale della popolazione e delle abitazioni" al fine di effettuare una valutazione della copertura del Censimento.

All'Istat, a livello nazionale e in generale durante le fasi di rilevazione spettano i seguenti compiti:

- ▶ definizione e pianificazione degli aspetti organizzativi, tecnici e metodologici sovrintendendo alle operazioni;
- ▶ coordinamento delle operazioni di raccolta dei dati;
- ▶ adozione delle circolari, dei provvedimenti e delle misure necessarie per garantire il buon andamento della rilevazione;
- ▶ monitoraggio;
- ▶ trattamento informatico dei dati mediante l'applicazione di idonei metodi di controllo e correzione;
- ▶ validazione dei dati raccolti dagli organi di Censimento;
- ▶ diffusione e trasmissione dei risultati ad Eurostat.

Inoltre:

- ▶ definizione del piano di formazione e predisposizione del materiale didattico;
- ▶ organizzazione e svolgimento della formazione;
- ▶ realizzazione dell'architettura informatica di SGRPES attraverso la rete Internet e mantenimento del servizio in esercizio costante durante il periodo della rilevazione;
- ▶ svolgimento dell'attività di supporto tecnico agli organi di censimento coinvolti nella rilevazione di qualità anche mediante la gestione di un numero telefonico dedicato;
- ▶ vigilanza sul rispetto delle istruzioni tecniche e metodologiche impartite mediante le circolari e il manuale;
- ▶ intervento nelle eventuali situazioni di particolare difficoltà;
- ▶ trasferimento agli organi di censimento dei contributi spettanti in base a quanto previsto dal PGC e dalla circolare n. 17.

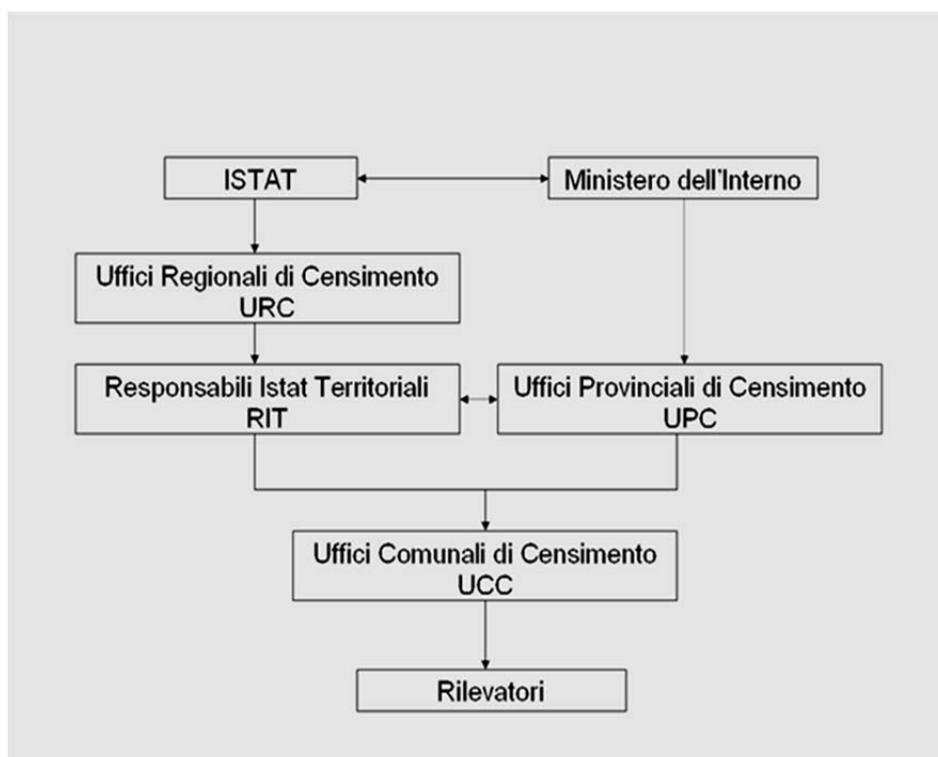
L'Istat ha anche il compito di predisporre il materiale della rilevazione e di assicurare il rispetto del segreto statistico e la tutela della riservatezza delle informazioni personali acquisite presso i rispondenti.

L'Istat, per svolgere il suo ruolo e assolvere ai suoi compiti, ha organizzato una rete di rilevazione a più livelli territoriali. A livello nazionale è stato affiancato dal Ministero dell'Interno²⁹ che ha avuto il ruolo di coordinamento delle attività svolte nell'ambito della rilevazione dagli Uffici Provinciali di Censimento costituiti presso le prefetture; a livello regionale è stata coadiuvata dagli Uffici Regionali di Censimento (URC), a livello provinciale hanno operato i Responsabili Istat Territoriali (RIT) e gli Uffici Provinciali di Censimento (UPC); infine, a livello comunale, gli Uffici Comunali di Censimento (UCC) e i Rilevatori (Figura 2.1).

²⁹ L'Ufficio di statistica del Ministero dell'Interno, in quanto appartenente al Sistema statistico nazionale, svolge correntemente le funzioni dettate dal D.Lgs. n. 322/89 e dalla direttiva del Comstat n. 5 del 15 ottobre 1991.

Il ruolo dell'Istat relativamente al Censimento della popolazione e delle abitazioni del 2011 (inclusa la Rilevazione PES) è stato di organizzare le operazioni censuarie disciplinate dal PGC, da apposite circolari, da specifiche intese e dall'art. 50 del D.lgs. n.78/2010, convertito con modificazioni in legge n. 122/2010 che lo ha indetto e finanziato, nonché di sovrintendere alle relative attività.

Figura 2-1 - Organizzazione della rete di rilevazione



Per organizzare e coordinare la PES, l'Istat ha costituito al suo interno un apposito Gruppo di Lavoro (GdL) con l'obiettivo di predisporre gli strumenti tecnici e gli atti organizzativi e di eseguire tutte le fasi di rilevazione fino all'elaborazione dei risultati finali³⁰.

Il gruppo di lavoro, interdipartimentale, costituito con delibera n. 11/DICA del 14/11/2011, è composto complessivamente da 41 membri, ricercatori, tecnologi e collaboratori tecnici. I membri del GdL appartengono al Dipartimento per i censimenti e gli archivi amministrativi e statistici (DICA), al Dipartimento per l'integrazione, la qualità e lo sviluppo delle reti di produzione e ricerca (DIQR) e alla Direzione centrale per le tecnologie dell'informazione e della comunicazione (DCIT). Tra i membri sono presenti due coordinatori: uno per gli aspetti organizzativi e gestionali e un altro per gli aspetti metodologici.

Il Gruppo di lavoro si articola in sei aree, ognuna delle quali ha un coordinatore scelto tra i membri della stessa:

³⁰ La progettazione della rilevazione, è stata curata da un Gruppo di lavoro, interdipartimentale, costituito, precedentemente, con delibera n. 144/DPTS del 21 luglio 2011, i cui membri appartengono alla Direzione Centrale dei Censimenti Generali (DCCG) e alla Direzione Centrale per le Tecnologie e il supporto metodologico (DCMT).

- ▶ l'area "*Organizzazione e comunicazione della rete*" cura gli aspetti organizzativi e la comunicazione di tutte le fasi della rilevazione. In particolare, la predisposizione delle apposite circolari; l'organizzazione della formazione e la predisposizione del materiale ad essa connesso e del manuale delle istruzioni relativamente all'organizzazione e ai compiti della rete di rilevazione; i rapporti/contratti con le società di servizi esterne coinvolte nella rilevazione; l'organizzazione della rete di rilevazione e i rapporti con essa (FAQ, Bacheche e numero telefonico dedicato); l'organizzazione della stampa e la spedizione del materiale alla rete, il suo ritiro e il relativo monitoraggio. Durante la rilevazione sul campo cura, mediante SGRPES, il monitoraggio dei questionari acquisiti e l'erogazione dei contributi ai comuni campione coinvolti nella rilevazione;
- ▶ l'area "*Produzione materiale censuario e conduzione sul campo*" cura la predisposizione della parte tecnica del materiale della formazione relativamente alle unità di rilevazione e ai contenuti del questionario; la predisposizione del questionario di rilevazione e del manuale delle istruzioni, quest'ultimo per la parte dei contenuti del questionario e della relativa compilazione, alcuni aspetti del monitoraggio durante la raccolta dei dati; fornisce supporto tecnico alla rete di rilevazione attraverso il numero telefonico dedicato, FAQ e Bacheche;
- ▶ l'area "*Georeferenziazione*" cura l'aggiornamento della cartografia a supporto della rilevazione e la relativa implementazione delle mappe, fisiche e digitali, nonché la generazione degli itinerari di sezione nel sistema di gestione della rilevazione;
- ▶ l'area "*Informatica (SGRPES)*" cura la predisposizione di parte del materiale della formazione e del manuale di istruzioni relativamente al Sistema di Gestione della Rilevazione a supporto dell'indagine; la progettazione e realizzazione dell'architettura informatica per la rilevazione; i test di prova del sistema con la rete di rilevazione; l'aggiornamento informatico in itinere; alcuni aspetti del monitoraggio della rilevazione sul campo;
- ▶ l'area "*Campionamento, stima e analisi dati*" cura gli aspetti metodologici della rilevazione ovvero l'estrazione del campione, l'errore campionario, le stime e i riporti all'universo di riferimento e l'analisi dei risultati ai fini della loro diffusione;
- ▶ l'area "*Record linkage e analisi dei dati*" cura il *match/linkage* dei risultati dell'indagine di copertura con i risultati del Censimento della Popolazione ai fini della valutazione dell'eventuale sottocopertura del Censimento stesso e l'analisi dei risultati.

2.3 Ruolo e compiti della rete e operazioni sul campo

La rete di rilevazione designata per l'esecuzione delle operazioni sul campo della Rilevazione PES, è stata costituita dagli Uffici Regionali di Censimento (URC) e dai loro Responsabili Istat Territoriali (RIT), dagli Uffici Provinciali di Censimento (UPC) e dagli Uffici Comunali di Censimento (UCC) in forma singola o associata.

Gli Uffici di Censimento, chiamati al rispetto delle modalità stabilite dalle Circolari n. 3 del 3 marzo 2011, prot. n.1896 e n. 17 del 23 gennaio 2012, prot. n. 1990, sono rimasti in funzione sino alla conclusione delle attività di rilevazione.

A livello regionale gli URC e i RIT hanno coadiuvato l'Istat nel sovrintendere al funzionamento della rete territoriale svolgendo i seguenti compiti:

hanno coordinato e partecipato alla formazione del personale degli organi di rilevazione;

- ▶ hanno vigilato sul regolare e tempestivo svolgimento delle operazioni di rilevazione nel territorio regionale di competenza;
- ▶ hanno supervisionato la corretta assegnazione dei rilevatori comunali alle sezioni di censimento;

- ▶ hanno monitorato l'andamento della rilevazione utilizzando le funzioni di reportistica previste in SGRPES;
- ▶ hanno coordinato le attività di rilevazione sul territorio di competenza in collaborazione con gli UPC.

Gli Uffici Regionali di Censimento, uno per ogni regione, sono stati costituiti in occasione del Censimento della popolazione presso ogni Ufficio Territoriale dell'Istat, ad eccezione della regione Valle d'Aosta e delle Province di Trento e di Bolzano³¹. L'URC della Valle d'Aosta è stato costituito presso l'Ufficio di statistica della Regione Valle d'Aosta a seguito dell'intesa tra Istat e Regione Autonoma Valle d'Aosta. Gli Uffici di Censimento delle Province di Trento e di Bolzano sono stati costituiti presso il Servizio Statistica delle rispettive Province autonome e hanno svolto per il territorio di competenza le funzioni di Ufficio Regionale di Censimento. I 21 URC hanno coinvolto nella rilevazione complessivamente 53 operatori³², di cui 32 Responsabili Istat Territoriali, ciascuno responsabile in media di otto Uffici Comunali di Censimento. I RIT lombardi hanno seguito, ciascuno, in media 11 UCC, seguiti dai RIT marchigiani e da quelli calabresi che ne hanno seguiti in media dieci. Nella Valle d'Aosta e nel Lazio si riscontrano le medie più basse: i RIT valdostani hanno seguito, ognuno, in media, tre UCC, mentre, i RIT laziali quattro (Tavola 2).

Tavola 2.2 - Operatori degli Uffici Regionali di Censimento e media UCC per RIT per regione e Italia (valori assoluti)

Regione	Operatori Uffici Regionali di Censimento (URC)			
	Responsabile	Responsabile Istat Territoriale (RIT)	Totale	Media UCC per RIT
Piemonte	1	2	3	8,0
Valle d'Aosta/ <i>Vallée d'Aoste</i>	1	1	2	3,0
Liguria	1	1	2	8,0
Lombardia	1	3	4	10,7
Provincia di Bolzano/ <i>Bozen</i>	1	1	2	6,0
Provincia di Trento	1	1	2	6,0
Veneto	1	2	3	9,5
Friuli-Venezia Giulia	1	1	2	9,0
Emilia-Romagna	1	2	3	8,0
Toscana	1	2	3	8,5
Umbria	1	1	2	7,0
Marche	1	1	2	10,0
Lazio	1	3	4	4,0
Abruzzo	1	1	2	8,0
Molise	1	1	2	5,0
Campania	1	2	3	9,5
Puglia	1	2	3	7,5
Basilicata	1	1	2	6,0
Calabria	1	1	2	10,0
Sicilia	1	2	3	8,5
Sardegna	1	1	2	9,0
ITALIA	21	32	53	7,8

³¹ Come stabilito dal PGC e dalla circolare n. 17 del 23 gennaio 2012 prot. n. 1990.

³² Con il termine operatori ci si riferisce al personale inserito nel Sistema di Gestione della Rilevazione (SGRPES) da ogni Responsabile di Ufficio di Censimento (Comunale, Provinciale e Regionale).

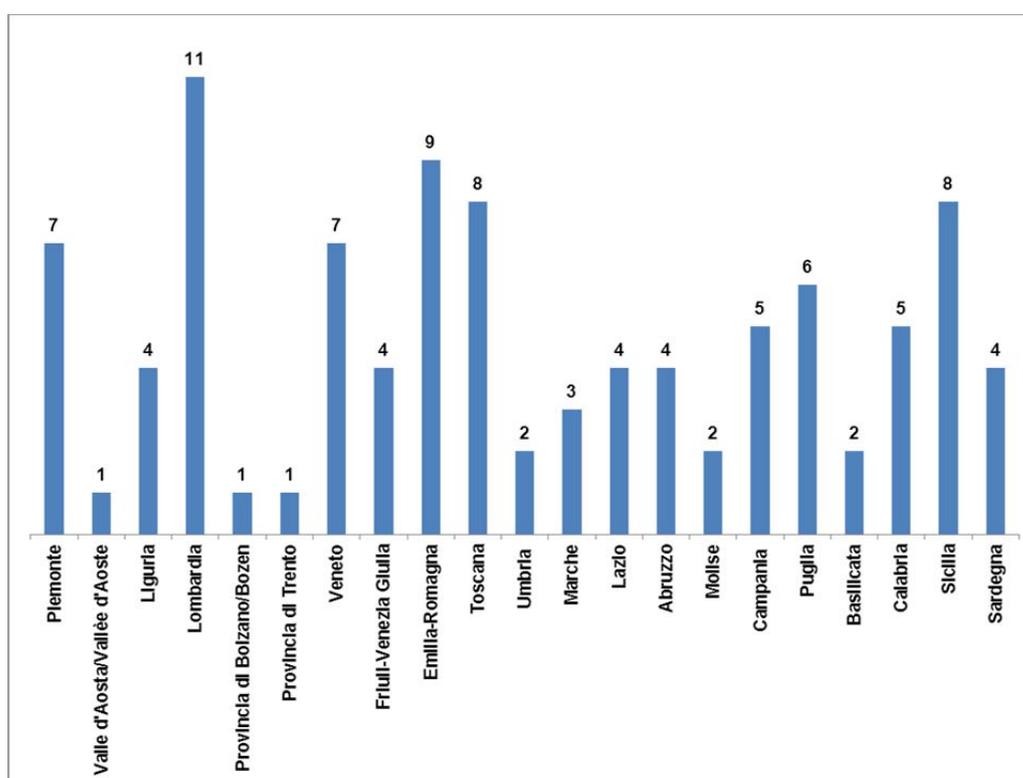
A livello provinciale gli UPC hanno supportato gli URC e i RIT nella conduzione della rilevazione e hanno assolto i seguenti compiti:

- ▶ hanno vigilato sul corretto svolgimento delle operazioni;
- ▶ hanno assistito gli UCC nella soluzione di eventi imprevisti o eccezionali;
- ▶ hanno monitorato l'andamento della rilevazione utilizzando le funzioni di reportistica previste in SGRPES;
- ▶ hanno coordinato le attività di rilevazione sul territorio di competenza in collaborazione con i RIT.

Gli Uffici Provinciali di Censimento sono stati costituiti presso gli Uffici di statistica delle Prefetture, ad eccezione della regione Valle d'Aosta, della provincia di Trento e della provincia di Bolzano³³. Per quanto riguarda la regione Valle d'Aosta l'URC è stato costituito presso l'Ufficio statistica della Regione Autonoma Valle d'Aosta a seguito dell'intesa tra Istat e Regione Autonoma Valle d'Aosta e ha assunto anche i compiti di UPC. Nelle Province di Trento e di Bolzano l'Ufficio di Censimento costituito presso il Servizio Statistica delle rispettive Province autonome, ha svolto le funzioni di UPC.

I 98 UPC hanno rappresentato quasi tutti gli UCC. Sono state infatti rappresentate tutte le province dei comuni campione, tranne la provincia sarda di Carbonia Iglesias, di recente costituzione e priva di Prefettura (Figura 2.2).

Figura 2-2 - Numero di Uffici Provinciali di Censimento per regione



³³ Come stabilito dal PGC e dalla circolare n. 17 del 23 gennaio 2012 prot. n. 1990.

A livello comunale gli UCC e i loro responsabili hanno svolto le fasi di rilevazione sul campo assolvendo i seguenti compiti:

- ▶ hanno selezionato, nominato e assegnato i rilevatori alle sezioni di censimento che costituiscono il campione di secondo stadio, rispettando i criteri di indipendenza dalla rilevazione censuaria;
- ▶ hanno organizzato l'attività di formazione ai rilevatori, in collaborazione con l'URC competente per territorio;
- ▶ hanno organizzato e svolto la rilevazione sul territorio nel rispetto delle modalità e dei tempi disposti dall'Istat;
- ▶ hanno utilizzato SGRPES per tutte le funzioni previste;
- ▶ hanno monitorato l'andamento della rilevazione sul campo e sono intervenuti nei casi di criticità;
- ▶ hanno assistito le famiglie nella compilazione dei questionari;
- ▶ hanno effettuato la revisione dei questionari;
- ▶ hanno curato il confezionamento del materiale cartaceo da consegnare al vettore dopo la conclusione della rilevazione secondo le indicazioni fornite;
- ▶ hanno assicurato il buon andamento delle operazioni di rilevazione nel territorio di competenza.

Gli Uffici Comunali di Censimento sono stati costituiti dai comuni, in forma singola o associata, per le specifiche finalità censuarie. Hanno partecipato alla rilevazione gli UCC sia in forma singola, sia associata con uno o più comuni campione. I 250 Uffici Comunali di Censimento presenti sul territorio nazionale hanno coinvolto nella rilevazione complessivamente 1.442 operatori: 250 Responsabili, 1.160 Rilevatori e 32 operatori impiegati in altre operazioni censuarie. Gli UCC della Lombardia e del Lazio, rispetto al complesso nazionale, hanno coinvolto la maggiore quota di operatori: essi hanno impegnato il 12,8 per cento del totale dei rilevatori in Lombardia e il 14,2 per cento nel Lazio.(Tavola 3).

Gli UCC hanno potuto servirsi degli stessi rilevatori che hanno partecipato alle operazioni del Censimento della popolazione. In tal caso, per assicurare l'indipendenza tra la PES e il Censimento è stato assolutamente necessario assegnare a ciascun Rilevatore una o più sezioni o parte di essa differenti, rispetto a quella/e assegnata/e durante le operazioni censuarie.

Il Responsabile dell'UCC, in considerazione del contenuto numero di sezioni di Censimento sottoposte a rilevazione in ciascun comune campione, ha svolto anche la funzione di Coordinatore. Il numero dei rilevatori selezionati è stato comunicato da ciascun Responsabile dell'UCC all'Istat attraverso la casella di posta elettronica dedicata.

Il Responsabile dell'UCC ha fornito a ciascun Rilevatore una *user-id* e una *password* personali con le quali accedere all'applicazione per registrare le informazioni raccolte durante la rilevazione.

I rilevatori hanno eseguito i seguenti compiti:

- ▶ hanno partecipato alle riunioni di formazione ed acquisito le necessarie conoscenze e abilità attraverso i diversi strumenti formativi messi a disposizione dall'Istat;
- ▶ hanno gestito quotidianamente, mediante l'uso di SGRPES, il Diario della/e sezione/i di Censimento ad essi assegnata/e dal Responsabile dell'UCC;
- ▶ hanno consegnato alle famiglie i questionari da compilare;
- ▶ hanno sollecitato le famiglie non ancora rispondenti nel rispetto dei tempi e delle modalità definite dal Responsabile dell'UCC;
- ▶ hanno provveduto alla raccolta dei questionari presso le famiglie, tenendo conto degli orari di presenza indicati dai rispondenti e fornendo loro informazioni su finalità e natura

obbligatoria della rilevazione, nonché assicurando loro assistenza alla compilazione del questionario, ove richiesta;

- ▶ hanno verificato la completezza e la correttezza dell'elenco degli indirizzi fornito dall'Istat (itinerari di sezione);
- ▶ hanno conteggiato le abitazioni non occupate e le abitazioni occupate esclusivamente da persone non abitualmente dimoranti;
- ▶ hanno revisionato i dati contenuti nei questionari e compilato i riquadri di propria pertinenza;
- ▶ hanno svolto ogni altro compito loro affidato dal Responsabile dell'UCC.

Tavola 2.3 - Operatori degli Uffici Comunali di Censimento per regione e Italia (valori assoluti e percentuali)

Regione e Italia	Operatori Uffici Comunali di Censimento (UCC)							
	Responsabile		Rilevatore		Altro personale		Totale	
	Valori assoluti	Valori percentuali	Valori assoluti	Valori percentuali	Valori assoluti	Valori percentuali	Valori assoluti	Valori percentuali
Piemonte	16	6,4	77	6,6	2	6,3	95	6,6
Valle d'Aosta/ <i>Vallée d'Aoste</i>	3	1,2	13	1,1	1	3,1	17	1,2
Liguria	8	3,2	32	2,8	0	0,0	40	2,8
Lombardia	32	12,8	148	12,8	8	24,9	188	12,9
Provincia di Bolzano/ <i>Bozen</i>	6	2,4	21	1,8	0	0,0	27	1,9
Provincia di Trento	6	2,4	15	1,3	0	0,0	21	1,5
Veneto	19	7,6	87	7,5	0	0,0	106	7,3
Friuli-Venezia Giulia	9	3,6	51	4,4	0	0,0	60	4,2
Emilia-Romagna	16	6,4	67	5,8	0	0,0	83	5,8
Toscana	17	6,8	62	5,3	7	21,9	86	6,0
Umbria	7	2,8	29	2,5	0	0,0	36	2,5
Marche	10	4,0	33	2,8	0	0,0	43	3,0
Lazio	12	4,8	164	14,2	8	25,0	184	12,7
Abruzzo	8	3,2	30	2,6	0	0,0	38	2,6
Molise	5	2,0	13	1,1	0	0,0	18	1,2
Campania	19	7,6	73	6,3	0	0,0	92	6,4
Puglia	15	6,0	64	5,5	0	0,0	79	5,5
Basilicata	6	2,4	37	3,2	2	6,3	45	3,1
Calabria	10	4,0	52	4,5	4	12,5	66	4,6
Sicilia	17	6,8	65	5,6	0	0,0	82	5,7
Sardegna	9	3,6	27	2,3	0	0,0	36	2,5
ITALIA	250	100,0	1.160	100,0	32	100,0	1.442	100,0

Nello svolgimento dei suddetti compiti, i Rilevatori si sono attenuti al rispetto della normativa in materia di protezione dei dati personali, secondo quanto stabilito e previsto dal “Manuale di istruzioni della Rilevazione”.

L'Istat, per tutelare i rilevatori, ha stipulato in loro favore una polizza di assicurazione infortuni integrativa connessa all'esercizio delle funzioni relative all'incarico di Rilevatore, e valida per i periodi di effettiva attività. Durante il periodo di attività sul campo non si è verificato nessun evento che richiedesse la suddetta copertura assicurativa.

Le operazioni sul campo, come la consegna dei questionari, l'aggiornamento del Diario di sezione in SGRPES e il ritiro dei modelli, hanno impegnato i Rilevatori che, quotidianamente, si sono recati nelle sezioni di censimento supportati da ausili forniti dall'Istat, come le mappe con ortofoto, le sezioni di censimento e il grafo stradale commerciale³⁴; l'elenco delle strade per sezione di censimento; il manuale della rilevazione e il Diario di sezione.

Come primo passo i rilevatori hanno proceduto alla verifica dei confini delle sezioni di censimento assegnate loro dal Responsabile dell'UCC. Successivamente, i responsabili degli UCC hanno consegnato ai rilevatori un numero congruo di questionari.

L'11 aprile 2012 è la data che ha segnato l'inizio del lavoro sul campo e che ha visto come protagonisti gli UCC, i rilevatori e i loro responsabili.

Da tale data e fino alla fine della rilevazione sul campo (30 giugno 2012), i rilevatori hanno percorso le vie della sezione di censimento consegnando alle famiglie i questionari e registrando le abitazioni; l'intestatario del Foglio di famiglia ha compilato il questionario (in alternativa, un'altra persona della famiglia o una persona temporaneamente presente nell'alloggio) e, quando richiesto, i rilevatori hanno offerto il loro supporto alla compilazione. Dopo avere ritirato e contestualmente revisionato i questionari, effettuando controlli di coerenza e completezza, i rilevatori, rientrati nella sede dell'UCC, hanno acquisito i questionari in SGRPES. Tutto ciò, mentre il Responsabile monitorava il percorso effettuato dal Rilevatore nella sezione (o sezioni) a lui assegnata, per verificare che fosse stata percorsa integralmente e che non fossero stati percorsi, erroneamente, strade o tratti appartenenti ad altre sezioni.

Le operazioni sul campo sono state condotte senza particolari problemi grazie al sistema di organizzazione della Rete, nonché grazie ai buoni rapporti di collaborazione esistenti tra i vari operatori coinvolti.

2.4 Formazione

Il processo formativo che l'Istat ha organizzato in occasione della rilevazione PES ha visto sia la predisposizione di incontri frontali in aula sia l'implementazione di strumenti in grado di garantire assistenza continua e formazione costante alla rete di rilevazione.

Sono stati organizzati due momenti di formazione frontale. Il primo, destinato agli Uffici Regionali di Censimento, si è tenuto a Roma nell'Aula Magna dell'Istat nei giorni 8 e 9 marzo 2012 e ha visto la partecipazione dei componenti degli URC, alcuni in aula, altri collegati in videoconferenza. Al secondo incontro, svoltosi a Castel San Pietro (Bo) nei giorni 20 e 21 marzo 2012 e destinato agli Uffici Provinciali di Censimento, hanno partecipato poco più della metà dei 98 UPC invitati. Il Piemonte è l'unica regione che ha registrato la partecipazione di tutti gli UPC, assenti invece quelli del Lazio, del Molise e della Basilicata. Gli UPC di Calabria, Campania e Emilia-Romagna hanno registrato le più basse percentuali di presenza in aula. Presenti gli UPC liguri, toscani, abruzzesi e sardi (per il 75 per cento); quelli della Lombardia (73 per cento circa), del Veneto (57 per cento),

³⁴ Non validato né dall'Istat né dal comune.

delle Marche e della Puglia (67 per cento ciascuna); la metà degli UPC del Friuli-Venezia Giulia, dell'Umbria e della Sicilia (Tavola 2.4).

Un caso particolare è rappresentato dalla regione Valle d'Aosta e dalle Province di Bolzano e di Trento i cui relativi Uffici di Censimento hanno svolto le funzioni di URC e UPC per i territori di competenza³⁵ e che, pertanto, hanno partecipato alla formazione, tenutasi a Roma, destinata agli URC.

Tavola 2.4 - Uffici Provinciali di Censimento presenti, assenti all'evento formativo per regione e Italia (valori percentuali)

Regione e Italia	Uffici Provinciali di Censimento (UPC)		Totale
	Incidenza di presenze alla formazione su totale UPC regionale	Incidenza di assenze alla formazione su totale UPC regionale	
Piemonte	100	0	100
Valle d'Aosta/Vallée d'Aoste (a)	0	100	100
Liguria	75	25	100
Lombardia	72,7	27,3	100
Provincia di Bolzano/ <i>Bozen</i> (a)	0	100	100
Provincia di Trento (a)	0	100	100
Veneto	57,1	42,9	100
Friuli-Venezia Giulia	50	50	100
Emilia-Romagna	33,3	66,7	100
Toscana	75	25	100
Umbria	50	50	100
Marche	66,7	33,3	100
Lazio	0	100	100
Abruzzo	75	25	100
Molise	0	100	100
Campania	40	60	100
Puglia	66,7	33,3	100
Basilicata	0	100	100
Calabria	20	80	100
Sicilia	50	50	100
Sardegna	75	25	100
ITALIA	54,1	45,9	100

(a) Gli Uffici di Censimento della regione Valle d'Aosta, delle Province di Bolzano e di Trento, svolgendo le funzioni sia di URC e sia di UPC per il territorio di loro competenza, hanno partecipato all'evento formativo che si è tenuto a Roma dedicato agli URC.

L'organizzazione della formazione a cascata ha consentito agli URC, in collaborazione con gli UPC, di organizzare giornate formative destinate agli Uffici Comunali di Censimento. Gli UCC, a loro volta, hanno fornito, nel corso dell'indagine, assistenza ai cittadini.

³⁵ Come stabilito dal PGC (paragrafo 3.1) e dalla Circolare n. 17 del 23 gennaio 2012 (paragrafo 3).

Le tematiche trattate in occasione dei due momenti formativi hanno principalmente riguardato i seguenti punti:

- ▶ l'organizzazione e i compiti della rete;
- ▶ l'analisi del questionario e le relative modalità di compilazione;
- ▶ il processo di rilevazione;
- ▶ il funzionamento di SGRPES.

Oltre a questi incontri, sono stati predisposti strumenti idonei a garantire all'intera rete di rilevazione una formazione continua e un'assistenza costante durante la rilevazione.

Più nello specifico, la rete di rilevazione ha potuto trovare risposta alle domande relative agli aspetti gestionali e tecnici della PES, tramite:

- ▶ FAQ (Frequently Asked Questions);
- ▶ Bacheche, all'indirizzo dedicato, sulle quali i RIT e i responsabili degli URC potevano porre domande e consultare le risposte fornite a quesiti posti da altri RIT o responsabili di URC;
- ▶ indirizzo e-mail dedicato;
- ▶ numero telefonico dedicato.

Attraverso questi canali sono stati richiesti chiarimenti relativi a 63 argomenti e inviati 171 messaggi tra quesiti e risposte.

Il numero di telefono dedicato alla rete di rilevazione è stato attivo dall'11 aprile 2012 al 31 maggio 2012, dal lunedì al venerdì, dalle ore 9:00 alle ore 13:00 e dalle ore 14:00 alle ore 16:00 ed ha ricevuto complessivamente 72 telefonate: 61 provenienti dagli UCC, 7 dagli URC e 4 da cittadini.

Il Piemonte è la regione dalla quale è arrivata la percentuale più elevata di chiamate, pari al 19 per cento circa, seguito dalla Campania con il 18 per cento di chiamate e dal Lazio con l'11 per cento circa. Le altre regioni hanno registrato tassi di chiamate inferiori al 10 per cento. Non sono pervenute telefonate dalla Liguria, dalla provincia di Bolzano, dalla Basilicata, dalla Sicilia e dalla Sardegna (). Nelle prime due settimane di rilevazione si sono registrate le percentuali più elevate di telefonate, rispettivamente, il 28 per cento e il 26 per cento circa. Nelle settimane successive, il tasso di telefonate è diminuito per poi iniziare ad aumentare nella settimana dal 7 all'11 maggio 2012 e crescere ulteriormente dal 14 al 18 maggio 2012, probabilmente per effetto delle comunicazioni inviate agli UCC al fine di sollecitare l'aggiornamento dello stato dei questionari in SGRPES. Nelle ultime due settimane di attività si sono registrate percentuali di chiamate molto basse.

Il maggior numero di telefonate si è registrato nella prima fascia oraria (dalle 9:00 alle 11:00) per poi decrescere sensibilmente nelle fasce orarie successive (51,4 per cento); nella seconda mattinata (dalle 11:01 alle 13:00) la percentuale è diminuita notevolmente (33,3 per cento), per poi ridursi sensibilmente nella fascia oraria pomeridiana in cui è pervenuto il 15,3 per cento circa di contatti.

Figura 2.3 - Telefonate ricevute dal numero dedicato alla rete di rilevazione per provenienza regionale (valori percentuali)

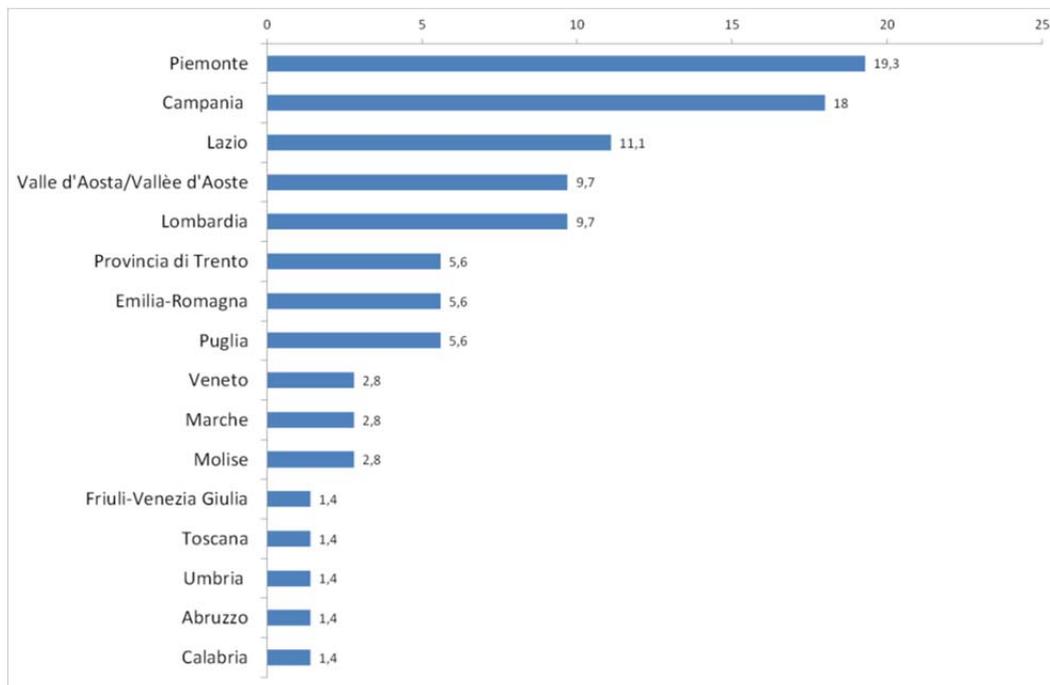


Figura 2.4 - Telefonate ricevute dal numero dedicato alla rete di rilevazione per settimana di rilevazione (valori percentuali)

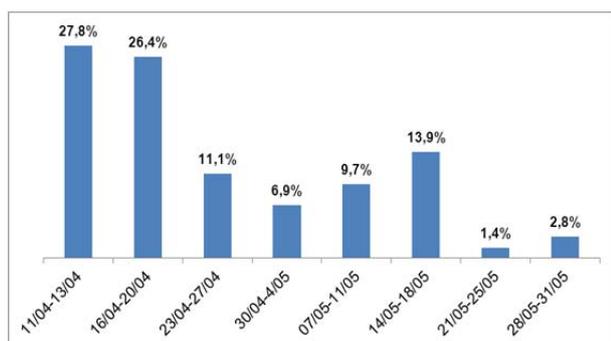
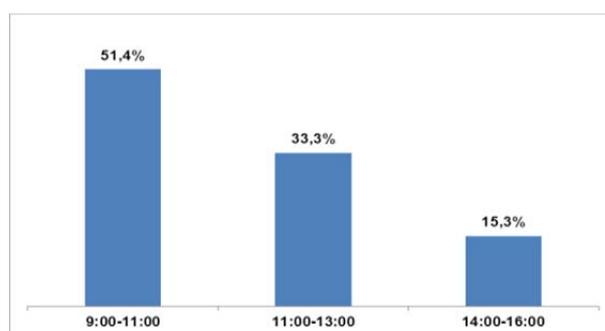


Figura 2.5 - Telefonate ricevute dal numero dedicato alla rete di rilevazione per fascia oraria di ricezione (valori percentuali)



Sia gli strumenti di assistenza sia gli incontri formativi hanno supportato proficuamente la rete di rilevazione, non solo durante la fase di raccolta dei dati, ma anche nei periodi immediatamente precedenti e successivi. La formazione PES, per le sue caratteristiche di formazione a cascata e formazione costante, ha permesso l'aggiornamento continuo della rete di rilevazione, e di conseguenza dei cittadini, rispondendo in tempo reale alle richieste di chiarimento e di approfondimento.

2.5 Monitoraggio del materiale della rilevazione

Per consentire lo svolgimento delle operazioni sul territorio, l'Istat, dopo aver predisposto il materiale della rilevazione, ha previsto il suo trasporto e la consegna presso gli Uffici Comunali di Censimento (UCC) e presso le altre strutture interessate della rete di rilevazione. L'Istat ha inoltre programmato il ritiro dei questionari compilati presso le sedi degli UCC e il conseguente trasporto presso la sede Istat di Roma, per lo svolgimento della successiva attività di registrazione manuale. Postel Spa, nel rispetto delle prescrizioni contrattuali³⁶, ha svolto attività di pre stampa, stampa, allestimento e confezionamento del materiale, mentre Poste Italiane Spa si è occupata delle attività di trasporto e consegna dei colli presso gli UCC e presso le altre strutture territoriali interessate (sede Istat di Roma e sede Anusca di Castel San Pietro, presso le quali si sono svolte due giornate formative, dedicate rispettivamente ai componenti degli Uffici Regionali di Censimento la prima e degli Uffici provinciali di Censimento la seconda); inoltre, ha effettuato il ritiro dei questionari compilati presso le sedi degli UCC e il trasporto presso la sede Istat di Roma. Nota a parte meritano i questionari predisposti a tutela delle minoranze linguistiche tedesca (nei comuni campione della provincia di Bolzano) e slovena (in alcuni comuni campione del Friuli-Venezia Giulia).

La traduzione in tedesco è stata condotta autonomamente dalla Provincia autonoma di Bolzano, mentre la traduzione in lingua slovena è stata gestita dall'Istat, servendosi delle competenze del personale interno. Ciò ha comportato la posticipazione della fase di spedizione dei questionari tradotti in sloveno³⁷. Tuttavia, nonostante le difficoltà incontrate, la rilevazione nei comuni caratterizzati dalla presenza della minoranza linguistica slovena si è conclusa nel rispetto dei termini fissati dall'Istat.

Il materiale inviato a ciascun Ufficio Comunale di Censimento consisteva in:

- ▶ questionari della rilevazione (Mod.Istat.PES.1);
- ▶ buste di restituzione dei questionari compilati;
- ▶ manuale della rilevazione;
- ▶ ricevute di avvenuta restituzione del questionario;
- ▶ tesserini di riconoscimento.

Le quantità inviate sono state calcolate in base a stime riguardanti il numero delle sezioni di Censimento e il numero delle famiglie³⁸, il numero presunto di rilevatori impiegati, il numero dei responsabili degli UPC e degli URC e, infine, il numero dei RIT coinvolti nella rilevazione.

L'Istat ha ordinato, complessivamente, oltre 183 mila fogli di famiglia e altrettante buste di restituzione e ricevute di ritorno, comprendendo un 20 per cento circa di materiale di scorta, oltre 3.700 manuali e oltre 1.800 tesserini di riconoscimento per i rilevatori. Nel corso della rilevazione sul campo, alcuni comuni hanno riscontrato una carenza di modelli di rilevazione derivante principalmente dallo smarrimento o dal deterioramento dei questionari da parte delle famiglie contattate dal Rilevatore. In tali casi, il Rilevatore ha provveduto alla sostituzione del modello, attingendo dalle scorte.

Nei casi in cui anche le scorte si sono rivelate insufficienti, si è resa necessaria un'integrazione dei quantitativi di materiale non sufficienti per far fronte alle richieste formulate dai responsabili degli UCC e dai RIT, che hanno svolto le funzioni di raccordo tra l'Istat centrale e gli UCC coinvolti. In particolare, le segnalazioni sono pervenute via mail, all'indirizzo di posta elettronica dedicato alla

³⁶ Per maggiori approfondimenti vedere il precedente paragrafo 0, 1.4 Gestione dei servizi in *outsourcing* .

³⁷ I questionari tradotti in sloveno sono stati recapitati ai Comuni interessati il 1° giugno 2012.

³⁸ Indicati nell'Allegato B alla Circolare n. 17 (prot. n. 1990 del 23 gennaio 2012).

rilevazione; telefonicamente, attraverso il numero telefonico dedicato e utilizzando il forum “Bacheche” riservato ai RIT e ai responsabili degli URC.

Per fare fronte alle richieste inoltrate dai comuni, l'Istat ha innanzitutto attinto dalle scorte in giacenza presso la sua sede di Roma³⁹, e quindi ha distribuito piccoli quantitativi di questionari e ha evaso tutte le richieste relative alla carenza di manuali e di ricevute di avvenuta restituzione dei questionari compilati.

Grazie all'intervento dei RIT, che hanno effettuato, presso gli UCC, il monitoraggio dei quantitativi e della tipologia del materiale necessario e del materiale inutilizzato, l'Istat ha potuto redistribuire i questionari tra comuni limitrofi. Infine, l'Istat ha richiesto⁴⁰ a Postel Spa una lavorazione aggiuntiva relativa alla stampa, all'allestimento e alla consegna di ulteriori 7 mila questionari. Non è stato invece necessario integrare proporzionalmente le quantità delle ricevute di avvenuta restituzione dei questionari compilati, in quanto i comuni si sono serviti delle scorte del Censimento rimaste inutilizzate.

Il materiale è stato consegnato ai comuni richiedenti nel rispetto del termine della rilevazione, prorogato al 30 giugno 2012 attraverso una comunicazione inviata via mail ai responsabili degli UCC e ai Responsabili degli altri organismi della rete di rilevazione, in data 29 maggio 2012. Nel complesso, il 38,2 per cento dei manuali prodotti è stato consegnato all'Istat che ha provveduto a distribuirli agli URC e agli UPC durante i due momenti formativi, tenendone una congrua quantità per uso interno. Agli UCC del Lazio, del Veneto e della Lombardia sono state consegnate le maggiori quantità di manuali e di tesserini di riconoscimento per i rilevatori: il 7,3 per cento e il 10,4 per cento in Lombardia, il 6,2 per cento e 15,7 per cento nel Lazio, il 5,1 per cento e il 10,6 per cento nel Veneto rispettivamente.

La quantità maggiore di fogli di famiglia, buste di restituzione e ricevute di ritorno è stata consegnata agli UCC della Lombardia (13 per cento circa di ogni tipologia di materiale), seguiti dagli UCC della Campania e del Lazio (8 per cento circa), della Sicilia e del Veneto (7 per cento circa).

A ridosso della conclusione delle operazioni di revisione e controllo dei questionari da parte degli UCC, l'Istat ha rilasciato un'informativa⁴¹ con la quale sono state fornite ai responsabili degli UCC le istruzioni per il confezionamento dei questionari cartacei e la consegna dei plichi al vettore Poste Italiane Spa incaricato del ritiro e del trasporto.

Il materiale è stato restituito da ciascun comune in un'unica soluzione, nella fascia temporale tra il 16 luglio e il 3 agosto 2012. In particolare, al termine delle operazioni di revisione, controllo e registrazione dei questionari in SGRPES, i questionari cartacei sono stati inseriti, a cura del Responsabile dell'UCC, in scatole di cartone. Da ogni questionario, nel rispetto delle norme poste a tutela del segreto statistico e della riservatezza dei dati personali, sono stati separati i primi due fogli (contenenti l'indirizzo dell'intestatario del foglio di famiglia e la lista con i dati anagrafici dei componenti della famiglia dimoranti abitualmente nell'alloggio) dalla restante parte del questionario, in modo da inserire i due plichi in scatole diverse.

³⁹ Le scorte a disposizione della sede Istat di Roma si sono arricchite, inoltre, del materiale restituito dai tre comuni campione che non hanno partecipato alla rilevazione (Formello, Tivoli e Veggiano).

⁴⁰ La richiesta di lavorazione aggiuntiva è stata inoltrata a Postel in data 5 giugno 2012.

⁴¹ Informativa n. 1 del 25 giugno 2012.

Tavola 2.5 - Allestimento stampa e consegna materiale per regione e Italia (valori percentuali)

Regione e Italia	Stampa e consegna materiale per rilevazione					
	Mod.Istat. Pes1 (a)	CP.RE. PES (b)	Mod.Istat. Pes.RIC (c)	Mod.Istat Pes1_SLV (d)	Mod.Istat. Pes_MAN (e)	Mod.Istat. Pes_RIL (f)
Piemonte	6,5	6,3	6,3	-	3,6	4,5
Valle d'Aosta/Vallée d'Aoste	1,2	1,2	1,2	-	0,8	1,5
Liguria	3,5	3,3	3,3	-	1,9	2,9
Lombardia	12,9	12,9	12,9	-	7,3	10,4
Provincia di Bolzano/Bozen	2,0	2,0	2,0	-	1,3	1,6
Provincia di Trento	1,8	1,9	1,9	-	1,3	1,5
Veneto	7,0	7,2	7,2	-	5,1	10,6
Friuli-Venezia Giulia	2,8	2,9	2,9	100,0	2,2	4,1
Emilia-Romagna	6,6	6,6	6,6	-	4,1	5,7
Toscana	5,6	5,8	5,8	-	3,7	4,2
Umbria	2,4	2,5	2,5	-	1,6	3,2
Marche	3,3	3,3	3,3	-	2,2	2,2
Lazio	7,7	8,0	8,0	-	6,2	15,7
Abruzzo	2,9	3,0	3,0	-	1,8	2,9
Molise	1,6	1,7	1,7	-	1,1	1,2
Campania	8,2	8,3	8,3	-	4,4	5,8
Puglia	6,3	6,2	6,2	-	3,5	6,4
Basilicata	2,1	2,1	2,1	-	1,5	3,3
Calabria	3,8	3,7	3,7	-	2,2	4,0
Sicilia	7,3	7,3	7,3	-	4,0	5,9
Sardegna	3,5	3,4	3,4	-	2,0	2,4
Istat Roma	1,0	0,4	0,4	-	38,2	0,0
ITALIA	100,0	100,0	100,0	100,0	100,0	100,0

Nel prospetto è compreso il materiale dei tre comuni campione caduti, non è compreso il materiale in lingua tedesca relativo alla provincia di Bolzano, predisposto autonomamente dalla Provincia autonoma, e non sono considerate le redistribuzioni intermedie.

(a) Mod.Istat.Pes1: foglio di famiglia comprensivo della lettera informativa e della guida alla compilazione;

(b) CP.RE.PES: buste di restituzione;

(c) Mod.Istat.Pes_RIC: ricevute di ritorno;

(d) Mod.Istat.Pes1 SLV: foglio di famiglia comprensivo della lettera informativa e della guida alla compilazione in lingua slovena;

(e) Mod.Istat.PES_MAN: manuale di rilevazione;

(f) Mod.Istat.Pes_RIL: tesserino per il Rilevatore.

Il Responsabile dell'UCC, inoltre, ha confezionato i pacchi seguendo la numerazione della sezione di censimento, avendo cura che ogni eventuale questionario aggiuntivo seguisse il principale. Per evitare danneggiamenti o perdite di materiale, anche in relazione alle successive verifiche per la liquidazione del saldo del contributo⁴², è stato espressamente chiesto ai comuni di non inserire nelle scatole questionari arrotolati, piegati o inseriti l'uno dentro l'altro, di non inserirli in buste, veline trasparenti, cartelline o simili e di non apporre in alcun caso punti metallici, fermagli e simili. Una volta inserito il materiale nelle scatole, il Responsabile dell'UCC ha identificato ciascuna di

⁴² Per maggiori approfondimenti circa la corresponsione dei contributi ai comuni vedi il precedente paragrafo 1.3 "Aspetti finanziari e amministrativi".

esse con una distinta, secondo il contenuto della scatola stessa (questionari Mod.Istat.PES.1 privi dei primi due fogli o i primi due fogli di ciascun questionario). In particolare, ciascuna distinta (Figura 2.6) è stata compilata informaticamente e stampata in duplice copia, una da apporre all'esterno della scatola e una da inserire all'interno della stessa. Le stampe delle distinte non dovevano contenere correzioni manuali alle informazioni in esse contenute.

Il ritiro dei plichi è stato identificato da un Documento di Trasporto (DDT) che il vettore ha consegnato al Responsabile dell'UCC, o ad un suo incaricato, contestualmente al ritiro dei plichi.

Si è raccomandata la massima attenzione nel confezionamento dei plichi e nella comunicazione a Poste Italiane Spa della quantità di plichi da ritirare, poiché il vettore non era autorizzato a ritirarne una quantità diversa rispetto a quella dichiarata.

I questionari e l'ulteriore materiale non utilizzati dall'UCC non sono stati inviati all'Istat, ma sono stati smaltiti a cura dell'UCC.

Il Responsabile dell'UCC è tenuto a conservare i Diari di sezione cartacei compilati dai Rilevatori durante la rilevazione, nel rispetto delle norme poste a tutela del segreto statistico e della riservatezza dei dati personali, fino alla diffusione, da parte dell'Istat, dei dati definitivi relativi all'indagine (31 dicembre 2014). Fino a tale data, infatti, è facoltà dell'Istat richiedere la consultazione di tali diari.

Le prescrizioni in materia di confezionamento dei plichi sono state applicate anche ai questionari in lingua slovena e tedesca.

Figura 2.6 - Distinta di consegna materiale da UCC a Istat

DISTINTA	
Codice Univoco Distinta	<input type="text"/>
Denominazione Comune	<input type="text"/>
Codice Provincia	<input type="text"/>
Codice Comune	<input type="text"/>
Tipologia scatola (Distinta 1 ; Distinta 2)	<input type="text"/>
<i>Le quantità indicate devono corrispondere a quelle del materiale effettivamente contenuto nella scatola</i>	
Mod. Istat.PES.1 (indicare solo se trattasi di Distinta 1)	N. <input type="text"/>
Primo foglio dei questionari (indicare solo se trattasi di Distinta 2)	N. <input type="text"/>
Lista (indicare solo se trattasi di Distinta 2)	N. <input type="text"/>

2.6 Monitoraggio della rilevazione sul campo

Il monitoraggio è un'attività che coinvolge tutte le fasi della rilevazione e che deve essere operato tenendo conto delle varie interconnessioni con l'aspetto qualitativo della rilevazione. L'osservazione costante dei risultati parziali dà modo di prevenire gli errori non campionari che possono essere generati, da vari fattori, durante la raccolta dei questionari.

Due peculiarità caratterizzano la rilevazione PES:

- ▶ è un'indagine campionaria areale;
- ▶ è indipendente dal Censimento della popolazione.

Essendo un'indagine campionaria areale, cioè avendo come campione di primo stadio i comuni e di secondo stadio le sezioni di censimento, non prevede un campione di famiglie eleggibili, perché l'universo delle famiglie rimane ignoto fino alla fine della rilevazione. L'indipendenza dell'indagine PES dal Censimento della popolazione ha come conseguenza il fatto di non potere considerare "universo" l'insieme delle famiglie rilevate durante il Censimento. Ne consegue che per monitorare le operazioni di rilevazione sul campo è stato utilizzato il numero di famiglie derivanti dalle Liste Anagrafiche Comunali (Lac) al 31 dicembre 2010. Da ciò deriva che a ogni Responsabile di UCC campione sono state rese note le sezioni di Censimento estratte e il corrispettivo numero di famiglie derivante dalla Lista Anagrafica Comunale (Lac); a sua volta il Responsabile, dopo avere selezionato un congruo numero di rilevatori, ha assegnato loro una sezione o parte di essa oppure più sezioni di censimento (tenendo conto del principio di indipendenza secondo il quale ad ogni Rilevatore dovevano essere assegnate sezioni diverse rispetto a quelle assegnate durante il Censimento). Il Rilevatore si è recato nella sezione a lui assegnata e, percorrendola, ha registrato le informazioni riscontrate. Per non perdere le informazioni relative ai componenti di alcuna famiglia presente sul territorio e interessata dalla rilevazione, l'Istat ha adottato diverse misure rivolte a minimizzare i rifiuti a rispondere. Per prevenire forme di diffidenza nei confronti dei rilevatori che si sarebbero presentati alla famiglia per la consegna del questionario, i comuni potevano affiggere manifesti sulle bacheche comunali o inserire annunci su quotidiani locali relativi alla rilevazione e i rilevatori si sono presentati alle famiglie con un tesserino di riconoscimento ben visibile. Nel caso di rifiuto al primo contatto con la famiglia, i rilevatori dovevano prendere un appuntamento successivo per dare modo alla famiglia di informarsi sulla veridicità della rilevazione e delle generalità del Rilevatore. Diversamente, nel caso di rifiuto da parte della famiglia per "momento poco opportuno", "mancanza di tempo", "non interessato", ecc., i rilevatori dovevano usare abilità persuasive; fissare un appuntamento per ritornare in un momento più opportuno per la famiglia; ritornare dopo qualche giorno per incontrare un componente della famiglia più disponibile. Di fronte ad un'abitazione in cui non fosse presente alcuna persona, i rilevatori dovevano ritornare in giornate ed orari diversi con l'intento di incontrare qualcuno e/o informarsi presso il portiere o i vicini sull'orario di rientro della famiglia. Nel caso in cui il componente della famiglia non ricordasse tutte le informazioni relative agli altri componenti, i rilevatori dovevano fissare un appuntamento per il ritiro del questionario. Sempre per minimizzare gli errori non campionari, il Rilevatore ha operato la revisione, in presenza della famiglia, attraverso controlli di coerenza e completezza; dopodiché ha inserito alcune informazioni relative ai questionari in SGRPES⁴³.

Al fine di poter rilevare tutte le famiglie, il periodo di rilevazione sul campo ha tenuto conto della chiusura delle operazioni censuarie e della chiusura della PES prima del mese di agosto, periodo durante il quale la reperibilità delle famiglie in casa è bassa. Per minimizzare gli errori dovuti all'effetto ricordo da parte dei rispondenti, il periodo di rilevazione della PES non si è allontanato troppo dal 9 ottobre 2011, data di riferimento del Censimento a cui si riferivano anche alcune domande del questionario PES. Non si è potuto evitare il ritardo nell'avvio della rilevazione da parte di alcuni UCC, in quanto occupati in incombenze amministrative (chiusura del Censimento della popolazione, elezioni amministrative, altre indagini Istat, ecc.).

⁴³ Per approfondimenti sull'argomento vedi cap.5.

Il monitoraggio della rilevazione sul campo, attraverso SGRPES, ha coinvolto contemporaneamente a vari livelli e rispetto al territorio di competenza i diversi soggetti coinvolti nella rilevazione: il Responsabile dell'UCC, l'UPC, l'URC, e l'Istat centrale. La prima fase di monitoraggio della rilevazione sul campo, ad opera dell'Istat centrale e degli URC, ha riguardato la verifica della registrazione su SGRPES da parte del Responsabile dell'UCC del suo nominativo e di quello dei rilevatori e delle sezioni a loro assegnate. I responsabili, che a rilevazione iniziata non avevano ancora adempiuto alla registrazione, sono stati sollecitati a farlo. Le procedure di monitoraggio da parte dell'Istat, per quanto concerne l'organizzazione, sono consistite nell'analisi periodica e costante di report riguardanti informazioni presenti in SGRPES e loro eventuali elaborazioni relative alle famiglie coinvolte nell'indagine PES e alle abitazioni non occupate, utili al controllo quantitativo dei questionari. I report riguardavano: distribuzioni di frequenza periodiche e cumulate (assolute e percentuali) delle famiglie da PES (Lista Famiglia + rifiuto + temporaneamente assenti); i report periodici contenevano i risultati registrati per ogni intervallo temporale tra un'analisi e l'altra, mentre i report cumulati si riferivano ai risultati raccolti dal primo all'ultimo giorno di rilevazione effettuato.

Le variabili di incrocio territoriali e temporali per la costruzione dei report sono state rispettivamente: regione, provincia autonoma, provincia e comune; periodo di riferimento dei dati settimanale e frazione di settimana. Le distribuzioni di frequenza cumulate percentuali per comune delle famiglie da PES (Lista Famiglia + rifiuto + temporaneamente assenti) rispetto al complesso comunale delle famiglie da Lac⁴⁴, ha dato modo di monitorare, per ogni comune, la quantità di famiglie che durante la rilevazione sono state rilevate nel tempo (Tavola 2-6).

A livello nazionale, nei primi dieci giorni di rilevazione, cioè dall'11 aprile al 20 aprile 2012, oltre il 50 per cento di comuni non aveva contattato alcuna famiglia e un po' meno della metà ne aveva rilevate non oltre il 25 per cento. Dopo ventinove giorni di rilevazione rimanevano ancora l'11,5 per cento di comuni che non aveva contattato alcuna famiglia, mentre i rimanenti comuni stavano incrementando la percentuale di famiglie rilevate con vario esito.

Dopo avere sollecitato gli UCC che non avevano ancora contattato alcuna famiglia del proprio comune, si è constatato che soltanto una parte residuale di comuni non aveva ancora avuto alcun contatto con le famiglie, gli altri facenti parte di quell'11,5 per cento, avevano contattato le famiglie e ritirato i questionari, senza però registrarli quotidianamente. Quindi il 10 maggio 2012 è stata inviata a tutti gli UCC una mail con la quale si sollecitava l'aggiornamento tempestivo dello stato dei questionari in SGRPES.

