



**File ad uso pubblico
micro.STAT**

Indagine sull'integrazione delle seconde generazioni

Periodo di riferimento: anno 2015

Aspetti metodologici dell'indagine

INDICE

1. Introduzione.....	3
2. La popolazione di riferimento	4
3. Il disegno campionario	6
4. La rilevazione e il trattamento dei dati.....	8
5. La metodologia di calcolo dei pesi campionari.....	11
6. Valutazione del livello di precisione delle stime.....	14
7. La diffusione dei risultati dell'indagine.....	18
8. Riferimenti bibliografici	18
9. Contatti.....	19

1. Introduzione

L'Istat da anni considera, ormai, bambini e ragazzi, non solo come una classe di età, ma come un vero e proprio soggetto sociale, portatore di specifiche esigenze da studiare e alle quali prestare attenzione nelle indagini per consentire la programmazione di opportune politiche di intervento. L'attenzione per l'infanzia e l'adolescenza sono cresciuti nel tempo facendo diventare i ragazzi protagonisti di specifiche rilevazioni.

Anche per quanto riguarda la scuola le rilevazioni e le analisi dei dati dell'Istat e del MIUR offrono una crescente quantità di informazioni., sia attraverso la realizzazione di indagini campionarie, sia attraverso la valorizzazione dei dati amministrativi.

I ragazzi rappresentano però un soggetto sociale allo stesso tempo fondamentale (per la costruzione del futuro) e vulnerabile. Per questo si impone la necessità di arricchire, in un'epoca di crisi e cambiamento, le informazioni statistiche raccolte e analizzate. È evidente, infatti, che il superamento della crisi che l'Italia sta vivendo deve avvenire anche attraverso la realizzazione di politiche che consentano di valorizzare il capitale umano dei giovanissimi. È per questo che statistiche sulla scuola e sui ragazzi appaiono oggi più utili che mai anche in un'ottica di miglioramento generale delle condizioni di vita.

I ragazzi con background migratorio rappresentano poi l'anello più vulnerabile, all'interno di questa categoria sociale. La storia di paesi (come la Francia e la Gran Bretagna) che prima di noi hanno gestito i fenomeni migratori ci ha insegnato che le dinamiche di integrazione delle seconde generazioni e dei ragazzi immigrati sono molto diverse rispetto a quelle vissute dai migranti adulti e che possono portare, oltre che a un disagio generalizzato, a gravi episodi di tensione. Questi paesi, tra l'altro, si sono trovati a dover gestire l'integrazione dei ragazzi stranieri in momenti storici diversi da quello attuale, caratterizzati da crescita economica. Nel caso dell'Italia la situazione è particolarmente delicata perché ci troviamo a dover gestire l'ingresso nel mercato del lavoro di tanti giovani con background migratorio in un momento di crisi.

Per questo è importante che la statistica si adegui rapidamente e porti sul tavolo dei policy makers informazioni in grado di mettere in evidenza le criticità dei processi di integrazione di questi ragazzi, ma anche le esperienze positive delle quali fare tesoro per una scuola che includa davvero e che possa essere risorsa preziosa per accompagnare un più ampio inserimento positivo nella società di accoglienza.

Nell'ambito della Convenzione stipulata fra l'Istat e il Ministero dell'Interno, che si inquadra nel contesto dei Progetti finanziati dal Fondo Europeo per l'Integrazione (Fondo FEI), l'Istat ha svolto tra marzo e giugno 2015 l'indagine sull'Integrazione delle seconde

generazioni che ha previsto la collaborazione del Ministero dell'Istruzione, dell'Università e della Ricerca (MIUR). Le seconde generazioni e in generale i ragazzi figli di stranieri, rappresentano sia un punto di vista avanzato sui processi di integrazione, sia una vera e propria sfida per l'attività di progettazione degli interventi politici per favorire la buona riuscita dell'inserimento sociale. La scuola è apparsa l'ambiente migliore per effettuare la rilevazione in quanto è un luogo fondamentale di socializzazione oltre che di apprendimento.