Nell'arco temporale dal 10 maggio al 1° giugno 2012 (il 31 maggio era prevista la conclusione della raccolta dei questionari) solo lo 0,4 per cento di comuni non aveva ancora contattato alcuna famiglia, mentre gli altri comuni si spostavano sempre più verso una percentuale consistente di famiglie rilevate. Infatti, il 40,5 per cento dei comuni aveva rilevato oltre il 75 per cento di famiglie e il 20,6 per cento aveva chiuso la raccolta (il numero di famiglie PES era inferiore a quello Lac) oppure la percentuale delle famiglie rilevate aveva raggiunto il 100 per cento delle famiglie da Lac e, quindi, continuava la rilevazione per le famiglie non presenti in Lac. Successivamente sono state concesse proroghe per la raccolta dei questionari fino al 30 giugno. Alcuni comuni hanno utilizzato parte del tempo riservato per il confezionamento dei plichi per ultimare la raccolta dei questionari (dal 30 giugno al 15 luglio). Il 13 luglio, solo cinque comuni (pari al 2 per cento) non avevano ancora terminato la raccolta che si è protratta, esclusivamente per loro, fino al 27 luglio 2012.

⁴⁴ Liste Anagrafiche Comunali (Lac) al 31 dicembre 2010.

Il Responsabile dell'UCC dei comuni che avevano chiuso le fasi di revisione, controllo e registrazione dei questionari, dal 12 giugno 2012, tramite la funzione "Chiusura indagine" di SGRPES, doveva dichiarare la fine della rilevazione. Il sistema, di conseguenza, non avrebbe consentito ulteriori modifiche ai dati immessi, ma solo la visualizzazione. Tale operazione doveva concludersi entro il 30 giugno, ma visto il protrarsi delle operazioni di raccolta si è protratta fino al 27 luglio. Questa fase è stata monitorata attraverso report periodici contenenti la distribuzione dei comuni per regione rispetto alla dichiarazione di conclusione della rilevazione. Tali report venivano inviati via mail agli URC che a loro volta monitoravano e sollecitavano i comuni del territorio di loro competenza a concludere le operazioni.

Ponendo come data di scadenza delle operazioni di raccolta dei questionari il 15 giugno 2012 (scadenza della prima proroga), si sono distribuiti i comuni rispetto ai giorni intercorsi tra tale data e la data in cui il comune ha dichiarato chiusa la rilevazione in SGRPES (Tavola 2-7). A livello nazionale, la maggior parte dei comuni campione (83,7 per cento) ha dichiarato conclusa la rilevazione entro il primo mese; il 50 per cento dei comuni entro ventuno giorni e il 20,2 per cento entro sette giorni dalla data di scadenza. Il Molise è la regione in cui tutti i comuni hanno concluso la rilevazione entro sette giorni. L'83,3 per cento dei comuni della provincia di Bolzano ha concluso entro due settimane e quasi la metà (49,9 per cento) entro la data prevista. Il 77,8 per cento dei comuni del Friuli-Venezia Giulia ha dichiarato conclusa la rilevazione entro ventuno giorni dalla scadenza. A 28 giorni dalla scadenza, il 50 per cento dei comuni liguri, il 73,7 per cento dei comuni veneti e campani e il 70,5 per cento dei comuni siciliani ha registrato la conclusione della rilevazione. La maggior parte dei comuni delle altre regioni ha concluso la rilevazione entro ventotto giorni dalla scadenza, tra questi, però il 30 per cento dei comuni calabresi aveva concluso entro la data prevista.

Tavola 2.6 - Comuni campione per intervalli temporali, classi di famiglie rilevate, ripartizioni territoriali e Italia (valori percentuali cumulati)

Classi di famiglie rilevate, ripartizioni territoriali	Comuni campione (a)				
	da 1/4 a 20/4/2012	da 21/4 a 9/5/2012	da 10/5 a 1/6/2012	da 2/6 a 21/6/2012	da 22/6 a 13/7/2012 (c)
0%	55,9	13,6	1,7	0,0	0,0
da 0,1% a 25%	44,1	44,0	11,9	3,4	0,0
da 25,1% a 50%	0,0	35,6	6,8	6,8	3,4
da 50,1% a 75%	0,0	5,1	25,4	11,9	1,7
da 75,1% a 100%	0,0	1,7	45,7	45,7	13,6
famiglie < Lac = fine raccolta; famiglie > Lac (b)	0,0	0,0	8,5	32,2	81,3
Nord-ovest	100,0	100,0	100,0	100,0	100,0
0%	37,9	6,9	0,0	0,0	0,0
da 0,1% a 25%	58,7	24,1	5,2	0,0	0,0
da 25,1% a 50%	3,4	38,0	5,2	1,7	0,0
da 50,1% a 75%	0,0	24,1	20,7	15,5	0,0
da 75,1% a 100%	0,0	6,9	41,3	36,2	15,5
famiglie < Lac = fine raccolta; famiglie > Lac (b)	0,0	0,0	27,6	46,6	84,5
Nord-est	100,0	100,0	100,0	100,0	100,0

segue

Tavola 2.6 - Comuni campione per intervalli temporali, classi di famiglie rilevate, ripartizioni territoriali e Italia (valori percentuali cumulati)

Classi di famiglie rilevate, ripartizioni territoriali	Comuni campione (a)				
	da 1/4 a 20/4/2012	da 21/4 a 9/5/2012	da 10/5 a 1/6/ 2012	da 2 /6 a 21 /6 /2012	da 22 /6 a 13/7/2012 (c)
0%	47,8	8,7	0,0	0,0	0,0
da 0,1% a 25%	47,8	45,7	2,2	2,2	0,0
da 25,1% a 50%	4,4	34,8	8,7	4,3	0,0
da 50,1% a 75%	0,0	6,5	23,9	8,7	0,0
da 75,1% a 100%	0,0	4,3	52,2	41,3	8,7
famiglie < Lac = fine raccolta; famiglie > Lac (b)	0,0	0,0	13,0	43,5	91,3
Centro	100,0	100,0	100,0	100,0	100,0
0%	68,3	19,0	0,0	0,0	0,0
da 0,1% a 25%	31,7	46,1	3,2	1,6	0,0
da 25,1% a 50%	0,0	25,4	14,3	4,8	0,0
da 50,1% a 75%	0,0	7,9	30,2	19,0	3,2
da 75,1% a 100%	0,0	0,0	30,2	34,9	6,3
famiglie < Lac = fine raccolta; famiglie > Lac (b)	0,0	1,6	22,1	39,7	90,5
Sud	100,0	100,0	100,0	100,0	100,0
0%	50,0	3,8	0,0	0,0	0,0
da 0,1% a 25%	42,3	42,4	7,7	0,0	0,0
da 25,1% a 50%	7,7	34,6	0,0	3,8	0,0
da 50,1% a 75%	0,0	3,8	19,2	7,7	0,0
da 75,1% a 100%	0,0	15,4	30,8	30,8	3,8
famiglie < Lac = fine raccolta; famiglie > Lac (b)	0,0	0,0	42,3	57,7	96,2
Isole	100,0	100,0	100,0	100,0	100,0
0%	52,8	11,5	0,4	0,0	0,0
da 0,1% a 25%	44,8	40,1	6,0	1,6	0,0
da 25,1% a 50%	2,4	33,3	7,9	4,4	0,8
da 50,1% a 75%	0,0	10,3	24,6	13,5	1,2
da 75,1% a 100%	0,0	4,4	40,5	38,5	10,3
famiglie < Lac = fine raccolta; famiglie > Lac (b)	0,0	0,4	20,6	42,0	87,7
Italia	100,0	100,0	100,0	100,0	100,0

Lac al 31 dicembre 2010.

- (a) Comuni campione: rapporto percentuale tra il numero di famiglie rilevate + il numero di famiglie temporaneamente assenti + rifiuti e il totale famiglie da Lac.
- (b) Classe di famiglie che comprende comuni con numero famiglie minore del numero famiglie da Lac che, quindi, hanno finito la raccolta dei questionari e comuni con numero famiglie maggiore del numero famiglie da Lac che, continuano la rilevazione.
- (c) Nell'intervallo temporale dal 22 giugno al 13 luglio 2012 la maggior parte dei comuni ha concluso la raccolta dei questionari, solo alcuni comuni terminano la raccolta tra il 16 luglio e il 27 luglio 2012.

Tavola 2.7 - Comuni campione per intervalli temporali di chiusura della rilevazione, regione, Province di Bolzano e Trento e Italia (valori percentuali)

Regione, Province di Bolzano e Trento e Italia	Classe temporale di chiusura della rilevazione						Totale
	Entro il 15/6	Tra 1 e 7 gg	Tra 8 e 14 gg	Tra 15 e 21 gg	Tra 22 e 28 gg	Oltre 28 gg	
Piemonte	0,0	12,5	25,0	6,3	43,7	12,5	100,0
Valle d'Aosta/ <i>Vallée d'Aoste</i>	0,0	0,0	33,3	0,0	66,7	0,0	100,0
Liguria	12,5	12,5	12,5	0,0	12,5	50,0	100,0
Lombardia	9,4	6,3	12,5	21,9	28,0	21,9	100,0
Provincia di Bolzano/ <i>Bozen</i>	49,9	16,7	16,7	16,7	0,0	0,0	100,0
Provincia di Trento	0,0	16,7	33,4	0,0	50,0	0,0	100,0
Veneto	5,3	5,3	10,5	15,8	36,8	26,3	100,0
Friuli-Venezia Giulia	0,0	11,1	33,3	33,3	22,2	0,0	100,0
Emilia-Romagna	0,0	22,2	16,7	5,6	33,3	22,2	100,0
Toscana	17,6	5,9	11,8	11,8	52,9	0,0	100,0
Umbria	0,0	0,0	14,3	28,6	42,8	14,3	100,0
Marche	10,0	20,0	10,0	10,0	40,0	10,0	100,0
Lazio	0,0	16,7	0,0	16,7	58,3	8,3	100,0
Abruzzo	0,0	0,0	25,0	37,5	12,5	25,0	100,0
Molise	0,0	100,0	0,0	0,0	0,0	0,0	100,0
Campania	0,0	5,3	5,3	26,3	36,8	26,3	100,0
Puglia	6,7	6,7	6,7	26,6	46,6	6,7	100,0
Basilicata	16,7	0,0	0,0	33,3	33,3	16,7	100,0
Calabria	30,0	20,0	0,0	20,0	20,0	10,0	100,0
Sicilia	23,5	5,9	5,9	17,6	17,6	29,5	100,0
Sardegna	11,1	11,1	22,2	11,1	33,4	11,1	100,0
ITALIA	8,7	11,5	12,7	17,1	33,7	16,3	100,0

CAPITOLO 3

Rilevazione sul campo: metodi e strumenti ⁴⁵

Durante la rilevazione sul campo è stata data particolare attenzione all'aspetto della comprensione, da parte di tutti gli attori, delle modalità e delle singole operazioni da effettuare. Attraverso un continuo monitoraggio delle criticità utilizzando molteplici strumenti, l'Istat ha affinato le informazioni e precisato definizioni e metodi della rilevazione. Il principale obiettivo è stato quello di chiarire le differenze tra la PES e il Censimento e di sottolineare le peculiarità e criticità dell'indagine di copertura.

Molti dubbi hanno riguardato le unità di rilevazione, ovvero il campo di osservazione dell'indagine (cfr. par. 3.1). La somministrazione del questionario (cfr. par. 3.2) presenta alcune differenze sostanziali rispetto al Censimento, legate alla differenza nelle unità da rilevare e alla presenza di alcune domande specifiche. Le istruzioni contenute nel manuale (cfr. par. 3.3), nonché alcuni strumenti ad hoc, quali FAQ, assistenza telefonica, bacheche on line (cfr. par. 3.4) hanno contribuito a risolvere, in tempo reale o quasi, i problemi sollevati dagli uffici e dalle diverse realtà coinvolte direttamente nell'indagine (Uffici territoriali dell'Istat, Uffici Comunali, Rilevatori, cittadini).

3.1 Unità di rilevazione

Una delle questioni più complesse che ha riguardato la fase di formazione e di rilevazione sul campo riguarda il campo di osservazione oggetto di indagine della PES.

Infatti le unità di rilevazione non erano le stesse del 15° Censimento generale della popolazione e delle abitazioni. Poiché la gran parte del personale coinvolto nell'indagine aveva partecipato anche al Censimento fino a poco tempo prima, era di primaria importanza che venisse compresa la differenza rispetto alle unità da intervistare. Solo le persone che dimoravano abitualmente nelle sezioni campione erano oggetto della rilevazione PES. Questo ha originato alcuni dubbi quando sono state trovate sul campo persone che, pur vivendo all'interno delle zone selezionate nel campione areale, non avevano lì la dimora abituale.

Nella PES sono state escluse le famiglie e le persone dimoranti in convivenza o in altri tipi di alloggio, così come i senza fissa dimora e i senza tetto. Anche in questo caso alcune incomprensioni sono state dovute a cambi di residenza tra la data del Censimento e quella della PES.

Le abitazioni sono state conteggiate (e non censite con un questionario, come era avvenuto nel Censimento), includendo anche quelle non occupate e quelle occupate esclusivamente da persone che non avevano dimora abituale nell'alloggio, ovvero quelle temporaneamente presenti.

Nel caso di cittadini italiani, sono stati inclusi dalla rilevazione quelli che avevano dimora abituale nell'alloggio, anche se temporaneamente assenti. I cittadini stranieri sono stati oggetto di rilevazione se avevano dimora abituale nell'alloggio ed erano iscritti in anagrafe, o se in possesso di regolare titolo di soggiorno in Italia, anche se temporaneamente assenti.

Sul campo si sono risolti ulteriori dubbi, generalmente legati allo sfasamento temporale tra PES e data di riferimento (9 ottobre 2011). Le famiglie infatti, nel rispondere ai quesiti del questionario, dovevano riferirsi alla data del Censimento, ossia al 9 ottobre 2011.

⁴⁵ I paragrafi sono a cura di:

3.1 Unità di rilevazione: Lorenzo Cassata

3.2 Questionario: Giacomo Maria Ricci

3.3 Manuale: Silvia Farano

3.4 Strumenti ed operazioni di ausilio della rilevazione sul campo: Federico Benassi.

È stato quindi necessario chiarire che facevano parte del campo di osservazione PES anche gli individui nati dopo il 9 ottobre (che però dovevano rispondere solo alle domande da 1.1 a 1.4, rispettivamente relative alla relazione di parentela, sesso, data e luogo di nascita) e le persone trasferite nelle sezioni campione dopo il 9 ottobre (che rispondevano a tutto il questionario). Peraltro l'inserimento dei nati dopo il 9 ottobre 2011 ha richiesto una modifica "in corso" in SGRPES (cfr. cap. 4), visto che originariamente non era stata prevista la possibilità di inserire date posteriori a quella del Censimento. L'ambiguità riguardava la dimora abituale (caratteristica da possedere al momento della rilevazione PES) e le risposte alle domande del questionario (al momento del Censimento). È stato necessario precisare quindi che immigrati e nati dopo il 9 ottobre fino al momento della rilevazione dovevano essere rilevati, mentre al contrario emigrati e deceduti dopo il 9 ottobre non andavano inclusi nella lista del questionario.

Un'eccezione importante alla regola di riferimento alla data del Censimento (9 ottobre 2011) era la relazione di parentela (domanda 1.1) dei nati dopo il 9 ottobre, che andava riferita alla data della PES e non al 9 ottobre (come indicato in grassetto nella guida alla compilazione), essendo ovviamente impossibile in quel caso riferirsi a un giorno precedente alla nascita.

Le persone che, pur convivendo già il 9 ottobre, hanno contratto matrimonio fra la data del Censimento e la rilevazione PES hanno avuto difficoltà di comprensione su come compilare le domande sulla relazione di parentela e sullo stato civile. La relazione di parentela dei due coniugi da inserire era infatti quella al 9 ottobre (ovvero ad es. "convivente in coppia con l'intestatario") e così pure per lo stato civile (ad es. "celibe/nubile").

3.2 Questionario

Per la rilevazione di controllo della copertura del 15° Censimento generale della popolazione e delle abitazioni è stato predisposto il Foglio di famiglia (Mod. Istat.PES.1).

Il questionario è stato progettato sullo schema di quello breve utilizzato in occasione del Censimento della popolazione (Mod. Istat CP.1B_6P), ma in forma più semplificata, tenendo conto del duplice scopo dell'indagine: misurare l'errore di copertura del Censimento e analizzare la qualità delle risposte.

L'esigenza dell'autocompilazione ha poi portato a scelte indirizzate a semplificare il più possibile il questionario, mediante la collocazione di filtri e richiami a chiarimenti e avvertenze. In tal modo è stato predisposto uno strumento il più possibile autoesplicativo e comprensibile in ogni sua parte. Nello specifico, il questionario PES è stato strutturato nelle seguenti parti:

un frontespizio esplicativo, con un prospetto da compilare a cura del Rilevatore;

- ▶ la lettera informativa sull'indagine, a firma del Presidente dell'Istat;
- ▶ la Lista dei componenti della famiglia (LISTA), ovvero di tutte le persone dimoranti abitualmente nell'alloggio, anche se temporaneamente assenti;
- ▶ i Fogli individuali, in cui riportare le notizie su ciascuna persona della famiglia;
- ▶ la Guida alla compilazione;
- ▶ una pagina finale contenente un riquadro da compilare a cura del Comune e del Rilevatore.

L'indagine era rivolta alle famiglie (non alle convivenze), perciò il Foglio di famiglia doveva essere compilato, con riferimento al 9 ottobre 2011, dall'intestatario del Foglio di famiglia (cioè dalla persona alla quale era intestata la scheda di famiglia in anagrafe) o, in caso di impossibilità, da un'altra persona della famiglia o da una persona temporaneamente presente nell'alloggio con riferimento ai soli componenti con dimora abituale. Nella Lista e nei Fogli individuali dovevano cioè essere riportate le informazioni di ogni individuo appartenente alla famiglia con dimora abituale nell'alloggio, mentre gli eventuali ospiti e in generale le persone presenti nell'alloggio per un breve periodo non dovevano essere considerate nel questionario.

La definizione adottata di famiglia è stata la stessa utilizzata per il Censimento. Nel caso di due o più famiglie presenti nell'alloggio (famiglie coabitanti) ogni famiglia doveva compilare un questionario diverso. Nel caso di famiglie numerose (con più di 6 componenti) era necessario consegnare uno o più questionari aggiuntivi, ovvero in numero sufficiente a comprendere tutte le persone della famiglia. L'utilizzo di più questionari e/o la presenza di famiglie coabitanti sono stati segnalati dal Rilevatore nei riquadri A, B e/o C del prospetto nell'ultima pagina del modello.

Nella LISTA, per ciascun componente della famiglia dovevano essere indicati il nome, il cognome, il sesso, la data e il luogo di nascita. Bisognava compilare una riga per ciascun componente, a partire dall'intestatario del Foglio di famiglia e proseguendo secondo l'ordine indicato nel questionario stesso, ordine legato al grado di parentela con l'intestatario. Nella LISTA, per ogni componente è stato indicato il riferimento alle pagine del questionario da completare, cioè al proprio Foglio individuale. Ciascun membro della famiglia doveva compilare il proprio Foglio individuale suddiviso in quattro sezioni relative a: Notizie anagrafiche e stato abitativo al 9 ottobre 2011, Stato civile e Matrimonio, Cittadinanza ed Istruzione. L'ordine dei Fogli individuali doveva essere lo stesso seguito per elencare i componenti della famiglia nella LISTA.

Nell'ambito della prima e della terza sezione erano presenti anche quesiti che riguardavano notizie anagrafiche già richieste nella compilazione della LISTA. La ripetizione di queste informazioni aveva finalità di fornire uno strumento di controllo sulla correttezza delle informazioni e sull'esatto ordine di compilazione dei fogli individuali rispetto alla Lista. Nella prima sezione erano presenti in particolare le domande 1.5 e 1.6, le più importanti per la riuscita della rilevazione PES, attraverso le quali si poteva capire se la persona era stata o meno censita in quell'alloggio in occasione del Censimento e se era stata censita correttamente. In particolare si chiedeva di indicare dove la persona possedeva la dimora abituale alla data del Censimento. Le persone che dichiaravano di avere avuto la dimora abituale in un altro alloggio dello stesso comune o in un altro comune, dovevano specificare l'indirizzo o il comune, questo avrebbe permesso di individuarne la presenza nel caso in cui l'indirizzo della dimora abituale alla data del censimento fosse appartenuta al campione sottoposto all'indagine. Per le persone con dimora abituale nello stesso alloggio, si chiedeva di specificare se e come erano state censite, se in lista A (che nel mod.Istat.CP1 del Censimento corrispondeva alla attuale Lista delle persone che hanno dimora abituale nell'alloggio) o in Lista B (in cui invece andavano riportate alcune informazioni sulle Persone che NON avevano dimora abituale nell'alloggio). Sulla compilazione di queste due domande si è concentrata una parte importante della formazione, nonché specifiche istruzioni all'interno del manuale. Nonostante ciò, per alcuni casi particolari è stato necessario ribadire e specificare le modalità di compilazione anche durante la fase di rilevazione sul campo.

Nelle altre tre sezioni erano presenti complessivamente nove domande, ripetizioni di quesiti presenti nel modello utilizzato per il Censimento. Lo scopo di tali domande era proprio la valutazione della qualità dei dati censuari, mediante il confronto tra le risposte alla PES e al Censimento.

Il questionario era destinato ad essere processato attraverso un *data entry* manuale: questa modalità rende meno stringenti le regole di compilazione presenti nel questionario rispetto a quello distribuito a tutte le famiglie italiane nell'ottobre 2011, nell'occasione del Censimento.

3.3 Manuale

Uno degli strumenti predisposti dall'Istat e utilizzati in occasione della rilevazione PES, oltre al questionario, è il manuale della rilevazione, che è stato inviato agli UCC e distribuito ai Rilevatori impegnati sul campo. Nel manuale viene delineato il processo di rilevazione, a partire dalla definizione del campo di osservazione, passando per la presentazione del questionario, fino alla

descrizione delle varie attività svolte dalla rete di rilevazione, con un approfondimento dedicato al sistema informatico utilizzato per la gestione e il monitoraggio delle operazioni (SGRPES).

Il manuale, che è stato uno strumento fondamentale per la formazione di tutti i soggetti coinvolti nell'operazione post-censuaria, è composto da cinque capitoli: nei primi due vengono illustrati gli obiettivi, le unità di rilevazione, i tempi e le modalità di svolgimento dell'indagine, mentre nei successivi tre capitoli vengono descritte le attività che dovevano svolgere gli operatori della rete durante le diverse fasi della rilevazione e gli strumenti di ausilio all'indagine. In particolare, il terzo capitolo è dedicato alle istruzioni per i rilevatori nella fase di consegna e di ritiro dei questionari.

Ad ogni Rilevatore impegnato sul campo venivano assegnate un'utenza per accedere al sistema di gestione della rilevazione (SGRPES) e un numero di sezioni da lavorare. Per ogni sezione di sua competenza, il Rilevatore disponeva di un itinerario di sezione, sotto forma di elenco di indirizzi, una mappa cartografica della sezione, che doveva aiutarlo nella fase della ricognizione sul campo, e la versione stampata del Diario di sezione, dove erano indicati i codici da assegnare ai questionari consegnati e dove andava riportato l'esito del contatto e i dati della famiglia individuata. Nella fase di consegna dei questionari, il Rilevatore percorreva il territorio della sezione che gli era stata assegnata, recandosi presso ogni accesso esterno di edificio che incontrava e individuando tutte le abitazioni esistenti all'interno. Se l'abitazione era occupata, il Rilevatore doveva accertare la dimora abituale delle persone trovate, consegnare uno o più questionari a seconda del numero di persone o di famiglie presenti, e procedere quindi con l'interno successivo. Al momento della consegna, il Rilevatore doveva assegnare un codice al questionario (tra quelli presenti nella versione stampata del Diario di sezione), trascrivere su di esso i dati dell'intestatario e i riferimenti dell'UCC al quale la famiglia poteva rivolgersi per ricevere assistenza in caso di difficoltà nella compilazione e, infine, annotare sulla stampa del Diario l'esito del contatto con la famiglia, che successivamente doveva essere registrato in SGRPES. L'aggiornamento costante e tempestivo di SGRPES con gli esiti dei contatti con le famiglie consentiva al Responsabile dell'UCC di monitorare l'andamento delle operazioni di consegna e di ritiro dei questionari ed eventualmente di gestire i casi più problematici. Se la famiglia decideva di compilare il questionario con la collaborazione del Rilevatore, quest'ultimo ne effettuava la consegna e il ritiro contestuale. Durante la fase del ritiro, il Rilevatore doveva percorrere nuovamente l'itinerario della sezione a lui assegnata, ritirare i questionari compilati, revisionarli, aggiornare il Diario di sezione e consegnare i questionari all'UCC di appartenenza. La revisione del questionario consisteva nel verificare la correttezza e la completezza delle informazioni contenute nei questionari restituiti e nell'eventuale correzione degli errori o delle incoerenze presenti, avvalendosi dell'aiuto del rispondente o di un componente della famiglia presente al momento del ritiro. Se per qualche motivo non era possibile effettuare la revisione durante la fase del ritiro del questionario, il Rilevatore doveva almeno assicurarsi di avere tutte le informazioni necessarie per poter ricontattare la famiglia. Se il Rilevatore accertava che l'abitazione non era occupata, oppure era occupata da persone non abitualmente dimoranti o temporaneamente assenti, si limitava a inserire questa informazione in SGRPES, senza compilare alcun questionario. Un'altra attività che il Rilevatore doveva curare durante la rilevazione sul campo era quella relativa alla compilazione dei riquadri riassuntivi presenti nell'ultima pagina del questionario, tenendo conto delle indicazioni fornite nel manuale della rilevazione.

L'organizzazione della rilevazione e la definizione delle funzioni e dei compiti assegnati ai vari organi coinvolti nel processo di raccolta dei dati sono trattati nel quarto capitolo del manuale. In questo capitolo vengono descritte le attività che dovevano svolgere gli Uffici territoriali di Censimento nel corso delle diverse fasi della rilevazione, a partire dalla selezione e dalla formazione del personale, fino al coordinamento e al controllo delle operazioni sul campo. Infine,

l'ultimo capitolo del manuale è dedicato alla presentazione di SGRPES che ha costituito lo strumento informatico di supporto alla conduzione della rilevazione. In SGRPES confluivano, infatti, tutte le informazioni relative al processo di rilevazione, grazie all'aggiornamento continuo effettuato dai rilevatori, ed era consultabile dagli operatori della rete che a vario titolo vi potevano accedere. Il Sistema consentiva di gestire e monitorare tutte le attività connesse al processo di rilevazione mediante una serie di funzioni suddivise in macroaree (Operatori, Sezione, Questionari, Rapporti e Utilità) e organizzate con dei menù a tendina, facilmente navigabili. Con SGRPES era possibile controllare lo stato di lavorazione dei questionari, quantificare il numero dei questionari chiusi per tipologia, verificare il lavoro dei singoli operatori della rete e visualizzare i modelli riepilogativi inseriti dall'operatore.

3.4 Strumenti ed operazioni di ausilio alla rilevazione sul campo

L'errore di rilevazione di un'indagine statistica (sia essa campionaria o meno) può essere sensibilmente ridotto facilitando lo scambio di informazioni tra i diversi soggetti componenti la rete di rilevazione circa la corretta interpretazione della procedura di rilevazione, degli aspetti definitivi nonché di quelli tecnico/informatici.

Nel caso della PES, sono stati utilizzati tre strumenti specifici: le *Frequently Asked Questions* (FAQ), le bacheche *on line* e il supporto telefonico. Questi strumenti si differenziano in modo sostanziale l'uno dall'altro, sia in riferimento alle modalità di strutturazione e di diffusione dei contenuti informativi che ai potenziali fruitori.

Con riferimento alle modalità di strutturazione e di diffusione dei contenuti informativi, le FAQ sono un insieme di domande con relative risposte ritenute "frequenti". Hanno dunque un contenuto informativo strutturato a priori e, con specifico riferimento alla PES, sono state organizzate su base tematica, implementate direttamente dai responsabili di ogni area, per essere poi diffuse su Internet tramite il Sistema di Gestione della Rilevazione (SGRPES).

La diffusione dei contenuti informativi attraverso lo strumento FAQ, teoricamente non dinamica, poiché i contenuti sono definiti a priori e rimangono sostanzialmente tali per tutta la durata della rilevazione, ha avuto, soprattutto nel periodo iniziale della rilevazione, natura progressiva, essendo stato arricchito con alcune domande/risposte riferite a quanto accadeva sul campo e dalle segnalazioni mano a mano pervenute dalla rete di rilevazione. Le bacheche sono invece un sistema dinamico a contenuto informativo aperto (o comunque non strutturato a priori). Ciascun utente abilitato ad accedere e a scrivere nelle bacheche può infatti postare liberamente il proprio quesito (o suggerimento) relativamente a tutti gli aspetti della rilevazione o a questa connessi.

Questo strumento, nel caso dell'indagine PES, è stato implementato all'interno della *Intranet* dell'Istituto di statistica. L'accesso era riservato ai responsabili territoriali della rilevazione (dipendenti degli uffici regionali dell'Istat), che ponevano i quesiti, spesso facendosi portavoce delle istanze di uffici comunali e/o singoli rilevatori, e ai referenti del Dipartimento Censimenti per le varie tematiche, che si occupavano di rispondere alle domande nel modo più completo e tempestivo possibile nonché di proporre essi stessi chiarimenti o novità rilevanti ai fini dell'indagine. Il sistema delle bacheche ha inoltre in parte contribuito ad alimentare le FAQ.

Anche il supporto telefonico è un sistema dinamico a contenuto informativo aperto (o comunque non strutturato a priori). Ciascun utente abilitato era infatti libero di chiamare il numero telefonico preposto all'interno delle fasce orarie indicate e poteva sottoporre all'operatore i quesiti che riteneva più utili. Questo strumento di ausilio alla rilevazione ha visto un numero totale di chiamate prese in carico pari a circa 70. Per una visione più di dettaglio relativa a questo strumento si rimanda al precedente capitolo 2.

Con riferimento ai fruitori potenziali di ciascuno strumento di supporto, è necessario fare alcune distinzioni ulteriori.

Essendo le FAQ scaricabili direttamente da SGRPES sotto forma di un file *Microsoft Excel*, i fruitori di tale strumento erano tutti gli utenti abilitati ad accedere al Sistema di Gestione della Rilevazione (cfr. cap. 5), ovvero tutti gli appartenenti alla rete della rilevazione ai vari livelli.

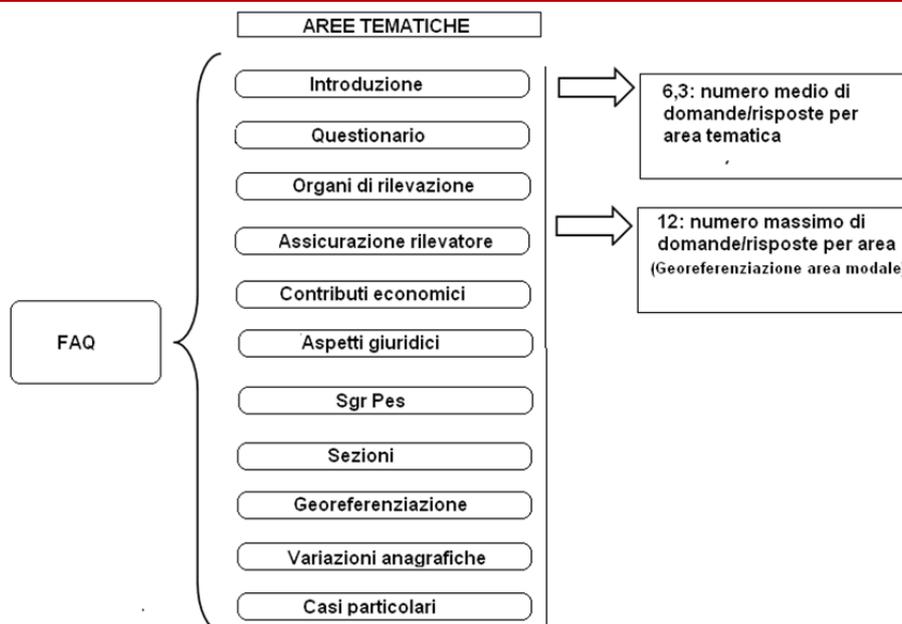
Al contrario le bacheche sono uno strumento di natura telematica ad uso esclusivo Istat, pensato per favorire la condivisione di informazioni tra sedi territoriali e sede centrale (limitatamente ai soggetti coinvolti nelle diverse rilevazioni e abilitati ad accedere a questo strumento)⁴⁶.

Il supporto telefonico era anch'esso ristretto a un insieme di utenti (i componenti di ciascun UCC) anche se, seppur sporadicamente, ne hanno fatto uso anche singoli cittadini.

Le FAQ pensate per la PES sono state complessivamente 69, organizzate in 11 aree tematiche: introduzione, questionario, organi di rilevazione, assicurazione Rilevatore, contributi economici, aspetti giuridici, SGRPES, sezioni, georeferenziazione, variazioni anagrafiche, casi particolari (Fig. 3.1). Ciascuna area tematica era contenuta in un singolo foglio elettronico di un unico file Excel scaricabile in locale da ogni utente abilitato ad accedere ad SGRPES. Il numero medio di domande/risposte per area tematica è risultato pari a 6,3. L'area tematica contenente il maggior numero di domande/risposte è l'area "georeferenziazione", con 12 domande/risposte.

Le bacheche hanno registrato l'inserimento di 63 argomenti tematici per un numero complessivo di risposte pari a 102. Notevolmente maggiore è stato ovviamente il numero complessivo di visite che, ottenuto come sommatoria del numero di visite per ciascun argomento⁴⁷, risulta pari a 4.675.

Figura 3.1 - Aree tematiche delle FAQ, numero medio di domande/risposte per area tematica e area tematica modale

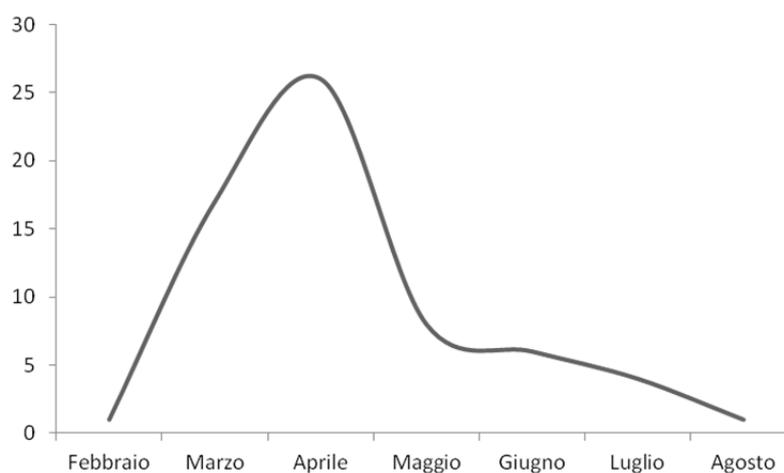


⁴⁶ All'interno del sistema di bacheche dell'Istituto Nazionale di Statistica sono contenute una molteplicità di indagini. L'indagine PES, nello specifico, è collocata all'interno di Popolazione e Abitazioni, a sua volta appartenente alla categoria Censimenti (Censimenti -> Popolazione e Abitazioni -> Indagine Pes).

⁴⁷ Una stessa persona può visitare più volte lo stesso argomento tematico.

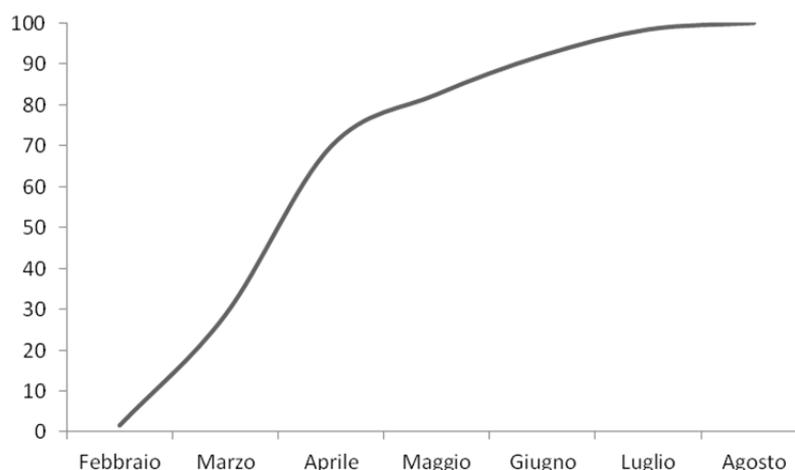
Ciascun argomento tematico è stato visitato in media 74 volte, mentre l'argomento che ha registrato il maggior numero di visite in assoluto è "relazioni di parentela", con 236 visite. Si riporta in Figura 3.1 la distribuzione degli argomenti per data di inserimento in valore assoluto e percentuale cumulata in Figura 3.2. Tenendo conto che il primo argomento, "info per organizzazione e scelta dei rilevatori" è stato postato nel mese di Febbraio 2012 e che l'ultimo argomento, "pagamento PES", è datato 13 Agosto 2012, nel mese di Aprile erano già stati inseriti un numero di argomenti pari a circa il 70 per cento del totale degli argomenti inseriti. Il mese in cui è stato inserito il maggior numero di argomenti risulta aprile, con 26 argomenti contro una media di argomenti per mese di poco inferiore a 9.

Figura 3.2 - Distribuzione degli argomenti inseriti nelle bacheche per mese di inserimento, frequenze assolute



Fonte: nostre elaborazioni su dati Istat/Sistema bacheche

Figura 3.3 - Distribuzione degli argomenti inseriti nelle bacheche per mese di inserimento, frequenze cumulate percentuali e



Fonte: nostre elaborazioni su dati Istat/Sistema bacheche

CAPITOLO 4

Applicazioni geografiche per la rilevazione⁴⁸

4.1 Ripartizione della popolazione censita al 2001 nelle sezioni del 2011

La valutazione del grado di copertura della rilevazione censuaria richiede l'estrazione di un campione di sezioni su cui effettuare nuovamente la rilevazione per valutarne la copertura rispetto al Censimento. Il campione deve essere opportunamente pesato in base alla popolazione residente, ma per rispettare i tempi previsti non si è potuto attendere la conclusione delle operazioni censuarie, dovendo ricorrere alla ripartizione della popolazione residente del 2001 nelle nuove sezioni. Dal momento che le due basi cartografiche divergono sia per geometria che per codifica, si sono dovute implementare procedure ad hoc per effettuare la ripartizione richiesta. Illustrate sommariamente le variazioni intercorse, si procederà ad illustrare i vari metodi proposti e le diverse problematiche connesse per arrivare ad illustrare nel dettaglio la procedura adottata per la preparazione dei dati ai fini del campionamento.

La necessità del confronto tra i censimenti in ambito PES

Per poter effettuare la rilevazione post censuaria, si è reso necessario estrarre un campione di sezioni di censimento del 2011 su cui effettuare nuovamente la rilevazione e valutare in questo modo il grado di copertura raggiunto con la rilevazione censuaria. È stato attribuito ad ogni sezione un peso pari alla popolazione residente. A causa dei vincoli temporali di progetto, non si è potuto attendere lo svolgimento delle operazioni censuarie, quindi si è scelto di ripartire la popolazione residente del 2001 nelle nuove sezioni, poiché la popolazione residente è un dato areale non è possibile automatizzare l'attribuzione alle nuove sezioni di censimento con una semplice sovrapposizione spaziale. È opportuno soffermarsi sul contenuto informativo delle sezioni per comprendere come sia stato possibile confrontare le due basi territoriali e procedere all'attribuzione della popolazione residente, anche se i procedimenti adottati contengono degli errori di stima. Tuttavia, questi errori non sono dovuti al metodo ma alla risoluzione areale del dato.

Le sezioni di Censimento: esigenze cognitive e comparabilità tra basi territoriali

La preparazione delle basi territoriali è una delle principali fasi pre-censuarie in cui si analizza visivamente il territorio ripartendolo in aree omogenee, sulla base del numero di edifici contenuti nelle sezioni (Istat, 2009b). La costruzione delle basi cartografiche del 2011 è stata effettuata secondo il Progetto Census al fine di dar seguito alle attività preparatorie dei censimenti che prevedono sia l'aggiornamento delle basi territoriali che gli adempimenti in materia ecografica, onomastica stradale e numerazione civica che hanno una rilevante importanza per la completa riuscita della rilevazione, in relazione a quanto previsto dagli articoli 9 e 10 della Legge 1.228/54 (Istat s.d.). La sezione viene costruita per contenere un numero congruo di edifici e di conseguenza un quantità di famiglie tale da costituire un carico di lavoro sostenibile per un unico

⁴⁸ I paragrafi sono a cura di:

- 4.1 Ripartizione della popolazione censita al 2001 nelle sezioni del 2011: Gianluigi Salvucci
- 4.2 Abbinamento delle sezioni 2001-2011 per la preparazione cartografica: Valerio Vitale
- 4.3 Supporti cartografici e basi di dati geografiche: Marina Bertollini
- 4.4 Elaborati cartografici: Marina Bertollini e Antonietta Zullo
- 4.5 Itinerari di sezione a supporto della rilevazione: Antonietta Zullo

Rilevatore. In realtà le sezioni di Censimento hanno un elevatissimo potenziale informativo derivante dalla loro classificazione in diverse tipologie di località (Istat, 2009c). Le tipologie di località adottate dall'Istat suddividono il territorio per macro-categorie in funzione dell'uso prevalente, che si desume dalla distanza tra gli edifici (Prospetto 4.1).

Prospetto 4.1 - Declaratoria tipologie località Istat (Istat 2009b)

Tipo località	Declaratoria
1 - Centro	Gruppo di case contigue o vicine con interposte strade, piazze e simili, privo del luogo di raccolta che caratterizza i centri. Per qualificare la vicinanza viene fissata una idonea distanza: 70 metri per i centri abitati e 30 metri per i nuclei
2 - Nucleo	
3 - Località produttiva	Aree in ambito extra-urbano non comprese nei centri e nei nuclei abitati nelle quali siano presenti almeno 10 unità locali o un numero di addetti superiore a 200 e la cui superficie sia almeno di 5 ettari.
4 - Case sparse	Case disseminate nel territorio comunale a distanza tale da non poter costituire nemmeno un nucleo abitato.

Fonte: Descrizione dei dati geografici e delle variabili censuarie per sezione di Censimento Anni 1991, 2001, 2011 (pubblicato on line)

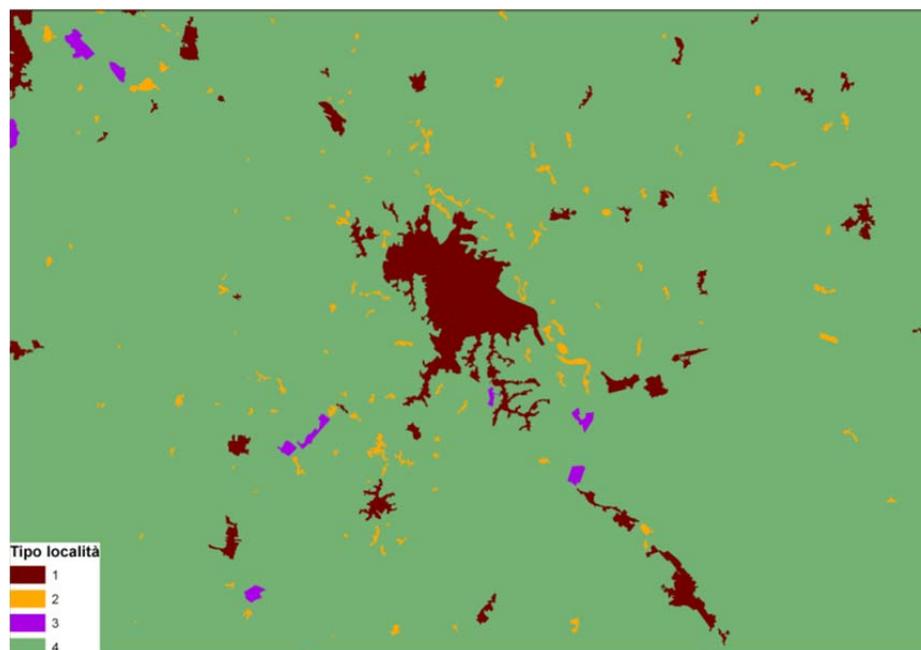
I *macro-usi* individuati dalle sezioni sono caratterizzati da diverse densità della popolazione, che va diminuendo dal centro verso la periferia configurando un assetto urbano gerarchico in cui la popolazione si concentra prevalentemente in località "1 – Centro", mentre le località di tipo "2 – Nucleo" si ritrovano immerse nelle case sparse a corona dei centri maggiori e quindi sono collocate ad un livello inferiore rispetto le prime, come si può vedere, a titolo di esempio, guardando il comune di Siena in Figura 4.1.

Il modello urbano descritto dalla ripartizione precensuaria rispecchia pertanto le logiche classiche del gradiente urbano della densità della popolazione (Clark, 1951) e restituisce una sorta di uso del suolo antropico ad una risoluzione intorno a 1:20 mila, molto più dettagliata dei più utilizzati usi del suolo (uno per tutti Corine Land Cover) che solitamente sono restituiti 1:100 mila. Si tratta sostanzialmente di un uso del suolo antropico con "*macro-usi*" che derivano da esigenze censuarie, in particolare le sezioni di località 1 sono quelle che identificano la "*città consolidata*", per cui a seguito di espansioni urbane si verificheranno dei passaggi di superfici dalla località "4" case sparse alla località "1" perché questa è costituita da una sorta di "*vuoto*" che attende di essere occupato da attività socio-economiche (Celant, 1988), allo stesso modo si deve interpretare il passaggio dalla località "4" alla "2" (nuclei) o alla "3" (aree produttive).

Anche se le sezioni di censimento non distinguono i diversi uso del suolo del tessuto naturale, offrono comunque una elevata possibilità di analisi in merito alla diffusione della concentrazione della popolazione (Abbate, Salvucci 2011), molto utile ai fini del campionamento della PES, ma in generale anche della valutazione della preparazione dell'indagine. È infatti opportuno analizzare i mutamenti intercorsi nella frangia urbana, ovvero il confine tra località di tipo "1" Centro e "4" Case sparse, perché la frangia urbana costituisce la zona di maggior vitalità dell'evoluzione urbana, caratterizzata spesso da fenomeni di *urban sprawl* (Salvati, Salvucci, Abbate Corrado, ZITTI, & Roberta GEMMITI, 2010).

Per tali motivi, ad ogni tornata censuaria si rende necessario intervenire sulle basi territoriali per verificarne l'adeguatezza alla nuova realtà dei luoghi. Pertanto, risulta altamente probabile che le sezioni del tessuto urbano consolidato non subiscano eccessivi mutamenti, al contrario di quanto avviene nelle frange urbane, dove il processo di urbanizzazione è in continua evoluzione.

Figura 4.1 - Distribuzione delle sezioni del comune di Siena per tipo di località (elaborazione su dati Istat, 2011)



Tra i mutamenti territoriali vanno annoverati quelli di natura amministrativa. Gli enti locali possono mutare nel tempo, con la nascita di nuovi comuni o province, oppure per effetto del trasferimento di territorio da un comune all'altro e, più raramente, da una regione all'altra. Il contenuto informativo della sezione attraverso il suo codice tiene conto dell'appartenenza di quella porzione di territorio ad un determinato comune, infatti il codice di tredici caratteri, attribuito ad ogni sezione, riporta partendo da sinistra verso destra, nei primi tre caratteri l'identificativo provinciale, seguito da tre cifre indicanti il comune, e i rimanenti sette caratteri rappresentano la sezione in maniera univoca all'interno del comune stesso. Per alcuni comuni: Roma, Milano, Napoli, gli ultimi sette caratteri indicanti la sezione sono composti dai primi tre (da sinistra) con i quali viene individuata la zona toponomastica e dagli ultimi quattro caratteri composti dalla numerazione delle sezioni. Per zona toponomastica si intende un agglomerato di sezioni contigue accomunate da un'identità storica che ne identifichi un quartiere: nel caso di Roma si tratta di Rioni, Quartieri, Suburbi e Zone (Roma, 1960) e non devono essere confuse con le "Aree di Censimento" (Ace) che hanno natura statistica e servono "come zone di riferimento per la diffusione dei dati risultanti dal Censimento in oggetto, raccolti su un campione di famiglie a cui viene somministrato il questionario in forma completa" (Istat, 2011).

I molteplici mutamenti che possono intercorrere tra le basi censuarie, tra un Censimento e l'altro, pongono dei limiti nel procedere ad una sovrapposizione geografica delle basi territoriali per l'individuazione delle stesse porzioni di territorio, a prescindere dalle variazioni di forma delle sezioni. Quindi, nella costruzione dell'impianto cartografico, si cerca di contenere per quanto possibile le variazioni, al fine di mantenere la comparabilità più elevata tra i censimenti. I limiti sopradetti vengono superati attraverso la comparazione visiva delle due basi cartografiche, ma è comprensibile che questo non possa avvenire per ogni sezione.

Le variazioni delle basi territoriali 2001-2011

Dall'analisi della distribuzione delle sezioni per tipologia di località si evince che il maggior numero di sezioni è quello di località "Centro" (Tavola 4.1). L'elevata concentrazione di persone in queste aree, richiede sezioni di dimensioni più piccole per mantenere costanti i carichi di lavoro dei rilevatori. L'espansione urbana e i miglioramenti di forma apportati hanno indotto ad un aumento di 13.196 sezioni con tipologia Centro abitato tra il Censimento 2001 e il Censimento del 2011. Le trasformazioni sono avvenute prevalentemente per gemmazione producendo nuove sezioni nelle altre tipologie. In definitiva la composizione del territorio nazionale rimane inalterata tra i due censimenti, con un aumento del 67 per cento circa del numero di sezioni di località Centro Abitato "1" (Tavola 4.2).

Tavola 4.1 - Distribuzione delle sezioni dell'intero territorio nazionale per tipologia alle diverse epoche

Tipo località	Numero sezioni		% sezioni	
	2001	2011	2001	2011
1 - Centro	258.656	270.926	67,62%	67,43%
2 - Nucleo	39.449	41.307	10,31%	10,28%
3 - Località produttiva	2.589	3.397	0,68%	0,85%
4 - Case sparse	81.840	86.151	21,39%	21,44%
Italia	382.534	401.781	100,00%	100,00%

Dal confronto delle principali caratteristiche geografiche delle sezioni di Censimento 2001-2011 (Tavola 4.2 e tavola 4.3) si evince come le nuove sezioni di centro abitato abbiano rilevato un miglioramento sostanziale nel disegno, verificatosi a seguito della diminuzione della loro superficie media e dall'aumento del loro perimetro, segni inconfutabili di un aumento del dettaglio con le quali sono state aggiornate.

Tavola 4.2 - Caratteristiche geografiche principali delle sezioni di Censimento del 2001

Tipo località	n sezioni	perimetro medio m	sup media mq
1 - Centro	258.656	1.218,46	65.748,60
2 - Nucleo	39.449	1.145,00	40.619,30
3 - Località produttiva	2.589	2.557,13	293.329,40
4 - Case sparse	81.840	10.704,61	3.454.323,81
Italia	382.534	3.249,43	789.655,17

I problemi sulla rappresentatività del campione saranno rilevati maggiormente nelle aree in evoluzione, pertanto è stata valutata la variazione per macro uso delle superfici attraverso una matrice di transizione. Nella tavola 4.5, sono evidenziati i passaggi di località per Km² avvenuti nel 2011. I passaggi di località da un lato sono un fenomeno automatico, dovuto all'evoluzione territoriale, ma dall'altro è il risultato di un miglioramento della risoluzione delle sezioni da cui deriva la mancata sovrapposizione di alcune aree, soprattutto legate al miglioramento della linea di costa. Le trasformazioni intercorse diventano rilevanti ai fini della valutazione del campionamento per la PES, ad esempio la trasformazione di una località di case sparse in una di centro implica un

Tavola 4.3 - Caratteristiche geografiche principali delle sezioni di Censimento del 2011

Tipo località	n sezioni	perimetro medio m	sup media mq
1 - Centro	270.926	1.234,53	64.899,97
2 - Nucleo	41.307	1.201,58	42.890,81
3 - Località produttiva	3.397	2.499,81	277.988,88
4 - Case sparse	86.151	10.479,22	3.270.682,89
Italia	401.781	3.224,11	751.831,75

aumento consistente della popolazione residente in quell'area. Al contrario se il passaggio è inverso ci si trova di fronte un miglioramento dei perimetri delle località e quindi all'assegnazione alle case sparse di un'area vuota presumibilmente vuota in partenza. Una parte delle superfici non è sovrapponibile, soprattutto lungo le coste, ciò è dovuto a miglioramenti della risoluzione grafica. Nonostante l'evoluzione urbana, la situazione complessiva nel corso dell'ultimo decennio è abbastanza stabile, considerando che la somma delle percentuali lungo la direttrice principali è pari al 99.116 per cento, a rappresentare una sostanziale stabilità nel corso dell'ultimo decennio (Tavola 4.4).

Tavola 4.4 - Matrice di transizione della tipologia di località dal 2001 al 2011 per Km² (valori assoluti)

Tipo località	2011					Totale
	non sovrapponibile	1	2	3	4	
non sovrapponibile	0,00	5,89	0,59	0,28	13,37	20,14
2001 1 - Centro	5,10	16.339,61	18,44	13,24	629,92	17.006,30
2 - Nucleo	0,27	58,73	1.441,30	5,26	96,88	1.602,44
3 - Località produttiva	0,20	73,03	19,76	636,25	29,27	758,51
4 - Case sparse	13,01	1.106,01	291,26	288,04	280.846,82	282.545,13
Totale	18,57	17.583,27	1.771,35	943,07	281.616,26	301.932,52

Tavola 4.5 - Matrice di transizione della tipologia di località dal 2001 al 2011 (valori percentuali)

Tipo località	2011					Totale
	non sovrapponibile	1	2	3	4	
non sovrapponibile	0,000%	0,002%	0,000%	0,000%	0,004%	0,007%
2001 1 - Centro	0,002%	5,412%	0,006%	0,004%	0,209%	5,632%
2 - Nucleo	0,000%	0,019%	0,477%	0,002%	0,032%	0,531%
3 - Località produttiva	0,000%	0,024%	0,007%	0,211%	0,010%	0,251%
4 - Case sparse	0,004%	0,366%	0,096%	0,095%	93,016%	93,579%
Totale	0,006%	5,824%	0,587%	0,312%	93,271%	100,000%

Il problema informativo ai fini del campionamento

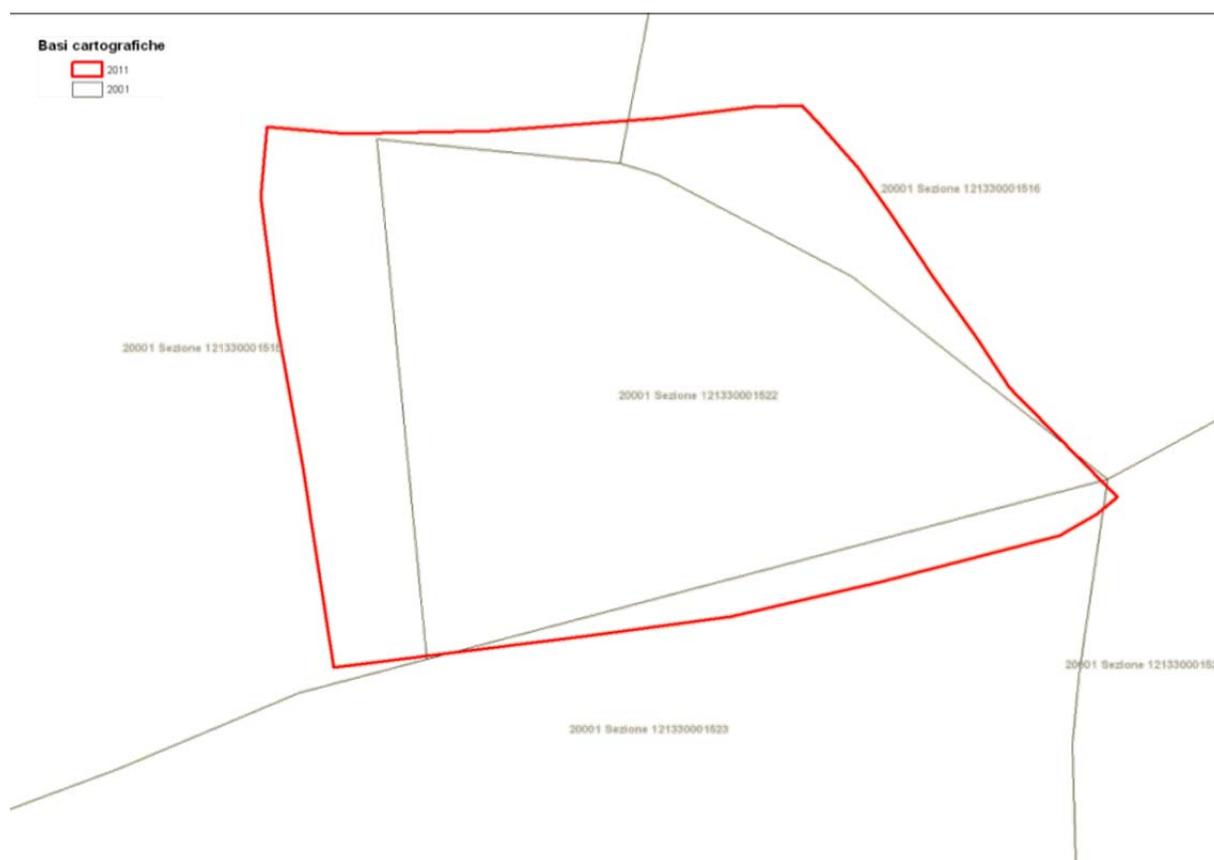
Nonostante la stabilità nell'evoluzione temporale del territorio, il confronto dei dati per sezione di censimento non può avvenire passando esclusivamente attraverso un *record-linkage* tabellare, in quanto una parte delle sezioni, pur non essendo mutate nella forma, hanno mutato il loro codice in seguito a variazioni amministrative. Ragionando, quindi, in termini geografici, il *record-linkage* si trasforma in sovrapposizione delle sezioni, infatti, se si conoscesse l'esatta ubicazione della popolazione all'interno della sezione, si potrebbe sovrapporla al nuovo tracciato delle basi territoriali. Purtroppo, avendo a disposizione un dato areale, la certezza dei risultati si ha solo nei casi di invarianza delle sezioni, oppure in quelli di fusione di intere sezioni del 2001, ma questi casi sono pochissimi rispetto a quelli di scissione. Nello specifico, le sezioni che non hanno subito variazioni di superficie, né cambiamento di codice sono appena 1.106 su 401.781 sezioni complessive; alle 1.106 sezioni se ne possono aggiungere 6 che, pur avendo cambiato il codice, identificano la stessa area. Solo 208 casi contemplano nuove sezioni generate da fusione di vecchie, ma in realtà in appena 5 casi la superficie della nuova sezione corrisponde al 90 per cento della somma di quelle originarie. È evidente che si è dovuto procedere ad un'imputazione della popolazione seguendo altre vie.

Sono state vagliate una serie d'ipotesi per rendere il più verosimile possibile la stima effettuata, in considerazione delle procedure di aggiornamento delle basi territoriali e della necessità di mantenere costanti i totali per comune ai fini del campionamento.

- a) Partendo dalla considerazione che la maggior parte delle nuove sezioni si ottiene per gemmazione, si potrebbe lasciare inalterata la popolazione della sezione madre per attribuire alla sezione figlia la popolazione media della località di centro abitato del comune. Questo modo di ragionare sarebbe molto aderente alla realtà dei luoghi, perché la nuova sezione potrebbe essere un nuovo edificato, ma bisogna considerare che questa metodologia altera i totali comunali del 2001 in modo imprevedibile, non potendo decurtare dalla sezione madre la parte di popolazione residente nella sezione figlia, creando una duplicazione;
- b) L'introduzione di una sovrapposizione geografica risolve il problema dell'attribuzione della sezione 2001 al 2011 a condizione di stabilire una soglia di verosimiglianza tra le due. Vale a dire utilizzando la superficie di massima sovrapposizione delle aree attribuendo la popolazione intera della sezione 2001 a quella che risulta maggiormente aderente, escludendo i casi di sovrapposizione inferiore all'85 per cento. La popolazione non attribuita del comune sarebbe ripartita per superficie tra le sezioni rimanenti. Anche questa metodologia altera il totale comunale che per esigenze di stratificazione deve rimanere inalterato;
- c) In assenza di archivi amministrativi capaci di stabilire l'esatta ubicazione della popolazione residente all'interno della sezione, si sarebbero potute applicare delle tecniche di downscaling utilizzando, ad esempio, la densità media per uso del suolo. Tuttavia, a questa risoluzione geografica non sono ancora disponibili dati geografici così dettagliati. Ad esempio il progetto Corine Land Cover utilizza una risoluzione con dettaglio in scala 1:100 mila mentre, le sezioni di censimento sono prossime a 1:20 mila, vale a dire 5 volte più dettagliate. Da ciò deriva che non sarebbe corretto ripartire la popolazione media per unità spaziale per uso del suolo su una sezione con un dettaglio più elevato perché, inevitabilmente, si commette un errore di stima. Si commette un errore logicamente minore se si considera l'assenza di concentrazione di popolazione all'interno della sezione, infatti per la costruzione della sezione è indispensabile operare secondo criteri di omogeneità del territorio e di conseguenza il fatto che la popolazione sia uniformemente distribuita all'interno della sezione è determinato dall'esigenza precensuaria;

- d) Se così non fosse, si rileverebbe una carenza di rappresentatività della forma della sezione che include situazioni non omogenee al suo interno. Ad esempio una sezione di case sparse che contenesse un nucleo al suo interno conterrebbe un errore non sanabile. In questo caso l'informazione sarebbe perduta perché non è ricavabile se non attraverso un esame visivo dell'ortofoto, quindi ogni sezione ha una sua densità caratteristica;
- e) La sovrapposizione delle sezioni del 2011 a quelle del 2001 consente di individuare le porzioni di superficie del Censimento precedente che, moltiplicate per la densità media della popolazione della sezione al 2001, offrono una stima dei residenti nel confine di quella al 2011 che lascia inalterato il totale comunale. Le potenzialità del metodo sono ravvisabili nella velocità di applicazione, nel mantenimento dei totali, nel rispetto della corrispondenza geografica delle sezioni e della loro sovrapposizione. È evidente che anche piccole variazioni della forma incidano sul risultato. Il metodo proposto non è scevro di errate interpretazioni.

Figura 4.2 - Confronto tra le basi territoriali del 2001 e quelle del 2011 per la sezione 1522 del comune di Varese



Esemplificazione del calcolo della popolazione stimata su una sezione tipo

A mero scopo esemplificativo si rappresenta l'applicazione del metodo utilizzato sulla sezione 1522 del comune di Varese (cod Istat 12133), che risulta in gran parte proveniente dalla sezione con il medesimo codice del 2001 integrata da porzioni di altre sezioni come si vede in Figura 4.2. Nel 2011 la sezione 1522 è stata ampliata aggregando parti delle sezioni limitrofe, infatti se il 65 per cento della sezione 2011 è costituito dalla sezione 1522, questa porzione corrisponde quasi

per intero alla sezione iniziale. Conseguentemente la popolazione stimata viene ripartita sulla base della corrispondenza dei poligoni che compongono la sezione in percentuale della superficie iniziale (b/a di Tavola 4-6). Si suppone che al raddoppio della superficie iniziale non corrisponda un aumento della popolazione, alle 10 unità iniziali se ne aggiungono solamente 3 sulla base della densità 2001 (d di Tavola 4-6) per le superfici annesse.

Tavola 4.6 - Ripartizioni delle superfici iniziali per la costruzione della sezione 2011 n 1522

	a	b	b/sup. totale 2011	e = b/a	c	d=(c/a)x 100.000	(d x e)/ 100.000
2001	superficie sezione 2001 (mq)	superficie dell'intersezione (mq)	% 2001	% corrispondenza dell'intersezione con la sezione iniziale	residenti al 2001	densità 2001 per ettaro quadro	persone residenti attribuite
1515	64.010,60	782,10	18,67%	1,22%	137	21	2
1516	10.351,97	411,64	9,82%	3,98%	8	8	0
1522	2.741,82	2.739,63	65,38%	99,92%	10	36	10
1524	7.389,05	1,67	0,04%	0,02%	16	22	0
1523	6.837,82	255,15	6,09%	3,73%	39	57	1
totale superficie al 2011: 4.190,18							
totale persone attribuite alla nuova sezione: 13							

4.2 Abbinamento delle sezioni 2001-2011 per la preparazione cartografica a supporto della PES

Dal 2001 al 2011 la composizione del territorio italiano per macro-uso individuato dall'Istat attraverso le tipologie di località è rimasta sostanzialmente invariata. Come evidenziato nel paragrafo precedente il 99,11 per cento della superficie nazionale non ha modificato la tipologia rispetto al 2011. Fenomeni di crescita urbana hanno richiesto variazioni della perimetrazione delle località e conseguente modifica delle sezioni che le compongono. Inoltre, per la costruzione delle basi cartografiche del 2011, sono state messe a disposizione da parte dell'AGEA le ortofoto a colori di qualità superiore rispetto a quelle utilizzate in passato (Di Pede, Lipizzi 2009).

In questo paragrafo vengono studiati i principali casi di variazioni intercorse nel tempo, attraverso l'analisi visiva della variazione territoriale, per meglio comprendere eventuali problematiche connesse alla stima della popolazione necessaria per il campionamento della PES.

Il cambiamento dei confini delle sezioni, implicazioni sulla riattribuzione della popolazione

L'utilizzo di ortofoto a colori di qualità superiore a quelle utilizzate nel 2001 consente di distinguere meglio i segni del territorio quali elementi morfologici e infrastrutture che delimitano idealmente le sezioni.

Le difficoltà di sovrapposizione del diverso reticolo delle basi territoriali devono essere interpretate visivamente per offrire un contributo coerente alle fasi di preparazione dell'indagine, ne è un esempio il nuovo limite della linea di costa che individua superfici già esistenti al 2001 ma non rilevate per carenza di risoluzione delle ortofoto disponibili all'epoca (Cruciani, 2010).

Situazioni simili si riscontrano nel passaggio da località di tipo 1 a località di tipo 4 (0,21 per cento della superficie nazionale), dal momento che è difficile ipotizzare un arretramento del tessuto

urbano consolidato, verificatosi solo per un miglioramento della risoluzione che avvicina di più il limite della località a quello effettivo dell'urbanizzato.

In questi casi la variazione di superficie non dovrebbe comportare variazione di popolazione, ma esse devono essere comunque validate visivamente per verificare che l'edificato contenuto non muta nelle successive versioni delle basi territoriali. Vincoli temporali ed economici hanno impedito il controllo puntuale, quindi si è proceduto accettando la miglior stima possibile.

I casi di espansione urbana

Ai fini del campionamento, i casi di espansione urbana sono i più delicati da osservare e quelli più critici per la realizzazione del piano campionario. Bisogna considerare che le aree di case sparse hanno una bassissima densità di popolazione, di conseguenza un peso campionario irrisorio.

Secondo la metodologia di campionamento approvata, era necessario estrarre sezioni in maniera indipendente dalla rilevazione censuaria, pertanto, è stata assegnata la popolazione residente al 2001 nelle nuove partizioni di censimento del 2011. Alle nuove sezioni di località "1" Centro, derivanti da un "vuoto", per garantire l'indipendenza tra la rilevazione censuaria e la rilevazione di controllo e copertura, non deve essere attribuita una popolazione residente che le faccia pesare, ai fini del campionamento, più di quanto pesavano nel 2001.

Nell'esempio in Figura si presenta un classico caso di compattazione ed espansione dell'urbanizzato, dove la sezione originaria presenta una bassissima densità di popolazione con uso del suolo prevalentemente agricolo. Nel corso del tempo è avvenuta una naturale trasformazione dei terreni agricoli adiacenti all'urbanizzato in tessuto urbano continuo (modello di Sinclair, 1967 in Lloyd 1986). La nuova sezione "figlia" evidenziata in giallo, viene urbanizzata e compattata dagli interstizi dell'edificato precedente.

Dal confronto temporale delle ortofoto si evince chiaramente che la popolazione al 2001 era sostanzialmente identica a quella della sezione "madre", per cui, nonostante il processo di urbanizzazione le viene assegnato un peso infinitesimale.

I casi di gemmazione delle sezioni originarie, utilizzo dei dati della Rilevazione dei Numeri Civici

La situazione più complessa da risolvere nel metodo applicato è sicuramente lo scorporo di territorio da parte di una sezione. È noto infatti che nei casi di *downscaling* si rende necessario un aumento della risoluzione dei dati per potersi avvicinare il più possibile alla realtà dei luoghi descritti. Disponendo di dati puntuali provenienti dalla Rilevazione dei Numeri Civici, è possibile al contrario determinare con esattezza la sovrapposizione dei dati alle diverse sezioni del 2011. In occasione della PES è stato pertanto molto utile poter utilizzare i dati relativi all'ubicazione degli edifici, aumentando quindi la risoluzione del dato del precedente Censimento riferito all'area della sezione. In questo modo si è potuto verificare il risultato della ripartizione della popolazione effettuata con il metodo della densità abitativa, come descritto precedentemente.

La figura mostra come il comune di Firenze ha completamente stravolto il disegno iniziale delle basi territoriali, suddividendo le sezioni con nuove partizioni più piccole. Il caso in esame è abbastanza critico da stimare, perché è evidente la concentrazione a sud-est dell'edificato urbano che sarebbe ripartita omogeneamente, falsando la restituzione della realtà con conseguenze negative sul piano di campionamento.

Fusione di sezioni originarie

Quando la fusione aggrega per intero sezioni del 2001 la popolazione complessiva non risulta alterata. Si verificano pochi casi di questo tipo e il metodo adottato non altera i totali poiché si

considerano per intero le sezioni del 2001. Nell'esempio l'area in esame è troppo frammentata, per cui si è reso necessario procedere alla fusione delle sezioni.

Figura 4.3 - Esempio di riduzione di località di centro abitato a seguito di revisione delle basi territoriali del 2011, evidenziato con retinato blu.



Figura 4.4 - Processo di gemmazione di una sezione di località di centro abitato da una di case sparse.

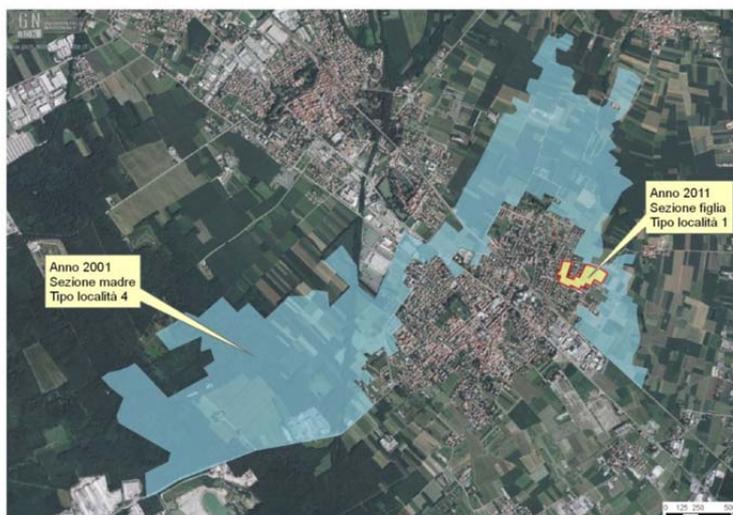


Figura 4.5 - Confronto temporale dell'urbanizzato nella nuova sezione di centro abitato.

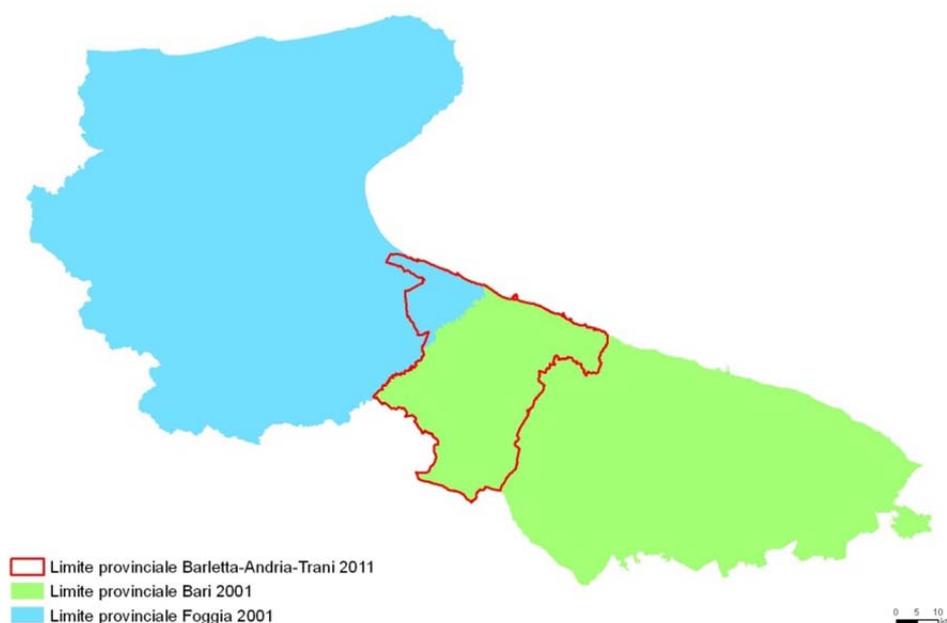


Figura 4.6 - Scomposizione di una sezione del 2001 in una moltitudine di sezioni al 2011 (comune di Firenze)**Figura 4.7 - Esempio di fusione di sezioni 2001 (blu) nell'unica sezione 2011 (rosso)**

Variazioni amministrative, riattribuzione spaziale della popolazione ai nuovi confini amministrativi

È noto che nel codice identificativo della sezione è presente il riferimento alla provincia e al comune (PROCOM). Ne deriva che ad ogni mutamento territoriale o ad ogni costituzione di nuovi enti locali (province e comuni) il nuovo codice identificativo non corrisponderà al precedente⁴⁹. Considerando, inoltre, che potrebbero verificarsi anche mutamenti geometrici delle sezioni, si comprende che per il confronto territoriale intercensuario a livello di sezione non è sufficiente adottare una tabella di transcodifica dei PROCOM, ma è necessario effettuare l'aggancio attraverso sovrapposizione geografica. Ad esempio la nuova provincia di Barletta-Andria-Trani deriva in parte dalla provincia di Foggia (codice provincia: 71) e di parte della provincia di Bari (codice provincia: 72), come visibile in Figura 4.8.

Figura 4.8 - Esempio di variazione di limiti provinciali



Ad una risoluzione più elevata, entrando nello specifico, la superficie della sezione n. 720050000964 apparteneva nel 2001 al comune Andria (PROCOM 72005), trasformato nel 2011 in 110001. Inoltre, le nuove basi territoriali non consentono di attribuire la superficie della sezione 964 alla nuova omonima, la sezione in esame si sovrappone a 6 sezioni del 2011, con diverse superfici, come visibile in Tavola 4.6. Esaminando l'urbanizzato del 2001 nel dettaglio questo ricade principalmente nell'intersezione 2001-2011 A e B.

In questa situazione, l'attribuzione della popolazione in misura proporzionale alle superfici è condivisibile.

⁴⁹ Per assurdo lo spostamento di una intera provincia da una regione all'altra non comporterebbe un mutamento di PROCOM, mentre è evidente che il passaggio di un comune da una regione all'altra produce il cambiamento perché inevitabilmente viene assegnato ad una provincia diversa da quella originaria.

Figura 4.9 - Sezioni del 2011 sovrapponibili alla 720050000964 del 2001

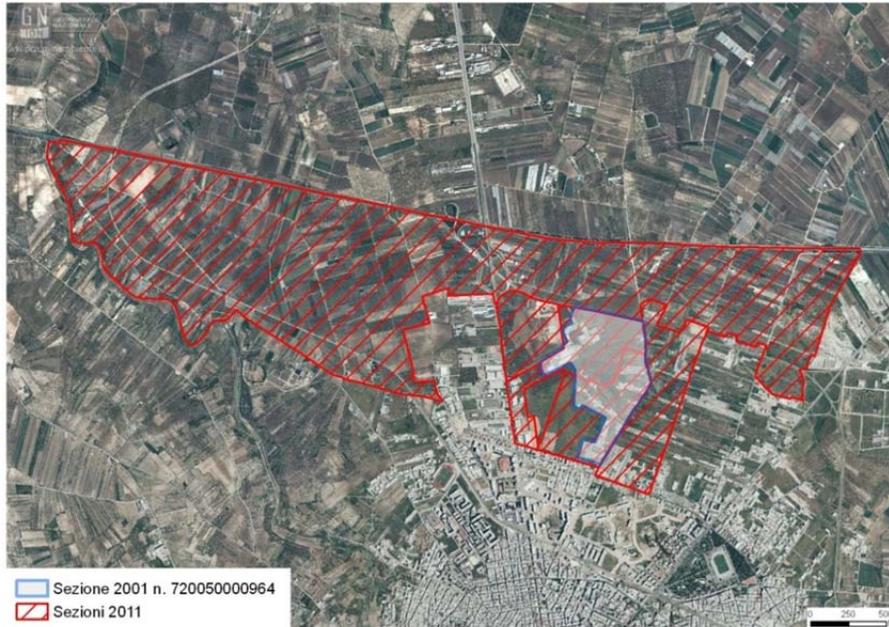


Figura 4.10 - Analisi visiva dell'edificato urbano del 2001 rispetto il tracciato delle sezioni del 2011



Tavola 4.7 - Ripartizione della superficie della sezione 720050000964 del 2001 nelle diverse sezioni del 2011 cui si sovrappone

sezioni 2011 cui si sovrappone	Ha
1100010001076	20,59
1100010000978	0,23
1100010000926	0,03
1100010000964	18,13
1100010000933	0,07
1100010001078	0,15
totale superficie ex sezione 720050000964	39,20

4.3 Strumenti cartografici e basi di dati geografiche a supporto della rilevazione

Sull'utilizzo degli strumenti informativi geografici per l'ausilio dell'estrazione delle unità statistiche da inserire nel campione si è appena scritto sopra. Un'altra funzione che tali strumenti hanno rivestito in questa rilevazione è stata quella di supporto ai comuni per il miglioramento della qualità e del grado di completezza dei dati raccolti.

Dal momento che il concetto di popolazione residente riguarda il territorio, attraverso il legame esistente fra l'individuo e l'indirizzo presso il quale questo risiede, anche i concetti di sovra e sottocopertura dipendono dalla porzione di territorio che viene considerata ai fini della stima (*Istat 2009*).

Pertanto, con l'intento prioritario di ridurre il rischio di individuare in maniera non corretta il territorio di propria competenza e quindi, ad esempio, di travalicare il territorio della sezione di censimento in oggetto, o viceversa di non considerare nella rilevazione intere aree ipotizzando erroneamente una loro appartenenza a sezioni adiacenti, sono stati generati per ciascuna sezione di censimento:

- ▶ una mappa cartografica con l'inquadramento territoriale alla scala comunale dei limiti della sezione (Mappa dei Limiti Comunali e Sezioni);
- ▶ una mappa cartografica dei limiti della sezione (Mappa della Sezione);
- ▶ l'itinerario di sezione (CP.5), cioè la lista degli indirizzi composti da specie, denominazione, numero civico ed eventuale esponente, ricadenti nella sezione.

Basandosi l'indagine di copertura su uno schema campionario di tipo areale, il metodo di raccolta dei dati ha previsto che le sezioni di censimento ricadenti nel campione dovessero essere interamente percorse dal Rilevatore per censire tutte le famiglie e le persone che vivessero in edifici di tipo residenziale.

I responsabili degli Uffici di Censimento si sono potuti avvalere, per coadiuvare il Rilevatore nella localizzazione della sezione di censimento e delle abitazioni eventualmente presenti al suo interno, di strumenti che solo parzialmente erano disponibili nell'Indagine di Copertura del Censimento Generale della Popolazione del 2001.

Nell'indagine di copertura del 2011 è stato infatti possibile verificare in maniera tempestiva sia l'ubicazione che l'esattezza dei confini della sezione di censimento oggetto della ricognizione, nonché la presenza di riferimenti utili alla individuazione delle abitazioni, delle famiglie e degli individui in essa presenti. Ciò è avvenuto grazie alla produzione di strumenti di carattere prettamente geografico pubblicati e resi disponibili via web ai comuni, tramite il nuovo sistema di gestione della rilevazione dell'indagine di copertura del censimento della popolazione, SGRPES.

Si tratta in particolare della mappa geografica di inquadramento della specifica sezione all'interno del territorio comunale, della mappa di dettaglio fornita per ciascuna sezione ed infine dell'itinerario di sezione, ove disponibile. Per ciascuna sezione, tramite SGRPES è stato possibile visualizzare, scaricare e stampare i suddetti elaborati, resi fruibili in formato PDF.

La "Mappa dei Limiti Comunali e delle Sezioni" fornisce il limite amministrativo comunale ed inquadra al suo interno la sezione di Censimento selezionata. L'obiettivo della mappa è quello di orientare il Rilevatore nella localizzazione della sezione di censimento all'interno dei confini amministrativi del comune.

Sulla base di una ortofoto a colori è stato prodotto uno zoom del limite amministrativo, con una scala variabile per ciascun comune in funzione dell'estensione del territorio comunale. La sezione selezionata è stata evidenziata graficamente rispetto alle altre sezioni di Censimento del comune ed è stata individuata mediante un'etichetta recante il suo codice (Figura 4.11).

Figura 4.11 - Visualizzazione di una "Mappa dei Limiti Comunali e delle Sezioni"



Trattandosi di una mappa alla scala comunale, sono stati inseriti gli elementi relativi alle reti e alle sezioni di censimento nei loro tratti essenziali, in particolare:

- ▶ la rete stradale principale (strade provinciali, statali, ecc.) di fonte commerciale TeleAtlas specificata con i relativi toponimi (es.: SS1, SP140, A1);
- ▶ i limiti delle altre sezioni di censimento del Comune, senza ulteriori attributi per non generare confusione nella visualizzazione;
- ▶ le ortofoto.

Le ortofoto impiegate nelle mappe sono quelle disponibili in Istituto, ovvero le medesime impiegate per l'aggiornamento delle basi territoriali dell'Istat nel corso del progetto Census2010⁵⁰. Pertanto si

⁵⁰ Progetto condotto dall'Istat con la collaborazione dei comuni per l'aggiornamento delle basi territoriali, in preparazione del 15° Censimento della Popolazione e delle Abitazioni, tenendo conto delle modifiche intervenute nel territorio comunale durante gli anni successivi al 2001 secondo gli adempimenti previsti dall'art. 39 del D.P.R. 223/89.

Figura 4.13 - Visualizzazione di una “Mappa della Sezione di Censimento” priva di ortofoto



Figura 4.14 - Visualizzazione di una “Mappa della Sezione di Censimento” relativa ad una località di tipo speciale (montagna).



Nei casi in cui la qualità dell'immagine non era tale da consentire la generazione delle mappe (es. mappe di sezioni con estensione territoriale corrispondente all'isolato), sono state utilizzate le ortofoto pubblicate on-line dal Ministero dell'Ambiente sul Portale Cartografico Nazionale. In tali casi è stata evidenziata la fonte, riportando il logo del Portale Cartografico Nazionale nella parte in alto a sinistra della mappa.

Per quanto riguarda la non corrispondenza del perimetro dei limiti della sezione di Censimento riportati nelle mappe fornite dall'Istat con quelli in uso presso il Comune, la causa va rintracciata nel fatto che le sezioni fornite per condurre la PES, così come quelle fornite per il 15° Censimento generale della popolazione e delle abitazioni, non contengono le eventuali modifiche richieste dai comuni dopo la chiusura della finestra di aggiornamento a loro assegnata con il progetto Census2010. Le modifiche concordate con l'Istat sono state infatti acquisite e diffuse alla chiusura del processo di controllo e correzione dei dati rilevati nel corso del 15° Censimento generale della popolazione e delle abitazioni.

Si è presentato anche il caso in cui i limiti della sezione di censimento riportati nelle mappe non coincidono perfettamente con gli elementi fisici

utilizzati per identificare univocamente la sezione stessa (mezzeria di strade, ferrovie, fiumi, etc.), a causa del fatto che le rappresentazioni grafiche delle sezioni di censimento del 2011 derivano, nei Centri Abitati, essenzialmente da quelle del 2001. Queste ultime sono state prodotte tramite digitalizzazione sulla base di foto aeree o di altre carte topografiche, a scale di restituzione inferiori rispetto a quelle impiegate per la produzione delle mappe utilizzate in questa rilevazione. Quindi, la precisione del disegno delle sezioni di Censimento talvolta non è stata sufficiente a determinare univocamente la sezione di Censimento da assegnare ad un edificio attraversato dai limiti di sezione. In questi casi, il Rilevatore ha comunque potuto fare riferimento agli itinerari di sezione per verificare la corretta sezione di appartenenza.

Da ultimo si segnala il caso di mancata corrispondenza tra i toponimi visibili nelle mappe ed i toponimi e la numerazione civica presente sul territorio. Per la produzione delle mappe è stato utilizzato il grafo stradale commerciale TeleAtlas, aggiornato al 2010, poiché disponibile in Istituto in formato vettoriale e quindi adatto alla produzione di immagini cartografiche. Gli archivi di provenienza comunale acquisiti tramite le rilevazioni ANSC o RNC, invece, al momento della PES erano disponibili soltanto in formato tabellare. Per tale motivo i nomi delle strade visibili sulle mappe possono non coincidere con la realtà presente sul territorio. Si tenga presente che il grafo è stato rappresentato nella mappa solo come supporto all'inquadramento delle sezioni e perciò non è stato oggetto di controllo e correzione da parte del Rilevatore.

Per ognuna delle 2.507 sezioni ricadenti nel campione sono state prodotte e verificate 3 mappe, per un totale di 7.521 mappe.

4.4 Elaborati cartografici

Per la fase di produzione delle mappe ci si è avvalsi di procedure automatiche appositamente sviluppate. La fase di controllo di eventuali anomalie presenti in mappa è stata invece condotta in modo manuale, andando ad intervenire, lì dove necessitava una correzione, sui parametri da modificare nelle procedure per la generazione della nuova mappa.

Si è scelto di creare tutte le mappe e di renderle scaricabili in formato file PDF da SGRPES, piuttosto che dare la possibilità ai comuni di richiedere online la generazione contestuale delle mappe. Ciò per rendere disponibili in maniera immediata, a richiesta, gli strumenti geografici di supporto alla rilevazione eliminando ogni attesa dovuta all'elaborazione del processo.

L'ambiente di sviluppo delle procedure di elaborazione delle mappe è costituito dal software ArcGIS, il sistema informativo geografico (GIS) della società Esri. Tale sistema viene impiegato per la creazione di cartografia digitale, la condivisione di informazioni geografiche e la loro gestione ed organizzazione, l'analisi spaziale per l'estrazione di nuove informazioni da una base di dati geografici.

Nel pacchetto ArcGIS sono disponibili diversi ambienti di sviluppo software; quello utilizzato in questo caso è Visual Basic. Le procedure richiamano al loro interno le funzioni primitive di ArcGIS per realizzare le elaborazioni cartografiche necessarie.

Le procedure sono richiamabili tramite un "Pulsante di controllo" presente sulla barra degli strumenti del modulo ArcMap e, a partire da due file di input, esse implementano un ciclo principale che esplora le unità di elaborazione. Il primo file di input è costituito dalle unità di primo stadio e di secondo stadio del campione (comuni e sezioni di censimento) ed il secondo da codici e parametri relativi agli elementi geografici, propri di ciascun comune, utilizzati dalle procedure per la produzione cartografica.

All'interno del ciclo principale, le due procedure realizzate rispettivamente per la creazione della mappa di inquadramento territoriale e per le due mappe di dettaglio della sezione, implementano le fasi di elaborazione descritte di seguito.

La prima operazione consiste nel caricamento dei diversi strati geografici (layer): il file vettoriale⁵² delle sezioni di censimento definite dalle Basi Territoriali Istat; il grafo stradale commerciale; le ortofoto del catalogo Istat e del Portale Cartografico Nazionale (PCN). Nel caso della “Mappa dei Limiti Comunali e delle Sezioni” viene caricato anche lo strato relativo ai limiti amministrativi comunali. Nel caso della “Mappa della Sezione di Censimento” vengono caricati, oltre a quelli già citati, anche gli strati “Ferrovie”, “Laghi e mari” e “Fiumi”. L’accesso ai dati sul PCN avviene tramite tecnica Web Map Service (WMS), ossia tramite un protocollo standard di interscambio, definito a livello internazionale dall’Open Gis Consortium (OGC), per la visualizzazione dinamica di dati geografici residenti su server, senza occupare spazio disco sul PC dell’utente. Questo servizio consente agli utenti finali di effettuare sui dataset geografici selezionati le operazioni più elementari quali visualizzare, navigare, effettuare zoom, sovrapporre altri layer, consentendo anche di esplorare i relativi metadati.

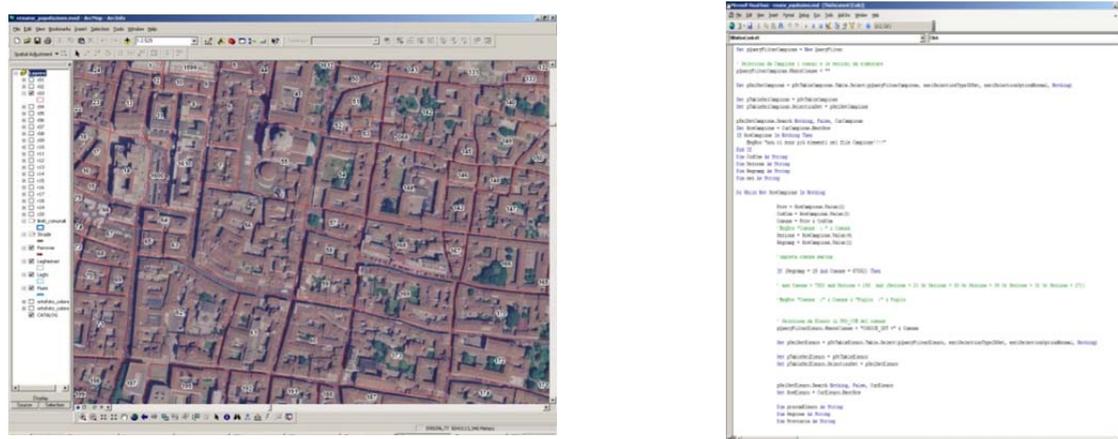
Per rappresentare graficamente i diversi strati geografici sulle mappe, si utilizzano tematismi di colori e forme diverse, riportati in una legenda pubblicata su SGRPES e disponibile per la consultazione ed il download unitamente alle mappe.

Una volta caricati tutti i layer si procede dapprima alla selezione spaziale della/e unità di visualizzazione (limiti comunali o sezione di censimento) e poi allo zoom effettuato alla relativa estensione territoriale.

Per evidenziare la sezione su cui si punta l’attenzione e distinguerla dalle sezioni circostanti presenti nelle mappe, oltre all’adozione di uno specifico colore per la selezione del suo perimetro, le procedure appongono una etichetta standard che riporta il codice della sezione.

Al termine di ogni operazione le procedure effettuano opportune azioni di *refresh* delle mappe. Queste consistono nell’aggiornamento delle strutture di rappresentazione dei dati di ArcMap e della visualizzazione, all’interno dell’area di lavoro (Data View), degli esiti delle operazioni (Figura 4.15).

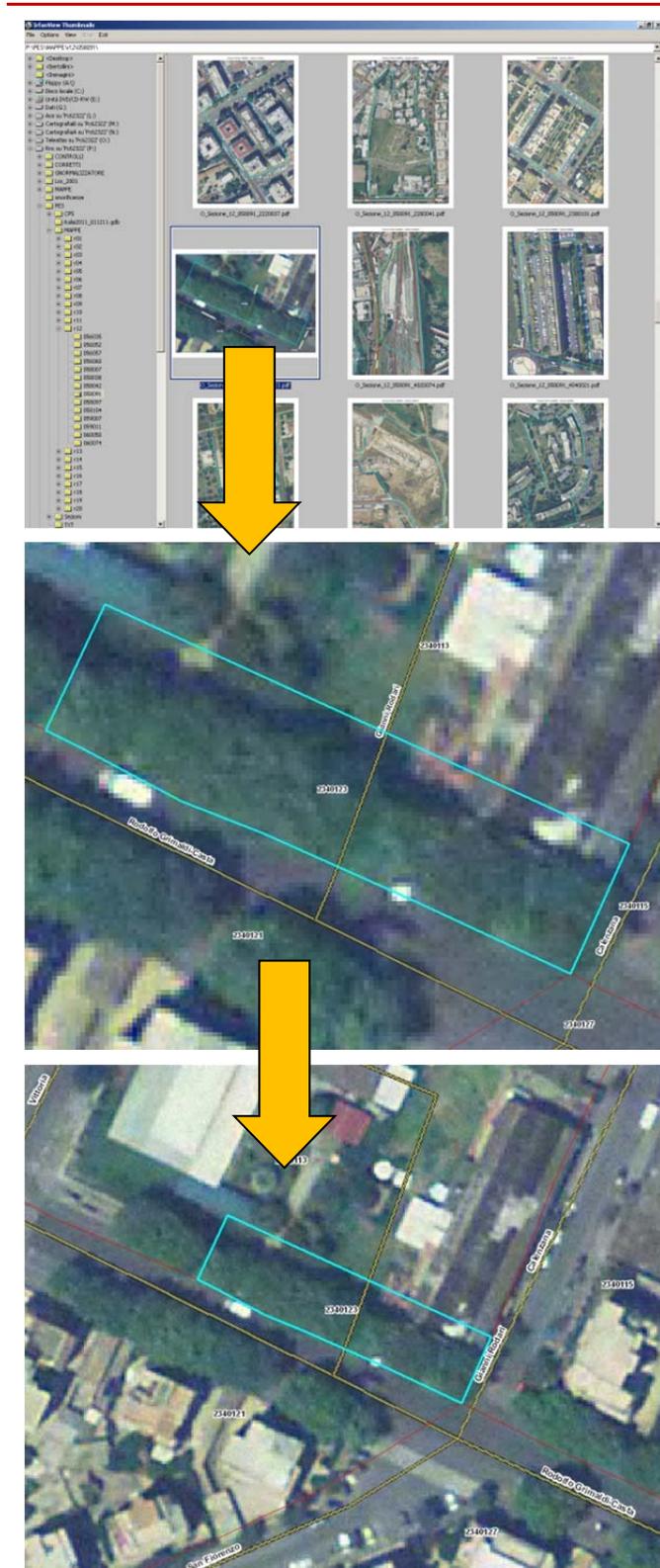
Figura 4.15 - Visualizzazione di una fase di produzione delle mappe



Le mappe elaborate a questo punto devono essere stampate, operazione che viene compiuta passando tramite la procedura all’area di stampa di ArcMap (Layout View) ed effettuando la preparazione del layout tramite la definizione delle proprietà di stampa. Il formato di stampa previ-

⁵² I dati vettoriali sono costituiti dagli elementi geometrici, punti, linee e poligoni, codificati e memorizzati sulla base delle loro coordinate. A ciascun elemento è associato un record del database geografico che contiene tutti gli attributi dell’oggetto rappresentato, compresi quelli geometrici (posizione, lunghezza, area, ecc.).

Figura 4.16 - Esempio di controllo e correzione delle mappe. Passaggio dalla impostazione automatica dello zoom (poco chiara) a quella manuale (più chiara).



sto è A4. In base alle dimensioni dell'entità territoriale da rappresentare, si definisce l'orientamento orizzontale o verticale e viene determinata la risoluzione grafica in funzione della quantità di informazioni geografiche riportate in mappa. Viene quindi impostato il titolo della mappa, composto da una parte costante che specifica il tipo di mappa e da una parte dinamica che specifica la denominazione del comune, il codice Istat del comune ed il codice di sezione cui si riferisce. Infine la mappa viene esportata in formato PDF. Al termine della produzione dei file PDF delle mappe, prima di iniziare il ciclo per un'altra sezione e/o comune, le procedure effettuano l'operazione di deselegazione di tutte le *feature* (*Clear selected features*) per preparare il sistema ad un nuovo ciclo. A valle della creazione delle mappe è seguita una fase di controllo ed eventuale correzione delle stesse. Si è trattato di un controllo a video, effettuato tramite un software open source di visualizzazione delle anteprime di stampa (*IrfanView Thumbnails*). Si sono riscontrate diverse problematiche. Le più frequenti sono state: la mancanza di ortofoto del catalogo Istat, la bassa risoluzione dell'immagine, la creazione dell'etichetta della sezione al di fuori dell'area di stampa. Il caso della mancanza di ortofoto è stato risolto tramite l'acquisizione delle stesse attraverso il servizio WMS del portale PCN, come già descritto in precedenza. Poiché oltre una certa scala il servizio non offre la visualizzazione delle immagini, in alcuni casi è stato necessario "fissare" manualmente la scala dei limiti comunali, così da rendere visibile l'immagine acquisita via web. Anche nel caso della risoluzione non soddisfacente di alcune immagini dovuta ad uno zoom ad una scala troppo grande, ad esempio nel caso di sezioni di limitate dimensioni, si è intervenuti manualmente sull'impostazione dello zoom, così da rendere significativa la produzione della mappa (Figura 4.16).

Per quanto riguarda l'ultimo aspetto, essendo fissata automaticamente la posizione dell'etichetta del codice della sezione all'interno dei limiti comunali, si è proceduto alla modifica manuale di tale posizione, così da renderla visibile anche per quelle sezioni poste al margine dei confini del territorio comunale.

Per ciascuna di tali modifiche sono state ri-eseguite le procedure a seconda dell'intervento descritto o per l'intera regione o per provincia, o per comune o anche per una o più sezioni.

Infine, le 7.521 mappe prodotte e verificate sono state rese disponibili per il popolamento di SGRPES e la loro pubblicazione on line.

4.5 Itinerari di sezione a supporto della rilevazione

L'informazione geografica come ausilio alla rilevazione contribuisce alla conoscenza del territorio, alla verifica della congruenza dei vari archivi e quindi al miglioramento della qualità dei dati raccolti. L'informazione dell'indirizzo assume un ruolo rilevante nelle rilevazioni censuarie e in tutte le analisi (economiche, sociali, demografiche, ambientali, urbanistiche ecc.) che hanno come base anche un riferimento territoriale e, più in generale, integrata con altri dati (catasto, anagrafe, motorizzazione, ufficio del registro ecc.) consente un largo impiego condiviso dell'informazione geografica, come base di interoperabilità tra i diversi archivi amministrativi di fonte centrale e locale.

Il dato relativo all'indirizzo, ancor più associato all'edificio, consente di geolocalizzare la famiglia o l'individuo che in esso dimora, di verificare la sovra e sottocopertura delle anagrafi, di verificare le mancate risposte dei cittadini ed infine essere d'aiuto ai rilevatori sul campo.

Per l'attuazione di questi obiettivi l'Istat ha realizzato, tra il 2010 ed il 2011, la "Rilevazione dei Numeri Civici" (RNC) nei centri abitati dei comuni capoluogo di provincia o con dimensione demografica superiore ai ventimila abitanti, in collaborazione con l'Agenzia del Territorio, l'"Archivio nazionale degli stradari e dei numeri civici" (ANSC) per i restanti comuni.

La validazione e l'integrazione di tali archivi ha consentito di predisporre il database definitivo di indirizzi dal quale ha avuto origine l'itinerario di sezione della PES.

L'itinerario di sezione è la lista ordinata degli indirizzi esistenti all'interno di ciascun arco di strada compreso nella sezione di Censimento.

Gli itinerari di sezione prima della rilevazione sono stati predisposti e pubblicati su SGRPES in formato di file disponibili per il download immediato da parte del comune, per facilitare e velocizzare le operazioni di preparazione degli itinerari da parte dei singoli rilevatori.

Sono state realizzate due applicazioni, una per la produzione degli itinerari di sezione dei comuni RNC ed una per i comuni ANSC.

Le procedure implementate, effettuando operazioni di *join multiple*, selezionano per ogni comune del campione PES: gli indirizzi, le sezioni ed i relativi codici territoriali della regione, della provincia e del comune e popolano un archivio PES che in dettaglio si compone dei campi sotto descritti:

- ▶ campi relativi ai codici identificativi di ciascuna unità amministrativa d'analisi:
- ▶ COD_REG, per il codice di regione;
- ▶ COD_PRO, per il codice della provincia;
- ▶ COD_COM, per il codice del comune di interesse;
- ▶ campi relativi alla descrizione di ciascuna unità territoriale:
- ▶ NOME_COMUNE, denominazione del comune.
- ▶ NOME_LOCALITA', denominazione della località;
- ▶ TIPO_LOC, tipo della località ("Centro abitato", "Nucleo Abitato", "Località produttiva" e "Case sparse");
- ▶ campi relativi all'indirizzo ed alla sezione di Censimento di appartenenza:

- ▶ SEZ2011, il numero della sezione;
- ▶ DUG, Denominazione Urbanistica Generica dell'area di circolazione (es. Via, Piazza, Viale, ecc.);
- ▶ TOPONIMO per la Denominazione dell'Indirizzo (es. Giuseppe Garibaldi);
- ▶ Numero civico;
- ▶ Esponente del numero civico.

Dall'archivio ANSC, presente sul portale dell'Agenzia del Territorio <http://www.agenziaterritorio.it>, si è effettuato il download, per ogni comune del campione, di 2 file di dati in formato predefinito, uno l'archivio pre-censuario degli indirizzi di fonte comunale e l'altro l'archivio degli stradari provenienti dagli indirizzi delle unità immobiliari registrate nel sistema informativo nazionale del catasto, aggiornati alla data del 1 marzo 2012.

Le procedure realizzate consentono l'elaborazione ed il caricamento di CP5 e STRADARIO su un database relazionale e successivamente l'estrapolazione delle informazioni di interesse per la produzione dell'itinerario di sezione e per il popolamento dell'archivio SGRPES in un nuovo archivio.

Il database risultante ANSC_PES contiene gli indirizzi, la sezione ed i codici territoriali della regione, della provincia e del comune ed è composto degli stessi campi dell'archivio PES descritto in dettaglio sopra.

Per la preparazione dei dati dei comuni appartenenti alle province di Trento e Bolzano sono state realizzate procedure distinte per poter gestire il diverso formato dei file dati e consentire la nomenclatura degli indirizzi in doppia lingua: italiano e tedesco.

I dati così preparati nei due archivi descritti sono diventati fonte per il popolamento del database di SGRPES, unitamente alla tabella dei Toponimi presenti in ogni comune.

Per soddisfare la richiesta dei comuni ANSC di poter effettuare il download dei dati relativi agli indirizzi del proprio ambito territoriale, sono stati prodotti e pubblicati online anche i relativi file in formato txt.

Gli Itinerari di Sezione (Figura 4.17) sono stati prodotti come report in formato PDF la cui struttura prevede le seguenti sezioni:

- ▶ intestazione report: specifica il titolo del report.
- ▶ intestazione pagina: visualizza le informazioni comuni a un gruppo di record che riflettono il raggruppamento effettuato sui dati: gli indirizzi di una singola sezione di censimento nell'ambito di un comune e riporta in ogni pagina:
 - la denominazione del Comune;
 - il suo codice Istat composto da codice Provincia e codice Comune;
 - il numero della sezione a cui si riferisce;
 - la denominazione della località in cui ricade la sezione ed il tipo di località (Centro abitato, Nucleo abitato, Località produttiva, Case sparse);
- ▶ corpo del report: visualizza le informazioni di dettaglio del report: l'archivio degli indirizzi della sezione nel formato tabellare con la successione di campi:
 - DUG per la Specie;
 - TOPONIMO per la Denominazione dell'Indirizzo;
 - Numero civico;
 - Esponente del numero civico.

La conversione in formato PDF del report è stata realizzata con Pdf995, un'applicazione freeware che si compone di un driver da installare, che crea una stampante virtuale, e di primitive,

richiamabili dalle procedure automatiche, di configurazione e di conversione che consentono di lavorare sul documento con le impostazioni di pagina, la conversione e la produzione del file PDF.

Di seguito un esempio di un Itinerario di sezione per la sezione 5802 del Comune di Milano.

Il controllo di eventuali anomalie presenti nell'itinerario di sezione è stata effettuato in modo manuale.

L'analisi semantica sui dati (DUG e denominazione della strada, numero civico ed esponente) è stata condotta non al fine di apportare in questo contesto modifiche agli archivi e rielaborare il report, ma al fine di definire una casistica di errori per tipologia di cui tenere conto, successivamente, nella fase di standardizzazione dei dati nell'ambito del progetto di evoluzione degli archivi RNC/ANSC.

In una sezione in cui erano assenti, completamente, oppure in parte, i relativi indirizzi è stata eseguita una verifica puntuale sui dati per accertare la causa di tale anomalia. Le cause delle anomalie sono state poi raggruppate in classi:

- ▶ Per un comune che ha partecipato alla rilevazione RNC, le sezioni di Censimento di tipo località 2 (Nucleo Abitato) , 3 (Località produttiva), 4 (Case sparse) non contengono gli indirizzi perché la RNC ha rilevato solo le sezioni di tipo località 1 (Centro abitato);
- ▶ La sezione appartiene ad un comune ANSC che non ha partecipato al progetto ANSC oppure, pur avendo partecipato, non ha inserito i dati relativi agli stradari e all'indirizzo specifico per la sezione campione;
- ▶ Gli indirizzi relativi ai fabbricati edificati in data successiva alle rilevazioni suddette (RNC; ANSC) non sono disponibili.

Sono stati prodotti circa 2.507 reports in PDF relativi ai 252 comuni del campione PES, pubblicati su SGRPES e resi disponibili per la consultazione ed il download da parte del Comune di pertinenza. Le anomalie segnalate dai comuni durante la rilevazione sono state raggruppate in tre classi.

- ▶ Indirizzi presenti nell'itinerario di sezione di una sezione diversa da quella risultante all'ufficio anagrafe del Comune per via di un disallineamento tra l'ufficio anagrafe e l'ufficio toponomastica. L'itinerario di sezione riporta i dati inseriti e validati dagli uffici di toponomastica dei comuni negli archivi RNC/ANSC; quindi, non è possibile modificare tali archivi e rielaborare l'itinerario di sezione. La soluzione proposta ai comuni in questi casi è di risolvere i disallineamenti con l'ufficio di toponomastica; di proporre l'inserimento dell'indirizzo, come nuovo, nella sezione in cui mancava su SGRPES; di ignorare l'indirizzo dall'itinerario della sezione in cui era presente e di provvedere, quanto prima, alla modifica dei dati sugli archivi RNC/ANSC;
- ▶ Indirizzi o itinerari di sezione mancanti. In questi casi il comune non aveva partecipato in parte o totalmente al progetto ANSC. La soluzione suggerita ai comuni è stata di proporre l'inserimento dell'indirizzo, come nuovo, su SGRPES e di provvedere all'aggiornamento dei dati su ANSC;
- ▶ Disallineamenti su DUG e toponimi dovuti a modifiche avvenute sul territorio nell'arco di tempo tra le rilevazioni RNC/ANSC e PES, oppure ad errori di imputazione dei comuni nelle rilevazioni RNC/ANSC. È stato suggerito ai comuni di inserire le DUG ed i toponimi corretti come nuove entità e di provvedere all'aggiornamento degli archivi RNC/ANSC.

I miglioramenti possibili, al fine di superare le limitazioni riscontrate in questa rilevazione, riguardano principalmente due aspetti: la mancata localizzazione geografica sulla mappa degli indirizzi presenti nell'itinerario di sezione e la non completezza degli archivi fonte dei dati.

Per quanto riguarda il primo punto non è stato possibile collocare sulla mappa l'indirizzo presente nell'itinerario perché non si dispone delle coordinate geografiche di ogni indirizzo. Al momento, per

i comuni RNC, si hanno le coordinate geografiche per una parte limitata di civici e per un'ampia parte di indirizzi le coordinate geometriche, che indicano la posizione del centroide dell'edificio a cui si riferiscono. Per i comuni ANSC non si hanno informazioni circa la loro localizzazione geografica. Di qui la scelta di riportare in mappa i toponimi delle principali strade del grafo stradale. Relativamente al secondo punto, ci si riferisce al fatto che negli archivi RNC non sono stati rilevati gli indirizzi e gli edifici situati nelle località non di centro (tipo località 2,3,4), mentre negli archivi ANSC oltre alle informazioni geografiche relative agli indirizzi e agli edifici mancano alcuni comuni o sezioni di alcuni comuni.

Figura 4.17 - Esempio di Itinerario di sezione (sezione 5802 del Comune di Milano)

Istat		Indagine di Copertura 15° Censimento della Popolazione e delle Abitazioni Nuovi Itinerari di Sezione - Mod. Istat CP.5		
Comune di: MILANO (15146)		Località: Milano (Centro abitato)		Pagina 1 di 2
Sezione di Censimento: 5802				
Specie (DUG)	Denominazione	Civico	Esponente	Note
Via	CARLOTTA MARCHIONNI		2	
Via	CARLOTTA MARCHIONNI		2	SNC
Via	CARLOTTA MARCHIONNI		2	SNC1
Via	CARLOTTA MARCHIONNI		2	SNC2
Via	CARLOTTA MARCHIONNI		2	SNC3
Via	CARLOTTA MARCHIONNI		8	
Via	CARLOTTA MARCHIONNI		8	SNC
Via	CARLOTTA MARCHIONNI		8	SNC1
Via	CARLOTTA MARCHIONNI		8	SNC2
Via	CARLOTTA MARCHIONNI		8	SNC3
Via	CARLOTTA MARCHIONNI		8	SNC4
Via	CARLOTTA MARCHIONNI		8	SNC5
Via	CARLOTTA MARCHIONNI		8	SNC6
Via	CARLOTTA MARCHIONNI		8	SNC7
Via	LUIGI ANGELONI		15	
Via	LUIGI ANGELONI		15	SNC
Via	LUIGI ANGELONI		15	SNC1
Via	LUIGI ANGELONI		15	SNC2
Via	LUIGI ANGELONI		15	SNC3
Via	LUIGI ANGELONI		15	SNC4
Via	LUIGI ANGELONI		15	SNC5
Via	LUIGI ANGELONI		17	
Via	LUIGI ANGELONI		19	
Via	LUIGI ANGELONI		19	SNC
Via	LUIGI ANGELONI		21	
Via	LUIGI ANGELONI		21	SNC
Via	LUIGI ANGELONI		21	SNC1
Via	LUIGI ANGELONI		23	
Via	LUIGI ANGELONI		25	
Via	LUIGI ANGELONI		27	
Via	LUIGI ANGELONI		27	SNC
Viale	MARIO RAPISARDI		4	
Viale	MARIO RAPISARDI		4	SNC
Viale	MARIO RAPISARDI		4	SNC1
Viale	MARIO RAPISARDI		4	SNC2
Viale	MARIO RAPISARDI		4	SNC3
Viale	MARIO RAPISARDI		4	SNC4
Viale	MARIO RAPISARDI		6	
Viale	MARIO RAPISARDI		8	

Tali limiti saranno superati dall'evoluzione e l'integrazione degli archivi RNC/ANSC con gli archivi del 15° Censimento della popolazione e delle abitazioni. Un primo progetto è in corso di definizione e costituirà l'ANNCSU (Archivio Nazionale dei Numeri Civici e delle Strade Urbane) presente nel decreto convertito in legge L. 17 dicembre 2012, n. 221 "Ulteriori misure urgenti per la crescita del Paese".

La base dati sarà ampliata e sarà costituita dai dati risultanti dal *linkage* degli archivi:

- ▶ RNC;
- ▶ ANSC;
- ▶ Archivio degli indirizzi delle famiglie rilevati nel Censimento della popolazione e delle abitazione del 2011;
- ▶ Archivio degli indirizzi degli edifici rilevati nel Censimento della popolazione e delle abitazione del 2011.

Nel processo di *linkage* sono previste operazioni di bonifica, di standardizzazione delle specie e delle denominazioni così da realizzare anche i rispettivi dizionari di riferimento delle stesse, processi di eliminazione dei duplicati e di normalizzazione degli indirizzi e la geocodifica dell'indirizzo alla sezione di censimento.

In tal modo anche tutte le anomalie derivanti da errata composizione dell'indirizzo nelle varie fonti di dati saranno eliminate.

CAPITOLO 5

Sistema informatico a supporto della rilevazione (SGRPES)⁵³

5.1 Introduzione al sistema

Lo svolgimento dell'indagine di copertura del Censimento della popolazione è stato supportato dal sistema informatico denominato Sistema di gestione della Rilevazione PES (SGRPES). Si tratta di un'applicazione disponibile via web, che consente agli operatori di lavorare da qualunque postazione con accesso ad Internet. Il sito è accessibile da diversi browser disponibili sul mercato quali Mozilla, Internet Explorer, Google Chrome e Safari. Ovviamente SGRPES è un'applicazione protetta, che richiede credenziali di accesso, create al momento della registrazione, per poter accedere ed operare nel sistema.

Al fine di garantire la massima sicurezza dei dati, sia in termini di trasmissione che della conservazione degli stessi, è stato adottato il protocollo standard dell'Istituto nazionale di statistica. Inoltre sono stati effettuati accurati test sull'applicazione da un gruppo autonomo rispetto agli sviluppatori per evitare che problemi di software potessero minare la sicurezza dell'applicazione.

L'obiettivo di SGRPES è quello di garantire un continuo monitoraggio della rilevazione ed un'attenta gestione di ogni singolo questionario in ogni fase di lavorazione.

L'indagine di copertura del Censimento della popolazione è stata effettuata a valle dell'operazione censuaria ed ha richiesto ad un campione di comuni di svolgere attività simili a quelle del Censimento. È venuto quindi naturale mettere a disposizione della rete di rilevazione, analogamente con quanto fatto per il Censimento della popolazione, uno strumento che coadiuvasse gli operatori nelle diverse fasi del processo. L'applicazione web è stata sviluppata con l'obiettivo di renderla il più simile possibile a quella messa a disposizione del Censimento, denominata SGR, questo ha portato dei vantaggi sia alla rete di rilevazione sia all'Istituto:

- ▶ la formazione è stata focalizzata solo sulle nuove funzionalità, in quanto la rete conosceva già come muoversi nel sistema;
- ▶ il riuso del software del Censimento della popolazione ha permesso di ridurre drasticamente i costi di sviluppo;
- ▶ le osservazioni e le criticità emerse durante il Censimento della popolazione sono state accolte e risolte.

La diversa mole di dati tra le due indagini e le differenti fasi del processo hanno inoltre permesso di semplificare l'architettura del sistema ed il flusso di lavorazione del questionario, rendendo SGRPES uno strumento molto flessibile. Il sistema di gestione della rilevazione, non prevedendo le interazioni con il vettore postale, l'uso di liste pre-censuarie né la sottocopertura mirata, è risultato molto più semplice e snello rispetto all'analogo strumento predisposto per il Censimento della popolazione. D'altra parte, proprio per aiutare lo svolgersi dell'indagine in modo tradizionale, è stata data particolare attenzione alla fase di lavoro sul campo, attraverso il caricamento all'interno del sistema di mappe e itinerari di sezione, nonché attraverso l'arricchimento del "diario

⁵³ I paragrafi sono a cura di:

- 5.1 Introduzione al sistema: Maura Giacummo, Manuela Marrone e Raffaella Rosati
- 5.2 Profili applicativi e visibilità territoriale: Eleonora Sibilio
- 5.3 Flusso di lavorazione del questionario: Giulia Vaste
- 5.4 Lista e modelli riepilogativi: Marco Amato
- 5.5 Monitoraggio del lavoro sul campo: Lucia Toti
- 5.6 Rapporti riassuntivi: Enrico Orsini
- 5.7 Test: Manlio Salvatore Paterniti
- 5.8 Utilizzo del sistema: Linda Corradi

di sezione”, già previsto per il Censimento della popolazione, qui reso ancora più *completo*, in modo da supportare il Rilevatore in tutte le operazioni di lavorazione del questionario.

Per poter ottenere i dati necessari per effettuare operazioni di linkage tra gli individui registrati nel Censimento e nell'indagine di copertura i modelli riepilogativi sono stati dotati di nuove funzionalità ed hanno permesso la raccolta dei dati di anagrafica di ciascun individuo.