2. La popolazione di riferimento

L'indagine è basata su un campione di scuole secondarie (primo e secondo grado) statali con almeno 5 alunni stranieri. Le scuole sono state estratte sulla base dei dati forniti dall'anagrafe degli studenti del MIUR in base ai dati provvisori riferiti a metà dicembre 2014 (per Aosta e Bolzano il dato di riferimento è stato quello dell'a.s. 2012/2013). L'indagine è stata rivolta allo studio delle seconde generazioni in senso lato e quindi, più propriamente, ai ragazzi con un background migratorio. Sono stati intervistati i ragazzi che hanno una cittadinanza diversa da quella italiana (compresi gli apolidi), seguendo il criterio della cittadinanza e non quello del Paese di nascita. Si sottolinea che, in accordo con quanto previsto dalla normativa italiana, sono stati considerati stranieri anche i ragazzi nati in Italia da genitori stranieri. Il questionario ha consentito di identificare le differenti tipologie e le diverse generazioni (generazione 2, generazione 1,5, etc.). I ragazzi nati all'estero che hanno acquisito la cittadinanza italiana sono stati considerati italiani. Nel caso di doppia cittadinanza di cui almeno una sia italiana il ragazzo è stato considerato italiano. Nella fase iniziale di campionamento la cittadinanza attribuita agli allievi è stata quella registrata dal MIUR nell'anagrafe degli studenti. A seguito di alcune segnalazioni di irregolarità ricevute dai comuni rispetto ad errata classificazione degli allievi in base alla cittadinanza si è proceduto a dei controlli. Le verifiche incrociate effettuate con le liste anagrafiche comunali hanno consentito di individuare imprecisioni e mancati aggiornamenti delle situazioni scolastiche, con conseguente revisione del campione teorico. Essendo l'obiettivo principale lo studio dell'integrazione delle persone con background migratorio è stato necessario, per una complessiva comprensione del fenomeno, coinvolgere anche alunni italiani (in qualità di gruppo di controllo ne sono stati intervistati un numero pari a quello degli stranieri presenti in ogni classe).

L'indagine ha avuto come ulteriori target di riferimento:

- i dirigenti scolastici delle scuole campione;
- gli Insegnanti di italiano e matematica delle scuole campione che insegnano in classi con stranieri.

I principali contenuti informativi dell'indagine sono per quanto riguarda gli allievi (sia stranieri che italiani) :

- la storia migratoria (se migranti)
- la conoscenza e l'uso della lingua italiana (per gli stranieri)
- la scuola, gli insegnanti e i compagni
- il tempo libero e gli amici
- la famiglia e l'abitazione

Per quanto riguarda i dirigenti scolastici e gli insegnanti i temi principalmente approfonditi sono:

- l'accoglienza dei ragazzi stranieri a scuola
- le attività a favore dell'integrazione
- la formazione rispetto al tema della presenza straniera .

Ai dirigenti scolastici è stata proposta un'intervista strutturata da un rilevatore. Gli alunni e i docenti hanno compilato un questionario online. L'accesso al questionario era possibile attraverso utenze e password riservate. Gli alunni hanno provveduto alla auto-compilazione del modello nelle aule scolastiche con l'eventuale supporto di un rilevatore comunale. Gli insegnanti potevano auto-compilare il questionario in qualsiasi momento durante il periodo di indagine.

I comuni interessati dalla rilevazione sono stati 821. I comuni sono stati coinvolti con i loro uffici di statistica come riferimento per gli uffici Regionali dell'Istat, che hanno avuto la fondamentale funzione di seguire sul territorio la rilevazione.

Hanno risposto all'indagine oltre 1400 scuole su tutto il territorio nazionale.

3. Il disegno campionario

La popolazione di interesse dell'indagine, ossia l'insieme delle unità statistiche intorno alle quali si intende investigare, è costituita dagli alunni stranieri che frequentano le scuole medie, i licei e gli istituti tecnico/professionali italiani (Anagrafe degli studenti del MIUR) in cui si registra un numero minimo di presenza di alunni stranieri pari a 5. L'introduzione di una soglia circa la presenza di alunni stranieri nelle scuole è stata determinata dal fatto che numerose scuole, soprattutto al Sud, presentavano un numero esiguo di alunni stranieri.

L'archivio delle scuole così delimitato costituisce l'universo di riferimento per la definizione del disegno di campionamento. L'archivio è composto complessivamente da 9.386 scuole e contiene informazioni relative alle scuole stesse (tipologia, localizzazione) e informazioni sugli alunni stranieri che le frequentano (numerosità, sesso, cittadinanza, ecc.). Nell'archivio è disponibile anche l'informazione sul numero di alunni italiani nelle classi in cui sono presenti alunni stranieri; tale popolazione è altresì d'interesse per l'indagine almeno per la stima di alcuni parametri di popolazione.

Poiché i principali parametri oggetto di stima dell'indagine, totali e frequenze relative, sono riferiti alla popolazione degli alunni stranieri, il disegno di campionamento è stato progettato con riferimento a tale popolazione.

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

- Regione (21 modalità, con Trento e Bolzano distinte);
- Regione (19 modalità, le regioni di piccole dimensioni del Sud – Molise, Basilicata e Sardegna – sono state aggregate in un'unica regione fittizia) e tipologia di comune (2 modalità: comuni del Centro-Nord con più di 250.000 abitanti e il comune di Napoli del Sud Italia; altri comuni);
- Regione (19 modalità) e tipologia di scuola (2 modalità: medie e superiori);
- Ripartizione geografica (4 modalità: Nord-Ovest, Nord-Est, Centro e Sud-Isole) e tipologia di scuola (3 modalità: medie, licei, istituti tecnico/professionali);

Per l'indagine in oggetto è stato adottato un disegno di campionamento stratificato a grappoli in cui le unità primarie (i grappoli) sono le scuole e le unità finali sono gli alunni stranieri. L'universo di scuole è stato stratificato tenendo conto delle modalità delle seguenti variabili:

- Regione (21 modalità);
- Tipologia di comune (2 modalità: comuni del Centro-Nord con più di 250.000 abitanti e il comune di Napoli del Sud Italia; altri comuni);
- Tipologia di scuola (3 modalità: medie, licei, istituti tecnico/professionali);
- Incidenza di alunni stranieri (3 classi definite sulla base dei terzili delle distribuzioni di alunni stranieri presenti nelle scuole delle regioni del Centro-Nord e nelle grandi regioni del Sud; 2 classi definite sulla base della mediana delle distribuzioni di alunni stranieri presenti nelle scuole delle regioni Molise, Basilicata e Sardegna).

La numerosità del campione di scuole è stata determinata in modo indiretto: è stata fissata dapprima la numerosità complessiva del campione di alunni stranieri e determinata l'allocazione di tale numerosità negli strati ottenuti dall'incrocio delle variabili di stratificazione definite sull'universo delle scuole; successivamente è stato calcolato il numero di scuole campione da selezionare da ciascuno strato sulla base del numero medio di alunni stranieri per scuola.

L'allocazione di alunni stranieri negli strati definiti sull'universo delle scuole è stata ottenuta utilizzando una metodologia di allocazione ottima multivariata e multi-dominio (Bethel, 1989 ; Falorsi et al., 1998) che consente di ottenere la minima numerosità del campione che rispetta dei vincoli di errore campionario prefissati a livello dei domini di stima definiti. A tal fine è stato utilizzato il software MAUSS disponibile nell'Istituto.

Poiché non erano disponibili informazioni sulla variabilità dei fenomeni di interesse dell'indagine, si è proceduto utilizzando come stima di interesse una prevalenza generica pari al 10%; i vincoli sugli errori campionari sono stati fissati in modo differenziato per i domini di stima sopra definiti.

Il campione di scuole è risultato composto da 1.419 unità. La selezione delle scuole campione è stata effettuata in modo casuale con probabilità uguali all'interno di ciascuno strato. Con riferimento ad una generica scuola campione i appartenente allo strato h ($h=1, \dots, H$) la probabilità di inclusione è espressa da:

$$P_{hi} = \frac{n_h}{N_h}$$

dove n_h e N_h rappresentano rispettivamente il numero di scuole campione e il numero di scuole universo nello strato h .