Nei paragrafi successivi è descritta in modo dettagliato la base dati del sistema, ponendo particolare attenzione alla gestione del territorio, e la metodologia utilizzata per i test effettuati sull'applicazione

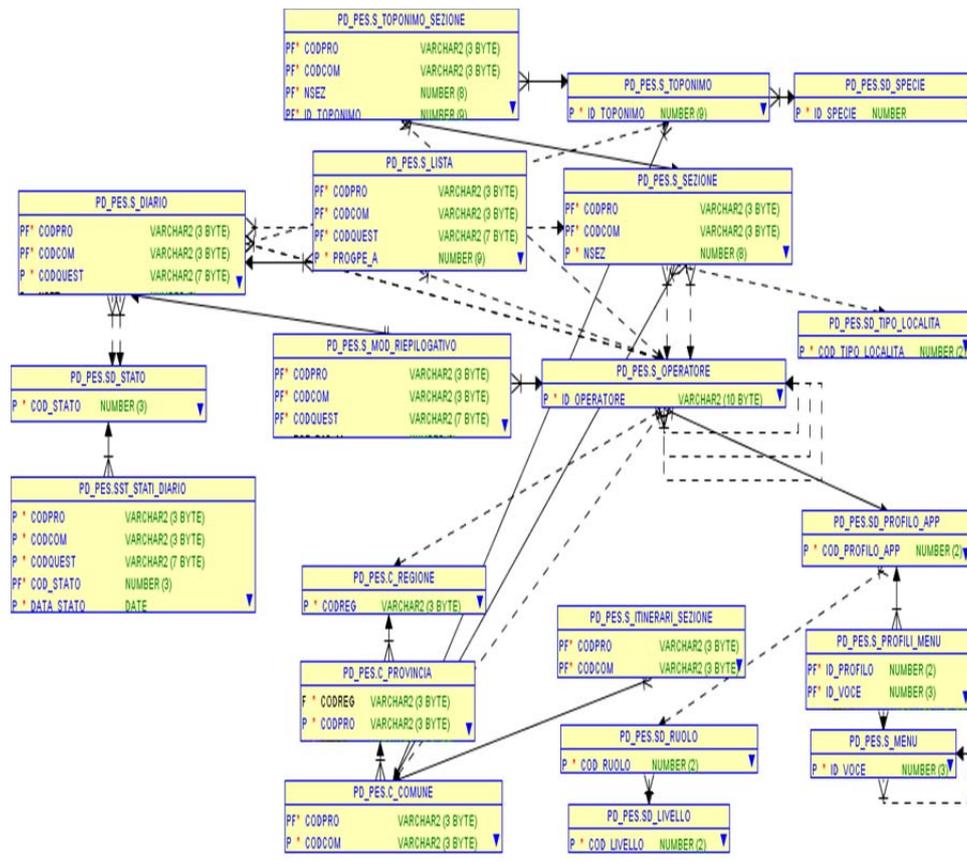
La Base Dati

Il data base di SGRPES è di tipo relazionale, ossia le tabelle sono correlate tra loro attraverso campi chiave. La base dati, schematizzata in figura può essere raggruppata nelle seguenti strutture dati, tutte precaricate prima dell'inizio della rilevazione:

- ▶ **Tabelle territoriali:** contengono le informazioni di decodifica del territorio (regioni, province, comuni e sezioni), nonché tutte le informazioni riguardanti le mappe. In un'apposita tabella si tiene traccia dell'assegnazione delle sezioni ai rilevatori, oltre a memorizzare quali operazioni, e in che data, siano state effettuate dall'operatore per le sezioni di competenza;
- ▶ **Tabelle anagrafica:** contengono i dati anagrafici delle utenze per accedere al sistema e le password criptate degli operatori. Queste tabelle sono precaricate con i dati dei responsabili degli Uffici Comunali di Censimento e con i dati del personale Istat coinvolto nella rilevazione e sono popolate dall'applicazione con la registrazione da parte del responsabile degli operatori della rete;
- ▶ **Tabelle di decodifica:** contengono la descrizione delle variabili di SGR, in questo gruppo sono incluse anche le tabelle lette dall'applicativo che permettono la costruzione dinamica delle pagine del sito. Sono identificabili all'interno della base dati perché posseggono tutte il prefisso SD_ ed il nome del campo per le quali è descritta la decodifica, inoltre hanno una struttura standard composta di due colonne, la prima contenente i valori che la variabile di SGRPES può assumere e la seconda la descrizione nello specifico;
- ▶ **Tabelle dei dati:** contengono i dati raccolti tramite l'applicazione. Le tabelle principali di questo gruppo sono una tabella specifica per questionario, contenente le informazioni sull'intestatario del questionario e sul flusso di lavorazione dello stesso, e una tabella che contiene informazioni sui singoli individui identificati all'interno del questionario con le principali informazioni di anagrafica. Tutte le tabelle dei dati contengono informazioni su chi effettua l'operazione ed in quale data ciascuna operazione viene svolta e sono popolate esclusivamente dagli operatori della rete nel momento in cui individuano una famiglia inserendone i componenti;
- ▶ **Tabelle di gestione:** si riferiscono a tutte le tabelle che consentono all'applicazione di funzionare in modo parametrico e dinamico. Appartengono a questa categoria le tabelle che contengono le informazioni per la creazione dinamica del menù in base all'utente connesso o che gestiscono la visibilità territoriale. Queste tabelle sono tutte precaricate;
- ▶ **Tabelle storiche:** i dati più significativi sono salvati in tabelle storiche ad ogni modifica che viene effettuata. Questo consente di ripristinare in caso di problemi dell'applicazione la situazione immediatamente precedente. Sono storicizzate le tabelle con i microdati e le operazioni di assegnazioni ai rilevatori delle sezioni di competenza. I dati sullo stato inoltre permettono alla rete di visualizzare tutti i cambiamenti di stato subiti dai questionari.

- Viste materializzate: un insieme di oggetti, propri del DBMS Oracle, che sono state utilizzate per la gestione dei rapporti riassuntivi. Tali viste si aggiornano in modo automatico tramite procedure.

Figura 5.1 - Schema E/R: struttura della base dati, con le principali entità



La base dati di SGRPES, oltre che da tabelle, è costituita da *Trigger* e procedure che consentono la gestione completa dell'applicazione. In particolare i *trigger* permettono di automatizzare alcune operazioni senza demandare il controllo solo ed esclusivamente all'applicativo. Tutte le operazioni di storicizzazione sono demandate a questi oggetti. Le procedure Oracle invece sono state sviluppate per il popolamento delle tabelle, e soprattutto per la gestione dei rapporti riassuntivi che consentono il monitoraggio dell'indagine.

La gestione del territorio

All'interno del sistema si deve mettere a disposizione del Rilevatore il materiale cartografico e gli stradari, di cui si è ampiamente discusso al paragrafo 4.4, che vengono pubblicati sul portale SGRPES e possono essere scaricati e stampati dall'apposito menù (Sezione/Itinerario di sezione). Possono accedere alle funzioni tutti i profili previsti dal sistema, in particolare il Rilevatore dispone di una lista degli indirizzi (Itinerario di sezione) e di un insieme di mappe cartografiche relative a tutto il comune di competenza, e non alle sole sezioni che deve gestire. Questa scelta è stata fatta per evitare che l'assegnazione delle sezioni potesse diventare bloccante rispetto al lavoro sul campo.

Le mappe cartografiche e gli itinerari di sezione sono di fatto dei file rispettivamente in formato PDF e JPEG. Per poterli caricare su Oracle senza alterarne la struttura, si è utilizzata la tipologia di dato denominata "BLOB". Oracle, infatti, tratta questi dati come file binari deputati a memorizzare file immagini ma anche file audio e video. Il caricamento di questo tipo di dato deve essere fatto utilizzando appositi package Oracle, sono state quindi sviluppate procedure ad hoc che hanno permesso lo spostamento dei file dei client, nei quali venivano temporaneamente conservati i dati, al file system del server Oracle. I file sono stati caricati prima dell'inizio della rilevazione e tutte le informazioni sono conservate in una sola tabella Oracle chiamata "S_ITINERARI_SEZIONE".

Prima dell'inizio delle operazioni sul campo, inoltre, tutti gli indirizzi (toponimo e DUG), provenienti dagli itinerari di sezione sono stati caricati in apposite tabelle territoriali e sono stati agganciati alle sezioni nelle quali ricadevano; questa operazione è stata effettuata per disporre di un insieme di indirizzi per permettere all'applicazione web di proporre, al momento dell'inserimento delle informazioni anagrafiche della famiglia, solo indirizzi normalizzati e controllati, legati ad una specifica sezione di censimento. In questo modo il Rilevatore ha avuto piena corrispondenza tra gli indirizzi proposti dal sistema e quelli trovati nei PDF degli itinerari di sezione. Questa scelta è stata divergente rispetto a quanto fatto per il Censimento della popolazione, infatti per quest'ultimo sono stati caricati i dati provenienti dal vettore postale, che spesso hanno generato incongruenze, arrecando un notevole disagio per la rete di rilevazione.

5.2 Profili applicativi e visibilità territoriale

Una delle principali funzionalità garantite dall'applicazione è stata quella della gestione della rete di rilevazione. Il responsabile dell'ufficio comunale poteva gestire in totale autonomia l'inserimento degli operatori che avevano il compito di coadiuvarlo nelle operazioni sul campo, la gerarchia territoriale e la divisione del lavoro tra i rilevatori. I profili applicativi abilitati previsti per l'indagine sono stati i seguenti:

- ▶ Utente ISTAT (ISTAT);
- ▶ Responsabile Ufficio Regionale di Censimento (URC);
- ▶ Responsabile Istat Territoriale (RIT);
- ▶ Ufficio Provinciale di Censimento (UPC);
- ▶ Responsabile Ufficio Comunale di Censimento (UCC);
- ▶ Rilevatore (RIL);
- ▶ Responsabile di Prefettura (PRE).

Il profilo applicativo consente la visibilità territoriale degli operatori secondo la loro competenza; nello specifico l'ISTAT aveva visibilità nazionale su tutto il processo, mentre il responsabile di Prefettura poteva accedere al territorio della provincia, salvo eccezioni. Allo stesso modo l'URC e il RIT avevano visibilità regionale, l'UPC provinciale, mentre l'UCC e il RIL avevano visibilità comunale su tutte le funzioni ad eccezione del diario di sezione. Vista l'importanza di tale funzione, infatti, la visibilità della stessa per i rilevatori era subordinata all'assegnazione. In particolare i rilevatori potevano vedere le sole sezioni a loro assegnate.

Il Responsabile UCC poteva creare in modo dinamico ed indipendente la propria rete di rilevazione. Attraverso il menù "Operatori" provvedeva:

- ▶ alla variazione dei suoi dati personali;
- ▶ all'inserimento dei dati di tutti i rilevatori;
- ▶ all'assegnazione delle sezioni ai rilevatori.

Ogni operatore registrato nell'applicazione doveva possedere le credenziali di accesso, ossia un codice che lo identificava in modo univoco. La user-id e la password erano generate in modo automatico dall'applicazione nel momento in cui veniva individuato l'operatore. Nella maschera di inserimento dei rilevatori, riportata in figura, vengono richieste informazioni anagrafiche ed indirizzo di posta per il successivo invio delle credenziali di accesso tramite e-mail.

Prospetto 5.1 - Macro aree e profilo applicativi secondo il ruolo e mansioni corrispondenti

Macro Aree	Funzioni	Profili applicativi abilitati
Operatori	Variazione responsabile UCC	UCC
	Gestione Rilevatore: Inserimento	UCC
	Gestione Rilevatore: Modifica/Cancelazione	UCC
	Assegnazioni: Sezioni ai Rilevatori	UCC
	Composizione della rete di rilevazione	ISTAT, URC, RIT, UPC, UCC
	Cerca operatore	ISTAT, URC, RIT, UPC, UCC
Rapporti Riassuntivi	Stato lavorazione questionari	ISTAT, URC, RIT, UPC, PRE, UCC
	Questionari chiusi per tipologia	ISTAT, URC, RIT, UPC, PRE, UCC
	Lavoro dell'operatore	ISTAT, URC, RIT, UPC, PRE, UCC
	Modelli riepilogativi	ISTAT, URC, RIT, UPC, PRE, UCC
	Indicatore copertura popolazione	ISTAT, URC, RIT
Sezione	Diario: Visualizzazione diario	ISTAT, URC, RIT, UPC,UCC, RIL
	Diario: Aggiornamento diario	UCC, RIL
	Toponimo: Inserimento nuovo toponimo	UCC
	Toponimo: Eliminazione nuovo toponimo	UCC
	Genera codici questionario	ISTAT, URC, RIT, UPC, UCC, RIL
	Itinerari Sezione	URC, RIT, UPC, UCC, RIL
Questionari	Registrazione arrivo questionario cartaceo	URC,RIT,UPC,UCC, RIL
	Eliminazione registrazione arrivo questionario cartaceo	URC,RIT,UPC ,UCC, RIL
	Modelli riepilogativi: Inserimento modelli riepilogativi	ISTAT, URC, RIT, UPC, UCC, RIL
	Modelli riepilogativi: Modifica/Eliminazione	ISTAT, URC, RIT, UPC, UCC, RIL
Utilità	Ricerca questionario	ISTAT, URC, RIT, UPC, UCC, RIL
	Sostituzione questionario per smarrimento	UCC, RIL
	Elimina sostituzione questionario per smarrimento	UCC, RIL
	Inserimento riga	UCC, RIL
	Eliminazione riga	UCC, RIL

L'assegnazione di una o più sezioni del comune di competenza a ciascun Rilevatore avveniva attraverso diverse maschere che facilitavano tale operazione permettendo la ricerca dell'operatore o per Userid assegnata o per cognome (Figura 5.3). Qualora non si fossero possedute queste informazioni era sempre possibile visualizzare l'elenco completo dei Rilevatori. Successivamente era possibile selezionare e/o de-selezionare le sezioni presentate nella pagina e tramite il pulsante "Conferma" effettuare il salvataggio. Durante tutta la rilevazione è stato possibile effettuare le

modifiche di assegnazione, questo per poter gestire eventuali cambi di personale. Per sezioni particolarmente problematiche, in termini di presunta popolosità o per collocazione geografica, era possibile la creazione di sottosezioni ad hoc. Queste modifiche sono state gestite esternamente all'applicazione, tramite procedure Oracle.

Era possibile infine consultare, in ogni momento, tutta le rete di rilevazione, con i dettagli relativi, attraverso la selezione di filtri quali l'identificativo, la regione, la provincia, il comune, il cognome, il ruolo.

Figura 5.2 - Maschera di inserimento dei Rilevatori

ucc roma - Responsabile UCC | [v528 - June 25 2012] | Manuale | Manuale Rilevazione | Informativa | Domande&Risposte | Nota informativa ritiro pacchi | Format per ritiri |

15° CENSIMENTO GENERALE DELLA POPOLAZIONE E DELLE ABITAZIONI 2011

SGR PES

SGR Post Enumeration Survey

Distinta Istat.it

Home Comunicazioni Logout

Operatori | Rapporti | Sezione | Questionari | Utilità

Inserimento Rilevatori

Nome*

Cognome*

Data di nascita (gg/mm/aaaa)*

eMail*

Cellulare

*campo obbligatorio

Conferma Annulla

Figura 5.3 - Maschera di assegnazione delle sezioni agli operatori

ucc roma - Responsabile UCC | [v528 - June 25 2012] | Manuale | Manuale Rilevazione | Informativa | Domande&Risposte | Nota informativa ritiro pacchi | Format per ritiri |

15° CENSIMENTO GENERALE DELLA POPOLAZIONE E DELLE ABITAZIONI 2011

SGR PES

SGR Post Enumeration Survey

Distinta Istat.it

Home Comunicazioni Logout

Operatori | Rapporti | Sezione | Questionari | Utilità

Assegnazione Sezioni al Rilevatore:

0580910281 - agiudici maria luisa

10 record trovati, visualizzati tutti i record. Pag. 1

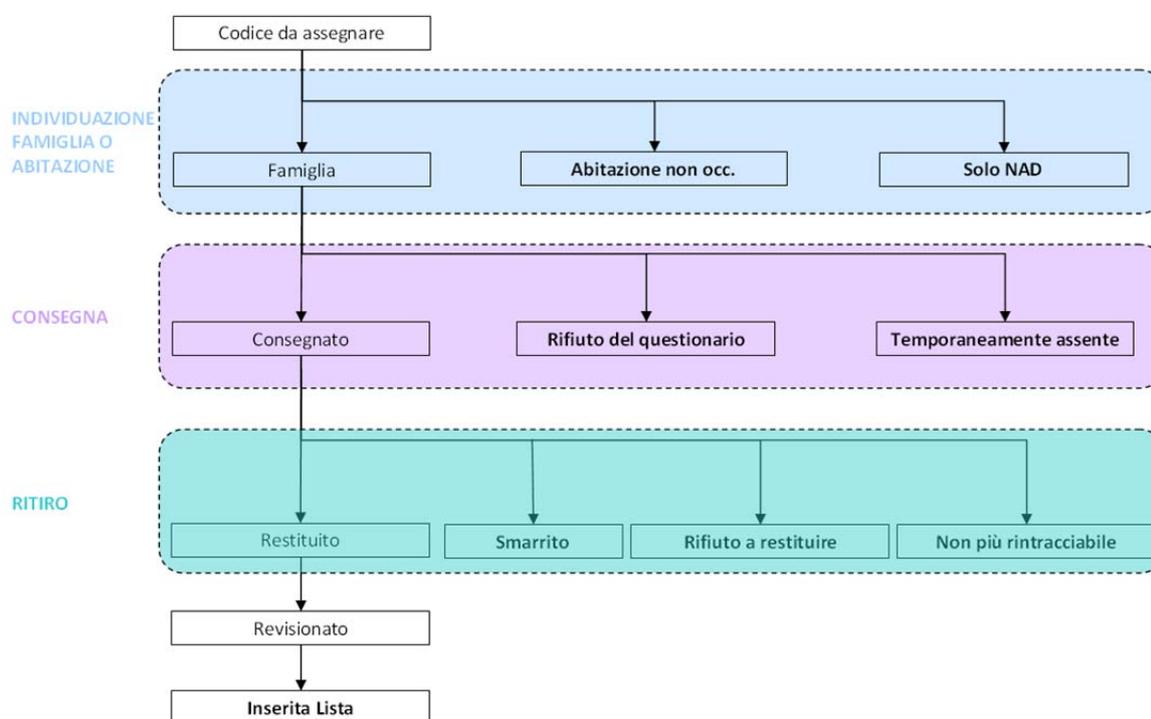
Num. Sezione	Id rilevatore	Cognome rilevatore	Assegnazione
	1010005		<input type="checkbox"/>
	1010008		<input type="checkbox"/>
	1010010		<input type="checkbox"/>
	1010015		<input type="checkbox"/>
	1010044		<input type="checkbox"/>
	1010047		<input type="checkbox"/>
	1010053		<input type="checkbox"/>
	1010056		<input type="checkbox"/>
	1010057		<input type="checkbox"/>
	1010059		<input type="checkbox"/>

Conferma

5.3 Flusso di lavorazione del questionario

La lavorazione dei questionari, analogamente a quanto previsto nel sistema di gestione della rilevazione utilizzato per il Censimento, veniva tracciata attraverso lo stato del questionario: ogni stato rappresentava una fase di lavorazione del questionario o la sua chiusura con relativo esito della rilevazione. Il flusso degli stati del questionario era predefinito e guidato dalle funzioni del sistema. Viene di seguito illustrato il significato di ogni stato del questionario, esplicitando la fase della rilevazione che esso rappresenta; vengono inoltre descritte le funzioni del sistema informatico che realizzavano le transizioni di stato.

Figura 5.4 - I diversi stati del questionario durante le fasi della rilevazione



Ogni questionario cartaceo doveva possedere un identificativo univoco che consentisse l'abbinamento all'interno del sistema informatico e lo collegasse al questionario di famiglia o all'abitazione cui era stato associato. Il sistema offriva quindi la possibilità di generare codici questionario che solo successivamente, nel momento in cui il Rilevatore avesse individuato una famiglia o un'abitazione sul campo, venivano utilizzati. Lo stato "Codice da assegnare" rappresentava, appunto, un codice questionario generato non ancora associato a un cartaceo. Quando il Rilevatore, percorrendo la sezione di Censimento di sua pertinenza, individuava una famiglia, un'abitazione non occupata o un'abitazione occupata solo da persone non abitualmente dimoranti, associava il codice questionario alla famiglia o all'abitazione individuata e poneva il questionario rispettivamente negli stati "Famiglia", "Abitazione non occupata", "Solo NAD". Il sistema consentiva la memorizzazione delle informazioni di interesse per questi stati: nel caso in cui era stata individuata una famiglia era possibile inserire nome e cognome dell'intestatario e indirizzo dell'abitazione; in caso di un'abitazione non occupata o un'abitazione occupata solo da persone non abitualmente dimoranti, era richiesto solo l'indirizzo. Una volta individuata la famiglia, il Rilevatore provvedeva alla consegna del questionario: se riusciva a consegnare il questionario,

lo poneva in stato "Consegnato"; se non riusciva a consegnare il questionario perché la famiglia, abitualmente dimorante nell'abitazione, risultava assente, poneva il questionario in stato "Temporaneamente assente"; se la famiglia rifiutava il questionario, lo stato veniva posto in "Rifiuto del questionario". I questionari consegnati, venivano restituiti dalle famiglie direttamente al Rilevatore, unico canale di restituzione previsto. La fase di ritiro prevedeva pertanto lo stato "Restituito" se il Rilevatore riceveva il questionario, "Rifiuto a restituire" se la famiglia non restituiva il questionario compilato, "Non più rintracciabile" se il Rilevatore non riusciva a reperire la famiglia cui aveva consegnato il questionario, "Smarrito" se la famiglia aveva smarrito il questionario. In quest'ultimo caso il sistema metteva a disposizione un'apposita funzione per assegnare un nuovo codice questionario alla famiglia e registrare lo smarrimento del vecchio questionario. I questionari restituiti dovevano poi essere revisionati dal Rilevatore (stato "Revisionato"), che doveva anche inserire nel sistema la lista dei componenti della famiglia (stato "Inserita Lista"). Lo stato "Inserita Lista" indicava un esito positivo della rilevazione: il questionario è stato compilato dalla famiglia, restituito al Rilevatore, che lo ha revisionato e ha inserito nel sistema la lista dei componenti della famiglia. Gli altri stati di chiusura del questionario, evidenziati in grassetto, rappresentavano invece casi in cui non si disponeva di un questionario compilato: "abitazione non occupata" e "solo NAD" consentivano di rilevare la presenza di abitazioni in cui non erano presenti famiglie abitualmente dimoranti; gli stati "rifiuto del questionario" e "temporaneamente assente" indicavano l'impossibilità da parte del Rilevatore di consegnare il questionario alla famiglia individuata; gli stati "rifiuto a restituire", "smarrito" e "non più rintracciabile" testimoniavano l'impossibilità da parte del Rilevatore di recuperare il questionario che era stato precedentemente consegnato alla famiglia.

SGRPES consentiva la gestione di tutte le suddette fasi di lavorazione del questionario, direttamente all'interno del "diario di sezione" o attraverso funzioni dedicate. Scendendo nel dettaglio delle operazioni disponibili, il sistema supportava già le fasi *pre-rilevazione*, ovvero la predisposizione di codici questionario da utilizzare durante la rilevazione sul campo, nonché la visualizzazione degli itinerari di sezione. La rilevazione sul campo era supportata principalmente dal "diario di sezione", strumento cardine per il lavoro del Rilevatore: esso mostrava, infatti, per sezione di censimento, le informazioni sui questionari, quali codice, nome e cognome dell'intestatario, indirizzo e stato. Ogni questionario poteva essere lavorato procedendo alla modifica del suo stato e all'inserimento delle relative informazioni. All'interno del "diario di sezione" era possibile visualizzare anche lo storico del questionario, ovvero gli stati per i quali è transitato. Tutto il ciclo di vita del questionario poteva essere, di fatto, gestito all'interno di tale pannello di controllo, dalla fase di individuazione della famiglia all'inserimento della lista. Il Rilevatore aveva pertanto piena operatività sui questionari della propria sezione di censimento, sui quali aveva sempre a disposizione un quadro completo e aggiornato. Per supportare anche il lavoro degli operatori di back-office, che non svolgono direttamente il lavoro sul campo e non hanno sezioni di censimento assegnate, sono state previste funzioni, esterne al "diario di sezione", dedicate all'espletamento di alcune fasi di lavorazione del questionario, quali la registrazione della presenza presso l'ufficio comunale del cartaceo compilato o l'inserimento dei componenti della lista. Il sistema offriva, infine, alcune funzioni di utilità che supportavano casistiche particolari, quali ad esempio lo smarrimento di un questionario o l'individuazione e l'inserimento di un nuovo toponimo all'interno di una sezione di Censimento.

5.4 Lista e modelli riepilogativi

I modelli riepilogativi permettevano di inserire, per ciascuna famiglia trovata sul territorio, i corrispondenti individui dimoranti abitualmente nell'abitazione. Per ogni individuo erano richiesti i principali dati di anagrafica, ovvero nome, cognome, data di nascita e cittadinanza. Le informazioni

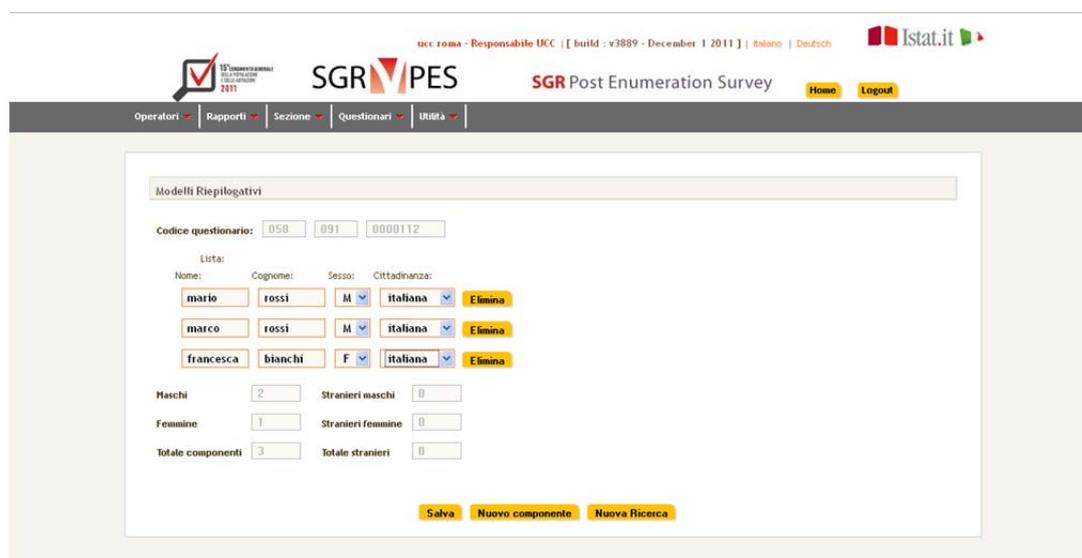
raccolte in questa fase, oltre per il calcolo dei valori aggregati per questionario, sono state poi utilizzate per il *linkage* con i dati provenienti dal Censimento della popolazione. Durante la rilevazione era sempre possibile cambiare i dati inseriti e modificare ogni singolo componente della lista. Compatibilmente con il profilo dell'utente connesso, era possibile accedere ai “Modelli Riepilogativi” per “Inserimento”, “Modifica” ed “Eliminazione”. Relativamente a ciascun Comune, l'utente doveva digitare il codice questionario (Figura 5.5).

Figura 5.5 - Maschera di gestione dei modelli riepilogativi



Nel caso in cui il codice questionario digitato fosse di sua competenza, l'operatore poteva inserire, modificare o eliminare la lista e i modelli riepilogativi. Nella prima schermata l'utente poteva aggiungere o eliminare singoli componenti della famiglia. Il sistema, una volta inserito o eliminato un componente, aggiornava automaticamente il numero totale dei maschi, delle femmine, degli stranieri suddivisi in maschi e femmine. Con la funzione “Eliminazione”, il sistema permetteva la cancellazione dell'intera lista inserita e del modello riepilogativo.

Figura 5.6 - Maschera di gestione dei modelli riepilogativi funzione “eliminazione”



5.5 Monitoraggio del lavoro sul campo

Controllare e gestire lo stato del questionario è la funzione centrale del sistema di monitoraggio della rilevazione. A tal fine, nell'applicazione sono presenti sia operazioni specifiche, che consentono il passaggio 'guidato' da uno specifico stato a un altro; e operazioni di dettaglio come la funzione 'Diario', che consente di guidare le varie operazioni sul campo e di supervisionare l'esito della rilevazione su ciascuna unità. Attraverso la funzione 'Diario' di Sezione è possibile, a fronte di una ricerca per codice questionario e/o sezione di competenza (Figura 5.7), visualizzare i dati di uno o più questionari presenti nel sistema. Ai fini del monitoraggio, risulta molto utile poter visualizzare tutte le operazioni registrate dal sistema ed effettuate su quel particolare questionario, attraverso l'icona di "Storico del questionario" e modificare lo stato del questionario (per la compatibilità dei vari stati si rimanda al paragrafo 5.3) attraverso l'icona "Modifica" qualora il Rilevatore individui un'abitazione non occupata (o occupata solo da persone non abitualmente dimoranti) attribuisce il codice questionario al solo indirizzo dell'abitazione e modifica rispettivamente lo stato del questionario in "Abitazione non occupata" (operazione effettuabile anche attraverso la specifica funzione "Utilità/Inserimento nuova riga").

Individuata la famiglia, il Rilevatore associa al codice questionario nome e cognome dell'intestatario, indirizzo dell'abitazione e aggiorna lo stato del questionario. Lo stato "Famiglia" rappresenta l'avvenuta associazione tra il codice questionario e un intestatario: sia nel caso in cui il questionario venga effettivamente consegnato ad una famiglia, ma anche in caso di mancata consegna, rifiuto a rispondere o assenza della famiglia dall'abitazione. L'esito della consegna del questionario alla famiglia viene registrato attraverso l'icona "Modifica", che consente anche il passaggio dallo stato "Famiglia" agli stati "Consegnato", "Rifiuto del questionario" o "Temporaneamente assente".

Se il questionario viene compilato dalla famiglia e restituito al Rilevatore, lo stesso provvede a modificare lo stato del questionario in "Restituito". Viceversa, nel caso in cui il Rilevatore non riceva il questionario compilato, modifica lo stato del questionario in "Rifiuto a restituire" o "Non più rintracciabile".

A restituzione avvenuta, l'operatore deve revisionarlo, modificandone lo stato in "revisionato": per i questionari in stato "revisionato", è disponibile l'icona per accedere alla funzione di inserimento dei modelli riepilogativi e della lista dei componenti della famiglia.

In caso di errore, è possibile ripristinare lo stato precedente del questionario, attraverso l'apposita icona "Ripristina stato".

Figura 5.7 - Maschera di monitoraggio funzione "Diario"

Diario di Sezione: Visualizzazione

Questionario

Regione

Provincia

Comune

Sezione

Figura 5.8 - Maschera di monitoraggio funzione “Diario” particolare della funzione “Aggiornamento”

Diario di Sezione: Aggiornamento

Parametri della ricerca
 Sezione: 1010005 , Rilevatore: bianchi mario - 0580910963

Filtri per la ricerca
 Stato questionario: Tutti
 Indirizzo: Tutti

90 record trovati, visualizzati 41 - 90. [Prima/Prec.] 1, 2, 3, 4, 5, 6, 7, 8 [Succ./Ultima]

Codice questionario	Cognome intestatario	Nome intestatario	Indirizzo	Civico	Esponente	Interno	Stato questionario	Data questionario	Storico del questionario	Modifica Ripristina Stato	Modello reipilogativo
0580910000301							Codice da assegnare	02-02-2012			
0580910000302							Codice da assegnare	02-02-2012			
0580910000303							Codice da assegnare	02-02-2012			
0580910000304							Codice da assegnare	02-02-2012			
0580910000305							Codice da assegnare	02-02-2012			
0580910000310							Codice da assegnare	07-02-2012			
0580910000311							Codice da assegnare	07-02-2012			
0580910000312							Codice da assegnare	07-02-2012			
0580910000313							Codice da assegnare	07-02-2012			
0580910000314							Revisionato	05-04-2012			

Figura 5.9 - Maschera della funzione “Storico del questionario”

Storico questionario

Codice questionario	Stato questionario	Data questionario
0580910000314	Inserita Lista	02-04-2012
0580910000314	Codice da assegnare	07-02-2012
0580910000314	Revisionato	07-02-2012

Esporta dati:

Figura 5.10 - Maschera della funzione “Modifica”

Questionario

Codice questionario: 058 091 0000314

Indirizzo: ACCETTA CRISTOFORO COLOMBO

Civico:

Esponente:

Interno:

Cognome:

Nome:

Stato questionario: Restituito

5.6 Rapporti riassuntivi

I rapporti riassuntivi sono le funzioni di monitoraggio della rilevazione, e il loro fine è consentire una valutazione in tempo quasi reale dell'andamento delle operazioni sul campo. Queste funzioni visualizzano tabelle di dati aggregati che riassumono l'evoluzione della rilevazione in termini di quantità di questionari, suddivisi per fase della lavorazione (questionari in lavorazione, chiusi e riepilogo), stato del questionario e utente incaricato alla lavorazione. I dati sono aggregati per i diversi livelli territoriali previsti e possono essere navigati, compatibilmente con il profilo dell'operatore connesso, mediante la tecnica del *Drill-down*⁵⁴ e del *Roll-back*⁵⁵.

I livelli territoriali previsti sono:

- ▶ Intero territorio nazionale: l'operatore visualizza tutte le informazioni suddivise per Regione, Provincia, Comune e Sezione;
- ▶ Regionale: l'utente visualizza le informazioni suddivise per Province, Comuni e Sezioni della regione di competenza;
- ▶ Prefettura: l'utente visualizza le informazioni suddivise per Comuni e Sezioni della Prefettura di competenza;
- ▶ Provinciale: l'utente visualizza le informazioni suddivise per Comuni e Sezioni della provincia di competenza;
- ▶ Comunale: l'utente visualizza le informazioni suddivise per le Sezioni del comune di competenza.

A cui si aggiunge un livello non-territoriale, ma basato sull'operatore della rete che visualizza le informazioni, suddivise per singolo operatore del comune selezionato.

Le funzioni di monitoraggio in SGRPES sono:

- ▶ Stato lavorazione questionari;
- ▶ Questionari chiusi per tipologia;
- ▶ Lavoro dell'Operatore;
- ▶ Modelli riepilogativi.

Nelle funzioni di monitoraggio "Stato Lavorazione questionari", "Questionari chiusi per tipologia" e "Modelli riepilogativi", il massimo livello di dettaglio è la sezione (tutti i livelli sono territoriali), mentre nella funzione "Lavoro dell'operatore" le informazioni al massimo livello di dettaglio sono quelle relative al singolo operatore (i precedenti livelli sono territoriali). Per ogni funzione è presente lo scarico nei formati *Portable Document Format* (PDF)⁵⁶ ed Excel (.xls)⁵⁷ dei dati visualizzati nella pagina web. Mentre, per i soli rapporti riassuntivi territoriali, è presente lo scarico nel formato *Comma-Separated Values* (CSV)⁵⁸, di tutti i comuni della regione e di tutte le sezioni della provincia. Nei prospetti successivi (da Prospetto 5.2 fino a Prospetto 5.5) sono riportate le descrizioni delle informazioni specifiche mostrate da ogni funzione di monitoraggio.

⁵⁴ La tecnica del Drill-down permette la visualizzazione e/o la ricerca di informazioni gerarchizzate.

⁵⁵ La tecnica Roll-back è un'operazione che permette di riportare il database a una versione precedente.

⁵⁶ Il formato Portable Document Format (.pdf) è un formato di file basato su un linguaggio di descrizione pagina.

⁵⁷ Il formato Excel (.xls) è un formato di file basato su un foglio elettronico.

⁵⁸ Il formato Comma-Separated Values (.csv) è un formato di file basato su file di testo utilizzato per l'importazione e l'esportazione di una tabella di dati.

Prospetto 5.2 - Stato lavorazione questionari per sezione

Colonna	Colonna corrispondente nel rapporto riassuntivo
Regione/Provincia/Comune/Sezione	Indicazione del territorio geografico (nome regione, nome provincia, nome comune , numero sezione)
Righe Diario	Nr. totale di righe del diario con stato diverso da 'Codice da assegnare'
Questionari consegnati	Nr. di questionari nello stato 'consegnato'.
Famiglie individuate	Nr di questionari per i quali lo stato nel diario è 'famiglia'.
Questionari restituiti	Nr. di questionari per i quali lo stato nel diario è 'restituito'
Revisionati	Nr. di questionari per i quali lo stato nel diario è 'revisionato'.
Righe chiuse	Nr. totale di questionari negli stati finali di lavorazione (Inserita lista, Smarrito, Rifiuto a restituire, non più rintracciabile, rifiuto del questionario, Temporaneamente assente, Abitazione non occupata, solo NAD).

Prospetto 5.3 - Questionari chiusi per tipologia

Colonna	Colonna corrispondente nel rapporto riassuntivo
Regione/Provincia/Comune/Sezione	Indicazione del territorio geografico (nome regione, nome provincia, nome comune, nominativo Rilevatore)
Righe Chiuse	Nr. totale di questionari negli stati finali di lavorazione (Inserita lista, Smarrito, Rifiuto a restituire, non più rintracciabile, rifiuto del questionario, Temporaneamente assente, Abitazione non occupata, solo NAD).
Rifiuto	Nr. totale di questionari rifiutati (include gli stati 'rifiuto del questionario' che 'rifiuto a rispondere')
Temporaneamente assente	Nr. di questionari lavorati nello stato 'Temporaneamente assente'.
Solo NAD	Nr. di questionari nello stato 'Solo Nad'.
Lista	Nr. di questionari per i quali sono compilate la lista ed i modelli riepilogativi, ovvero nello stato 'Inserita lista'.
Abitazioni non occupate	Nr. di questionari non consegnati presso l'abitazione perché individuo assente, ovvero nello stato 'Abitazione non occupata'.
Smarriti	Nr. di questionari nello stato 'Smarrito'.
Non più rintracciabile	Nr. di questionari nello stato 'non più rintracciabile'.

Prospetto 5.4 - Lavoro dell'Operatore

Colonna	Colonna corrispondente nel rapporto riassuntivo
Regione/Provincia/Comune/Rilevatore	Indicazione del territorio geografico (nome regione, nome provincia, nome comune, nominativo Rilevatore)
Consegnato	Nr. di questionari lavorati dall'operatore nello stato 'Consegnato'.
Restituiti	Nr. di questionari lavorati dall'operatore nello stato 'Restituito'.
Rifiuto	Nr. totale di questionari nello stato di 'Rifiuto' (include gli stati 'Rifiuto del questionario' che 'Rifiuto a rispondere').
Temporaneamente assente	Nr. di questionari lavorati dall'operatore nello stato 'Temporaneamente assente'.
Solo NAD	Nr. di questionari lavorati dall'operatore nello stato 'Solo NAD'.
Lista	Nr. di questionari per i quali sono stati inseriti dall'operatore i componenti della famiglia, ovvero nello stato 'Inserita lista'.
Abitazioni non occupate	Nr. di questionari non consegnati presso l'abitazione perché individuo assente, ovvero nello stato 'Abitazione non occupata'.
Smarriti	Nr. di questionari nello stato 'Smarrito'.
Non più rintracciabile	Nr. di questionari nello stato 'Non più rintracciabile'

Prospetto 5.5 - Modelli riepilogativi

Colonna	Colonna corrispondente nel rapporto riassuntivo
Nr. Famiglie	Indicazione del territorio geografico (nome regione, nome provincia, nome comune, numero sezione)
Persone abitualmente dimoranti Maschi	Nr. totale di righe nello stato 'Inserita lista'.
Persone abitualmente dimoranti Femmine	Nr. totale di individui di sesso maschile abitualmente dimoranti.
Persone abitualmente dimoranti Totale	Nr. totale di individui di sesso femminile abitualmente dimoranti .
Persone abitualmente dimoranti Di cui stranieri maschi	Nr. totale di individui abitualmente dimoranti .
Persone abitualmente dimoranti Di cui stranieri femmine	Nr. totale di individui stranieri abitualmente dimoranti maschi femmine
Totale stranieri	Nr. totale di individui stranieri abitualmente dimoranti femmine

5.7 Test

Il test ha come principale obiettivo la verifica della correttezza funzionale di una applicazione rispetto ai requisiti del committente (*test funzionali*), ma sempre più spesso viene usato come tecnica di base per la realizzazione di controlli mirati alla valutazione di altri fattori di qualità come ad esempio affidabilità, usabilità (si vuole valutare la facilità d'uso dell'applicazione da parte dell'utente finale), efficienza, sicurezza, interoperabilità, prestazioni (*test di carico*: l'applicazione è sottoposta al carico di lavoro massimo previsto dai requisiti e le sue funzionalità sono controllate in

queste condizioni), robustezza (*stress test*: l'applicazione è sottoposta a carichi di lavoro superiori a quelli previsti dai requisiti o è portata in condizioni operative eccezionali – in genere sottraendogli risorse di memoria e di calcolo per controllare la capacità di “*recovery*” (recupero) del sistema dopo un fallimento) in modo da garantire un prefissato livello di qualità del prodotto.

Strategia generale di test

La metodologia di progettazione dei test utilizzata per il progetto SGRPES si riassume brevemente nei seguenti punti:

1. Analisi dei documenti delle specifiche funzionali e dei requisiti e dei documenti di progetto;
2. Individuazione delle criticità per ogni funzionalità prevista;
3. Definizione dei test case tesi alla valutazione delle singole funzionalità dell'applicazione. Per ogni test case, in base alle specifiche di progetto, occorre identificare le pre-condizioni che devono essere soddisfatte prima dell'esecuzione del test, l'insieme di azioni (actions) da intraprendere per effettuare il test e il comportamento atteso dall'applicazione in seguito a tali azioni (expected outcome);
4. Aggregazione di più test case in modo da formare un test plan specifico per ciascuna funzionalità completa;
5. Avvio della fase iterativa di testing e tracciamento dei test effettuati (test run): occorre tracciare l'identificativo del test case in esame, e il risultato (positivo o negativo) corredato da alcune informazioni aggiuntive (data, versione del prodotto);
6. Implementazione di procedure semiautomatiche, quando è possibile, per la ripetibilità dei test e la validazione di casi limite;
7. In caso di riscontro di una anomalia (un discostamento fra il comportamento desiderato per l'applicazione e quello effettivo, cioè quello previsto dalle specifiche) occorre inviare una segnalazione (in breve bug), corredata da tutte le informazioni, ricavabili dal test case e utili allo sviluppatore per ripetere l'anomalia (pre-condizioni, azioni, expected outcome, browser utilizzato, versione del software), insieme alla descrizione del comportamento errato (actual outcome). Il bug viene inoltrato, per e-mail, sia allo sviluppatore che ha implementato la funzionalità per la correzione, sia agli analisti che ne hanno redatto i documenti di analisi per una loro supervisione. Il raffronto con gli analisti infatti è una assicurazione contro una non corretta interpretazione delle specifiche da parte dei tester o degli sviluppatori;
8. I test, in presenza di eventuali variazioni delle specifiche, vanno rivisti sin dalla definizione dei test case e vanno ripetuti una volta corretta l'anomalia e in generale ad ogni nuovo rilascio parziale o definitivo, in modo da verificare la correzione di errori già segnalati o l'introduzione di nuovi errori (test di regressione);
9. Individuazione degli stakeholder (sedi territoriali Istat, comuni, province, rete di rilevazione ecc.) per effettuare il beta testing⁵⁹: una volta che i moduli che compongono l'applicazione sono stabilizzati, il “cliente”, cioè un campione degli utenti finali comincia a prendere visione del prodotto. In questa fase infatti può dare dei suggerimenti utili per migliorare ulteriormente la qualità dell'applicazione in termini di usabilità;
10. Rilascio della documentazione relativa ai risultati delle campagne di test.

⁵⁹ Una terminologia molto diffusa divide i controlli finali in alfa-test e beta-test. I due termini distinguono i controlli rispetto al soggetto che li esegue. Il primo stadio di testing (alfa) sarà svolto dal gruppo di testing con il supporto dei coordinatori dei progetti di sviluppo. Seguirà il beta testing in cui verranno coinvolti attivamente i rappresentanti degli stakeholder.

Chi si occupa dei test oltre alle verifiche nell'ambito degli scenari previsti dall'analisi, deve saper prevedere il comportamento di utenti maliziosi, che possono sfruttare eventuali punti deboli della sicurezza per fini non corretti, e il comportamento di utenti distratti, che possono compiere azioni completamente avulse dagli scenari previsti in fase di analisi, che magari non hanno senso dal punto di vista logico, ma che l'applicazione permette (e se l'applicazione permette una determinata azione, prima o poi qualcuno la compie) con delle conseguenze che possono manifestarsi quando il software è in pieno regime d'uso, danneggiando l'utente e squalificando l'immagine del fornitore (*difetti quiescenti*).

Bisogna tener presente infine che il test può indicare la presenza di errori, ma non ne può garantire l'assenza, il software error-free non esiste o non è certificabile.

La Base Strumenti a supporto dei test:

Bugzilla è un programma di gestione dei ticket. L'esperienza mostra come questo programma, in unione ad un suo add-on Testopia, sia di fondamentale importanza, in quanto soddisfa molto validamente parecchie esigenze:

- ▶ la definizione dei test plan e dei test case (pre-condizioni, actions ed expected outcome);
- ▶ il tracciamento dei test run e la redazione assistita delle segnalazioni in caso di anomalie: il ticket che descrive l'anomalia infatti è pre-compilato a partire dal relativo test case;
- ▶ la gestione in modo agile e puntuale delle comunicazioni fra tester, analisti e sviluppatori: Bugzilla permette infatti, tramite una interfaccia di amministrazione, di mappare ogni singola funzionalità dell'applicazione da testare con lo sviluppatore che l'ha implementata e l'analista che ne ha redatto le specifiche, in modo tale che in modo automatico ogni singola segnalazione venga inoltrata, tramite e-mail, solo alle persone direttamente interessate. Ad ogni segnalazione inoltre viene associato un livello di priorità (critica, normale, minore, suggerimento, ecc.). tutti i soggetti interessati infine possono aggiungere le loro osservazioni e i loro commenti;
- ▶ la possibilità di far tesoro delle esperienze di ogni operatore: una volta che il bug è risolto lo sviluppatore può aggiungere un suo commento con la descrizione della soluzione adottata. Queste informazioni possono essere utili ad altri soggetti che prendono parte al progetto o ad altri progetti simili;
- ▶ la gestione condivisa online del workflow del bug.

Bugzilla offre inoltre ai tester un' interfaccia di amministrazione molto flessibile, per gestire, fra l'altro, utenze e relativi permessi, configurare il *workflow* dei bug, configurare alcuni servizi di ricerca, ecc.

Selenium è un add-on del browser Firefox. Quando si lavora con il browser, Selenium offre la possibilità di registrare delle macro, cioè di registrare una sequenza di azioni che da tastiera o da mouse si fanno lavorando con il browser sull'applicazione. La macro può in seguito esser riavviata in modo da rieseguire in automatico tutta la sequenza di azioni.

Le macro inoltre sono modificabili e configurabili anche attraverso il codice script ad esse associato, permettendo di ripetere un determinato test con l'introduzione di alcune varianti (per es. con diversi profili utente, o lavorando su questionari di province diverse).

Le macro di Selenium quindi, con un intervento minimo dell'operatore umano, si prestano bene a ripetere periodicamente alcuni tipi di test *run*, consentendo un notevole risparmio di tempo e risultando particolarmente utili per i test di regressione. Infine sono cross-browser, cioè possono essere utilizzate con i vari browser (Firefox, Internet explore, Chrome, ecc) nelle varie versioni.

Tamper data è un add-on del browser Firefox. Esso è utile nei test di sicurezza in quanto mette in evidenza eventuali punti deboli nelle comunicazioni fra il client (il browser) e il server

(l'applicazione), consentendo ai tester di simulare attacchi da parte di utenti maliziosi. Tamper data permette inoltre di verificare se informazioni che dovrebbero essere criptate sono invece inviate al server in chiaro.

IETester , Windows virtual PC e Oracle virtual box. Il cross-browser.

Quando si sviluppa un'applicazione web accessibile tramite un browser, un problema che non va sottovalutato è quello del cosiddetto *cross-browser*: il comportamento dell'applicazione deve essere lo stesso al variare del browser che si utilizza fra quelli di uso comune (Firefox, Internet explorer, Chrome, nel caso del Censimento).

Il cross-browser è quindi un problema che va affrontato in modo sistematico anche dal gruppo di testing.

Fra gli strumenti utili si segnalano IETester, un programma che simula il comportamento di varie versioni del browser Internet explorer (dalla IE10 preview, IE9, IE8, IE7 IE6 e IE5.5) e soprattutto Windows virtual PC ed Oracle virtual box, che simulano rispettivamente gli ambienti Windows XP e Ubuntu, su cui installare una determinata versione del browser.

5.8 Utilizzo del sistema

La rete di rilevazione era formata da 1.667 operatori, come ci si attendeva, il numero maggiore di operatori è da identificarsi tra i Rilevatori, che hanno rappresentato il 73 per cento del totale, mentre gli UCC salvo pochi casi sono rimasti con un solo responsabile per comune. La tabella sottostante mostra in dettaglio il numero di operatori suddivisi per ruoli.

Tavola 5.1 - Distribuzione degli operatori della rete

Ruolo	Numero Operatori	
	ass.	%
Rilevatore	1.208	72,5%
UCC	287	17,2%
Ufficio Provinciale	100	6,0%
RIT Responsabile Istat regionale Territoriale	32	1,9%
RURC Responsabile Ufficio Regionale	21	1,3%
Istat Nazionale	19	1,1%
totale	1667	100%

La distribuzione degli operatori per regione rispecchia la presenza nel campione dei comuni metropolitani, e quindi il maggior numero di operatori è presente nel Lazio, Lombardia e Campania. La Tavola 5-2 riporta la distribuzione per territorio di competenza degli operatori creati in SGR, in ordine di numerosità decrescente.

Il sistema è stato in linea dal 4 aprile al 26 luglio del 2012 e durante tale periodo sono state effettuate 983.999 operazioni, dato che è comunque una sottostima rispetto al reale utilizzo, in quanto considera solo le operazioni tracciabili e quindi esclude ad esempio tutta la fase di solo monitoraggio. La tabella seguente mostra le operazioni suddivise in quattro sotto periodi. Il mese con maggior utilizzo dell'applicazione è maggio, con 600.690 operazioni; in particolare, il 35 per cento del lavoro si è svolto durante la terza decade del mese.

Il numero maggiore di lavorazioni SGR, dal punto di vista del territorio, sono state svolte da Roma, che con un totale di 35.990 ha effettuato il 4 per cento delle operazioni totali; segue il comune di

Milano con 20.051 operazioni. Per valutare il numero di operazioni, si tenga conto che mediamente ogni questionario è stato lavorato in quasi 5 operazioni, quindi in genere si è seguito il flusso standard previsto per la rilevazione (si veda il paragrafo 4.3). A fine rilevazione il numero totale dei questionari lavorati è 184.282, così distribuiti per stato finale (Tavola 5.4).

Tavola 5.2 - Distribuzione degli operatori per regione, al netto degli operatori nazionali

Regione	Numero Operatori	
	ass.	%
Lombardia	206	12,9
Lazio	205	12,8
Veneto	115	7,2
Piemonte	111	7
Campania	101	6,4
Toscana	98	6,1
Emilia-Romagna	96	6
Sicilia	91	5,7
Puglia	89	5,6
Calabria	71	4,4
Friuli Venezia Giulia	69	4,3
Basilicata	47	2,9
Marche	46	2,9
Liguria	44	2,8
Abruzzo	42	2,6
Sardegna	40	2,5
Umbria	38	2,4
Bolzano-Bozen	27	1,7
Trento	21	1,3
Molise	20	1,3
Valle D'Aosta	19	1,2
Totale	1.596	100,0%

Tavola 5.3 - Distribuzione temporale delle operazioni statistiche espletate in SGR

Periodo	Numero Operatori	
	ass.	%
Dal 4 al 30 Aprile	214.249	21,8%
Dal 1 al 31 Maggio	600.690	61,0%
Dal 1 al 30 Giugno	138.909	14,1%
Dal 1 al 16 Luglio	30.151	3,1%
totale	983.999	100,0%

Tavola 5.4 - Distribuzione dei questionari lavorati per stato finale

Stato	Numero questionari	
	ass.	%
Inserita Lista	137.771	74,8%
Abitazione non occupata	37.733	20,5%
Temporaneamente assente	3.583	1,9%
Solo NAD	2.992	1,6%
Rifiuto	1.662	0,9%
Non più rintracciabile	541	0,3%
totale	184.282	100,0%

PARTE SECONDA

GLI ASPETTI METODOLOGICI E I RISULTATI

CAPITOLO 1

Disegno di campionamento, processo di stima e attendibilità delle stime prodotte⁶⁰

1.1 Obiettivi dell'indagine e parametri di interesse

Le informazioni raccolte durante le operazioni di enumerazione censuaria della popolazione sono inevitabilmente soggette ad errore. Le indagini di copertura perseguono l'obiettivo primario di produrre una valutazione del livello di accuratezza di tali informazioni (con riferimento, generalmente, a più popolazioni di interesse) basata sul confronto tra i valori dei parametri di interesse forniti dal Censimento e quelli delle corrispondenti stime ottenute tramite un'indagine di tipo campionario.

La stima dell'errore di sotto-copertura è ottenuta con un modello *dual system* (o cattura-ricattura⁶¹) il quale si avvale di più occasioni di conteggio⁶² (o "cattura", appunto) al fine di stimare il reale ammontare della popolazione. Sebbene la sotto-copertura sia la principale fonte di errore, la sfida chiave della PES 2011 è stata la ricerca di una metodologia che fosse in grado di produrre una stima attendibile del livello di sovra-copertura del Censimento e di come includere tale stima nel modello *dual system*.

Per i fini sopra delineati, i parametri di popolazione oggetto di stima più importanti sono:

- ▶ il tasso di copertura, espresso come rapporto tra il numero di unità enumerate al Censimento (al netto della sovra-copertura) e la dimensione effettiva (incognita) della popolazione,
- ▶ il tasso di sotto-copertura, dato dal rapporto tra il numero di unità sfuggite all'enumerazione censuaria (al netto della sovra-copertura) e la dimensione effettiva (incognita) della popolazione. suggerimenti per gli utili commenti, i preziosi suggerimenti e le segnalazioni di imprecisioni fondamentale.

Essendo non nota la dimensione effettiva della popolazione, per il calcolo dei suddetti parametri, come già evidenziato, è necessario ricorrere ad una procedura di stima basata sull'utilizzo di appositi modelli probabilistici fondati sulla mancata enumerazione delle unità (che può verificarsi sia al Censimento che all'indagine di copertura). La metodologia di stima scelta in questa indagine verrà illustrata in dettaglio nel successivo paragrafo 1.5.

1.2 Popolazione obiettivo e domini di stima

La popolazione di riferimento della PES 2011 è costituita dagli individui residenti nelle abitazioni esistenti nel territorio nazionale (escluse le convivenze) alla data del Censimento (9 ottobre 2011). La finalità primaria dell'indagine in esame è l'ottenimento di stime affidabili dei predetti parametri con riferimento ai seguenti ambiti territoriali e socio-demografici:

- ▶ l'intero territorio nazionale;

⁶⁰ I paragrafi sono tutti a cura di Monica Russo ad eccezione del paragrafo: 1.6 HtC Index a cura di Andrea Fasulo e Marco Dionisio Terribili.

Rivolgo un vivo ringraziamento ai dottori Owen Abbott e Amy Large dell'Office for National Statistics (ONS) per gli utili commenti, i preziosi suggerimenti e le segnalazioni di imprecisioni di cui ho tenuto conto nella redazione del testo della presente nota e, in particolare, del paragrafo 1.5.

⁶¹ Sekar, C.C., Deming W.E. (1949).

⁶² Nel presente contesto d'indagine, le occasioni di cattura sono il Censimento e l'indagine di copertura.

- ▶ le regioni geografiche e le province autonome di Trento e Bolzano;
- ▶ le classi di età: 0-14, 15-29, 30-49, 50-64, 65 e più;
- ▶ la nazionalità, secondo le modalità: italiana, straniera.

I primi due domini a differenza degli ultimi due presentano l'importante caratteristica di essere domini pianificati, essendo le loro dimensioni campionarie definite a priori; ciò nonostante, la dimensione campionaria e la robustezza del processo di stima scelto per l'ottenimento delle stime dei parametri di interesse hanno permesso di ottenere stime attendibili anche per i domini definiti dalle classi di età e dalle nazionalità.

Anche se non presenti in questo volume, con i dati disponibili sarà possibile determinare le stime in oggetto anche con riferimento a domini territoriali sub-regionali (le province) e sub-provinciali (i tredici comuni metropolitani). Essendo questi ultimi domini non pianificati, la precisione delle stime ad essi inerenti può essere migliorata sia introducendo una post-stratificazione nello stimatore studiato per l'indagine, sia attraverso l'adozione di stimatori speciali per piccole aree. L'ottenimento delle stime in questione rappresenta comunque un obiettivo non richiesto dai regolamenti.

1.3 Errori di copertura

I principali tipi di errore di cui può essere affetta l'operazione di enumerazione censuaria sono:

1. *l'errore di sotto-copertura*, che si verifica quando una unità residente in Italia alla data del 9 ottobre 2011 non viene rilevata dal Censimento.
2. *l'errore di sovra-copertura*, che può manifestarsi nelle seguenti quattro forme:
 - *tipo 1: duplicati relativi alla stessa sezione di Censimento*, che si verifica quando per lo stesso individuo esistono due o più ritorni censuari nella stessa sezione di Censimento (ad esempio, una famiglia compila il questionario sia online che cartaceo);
 - *tipo 2: duplicati relativi a differenti sezioni di Censimento*, che si verifica quando per lo stesso individuo si hanno due ritorni censuari ma in sezioni di Censimento diverse (ad esempio, uno studente enumerato nella sezione di Censimento della famiglia di origine (conteggio errato) e nel luogo dove egli si è trasferito per studiare (conteggio corretto));
 - *tipo 3: conteggi nella sezione di Censimento errata*, che si verifica quando un individuo viene enumerato solo all'indirizzo errato; quindi, è presente solo la metà errata della duplicazione nei ritorni censuari (ad esempio, uno studente enumerato solo all'indirizzo della famiglia di origine o una persona che ha traslocato dopo la data di riferimento del Censimento conteggiata al nuovo indirizzo);
 - *tipo 4: rilevazioni erronee*, è un ritorno censuario che non avrebbe mai dovuto verificarsi poiché fittizio o perché non facente parte della popolazione di riferimento (ad esempio, un bambino nato dopo la data di riferimento del Censimento, una persona morta prima della data di riferimento del Censimento, una persona che non era residente in Italia alla data del 9 ottobre 2011, una persona il cui nome è un palese errore, tipo Felix the Cat).

In fase di stima si considerano solo gli errori di tipo 2 e 3 che sono stimati separatamente⁶³ ed inseriti in un fattore di aggiustamento applicato al modello di stima. Gli errori di tipo 1, infatti, sono corretti durante il processamento dei dati censuari attraverso procedure *ad hoc* mirate a rettificare

⁶³ Per una descrizione dettagliata del processo di stima si veda il paragrafo 1.5.3.

e risolvere per quanto possibile questo tipo di sovra-copertura, mentre quelli di tipo 4 sono identificati con ulteriore lavoro sul campo⁶⁴.

Storicamente, in Italia la sovra-copertura è stata considerata un problema meno rilevante nei censimenti passati; infatti, essa è stata stimata solo a partire dalla precedente tornata censuaria⁶⁵. La sua stima per il 2001 è risultata pari a 0,32 per cento a livello nazionale, registrando, quindi, un valore sensibilmente più basso di quello ottenuto per il tasso di sotto-copertura (1,45 per cento la sotto-copertura lorda e 1,12 per cento quella al netto della sovra-copertura). Conseguentemente, sebbene tutti i tipi di errore su descritti inficiano l'attendibilità del conteggio effettuato al Censimento, quello che maggiormente incide negativamente è l'errore di sotto-copertura, in quanto l'errore di sovra-copertura presenta in genere un'influenza di tipo residuale.