Tutti gli alunni stranieri appartenenti alle scuole campione sono stati intervistati. Inoltre in ogni classe in cui erano presenti alunni stranieri è stato selezionato casualmente un campione di alunni italiani di numerosità pari a quella degli alunni stranieri della classe.

La dimensione complessiva del campione di alunni stranieri appartenenti al campione selezionato di scuole è di 42.239 unità.

4. La rilevazione e il trattamento dei dati

La tecnica Cawi è una metodologia d'indagine orientata ai rispondenti (friendly) che si basa sulla possibilità di compilare e inviare, da parte dei rispondenti, un questionario online cui possono accedere per mezzo di un browser, attraverso l'indirizzo del server fornito dall'Istituto. Nel caso dell'indagine sull'integrazione delle seconde generazioni, questa tecnica è sembrata essere particolarmente utile in quanto si trattava di raggiungere rispondenti giovani, abituati all'utilizzo delle nuove tecnologie e si poteva contare sulla disponibilità, nella quasi totalità delle scuole, di dotazioni informatiche adeguate per lo svolgimento della rilevazione. L'intento, fin dalla fase di progettazione, è stato di pervenire alla definizione di uno strumento di rilevazione delle informazioni in cui i problemi di compilazione fossero ridotti al minimo per evitare un effetto questionario come determinante del successo o meno dell'indagine.

La rilevazione è avvenuta, durante l'orario scolastico, nella aule/laboratori scolastici alla presenza di un rilevatore comunale che aveva la funzione di introdurre l'indagine e risolvere eventuali problematiche, ma non somministrava l'intervista.

Le modalità di conduzione e la tempistica prevista per la rilevazione hanno reso necessario e possibile un monitoraggio giornaliero dell'andamento delle operazioni; non solo è stato possibile seguire centralmente, momento per momento l'andamento della rilevazione, sapendo in tempo reale quante persone stavano compilando il questionario e in quali scuole; la tecnica di rilevazione ha inoltre consentito che, mentre gli studenti si collegavano al sito accedendo con la propria utenza e iniziando a rispondere ai quesiti, i dati venissero immediatamente salvati su un server dell'Istat ed era quindi possibile effettuare il download dei questionari che man mano venivano compilati. Quotidianamente, attraverso programmi SAS opportunamente predisposti, i questionari scaricati venivano sottoposti a controllo al fine di verificare la partecipazione all'indagine e la funzionalità della procedura di acquisizione e, al contempo, monitorare situazioni potenzialmente critiche. Grazie alle procedure di monitoraggio, ad esempio, è stato possibile segnalare e

risolvere tempestivamente un malfunzionamento temporaneo del processo di salvataggio dei questionari sul server predisposto. La procedura di monitoraggio, inoltre, ha consentito di evidenziare in tempo utile eventuali strati sofferenti e ha permesso il sollecito mirato di quelle scuole che non avevano ancora provveduto ad organizzare la rilevazione dei propri studenti. La risposta degli istituti scolastici in generale è stata molto buona, grazie anche all'attivo coinvolgimento del Ministero dell'Istruzione a sostegno della rilevazione e all'attività dei rilevatori comunali: poco meno del 98 per cento delle scuole ha partecipato all'indagine. Di queste ultime, la quasi totalità ha utilizzato la modalità di compilazione via web e solo il 2,5 per cento - a causa di difficoltà tecniche o carenza di strumentazione informatica (computer e/o connessione internet) – ha fatto compilare i questionari cartacei ai propri alunni.

La rilevazione è iniziata il 9 marzo e si è chiusa il 9 giugno 2015. Per misurare il tempo di risposta medio delle scuole del campione selezionato è possibile osservare la distribuzione del giorno di inizio di compilazione, considerando la distanza dal primo giorno dell'indagine a quello del primo questionario compilato da uno qualsiasi dei suoi studenti.