Per meglio comprendere quanto appena affermato, nella tavola 1.1 sono riportati le stime dei tassi di sovra-copertura e sotto-copertura (al lordo e al netto della sovra-copertura) relativi a cinque Paesi e a differenti anni.

Tavola 1.1 - Comparazione internazionale delle stime dei tassi di sovra-copertura, sotto-copertura e sotto-copertura netta

Paese	Anno	Tasso di Sovra-copertura (%)	Tasso di Sotto-copertura (%)	Tasso di Sotto-copertura netta (%)
Svizzera	2000	0,3	1,6	1,4
Canada	1996	0,7	3,2	2,4
	2001	1,0	4,0	3,0
	2006	1,6	4,3	2,7
	2011	1,9	4,1	2,2
Stati Uniti	1990	3,1	4,7	1,6
	2000	-	-	-0,5
	2010	-	-	-0,01
Australia	1996	0,2	1,8	1,6
	2001	0,9	2,7	1,8
	2006	-	-	2,7
	2011	-	-	1,7
Regno Unito	2001	0,4*	6,1	5,7
Inghilterra e Galles	2011	0,6	7,2	6,6

Fonte: Adattata da Renaud, A. (2007) e aggiornata con i dati del Censimento 2011.

*Stima fornita dall'ONS-Longitudinal Study 2001.

⁶⁴ Il numero di ritorni censuari errati dovrebbe essere trascurabile tenuto conto delle precauzioni prese per soddisfare l'ipotesi che la popolazione sia chiusa, così come descritto nel paragrafo 1.5.3 e 1.5.4.

⁶⁵ Le stime prodotte dall'indagine di copertura del 2001, tuttavia, devono essere considerate solo sperimentali in quanto esse non sono state ottenute attraverso un *record linkage*, ma ricorrendo a una serie di domande somministrate direttamente agli individui, dal momento che sarebbe altrimenti stato necessario disporre dei nominativi di tutti i censiti e non solo di quelli enumerati nelle sezioni campione per l'indagine di copertura (Istat, 2009).

Dai dati⁶⁶ si evince che la sovra-copertura è generalmente bassa, inferiore all'1 per cento; tuttavia, concentrando l'attenzione sulla serie storica del Canada e dell'Australia, risulta evidente che i valori dei tassi hanno una tendenza crescente nel tempo. In Italia, per il Censimento del 2011, con i cambiamenti nella metodologia di raccolta dei dati e del comportamento sociale, era possibile aspettarsi che la sovra-copertura sarebbe stata più elevata⁶⁷.

Si indichino ora con: N_{CENS} il totale della popolazione residente (in termini di individui); ${}_L\hat{N}_{PES}$ e ${}_N\hat{N}_{PES}$ le stime del numero di individui fornite dalla PES rispettivamente al lordo e al netto della sovra-copertura; ${}_L S^T$ e ${}_N S^T$ il numero di casi di sotto-copertura del Censimento al lordo e al netto della sovra-copertura; S^V la sovra-copertura del Censimento; ${}_L TS^T$ e ${}_N TS^T$ i tassi di sotto-copertura del Censimento al lordo e quello al netto della sovra-copertura; TS^V il tasso di sovra-copertura del Censimento. La dimensione effettiva della popolazione residente può essere espressa nella forma

$$N = {}_N\hat{N}_{PES} = N_{CENS} + ({}_L S^T - S^V) = N_{CENS} + {}_N S^T \quad (1)$$

in cui

$${}_N S^T = {}_L S^T - S^V \quad (2)$$

è detto *errore netto di sotto-copertura*; inoltre, sono valide le seguenti relazioni

$${}_L S^T = {}_L\hat{N}_{PES} - N_{CENS} \quad (3)$$

$${}_L TS^T = \left(1 - \frac{N_{CENS}}{{}_L\hat{N}_{PES}}\right) 100 = \frac{{}_L S^T}{{}_L\hat{N}_{PES}} 100 \quad (4)$$

$${}_N TS^T = \left(1 - \frac{N_{CENS}}{{}_N\hat{N}_{PES}}\right) 100 = \frac{{}_N S^T}{{}_N\hat{N}_{PES}} 100 \quad (5)$$

$$TS^V = {}_L TS^T - {}_N TS^T. \quad (6)$$

Da quanto sopra esposto, discende che la popolazione conteggiata al Censimento coincide con la dimensione effettiva incognita della popolazione solo nel caso (alquanto inverosimile) in cui l'errore netto di sotto-copertura sia nullo.

1.4 Disegno di campionamento

1.4.1 Caratteristiche generali

In generale, i Censimenti sono strettamente connessi col territorio sul quale insiste la popolazione con riferimento alla quale se ne vuole determinare l'ammontare ed acquisire notizie su vari caratteri di diversa natura (sociale, economica, ecc.).

Quelli eseguiti in Italia sulla popolazione e sulle abitazioni si fondano su una partizione in sub-aree (dette sezioni di Censimento) dell'intero territorio nazionale.

L'indagine di copertura consiste in una ripetizione delle operazioni di enumerazione su un campione casuale delle suddette sezioni che, nel disegno di campionamento su cui si basa la PES 2011, hanno il ruolo di unità finali di campionamento.

⁶⁶ Va notato che le variazioni sia tra Paesi che tra anni possono essere dovute, in parte, anche, ad esempio, a differenti metodologie adottate per il Censimento.

⁶⁷ I risultati della PES 2011 sono riportati nel capitolo quarto.

Peraltro, non sarebbe stato possibile utilizzare altri tipi di aree (segmenti o parti di sezioni), in quanto – al momento della progettazione della PES 2011 – non erano disponibili informazioni inerenti a sub-aree delle sezioni di Censimento, necessarie ai fini dell'abbinamento tra i dati del Censimento con quelli rilevati dall'indagine di copertura.

Il disegno di campionamento in esame è di tipo areale a due stadi di selezione: le unità primarie (UP) sono costituite dai comuni, quelle secondarie (US) dalle sezioni di Censimento; tutti gli individui appartenenti alle sezioni campione vengono enumerati.

Le ragioni alla base della scelta di questo disegno sono diverse: necessità, o opportunità, di tipo organizzativo; riduzione dei costi e dei tempi tecnici di esecuzione dell'indagine.

Infatti, l'introduzione dei comuni come primo stadio di campionamento consente di tenere sotto controllo il numero dei comuni coinvolti nell'indagine; il disegno alternativo consistente nella formazione di un campione di sezioni di Censimento avrebbe comportato un aumento del numero di comuni campione e, di conseguenza, un aumento dei costi dell'indagine.

Il disegno di campionamento in esame presenta, inoltre, le due seguenti importanti caratteristiche:

- ▶ sia le unità primarie che quelle secondarie sono stratificate;
- ▶ il meccanismo probabilistico di formazione del campione consiste nell'estrazione delle unità primarie con probabilità variabili senza ripetizione e nell'estrazione delle unità secondarie con probabilità uguali senza ripetizione.

L'indagine ha coinvolto 252 comuni campione e 2.488 sezioni⁶⁸, per un totale di 332.710 individui intervistati.

1.4.2 Scelta delle variabili di stratificazione

Al fine di introdurre ulteriori guadagni di efficienza (a parte quelli dovuti all'utilizzo di probabilità variabili e di stimatori speciali) nelle stime uscenti dall'indagine in discussione, si è proceduto alla stratificazione sia delle unità primarie che delle unità secondarie.

Per quanto concerne il primo stadio di campionamento, tutti i comuni esistenti nell'intero territorio nazionale sono stati ripartiti in primo luogo in centocinque strati – nel prosieguo, verranno richiamati con la locuzione *strati dell'indagine* – identificati mediante il concatenamento delle regioni geografiche (le province autonome di Trento e Bolzano sono trattate come regioni a sé) e delle cinque classi di dimensione demografica dei comuni secondo le modalità:

- ▶ meno di 5 mila ab.
- ▶ tra 5 mila e 10 mila ab.
- ▶ tra 10 mila e 20 mila ab.
- ▶ tra 20 mila e 100 mila ab.
- ▶ oltre 100 mila ab. e i 13 comuni metropolitani (Torino, Genova, Milano, Venezia, Bologna, Firenze, Roma, Napoli, Bari, Palermo, Catania, Cagliari e Messina).

In realtà, il numero effettivo di strati dell'indagine è pari a cento, poiché cinque di essi sono risultati privi di unità.

Per la stratificazione dei comuni e delle sezioni dentro ciascuno degli strati dell'indagine h ($h = 1, \dots, H = 100$) vale la pena ricordare che da precedenti studi condotti per la progettazione dell'indagine di copertura del 14° Censimento della popolazione⁶⁹ (Di Consiglio, Falorsi, 2003) è

⁶⁸ In realtà, il campione teorico di primo stadio era costituito da 255 comuni e quello di secondo stadio da 2.507 sezioni; tuttavia, poiché tre comuni non hanno partecipato all'indagine, il numero di sezioni interessate è stato di 2.488.

⁶⁹ In particolare, dal lavoro di Di Consiglio e Falorsi (2003) si evince che:
 «Per la definizione della stratificazione dei comuni e delle sezioni all'interno di ogni T_h ($h=1, \dots, 20$) è stata condotta un'analisi volta a valutare la relazione tra il tasso di copertura (desunto dall'indagine di copertura del Censimento della

emerso che tra le variabili ausiliarie, considerate per i fini sopra delineati, quelle maggiormente atte a migliorare la qualità delle stime dell'indagine sono: la dimensione demografica del comune e il tipo di sezione.

Questi risultati, insieme a quelli ottenuti con altri studi eseguiti sulle indagini Istat sugli individui e sulle famiglie, hanno guidato verso la scelta di usare – anche per l'indagine in esame – la dimensione demografica per la stratificazione dei comuni dentro ogni *h*.

Non essendo disponibili al momento della progettazione del campione di secondo stadio informazioni inerenti alla variabile *tipo di sezione* per tutte le sezioni di Censimento, per il suddetto stadio di campionamento si è deciso di stratificare le sezioni secondo le tre modalità corrispondenti ai terzili della distribuzione della popolazione delle sezioni espressa in funzione della dimensione demografica delle sezioni stesse. Tale circostanza, tuttavia, non comporta una perdita di efficienza poiché la dimensione demografica delle sezioni (espressa in termini di individui) costituisce una variabile strettamente correlata con il *tipo di sezione*.

1.4.3 Determinazione della numerosità campionaria e l'allocazione del campione tra gli strati di primo stadio

Per ragioni operative, organizzative e di costo e in base agli errori attesi delle stime del tasso di copertura, si è stabilito che il numero complessivo di comuni campione e la numerosità complessiva di individui campione dovessero essere rispettivamente pari a 255 unità e a circa 366.670 unità (corrispondenti a circa 152.227 famiglie).

Per quanto riguarda la determinazione delle numerosità campionarie per ciascuno dei 100 strati dell'indagine, occorre osservare che l'indagine in esame ha il duplice obiettivo di produrre stime sia nazionali che regionali⁷⁰. In tali circostanze, per allocare il campione occorre trovare un compromesso ragionevole tra le due allocazioni ottimali corrispondenti ai due obiettivi separatamente considerati: da una parte, le stime nazionali e, dall'altra, le stime regionali. In assenza di informazioni aggiornate sulla variabilità del tasso di copertura tra gli strati dell'indagine e non avendo avuto la possibilità di effettuare l'analisi sui dati della precedente indagine di copertura (in quanto non disponibili), non è stato possibile realizzarla mediante l'allocazione ottimale di Neyman. Come noto, sotto l'ipotesi di invariabilità del tasso di copertura o in assenza di informazioni circa la variabilità della variabile d'interesse, la migliore allocazione tra gli strati dell'indagine – finalizzata all'ottenimento della stima a livello nazionale – è quella proporzionale alle dimensioni demografiche degli strati dell'indagine; in ogni modo, tale allocazione comporta

Popolazione 1991) con alcune variabili ottenute sia dal Censimento 1991 che dalle indagini ISTAT sui dati di qualità delle anagrafi comunali dell'anno 1998. Le variabili considerate nello studio sono: (1) la collocazione geografica della sezione; (2) la dimensione demografica del comune; (3) un indicatore di vicinanza del comune alla soglie *critiche* di popolazione; (4) un indicatore della qualità del lavoro del comune in relazione alle rilevazioni statistiche in esso condotte; (5) la tipologia della sezione censuaria (centro, nucleo e case sparse). [...]

La finalità dello studio è stata quella di individuare l'insieme delle variabili tra quelle in esame che sono maggiormente legate al tasso di copertura da utilizzare per la stratificazione dei comuni. A tale scopo sono stati studiati dei modelli che, per ciascuna sezione campione dell'indagine di copertura 1991, mettono in relazione il tasso di copertura con le informazioni sopra elencate. [...]

Dall'analisi dei risultati ottenuti emerge che le variabili utili alla stratificazione sono la popolazione del comune e il tipo di sezione, inoltre risulta essere significativa anche la considerazione congiunta della ripartizione con la classe di popolazione e della ripartizione con il tipo di sezione. [...]

In conclusione, poiché le celle ottenute dall'intersezione delle modalità delle variabili ripartizione e classi di popolazione sono già incluse tra le variabili da utilizzare per la stratificazione, in quanto definiscono gli strati (in questa nota, denominati *strati d'indagine*) che rappresentano i domini principali di interesse per le stime dell'indagine, si è considerata, pertanto, solo la variabile tipo di sezione come ulteriore criterio di stratificazione delle sezioni all'interno di ciascuno strato T_h ($h=1,2,\dots,20$)...».

⁷⁰ In realtà, la PES ha anche l'obiettivo di produrre stime per classi di età e per nazionalità (vedi paragrafo 1.2); tali domini di stima, tuttavia, sono non pianificati, non essendo le loro dimensioni campionarie definite a priori.

coefficienti di variazione delle stime del tasso di copertura sensibilmente diversi tra i differenti domini di studio. Al contrario, al fine di rendere approssimativamente uguali i coefficienti di variazione in oggetto, l'allocazione più opportuna è ovviamente quella uniforme (ossia, assegnazione di un ugual numero di unità campionarie ai vari strati dell'indagine) che, tuttavia, conferisce alle stime in esame livelli di precisione sensibilmente diversi da quelli ottenibili con un'allocazione proporzionale. Per le ragioni sopra esposte, al fine di garantire un certa comparabilità nei livelli di precisione in discussione, senza penalizzare troppo la stima nazionale, è stata adottata un'allocazione di compromesso tra quella uniforme e quella proporzionale, in base alla quale le dimensioni campionarie espresse in termini di individui da intervistare a livello regionale sono state definite come combinazioni lineari convesse delle numerosità relative alle due allocazioni ottimali. In particolare, la numerosità totale del campione n è stata allocata tra le regioni geografiche e province autonome secondo la relazione

$$n_r = \left(0,30 \cdot \frac{n}{21}\right) + \left(0,70 \cdot n \frac{N_r}{N}\right) \quad (7)$$

in cui n_r , N_r indicano rispettivamente il numero di individui campione e quello della popolazione inerenti alla regione o provincia autonoma r , mentre N è la popolazione in termini di individui a livello nazionale. Date le numerosità campionarie per regione, le stesse sono state allocate tra le classi di dimensione demografica dei comuni in base alla relazione

$$n_{r,cp} = \left(0,40 \cdot \frac{n_r}{CP_r}\right) + \left(0,60 \cdot n_r \frac{N_{r,cp}}{N_r}\right) \quad (8)$$

in cui: $n_{r,cp}$ è il numero di individui campione nella classe di dimensione demografica cp della regione geografica r , CP_r indica il numero di classi di dimensione demografica nella regione r , $N_{r,cp}$ denota il numero di individui della popolazione nella classe di dimensione demografica cp della regione geografica r .

1.4.4 Formazione del campione di primo stadio

Tenendo presenti le conclusioni del paragrafo 1.4.2, si è ritenuto opportuno attuare la stratificazione dei comuni appartenenti a ciascuno degli strati d'indagine h ($h = 1, \dots, H$) in base alla sola dimensione demografica dei comuni stessi e nel rispetto delle seguenti condizioni:

- autoponderazione del campione a livello di ciascuno degli strati d'indagine h ;
- suddivisione dei comuni di h nei due insiemi: comuni la cui popolazione è superiore ad una prefissata *soglia* (comuni *Auto Rappresentativi* o AR); restanti comuni (comuni *Non Auto Rappresentativi* o NAR);
- ciascun comune AR costituisce strato a sé;
- i comuni dell'insieme NAR sono suddivisi in strati di ampiezza approssimativamente costante in termini di popolazione ivi residente, dopo averli ordinati secondo una graduatoria decrescente in funzione della popolazione degli stessi;
- selezione di un solo comune campione in ognuno degli strati NAR.

Dunque, ignorando per un momento la stratificazione delle sezioni di Censimento di cui si discuterà nel prosieguo e alla luce di quanto sopra esposto, discende che:

- il disegno di campionamento inerente alla parte AR è del tipo ad uno stadio stratificato in cui le unità di campionamento sono costituite dalle sezioni di Censimento, definibili come grappoli di individui;

ii) il disegno di campionamento relativo alla parte NAR è del tipo a due stadi stratificato in cui le unità primarie sono costituite dai comuni e le unità secondarie dalle sezioni di Censimento. Al fine di conferire maggiore chiarezza ai successivi sviluppi algebrici, conviene introdurre le seguenti notazioni simboliche. Con riferimento al generico strato d'indagine h ($h=1, \dots, H$), si denoti con: l ($l=1, \dots, L_h$), c e j rispettivamente gli indici di strato di comune, comune e sezione⁷¹. Inoltre: N_{hlcj} è il numero di individui residenti nella generica sezione j del comune c appartenente allo strato l ; N_{hlc} ed S_{hlc} indicano rispettivamente il numero di individui residenti e di sezioni nel generico comune c dello strato l ; N_{hl} ed S_{hl} denotano il numero di individui e di sezioni dello strato l , mentre N_h e S_h sono le analoghe quantità marginali riferite a h . Siano ancora m_{hl} il numero di comuni campione nello strato l , n_h il numero pianificato di individui campione e

$$n = \sum_{h=1}^H n_h \quad (9)$$

il numero pianificato di individui campione a livello nazionale.

Tenendo presente quanto sopra premesso, si riassumono ora le linee essenziali della metodologia di formazione del campione di primo stadio.

Determinate le numerosità, n e n_h ($h=1, \dots, H$), in funzione degli errori attesi delle stime del tasso di copertura (si veda la parte concernente la valutazione degli errori nel paragrafo 1.4.5) e in accordo con ragioni operative, organizzative e di costo, la metodologia in questione si articola nei seguenti passi:

- a) per ogni h ($h=1, \dots, H$) si definisce il numero di comuni campione, m_h , in modo da rispettare la condizione che il valore atteso del numero di individui campione relativo a h sia uguale al numero programmato; in simboli

$$\sum_{c \in h} \sum_{j \in c} N_{hcj} \pi_{hcj} = n_h \quad (10)$$

in cui: $\pi_{hcj} = (m_h N_{hc} / N_h) (s_{hc} / S_{hc})$ è la probabilità di inclusione della sezione j appartenente al comune c dello strato h ; $s_{hc} = s_h = \dot{n}_h / (N_h / S_h)$ è il numero minimo di sezioni campione per comune campione⁷², determinato attraverso una procedura di tipo iterativo, facendo variare il numero minimo di individui da intervistare in ciascun comune campione, \dot{n}_h , in modo da rispettare il numero complessivo desiderato di comuni che si intende far partecipare all'indagine; N_h / S_h è il numero medio di individui per sezione riferito ad h ; inoltre, si ha $\pi_{hcj} = \pi_{hc}$ poiché tutte le sezioni del comune c hanno la stessa probabilità di inclusione. Risolvendo l'equazione (10) rispetto all'incognita m_h si ottiene:

$$m_h = \frac{n_h}{s_h} \left(\frac{1}{N_h} \sum_{c \in h} N_{hc} \frac{1}{S_{hc}} \sum_{j \in c} N_{hcj} \right)^{-1} \quad (11)$$

Ci preme sottolineare che è stato volutamente ignorato l'indice di strato l in quanto questo passo è antecedente alla stratificazione dei comuni in base alla dimensione demografica

⁷¹ Adottiamo qui e nel seguito la convenzione che laddove un indice è assente, la quantità deve essere intesa come marginale rispetto a quell'indice.

⁷² L'uguaglianza $s_{hc} = s_h$ discende dalla circostanza che il numero minimo di sezione campione per comune campione è fisso nello strato d'indagine h .

degli stessi ed è effettuato con lo scopo di avere una prima valutazione sul numero di comuni coinvolti nell'indagine;

b) si attua il processo di stratificazione dei comuni

b1) si determina il valore della soglia G_h ($h=1, \dots, H$), mediante la relazione

$G_h = (n_h/m_h)/F_h$, in cui il numeratore esprime il numero minimo di individui campione per comune campione e il denominatore la frazione di campionamento inerente a h , espressa in termini di individui. Vale la pena osservare che – a differenza di quanto avviene solitamente nelle indagini sulle famiglie e gli individui (basate su disegni di campionamento a due stadi: comuni, famiglie), in cui il numeratore della soglia è dato dal prodotto del numero minimo di unità campione di secondo stadio per ciascuna unità campione di primo stadio moltiplicato per il numero medio di unità finali (individui) per unità di secondo stadio – nell'indagine qui in discussione non si è potuto procedere nello stesso modo a causa dell'elevata variabilità del numero medio di individui per sezione di Censimento;

b2) si ordinano, in modo decrescente, i comuni di ogni h secondo la dimensione demografica. I comuni con dimensione superiore o uguale a G_h e i comuni metropolitani sono classificati come comuni AR; i restanti, come comuni NAR;

b3) i comuni di tipo NAR sono suddivisi in L_h strati, di dimensione approssimativamente uguale al prodotto $m_h \times G_h$;

c) si seleziona un solo comune campione ($m_{hl}=1$) da ciascuno degli strati⁷³ NAR con probabilità $z_{hlc} = N_{hlc}/N_{hl}$;

d) per ognuno dei comuni selezionati si definisce il numero atteso di individui campione mediante la relazione⁷⁴ $n_{hlc} = F_h N_{hl}/m_h$ per i comuni NAR e $n_{hl} = F_h N_{hl}$ per i comuni AR, basata sulla condizione di autoponderazione:

$$\frac{n_{hl}}{N_{hl}} = \frac{n_h}{N_h} = F_h .$$

Per la stratificazione dei comuni e per la selezione dei comuni campione con probabilità variabili sono state utilizzate le informazioni più recenti sulla popolazione residente dei comuni, riferite alla data del 31 dicembre 2010. Inoltre, è importante aggiungere che al momento della progettazione del campione di primo stadio non erano disponibili informazioni inerenti alle nuove sezioni di Censimento, la cui definizione era in corso di ultimazione. Tale circostanza ha comportato nello studio del campione di primo stadio (in termini di stratificazione dei comuni e selezione dei comuni campione) la necessità di ipotizzare per il secondo stadio la selezione di un campione casuale semplice di individui (definito in termini di numerosità attesa di individui), anziché un campione casuale di sezioni. Tuttavia, come illustrato nel paragrafo 1.4.1, il disegno effettivo prevede, al secondo stadio, la selezione di un campione di sezioni, all'interno delle quali il numero di individui da enumerare presenta la natura di variabile casuale. Pertanto, anche se il numero di sezioni campione è stato successivamente definito sotto il vincolo che il numero effettivo di individui enumerati fosse il più possibile prossimo a quello definito nella fase di definizione del disegno di primo stadio, il numero effettivo di individui campione e quello atteso possono risultare differenti.

⁷³ Per l'estrazione delle UP si può ricorrere alla procedura di selezione sistematica, suggerita da Madow (1949) e Murthy (1967), che presenta le seguenti caratteristiche: (i) assegna una probabilità di inclusione espressa nella forma $\pi_{hlc} = m_h z_{hlc}$; (ii) la sua implementazione è estremamente semplice; (iii) permette di ottenere stime generalmente più efficienti rispetto a quelle che si ottengono con altre procedure di selezione (Cicchitelli *et al.*, 1992; Fabbris, 1991).

⁷⁴ Negli strati AR si ha m_h e $z_{hlc} = 1$.

1.4.5 Valutazione degli errori relativi attesi

Successivamente all'estrazione del campione di primo stadio, descritta in dettaglio nel precedente paragrafo 1.4.4, è stata effettuata una valutazione degli errori relativi per i diversi domini di stima relativamente al tasso di copertura. A riguardo, ci preme sottolineare che la variabilità della suddetta stima (come anche quella della stima del totale della popolazione) ha una componente dovuta al modello e una componente dovuta al disegno⁷⁵. In questo documento l'attenzione è stata limitata alla seconda componente, che ha rilievo nel disegno di campionamento. Inoltre, occorre osservare che si tratta di una valutazione approssimativa dei coefficienti di variazione in quanto: (i) le stime del tasso di copertura non sono quelle dell'indagine in esame, ma quelle comunali basate su dati 2001 elaborati nell'ambito di una convenzione tra l'Istat e il Dipartimento di Scienze Statistiche dell'Università di Bologna; (ii) la variabilità considerata è quella tra le unità di primo stadio e quella dentro ogni comune estratto nell'ambito del quale si è ipotizzato un campionamento casuale semplice di individui, non essendo stato possibile al momento della valutazione in oggetto fare riferimento ad un campionamento di secondo stadio concernente le sezioni di Censimento. Si indichi con d ($d=1, \dots, D$) il generico dominio di stima e con $\hat{\tau}_d$ la stima del tasso di copertura inerente al suddetto dominio; per ogni dominio di stima, una approssimazione della stima della varianza di $\hat{\tau}_d$ può essere fornita dall'espressione:

$$\widehat{VAR}(\hat{\tau}_d) = \frac{1}{N_d^2} \left[\widehat{VAR}({}_{AR}\tilde{N}_d^{Cen}) + \widehat{VAR}({}_{NAR}\tilde{N}_d^{Cen}) \right] \quad (12)$$

in cui:

$$\begin{aligned} \blacktriangleright \hat{\tau}_d &= \frac{\tilde{N}_d^{Cen}}{N_d} = \frac{{}_{AR}\tilde{N}_d^{Cen} + {}_{NAR}\tilde{N}_d^{Cen}}{N_d} = \\ &= \frac{1}{\sum_{h=1}^{H_d} \sum_{l=1}^{L_h} \sum_{c=1}^{M_{hl}} N_{hlc}} \left[\sum_{h=1}^{H_d} \left[\sum_{l=1}^{AR L_h} \tilde{\tau}_{hl}^{Cen} N_{hl} + \sum_{l=1}^{NAR L_h} \tilde{\tau}_{hlc}^{Cen} \frac{N_{hlc}}{z_{hlc}} \right] \right] \end{aligned}$$

è la stima del tasso di copertura riferita al dominio d

$$\begin{aligned} \blacktriangleright \widehat{VAR}({}_{AR}\tilde{N}_d^{Cen}) &= \sum_{h=1}^{H_d} \sum_{l=1}^{AR L_h} N_{hl}^2 \frac{N_{hl} - n_{hl}}{N_{hl} - 1} \frac{\tilde{\tau}_{hl}^{Cen} (1 - \tilde{\tau}_{hl}^{Cen})}{n_{hl}} \\ \blacktriangleright \widehat{VAR}({}_{NAR}\tilde{N}_d^{Cen}) &= \sum_{t=1}^{T_d} \frac{{}_{NAR}L_t}{{}_{NAR}L_t - 1} \sum_{l=1}^{NAR L_t} \left(\tilde{N}_{tl}^{Cen} - \frac{\tilde{N}_t^{Cen}}{{}_{NAR}L_t} \right)^2 + \\ &+ \sum_{h=1}^{H_d} \sum_{l=1}^{NAR L_h} \frac{N_{hlc}^2}{z_{hlc}} \frac{N_{hlc} - n_{hlc}}{N_{hlc} - 1} \frac{\tilde{\tau}_{hlc}^{Cen} (1 - \tilde{\tau}_{hlc}^{Cen})}{n_{hlc}} \end{aligned}$$

- ▶ t è l'indice di superstrato ($t=1, \dots, T_d$)
- ▶ T_d è il numero di superstrati t nel dominio d
- ▶ N_d denota il numero di individui della popolazione nel dominio d
- ▶ \tilde{N}_d^{Cens} , ${}_{AR}\tilde{N}_d^{Cens}$ e ${}_{NAR}\tilde{N}_d^{Cens}$ indicano rispettivamente la stima del numero di individui coperti al Censimento nel dominio d , nei comuni AR e NAR appartenenti al dominio d
- ▶ H_d è il numero di strati dell'indagine, h , appartenenti al dominio d

⁷⁵ Le espressioni analitiche degli stimatori del tasso di copertura e del totale della popolazione a cui ci si riferisce sono date dalle formule (25) e (26) riportate nel paragrafo 1.5.4.

- ▶ ${}_{AR}L_h$ ed ${}_{NAR}L_h$ sono rispettivamente il numero di strati AR e di strati NAR nello strato dell'indagine h
- ▶ ${}_{NAR}L_t$ denota il numero di strati NAR nel superstrato t ($t=1, \dots, T_d$)
- ▶ M_{hl} è il numero di comuni nello strato l appartenente allo strato dell'indagine h
- ▶ $\tilde{\tau}_{hlc}^{Cen}$ è il tasso di copertura stimato nel comune c appartenente allo strato l inerente allo strato dell'indagine h (nei comuni AR si ha l'identità $\tilde{\tau}_{hlc}^{Cen} = \tilde{\tau}_{hl}^{Cen}$)
- ▶ \tilde{N}_{hl}^{Cen} e \tilde{N}_t^{Cen} indicano la stima del numero di individui coperti al Censimento rispettivamente nello strato l appartenente al superstrato t e nel superstrato t .

È importante, altresì, osservare che il disegno di campionamento in esame – essendo basato sulla selezione di un solo comune per strato – se da una parte introduce un guadagno di efficienza nelle stime dell'indagine, dall'altra non consente di stimare correttamente la varianza delle stime, poiché sono nulle le probabilità di inclusione del secondo ordine all'interno degli strati. Per tale ragione, ai fini del calcolo della (12), relativamente alla stima della varianza di ${}_{NAR}\tilde{N}_d^{Cen}$, è stato necessario ricorrere alla tecnica di *collassamento degli strati* che consiste nel raggruppare gli ${}_{NAR}L$ (${}_{NAR}L = {}_{NAR}L_h \cdot H$) strati in T_d ($T_d < {}_{NAR}L$) superstrati, ciascuno dei quali contenente ${}_{NAR}L_t$ ($t=1, \dots, T$) strati (in cui T indica il numero totale di superstrati).

Gli errori relativi sono stati quindi calcolati secondo la ben nota formula:

$$CV(\hat{\tau}_d) = \frac{\sqrt{VAR(\hat{\tau}_d)}}{\hat{\tau}_d}.$$

In Appendice sono riportati per ogni dominio di stima le numerosità campionarie attese in termini di comuni, di individui e di sezioni di Censimento, i valori delle stime dei tassi di copertura $\hat{\tau}_d$ ottenuti utilizzando i tassi di copertura stimati a livello comunale su dati 2001 (nell'ambito della già citata convenzione tra l'Istat e il Dipartimento di Scienze Statistiche dell'Università di Bologna) ed i coefficienti di variazione inerenti alle stime dei tassi in oggetto.

Si precisa, infine, che le numerosità attese del campione in termini di individui sono calcolate come somma dei n_{hl} (nei comuni AR) e dei n_{hlc} (nei comuni NAR), la cui derivazione è descritta nel paragrafo 1.4.4, mentre il numero atteso di sezioni campione per comune è calcolato dividendo le suddette quantità per il numero medio di individui per sezione a livello comunale ottenuto dividendo la popolazione residente del comune al 31 dicembre 2010 per il numero di sezioni per comune desunto dal geo-database creato da Basi Territoriali aggiornato al 28 settembre 2011.

1.4.6 Formazione del campione di secondo stadio

Il secondo stadio di campionamento è costituito dalle sezioni di Censimento interpretabili come grappoli di individui.

Per la formazione del campione di sezioni si è ritenuto conveniente stratificare la popolazione delle sezioni in base alla dimensione demografica delle stesse, secondo le tre modalità definite dai terzili della distribuzione della popolazione in oggetto.

Per ciò che riguarda l'allocazione della numerosità campionaria comunale nei suddetti strati, in assenza di informazioni aggiornate ed affidabili sulle varianze di strato, si è deciso di effettuarla in modo proporzionale alla dimensione demografica dello strato nel comune. Se q indica la popolazione delle sezioni suddivisa in terzili ($q=1,2,3$; $q=1$: 1° terzile, $q=2$: 2° terzile, $q=3$: 3° terzile), in simboli si può scrivere

$$n_{hlq} = n_{hl} \frac{N_{hlq}}{N_{hl}} \quad \text{per i comuni AR} \quad (13)$$

$$n_{hlcq} = n_{hlc} \frac{N_{hlcq}}{N_{hlc}} \quad \text{per i comuni NAR}$$

in cui, con riferimento al generico strato d'indagine h ($h=1, \dots, H$): N_{hlq} e n_{hlq} rappresentano rispettivamente il numero di individui della popolazione e del campione appartenenti allo strato q dello strato l ; N_{hlcq} e n_{hlcq} denotano rispettivamente il numero di individui della popolazione e del campione appartenenti allo strato q del comune c nello strato l ; n_{hl} e n_{hlc} , infine, indicano rispettivamente il numero di individui campione nello strato l e nel comune c nello strato l .

Effettuata l'allocazione del campione tra gli strati, le unità di secondo stadio sono state selezionate con probabilità uguali e senza ripetizione. Con riferimento al generico strato l nello strato d'indagine h , si denoti⁷⁶ con: S_{hlcq} e s_{hlcq} rispettivamente il numero di sezioni nella popolazione e nel campione appartenenti allo strato q nel comune campione c ; S_{hlq} e s_{hlq} rispettivamente il numero di sezioni nella popolazione e nel campione appartenenti allo strato q . La probabilità di inclusione della generica sezione j è dunque data da:

$$\pi_{hlqj} = S_{hlq} \frac{1}{S_{hlq}} \quad \text{per i comuni AR} \quad (14)$$

$$\pi_{qj|hlc} = S_{hlcq} \frac{1}{S_{hlcq}} \quad \text{per i comuni NAR.}$$

Per i comuni NAR la probabilità di inclusione della sezione j dello strato q appartenente al comune c è condizionata alla circostanza che il comune c sia stato incluso nel campione di primo stadio.

Per ogni strato q definito dentro ciascun comune campione si determina il numero di sezioni di Censimento da selezionare al secondo stadio imponendo la condizione di autoponderazione a livello di strato dell'indagine h

$$\pi_{hlqj} = F_h \quad \text{per i comuni AR}$$

$$\pi_{hlcqj} = F_h \quad \text{per i comuni NAR}$$

in cui π_{hlqj} e π_{hlcqj} indicano le probabilità d'inclusione della sezione j rispettivamente nei comuni AR e NAR.

Sostituendo π_{hlqj} , π_{hlcqj} e F_h con le rispettive espressioni, si ottiene

$$\frac{S_{hlq}}{S_{hlq}} = \frac{n_h}{N_h} \quad \text{per i comuni AR} \quad (15)$$

$$m_h \frac{N_{hlc}}{N_{hl}} \frac{S_{hlcq}}{S_{hlcq}} = \frac{n_h}{N_h} \quad \text{per i comuni NAR.}$$

Risolviendo le (15) rispetto alle incognite s_{hlq} e s_{hlcq} si ricava

⁷⁶ Si ricorda che, al fine di non appesantire la notazione introdotta, si adotta la convenzione che laddove un indice è assente la quantità deve essere intesa come marginale rispetto allo stesso indice.

$$s_{hlq} = \frac{n_{hlq}}{(N_{hlq}/S_{hlq})} \quad \text{per i comuni AR} \quad (16)$$

$$s_{hlcq} = \frac{n_{hlcq}}{(N_{hlcq}/S_{hlcq})} \quad \text{per i comuni NAR.}$$

La probabilità di inclusione finale, ossia comprensiva della probabilità di inclusione del comune c al quale afferisce la sezione di Censimento j , è dunque uguale alla prima delle espressioni (14) per le sezioni appartenenti ai comuni AR, mentre è pari a

$$\pi_{hlcqj} = \pi_{hlc} \pi_{qj|hlc} = m_h \frac{N_{hlc}}{N_{hl}} \frac{s_{hlcq}}{S_{hlcq}} \quad (17)$$

per le sezioni selezionate nei comuni NAR.

Nell'Appendice 2A sono illustrati per ogni dominio di stima le numerosità campionarie teoriche e realizzate in termini di comuni, di individui e di sezioni di Censimento.

1.5 Procedura di stima

1.5.1 Caratteristiche generali

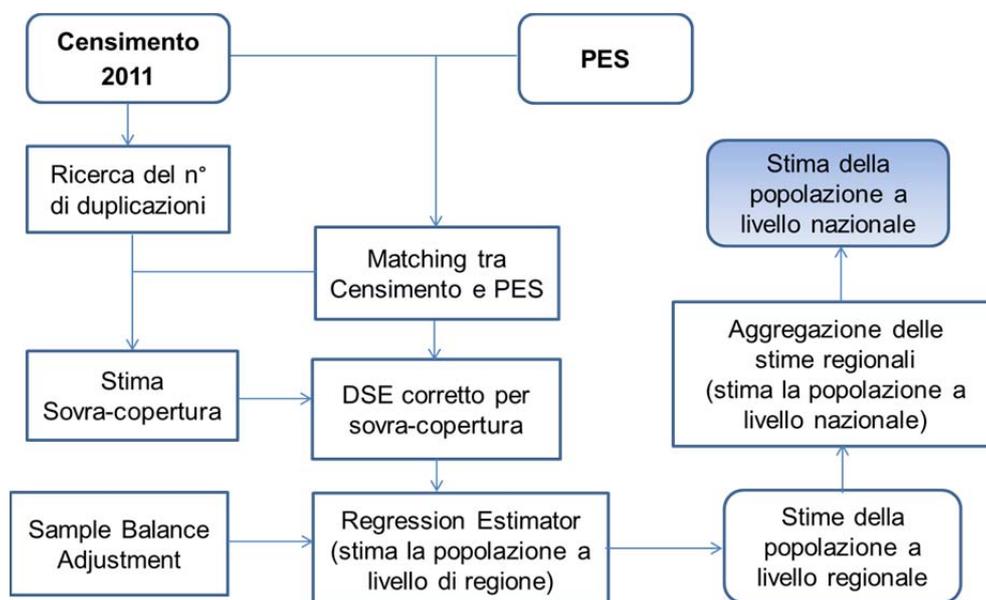
L'obiettivo della procedura di stima in oggetto non è l'ottenimento di una misura dell'errore di copertura attraverso la quantificazione della sotto-copertura e della sovra-copertura, ma è quello di produrre degli aggiustamenti netti: (i) riducendo le stime Dual-System mediante una stima della sovra-copertura, (ii) imputando il minor numero di individui, (iii) non rimuovendo i duplicati.

Essa si articola nelle fasi seguenti:

- a) costruzione di un sistema di pesi campionari finali e loro assegnazione agli individui rispondenti per l'ottenimento delle stime necessarie ai fini della determinazione delle stime da modello dei tassi di copertura e sotto-copertura;
- b) definizione di una misura della propensione alla sovra-copertura – basata sui conteggi censuari corretti e sui conteggi censuari errati – e studio dell'espressione per il calcolo di una stima della suddetta misura;
- c) stima del Dual-System Estimator (DSE) corretto per sovra-copertura, distintamente secondo classi di età e separatamente per i due sessi, per ciascun comune campione selezionato in ognuno degli strati HtC, dentro ciascuna regione geografica;
- d) calcolo di una stima della popolazione a livello di regione geografica, distintamente per classi di età, sesso e modalità dell'indice HtC, attraverso l'utilizzo di uno stimatore del tipo rapporto;
- e) allo scopo di conferire alle stime di cui al precedente punto d) un migliore livello di attendibilità, applicazione del procedimento Sample Balance Adjustment (SBA) in tutti gli strati HtC nei quali i campioni in essi selezionati risultino essere outlier;
- f) infine, utilizzando le stime prodotte fino a questo punto, si possono immediatamente ottenere le stime dei totali *veri* (incogniti) delle popolazioni regionali e nazionale.

I risultati ottenuti applicando la presente procedura di stima, ossia le stime dei tassi di copertura distintamente in ogni dominio di stima, sono descritti nel capitolo quarto.

Figura 1.1 - Flowchart del processo di stima



1.5.2 Determinazione dei pesi campionari

In questo paragrafo è descritta la procedura utilizzata per il calcolo dei pesi campionari da attribuire a ciascun individuo rispondente; essa è articolata in passi successivi.

Inizialmente è stato determinato il peso diretto teorico di ogni sezione di Censimento campione come reciproco della probabilità di inclusione espressa dalla prima delle espressioni (14) per le sezioni appartenenti ai comuni AR, mentre è data dalla (17) per le sezioni afferenti ai comuni NAR; in simboli⁷⁷: $d_{hlcqj} = 1/\pi_{hlcqj}$. Successivamente, poiché in fase di rilevazione si è verificata una mancata risposta di primo stadio (ossia, alcuni comuni campione non hanno partecipato all'indagine), è stato necessario correggere i pesi diretti teorici assegnati alle sezioni di Censimento per tenere conto di tale circostanza. A tal fine, è stato calcolato un fattore correttivo espresso dal rapporto tra il numero m_h di comuni campione dello strato h e il corrispondente numero \bar{m}_h di comuni che hanno partecipato all'indagine. Tale fattore è stato poi moltiplicato per il peso diretto teorico d_{hlcqj} , ottenendo così il peso campionario corretto per mancata risposta relativo alla sezione j

$$w_{hlcqj} = d_{hlcqj} \frac{m_h}{\bar{m}_h} . \quad (18)$$

Naturalmente, ad ogni individuo rilevato nella sezione di Censimento campione j compete un peso uguale a quello attribuito alla sezione j .

Il peso finale definito dalla (18) è stato utilizzato per ottenere tutte le stime campionarie necessarie ai fini della determinazione delle stime da modello dei tassi di copertura e sotto-copertura che verranno illustrate nel prosieguo.

⁷⁷ Al fine di non appesantire la comprensione dei successivi sviluppi algebrici, nel prosieguo il pedice relativo alla sezione j sarà sempre indicato con $hlcqj$ sia che la sezione j appartenga ad un comune AR che, invece, ricada in un comune NAR.

1.5.3 Modello di stima dell'errore di sovra-copertura

Le indagini di qualità sono state condotte dall'Istat in connessione degli ultimi quattro Censimenti della popolazione e delle abitazioni (1981, 1991, 2001 e 2011), ma solo nel 2001, per la prima volta, sono state prodotte stime della sovra-copertura. Esse erano solo indicative del fenomeno e non erano considerate molto affidabili, in quanto ottenute dai risultati dei questionari autocompilati in occasione dell'indagine di copertura e non tramite il *record linkage* (Istat, 2009). Da questa tornata censuaria, il regolamento comunitario impone una più precisa e dettagliata stima della popolazione ed include nei vincoli anche una misurazione della sovra-copertura⁷⁸.

L'errore di sovra-copertura, di segno opposto a quello di sotto-copertura, è solitamente considerato di minore entità. Dagli studi effettuati dai principali Istituti di Statistica a livello internazionale, infatti, si evince che l'errore di sovra-copertura è in un rapporto di uno a tre rispetto a quello di sotto-copertura (in alcuni Paesi, tale rapporto è risultato anche di uno a cinque).

Il fenomeno della sovra-copertura trae origine da un insieme di errori consistenti in duplicazioni, conteggi nel posto sbagliato ed enumerazioni errate; al riguardo, nel paragrafo 1.3 è stata illustrata una classificazione che suddivide gli errori di sovra-copertura che possono verificarsi al Censimento in quattro tipologie.

Tuttavia, oltre a questa suddivisione, si può altresì differenziare la sovra-copertura a seconda che essa si manifesti a livello nazionale o locale.

La sovra-copertura nazionale (solitamente detta *national list inflation*, nella terminologia anglosassone) si riferisce alla circostanza in cui nei dati provenienti da fonti amministrative, relativi a una determinata area territoriale, risultino più individui di quanti ne risiedono realmente in quell'area; si tratta, pertanto, di record errati generati da individui che sono presenti nella fonte dei dati ma che non dovrebbero essere inclusi nella popolazione d'interesse. Esempi di questa forma di sovra-copertura sono: un bambino che è stato registrato nel questionario censuario di entrambi i genitori separati, un individuo emigrato prima della data di riferimento del Censimento, un individuo deceduto ma il cui decesso non sia ancora stato registrato, ecc. Come si vedrà nel successivo paragrafo 1.5.4, la sovra-copertura nazionale è considerata trascurabile come conseguenza della verifica di una delle ipotesi del modello *dual system* scelto ai fini della stima dei parametri d'interesse per l'indagine in oggetto.

La sovra-copertura locale, invece, è generata da individui rilevati dal Censimento in un indirizzo ma che in realtà risiedono altrove; cioè, individui conteggiati nel luogo sbagliato. Nell'ipotesi che non ci siano record duplicati, ossia quando uno stesso individuo viene rilevato più di una volta nell'ambito del Censimento, questa forma di sovra-copertura bilancia la sotto-copertura locale poiché un individuo enumerato erroneamente in un indirizzo, per definizione, non è conteggiato nella sua corretta localizzazione. Ai fini della stima del corretto ammontare della popolazione, occorre precisare che il bilanciamento tra i due tipi di errore (sovra-copertura locale e sotto-copertura locale) avviene se essi si verificano nello stesso dominio di stima rispetto al quale è calcolata la DSE; in altre parole, se la DSE è calcolata a livello di sezione di Censimento e l'indirizzo in cui l'individuo viene erroneamente conteggiato appartiene alla stessa sezione in cui ricade l'indirizzo dove di fatto risiede, ma nel quale egli non viene rilevato dal Censimento, allora i due errori si bilanciano. Una considerazione che scaturisce da quanto appena illustrato è che la sotto-copertura nazionale è, quindi, generata da individui completamente sfuggiti al processo di rilevazione, ma che dovrebbero essere inclusi nella popolazione.

Nel caso in cui l'ipotesi su menzionata di assenza di record duplicati non si ritenga più valida, l'individuo è rilevato dal Censimento sia all'indirizzo corretto che a quello errato; in tale circostanza,

⁷⁸ La PES è normata dalla Commissione Europea (Commission Regulation n. 1151/2010).

si verifica una sovra-copertura dell'individuo nel dominio di stima in cui ricade l'indirizzo in cui egli è erroneamente rilevato (pur non essendosi verificata una sotto-copertura in alcun dominio di stima). Le innovazioni di metodo e tecniche introdotte con il Censimento del 2011, relative all'uso di liste pre-censuarie e alla pluralità dei canali di consegna⁷⁹, compilazione e restituzione dei questionari, hanno accresciuto rispetto al passato il rischio di duplicazione.

Nel paragrafo 1.5.1 si è visto che l'approccio adottato per trattare la sovra-copertura è di stimare l'ordine di grandezza del fenomeno e quindi di includere tale valutazione in un aggiustamento netto per la copertura. I tipi di errore di sovra-copertura considerati in fase di stima sono il 2 e il 3.

La metodologia utilizzata per la stima della sovra-copertura nella PES 2011 è costituita dalle tre seguenti fasi:

1. stima del numero di duplicazioni attraverso la PES
2. stima dei conteggi nel posto sbagliato mediante la PES
3. calibrazione delle stime di cui al punto 1), utilizzando il numero di duplicazioni rilevate al Censimento, al fine di migliorare l'accuratezza delle stime di sovra-copertura (il campione PES non è disegnato per ottenere stime del numero di duplicazioni).

Il processo di stima sopra riassunto fornisce dei pesi di sovra-copertura per classi di età, ripartizione geografica e indice HtC. Tali pesi sono successivamente applicati ai record individuali del Censimento, la cui somma costituisce il totale censuario coinvolto nel DSE, così da ridurre il contributo di ciascuno di essi all'ammontare di popolazione risultante dalle operazioni di rilevazione censuaria e, pertanto, il totale stesso.

Per quanto sopra esposto, appare evidente che una assunzione chiave del procedimento di stima della sovra-copertura è la definizione della corretta localizzazione dell'individuo alla data di riferimento del Censimento. Nell'indagine in discussione si è assunto che:

“la PES definisce la corretta localizzazione dell'individuo alla data del 9 ottobre 2011 attraverso la risposta dell'intervistato alla domanda 1.5 del questionario PES”

ossia, l'esatta localizzazione dell'individuo alla data di riferimento del Censimento (9 ottobre 2011) non è data dal luogo presente nel file del ritorno censuario, né dove è stato intervistato dal rilevatore della PES, ma è il luogo dichiarato dall'individuo in fase di compilazione del questionario della PES in risposta alla domanda “Dove dimorava alla data del Censimento (9 ottobre 2011)?” presente nel questionario medesimo.

Il parametro oggetto di stima è la *propensione alla sovra-copertura*: essa è definita come l'attitudine dell'individuo ad essere sovra enumerato; dunque, è una caratteristica insita nell'individuo. Da studi effettuati dall'*Office for National Statistics* (ONS)⁸⁰ si evince che i tassi di sovra-copertura tendono a variare tra gruppi di popolazione definiti dall'incrocio delle modalità di alcune variabili, quali, ad esempio, il luogo in cui risiede l'individuo, l'età e l'essere o meno studente. Infatti, per ciò che concerne l'età, risulta che la mobilità nella residenza è concentrata in gruppi di età giovane e di conseguenza i tassi di sovra-copertura nella fascia d'età 18-25 anni sono generalmente più elevati. Anche la condizione di studente inflaziona i tassi di sovra-copertura in conseguenza del fatto che gli studenti presentano generalmente una probabilità elevata di essere conteggiati sia nella residenza di origine (ossia, quella in cui vivevano con i genitori) sia nel luogo dove si sono trasferiti per effettuare gli studi (causa principale che determina gli errori di sovra-copertura di tipo 2, ossia le duplicazioni). Il sesso, invece, non sembra essere determinante ai fini della sovra-copertura. Dal figura 1.2 (Brown *et al.*, 2009), che visualizza i tassi di sovra-copertura

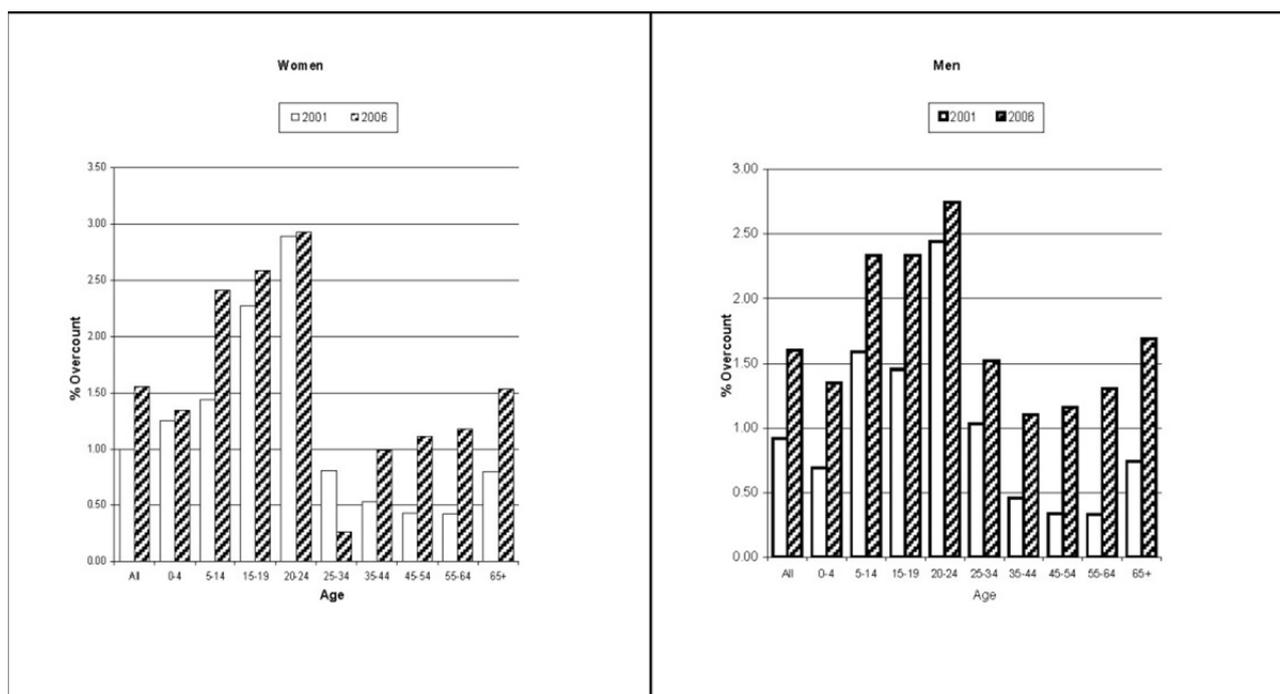
⁷⁹ In occasione del XV Censimento della popolazione e delle abitazioni, la percentuale di persone che hanno inviato il questionario compilato via web è stata elevata.

⁸⁰ Large *et al.*, 2011; Brown *et al.*, 2009.

per età e sesso nel Censimento canadese del 2001 e 2006, emerge che il tasso di sovra-copertura totale è pari a 0,96 per cento nel 2001 e 1,50 per cento nel 2006; inoltre, esso varia considerevolmente per età (la fascia d'età 20-24 è quella che presenta il valore più elevato del suddetto tasso), mentre è pressoché uguale tra i due sessi.

Ciò mette in luce la necessità di stimare la sovra-copertura identificando come domini di stima i gruppi socio-demografici che hanno una maggiore propensione ad essere sovra enumerati.

Figura 1.2 - Tassi di sovra-copertura per età e sesso. Canadian Census 2001 e 2006



Nella definizione dell'ambito geografico e delle classi d'età a cui riferire la stima della sovra-copertura si è tenuto in considerazione che, a differenza della sotto-copertura, la sovra-copertura non è un fenomeno diffuso; quindi, per poterne calcolare un stima statisticamente attendibile, il dominio di stima deve essere ampio.

Non essendo disponibile la variabile studente al momento della determinazione della suddetta propensione, si è pertanto deciso di calcolarla secondo:

- ▶ le ripartizioni geografiche (Nord-Ovest, Nord-Est, Centro, Sud, Isole);
- ▶ le modalità dell'indice *Hard to Count* (HtC) ($p = 1$ 'enumerazione facile', $p = 2$ 'enumerazione di media difficoltà', $p = 3$ 'enumerazione difficile');
- ▶ le età in classi (0-2 e 25-60 anni, 3-17 anni, 18-24 anni, 61 e più anni).

Introduciamo ora le seguenti notazioni simboliche utili per una migliore comprensione degli sviluppi algebrici che seguono. Si indichi con:

- i il generico individuo
- g, b due generiche tra le modalità congiunte della distribuzione doppia: {ripartizione geografica, indice HtC}, con g che può coincidere con b
- a_w è l'età in classi
- k la generica sezione di Censimento appartenente a b

s_g, s_b	il numero di sezioni campione rispettivamente in g e in b
S_g, S_b	il numero di sezioni della popolazione rispettivamente in g e in b
w_{jg}, w_{kb}	i pesi campionari assegnati rispettivamente alla sezione di Censimento j appartenente a g e alla sezione di Censimento k appartenente a b
$c_{a_w, i, jg}$	una variabile indicatrice che assume il valore 1 se l'individuo i , appartenente alla classe d'età a_w , è correttamente conteggiato dal Censimento nella sezione j dell'area g e 0 altrimenti
$o_{a_w, i, kb, jg}$	una variabile indicatrice che assume il valore 1 se l'individuo i , appartenente alla classe d'età a_w , è conteggiato erroneamente dal Censimento nella sezione j dell'area g poiché la sua corretta enumerazione è la sezione k dell'area b (con $k \neq j$ e b che può anche coincidere con g) e 0 altrimenti.

Una misura della propensione alla sovra-copertura è fornita dal rapporto

$$\gamma_{a_w, g} = \frac{\text{conteggi censuari totali in } a_w \text{ e } g}{\text{numero di individui univocamente e correttamente conteggiati dal Censimento in } a_w \text{ e } g} =$$

$$= \frac{\text{conteggi censuari corretti in } a_w \text{ e } g + \text{conteggi censuari errati in } a_w \text{ e } g}{\text{conteggi censuari corretti in } a_w \text{ e } g},$$

o, in simboli,

$$\gamma_{a_w, g} = \frac{X_{a_w, g}}{Y_{a_w, g}} = \frac{Y_{a_w, g} + E_{a_w, g}}{Y_{a_w, g}} \quad (19)$$

in cui, con riferimento alla classe d'età a_w e all'area g :

$$Y_{a_w, g} = \sum_{j \in S_g} \sum_{i \in j} c_{a_w, i, jg}$$

rappresenta il numero di conteggi esatti effettuati al Censimento,

$$E_{a_w, g} = \sum_{j \in S_g} \sum_b \sum_{\substack{k \in S_b \\ k \neq j}} \sum_{i \in k} o_{a_w, i, kb, jg}$$

indica il numero di conteggi errati (sovra-copertura) occorsi al Censimento,

$$X_{a_w, g} = \sum_{j \in S_g} \sum_{i \in j} c_{a_w, i, jg} + \sum_{j \in S_g} \sum_b \sum_{\substack{k \in S_b \\ k \neq j}} \sum_{i \in k} o_{a_w, i, kb, jg}$$

denota il numero totale di conteggi censuari.

L'inverso della propensione alla sovra-copertura è detto *fattore di sovra-copertura*

$$\frac{1}{\gamma_{a_w, g}}$$

e fornisce una misura di quanto si riduce il contributo di ogni individuo alla stima *Dual-System*, poiché agisce sul totale censuario e ne diminuisce il valore.

Chiaramente, se $E_{a_w, g} = 0$, si avrà $\gamma_{a_w, g} = 1$ e, quindi, l'individuo conteggiato al Censimento contribuisce con tutto il suo peso alla DSE; viceversa, quanto più elevato è $E_{a_w, g}$, tanto più $\gamma_{a_w, g}$

risulterà maggiore di uno, e tanto minore sarà il contributo dell'individuo alla stima della dimensione della popolazione.

Si denoti ora con

$$\hat{D}_{jg} = \sum_b \sum_{\substack{k \in S_b \\ k \neq j}} w_{kb} \sum_{i \in k} o_{i, kb, jg} c_{i, jg} \quad (20)$$

la stima del numero di ritorni errati del Censimento attribuiti alla sezione j laddove il ritorno esatto esiste nella PES nella sezione k ; l'espressione

$$\frac{\tilde{P}}{\hat{D}} = \frac{\tilde{P}}{\sum_g \sum_{j \in S_g} w_{jg} \hat{D}_{jg}} \quad (21)$$

rappresenta la stima del *tasso di duplicazione*, dato, pertanto, dal rapporto tra il numero delle duplicazioni individuate nei conteggi censuari, \tilde{P} , e la stima PES dello stesso.

Una stima della propensione alla sovra-copertura è fornita da

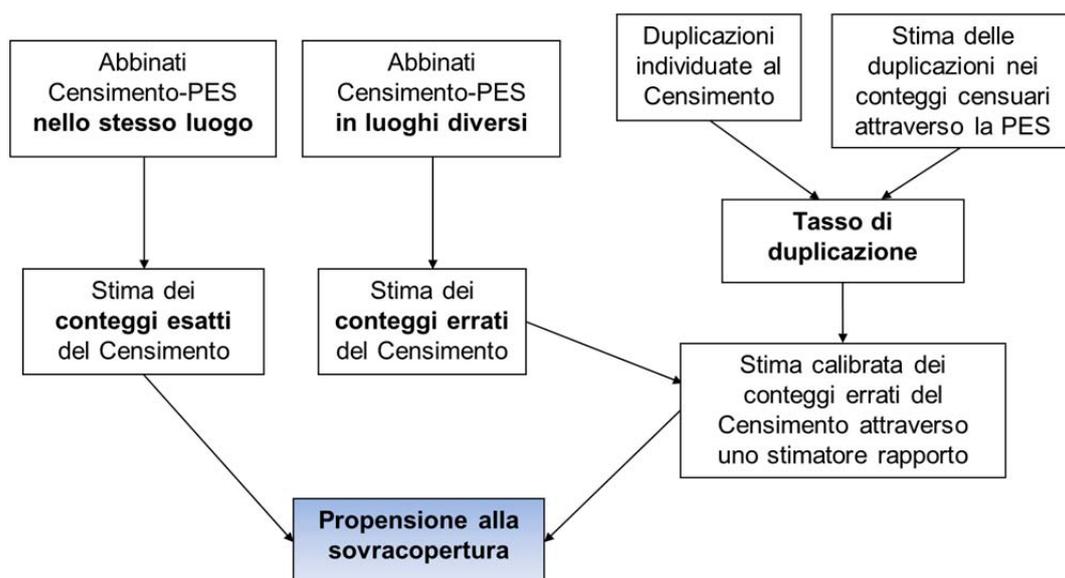
$$\hat{\gamma}_{a_w, g} = \frac{\hat{Y}_{a_w, g} + \hat{E}_{a_w, g}}{\hat{Y}_{a_w, g}} \quad (22)$$

dove:

$$\hat{Y}_{a_w, g} = \sum_{j \in S_g} w_{jg} \sum_{i \in j} c_{a_w, i, jg} \quad (23)$$

$$\hat{E}_{a_w, g} = \sum_{j \in S_g} \frac{\tilde{P}}{\hat{D}} \sum_b \sum_{\substack{k \in S_b \\ k \neq j}} w_{kb} \sum_{i \in k} o_{a_w, i, kb, jg} \cdot \quad (24)$$

Figura 1.3 - Flowchart del processo di stima della sovra-copertura



1.5.4 Dual-System Estimator corretto per sovra-copertura

Nel paragrafo 1.1 è stato evidenziato che, nel caso in cui la dimensione effettiva della popolazione è incognita, per determinare i tassi di copertura e di sotto-copertura, è necessario ricorrere ad un

metodo di stima basato sull'impiego di particolari modelli probabilistici in grado di formalizzare la mancata enumerazione delle unità che può verificarsi al Censimento e/o all'indagine di copertura. Il modello in esame rappresenta una variante del *modello di Petersen* o di *omogeneità entro le liste* (che, opportunamente adattato, costituisce una ragionevole approssimazione della realtà), in cui il totale censuario è corretto per tener conto della sovra-copertura; inoltre, al fine di migliorarlo, si è ritenuto opportuno introdurre in esso la correzione di *Chapman*, per applicazioni del DSE a piccole popolazioni.

Di seguito si riporta una descrizione del *Modello di Petersen*, che costituisce, pertanto, il punto di partenza del processo di stima.

Per illustrare brevemente il suddetto modello, si suppone che l'indagine di copertura ripeta le operazioni di Censimento su tutto il territorio nazionale. In tal caso, il Censimento della popolazione e l'indagine di copertura consentono la formazione di due distinte liste della popolazione, entrambe soggette, generalmente, ad errore di copertura.

Più precisamente, si indichi con *C* la lista ottenuta con il Censimento e con *I* quella realizzata con l'indagine.

Per la valutazione della numerosità *N* della popolazione è possibile utilizzare le due liste integrando l'informazione fornita da entrambe per mezzo di un modello di rappresentazione dell'errore di copertura.

Le ipotesi alla base del modello, che devono essere verificate al fine di prevenire una distorsione nei risultati, sono (Wolter, 1986):

1. la popolazione di riferimento è chiusa e di dimensione fissata pari ad *N*;
2. le probabilità che l'unità *i* appartenga o meno alla lista *C* e che appartenga o meno alla lista *I* sono espresse da una distribuzione multinomiale in cui le probabilità congiunte e quelle marginali possono essere visualizzate nella seguente forma tabellare:

		Lista I		
		Sì	No	
Lista C	Sì	φ_{i11}	φ_{i12}	φ_{i1+}
	No	φ_{i21}	φ_{i22}	φ_{i2+}
		φ_{i+1}	φ_{i+2}	1

3. le due liste, *C* e *I*, sono considerate il risultato di *N* prove mutualmente indipendenti usando le distribuzioni multinomiali descritte al punto 2. Per ogni singola unità *i* si definisce una variabile x_{iab} pari ad 1 se l'unità *i* cade nella cella *ab* e 0 altrimenti ($a, b = 1, 2$). Dunque, le unità della popolazione, a seconda dell'appartenenza ad una delle due liste oppure ad entrambe, possono essere classificate come nel prospetto:

		Lista I		
		Sì	No	
Lista C	Sì	N_{11}	N_{12}	N_{1+}
	No	N_{21}	N_{22}	N_{2+}
		N_{+1}	N_{+2}	N

dove: $N_{ab} = \sum_{i=1}^N x_{iab}$ è il numero di unità nella cella ab ($a, b = 1, 2$), $N_{a+} = \sum_{i=1}^N \sum_{b=1}^2 x_{iab}$ ($a = 1, 2$) e $N_{+b} = \sum_{i=1}^N \sum_{a=1}^2 x_{iab}$ ($b = 1, 2$) sono il numero di unità nelle celle marginali rispettivamente a_+ e $_+b$; ovviamente, le quantità N_{22} e N non sono osservabili;

4. è possibile determinare senza errore quali unità registrate nella lista I sono presenti nella lista C e quali no (ossia, non sono presenti errori di abbinamento);
5. entrambe le liste sono depurate da errori di registrazione e duplicazione;
6. le probabilità che le unità incluse nella lista C, φ_{i1+} ($i = 1, \dots, N$), e le probabilità che le unità siano incluse nella lista I, φ_{i+1} ($i = 1, \dots, N$), sono costanti per ciascuna lista, ossia soddisfano le condizioni $\varphi_{i1+} = \varphi_{1+}$ e $\varphi_{i+1} = \varphi_{+1}$; d'altra parte le due probabilità φ_{1+} e φ_{+1} sono generalmente differenti.

La validità delle suddette assunzioni è un aspetto critico per la applicabilità del modello; chiariamo, quindi, come esse sono state verificate:

- i) la prima assume innanzitutto che non ci siano “nascite”⁸¹ o “morti”⁸² nella popolazione (in modo che essa sia enumerata in entrambe le occasioni di rilevazione “Censimento e indagine” sotto le stesse condizioni) e poi che l'ammontare della popolazione oggetto di studio sia ignoto ma fissato, ossia che non sia una variabile aleatoria. Questa ipotesi è solitamente realizzata fissando una data di riferimento e conducendo l'indagine di copertura in tempi sufficientemente ravvicinati rispetto al Censimento il che: da una parte, consente di minimizzare i cambiamenti che possono avvenire nella popolazione nel periodo post-censuario e, dall'altra, mette gli individui in una condizione migliore per ricordare più facilmente le risposte corrette inerenti alla data di riferimento del Censimento. La data di riferimento dell'indagine (Istat 2012) è, analogamente al Censimento, quella del 9 ottobre 2011; tuttavia, per intervenire sulla stessa popolazione, evitando al contempo di interferire con le normali operazioni censuarie, si è scelto di condurre la rilevazione in un momento immediatamente successivo alle operazioni sul campo e, nello specifico, a partire dal 1° marzo 2012;
- ii) la terza ipotesi è inerente all'indipendenza tra le due indagini. La rilevazione dell'indagine di copertura è stata progettata in modo da garantire l'indipendenza tra le indagini (Censimento e PES); inoltre, a tal fine, i rilevatori impiegati nell'indagine di copertura, scelti tra quelli più esperti, sono stati inviati in sezioni differenti da quelle del Censimento (Istat, 2012);
- iii) la quarta ipotesi assume che ci sia un abbinamento perfetto tra gli individui rilevati nelle due indagini. Gli abbinamenti “falsi positivi” porterebbero ad una sottostima della popolazione, mentre i “falsi negativi” ad una sovrastima della stessa. Per assicurare la massima correttezza degli abbinamenti è stata adottata una procedura di *record linkage* strutturata in diverse fasi e iterazioni; inoltre, sono state adottate tecniche di *record linkage* sia deterministiche che probabilistiche e sono state utilizzate differenti regole di confronto nelle varie fasi della procedura⁸³;
- iv) da questa ipotesi si evince che errori quali: duplicazioni nei risultati del Censimento, ritorni errati in entrambe le rilevazioni in quanto fittizi e rilevazioni erronee nel Censimento perché non facenti parte della popolazione di riferimento devono essere corretti prima della fase di

⁸¹ Le “nascite” comprendono sia letteralmente gli eventi di nascita che le persone immigrate.

⁸² Le “morti” includono sia testualmente gli eventi di morte che le persone emigrate.

⁸³ Per una descrizione dettagliata di questa fase si veda il capitolo terzo.

stima (Wolter, 1986). Una considerazione da fare è che essa fa riferimento alla sovra-copertura nazionale e, comunque, ad errori di sovra-copertura di tipo 1, 2 e 4; non rientrano in questa assunzione gli errori di sovra-copertura locale (gli individui conteggiati all'indirizzo sbagliato, ossia la sovra-copertura di tipo 3). L'ipotesi è verificata e, quindi, la suddetta sovra-copertura può ritenersi trascurabile, in parte come conseguenza delle precauzioni prese per realizzare l'ipotesi 1 e, in aggiunta, prevedendo per il Censimento e per la PES un'accurata fase di controllo e correzione dei dati. La procedura di stima studiata per la PES 2011, inoltre, prevede di depurare la stima dell'ammontare incognito della popolazione anche dagli errori di sovra-copertura di tipo 2 e 3 attraverso un fattore di aggiustamento applicato alla DSE;

- v) ultima condizione – volta a cautelarsi contro il rischio che la stima di N risulti distorta – impone che la probabilità di cattura sia costante per ogni individuo in ciascuna delle due catture. Il modo più agevole di rispettare tale ipotesi è costituito dall'introduzione di una stratificazione delle unità della popolazione basata su un insieme di variabili scelte in modo tale da massimizzare l'omogeneità delle probabilità di cattura, e calcolare quindi la popolazione ignota come somma delle popolazioni stimate per ciascuno strato.

Sulla base della simbologia precedentemente introdotta, in assenza di sovra-copertura, il tasso di copertura è dato da $\tilde{\tau} = (N_{1+}/N)$. Tuttavia, tale formula non è immediatamente utilizzabile in quanto N è incognito; ma, avendo assunto l'indipendenza delle due rilevazioni, il tasso di copertura può essere stimato come $\tilde{\tau} = (N_{11}/N_{+1})$, mentre una stima della numerosità della popolazione è data da

$$\tilde{N} = N_{1+} + N_{1+} \frac{N_{21}}{N_{11}} = \frac{N_{+1} N_{1+}}{N_{11}} . \quad (25)$$

Naturalmente, i totali N_{+1} e N_{11} non sono noti; tuttavia, essi sono stimabili utilizzando le osservazioni campionarie desumibili dalla PES. Pertanto, denotando con j la generica sezione di Censimento inclusa nel campione, le precedenti relazioni possono essere adottate a livello delle su menzionate sezione censuarie, per le quali le quantità N_{j+1} e N_{j11} sono osservabili, essendo l'operazione di conteggio completa entro le sezioni campione. Il tasso di copertura $\tilde{\tau}_j$ della sezione j è, dunque, stimato da $\tilde{\tau}_j = (N_{j11}/N_{j+1})$.

Per ciò che concerne, invece, il tasso di copertura riferito ad un generico dominio di interesse (ad esempio, l'intero territorio nazionale, la regione, la classe di età o la nazionalità), si possono stimare le quantità a numeratore e denominatore della (25) sulla base delle osservazioni campionarie:

$$\hat{\tau} = \frac{\hat{N}_{11}}{\hat{N}_{+1}} \quad (26)$$

in cui

$$\hat{N}_{11} = \sum_j N_{j11} w_j \quad \text{e} \quad \hat{N}_{+1} = \sum_j N_{j+1} w_j$$

sono rispettivamente le stime campionarie delle quantità N_{11} e N_{+1} , essendo w_j il peso finale attribuito alla sezione campione j espresso dalla (18). Parimenti, la stima di N è ottenibile mediante la relazione

$$\hat{N} = \frac{\hat{N}_{+1} N_{1+}}{\hat{N}_{11}} . \quad (27)$$

Dopo aver descritto il *modello di Petersen* e gli accorgimenti metodologici ed operativi presi al fine di verificare le ipotesi alla base del modello, passiamo ora ad illustrare la variante del suddetto modello adottata per la PES 2011.

Indichiamo con

- a l'età secondo le classi di modalità: 0-2, 3-7, 8-17, 18-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60-64, 65-69, 70-74, 75-79, 80-84, 85 e più, separatamente per i due sessi
- p le modalità dell'indice HtC ($p = 1,2,3$)
- c il generico comune ($c = 1, \dots, r m_{ap}$)
- r la regione geografica ($r = 1, \dots, 21$). Trento e Bolzano sono considerate come due distinte regioni,

e supponiamo ancora che l'indagine di copertura ripeta le operazioni di Censimento su tutto il territorio nazionale. Il DSE di ${}_r N_{apc}$, ossia del totale di popolazione vero incognito riferito al gruppo di individui del comune c selezionato nella modalità p dell'indice HtC della regione r con età e sesso definiti dalla modalità a , è dato da

$${}_r \tilde{N}_{apc} = \frac{{}_r N_{+1,apc} \overline{\overline{{}_r N_{1+,apc}}}}{{}_r N_{11,apc}} \quad (28)$$

in cui $\overline{\overline{{}_r N_{1+,apc}}}$, ${}_r N_{+1,apc}$ e ${}_r N_{11,apc}$ denotano rispettivamente il numero di conteggi esatti effettuati al Censimento, il numero di conteggi occorsi alla PES e il numero delle persone comuni al Censimento e alla PES, riferito allo stesso gruppo di individui.

Tuttavia, $\overline{\overline{{}_r N_{1+,apc}}}$ non è noto; al Censimento, infatti, si determina l'ammontare ${}_r N_{1+,apc}$, che è il numero di conteggi censuari comprensivi degli errori di enumerazione dovuti alla sovra-copertura. Al fine di correggere una possibile distorsione nei risultati, tenendo presente il fatto che la propensione alla sovra-copertura è costante nell'area g^{84} , e, quindi, che

$$E \left[\left(\frac{{}_r N_{1+,apc}}{\hat{\gamma}_{a_w, g}} \right) \middle| {}_r N_{1+,apc} \right] \cong \overline{\overline{{}_r N_{1+,apc}}}$$

si è deciso di adottare uno stimatore della forma

$${}_r \tilde{\tilde{N}}_{apc} = \frac{{}_r N_{+1,apc} \frac{{}_r N_{1+,apc}}{\hat{\gamma}_{a_w, g}}}{{}_r N_{11,apc}}$$

che risulta essere – per quanto appena visto – uno stimatore approssimativamente corretto di ${}_r \tilde{N}_{apc}$ dati ${}_r N_{+1,apc}$, ${}_r N_{11,apc}$ e ${}_r N_{1+,apc}$.

Oltre a questo aggiustamento del DSE per far fronte al fenomeno della sovra-copertura, è stata introdotta la correzione di Chapman per applicazioni del DSE a piccole popolazioni, le quali consentono un migliore rispetto dell'ipotesi di omogeneità della probabilità di cattura del *modello di Petersen*.

Riprendendo la formula (27), si ha

$${}_r \tilde{N}_{apc}^C = \frac{({}_r N_{+1,apc} + 1) (\overline{\overline{{}_r N_{1+,apc}}} + 1)}{({}_r N_{11,apc} + 1)} - 1 \quad (29)$$

la cui approssimazione del primo ordine è data da

⁸⁴ Per la definizione di g e a_w si veda il paragrafo 1.5.3.

$${}_r\tilde{N}_{apc}^C \cong {}_r\tilde{N}_{apc} \left(1 + \frac{1}{{}_rN_{+1,apc}} + \frac{1}{{}_r\tilde{N}_{1+,apc}} \right) \left(1 - \frac{1}{{}_rN_{11,apc}} \right) - \left(1 - \frac{1}{{}_rN_{11,apc}} \right). \quad (30)$$

Tuttavia, poiché il totale censuario osservato è ${}_rN_{1+,apc}$, il DSE corretto applicato per l'indagine in oggetto assume la forma

$$\begin{aligned} {}_r\tilde{N}_{apc}^C &= \frac{({}_rN_{+1,apc} + 1) \left(\frac{{}_rN_{1+,apc}}{\hat{\gamma}_{a_w,g}} + 1 \right)}{({}_rN_{11,apc} + 1)} - 1 \cong \\ &\cong {}_r\tilde{N}_{apc} \left(1 + \frac{1}{{}_rN_{+1,apc}} + \frac{\hat{\gamma}_{a_w,g}}{{}_rN_{1+,apc}} \right) \left(1 - \frac{1}{{}_rN_{11,apc}} \right) - \left(1 - \frac{1}{{}_rN_{11,apc}} \right) \end{aligned} \quad (31)$$

Il che consente di concludere che ${}_r\tilde{N}_{apc}^C$ è uno stimatore approssimativamente corretto di ${}_r\tilde{N}_{apc}$, dati ${}_rN_{+1,apc}$, ${}_rN_{11,apc}$ e ${}_rN_{1+,apc}$.

Poiché, però, la PES non ripete le operazioni di conteggio su tutte le sezioni del territorio nazionale ma soltanto su un campione di esse, i totali ${}_rN_{+1,apc}$ e ${}_rN_{11,apc}$ devono essere stimati attraverso il campione PES.

Si indichi, quindi, con ${}_r m_{ap}$ il numero di comuni campione selezionati nello strato p dell'indice HtC della regione r aventi popolazione appartenente alla classe di età e sesso definiti dalla modalità a . Dopo aver stimato il fattore di sovra-copertura, $1/\gamma_{a_w,g}$, si determina – per ciascuno degli ${}_r m_{ap}$ comuni – il *DSE corretto per sovra-copertura* in base all'espressione

$${}_r DSE_{apc} = {}_r\hat{N}_{apc}^C = \frac{({}_r\hat{N}_{+1,apc} + 1) \left(\frac{{}_rN_{1+,apc}}{\hat{\gamma}_{a_w,g}} + 1 \right)}{({}_r\hat{N}_{11,apc} + 1)} - 1. \quad (32)$$

1.5.5 Regression Estimator

Con riferimento al gruppo di individui censiti nella regione r con età e sesso definiti dalla modalità a si dispone dello *scatter* costituito dalle

$${}_r m_a = \sum_p {}_r m_{ap}$$

coppie di punti

$$\left({}_r x_{apc}, {}_r DSE_{apc} \right)$$

in cui: ${}_r x_{apc}$ è il numero di persone enumerate dal Censimento nel comune campione c selezionato nello strato p dell'indice HtC della regione r ed aventi età e sesso definiti dalla modalità a ; ${}_r DSE_{apc}$ è il corrispondente DSE – calcolato secondo la metodologia sopra illustrata – riferito allo stesso gruppo di individui.

Si consideri il modello

$${}_r DSE_{apc} = {}_r \beta_a {}_r x_{apc} + {}_r \varepsilon_{apc}$$

in cui ${}_r\beta_a$ è il coefficiente di regressione riferito al gruppo di individui della regione r con età e sesso definiti dalla modalità a e ${}_r\varepsilon_{apc}$ si può definire termine d'errore espresso come differenza tra la variabile e la sua media subordinata:

$${}_r\varepsilon_{apc} = {}^oDSE_{apc} - E({}^oDSE_{apc} | \mathbf{x}) = {}^oDSE_{apc} - {}_r\beta_a \cdot {}_rX_{apc}$$

con

$$\mathbf{x}' = ({}_rX_{a11}, \dots, {}_rX_{a1 \ r, m_{a1}}, \dots, {}_rX_{aP1}, \dots, {}_rX_{aP \ r, m_{aP}}).$$

Indicando con ${}_r\hat{\beta}_a$ lo stimatore dei minimi quadrati di ${}_r\beta_a$, si può scrivere la retta di regressione stimata

$${}^oD\hat{S}E_{apc} = {}_r\hat{\beta}_a \cdot {}_rX_{apc}.$$

Dunque, una stima della popolazione riferita alla modalità p dell'indice HtC della regione geografica r ed inerente al gruppo di individui con età e sesso definiti dalla modalità a può essere ottenuta attraverso la relazione

$${}_r\ddot{N}_{ap} = {}_r\hat{\beta}_a \cdot {}_rX_{ap} \quad (33)$$

dove ${}_rX_{ap}$ indica il totale censuario corrispondente a tutti i comuni nello strato p dell'indice HtC della regione r ed inerente al gruppo di individui con età e sesso definiti dalla modalità a .

1.5.6 Sample Balance Adjustment

L'efficienza delle stime fornite dalla (33) può essere migliorata applicando la metodologia di seguito esposta nei casi in cui i campioni, utilizzati per il loro ottenimento, risultino non sufficientemente rappresentativi delle rispettive popolazioni d'interesse.

Sotto le usuali assunzioni di normalità, l'attesa è che i campioni su cui si fonda la PES 2011 presentino il rilevante requisito di una buona rappresentatività della copertura del Censimento.

Al riguardo, tuttavia, è importante osservare che alla predisposizione del piano di campionamento segue logicamente la scelta di un campione tra tutti i possibili *a priori*; pertanto, non si può escludere il rischio che il campione selezionato per l'indagine sia un *outlier* tra tutti quelli costituenti l'universo dei campioni. Più concretamente, il campione PES potrebbe (per puro effetto del caso) risultare costituito da sezioni di Censimento in ciascuna delle quali il Censimento ha conteggiato l'intera popolazione; conseguentemente, il campione in questione produrrebbe sotto-copertura zero.

La metodologia in discussione, nota nella letteratura specializzata sull'argomento con il nome *Sample Balance Adjustment* (SBA), si prefigge l'obiettivo di valutare se il campione PES sia sufficientemente rappresentativo (in caso contrario, le stime dei tassi di copertura avrebbero una distribuzione asimmetrica) e – nel caso in cui il campione non soddisfi il suddetto requisito – correggere il valore della stima ${}_r\ddot{N}_{ap}$ moltiplicandola per il fattore ${}_rF_p$ il cui valore (${}_rF_p > 1$ o ${}_rF_p < 1$) esprime il grado di *prossimità* del campione PES rispetto alla condizione ideale di campione bilanciato (${}_rF_p = 1$).

Per i fini sopra delineati, si possono utilizzare i questionari *dummy* corrispondenti alle famiglie per le quali non è stato ricevuto un ritorno censuario; essi costituiscono la migliore *proxy* della copertura.

La procedura consta dei seguenti passi:

1. si analizza la correlazione tra le ${}_r m_p$ coppie $({}_r\bar{q}_{pc}, {}_r\hat{q}_{pc})$ – in cui ${}_r\bar{q}_{pc}$ e ${}_r\hat{q}_{pc}$ rappresentano rispettivamente il *non-response rate* (ovvero, il rapporto tra il numero di

questionari censuari *dummy* e il numero complessivo di questionari censuari) e la sua stima fornita dalla PES – al fine di stabilire se la variabile censuaria *dummy* è una buona *proxy* della non risposta (quindi, della copertura); essa può, pertanto, essere usata se il campione risulta non rappresentativo. In tal caso (ossia, se la correlazione è maggiore di 0.5), un aggiustamento basato su di essi migliorerebbe lo stimatore regressione se il campione non è bilanciato;

2. si confronta il numero di questionari censuari *dummy* nella regione r ed inerenti alla modalità p dell'indice HtC, ${}_r Y_p$, noto da Censimento, e la sua stima, ${}_r \hat{Y}_p$, ottenuta attraverso la PES utilizzando uno stimatore rapporto.

Se ${}_r \hat{Y}_p$ risulta significativamente diverso da ${}_r Y_p$, allora il campione PES relativo alla regione r e alla modalità p dell'indice HtC è verosimilmente non bilanciato. Tuttavia, per una ulteriore e più valida verifica di tale circostanza, è opportuno ricorrere all'impiego del test illustrato nel punto che segue;

3. essendo nota la distribuzione del numero reale di questionari censuari *dummy* relativa a tutti i comuni della regione r e aventi modalità p dell'indice HtC, si determina la varianza (e non la stima) di ${}_r \hat{Y}_p$, ossia $Var({}_r \hat{Y}_p)$.

Si costruisce, quindi, l'intervallo di confidenza al 95 per cento intorno al valore vero ${}_r Y_p$:

$${}_r Y_p - 2\sqrt{Var({}_r \hat{Y}_p)} \leq {}_r Y_p \leq {}_r Y_p + 2\sqrt{Var({}_r \hat{Y}_p)}. \quad (34)$$

Se la stima cade all'interno dell'intervallo, allora non c'è evidenza che il campione sia *outlier*; se, invece, cade in una delle due code, allora il campione è un *outlier*. In quest'ultimo caso è opportuno applicare un fattore di aggiustamento per migliorare il bilanciamento del campione;

4. si calcola il fattore

$${}_r F_p = \left[\frac{{}_r Y_p + {}_r X_p}{{}_r X_p} \right] \cdot \left[\frac{\frac{{}_r M_p}{r m_p} \sum_{c=1}^{r m_p} ({}_r Y_{pc} + {}_r X_{pc})}{\frac{{}_r M_p}{r m_p} \sum_{c=1}^{r m_p} {}_r X_{pc}} \right]^{-1}. \quad (35)$$

Se il campione è bilanciato, allora ${}_r F_p = 1$; se, invece, il campione sottostima, si avrà ${}_r F_p > 1$; infine, se il campione sovrastima, risulterà ${}_r F_p < 1$.

Dato il fattore ${}_r F_p$, si determina la stima bilanciata di ${}_r N_{ap}$, attraverso la relazione:

$${}_r \ddot{N}_{ap} = {}_r F_p \cdot {}_r \hat{N}_{ap} = {}_r F_p \cdot {}_r \hat{\beta}_a \cdot {}_r X_{ap}. \quad (36)$$

1.5.7 Determinazione delle stime del totale "vero" (incognito) della popolazione residente a livello regionale e nazionale

Nelle precedenti pagine abbiamo discusso ed illustrato le molteplici fasi della metodologia in base alla quale sono ottenibili le stime di ${}_r N_{ap}$, simbolicamente esprimibili secondo le due espressioni finali:

$${}_r \ddot{N}_{ap} = {}_r \hat{\beta}_a \cdot {}_r X_{ap}$$

e

$${}_r\ddot{N}_{ap} = {}_rF_p \cdot {}_r\ddot{N}_{ap}$$

a seconda che il campione utilizzato per il calcolo delle stime di ${}_rN_{ap}$ risulti o meno bilanciato.

A questo punto, si possono immediatamente determinare tutte le altre stime che costituiscono la finalità primaria della indagine PES 2011.

Dunque, sommando rispetto alle modalità p dell'indice HtC, si ottengono le stime

$${}_r\ddot{N}_a = \sum_p {}_r\ddot{N}_{ap}$$

riferite ai gruppi di popolazione a della regione r .

Conseguentemente, sommando anche rispetto ai gruppi di popolazione si hanno le stime a livello regionale; in simboli

$${}_r\ddot{N} = \sum_a {}_r\ddot{N}_a = \sum_a \sum_p {}_r\ddot{N}_{ap} . \quad (37)$$

La stima a livello nazionale è, quindi, fornita dalla relazione

$$\ddot{N} = \sum_r {}_r\ddot{N} . \quad (38)$$

1.6 HtC Index

L'Hard To Count index (HtC) contribuisce ad individuare aree omogenee relativamente alla difficoltà di una sotto popolazione ad essere correttamente enumerata. Lo studio del modello, sulla base del quale l'indice è stato progettato, è stato condotto attraverso un'analisi di fattori sociali, economici e demografici, che influenzano significativamente la probabilità di un individuo di essere censito. Queste caratteristiche mettono in luce alcune differenze, relativamente ai livelli locali di mancata risposta.

Seguendo l'importante esperienza dell'istituto di statistica Inglese (ONS), che ha applicato l'HtC in occasione dei censimenti del 2001 e del 2011, un indice analogo è stato studiato per categorizzare i comuni italiani, con lo scopo di ottenere dei livelli attesi omogenei di corretta enumerazione degli individui.

Per studiare la propensione degli individui della popolazione ad essere correttamente censiti, si è partiti dalle informazioni rilevate all'indagine di copertura. È stato dunque adottato un modello predittivo con lo scopo di stimare le probabilità individuali di corretto conteggio; questo modello assume una funzione link tra una variabile dipendente ed alcune variabili ausiliarie, rilevate nel corso dell'indagine PES o disponibili da altre fonti. La variabile dipendente è dicotomica⁸⁵, indica il mancato abbinamento tra gli individui rilevati durante l'indagine di copertura e quelli rilevati al Censimento, le modalità che assume la variabile sono le seguenti:

$$Y \begin{cases} 0 \text{ abbinamento corretto} \\ 1 \text{ mancato abbinamento} \end{cases}$$

Essendo la variabile dipendente una binaria, i modelli implementati sono dei logistici.

In particolare essi si differenziano in modelli ad effetti fissi, formalizzabili come

$$\text{Logit } P(Y_i = 1 | X_i) = \text{Log} \frac{P(Y_i = 1 | X_i)}{1 - P(Y_i = 1 | X_i)} = \alpha + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_k X_{ki}$$

e modelli Logit ad effetti casuali.

$$\text{Logit } P(Y_{id} = 1 | X_{id}) = \text{Log} \frac{P(Y_{id} = 1 | X_{id})}{1 - P(Y_{id} = 1 | X_{id})} = \alpha + \beta_1 X_{1id} + \beta_2 X_{2id} + \dots + \beta_k X_{kid} + \gamma_d$$

⁸⁵ Variabile disponibile dopo la fase di record linkage (capitolo terzo).

Questi ultimi si differenziano dai primi poiché, tengono conto delle sezioni di Censimento (divisioni territoriali sub-comunali) con un'intercetta casuale di area.

Le variabili ausiliarie disponibili per le unità statistiche rilevate dalla PES, descrivono caratteristiche socio-demografiche degli individui, dei comuni e delle province a cui appartengono. La post-stratificazione permette di sfruttare a pieno la ricchezza informativa dell'indagine di copertura, le cui informazioni individuali aggiornate alla rilevazione, possono essere integrate con altre variabili di area, disponibili da archivio. In fase di studio del modello sono stati proposti tre diversi modelli:

- ▶ il primo tiene conto solo delle variabili individuali,
- ▶ il secondo oltre a quelle individuali sfrutta anche covariate di area,
- ▶ il terzo tiene conto anche delle interazioni tra alcune coppie di variabili.

Nella tavola 1.2, viene riportata la lista completa delle variabili ausiliarie, distinte per il livello informativo a cui fanno riferimento (individuale, comunale, provinciale ed interazioni).

Tavola 1.2 - Variabili ausiliarie per differenti livelli di dettaglio

Livello	Variabili ausiliare
Individuale	Età
	Età in classi
	Sesso
	Famiglie monocomponente
	Famiglie numerose (più di 7 individui)
	Flag straniero
	Single (single, separate, divorziati, vedovi)
	Proxy studenti (19≤età≤30 con almeno un diploma)
Comunale	Città universitaria
	Città costiera
	Zona altimetrica (in 5 modalità)
	Densità popolazione (pop. Per km ²)
Provinciale	Tasso di disoccupazione
Interazioni	Flag straniero * tasso di stranieri
	Single * Classe età 10÷29
	Città universitaria * Proxy studenti

La modellizzazione *multilevel* implica lo studio della variabilità a diversi livelli, quindi è consuetudine valutare preventivamente se tale variabilità può essere considerata significativa. Nel primo *step* sono state testate due differenti intercette casuali di area: una a livello comunale ed una al livello inferiore di disaggregazione territoriale, la sezione di Censimento. Lo scopo di questo test è di capire quanta parte della varianza totale è spiegata da questi gruppi. A tale scopo è stato possibile calcolare il Coefficiente di Correlazione Intraclasse (ICC) usando la formula:

$$ICC = \tau_{00}/(\tau_{00} + \sigma^2)$$

dove τ_{00} è la varianza tra i gruppi o varianza d'intercetta, mentre σ^2 rappresenta la varianza entro i gruppi o varianza residuale. Il valore ICC stimato, a livello municipale è di 0.009 mentre quello a livello di sezione di Censimento è di 0.032; quest'ultimo valore della statistica test ci porta a propendere per un livello di dettaglio più disaggregato dell'intercetta casuale di area.

Nel secondo ed ultimo *step*, la significatività della varianza d'intercetta è stata valutata attraverso un test del rapporto delle verosimiglianze, in cui il valore del *-2 log-likelihood* del modello nullo con intercetta casuale di area (-584,294) è stato rapportato a quello relativo al modello nullo senza intercetta casuale di area (-579,87). La differenza di 4,423 risulta statisticamente significativa su una mistura di distribuzione Chi-quadro. Questi risultati suggeriscono che un'intercetta casuale a livello di sezione censuaria produce un miglioramento sostanziale del modello applicato. È stato

stimato, attraverso l'applicazione di modelli *multilevel* di cui sopra, che il 3,2 per cento della varianza totale della probabilità di mancata risposta è dovuta proprio all'appartenenza di un individuo ad una certa sezione di Censimento piuttosto che ad un'altra.

Anche lo studio del modello è stato svolto in maniera *step-by-step*. Il modello è stato selezionato attraverso l'uso di criteri comunemente utilizzati per la massimizzazione della *goodness of fit*: la *log-likelihood*, l'AIC ed il BIC.

Nel primo modello proposto sono state usate solo le variabili del questionario, disponibili per ogni individuo intervistato. Il miglior modello individuato è quello le cui covariate utilizzate sono state classe d'età, sesso e cittadinanza; gli indicatori di bontà di adattamento del modello ai dati AIC, BIC e *log-likelihood* sono rispettivamente 29381, 29466 e -14682. Successivamente sono state aggiunte al modello, suddetto, delle covariate area-level (tasso di disoccupazione, comuni universitari, densità di popolazione e tasso di stranieri) che inducono un incremento della *goodness of fit* del modello proposto, riscontrabile da una riduzione degli indicatori: AIC, BIC e *log-likelihood* scendono rispettivamente 29196, 29324 e -14586. Infine sono state aggiunte alle variabili indipendenti viste finora, gli effetti combinati per alcune coppie di covariate, ritenute a priori interessanti; tra questa l'unica risultata statisticamente significativa è stata quella tra la cittadinanza ed il tasso di residenti stranieri a livello comunale. Anche aggiungendo questo effetto si riscontra una ulteriore riduzione di AIC, BIC e *log-likelihood* che passano rispettivamente a 29174, 29313 e -14574. La tavola 1.3 mostra i coefficienti di regressione per i tre modelli appena presentati.

Tavola 1-3 - Coefficienti di regressione dei modelli

Variabili ausiliarie	Modello con variabili individuali	Modello con variabili individuali e di area	Modello completo
Intercetta	-5,711	-6,905	-7,067
Classe di età 10-29	0,075	0,074	0,072
Classe di età 30-49	0,048	0,046	0,041
Classe di età 50-74	-0,555	-0,555	-0,564
Classe di età ≥75	-0,481	-0,480	-0,488
Sesso (femmine)	-0,164	-0,166	-0,168
Flag straniero	2,395	2,395	2,848
Tasso disoccupazione		10,411	10,489
Città universitaria		0,826	0,826
Densità di popolazione		9,505e-05	9,178e-05
Tasso stranieri		4,594	6,817
Flag straniero * Tasso stranieri			-5,795

Una volta calcolate le probabilità di esser stato conteggiato, o meno, al Censimento, esse sono state mediate a livello comunale, così da produrre un dettaglio locale dei valori d'interesse.

La distribuzione ordinata dei valori predetti, relativi ai 252 comuni campione, è stata suddivisa in tre categorie, sulla base del 40° e dell'80° percentile, ottenendo una partizione, del tipo, 40 per cento-40 per cento-20 per cento della distribuzione.

Pertanto i comuni più virtuosi (il 40 per cento del totale di quelli campione), dal punto di vista della corretta enumerazione degli individui abitualmente dimoranti, sono stati categorizzati con la modalità HtC $p = 1$, i comuni che si trovano in una situazione intermedia, con la modalità HtC $p = 2$, ed infine ai comuni che hanno incontrato maggiori difficoltà nelle operazioni censuarie, è

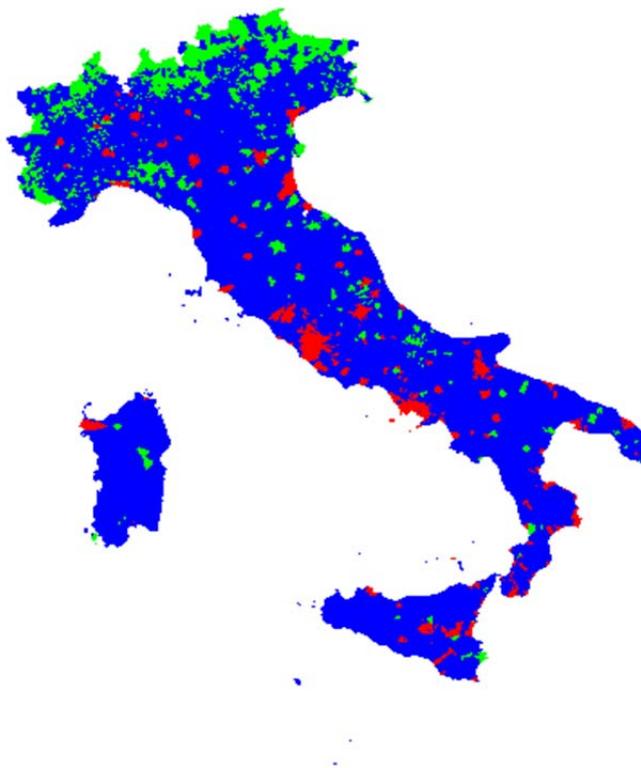
stata assegnata la modalità $p = 3$ dell'indice. Le modalità dell'indice HtC utilizzate nel modello sono:

- ▶ $p = 1$ 'enumerazione facile';
- ▶ $p = 2$ 'enumerazione di media difficoltà';
- ▶ $p = 3$ 'enumerazione difficile'.

Essendo interessati alla categorizzazione della totalità dei comuni italiani, questa categorizzazione delle probabilità predette è stata applicata anche alle probabilità dei comuni fuori dal campione selezionato, calcolate usando la sola parte sintetica del modello logistico *multilevel* descritto sopra. La ricchezza di informazioni disponibili ha permesso uno studio dettagliato, mai svolto finora, sugli individui più difficili da raggiungere nel corso del Censimento della popolazione. La figura 1.4 mostra la distribuzione dell'HtC nei comuni italiani:

Figura 1.4 - Distribuzione del HtC nei comuni italiani

Livello 1 in verde, livello 2 in blu, livello 3 in rosso.



I comuni più virtuosi, contrassegnati dal colore verde, sono distribuiti prevalentemente lungo le zone alpine ed appenniniche, in centri prevalentemente medio-piccoli. I comuni che si trovano in una situazione intermedia rispetto al fenomeno d'interesse, colorati in blu, sono i più numerosi e coprono la maggior parte del territorio nazionale. Le situazioni più problematiche, contraddistinte dal colore rosso nel cartogramma, mettono in evidenza come i grandi centri (Roma, Milano, Napoli, Firenze, Bari ecc.) abbiano più problemi di sovra e sotto copertura rispetto ai piccoli comuni. Dalla figura emerge anche una predominanza di comuni con HtC3 lungo le coste, soprattutto in zone balneari; tale problematica si collega a quella della copertura per evidenti motivi legati alle case di villeggiatura e alla mobilità dei lavoratori stagionali.

1.7 Attendibilità delle stime dei tassi di sotto e sovra-copertura

Le stime prodotte da un'indagine campionaria sono sempre affette da errore; per questo motivo, è necessario associare ad ogni stima una misura della sua precisione o incertezza. Limitatamente alla variabilità campionaria e quindi all'errore campionario⁸⁶, il problema della stima della precisione degli stimatori è correntemente noto come problema della *stima della varianza*, in considerazione della preminenza degli stimatori corretti o asintoticamente corretti. La determinazione di stime statisticamente e computazionalmente efficienti della varianza delle stime dei parametri di interesse è, pertanto, un aspetto cruciale per la valutazione della qualità delle stime stesse. A tale scopo, vengono generalmente utilizzati altri indicatori (coefficiente di variazione, intervalli di confidenza) immediatamente derivabili dalla varianza.

L'attendibilità delle stime dei parametri di interesse dell'indagine PES è stata valutata in base ai valori dei coefficienti di variazione e degli intervalli di confidenza, basati sulla stima della varianza delle stime prodotte dall'indagine calcolata applicando il *metodo bootstrap*.

Questo metodo, proposto solo di recente (Efron, 1979), è sostanzialmente un metodo di valutazione per via numerica, dalle caratteristiche distributive di uno stimatore di qualsiasi natura e, quindi, particolarmente utile nel caso di stimatori dalla struttura analitica complessa, con riferimento ai quali i metodi standard e il metodo di linearizzazione di Taylor richiedono complesse derivazioni matematiche per il calcolo della varianza delle stime. Esso, inoltre, è adattabile anche al contesto di piani di campionamento a più stadi di selezione (Efron, Tibshirani, 1993; Wolter, 2007).

Qui di seguito si descrivono le fasi metodologicamente più importanti per applicare il metodo bootstrap all'indagine in esame:

- a) la prima fase persegue la finalità di costruire una *pseudo-popolazione* dalla quale generare i *campioni bootstrap*.

A tale scopo, è necessario riflettere in essa la struttura della popolazione usata nella formazione del campione PES, che presenta in ciascuno degli *strati d'indagine* le seguenti caratteristiche essenziali:

- stratificazione delle unità di primo stadio, ossia dei comuni;
- selezione di un solo comune campione dagli strati NAR con probabilità variabile;
- stratificazione della popolazione delle unità di secondo stadio, costituite dalle sezioni di Censimento, sia dentro ciascun comune campione NAR che in ciascun comune AR costituente strato a sé stante, in base alla dimensione demografica delle sezioni di Censimento secondo le tre modalità definite dai terzili della distribuzione della popolazione in questione;
- selezione, con probabilità uguali e senza ripetizione, di un prefissato numero di sezioni di Censimento da ciascuno degli strati di cui al precedente punto.

Passiamo ora a descrivere i punti salienti della procedura utilizzata per la costruzione della pseudo-popolazione in discussione:

- in ciascuno degli strati NAR, si *copia* il corrispondente comune campione un numero di volte pari al rispettivo peso campionario (*pseudo-popolazione di primo stadio*).

⁸⁶ L'errore campionario deriva dall'aver osservato la o le variabili di interesse solo su una parte (campione) della popolazione a differenza dell'errore non campionario che deriva ad esempio da errori nelle liste della popolazione usate per selezionare le unità del campione o mancate risposte parziali dovute a risposte mancanti o non ammissibili a causa di errori di rilevazione o di registrazione.

- Si *copia* poi ciascuna delle sezioni di Censimento campione, afferenti ad ognuno degli strati di secondo stadio definiti nel suddetto comune, un numero di volte pari al rispettivo peso campionario (*pseudo-popolazione di secondo stadio*).
 - Riunendo la pseudo-popolazione di primo stadio e le tre di secondo stadio, si ottengono tre pseudo-popolazioni costituite da due stadi di selezione.
 - per ciascuno degli strati AR (ossia, per ogni comune AR), si costruiscono le pseudo-popolazioni relative ai tre strati di sezioni di Censimento con un procedimento analogo a quello esposto per il comune NAR.
- b) la seconda fase consiste nel costruire le *repliche bootstrap* estraendo – con ripetizione – dalle pseudo-popolazioni sopra descritte un certo numero di campioni di dimensioni uguali a quelle del campione PES.
- c) per ciascuno dei B ($B=1.000$) campioni così costruiti (campioni bootstrap) si calcolano le stime bootstrap $\hat{\theta}_b$ ($b=1, \dots, B$) del parametro θ mediante uno stimatore avente una forma funzionale uguale a quella dello stimatore $\hat{\theta}$ adottato per l'ottenimento delle stime dell'indagine (basate sull'utilizzo dei dati osservati con il campione originario).
- d) la varianza di campionamento bootstrap dello stimatore $\hat{\theta}$ è agevolmente ottenibile tramite l'utilizzo della formula

$$\hat{V}_{BS}(\hat{\theta}) = \frac{1}{B-1} \sum_{b=1}^B \left(\hat{\theta}_b - \frac{1}{B} \sum_{b=1}^B \hat{\theta}_b \right)^2. \quad (39)$$

- e) a questo punto, è possibile:
- calcolare il coefficiente di variazione percentuale espresso da

$$CV_{BS}(\hat{\theta}) = \frac{\sqrt{\hat{V}_{BS}(\hat{\theta})}}{\hat{\theta}} 100 \quad (40)$$

- definire un intervallo di confidenza *standard* al livello di fiducia del 95 per cento, assumendo che lo stimatore sia distribuito secondo una normale o in modo approssimativamente normale

$$\hat{\theta} \pm 1,96 \sqrt{\hat{V}_{BS}(\hat{\theta})}. \quad (41)$$

I coefficienti di variazione e gli intervalli di confidenza, inerenti alle stime dei tassi di copertura obiettivo della presente indagine, ottenuti attraverso la metodologia appena descritta, sono riportati nel capitolo quarto, per ogni dominio di stima.

CAPITOLO 2

Fasi post-rilevazione ⁸⁷

2.1 Obiettivi

La Post Enumeration Survey (PES) è un'indagine statistica che, come già detto, ha l'obiettivo di misurare la qualità del 15° Censimento generale della popolazione e delle abitazioni. Pertanto, ciascuna fase (progettazione, predisposizione dello strumento di rilevazione, rilevazione sul campo e monitoraggio della rilevazione) è stata condotta con estrema attenzione, in modo da garantire un elevato livello di accuratezza.

Tuttavia, al termine della fase di raccolta dei dati, ulteriori controlli ed aggiustamenti si sono resi necessari al fine di disporre di valori validi per ciascun individuo, eliminando gli errori e le incongruenze. Nello specifico i controlli sono stati finalizzati al soddisfacimento di due condizioni necessarie per il corretto prosieguo del lavoro:

- 1) assicurare che tutti i record individuali fossero classificabili nelle modalità rispetto alle quali i tassi di copertura sono stati prodotti e di non compromettere la correttezza e l'attendibilità delle stime di copertura⁸⁸;
- 2) assicurare una buona qualità delle variabili di abbinamento, dotate di un elevato potere identificativo degli individui rilevati⁸⁹, per una corretta conduzione della fase di record linkage tra i dati di indagine e quelli di Censimento. Infatti come già ricordato, tale fase è indispensabile per poter applicare il metodo cattura-ricattura e stimare l'ammontare della popolazione.

2.2 Costruzione della base dati 'PES'

Per la PES, analogamente a quanto avvenuto per il Censimento, è stato utilizzato il Sistema di Gestione della Rilevazione (SGRPES, descritto nella parte prima) implementato per la gestione dei rispondenti.

In SGRPES sono registrate solo alcune informazioni relative agli individui rispondenti; ci sono alcune informazioni relative alla localizzazione territoriale della famiglia (comune di residenza ed indirizzo dell'abitazione) e alcune informazioni necessarie per l'individuazione univoca dell'individuo (nome, cognome, data di nascita, sesso e luogo di nascita).

⁸⁷ I paragrafi sono a cura di:

- 2.1 Obiettivi: Federico Benassi, Antonella Bernardini
- 2.2 Costruzione della base dati PES: Francesca Tiero, Luca Valentino
- 2.3 Costruzione dell'archivio CEN: Luca Valentino
- 2.4 Controllo e correzione dei microdati: Marco Dionisio Terribili
- 2.4.1 Principali variabili: Marco Dionisio Terribili
- 2.4.2 Normalizzazione degli indirizzi: Andrea Fasulo
- 2.4.3 Corretta localizzazione degli individui: Andrea Fasulo
- 2.5 Analisi dell'eleggibilità: Antonella Bernardini
- 2.6 I numeri della rilevazione: Antonella Bernardini, Francesca Tiero

⁸⁸ Ad esempio, per poter produrre correttamente i tassi di copertura riferiti al genere, oppure alla cittadinanza, è stato necessario verificare che per tutti gli individui fosse attribuibile la variabile sesso e la variabile cittadinanza. Pertanto i dati da analizzare sono stati preventivamente sottoposti a una procedura di revisione, tale da attribuire un valore plausibile ai record che presentavano valori mancanti o non ammissibili. L'ammontare dei record sottoposti a correzione è stato comunque di dimensioni trascurabili rispetto alla dimensione dei dati trattati.

⁸⁹ Ad esempio nome, cognome, data di nascita ed indirizzo/i.

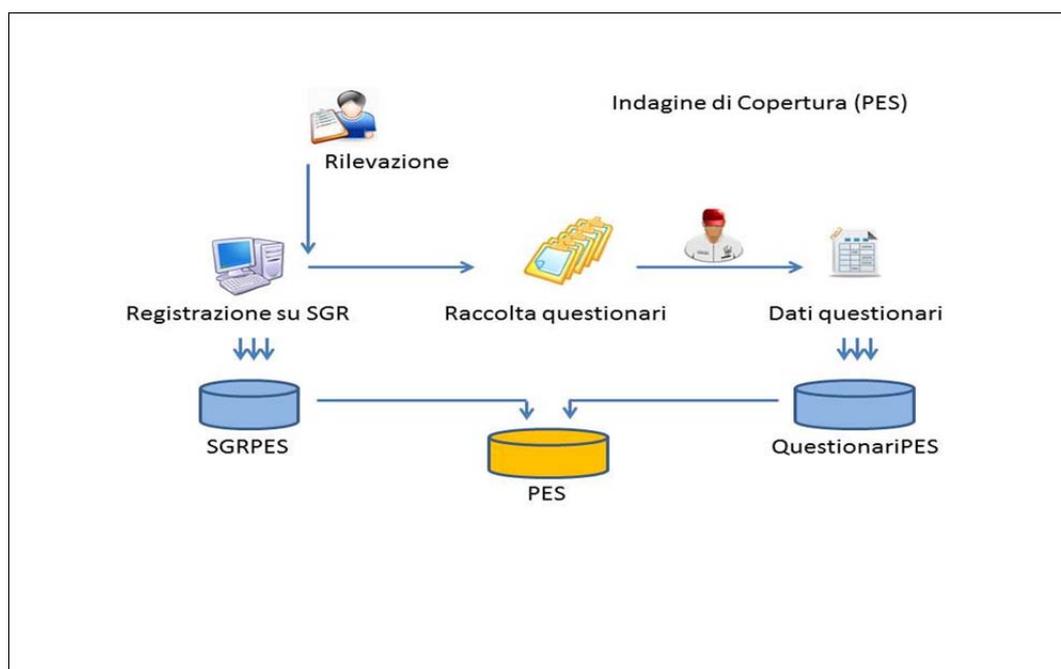
Il questionario sottoposto alla popolazione presente sul territorio, delle sezioni campione, rileva una serie di informazioni aggiuntive relative agli individui censiti. Alcuni di questi dati sono stati raccolti al fine esclusivo di valutare la qualità dei dati del Censimento (ad esempio grado di istruzione e stato civile), altre informazioni sono invece necessarie per le operazioni di abbinamento e quindi per la stima della popolazione e il calcolo dei tassi di sotto-copertura e sovra-copertura.

Conclusa la fase di rilevazione i questionari raccolti sono stati registrati in formato elettronico da una società esterna all'Istat.

Quindi per il database 'PES' si dispone sia della base dati di SGRPES che delle informazioni provenienti dai questionari cartacei, che hanno contribuito alla costruzione del database 'Questionario PES'.

Le informazioni provenienti dai due sistemi di registrazione possono essere integrate automaticamente poiché i dati di 'Questionario PES' contengono anche la chiave di aggancio a SGRPES. Nella figura 2.1 è rappresentato il processo che porta, in base a quanto detto, alla generazione della base dati 'PES'.

Figura 2.1 - Flusso di raccolta e registrazione dei dati della 'PES'



La costruzione del database 'Questionario PES', contenente le informazioni raccolte durante l'indagine di copertura (PES), è stata propedeutica ai fini dell'analisi dei dati stessi dell'indagine.

I dati raccolti tramite questionario, infatti, sono stati registrati e resi disponibili in diversi tracciati record, ognuno dei quali riferito ad una specifica sezione del questionario⁹⁰. Di seguito nella tavola 2.1 sono riprodotte le informazioni desunte dai questionari di indagine, associate ai diversi tracciati e la dimensione dei tracciati stessi.

⁹⁰ Per la struttura del questionario si rimanda al paragrafo 3.2 parte prima.

Tavola 2.1 - Riepilogo della dimensione delle sezioni del questionario cartaceo

Tracciato record	Sezione del questionario	Dimensione del tracciato
1	Primo foglio del questionario	138387
2	Lista "Persone della famiglia che hanno dimora abituale nell'alloggio"	332004
3	Informazioni inserite sui fogli individuali di ciascun componente la famiglia – domande 1.1 – 1.6	332450
4	Informazioni inserite sui fogli individuali di ciascun componente la famiglia – domande 2.1 – 4.2	332450
5	Informazioni sulla famiglia	138345
6	Informazioni sul pacco da cui è stato recuperato il questionario	1150
7	Riepilogativo del contenuto del pacco da cui è stato preso il questionario	415040

Le numerosità dei tracciati risultano, naturalmente, diverse. Ciò in ragione del fatto che quando i tracciati si riferiscono alle famiglie (tracciato 1) si ha un record per ogni questionario (ogni famiglia), mentre quando si riferiscono agli individui si hanno tanti record quanti sono i componenti della famiglia. Tuttavia anche tra i tracciati che si riferiscono agli individui si riscontrano differenze di numerosità.

Si è giunti alla definizione finale di un database unico, proveniente dai questionari cartacei, allineando il numero di record presenti nei diversi tracciati attraverso due chiavi, una associata univocamente ad un questionario cartaceo e dunque ad una singola famiglia, l'altra identificativa dell'individuo all'interno della famiglia stessa. Tale operazione di ricongiungimento delle informazioni di tutti i tracciati in record individuali, ha portato alla costruzione del database 'Questionario PES' di 331.575 record. Questa base di dati è stata poi integrata con le informazioni di SGRPES che conteneva 331.612 record.

Tale operazione ha richiesto ulteriori controlli sui singoli record per verificare le differenze di numerosità delle due basi di dati. La tavola seguente quantifica questa incongruenza.

Tavola 2.2 - Confronto tra la chiave di aggancio per i record di 'Questionario PES' e quelli di SGRPES

		In 'Questionario PES'		
		SI	NO	TOT
In SGRPES	SI	330.504	1.135	* 331.639
	NO	1.071	0	1.071
	TOT	331.575	1.135	332.710

* di cui 27 duplicati

La chiave che permetteva di collegare le due fonti, infatti, in 1.071 casi, su un totale di 331.575 record di 'Questionario PES', e in 1.135 casi, dei 331.612 di SGRPES, non era valorizzata (o riportava un dato non valido), inoltre in 27 casi la chiave nel database SGRPES risultava duplicata. Risolte le incongruenze attraverso l'analisi delle singole informazioni contenute nelle due fonti, è stata possibile l'integrazione completa e la costruzione dell'archivio 'PES' che contava 332.710 record.

L'analisi delle variabili comuni alle due basi di dati, per i record integrati, evidenzia una buona corrispondenza per i dati a livello di questionario (ad esempio, coincidono per tutti i record integrati i dati territoriali di provincia e comune e si riscontrano poche differenze sul codice sezione). Al contrario, in diversi casi, non c'è corrispondenza tra i dati individuali (tipo nome, cognome e data di nascita); questo, in massima parte, è dovuto non solo ad errori di digitazione quanto ad una non corrispondenza dell'ordine progressivo in cui i singoli individui appaiono nei questionari, come evidenziato nello schema seguente:

Esempio di corrispondenza tra i record di 'Questionario PES' e quelli di SGRPES

Questionario 101	Individuo1– Tizio Rossi		SGRPES_Q101	Individuo1 – Caio Rossi
Questionario 101	Individuo2 – Caio Rossi		SGRPES_Q101	Individuo2 – Tizio Rossi

In fase di costruzione dell'archivio 'PES', le informazioni desumibili da SGRPES sono risultate fondamentali per integrare le informazioni mancanti nonché per risolvere alcuni casi di disallineamento tra i vari tracciati. Queste operazioni sono state possibili grazie all'utilizzo della variabile Codice_questionario in SGRPES, una variabile composta da 13 cifre di cui le prime 6 identificative di provincia e comune e le restanti 7 generate in modo casuale dal sistema. Questa variabile ha consentito quindi di risalire al codice a barre identificativo del questionario e, dunque, al questionario stesso.

Nei tracciati, sia a livello individuale che a livello di famiglia, sono stati riscontrati dei duplicati. Nella maggior parte dei casi questi record contenevano le medesime informazioni.

Diverse sono le motivazioni che hanno portato alla presenza di record duplicati; a tale proposito, è utile ricordare che il questionario è stato predisposto per raccogliere i dati di una famiglia composta da, al più, cinque persone. Nel caso in cui una famiglia risultava più numerosa, era quindi necessario, durante la fase di rilevazione sul campo, utilizzare uno o più questionari aggiuntivi dello stesso tipo. Ciò ha determinato l'esigenza, limitatamente alle famiglie composte da almeno sei componenti, di riconoscere tutte le persone appartenenti ad uno stesso nucleo familiare (operazione di ricostruzione delle famiglie). Questa operazione è stata effettuata utilizzando il numero d'ordine definitivo, assegnato dal rilevatore a ogni famiglia della sezione campione.