Il questionario elettronico consente l'acquisizione controllata dei dati segnalando al rispondente già in fase di compilazione eventuali incongruenze tra variabili e valori fuori range per i quesiti chiusi. Inoltre è sempre garantito il rispetto dei salti di alcuni quesiti del questionario in funzione della risposta ad altri quesiti. Tuttavia non è sempre agevole riuscire a considerare tutte le possibili incompatibilità tra le variabili, né sarebbe opportuno vincolare la prosecuzione della compilazione del questionario a seguito di ciascuna potenziale incongruenza tra i quesiti. Nella predisposizione del piano di incompatibilità del questionario, infatti, si utilizzano sia regole hard, applicate solo ad alcune variabili la cui correttezza è ritenuta fondamentale per la prosecuzione delle risposte e nel caso di incompatibilità impediscono la prosecuzione fino alla risoluzione della stessa, sia regole soft, che avvisano il rispondente delle incongruenze tra i quesiti lasciando però all'intervistato la possibilità di confermare le sue risposte e proseguire nella compilazione. Vi sono poi quesiti aperti dove non sempre è possibile compilare regole di controllo per le risposte fornite. Quindi, anche in un'indagine che raccoglie le informazioni mediante questionario elettronico vi sono sempre un certo numero di errori da individuare e quindi correggere.

Il piano di controllo e correzione predisposto per l'indagine sull'integrazione delle seconde generazioni ha previsto due fasi: quella dell'individuazione dell'errore è stata seguita dalla fase di correzione vera e propria. Per individuare gli errori è stato predisposto un algoritmo compilato sul software statistico SAS in grado di evidenziare dapprima eventuali filtri non

funzionanti fra i quesiti che generano il mancato rispetto dei salti all'interno del questionario e quindi modalità di risposta non compatibili con altre selezionate in altri quesiti.

Anche la correzione degli errori ha seguito due passaggi sequenziali: ad una prima correzione di tipo deterministico ha fatto seguito l'approccio probabilistico per il quale è stato utilizzato un software generalizzato per l'individuazione e la correzione degli errori che si basa su tecniche di imputazione probabilistica e si applica alle variabili qualitative (SCIA¹).

Nel complesso, la procedura di correzione ha avuto un impatto limitato sui dati dell'indagine. Limitando l'analisi ai soli questionari ottenuti con tecnica CAWI (che rappresentano il 96,3 per cento del totale dei questionari), immaginando l'insieme dei dati come una matrice del tipo record (righe) per variabili (colonne) e computando le sole celle potenzialmente valorizzabili in funzione dei filtri ai quesiti, si ottiene una matrice di oltre 11 milioni di celle; di queste solamente lo 0,5 per cento ha subito una variazione nel confronto fra la matrice dei dati iniziali e quella finale di output del processo di correzione.

Analizzando l'impatto delle correzioni sui record, emerge che il 67,4 per cento dei questionari non ha necessitato di alcun tipo di intervento correttivo (Prospetto 1). La distribuzione dei record su cui è stata apportata almeno una correzione per numero di errori mostra che quasi il 45 per cento dei record ha necessitato di un solo intervento correttivo e il 22,5 per cento presentava due errori; meno del dieci per cento dei record, invece, ha richiesto correzioni su più di cinque variabili.

Prospetto 1 – Numero di errori per record

Numero di errori per record	Per 100 record in totale	Per 100 record corretti
0 errori	67,4	
1 errore	14,6	44,7
2 errori	7,3	22,5
3 errori	4,7	14,4
4-5 errori	2,9	8,8
6 errori o più	3,2	9,7
TOTALE	100,0	100,0

Fonte: Istat, Indagine sull'integrazione delle seconde generazioni

¹ Per una descrizione più ampia di Scia e della metodologia di Fellegi-Holt confronta: Riccini Margarucci, E. *Concord v. 1.01 Software generalizzato per il controllo e la correzione dei dati rilevati nelle indagini statistiche. Manuale utente e aspetti metodologici*. Documento ad uso interno: 2002. Fellegi, I. P. e D. Holt. "A Systematic Approach to Automatic Edit and Imputation". In *Journal of the American Statistical Association*, Vol. 71: 17-35. 1976.