Le attività di ricostruzione delle informazioni rilevate dall'Indagine di copertura sono state volte, pertanto, a garantire:

- ▶ l'attribuzione univoca di numeri d'ordine definitivi alle famiglie, quando mancanti o duplicati nell'ambito della stessa sezione campione;
- ▶ il riconoscimento di eventuali record duplicati, soprattutto nel caso di famiglie costituite da più di cinque persone, dove il capo famiglia si era registrato in più fogli. In questo caso la duplicazione era dunque attribuibile ad una doppia registrazione dello stesso individuo o della stessa famiglia.

Nel resto dei casi, invece, è accaduto che alla stessa chiave univoca di questionario fossero abbinate famiglie diverse. Questa problematica riguardava nuclei familiari differenti, ma spesso coabitanti, rilevati erroneamente con lo stesso questionario. La risoluzione di tali casi è avvenuta rendendo univoco l'identificativo familiare aggiungendo un digit in fondo al codice errato.

La quantificazione e l'eliminazione di queste tipologie di duplicazioni, insieme a quelle individuate nella prima fase del record linkage, sono descritte nel paragrafo 3.2.2.

Successivamente agli interventi descritti, si è proceduto ad unire le informazioni contenute nei tracciati 1, 2 e 3 ed aggiungere a queste quelle desumibili da SGRPES, per un totale di 332.710 record.

2.3 Costruzione dell'archivio 'CENS'

Per l'archivio Censimento ('CENS'), non avendo ancora a disposizione i dati dei questionari, l'unico data set di riferimento per le procedure di abbinamento è stato quello del Sistema di Gestione della Rilevazione, progettato e realizzato per il Censimento generale della popolazione e delle abitazioni, (SGR). L'archivio proveniente da SGR validato per la popolazione legale contava 59.433.744 individui (CENS). Considerando solo le sezioni di Censimento facenti parte del campione dell'Indagine di Copertura il data set, confrontabile con la base dati PES, che era di 352.106 individui; ha preso il nome di 'CENSPARZ' (da tener conto che in 'CENSPARZ' entrano solo i residenti in famiglia poiché i residenti in convivenza sono esclusi dall'Indagine di Copertura). La costruzione degli archivi 'CENS' e 'CENSPARZ' è stata strumentale alle operazioni di linkage descritte nel successivo capitolo terzo.

2.4 Controllo e correzione dei microdati

I dati contenuti in 'Questionario PES', integrati con le informazioni contenute in SGRPES, sono stati quindi controllati con lo scopo di disporre di un insieme di dati completo e coerente a livello individuale. Un insieme di dati, quindi, in cui ogni record avrebbe dovuto contenere valori delle variabili di interesse ammissibili dato un insieme finito di regole di compatibilità (in letteratura note con il nome di edits).

Si pensi, ad esempio, alle variabili età, stato civile e titolo di studio; è evidente che un bambino non può avere un titolo di studio elevato, ad esempio laurea o dottorato di ricerca, né tantomeno può risultare coniugato. Questo banale esempio chiarifica le incoerenze che sono state analizzate e corrette in questa fase seguendo l'approccio tipico adottato in Istituto e, nello specifico, i seguenti step:

1. *specificazione* delle relazioni logico/matematiche che i microdati devono rispettare;
2. *controllo dei record* ed individuazione di quelli che non soddisfano le relazioni definite al punto 1;
3. *localizzazione* degli errori, due o più variabili sono incoerenti fra loro: qual è quella giusta e qual è quella sbagliata?;
4. *correzione* degli errori con valori plausibili (tali da soddisfare tutti gli edit) provenienti da altre variabili dello stesso questionario o da altra fonte (SGRPES).

Le regole definite per il controllo e la correzione sono state esclusivamente regole intra-record, ossia volte alla verifica della compatibilità delle informazioni relative ad un solo individuo, senza alcun riferimento ad informazioni riguardanti altri individui (in questo caso avremmo parlato di regole inter-record). Le regole specificate sono state espresse mediante proposizioni logiche che si servono degli usuali operatori logici (AND, OR, NOT, etc.), ed hanno riguardato principalmente relazioni con la variabile età:

- ▶ if età ≤ 14 → stato civile = "non coniugato";
- ▶ if età ≤ 9 → titolo di studio ≤ "licenza elementare";
- ▶ if età ≤ 12 → titolo di studio ≤ "licenza media";
- ▶ if età ≤ 17 → titolo di studio ≤ "diploma di scuola media superiore";
- ▶ if età ≤ 20 → titolo di studio ≤ "laurea".

2.4.1 Principali variabili

Una volta identificati i record contenenti le incoerenze, si è proceduto a capire quale valore fosse quello più plausibile sulla base delle altre variabili ausiliarie disponibili, correggendo il valore definito errato.

Nella revisione dei microdati, si è attribuito un ordine gerarchico alle variabili per tener conto dell'impatto che l'avvenuta correzione di un campo potesse avere sul valore di un altro ad esso legato. Si è corretto quindi, in primo luogo, l'anno di nascita per due principali motivi:

- ▶ in funzione della data di nascita si stabilisce uno dei criteri di eleggibilità dell'unità;
- ▶ in base all'età calcolata sulla data di nascita si deve o meno fornire una risposta in merito al titolo di studio (solo per chi ha 6 anni o più) e alla condizione professionale (solo per chi ha 15 anni o più).

Successivamente sono state imputate le variabili di interesse con procedure sia di tipo deterministico che probabilistico, tenendo conto della diversa natura delle variabili esaminate. In questo modo ogni individuo rilevato all'indagine di copertura presenta valori compatibili per le variabili oggetto di stima.

Nella revisione della variabile anno di nascita, come anche per la variabile sesso, si è considerato in primo luogo il fatto che tali informazioni sono state rilevate più volte nel questionario. Pertanto, uniti i diversi tracciati e le informazioni contenute in SGRPES, in un unico database, è stata verificata la coerenza della stessa informazione ripetuta. In caso di discordanza, è stato attribuito un peso maggiore a quella riportata in SGRPES. Nel caso di informazione mancante nel questionario cartaceo, la stessa è stata imputata sulla base del valore presente in SGRPES. In generale, la procedura di correzione per l'anno di nascita e per il sesso ha riguardato una parte marginale dei record.

Corrette le variabili anno di nascita e sesso si è proseguito con il controllo e la correzione della variabile stato civile. Tale variabile, tuttavia, sarà utilizzata principalmente ai fini della stima dell'errore di misura del Censimento, non trattato in questo volume. Dove l'informazione era mancante o aveva un valore non ammissibile e si era in presenza di minori, è stata imputata la modalità celibe/nubile tramite un criterio deterministico; dove, invece, era presente la variabile anno di matrimonio è stato imputato un valore, dello stato civile, secondo la distribuzione dello stato civile condizionatamente alla presenza della data di matrimonio degli individui che avevano dichiarato entrambe le informazioni.

Un approccio diverso è stato adottato nel caso della variabile cittadinanza che, come prevedibile, presentava diversi casi mancanti o valori non ammissibili. È importante precisare che si sono scelte regole conservative, che tendessero a imputare il valore "straniero" solo nei casi in cui si avesse la quasi certezza della cittadinanza straniera. Questo per non introdurre distorsioni nella stima del tasso di copertura; è infatti ragionevole ritenere che gli stranieri abbiano una maggiore propensione/probabilità di sfuggire all'enumerazione censuaria.

Per poter attribuire la cittadinanza sono state seguite diverse procedure di imputazione, cercando in primo luogo di dedurre l'informazione mancante dalle altre variabili del record individuale. Il modello di rilevazione richiedeva, infatti, per coloro che avevano indicato la cittadinanza straniera, lo stato estero di cittadinanza in formato di stringa alfabetica. Inoltre, veniva chiesto a tutti il luogo di nascita, distinguendo se questo coincideva con il comune d'indagine o meno e, in quest'ultimo caso, specificando tra le modalità "altro comune italiano" (con indicazione del comune) e "all'estero" (con indicazione dello stato estero). La variabile stato estero di nascita, se presente e ammissibile, ha permesso di risalire alla cittadinanza, qualora non fosse stata indicata. Si è usato, quindi, il paese di nascita come *proxy* della cittadinanza. I restanti casi sono stati considerati sulla

base delle famiglia d'appartenenza. Nel caso di famiglie con più componenti sono state considerate le cittadinanze dichiarate all'interno della famiglia, imputando il valore prevalente.

2.4.2 Normalizzazione degli indirizzi

Un discorso specifico meritano gli indirizzi che localizzano la famiglia e i suoi componenti; in caso di discordanza tra i dati registrati nei tracciati e quelli provenienti da SGRPES è stato mantenuto il valore di SGRPES come valido, in quanto questo non veniva registrato ma acquisito da un archivio precedentemente caricato nel sistema.

Si è posto poi il problema della normalizzazione degli indirizzi: quelli provenienti dal tracciato 1 non erano stati normalizzati al contrario di quelli di SGRPES. Sono stati, quindi, normalizzati tutti gli indirizzi (sia del tracciato 1 che di SGRPES) in modo da garantire che i due gruppi di indirizzi fossero normalizzati secondo i medesimi criteri, anche rispetto ai dati del Censimento, per facilitare l'operazione di abbinamento.

Tuttavia, sempre al fine di avere più informazioni possibili per la fase di record linkage, dove si sono rilevate eventuali discordanze, si è individuato il valore corretto ma si è tenuto traccia anche del valore errato, in modo da poterlo eventualmente recuperare in fase di linkage.

Per quanto riguarda invece i codici di sezione e quelli dei comuni mancanti nei record di SGRPES (infatti il tracciato 1 aveva un numero di osservazioni superiore rispetto a quelli di SGRPES) sono state recuperate queste informazioni dal tracciato 6. La normalizzazione degli indirizzi è stata effettuata mediante l'utilizzo del software EGON.

2.4.3 Corretta localizzazione degli individui

L'ultimo passo di controllo e correzione dei dati si è focalizzato sullo studio delle risposte fornite alla domanda 1.5 del questionario PES. In tale quesito veniva richiesto di la dimora abituale alla data di riferimento del Censimento (9 ottobre 2011). I rispondenti potevano scegliere tra quattro possibili risposte:

1. In questo alloggio;
2. In un altro alloggio di questo comune, specificando l'indirizzo completo;
3. In un altro comune italiano, specificando il comune e la provincia;
4. All'estero.

La PES è stata svolta tra l'11 aprile e il 31 maggio 2012, al termine delle operazioni sul campo relative al Censimento, in modo tale che non ci fosse sovrapposizione tra le due rilevazioni. Nel pur ridotto lasso di tempo intercorrente tra le due rilevazioni (circa sei mesi), volto a ridurre al minimo la distorsione delle informazioni dovuta all'effetto memoria, possono essersi verificati un certo numero di eventi di natura demografica: movimenti migratori (interni al singolo comune, fra comuni all'interno del territorio italiano, da e per l'estero), eventi naturali (nascite e decessi), variazioni di status personale relativamente alla situazione sociale ed economica, di formazione (acquisizione di un nuovo titolo di studio), di stato civile (matrimonio o divorzio) e di cittadinanza.

Il quesito "Dove dimorava alla data del Censimento (9 ottobre 2011)?" è stato pertanto fondamentale per definire la "corretta localizzazione" dell'individuo alla data del Censimento e stimare la propensione alla sovra-copertura, cioè l'attitudine degli individui ad essere sovra-enumerati. Tale parametro è stato utilizzato per stimare successivamente il tasso di sovra-copertura e di sotto-copertura del Censimento (come descritto in paragrafo 1.5.3). A questo scopo si è cercato di garantire la maggior qualità possibile nelle risposte fornite al quesito attraverso un'attenta fase di revisione. Nella tavola 2.3 sono riportate le frequenze di risposta alla domanda 1.5 del questionario.

Tavola 2.3 - Domanda 1.5 Indicare in quale alloggio possedeva la dimora abituale alla data del Censimento

Items	Totale unità	di cui	
Mod.1 In questo alloggio	320.560		
Mod. 2 In un altro alloggio di questo comune	4.676	Sezione campione	850
		Sezione non campione	3.826
Mod.3 In un altro comune italiano	3.588	Sezione non disponibile	3.588
Mod.4 All'estero	599	Non applicabile	
Missing	328	Non applicabile	
Totale complessivo	329.751		

La “corretta localizzazione” è stata determinata a livello di sezione di Censimento, pertanto sono state prese in considerazione solo le unità che possedevano dimora abituale, alla data di riferimento del Censimento, in una sezione campione della PES. Come è evidente dai dati riportati in tavola 2.6, la quasi totalità degli individui, al momento della PES, si trovava nello stesso alloggio in cui possedeva dimora abituale al Censimento. Per le altre unità si è proceduto ad un secondo passaggio di geo-codifica degli indirizzi forniti. In particolare sono stati standardizzati gli indirizzi (in 4.676 casi) specificati dai rispondenti che hanno barrato la modalità 2 della domanda 1.5 e, attraverso un'importante fase di raccordo con l'unità operativa dell'Istituto “Archivio dei numeri civici e strumenti territoriali”, si è risaliti alla sezione di Censimento. Infatti i comuni sono costituiti da più sezioni di Censimento ed il fatto di risiedere nello stesso comune non garantiva necessariamente che la sezione fosse inclusa nel campione areale.

Alla casistica sopra specificata si sono aggiunti sporadici casi di incoerenza tra le informazioni contenute nei tracciati e quelle contenute in SGRPES. Dalle verifiche effettuate, tali incongruenze sono risultate quasi del tutto attribuibili ad errori di digitazione avvenuti nella fase di *data entry* dei questionari e risolte con questo secondo passo di standardizzazione e geo-codifica degli indirizzi.

Per gli individui, invece, che, al momento del Censimento, avevano dimora abituale in un altro comune, rispetto a quello in cui sono stati rilevati dalla PES (e allo stesso modo per quelli che non hanno fornito una risposta alla domanda 1.5), non è stato possibile attribuire una sezione di Censimento. Questi casi non sono entrati nel calcolo della probabilità che ha un individuo di essere sovra-coperto.

Allo stesso modo quelli che alla data del Censimento avevano dichiarato di avere dimora abituale all'estero sono stati definiti non eleggibili.

2.5 Analisi dell'eleggibilità

Il campo di osservazione della PES è un sottoinsieme di quello del 15° Censimento generale della popolazione e delle abitazioni. Sono oggetto di rilevazione, infatti, tutte le famiglie e le persone abitualmente dimoranti in alloggi nelle sezioni campione. Le famiglie, nel rispondere ai quesiti del questionario, devono riferirsi alla data del Censimento, ossia il 9 ottobre 2011. Per quanto riguarda le abitazioni, sono state conteggiate (e non censite con un questionario) quelle non occupate e quelle occupate esclusivamente da persone che non hanno dimora abituale nell'alloggio, ovvero quelle temporaneamente presenti. Nel caso di cittadini italiani, sono stati inclusi nella rilevazione quelli che hanno dimora abituale nell'alloggio, anche se temporaneamente assenti. I cittadini stranieri sono stati oggetto di rilevazione solo quelli con dimora abituale nell'alloggio e se iscritti in

anagrafe, o se in possesso di regolare titolo di soggiorno in Italia⁹¹, anche se temporaneamente assenti. La rilevazione campionaria non ha rilevato, invece, le famiglie e gli individui senza fissa dimora, i residenti in strutture residenziali collettive e in sedi diplomatiche e consolari, così come le persone e le famiglie in altro tipo di alloggio.

Pertanto, successivamente alla fase di controllo di coerenza dei singoli record, sono stati definiti i tre principali criteri di eleggibilità degli individui, eliminando quei record che erano:

1. non eleggibili per età dell'individuo;
2. non eleggibili per dimora abituale all'estero;
3. non eleggibili per assenza del tracciato del questionario.

Come prima cosa, è stata verificata l'appartenenza delle unità rilevate al campo di osservazione secondo l'età dell'individuo. Infatti, una volta revisionata la variabile anno di nascita, nonché controllata la coerenza del mese e giorno di nascita (tavola 2.4), è stato possibile stabilire l'eleggibilità dell'unità per l'indagine. A tale proposito, si ricorda infatti che sono eleggibili per l'indagine tutti e solo quegli individui che sono nati prima del 9 ottobre 2011, data di riferimento del Censimento. Relativamente a questo criterio di eleggibilità sono stati eliminati complessivamente 1.881 record, di cui per 455 individui non è stato possibile definirli eleggibili, secondo l'età, in quanto l'informazione era mancante, e 1.426 erano individui nati dopo la data di riferimento.

Per quanto riguarda la dimora abituale, invece, rientravano nel campo di osservazione coloro che erano presenti (dimoravano abitualmente) sul territorio nazionale alla data del Censimento, pertanto l'informazione rilevata attraverso la Domanda 1.5 è stata determinante per attribuire l'eleggibilità alle unità rilevate. In riferimento a questo secondo criterio, sono stati eliminati complessivamente 599 record. Per ciascuno di essi era infatti indicata alla Domanda 1.5 la modalità 2 corrispondente ad avere dimora abituale all'estero, alla data del Censimento, pur essendo al momento della PES presenti in Italia e raggiunti dalla stessa indagine.

Infine dopo l'operazione di costruzione del database, è emerso che per alcuni individui raggiunti dall'indagine PES nelle sezioni campione e registrati nel sistema SGRPES, non si disponeva delle relative informazioni provenienti dal questionario cartaceo (assenza dei tracciati), necessarie per la validazione dei record. Pertanto se l'individuo non ha volontariamente compilato il questionario, o non ha risposto a dei quesiti *core*, o ancora in fase di registrazione il questionario è andato smarrito, il rispettivo record con le sole informazioni provenienti da SGRPES non è stato validato. Per questa motivazione sono stati eliminati 636 record.

Tavola 2.4 - Eleggibilità degli individui rilevati alla PES

Criterio di eleggibilità	Variabili sottoposte a Controllo
Età individuo	Anno, Mese, Giorno di nascita
Dimora abituale	Domanda 1.5- Modalità Estero
Assenza tracciato	Presenza in SGRPES e nessuna informazione da questionario cartaceo

⁹¹ Per i cittadini di paesi non appartenenti all'Unione Europea, costituiscono titoli a soggiornare in Italia il permesso di soggiorno valido, il nulla osta all'ingresso in Italia per motivi di lavoro o di ricongiungimento familiare, la domanda di rinnovo del permesso di soggiorno e la domanda di rilascio del primo permesso. Qualora non in possesso di regolare titolo a soggiornare in Italia, i cittadini stranieri non comunitari non sono rilevati, essendo considerati come temporaneamente presenti.

Nella tavola seguente sono evidenziate le numerosità degli individui eleggibili e di quelli non eleggibili, rilevati alla PES, distinti per la motivazione di non eleggibilità.

Tavola 2.5 - Totale individui rilevato alla PES

Eleggibilità	N	%
Numero individui rilevati	332.710	100,00
Totale record validati	329.594	99,06
Totale record eliminati	3.116	0,94
di cui:		
- Non eleggibili per età di cui:	1.881	60,37
Nati dopo il 9 Ottobre 2010	1.426	45,76
Assenza informazione	455	14,61
- Non eleggibili per dimora abituale all'estero	599	19,22
- Non eleggibili per assenza tracciato	636	20,41

Come emerge dalla tavola 2.5, rispetto al totale degli individui rilevati, meno dell'1 per cento è stato definito non eleggibile. Un risultato più che accettabile considerato che, più del 45 per cento dei non eleggibili, erano individui che facevano parte correttamente della famiglia, ma non sarebbero dovuti essere registrati perché nati dopo la data di riferimento del Censimento.

2.6 I numeri della rilevazione

Nella figura 2.2 che segue sono riassunte tutte le operazioni condotte nella fase di validazione delle informazioni relative agli individui rilevati all'indagine PES, che a seguito della fase di Record Linkage, descritta nel capitolo seguente, hanno contribuito alla stima della popolazione residente alla data del 9 Ottobre 2011 sul territorio nazionale e al calcolo dei tassi di sotto-copertura e sovra-copertura.

La tavola 2.6 riporta il totale degli individui eleggibili rilevati all'indagine, distribuiti per regione; tale variabile rappresenta il livello di dettaglio territoriale in base al quale è stata richiesta la diffusione delle stime da parte di Eurostat.

Prima di procedere con l'abbinamento tra le due indagini (PES e Censimento) si è reso necessario depurare i dati da eventuali duplicazioni, come sarà descritto nel paragrafo 3.2.2. Complessivamente sono stati individuati ed eliminati 442 record duplicati dal totale dei record validati (329.594).

Figura 2.2 - Riepilogo di tutte le operazioni svolte

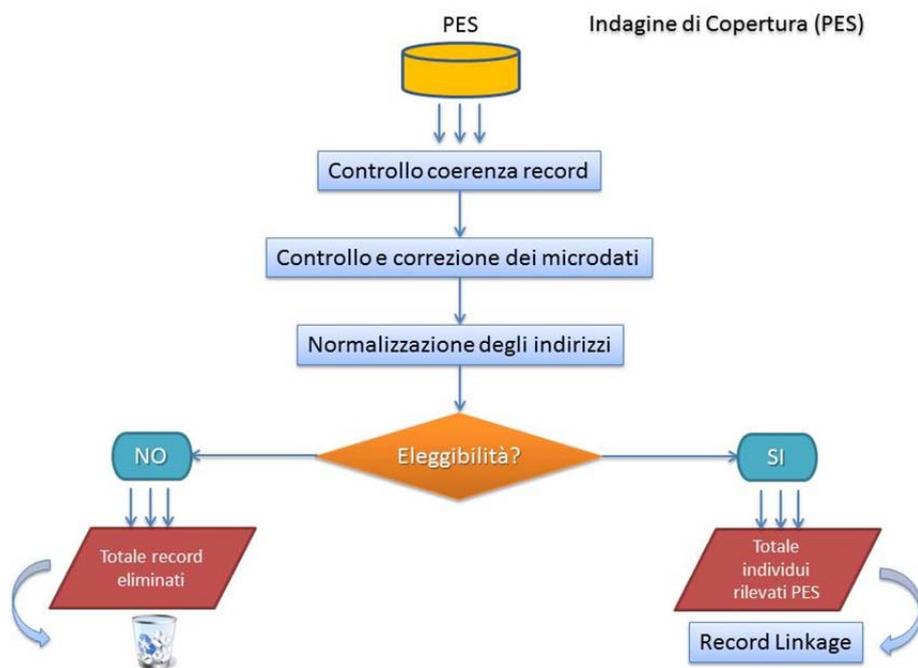


Tavola 2.6 - Totale individui eleggibili per regione

REGIONE	Numero di individui rilevati
Piemonte	21.757
Valle d'Aosta/Vallé d'Aoste	4.148
Lombardia	41.633
Bolzano	7.642
Trento	6.288
Veneto	24.386
Friuli Venezia Giulia	9.590
Liguria	9.853
Emilia Romagna	21.661
Toscana	19.351
Umbria	8.749
Marche	11.227
Lazio	22.537
Abruzzo	9.234
Molise	5.982
Campania	28.429
Puglia	22.251
Basilicata	6.559
Calabria	12.445
Sicilia	24.727
Sardegna	10.703
ITALIA	329.152

CAPITOLO 3

Le procedure di abbinamento⁹²

3.1 Quadro generale e sintesi della strategia

L'abbinamento esatto dei dati elementari del Censimento con quelli dell'indagine di copertura è un passaggio fondamentale per le stime del tasso di copertura del Censimento. Il termine "abbinamento esatto" (o "record linkage") si riferisce all'insieme di metodologie e tecniche che consentono di identificare record riferiti ad una stessa unità statistica, contenuti in archivi diversi, anche in assenza di una chiave identificativa comune, priva di errori.

Nel caso dell'indagine di copertura del 15° Censimento generale della Popolazione il processo di abbinamento ha avuto molteplici obiettivi, il principale è stato quello di determinare il numero di individui rilevati in entrambe le occasioni e, per differenza, il numero di individui sfuggiti ad una delle due rilevazioni. Inoltre, data la mobilità degli individui sul territorio, anche nel breve tempo intercorso tra le operazioni sul campo della rilevazione censuaria e quelle dell'indagine di copertura, il processo di abbinamento è stato finalizzato a riconoscere due classi di individui: quelli che in entrambe le occasioni sono stati rilevati allo stesso indirizzo e quelli che, al tempo della rilevazione sul campo dell'indagine di copertura, sono stati rilevati ad un indirizzo diverso da quello censuario. Come verrà illustrato in seguito, questa operazione è stata funzionale alla procedura per la stima della sovra-copertura e ha richiesto una complessa strategia di abbinamento. Un ulteriore obiettivo del processo di abbinamento è stato quello di ricercare, infine, le unità rilevate all'Indagine di Copertura per cui non si è trovata corrispondenza nelle unità registrate al Censimento, nelle fonti amministrative utilizzate per predisporre la lista pre-censuaria, al fine di fornire un'informazione terza e recuperata sul campo, al confronto tra il Censimento e le fonti amministrative usate in fase di progettazione. Questo aspetto verrà approfondito al paragrafo 3.6.

L'abbinamento tra i risultati definitivi del Censimento e i risultati dell'indagine di copertura è centrale per stimare i tassi di copertura del Censimento con il modello di Petersen (o dual-system o di tipo cattura-ricattura). Infatti, per stimare correttamente l'errore di sotto-copertura, la procedura di abbinamento che identifica le unità enumerate in entrambe le occasioni deve essere priva di errori, i quali, anche se molto contenuti, possono compromettere la correttezza delle stime del tasso di copertura del Censimento.

Pertanto, per assicurare la massima correttezza degli abbinamenti è stata adottata una procedura di record linkage strutturata in diverse fasi. Sono stati adottati modelli di record linkage probabilistici, iterati considerando differenti metriche per il confronto delle variabili comuni nei vari passi della procedura. I primi passi sono stati indirizzati all'identificazione degli abbinamenti più facili, per mezzo di procedure computazionali più semplici, lasciando gli abbinamenti più complessi

⁹² I paragrafi sono a cura di:

3.1 Quadro generale e sintesi della strategia: Tiziana Tuoto

3.2 Attività preliminari: Luca Valentino

3.3 I metodi di linkage: Nicoletta Cibella

3.3.1 Record linkage deterministico: Nicoletta Cibella

3.3.2 Record linkage probabilistico: Nicoletta Cibella

3.3.3 Revisione manuale dei casi ambigui e la ricerca manuale dei non abbinati: Tiziana Tuoto

3.4 RELAIS: uno strumento per l'abbinamento : Luca Valentino

3.5 Strategia di abbinamento adottata: Tiziana Tuoto

3.6 Confronto con gli archivi amministrativi e la lista pre-censuaria: Nicoletta Cibella

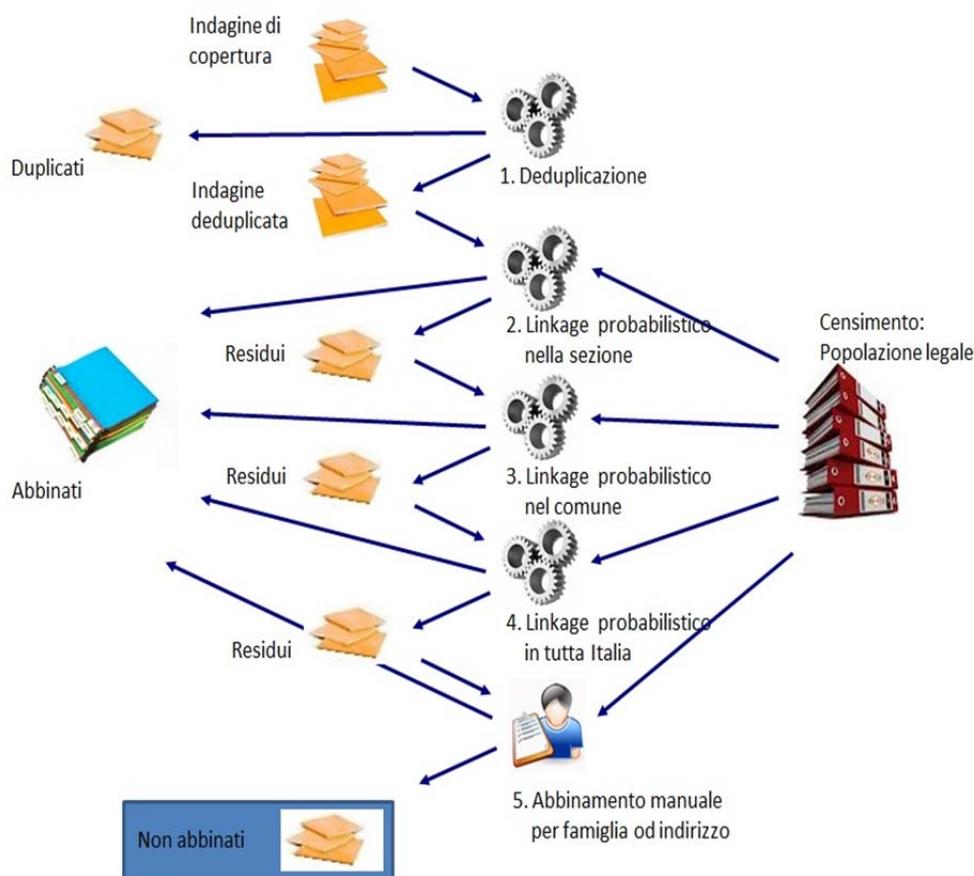
3.7 I risultati del linkage: Luca Valentino

alle fasi successive. Inoltre, un'accurata revisione manuale interattiva è stata applicata ai risultati più ambigui delle varie fasi per garantire la correttezza degli abbinamenti individuati attraverso i modelli statistici.

La complessa articolazione della strategia di linkage delineata è stata supportata e assistita dalla disponibilità e dall'utilizzo di strumenti informatici generalizzati, che implementano in maniera ottimale le sofisticate metodologie richieste.

Nella figura successiva è riportata una descrizione grafica della strategia complessiva.

Figura 3.1 - Strategia di linkage



Le procedure di abbinamento sono state realizzate in passi successivi, ad ogni passo sono stati considerati gli individui registrati all'indagine di copertura non abbinati ai passi precedenti.

I passi considerati prevedono:

1. Linkage deterministico sui soli dati dell'indagine di copertura per l'eliminazione dei duplicati;
2. Linkage probabilistico di individui all'interno della stessa sezione di Censimento campione con revisione manuale dei casi ambigui;
3. Linkage probabilistico di individui all'interno dello stesso comune campione, distinguendo i comuni metropolitani dagli altri comuni coinvolti nell'indagine di copertura, con revisione manuale dei casi ambigui;
4. Linkage probabilistico di individui sull'intero territorio nazionale
 - a. considerando la popolazione legale suddivisa in macro aree con successiva revisione manuale dei casi ambigui;

b. considerando gli individui di nazionalità straniera con successiva revisione manuale dei casi ambigui;

5. Ricerca manuale degli individui non abbinati.

Il linkage probabilistico, implementato in assenza o incompletezza del codice identificativo univoco, richiede l'applicazione di modelli statistici, basati sulla somiglianza degli attributi relativi alle unità provenienti dai due file. Tali modelli vengono sempre più spesso utilizzati sia per integrare informazioni provenienti da fonti diverse sia per eliminare duplicazioni all'interno di uno stesso file. A tal fine, già da qualche anno è disponibile in Istat il software open-source RELAIS; in questo contesto è stata utilizzata la versione di prova 3.0 che mette a disposizione le metodologie più consolidate per il record linkage probabilistico ma anche facilities per l'applicazione di innovativi metodi di riduzione dello spazio di ricerca. Alcune specifiche su questo strumento sono fornite nel paragrafo 3.4.

Per garantire la massima qualità del risultato dell'abbinamento, cioè per minimizzare l'errore di falso abbinamento e di mancato abbinamento, i casi ambigui (ossia le unità per cui il modello statistico fornisce una probabilità di abbinamento non elevata) sono stati revisionati manualmente da revisori esperti.

3.2 Attività preliminari

Prima di effettuare il processo di abbinamento è stato necessario realizzare delle delicate procedure di pre-processamento dei dati che, se non svolte, avrebbero inficiato la qualità del linkage e avrebbero potuto portare a classificare erroneamente come non abbinate coppie che si riferivano, in effetti, allo stesso individuo a causa di errori nelle variabili di linkage.

Oltre alle attività preliminari descritte nel paragrafo 2.2, prima di procedere alle operazioni di linkage vere e proprie è stato necessario operare fondamentalmente due interventi:

- ▶ la riconciliazione degli schemi
- ▶ la deduplicazione dei dati

Tali attività sono decisive per l'esito della fase di abbinamento in modo particolare quando attributi raccolti nel questionario in testo libero assumono il ruolo di variabili cardine per le procedure di linkage.

3.2.1 Riconciliazione degli schemi

L'archivio PES nasce dall'integrazione di due data set SGRPES e 'Questionario PES'; è necessario dunque scegliere da quale data set prendere le variabili da utilizzare nelle procedure di linkage.

Come già accennato nel paragrafo 2.1 a causa di alcuni disallineamenti tra i numeri progressivi degli individui nel questionario si è scelto di utilizzare solo i dati di 'Questionario PES' per le variabili individuali.

Nella tavola seguente vengono riportate tutte le variabili utilizzate nelle procedure di linkage con informazione sul data set di origine.

Tavola 3.1 -- Variabili di linkage

Variabile	Data set PES	Data set CENS	Note
Codice Provincia	SGRPES*	SGR	
Codice Comune	SGRPES*	SGR	
Sezione di Censimento	SGRPES*	SGR	
Indirizzo – via	SGRPES	SGR	
Indirizzo – num. civico	SGRPES	SGR	
Indirizzo - interno	SGRPES	SGR	
Nome	'Questionario PES'	SGR	
Cognome	'Questionario PES'	SGR	
Sesso	'Questionario PES'	SGR	
Cittadinanza	'Questionario PES'	SGR	
Anno di nascita	'Questionario PES'	SGR	
Mese di nascita	'Questionario PES'	SGR	
Giorno di nascita	'Questionario PES'	SGR	
Luogo di nascita	'Questionario PES'	SGR	
Flag ANAG	'Questionario PES'	n.d.	Indica se l'individuo alla data del Censimento risiedeva in questa abitazione in un'altra o all'estero
Indirizzo precedente	'Questionario PES'	n.d.	Eventuale indirizzo dell'abitazione di residenza alla data del Censimento

* dato verificato tramite confronto con 'Questionario PES'

3.2.2 Deduplicazione

Prima di effettuare le procedure di abbinamento tra gli individui rilevati alla PES e i record individuali del Censimento, nelle sezioni e comuni campione, si è provveduto a rimuovere i duplicati dalla indagine PES.

A tal fine ci si è serviti della strumento software Relais (le cui peculiarità sono descritte nel successivo paragrafo 3.4) ed è stata effettuata una procedura deterministica in cui venivano eliminati i record con le variabili identificative dell'unità in PES coincidenti.

In particolare, si sono eliminati dalla PES 442 individui con Nome, Cognome, Data di nascita, Sesso, Indirizzo coincidenti, sottoposti, comunque, anche a revisione manuale.

La procedura di deduplicazione è stata effettuata, ovviamente, sui soli dati dell'indagine PES in quanto la fonte censuaria era già priva di duplicazioni.

3.3 Metodi di linkage

Come già sottolineato nei paragrafi precedenti, la procedura di abbinamento è stata complessa ed è stata realizzata in diversi passi, con metodi di tipo deterministico essenzialmente per quanto concerne la deduplicazione dell'indagine PES, e con tecniche più sofisticate di tipo probabilistico, per gli abbinamenti tra la PES e il Censimento. Da ultimo si è provveduto alla ricerca manuale dei record della PES risultati non ancora abbinati. Nei paragrafi successivi si delineano brevemente le metodologie sottostanti, con specifici rimandi bibliografici per approfondimenti.

3.3.1 Record linkage deterministico

Il record linkage deterministico è stato adottato per eliminare dal file PES i record che erroneamente erano stati inclusi più di una volta nella PES, per errori nella rilevazione o nella raccolta e codifica dei questionari. Non essendo disponibile una chiave esatta come il codice fiscale nel questionario PES, si è considerata come chiave di aggancio per le coppie, al fine di identificare i duplicati, quella costituita dalla esatta coincidenza delle informazioni riportate in alcuni campi identificativi dell'unità. I file duplicati sono stati sottoposti, comunque, anche a revisione manuale. L'approccio deterministico è stato usato solo per questo passo della strategia complessiva.

3.3.2 Record linkage probabilistico

La procedura di abbinamento tra gli individui della PES e del Censimento è stata una fase complessa. Le variabili nome, cognome e data di nascita dell'individuo sono fortemente discriminanti per l'identificazione delle coppie e, quindi, potevano essere usate per implementare i metodi di record linkage probabilistico che riuscivano a discriminare bene le coppie dalle non coppie.

Nel modello probabilistico per l'individuazione degli abbinamenti, si ipotizza che la distribuzione del vettore dei confronti, γ , i cui elementi k sono il risultato del confronto tra le variabili di match comuni ai due insiemi e fortemente discriminanti per le unità, sia una mistura di due distribuzioni, una generata dalle coppie (a,b) che effettivamente rappresentano la stessa unità, distribuzione m , e una generata dalle coppie (a,b) che rappresentano unità diverse, distribuzione u . A partire dalla stima di tali distribuzioni, è possibile costruire il peso composto di abbinamento (Fellegi and Sunter, 1969), dato dal rapporto delle verosimiglianze

$$r = \frac{m(\gamma)}{u(\gamma)} = \frac{\Pr(\gamma | M)}{\Pr(\gamma | U)}$$

dove M è l'insieme delle coppie che rappresentano degli abbinamenti e U è l'insieme delle coppie che non rappresentano degli abbinamenti, con $M \cup U = \Omega$ e $M \cap U = \emptyset$. In generale, la stima dei parametri delle distribuzioni viene generalmente ottenuta per mezzo dell'applicazione dell'algoritmo EM (Jaro, 1985).

Sulla base del rapporto r , le coppie sono ordinate e sottoposte ad un processo di classificazione negli insiemi M ed U in modo tale da minimizzare gli errori di classificazione e la dimensione dell'area per cui non viene presa nessuna decisione. Le coppie che hanno una probabilità di abbinamento superiore ad una certa soglia, T_m , sono classificate come match, quelle al di sotto dalla soglia T_u come non match, per quelle con una probabilità compresa tra T_u e T_m non viene presa una decisione e sono mandate a revisione manuale.

In generale, la qualità dei risultati del processo di abbinamento può essere compromessa dalla elevata dimensione dei dati da abbinare e dalla presenza di errori nelle variabili testuali (principalmente nome e cognome) candidate ad essere variabili di matching. Tali problematiche possono essere affrontate con metodi specifici.

L'elevata dimensione dei dati causa, a sua volta, due generi di problemi: uno di tipo informatico dovuto all'eccessivo carico computazionale nella fase di confronto ed uno di tipo statistico, ossia la difficoltà di rilevare un fenomeno (quello di individuazione di M) che diventa molto raro se le coppie sono troppo poche rispetto alle non coppie. È stato in questo contesto necessario, dunque, diminuire la complessità del problema (ridurre lo spazio di ricerca di ricerca delle coppie) ed i metodi presi in considerazione sono stati essenzialmente tre: Blocking, Sorted Neighborhood e SimHash.

La presenza di errori nei valori testuali di alcuni attributi comporta confronti con esito negativo (non concordanza) a causa di minime differenze del tutto irrilevanti anche quando i due attributi sono riferiti alla stessa unità. Questi esiti possono causare un'errata classificazione della coppia dovuta a distorsioni nelle stime del modello. Il problema si supera utilizzando degli algoritmi che valutano la similitudine delle stringhe in modo che due valori molto simili anche se non identici risultino concordanti. In particolare sono state trovate utili 3 metriche Jaro, Soundex e 3grams. Di seguito una breve descrizione di tutti i metodi utilizzati.

Metodi di riduzione:

Blocking method: è il modo più semplice di realizzare una riduzione dello spazio di ricerca delle coppie, si limitano i confronti ai soli elementi delle liste, PES e Censimento, che presentano valori coincidenti per la variabile (le variabili) selezionata come variabile di blocco (numerica o alfanumerica).

Si introduce un elemento deterministico per cui tutte le unità che differiscono per la variabile di bloccaggio hanno probabilità pari a 1 di essere considerate non abbinamenti.

Sorted Neighborhood: il metodo prevede di ordinare le unità delle due liste considerate insieme (come fossero una unica lista) secondo la variabile di blocco (in questo caso più propriamente si tratta di variabile di ordinamento) e scorrere, in maniera sequenziale, l'ordinamento attraverso una finestra di ampiezza fissa w , la cui dimensione viene fissata dall'utente. I record all'interno della finestra sono confrontati tra loro e abbinati tra loro e solo queste coppie sono incluse nell'insieme delle coppie candidate. In questo caso quindi va specificata sia la variabile di blocco secondo cui fare l'ordinamento che l'ampiezza della finestra dei confronti. Questo metodo riduce il numero di confronti, rispetto al semplice bloccaggio, in quanto la scelta dell'ampiezza della finestra limita il numero dei possibili confronti per ogni record a $2w-1$.

Nel fissare l'ampiezza della finestra bisogna tener presente che si rischia di perdere veri abbinamenti se tale ampiezza è più piccola del numero di unità che presentano uno stesso valore nella variabile d'ordinamento.

Approfondimenti circa alcuni algoritmi sono disponibili in Baxter et al. (2003).

Simhash: è un metodo innovativo che si basa sull'analisi delle *impronte hash* dei valori della variabile in esame. Queste *impronte hash* sono delle stringhe di bit (ovvero concatenazione di 0 e 1) che hanno due caratteristiche fondamentali: 1) mantengono informazioni dei valori originali e 2) hanno una lunghezza fissa.

In pratica queste *impronte hash* derivano dalla combinazione di un decodifica in binario di singoli blocchi di caratteri che compongono la stringa di partenza. Confrontando due impronte con la distanza di Hamming (che dipende dalla frequenza di caratteri binari diversi nelle due sequenze) si dimostra che stringhe simili hanno distanze di Hamming basse.

Col metodo di riduzione Simhash entrano nello spazio di ricerca solo i record per i quali la distanza di Hamming della variabile di bloccaggio è inferiore ad una data soglia. Come nel metodo del Sorted Neighborhood per diminuire la complessità computazionale si prevede di ordinare preventivamente le unità delle due liste in modo da confrontare ogni record solo con quello adiacente della lista. Con il metodo Simhash l'ordinamento viene fatto sulle *impronte hash* ed inoltre, sfruttando il fatto che queste hanno una dimensione fissa, è possibile effettuare più iterazioni ordinando la lista su diverse permutazioni dell'impronta; in questo modo si rende la procedura meno sensibile alle differenze nei primi caratteri delle stringhe, rispetto al metodo del Sorted Neighborhood.

Per maggiori dettagli sulle parametrizzazioni del metodo Simhash in Relais o per approfondimenti sull'algoritmo si rimanda al manuale di Relais 3.0 (Relais 3.0 UserGuide) ed ai contributi di Mancini et al (2012) e Charikar (2002).

Metodi di valutazione della similitudine tra stringhe:

Jaro (e Jaro-Winkler): in breve, se si indicano con $s1$ e $s2$ le due stringhe da confrontare e con n_i il numero dei caratteri della stringa s_i ($i=1,2$), c il numero di caratteri comuni e t il numero di caratteri comuni trasposti, la funzione Jaro si ottiene come

$$Jaro(s1, s2) = \frac{1}{3} \left(\frac{c}{n_1} + \frac{c}{n_2} + \frac{c-t}{c} \right)$$

Una estensione molto usata della funzione di confronto Jaro è la funzione Jaro-Winkler, data da

$$JW(s1, s2) = Jaro(s1, s2) + (l * p * (1 - Jaro(s1, s2)))$$

dove p e l (numero dei caratteri comuni da inizio stringa) sono dei fattori che "premano" le stringhe più simili nei caratteri iniziali. Queste due funzioni conteggiano il numero di caratteri in comune tra le stringhe e il numero di interventi necessari a rendere uguali le stringhe.

Soundex: una funzione di confronto basata invece sulla pronuncia delle stringhe è la Soundex: alle stringhe viene associato un codice di 6 caratteri composto dal primo carattere della stringa e da 5 codici numerici che possono assumere valore da 0 a 6. Questi codici vengono assegnati in base alla classe di assonanza delle prime 5 consonanti della stringa.

3-Grams: la funzione 3-Grams invece lavora sui trigrammi che compongono la stringa, ossia sui sottoinsiemi di 3 caratteri consecutivi di una stringa. La funzione 3-Grams applicata alle stringhe $s1$ e $s2$ conteggia essenzialmente il numero dei trigrammi comuni sul numero totale di trigrammi di $s1$ e $s2$.

3.3.3 Revisione manuale dei casi ambigui e la ricerca manuale dei non abbinati

Come sottolineato più volte, l'abbinamento tra i dati del Censimento e dell'indagine di Copertura deve essere il più accurato possibile in quanto le stime del tasso di copertura del Censimento vengono prodotte adottando il modello cattura-ricattura che ipotizza che le procedure di abbinamento, usate per identificare le unità che sono state rilevate ad entrambe le rilevazioni, siano prive di errori; violazioni anche minime di questa ipotesi hanno impatti consistenti sulle stime prodotte.

Per tali ragioni, per realizzare l'abbinamento tra gli individui rilevati al Censimento e gli individui rilevati all'indagine di Copertura è stata adottata una procedura in più passi che si avvale al meglio della presenza di variabili con alto potere discriminante e della qualità delle informazioni presenti nelle due fonti. Tutte le coppie con probabilità di abbinamento non elevata, risultanti dai vari modelli probabilistici adottati nei diversi passi procedura di abbinamento, sono state sottoposte al lavoro di alcuni revisori esperti al fine di ridurre al minimo l'errore di linkage.

Gli esperti di dominio hanno selezionato un insieme di variabili comuni alle due fonti (Censimento e PES) ritenendole le più significative e discriminanti per giudicare se due record corrispondono effettivamente alla stessa unità di rilevazione. In particolare le variabili scelte come più rilevanti per gli individui sono: il nome e il cognome, la data di nascita, l'indirizzo comprensivo di tutti i campi che lo costituiscono, il comune e la provincia in cui l'individuo è stato rilevato ma tutte le variabili disponibili per la coppia concorrono a stabilire se si tratta, o meno, dello stesso individuo.

La revisione manuale consiste nel considerare i due record che compongono la coppia proposta, uno per ogni database di partenza, ed esaminare la coppia per confermare la decisione che si tratta di un corretto abbinamento o invece per decidere che si tratta di un non abbinamento.

Poiché lo scopo della revisione manuale è quello di determinare il vero stato di abbinamento della coppia, la decisione dovrebbe essere presa esclusivamente sulla base delle evidenze, senza applicare il 'principio di precauzione' per cui in caso di dubbio una coppia è dichiarata un non abbinamento. Per far sì che le decisioni prese dai revisori manuali non siano influenzate dai risultati ottenuti con le procedure statistiche, le coppie mandate a revisione non sono accompagnate dalla probabilità di corretto abbinamento risultata dal modello probabilistico e ai revisori non viene fornita informazione riguardo al fatto che la loro ricerca è finalizzata all'individuazione dei falsi positivi piuttosto che dei falsi negativi. In questo modo si cerca di garantire l'indipendenza del risultato della revisione manuale.

Il lavoro di revisione manuale è stato inoltre monitorato attraverso la riassegnazione di campioni di coppie da valutare a revisori diversi che lavorano in modo indipendente.

La revisione manuale è un processo diverso e non deve essere confuso con la ricerca manuale degli abbinamenti. Mentre la revisione manuale ha accompagnato i risultati dei modelli probabilistici applicati in ogni passo della complessa procedura di linkage, la ricerca manuale è stata applicata all'ultimo passo della strategia complessiva. La ricerca manuale consiste nel prendere un record non abbinato e cercarlo nel file per abbinarlo. La revisione manuale, invece, prende sempre e solo in considerazione coppie di record.

La qualità della procedura di linkage dipende dalla precisione delle decisioni prese nella revisione manuale. Pertanto, i revisori sono stati istruiti appositamente per questo lavoro e non sono state fatte pressioni per velocizzare i risultati.

3.4 RELAIS: uno strumento per l'abbinamento

RE.L.A.IS (Record linkage At Istat) è un software che nasce nel 2006 in Istat per rispondere alla necessità di avere uno strumento generalizzato per la fase di record linkage, ovvero per riconoscere se due record appartenenti a due diverse fonti di dati descrivono la stessa entità del mondo reale. RELAIS è un progetto open source, rilasciato con licenza EUPL. Tale scelta nasce dalla possibilità di riutilizzare soluzioni già disponibili nella comunità scientifica. Da un punto di vista tecnologico, RELAIS è stato implementato usando tecnologie open source quali Java e R, come linguaggi di programmazione, e MySQL, come sistema di gestione delle basi di dati.

RELAIS si basa sull'idea che un progetto di record linkage non è un processo rigido, ma deve essere costruito dinamicamente sulla base dei vincoli specifici della singola applicazione. In particolare, un processo di record linkage è suddivisibile nelle seguenti fasi: (i) pre-processamento; (ii) creazione dello spazio di ricerca; (iii) scelta delle variabili di Match; (iv) scelta delle funzioni di confronto; (v) scelta del modello decisionale; (vi) applicazione di vincoli sul linkage (linkage 1:1). Per ognuna di queste fasi RELAIS mette a disposizione diverse tecniche che possono essere diversamente combinate, consentendo di ottenere specifici workflow per ogni applicazione. La creazione dei workflow è facilitata dalla presenza di una semplice interfaccia grafica e dall'accessibilità dei menù delle diverse tecniche, resi disponibili solo quando le fasi propedeutiche si sono concluse con successo. Per facilitare ulteriormente l'utente nella costruzione del workflow migliore, la scelta delle variabili di Matching e di bloccaggio è coadiuvata da due utilità di profiling dei dati di input.

Per effettuare la procedura di abbinamento in questo contesto, ci si è avvalsi della versione 3.0 di RELAIS in cui sono implementate le metodologie più consolidate per il record linkage probabilistico. Inoltre la versione 3.0 dispone anche di facilities per l'applicazione di innovativi metodi di riduzione dello spazio di ricerca, particolarmente rilevanti in questo contesto.

3.5 Strategia di abbinamento adottata

Come già dichiarato all'inizio di questo capitolo, per assicurare la massima correttezza degli abbinamenti è stata adottata una procedura di record linkage strutturata in diverse fasi. I primi passi, caratterizzati da procedure meno impegnative sotto il profilo computazionale, sono stati dedicati all'identificazione degli abbinamenti più semplici, ossia abbinamenti di unità rappresentate in maniera identica o in modo molto simile e rilevate in entrambe le occasioni nella porzione di territorio circoscritta dal campione areale dell'indagine di copertura. I record non abbinati nei passi iniziali sono stati processati nelle fasi successive: la ricerca degli individui rilevati all'indagine di copertura viene estesa progressivamente a tutti gli individui censiti su tutto il territorio nazionale, per evitare che errori nella codifica delle variabili di localizzazione, o la mobilità sul territorio, provochino erroneamente dei mancati abbinamenti. Per l'elevata quantità di record da processare, circa 60 milioni per quanto riguarda la rilevazione censuaria, e per minimizzare il rischio di mancati abbinamenti, in queste fasi sono state applicate in parallelo diverse strategie di riduzione dello spazio di ricerca, basate su tecniche e variabili diverse. Inoltre, nei modelli statistici le variabili comuni sono state confrontate secondo metriche meno stringenti, così da individuare anche coppie di record in cui a causa di un elevato numero di errori (tipografici o di altro genere) la rappresentazione degli individui è meno simile. Questa procedura è particolarmente utile ad esempio per identificare gli abbinamenti che coinvolgono cittadini stranieri o per quei record con una bassa qualità nelle variabili identificatrici. Infine, gli abbinamenti identificati in queste fasi successive sono stati sottoposti ad un'accurata revisione manuale interattiva, per evitare che nei risultati fossero inclusi dei falsi abbinamenti.

Come accennato nel paragrafo 3.1, i passi implementati nella strategia di abbinamento sono i seguenti e ad ogni passo vengono processati gli individui registrati all'indagine di copertura non abbinati ai passi precedenti:

1. Linkage deterministico sui soli dati dell'indagine di copertura per l'eliminazione dei duplicati;
2. Linkage probabilistico di individui all'interno della stessa sezione di Censimento campione con revisione manuale dei casi ambigui;
3. Linkage probabilistico di individui all'interno dello stesso comune campione, distinguendo i comuni metropolitani dagli altri comuni coinvolti nell'indagine di copertura, con revisione manuale dei casi ambigui;
4. Linkage probabilistico di individui sull'intero territorio nazionale
 - a. Dividendo la legale in macro aree con successiva revisione manuale dei casi ambigui;
 - b. Elaborazione specifica per individui di nazionalità straniera con successiva revisione manuale dei casi ambigui;
5. Ricerca manuale degli individui non abbinati.

Nello specifico, al primo passo il linkage con il Censimento (il 2.) è stato eseguito tra tutti i 329.152 individui dell'indagine di copertura e i gli individui rilevati al Censimento nelle stesse sezioni appartenenti al campione dell'indagine di copertura. Infatti la variabile sezione è stata utilizzata come variabile di bloccaggio per partizionare le coppie. Sull'insieme delle coppie così individuate, sono stati applicati due modelli probabilistici, entrambi basati sulle variabili nome, cognome, giorno mese anno di nascita, luogo di nascita, che si differenziano per la metrica di confronto utilizzata sulle stringhe nome e cognome (Soundex e Jaro). In questa prima fase oltre 9'000 possibili abbinamenti sono stati inviati a revisione manuale anche se le probabilità stimate dal modello di corretto abbinamento sono alte, compresa tra 0,90 e 0,95. La revisione manuale ha confermato come corretti il 98 per cento degli abbinamenti proposti.

In totale gli abbinamenti trovati utilizzando il blocco nella sezione di Censimento sono 300.557, il 91,3 per cento del totale.

A questi vengono aggiunti altri abbinamenti risultanti dal confronto delle famiglie degli abbinati. In pratica sono stati confrontati tutti gli individui non abbinati appartenenti a famiglie abbinati (ovvero per cui almeno un componente è stato abbinato al passo precedente). Tale ricerca ha coinvolto oltre 4 mila individui dell'indagine di copertura che sono stati confrontati con oltre 6 mila del Censimento. Per la ricerca di questi individui sono state applicati 3 modelli basati su tecniche diverse di riduzione dello spazio di ricerca (bloccaggio con variabile comune e Sorted Neighbourhood su variabile cognome con finestra 50). Come variabili di abbinamento sono state considerate nome, cognome, giorno mese anno di nascita e indirizzo di residenza combinate in vario modo e con funzioni di confronto Jaro per le stringhe dei nomi e 3grams per gli indirizzi. Le coppie individuate da questi modelli, anche se risultano corrette abbinamento con probabilità superiori al 0,95, sono state passate alla revisione manuale e sono risultate corrette, in media, nel 90 per cento dei casi. In questo modo sono stati abbinati ulteriori 530 individui.

Il totale degli abbinati nella sezione è dunque pari a 301.087 e sono risultati ancora da abbinare e quindi da processare al passo successivo 28.065 record.

Nella seconda fase sono stati rilasciati i vincoli territoriali, ossia gli individui dell'indagine di copertura non abbinati al primo passo sono stati cercati all'interno dell'intero comune di residenza e non solo nella sezione campione. Per gestire la complessità computazionale derivante dal fatto che nel campione sono presenti anche comuni molto popolosi (Roma, Milano, Napoli, Torino), i record del Censimento appartenenti a questi comuni sono stati trattati con metodi di riduzione dello spazio di ricerca (Sorted Neighbourhood sulle variabili cognome e nome con finestra di ampiezza 150) diversi da quelli adottati per i comuni molto piccoli (in cui lo spazio di ricerca è stato creato tramite prodotto cartesiano con l'intero comune) e da quelli scelti per l'abbinamento all'interno dei comuni di dimensioni intermedie, per cui il confronto è stato fatto all'interno di blocchi individuati dalla prima lettera del cognome. Come variabili di abbinamento sono state usate ancora nome, cognome, giorno mese anno di nascita, combinate in vario modo, e con funzione di confronto Jaro per le stringhe dei nomi. In alcuni modelli è stata usata per il confronto anche la variabile indirizzo di residenza (con metrica 3grams) che ha permesso di individuare coppie di record per cui la codifica della sezione risulta diversa pur in coincidenza dello stesso indirizzo in formato testo.

Molti degli abbinamenti individuati in questa fase, selezionati sulla base della probabilità di abbinamenti sono stati sottoposti a verifica manuale, con i risultati riportati in tavola 3.2.

Tavola 3.2 - Risultati complessivi del processo di abbinamento, con dettaglio sui casi revisionati

Tipologia del comune campione per dimensione	Numero di comuni campione	Abbinamenti dichiarati validi dalla procedura	Abbinamenti incerti passati a revisione	Percentuale abbinamenti validi su quelli passati a revisione	Totale abbinamenti validi trovati
Comuni piccoli	127	2.630	968	36.6%	2.984
Comuni medi	88	8.866	296	91.9%	9.138
Comuni medio-grandi	27	4.280	422	44.3%	4.467
Comuni grandi	9	1.166	172	44.2%	1.242
Roma			2.413	30.3%	732
Milano			1.054	30.1%	317
Napoli			1.547	14.7%	227
Torino			834	17.0%	142
Totale					19.249

In questo secondo passo sono stati individuati ulteriori 19.249 per un totale di 320.336 abbinamenti mentre vengono passati alla fase successiva 8.816 individui per cui non è stato ancora possibile individuare il corrispondente elemento al Censimento.

Nel successivo terzo passo della procedura di linkage la ricerca degli abbinamenti è stata estesa a tutta Italia. Anche in questo caso non è stato possibile eseguire l'abbinamento in un unico processo a causa della dimensione eccessiva del data set del Censimento. Infatti tutti e 60 milioni di individui devono essere confrontati con i quasi 9 mila residui senza condizioni di blocco a livello territoriale. Si è preceduto preliminarmente a suddividere la popolazione italiana in blocchi da 5 milioni circa di record. La divisione è stata fatta principalmente su base regionale e si è tentato di abbinare gli 8.816 residui con tutti i blocchi ottenuti. Nei vari gruppi sono stati sperimentati modelli di linkage alternativi e alla fine è stato selezionato il più efficace che usa come metodo di riduzione dello spazio di ricerca il blocking-union sui primi tre caratteri del cognome e primi tre caratteri del nome, aggiungendo la variabile sesso nei gruppi di regioni più popolose, e come variabili di matching nome, cognome, giorno mese anno di nascita combinate in vario modo e con funzione di confronto Jaro per le stringhe. Gli abbinamenti individuati da questi modelli sono stati mandati tutti a revisione manuale, anche perché caratterizzati da probabilità finali di corretto abbinamento molto più basse rispetto ai passi precedenti. I risultati sono presentati in tavola 3.3.

Tavola 3.3 - Risultati complessivi del processo di abbinamento con dettaglio sulle regioni

Macro-aree	Numero di abbinamenti individuati dai modelli	Numero di abbinamenti confermati dalla revisione	Percentuale di abbinamenti corretti su quelli revisionati
Toscana Umbria Marche	753	442	58,7
Piemonte (tranne Torino) Val d'Aosta	487	328	67,4
Lombardia 1	478	355	74,3
Lombardia 2 (tranne Milano)	546	306	56,0
Veneto	438	404	92,2
Liguria Trentino Friuli Sardegna	740	520	70,3
Emilia Romagna	544	350	64,3
Lazio (tranne Roma)	271	189	69,7
Abruzzo Molise Basilicata Calabria	595	313	52,6
Campania (tranne Napoli)	746	327	43,8
Puglia	409	184	45,0
Sicilia	578	320	55,4
Totale	6.585	4.038	61,3

I 4.038 abbinamenti individuati in questa fase vanno decurtati di 76 unità, che corrispondono a individui dell'indagine di copertura per cui è stato trovato un legame con più di un individuo al Censimento, registrato in regioni diverse. Anche questi casi sono stati risolti con revisione manuale. Alla fine gli abbinati della fase tre sono 3.962.

Il quarto passo della procedura di linkage è stato dedicato specificatamente alla popolazione degli stranieri residenti in Italia, che costituiscono tradizionalmente una popolazione difficile da catturare e anche difficile da abbinare per errori presenti nelle variabili identificative. A tal fine sono stati selezionati dal Censimento tutti i cittadini di nazionalità straniera, circa 4 milioni di record, e

confrontati con tutti i record relativi a individui stranieri dell'indagine di copertura non ancora abbinati dopo i passi precedenti, poco più di 2.300 record. Per la riduzione dello spazio di ricerca è stato usato il metodo Simhash, che si è rivelato particolarmente utile in circostanze analoghe. Come variabili di match sono state utilizzate nome, cognome, sesso, e data di nascita, combinati in vario modo con funzione di confronto Jaro. Anche gli abbinamenti proposti da questi modelli sono stati passati a revisione manuale e hanno permesso di recuperare 239 nuovi abbinamenti.

Infine, i record rimasti non abbinati dopo l'esecuzione dei passi precedenti, 4.615, sono stati ricercati manualmente all'interno dell'intero database del Censimento, esaminando principalmente i record corrispondenti ad individui collocati allo stesso indirizzo di quello registrato all'indagine di copertura ma non ancora abbinati oppure quelli molto simili nelle variabili *nome* e *cognome*. Questa procedura ha permesso di recuperare ulteriori 677 abbinamenti che non era stato possibile individuare attraverso i modelli statistici applicati nei passi precedenti.

3.6 Confronto con gli archivi amministrativi e lista pre-censuaria

L'indagine di copertura è, come sottolineato più volte nel corso dei precedenti paragrafi, una indagine di alta qualità, indipendente dal Censimento ed è una indagine areale, che non parte, quindi, da nessuna lista di individui ma il rilevatore si reca sul territorio e rileva gli individui presenti. Per queste motivazioni, l'indagine di copertura è una indagine di notevole importanza per l'Istat, considerando anche che l'istituto sta progettando una grande rivoluzione per il prossimo Censimento che prevede un uso importante delle informazioni provenienti dagli archivi amministrativi, anche per minimizzare il disturbo sui rispondenti e i costi da sostenere per il Censimento, tra le altre motivazioni. Il Censimento della popolazione del 2011 è stato il primo Censimento della popolazione in Italia assistito da lista; ossia le informazioni presenti nelle liste anagrafiche comunali, le Lac, che rilevano i cittadini residenti in Italia, acquisite centralmente dall'Istat a partire dal 2011 con cadenza annuale, sono state adottate in diverse fasi del processo censuario:

- ▶ in fase di progettazione, per testare l'intero processo e valutare la bontà dell'informazione contenuta nei registri anche attraverso l'indagine pilota;
- ▶ come lista di partenza per la rilevazione censuaria vera e propria (agli individui in Lac sono stati inviati i questionari di Censimento);
- ▶ per verificare l'andamento della rilevazione censuaria e individuare la popolazione che aveva cambiato indirizzo di residenza rispetto alla Lac di partenza per il Censimento,
- ▶ per il processo di revisione anagrafica post-censuaria.

In questa ottica si è ritenuto utile analizzare i dati degli individui rilevati alla copertura e vedere se fossero stati rilevati o meno oltre che al Censimento, come richiesto dal il modello cattura-ricattura, anche nelle liste Lac.

In tavola 3.4, si riportano i dati, non pesati, del confronto tra gli individui rilevati alla PES, con la loro presenza o meno al Censimento e nelle Lac.

Considerando, quindi, pari a 100 il numero di individui rilevati alla PES, si riscontra che il 97,88 per cento di questi è presente sia al Censimento che nelle Lac (anche se non sempre allo stesso indirizzo), mentre lo 0,92 per cento degli intervistati è presente al Censimento ma non è iscritto in anagrafe. Similmente, lo 0,83 per cento degli individui rilevati alla PES era presente nelle Lac ma è non è stato trovato al Censimento. Infine, lo 0,37 per cento degli individui della PES non risulta né nelle Lac né al Censimento 2011 ma una parte di questi è stato trovato in altre fonti amministrative o a seguito della procedura di revisione tra Censimento ed anagrafe.

Tavola 3.4 - Classificazione degli individui rilevati alla PES, secondo la presenza o meno in Censimento e Lac

		Lac		
		si	no	
Censimento	si	322.178 (97,88%)	3.036 (0,92%)	
	no	2.733 (0,83%)	1.205 (0,37%)	
				329.152

Importante è considerare anche la localizzazione degli individui rilevati in più fonti, non solo se sono presenti, quindi, ma anche se coincidono gli indirizzi a cui vengono rilevati; tale informazione è importante quando la lista di partenza per la rilevazione degli individui è la fonte anagrafica-amministrativa e fornisce indicazioni sulla qualità delle informazioni presenti nella fonte.

In particolare, per gli individui eleggibili alla PES si ha che :

- ▶ 320.099 (99,35 per cento) rilevati allo stesso indirizzo delle Lac
- ▶ 2.079 (0,65 per cento) rilevati in altro indirizzo rispetto alle Lac
- ▶ la quasi totalità degli irreperibili al Censimento erano in PES allo stesso indirizzo delle Lac

L'analisi è stata poi approfondita e gli individui non trovati al Censimento sono stati abbinati, attraverso l'uso del software RELAIS, con l'archivio PES, con procedure di record linkage probabilistico. Si sono abbinati 4.360 individui. Il confronto è stato poi arricchito (Cibella e Tuoto et al, 2015) con le risultanze del processo di revisione anagrafica post-censuaria, presenti su una piattaforma informatica denominata SIREA (Sistema Revisione Anagrafica) e si sono anche considerate le informazioni derivanti da altri archivi amministrativi di fonte non anagrafica, organizzate in Istat in un Sistema Integrato di Microdati, denominato SIM (Istat, 2012). In SIM confluiscono diversi archivi tra cui le persone decedute risultanti nella base dati dell'Anagrafe Tributaria al 31.12.2011, gli iscritti all'Anagrafe Consolare all'8 ottobre 2011, le persone che hanno partecipato al mercato del lavoro in Italia, con versamenti contributivi all'INPS nel corso del 2011, i permessi di soggiorno validi al 31.12.2011, le persone per cui risulta un trasferimento di residenza, etc. Tale sistema permette di rinvenire, ad esempio, "segnali" di conferma dell'irreperibilità al Censimento su chi, per varie ragioni, non è stato censito e segnali di presunta assenza sul territorio su chi, pur essendo stato "recuperato" a seguito della revisione anagrafica post-censuaria, non è presente nelle fonti integrative utilizzate. I risultati del confronto sono riportati in tavola 3.5.

In tavola 3.5 è possibile avere un primo riscontro sulla qualità delle fonti amministrative con il cui ausilio si riesce a coprire la gran parte degli individui non trovati al Censimento, 3.512 (oltre l'80 per cento) hanno, infatti, almeno un segnale di presenza nelle fonti amministrative. Coloro che non presentano segnali ma sono in PES, 848 individui, vivono in famiglie sfuggite interamente al Censimento e 28 per cento si questi vive in famiglie monocomponenti. L'informazione sulla qualità delle Lac può essere desunta dal confronto con gli individui che non sono stati rilevati al Censimento ma erano in Lac, la metà di questi (51 per cento) ha effettuato un trasferimento di residenza a ridosso del Censimento. Vanno, di contro, analizzati meglio i non censiti che confermano la dimora abituale (62 per cento dei non trovati).

L'indagine campionaria areale di qualità è una ricchezza per la valutazione delle rilevazioni tradizionali ma anche degli archivi amministrativi e permette anche di mettere in luce criticità relative alla mappatura del territorio di alcune tipologie di popolazione particolarmente difficili da rilevare.

Tavola 3.5 - Esiti del confronto tra gli individui PES, non trovati al Censimento, i risultati della revisione censuaria (SIREA) e i segnali di presenza nel Sistema integrato dei microdati (SIM) al 31.12.2011. Valori assoluti e percentuali.

	PES		Totale
	Non trovati al Censimento	Non trovati al Censimento ma presenti in Lac	
Totale	2.973	1.387	4.360
di cui % su SIREA :			
Cancellati per irreperibilità	8,8	0,9	
Cancellati per trasferimento di residenza	27,0	51,0	
Non censiti che confermano dimora abituale	62,0	13,3	
Errori di lista	2,1	34,8	
Totale	100,0	100,0	
Nessun segnale in SIM	827 (27,8%)	21 (1,5%)	848
Almeno un segnale in SIM	2.146(72,2%)	1.366(98,5%)	3.512

Ulteriori analisi sono state effettuate per approfondire maggiormente le caratteristiche degli individui non censiti (luogo di residenza, presenza in più fonti alternative, ecc.), soprattutto per evidenziare se è possibile individuare, tra i segnali forniti dalle varie fonti, uno o più profili che permettano di identificare gli individui che sfuggono come residenti in un Comune (Cibella, Tuoto et al, SIEDS 2015).

3.7 I risultati del linkage

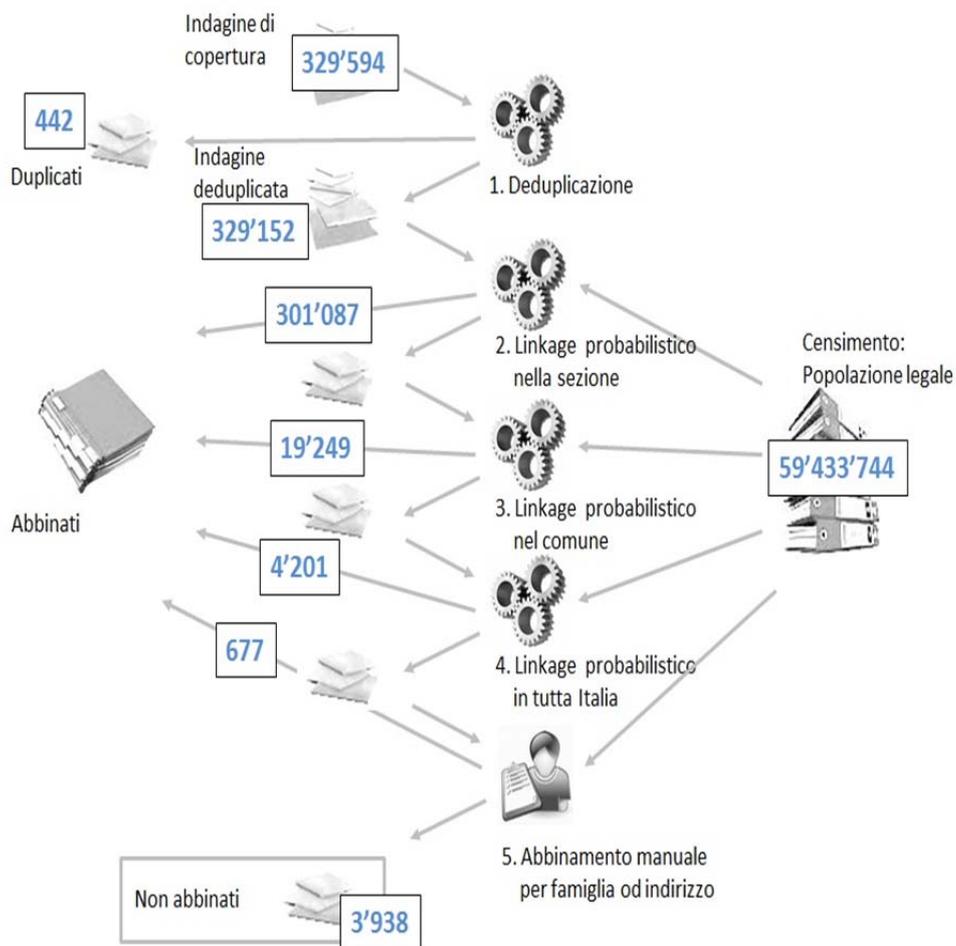
Nella tavola successiva sono richiamati i 5 passi della procedura di linkage, indicando le coppie individuate in ogni fase e le percentuali rispetto ai data set di partenza e al totale degli abbinati.

Tavola 3.6 - Sintesi dei risultati della procedura di abbinamento

Passo della procedura	Numero abbinamenti	Percentuale sui processati	Percentuale sul totale abbinati
1. Intra-sezione	301.087	91,47%	92,58%
2. intra-comune	19.249	68,59%	5,92%
3. su tutta Italia	3.962	44,94%	1,22%
4. stranieri	239	10,11%	0,07%
5. ricerca manuale	677	14,67%	0,21%
Totale abbinati	325.214	98,80%	
Non abbinati	3.938	1,20%	

Il risultato finale della procedura di abbinamento è che 3.938 individui rilevati con l'indagine di copertura non ha un corrispettivo tra la popolazione rilevata al 15° Censimento generale della Popolazione, e questi rappresentano 1,20 per cento del totale dei rilevati.

Figura 3.2 - Sintesi dei risultati della procedura di abbinamento



CAPITOLO 4

I risultati della PES⁹³

Nel presente capitolo sono illustrati i risultati dell'indagine per livello territoriale, ossia dapprima il nazionale e, nel paragrafo successivo, il dettaglio regionale. Oltre ai risultati riferiti a questi domini, pianificati in fase di progettazione dell'indagine, sono riportate le stime della popolazione e i tassi di sovra e sott-copertura riferiti ad alcune variabili sociodemografiche quali l'età e la cittadinanza. Nell'ultimo paragrafo sono riportati gli errori campionari sotto forma di coefficiente di variazione per la stima della popolazione 'vera' e di intervallo di confidenza per i tassi di sotto-copertura e sovra-copertura.

Al fine di agevolare la lettura delle tabelle si fornisce una spiegazione dettagliata delle variabili e dei tassi presenti nelle tabelle (si ricorda che tutti i valori sono da intendersi al netto delle convivenze).

“N Censimento” è la popolazione enumerata al 15° Censimento generale della popolazione e delle abitazioni; “N stimato al lordo della sovra-copertura” è la popolazione stimata dalla PES dalla quale non è stata decurtata la sovra-copertura; “N stimato al netto della sovra-copertura” è la popolazione stimata dalla PES decurtata dalla sovra-copertura; “Sotto-copertura lorda” è la stima della quota di popolazione “sfuggita” al Censimento dal quale non è stata decurtata la sovra-copertura; “Sovra-copertura” è la stima della quota di popolazione sovra coperta al Censimento; “Sotto-copertura netta” è la stima della quota di popolazione “sfuggita” al Censimento dalla quale è stata decurtata la quota di sovra-copertura; “Tasso di sotto-copertura lorda” è il rapporto percentuale tra la popolazione stimata dalla PES dalla quale non è stata decurtata la sovra-copertura e la popolazione stimata dalla PES dalla quale non è stata decurtata la sovra-copertura; “Tasso di sovra-copertura” è il rapporto percentuale tra la stima della quota di popolazione sovra coperta al Censimento e la popolazione stimata dalla PES dalla quale non è stata decurtata la sovra-copertura; “Tasso di sotto-copertura” è la differenza tra il tasso di sotto-copertura lorda e il tasso di sovra-copertura.

Ovviamente il tasso di sotto-copertura rappresenta l'errore di enumerazione del Censimento.

4.1 Livello nazionale

La tavola 4.1 presenta i risultati generali della PES a livello nazionale. La popolazione stimata dalla PES al lordo della sovra-copertura è pari a circa 60 milioni di unità; sottratta la sovra-copertura (228.855 unità) si arriva ad una popolazione netta pari a 59.774.142 unità. Se da quest'ultima quantità si sottrae la popolazione enumerata al Censimento si ottiene che la sotto-copertura, ossia le unità che la rilevazione primaria non ha contato, è pari a 642.097 individui pari ad un tasso di sotto-copertura dell'1,07 per cento.

La tavola 4.2 presenta i risultati dell'indagine di copertura per cittadinanza (italiana e straniera). La PES stima circa 4,5 milioni di stranieri a fronte di circa 4 milioni enumerati al Censimento. Il dato maggiormente significativo è certamente la componente di sotto-copertura netta dovuta agli stranieri: di 642.097 individui non enumerati al Censimento circa 500 mila sono stranieri; ciò evidenzia la difficoltà di cattura di questa sacca di popolazione. Infatti, gli individui di cittadinanza italiana che non sono stati “contati” dal Censimento sono appena 144.888 su 55.280.620 pari ad

⁹³ I paragrafi sono tutti a cura di Matteo Mazziotta ad eccezione del paragrafo:
 4.4 La PES in Italia, i confronti nel tempo a cura di Antonella Bernardini.

un tasso di sotto-copertura dello 0,26 per cento. Il tasso di sotto-copertura netta degli individui di cittadinanza straniera è pari all'11,07 per cento. Un altro valore interessante è la sovra-copertura dei cittadini stranieri che risulta essere pari solo a 8.204 unità: ciò significa che se è difficile contare una volta gli stranieri è certamente molto difficile contarli più volte.

Tavola 4.1 - Stima della popolazione e tassi di sovra e sotto copertura (Italia)

N Censimento (a)	N stimato al lordo della sovra-copertura (b)	N stimato al netto della sovra-copertura (c)	Sotto-copertura lorda (d=b-a)	Sovra-copertura (e)	Sotto-copertura netta f=(d-e)	Tasso di sotto-copertura lorda h=(d/b)*100	Tasso di sovra-copertura i=(e/b)*100	Tasso di sotto-copertura (h-i)@(f/b)*100
59.132.045	60.002.997	59.774.142	870.952	228.855	642.097	1,45	0,38	1,07

Tavola 4.2 – Stima della popolazione e tassi di sovra e sotto copertura per cittadinanza (Italia)

Cittadinanza	N Censimento (a)	N stimato al lordo della sovra-copertura (b)	N stimato al netto della sovra-copertura (c)	Sotto-copertura lorda (d=b-a)	Sovra-copertura (e)	Sotto-copertura netta f=(d-e)	Tasso di sotto-copertura lorda h=(d/b)*100	Tasso di sovra-copertura i=(e/b)*100	Tasso di sotto-copertura (h-i)@(f/b)*100
Italiana	55.135.732	55.501.271	55.280.620	365.539	220.651	144.888	0,66	0,40	0,26
Straniera	3.996.313	4.501.726	4.493.522	505.413	8.204	497.209	11,23	0,16	11,07
Totale	59.132.045	60.002.997	59.774.142	870.952	228.855	642.097	1,45	0,38	1,07

La tavola 4.3 presenta le stime della popolazione e i tassi di sovra e sotto-copertura per classi di età. I risultati recitano che è molto più semplice enumerare la popolazione compresa nella classe 50-64 (tasso di sotto-copertura pari allo 0,16 per cento) e quella compresa nella classe 65 e più (0,06 per cento). Risulta più difficile contare i bambini (1,50 per cento) a causa della difficoltà di rapportare la data di nascita a quella di riferimento del Censimento (9 ottobre 2011), i più giovani (classe di età 15-29) che devono spostarsi per lavorare e studiare e tutti gli individui in piena età lavorativa (30-49) che presentano un tasso di sotto-copertura pari al 2,23 per cento.

Tavola 4.3 - Stima della popolazione e tassi di sovra e sotto copertura per classi di età (Italia)

Classi di età	N Censimento (a)	N stimato al lordo della sovra-copertura (b)	N stimato al netto della sovra-copertura (c)	Sotto-copertura lorda (d=b-a)	Sovra-copertura (e)	Sotto-copertura netta f=(d-e)	Tasso di sotto-copertura lorda h=(d/b)*100	Tasso di sovra-copertura i=(e/b)*100	Tasso di sotto-copertura (h-i)@(f/b)*100
0-14	8.324.689	8.455.424	8.451.234	130.735	4.190	126.545	1,55	0,05	1,50
15-29	9.178.167	9.413.645	9.387.954	235.478	25.691	209.787	2,50	0,27	2,23
30-49	17.864.823	18.262.104	18.144.728	397.281	117.376	279.905	2,18	0,63	1,54
50-64	11.576.914	11.653.406	11.596.015	76.492	57.391	19.101	0,66	0,49	0,16
65 e più	12.187.452	12.218.418	12.194.210	30.966	24.208	6.758	0,25	0,20	0,06
Totale	59.132.045	60.002.997	59.774.142	870.952	228.855	642.097	1,45	0,38	1,07

4.2 Livello regionale

I risultati a livello regionale (pubblicati per la prima volta in Italia) confermano le difficoltà nelle regioni in cui sono presenti i grandi centri urbani, sia dal punto di vista della sotto-copertura che della sovra-copertura. Sembra costituire un caso particolare l'Abruzzo poiché, alla data del Censimento, sussistevano notevoli problemi di localizzazione degli individui legati al recente terremoto. I maggiori tassi di sotto-copertura si presentano in Toscana (1,69 per cento), Lazio (1,60 per cento), Liguria (1,30 per cento), Puglia (1,27), Piemonte e Sicilia (1,17 per cento), Emilia Romagna (1,03 per cento), Lombardia e Sardegna (appena sotto il tasso nazionale). In Molise (0,17 per cento), Basilicata (0,26 per cento), Friuli Venezia Giulia (0,32 per cento) e Valle d'Aosta (0,36 per cento) il Censimento presenta degli errori di copertura molto bassi. In merito ai tassi di sovra-copertura, la Lombardia e la Puglia (0,67 per cento), il Lazio (0,47 per cento), il Veneto (0,46 per cento), Trento (0,45 per cento) e la Calabria (0,43 per cento) presentano cifre più alte rispetto alle altre regioni.

Tavola 4.4 - Stima delle popolazioni e tassi di sovra e sotto copertura (regionale)

Regione	N Censimento (a)	N stimato al lordo della sovra-copertura (b)	N stimato al netto della sovra-copertura (c)	Sottocopertura lorda (d=b-a)	Sovra-copertura (e)	Sottocopertura netta f=(d-e)	Tasso di sottocopertura lorda h=(d/b)*100	Tasso di sovra-copertura i=(e/b)*100	Tasso di sottocopertura (h-i)@ (f/b)*100
Piemonte	4.330.669	4.388.399	4.382.103	57.730	6.296	51.434	1,32	0,14	1,17
Valle d'Aosta	125.986	126.782	126.444	796	338	458	0,63	0,27	0,36
Lombardia	9.648.023	9.805.379	9.739.184	157.356	66.195	91.161	1,60	0,67	0,94
Bolzano	499.353	504.164	502.397	4.811	1.767	3.044	0,95	0,35	0,61
Trento	518.964	524.666	522.310	5.702	2.356	3.346	1,09	0,45	0,64
Veneto	4.817.382	4.872.438	4.849.982	55.056	22.456	32.600	1,13	0,46	0,67
Friuli Venezia Giulia	1.208.411	1.214.516	1.212.259	6.105	2.257	3.848	0,50	0,19	0,32
Liguria	1.560.180	1.584.695	1.580.671	24.515	4.024	20.491	1,55	0,25	1,30
Emilia Romagna	4.317.113	4.369.882	4.362.076	52.769	7.806	44.963	1,21	0,18	1,03
Toscana	3.655.672	3.722.904	3.718.567	67.232	4.337	62.895	1,81	0,11	1,69
Umbria	879.370	886.506	885.409	7.136	1.097	6.039	0,80	0,12	0,68
Marche	1.534.536	1.545.618	1.543.295	11.082	2.323	8.759	0,72	0,15	0,57
Lazio	5.474.327	5.590.105	5.563.280	115.778	26.825	88.953	2,07	0,47	1,60
Abruzzo	1.303.335	1.337.381	1.327.144	34.046	10.237	23.809	2,55	0,75	1,79
Molise	312.530	313.377	313.052	847	325	522	0,27	0,10	0,17
Campania	5.753.564	5.835.421	5.815.646	81.857	19.775	62.082	1,40	0,34	1,07
Puglia	4.042.843	4.122.902	4.094.730	80.059	28.172	51.887	1,94	0,67	1,27
Basilicata	576.420	578.951	577.909	2.531	1.042	1.489	0,44	0,18	0,26
Calabria	1.953.284	1.971.938	1.963.319	18.654	8.619	10.035	0,95	0,43	0,51
Sicilia	4.986.669	5.056.508	5.045.604	69.839	10.904	58.935	1,38	0,21	1,17
Sardegna	1.633.414	1.650.465	1.648.761	17.051	1.704	15.347	1,03	0,10	0,93
Italia	59.132.045	60.002.997	59.774.142	870.952	228.855	642.097	1,45	0,38	1,07

4.3 Variabilità delle stime

Nella tavola 4.5 si può notare che gli errori di campionamento sono particolarmente contenuti. Il coefficiente di variazione (CV, espresso in percentuale) delle stime della popolazione (al netto della sovra-copertura) è sempre pari a valori inferiori allo 0,17 per cento. Ciò dimostra la bontà del piano di campionamento e della metodologia di stima adottata.

Tavola 4.5 - Stima degli errori campionari

Regione	N stimato al netto della sovracopertura	CV di N stimato al netto della sovracopertura (%)	Tasso di sottocopertura (%)	INF dell'intervallo di confidenza del Tasso di sottocopertura (%)	SUP dell'intervallo di confidenza del Tasso di sottocopertura (%)
Piemonte	4.382.103	0,09	1,17	1,00	1,35
Valle d'Aosta	126.444	0,11	0,36	0,15	0,58
Lombardia	9.739.184	0,11	0,94	0,72	1,16
Bolzano	502.397	0,15	0,61	0,30	0,92
Trento	522.310	0,08	0,64	0,48	0,80
Veneto	4.849.982	0,13	0,67	0,40	0,94
Friuli Venezia Giulia	1.212.259	0,07	0,32	0,17	0,46
Liguria	1.580.671	0,13	1,30	1,04	1,56
Emilia Romagna	4.362.076	0,13	1,03	0,76	1,30
Toscana	3.718.567	0,10	1,69	1,50	1,89
Umbria	885.409	0,10	0,68	0,48	0,89
Marche	1.543.295	0,10	0,57	0,36	0,77
Lazio	5.563.280	0,12	1,60	1,37	1,83
Abruzzo	1.327.144	0,09	1,79	1,62	1,97
Molise	313.052	0,11	0,17	0,00	0,38
Campania	5.815.646	0,14	1,07	0,79	1,34
Puglia	4.094.730	0,14	1,27	1,00	1,54
Basilicata	577.909	0,17	0,26	0,00	0,61
Calabria	1.963.319	0,06	0,51	0,39	0,63
Sicilia	5.045.604	0,16	1,17	0,85	1,48
Sardegna	1.648.761	0,12	0,93	0,68	1,18
Italia	59.774.142	0,06	1,07	0,96	1,19

4.4 La PES in Italia, i confronti nel tempo

La Post Enumeration Survey è stata condotta dall'Istat fin dal Censimento della popolazione e delle abitazioni del 1981. L'indagine ha subito delle profonde trasformazioni ed innovazioni, a partire dalla numerosità del campione dei comuni coinvolti.

L'indagine di copertura relativa al Censimento del 1981 ha prodotto tassi di copertura solo per l'universo delle famiglie ed ha coinvolto 66 comuni campione, rilevando 120 mila famiglie e 15 mila abitazioni non occupate.

La PES del 1991 ha coinvolto 85 comuni. È stato adottato un campione di tipo areale con la selezione dei comuni al di sopra dei 500 mila abitanti (autorappresentativi) e la stratificazione degli altri in base alla ripartizione geografica e all'ampiezza demografica. Al secondo stadio sono state selezionate con probabilità costante 638 sezioni di censimento dai comuni campione. Complessivamente sono state rilevate dall'indagine 52.844 famiglie, a cui si riferiscono i risultati, e 11.157 abitazioni non occupate.

Per l'indagine del 2001 il disegno di campionamento adottato ha invece uno schema complesso, è stato costituito dall'unione di un disegno a uno stadio stratificato nei comuni più grandi inclusi con certezza nel campione (parte autorappresentativa), in cui si sono estratte direttamente le sezioni di censimento, e da un disegno a due stadi stratificato (parte non autorappresentativa) in cui al primo stadio si è selezionato un campione di comuni e al secondo stadio, per ciascun comune, si è selezionato un campione di sezioni. In entrambi i casi tutte le famiglie appartenenti alle sezioni campione sono state enumerate. Complessivamente sono entrati a far parte del campione 98 comuni, di cui 14 autorappresentativi, dai quali sono state estratte 1.154 sezioni campione (di cui 1.098 con "unità eleggibili"). I comuni sono stati stratificati per ripartizione geografica e classe di dimensione demografica, che costituivano i principali domini di interesse, le unità di secondo stadio (le sezioni di censimento) sono state stratificate secondo la tipologia: centri; nuclei; case sparse/località produttive. Sono state rilevate dall'indagine oltre 68 mila famiglie.

La PES relativa al Censimento del 2011, come ampiamente descritto nel presente volume, si basa su un disegno campionario areale a due stadi stratificati in cui: il primo stadio è un campione di 252 comuni, stratificati per Regione e cinque classi di ampiezza demografica; il secondo stadio è composto da 2.500 sezioni di censimento stratificate per ampiezza demografica della sezione stessa. La PES ha intervistato un totale di 332.710 individui, dei quali circa 329 mila unità eleggibili.

È interessante analizzare i risultati nel corso del tempo, tuttavia vanno letti con estrema cautela, essi infatti risentono delle notevoli differenze nelle modalità con cui è stata svolta l'indagine nei decenni.

Tavola 4.6 - Tassi di copertura riferiti alle famiglie per ripartizione geografica e anno di censimento (valori percentuali)

Grado di copertura del Censimento	Anno Censimento			
	1981	1991	2001	2011
Tassi di copertura riferiti alle famiglie	96,45	99,10	98,51	
Tassi di copertura lorda riferiti agli individui		99,30	98,55	98,55
Tassi di copertura al netto della sovra-copertura riferiti agli individui			98,87	98,93
Tasso di sovra-copertura			0,32	0,38

* Attenzione i dati vanno letti tenendo conto delle differenti metodologie adottate nelle varie annualità

Per il 1981 i risultati si riferiscono, infatti, al campione e non sono riportati all'universo. In quella occasione, la stima della sotto-copertura è stata ottenuta abbinando le unità rilevate al Censimento e alla PES attraverso l'abbinamento dell'intestatario del Foglio di famiglia.

Nel 1991, nella fase di *record linkage* tra i dati censuari e quelli dell'indagine, gli identificativi utilizzati, poiché si riferivano alle famiglie e non agli individui, non sono risultati abbastanza discriminanti; considerando anche il fatto che come chiave identificativa non si era utilizzato il nominativo della persona di riferimento del Foglio di famiglia, ma le seguenti chiavi: provincia, comune, sezione di censimento, via, numero civico, piano, interno. In alcuni casi pertanto non è

stato possibile ottenere abbinamenti esatti e ciò sembra aver contribuito a una sottostima dell'errore di copertura.

Già nel 2001, come per l'ultima edizione, le stime si riferiscono direttamente agli individui e non solo alle famiglie. Al fine di avere identificativi altamente discriminanti per l'individuazione delle unità enumerate nelle due occasioni, evitando il più possibile errori di abbinamento, per la prima volta, si sono acquisiti i nominativi e gli indirizzi degli individui rilevati al Censimento e all'Indagine di copertura.

Per la PES del 15° Censimento generale della popolazione e delle abitazioni è stato progettato un questionario sullo schema di quello breve utilizzato in occasione del Censimento (Mod. Istat CP.1B_6P), ma in forma più semplificata, tenendo conto anche dell'esigenza dell'autocompilazione. In tal modo è stato predisposto uno strumento il più possibile autoesplicativo e comprensibile in ogni sua parte con l'obiettivo di disporre di identificativi degli individui ed informazioni completamente comparabili con quelle del Censimento. Questo è stato un grande punto di forza in quanto le informazioni disponibili hanno permesso di condurre la fase di record linkage, come illustrato nel paragrafo 2.3, priva di errori.

In Italia la sovra-copertura non è stata considerata un problema rilevante nei censimenti passati, infatti nel 2001, per la prima volta sono state prodotte stime della sovracopertura. La sua stima è risultata pari a 0,32 per cento a livello nazionale, registrando, quindi, un valore sensibilmente più basso di quello ottenuto per il tasso di sotto-copertura (1,45 per cento la sotto-copertura lorda e 1,13 per cento quella al netto della sovra-copertura). Tali stime però sono state ottenute non tramite il record linkage ma dai risultati dei questionari autocompilati in occasione dell'Indagine di copertura, attraverso una serie di domande somministrate direttamente agli individui, disponendo solo dei nominativi enumerati nelle sezioni campione per l'indagine di copertura e non di tutti i censiti (Istat, 2009). Pertanto tali stime possono ritenersi solo indicative del fenomeno e non essere considerate pienamente affidabili.

Dal 15° Censimento generale della popolazione e delle abitazioni, invece, il regolamento comunitario impone una più precisa e dettagliata stima della popolazione ed include nei vincoli anche una misurazione della sovra-copertura⁹⁴.

Come ampiamente descritto nel capitolo primo, la sovra-copertura nazionale è considerata trascurabile come conseguenza della verifica di una delle ipotesi del modello *dual system* scelto ai fini della stima dei parametri d'interesse per l'indagine in oggetto, mentre la sovra-copertura locale è generata da individui rilevati dal Censimento in un indirizzo ma che in realtà risiedono altrove; cioè, individui conteggiati nel luogo sbagliato. Sulla base del concetto di 'corretta localizzazione' dell'individuo alla data di riferimento del Censimento (9 ottobre 2011), determinata dalla PES è stato possibile calcolare la propensione di ciascun individuo ad essere sovra enumerato (*propensione alla sovra-copertura*).

⁹⁴ La PES è normata dalla Commissione Europea (Commission Regulation n. 1151/2010).

Conclusioni⁹⁵

I censimenti sono processi statistici molto complessi e costosi. Il 15° Censimento generale della popolazione e delle abitazioni ha coinvolto gli 8.092 comuni italiani e tutta la popolazione abitualmente dimorante in Italia alla data del 9 ottobre 2011, enumerando un totale di 59.433.744 unità, che costituiscono la “popolazione legale” (pubblicata su Supplemento ordinario n. 209 alla Gazzetta Ufficiale 294 del 18 dicembre 2012), ovvero il riferimento legale per il Paese valido fino alla successiva rilevazione.

Un conteggio esaustivo della popolazione sul territorio non può essere esente da errori di misurazione e di copertura, pertanto il Legislatore comunitario vincola ciascuno stato membro a fornire dettagliate stime della copertura censuaria.

L'Istat certifica l'accuratezza delle informazioni raccolte, in occasione del Censimento, attraverso la “Rilevazione campionaria di qualità della copertura del 15° Censimento generale della popolazione e delle abitazioni” (PES 2011) al fine di misurare gli errori per eccesso (tasso di sovra-copertura) e per difetto (tasso di sotto-copertura).

Tale rilevazione si basa su un campione areale tra i più grandi mai selezionati in Italia (circa 2.500 sezioni di censimento) e replica le operazioni di enumerazione delle famiglie abitualmente dimoranti nelle sezioni del campione.

Ciascuna delle fasi della PES, così come meglio descritto nella seconda parte del presente volume, è stata caratterizzata dall'applicazione di metodologie statistiche molto accurate e innovative che ne hanno esaltato non solo il profilo qualitativo ma anche aspetti di ricerca scientifica connessi al tema. Nella fase di determinazione della numerosità campionaria e di selezione del campione areale si sono sfruttate le informazioni geografiche elaborate dall'Istat nella fase di preparazione del Censimento della popolazione. Proprio l'informazione geografica è risultata strategica affinché fosse rispettata l'indipendenza tra il Censimento, assistito per la prima volta da Liste Anagrafiche Comunali (Lac), e la stessa PES 2011, assistita da archivi di indirizzi geocodificati a livello di sezione di censimento.

Per la PES il rilevatore è stato dotato di mappe cartografiche utili ad individuare con precisione i confini delle sezioni di censimento, area in cui effettuare la rilevazione, nonché punti di riferimento per la localizzazione delle abitazioni e, quindi, delle famiglie in esse presenti. Attraverso il sistema informatico SGRPES, per ciascuna sezione di censimento campionata, è stato possibile visualizzare, scaricare e stampare circa 2.500 itinerari di sezione utili al rilevatore nella percorrenza dell'area a lui assegnata. Il dettaglio di tali itinerari è stato decisamente elevato grazie all'utilizzo, limitatamente a 101 comuni, delle informazioni pervenute dalla Rilevazione dei Numeri Civici (RNC 2010). Attraverso il sistema SGRPES sono state gestite, inoltre, tutte le fasi di monitoraggio della rilevazione consentendo all'Istat una vista puntuale e sempre aggiornata dello stato di avanzamento dei lavori sul campo.

Un'accurata fase di controllo e correzione dei dati della PES 2011 ha preparato la matrice dei dati per una delle fasi maggiormente delicata di tutto il processo, il *record linkage* con il Censimento. I ricercatori dell'Istat, a questo punto, hanno fatto valere la grande esperienza maturata negli anni passati ed hanno messo a disposizione conoscenze approfondite e un software (RELAIS) implementato proprio all'interno dell'istituto.

La fase di stima è stata preceduta da un profondo studio della letteratura internazionale che ha posto i ricercatori davanti a scelte di natura metodologica molto complesse. L'approccio *dual system* (o cattura e ricattura) adottato dall'ONS è risultato il più adatto alla situazione della PES

⁹⁵ Autori: Patrizia Grossi e Matteo Mazziotta.

“italiana” e, dopo aver apportato alcune originali modifiche, è stato applicato in ciascuna delle sue numerose fasi di stima.

I risultati sono caratterizzati da un'elevata robustezza e presentano degli errori di campionamento molto contenuti.

La diffusione degli indicatori di sovra-copertura e sotto-copertura del Censimento 2011 e il loro inserimento all'interno degli ipercubi di qualità sino al livello territoriale regionale (NUTS 1 e NUTS 2), così come stabilito dal Regolamento (UE) n. 1151/2010 della Commissione Europea dell'8 dicembre 2010, sono avvenuti entro la scadenza prevista. Tali indicatori certificano l'elevata qualità del 15° Censimento generale della popolazione e delle abitazioni, hanno avvalorato, quindi, la scelta (vincente) di strutturare il censimento partendo dalle Lac e quindi da archivi amministrativi.

Infatti, le numerose innovazioni introdotte nel Censimento 2011 hanno decisamente accresciuto l'importanza della misurazione e valutazione della qualità del dato censuario. Verificare gli effetti di tali innovazioni, specialmente di quelle che hanno caratterizzato la preparazione della lista pre-censuaria, è stato considerato un obiettivo imprescindibile per l'Istat.

L'utilizzo delle fonti amministrative al fine di creare una lista valida per la conduzione di un Censimento rappresenta una novità assoluta che non rimarrà un esercizio isolato nel panorama delle statistiche sulla popolazione. L'imminente progettazione e la prossima esecuzione del censimento permanente devono necessariamente tenere in considerazione i risultati della PES che rappresentano un'esperienza fondamentale nella valutazione della qualità degli archivi amministrativi di popolazione. Proprio per questa ragione, sebbene l'obiettivo della PES sia stimare gli errori di copertura della rilevazione censuaria, è di fondamentale importanza verificare la copertura delle Liste Anagrafiche Comunali utilizzate al Censimento e con le quali è stato eseguito il confronto censimento/anagrafe. La PES rappresenta uno strumento di primaria importanza per testare la qualità delle Lac e procedere alla valutazione di tali archivi amministrativi in ottica futura.

Ormai il passo è stato fatto: il prossimo censimento della popolazione si baserà esclusivamente su archivi amministrativi e abbandonerà il metodo di rilevazione classico utilizzato nelle precedenti edizioni. Le PES condotte nel corso di tanti decenni e, l'ultima edizione in particolare, hanno dimostrato quanto tali rilevazioni siano fondamentali per comprendere meglio l'origine, la natura e l'entità dell'errore di copertura. In un'ottica di uso e riuso di archivi amministrativi al fine di enumerare la popolazione con frequenza annuale, sembra imprescindibile la conduzione di indagini di qualità come la PES in cui l'applicazione di metodologie statistiche innovative, sistemi di georeferenziazione e, soprattutto, l'utilizzo dei big data possano costituire la nuova frontiera per la valutazione della qualità dei censimenti moderni.

Riferimenti bibliografici

- Abbate C., Salvucci G. 2011. Population density in a city. In Aisberg (Ed.), Spatial2 Conference: Spatial Data Methods for Environmental and Ecological Processes. Foggia (IT), 1-2 September 2011. on line
- Abbott O. 2001. Hard to Count Index, One number census steering committee. <http://www.ons.gov.uk/ons/guide-method/census/census-2001/design-and-conduct/the-one-number-census/methodology/steering-committee/key-papers/hard-to-count-index.pdf>.
- Baxter R., Christen P., Churches T. 2003. A Comparison of Fast Blocking Methods for Record Linkage, *Proceedings of the ACM Workshop on Data Cleaning, Record Linkage and Object Identification*, Washington, DC, August 2003.
- Bernardini A., Fasulo A., Terribili M.D. 2015. A model-based categorisation of the Italian municipalities, based on non-response propensity in the 2011 census, *Giornate di Studio sulla popolazione*, 11^a edizione, AISP, Palermo, 4-6 Febbraio 2015.
- Brown J., Abbott O., Taylor A. 2009. Over-count in the 2011 Census: Estimation Issues. *Proceedings of the 17th meeting of the Government Statistical Service Methodology Advisory Committee*, Office for National Statistics: pp. 40-47.
- Charikar M.S. 2002. Similarity estimation techniques from rounding algorithms in *Proceedings ACM STOC 2002*, Montreal, Quebec, Canada, pp. 380–388.
- Cibella N., Fortini M., Ichim D., Tuoto T. 2010. Record linkage methods and techniques as proposed in RELAIS in *Proceedings of Statistics Canada Symposium 2010*, Social Statistics: The Interplay among Censuses, Surveys and Administrative Data.
- Cibella N., Gallo G., Pezone A., Tuoto T. 2015. Il Censimento assistito da liste: quali risposte fornisce l'integrazione dei dati individuali?, *Giornate di Studio sulla popolazione*, 11 edizione, AISP, Palermo, 4-6 Febbraio 2015.
- Cibella N., Gallo G., Pezone A., Tuoto T. 2015. L'integrazione tra l'indagine di copertura del Censimento della Popolazione 2011 e gli archivi amministrativi. Un'analisi sugli individui e le famiglie, *Riunione Scientifica SIEDS*, Ancona-Fermo, 28-29 Maggio 2015.
- Cicchitelli G., Herzel A., Montanari G.E. 1992. *Il campionamento statistico*, Il Mulino, Bologna.
- Celant A. 1988. Nuova città, nuova campagna Spazio fisico e territorio: atti del XXIV Congresso geografico italiano. Bologna, Pàtron ed.
- Clark C. 1951. Urban population density. *Journal of the Royal Statistical Society. Series A*, 114(4), pp. 490–496.
- Comune di Roma. 1960. Roma: popolazione e territorio dal 1860 al 1960: con la distribuzione territoriale dei risultati dei censimenti. Roma: Comune di Roma. Ufficio di statistica e censimento.
- Cruciani S. 2010. Le basi territoriali. Concetti, definizioni, prodotti per i censimenti, oltre i censimenti. In *L'informazione statistica per la conoscenza del territorio e il supporto alle decisioni*. Istat. Retrieved from <http://www.istat.it/it/archivio/9986>
- Di Consiglio L., Falorsi S. 2003. Alcuni aspetti metodologici relativi al disegno dell'indagine di copertura del Censimento Generale della Popolazione 2001. Documenti Istat: 11.
- Di Pede F., Lipizzi F. 2009. Nuovi strumenti territoriali per i Censimenti del 2010-2011. *Geoforus Geography and technology network*. Retrieved September 18, 2013, from

- http://www.geoforus.it/index.php?option=com_content&view=article&id=324:nuovi-strumenti-territoriali-per-i-censimenti-del-2010-2011&catid=23:archivio
- Efron B. 1979. Bootstrap Methods: Another Look at the Jackknife. *The Annals of Statistics*, 7 (1): pp. 1–26.
- Efron B., Tibshirani R.J. 1993. *An Introduction to the Bootstrap*. Chapman and Hall, New York.
- Fabbris L. 1991. Campioni di numerosità due o tre per strato selezionati con probabilità variabili: valutazione empirica di alcune stime di frequenze assolute. *Atti della giornata di studio sul campionamento statistico. Annali di statistica, Serie IX, Istat*.
- Fellegi I.P., Sunter A.B. 1969. A Theory for Record Linkage, *Journal of the American Statistical Association*, 64, pp. 1183-1210.
- Fortini M., Gallo G., Paluzzi E., Reale A., Silvestrini A. 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. Retrieved from http://www.istat.it/it/files/2011/04/2007_10.pdf
- Jaro M. 1985. Advances in Record Linkage Methodologies as Applied to Matching the 1985 Census of Tampa, Florida, *Journal of American Statistical Society*, 84(406), pp. 414-420.
- Grossi P., Mazziotta M. 2012. La progettazione dei censimenti generali 2010-2011 - 11: "La rilevazione campionaria areale di copertura e qualità del 15° Censimento generale della popolazione e delle abitazioni" Istat Working Paper n.16 del 2012. Istat. Retrieved from <http://www.istat.it/it/archivio/81920>
- Istat. 2009. La qualità dei dati – Conoscere il Censimento. 14° Censimento generale della popolazione e delle abitazioni. Roma.
- Istat. 2009a. La qualità dei dati. (Fernanda Panizon, Ed.) (Istat.). Roma: Istat. Retrieved from http://www3.istat.it/dati/catalogo/20091217_00/la_qualita_dei_dati_conoscere_il_censimento.pdf
- Istat. 2009b. Navigando tra le fonti demografiche e sociali. (Istat, Ed.). Roma.
- Istat. 2009c. <http://basiterritoriali.istat.it>. Retrieved from <http://basiterritoriali.istat.it>
- Istat. 2011. Circolare N°2, 15° Censimento generale della popolazione e delle abitazioni – Verifica e validazione delle Aree di Censimento. pubblicato on line.
- Istat. 2012. La progettazione dei censimenti generali 2010-11: "La rilevazione campionaria areale di copertura e qualità del 15° Censimento generale della popolazione e delle abitazioni", Istat Working Papers: 16.
- Istat. 2012. Esiti del confronto tra Censimento e anagrafe disponibile all' indirizzo web <http://www.istat.it/it/files/2012/12/scheda-confronto-Censimento-anagrafe.pdf>
- Large A., Brown J., Abbott O., Taylor A. 2011. Estimating and Correcting for Over-count in the 2011 Census. *Office for National Statistics Survey Methodology Bulletin*, 69 pp. 35-48.
- Lloyd P. 1986. Spazio e localizzazione: un'interpretazione geografica dell'economia ([5a ed.]). Milano: Franco Angeli.
- Madow W. G. 1949. On theory of systematic sampling. *Annals of Mathematical Statistics*, 20.
- Mancini L., Valentino L., Borrelli F., Marcone L. 2012. Record Linkage between Large Datasets: Evidence from the 15th Italian Population Census, in *Quaderni di statistica*, 05/2012, 14, pp. 149-152.
- Murthy M.N. 1967. *Sampling theory and methods*. Statistical Publishing Society, Calcutta.
- Office for National Statistics. 2011. ONS, London "Predicting patterns of household non response in the 2011 Census" <http://www.ons.gov.uk/ons/guide-method/census/2011/the-2011->

ensus/processing-the-information/statistical-methodology/predicting-patterns-of-household-non-response-in-the-2011-census.pdf

- Quintano C. 2000. Giornate di studio: La qualità dell'informazione statistica. In Quaderni di discussione 19. Istituto di Statistica e Matematica facoltà DI Economia Istituto Universitario Navale Napoli. Retrieved from [http://www.economia.uniparthenope.it/Siti_docenti/cquintano/pubblicazioni/Quintano Castellano SE7.pdf](http://www.economia.uniparthenope.it/Siti_docenti/cquintano/pubblicazioni/Quintano_Castellano_SE7.pdf)
- Relais. 2015 User's guide. Version 3.0, <http://www.istat.it/it/files/2011/03/Relais3.0UserGuide.pdf>
- Renaud A. 2007. Estimation of the 2000 census of population in Switzerland: Methods and results. *Survey Methodology*, 33: 199-210.
- Salvati L., Salvucci G., Abbate C., Zitti M., Gemmiti R. 2010. Lo sprawl urbano attraverso le sue manifestazioni insediative: il caso di Roma. In AISRE (Ed.), XXXI Conferenza Italiana di Scienze Regionali. Aosta.
- Sekar C. C., Deming W.E. 1949. On a Method of Estimating Birth and Death Rates and the Extent of Registration. *Journal of the American Statistical Association*, 44 (245): 101-115.
- Statistical Commission & Economic Commission For Europe. 2006. Conference of European Statistician. In Report of the fifty-fourth plenary session. Paris: United Nations. Retrieved from <http://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/70.e.pdf>
- United Nations Economic for Europe & Statistical Office of the European Communities (Eurostat). 2006. Conference of European Statisticians Recommendations for the 2010 Censuses of Population and Housing. New York, Geneva, UN. Retrieved from http://www.worldcat.org/title/conference-of-european-statisticians-recommendations-for-the-2010-censuses-of-population-and-housing/oclc/170896230&referer=brief_results
- Wolter K. 1986. Some coverage error models for census data. *Journal of the American Statistical Association*, 81(394). Retrieved from https://www.worldcat.org/title/some-coverage-error-models-for-census-data/oclc/114594061&referer=brief_results
- Wolter K. 2007. *Introduction to Variance Estimation*. 2nd Edition. New York: Springer.

Riferimenti legislativi

Normativa della statistica ufficiale:

Legge 30 luglio 2010 n. 122, "Piano Generale di Censimento".

Programma statistico nazionale 2011-2013.

Programma statistico nazionale aggiornamento 2012-2013.

Regolamenti e prassi internazionali:

Recommendations for the 2010 Censuses of Population and Housing, Appendix II, Alternative approaches to census taking.

Regolamento Ue n. 1151/2010 della Commissione Europea dell'8 dicembre 2010 – "Attuazione del regolamento (CE) n. 763/2008 del Parlamento europeo e del Consiglio relativo ai censimenti della popolazione e delle abitazioni per quanto riguarda le modalità e la struttura delle relazioni sulla qualità e il formato tecnico per la trasmissione dei dati".

Report of the fifty-fourth plenary session, Paris, 13-15 June 2006, ECE/CES/70, <http://www.unece.org/stats/documents/2006.06.ces.htm>.

Appendice 2.A

Nelle tavole 2.A.1, 2.A.2 e 2.A.3 sono riportate il numero di comuni campione e le numerosità attese del campione in termini di individui e sezioni in totale e separatamente per parte AR e NAR, per ripartizione territoriale, regione geografica, per classe di dimensione demografica dei comuni (1=meno di 5 mila abitanti, 2=tra 5 mila e 10 mila abitanti, 3=tra 10 mila e 20 mila abitanti, 4=tra 20 mila e 100 mila abitanti, 5=oltre 100 mila abitanti e comuni metropolitani) e a livello nazionale. Nelle successive tavole 2.A.4, 2.A.5 e 2.A.6 sono illustrati i valori delle stime dei tassi di copertura ottenuti così come descritto nel paragrafo 1.4.5 utilizzando informazioni relative al 2001 (fornite dallo studio effettuato nell'ambito della già citata convenzione tra l'Istat e il Dipartimento di Scienze Statistiche dell'Università di Bologna) ed i coefficienti di variazione relativa inerenti alle stime dei tassi in oggetto per i su citati domini.

Tavola 2.A.1 - Numerosità campionarie in termini di comuni, sezioni e di individui per ripartizione

REGIONE	N° di comuni	N° atteso di individui	N° atteso di sezioni	N° di comuni AR	N° atteso di individui (AR)	N° atteso di sezioni (AR)	N° di comuni NAR	N° atteso di individui (NAR)	N° atteso di sezioni (NAR)
Nord-Ovest	59	87.208	550	8	20.161	111	51	67.047	439
Nord-Est	59	76.205	609	11	14.263	105	48	61.942	504
Centro	48	72.097	491	9	19.892	117	39	52.205	374
Sud	63	91.852	561	6	12.095	71	57	79.757	490
Isole	26	39.308	287	6	8.692	56	20	30.616	231
Italia	255	366.670	2.498	40	75.103	460	215	291.567	2.038

Tavola 2.A.2 - Numerosità campionarie in termini di comuni, sezioni e di individui per regione

REGIONE	N° di comuni	N° atteso di individui	N° atteso di sezioni	N° di comuni AR	N° atteso di individui (AR)	N° atteso di sezioni (AR)	N° di comuni NAR	N° atteso di individui (NAR)	N° atteso di sezioni (NAR)
Piemonte	16	23.221	162	2	5.305	25	14	17.916	137
Valle D'Aosta	3	4.594	54	1	2.141	17	2	2.453	37
Lombardia	32	47.102	259	4	8.959	47	28	38.143	212
Bolzano	6	7.515	71	1	1.525	4	5	5.990	67
Trento	6	7.026	65	1	1.611	8	5	5.415	57
Veneto	20	26.596	199	4	4.923	49	16	21.673	150
Friuli V. G.	9	10.651	62	2	2.791	17	7	7.860	45
Liguria	8	12.291	75	1	3.756	22	7	8.535	53
Emilia Romagna	18	24.417	212	3	3.413	27	15	21.004	185
Toscana	17	21.479	165	4	4.538	36	13	16.941	129
Umbria	7	9.233	54	2	2.459	13	5	6.774	41
Marche	10	12.072	97	1	1.442	10	9	10.630	87
Lazio	14	29.313	175	2	11.453	58	12	17.860	117

segue

Tavola 2.A.2 - Numerosità campionarie in termini di comuni, sezioni e di individui per regione

REGIONE	N° di comuni	N° atteso di individui	N° atteso di sezioni	N° di comuni AR	N° atteso di individui (AR)	N° atteso di sezioni (AR)	N° di comuni NAR	N° atteso di individui (NAR)	N° atteso di sezioni (NAR)
Abruzzo	8	10.895	83	1	1.500	9	7	9.395	74
Molise	5	6.148	49	1	816	2	4	5.332	47
Campania	19	30.453	148	1	4.927	22	18	25.526	126
Puglia	15	22.950	133	2	2.954	15	13	19.996	118
Basilicata	6	7.859	65	0	0	0	6	7.859	65
Calabria	10	13.547	83	1	1.898	23	9	11.649	60
Sicilia	17	26.762	163	4	6.398	37	13	20.364	126
Sardegna	9	12.546	124	2	2.294	19	7	10.252	105
Italia	255	366.670	2.498	40	75.103	460	215	291.567	2.038

Tavola 2.A.3 - Numerosità campionarie in termini di comuni, sezioni e di individui per classi di dimensione demografica dei comuni

CLASSI DI DIMENSIONE DEMOGRAFICA	N° di comuni	N° atteso di individui	N° atteso di sezioni	N° di comuni AR	N° atteso di individui (AR)	N° atteso di sezioni (AR)	N° di comuni NAR	N° atteso di individui (NAR)	N° atteso di sezioni (NAR)
1	51	68.260	650	0	0	0	51	68.260	650
2	48	60.706	334	0	0	0	48	60.706	334
3	54	64.106	296	1	816	2	53	63.290	294
4	60	95.787	718	2	3.017	25	58	92.770	693
5	42	77.811	500	37	71.270	433	5	6.541	67
Italia	255	366.670	2.498	40	75.103	460	215	291.567	2.038

Tavola 2.A.4 - Stime dei tassi di copertura e coefficienti di variazione per ripartizione

RIPARTIZIONE	N° di individui della popolazione	Stima del tasso di copertura	CV
Nord-Ovest	16.120.067	0,985	0,005
Nord-Est	11.643.194	0,988	0,009
Centro	11.950.322	0,977	0,006
Sud	14.186.373	0,987	0,010
Isole	6.726.486	0,986	0,004
Italia	60.626.442	0,985	0,003

Tavola 2.A.5 - Stime dei tassi di copertura e coefficienti di variazione per regione

REGIONE	N° di individui della popolazione	Stima del tasso di copertura	CV
Piemonte	4.457.335	0,992	0,011
Valle D'Aosta	128.230	0,994	0,025
Lombardia	9.917.714	0,982	0,001
Bolzano	507.657	0,991	0,054
Trento	529.457	0,989	0,049
Veneto	4.937.854	0,990	0,004
Friuli V. G.	1.235.808	0,987	0,018
Liguria	1.616.788	0,985	0,042
Emilia Romagna	4.432.418	0,987	0,020
Toscana	3.749.813	0,985	0,009
Umbria	906.486	0,989	0,022
Marche	1.565.335	0,988	0,039
Lazio	5.728.688	0,967	0,003
Abruzzo	1.342.366	0,989	0,013
Molise	319.780	0,993	0,047
Campania	5.834.056	0,985	0,020
Puglia	4.091.259	0,987	0,017
Basilicata	587.517	0,991	0,078
Calabria	2.011.395	0,987	0,005
Sicilia	5.051.075	0,986	0,005
Sardegna	1.675.411	0,988	0,004
Italia	60.626.442	0,985	0,003

Tavola 2.A.6 - Stime dei tassi di copertura e coefficienti di variazione per classi di dimensione demografica dei comuni

CLASSI DI DIMENSIONE DEMOGRAFICA	N° di individui della popolazione	Stima del tasso di copertura	CV
1	10.358.869	0,995	0,001
2	8.458.578	0,991	0,002
3	9.671.759	0,991	0,002
4	17.904.697	0,987	0,009
5	14.232.539	0,966	0,010
Italia	60.626.442	0,985	0,003