Il tasso di risposta della rilevazione CAWI registrato per gli alunni stranieri² è stato nel complesso pari all'82,1 per cento, con valori più elevati per gli studenti delle scuole secondarie di secondo grado (83,2 per cento contro l'80,8 delle scuole secondarie di primo grado) (Prospetto 2).

Le mancate risposte al questionario online sono da attribuirsi principalmente alla irreperibilità o assenza temporanea degli alunni stranieri delle scuole selezionate nel campione (8,6 per cento in totale), in particolare nelle scuole secondarie di secondo grado (10,9 per cento). I rifiuti espliciti a partecipare alla rilevazione, invece, sono stati complessivamente piuttosto contenuti (3,1 per cento), con valori leggermente più elevati nelle scuole secondarie di primo grado (3,9 contro 2,4 per cento).

Prospetto 2 - Esiti delle interviste CAWI agli alunni stranieri e motivi di caduta (composizioni percentuali)

Esito dell'intervista	Per 100 alunni in totale		
	Scuole secondarie di I grado	Scuole secondarie di II grado	Totale
Intervista completa	80,8	83,2	82,1
Impossibilità a rispondere per gravi motivi	0,5	0,3	0,4
Intervista interrotta	0,4	0,6	0,5
Soggetto irreperibile	5,0	9,4	7,4
Straniero temporaneamente assente	0,9	1,5	1,2
Rifiuto	3,9	2,4	3,1
Esito non attribuito	8,4	2,5	5,2

Fonte: Istat, Indagine sull'integrazione delle seconde generazioni

5. La metodologia di calcolo dei pesi campionari

I pesi campionari degli alunni stranieri e degli alunni italiani in classi con presenza di stranieri sono stati definiti in modo differente. Poiché, nelle scuole campione, gli alunni stranieri sono stati censiti ad essi è stato assegnato il peso campionario della scuola di appartenenza.

Il peso campionario associato alla i -ma scuola campione nello strato h è determinato come inverso della probabilità di inclusione, p_{hi} , ovvero

² Il tasso di risposta non è calcolabile per gli alunni italiani in quanto questi ultimi sono stati intervistati come campione di controllo e in numero pari agli stranieri coinvolti nella rilevazione.

$$w_{hi} = \frac{N_h}{n_h}.$$

Per gli alunni italiani, per i quali è stata definita una numerosità campionaria per classe (classe con presenza di alunni stranieri nelle scuole campione) pari al numero degli alunni stranieri censiti nella stessa, il peso campionario è stato calcolato considerando anche un secondo stadio di selezione. Con riferimento alla j -ma classe della i -ma scuola campione dello strato h , la probabilità di inclusione della generica unità campionaria k è definita come:

$$P_{ijk} = \frac{m_{ij}}{M_{ij}},$$

dove m_{ij} e M_{ij} rappresentano rispettivamente il numero di alunni italiani campione e il numero di alunni italiani universo nella j -ma classe della i -ma scuola campione.

Il peso diretto associato al k -mo alunno italiano campione è espresso come inverso del prodotto delle probabilità di inclusione sopra definite:

$$w_{hijk} = \frac{N_h}{n_h} \frac{M_{ij}}{m_{ij}}.$$

A causa della caduta di alcune scuole campione il peso diretto ad esse associato è stato corretto, in ogni strato, con un fattore ottenuto come inverso del tasso di risposta delle scuole nello strato. Tale correzione si riflette anche sui pesi diretti associati agli alunni stranieri e agli alunni italiani.

Per il calcolo dei coefficienti di riporto all'universo è stata utilizzata una procedura generalizzata di stima, basata sull'uso di stimatori calibrati, o di ponderazione vincolata. La metodologia alla base di tali stimatori consente la determinazione di un unico coefficiente di riporto all'universo in grado di produrre stime coerenti a totali noti desunti da un archivio di riferimento.

La metodologia di calibrazione si basa sull'utilizzo di opportune informazioni ausiliarie, correlate con le principali variabili oggetto di indagine, che hanno la funzione di aumentare l'accuratezza delle stime. I vincoli sono proprio le condizioni di uguaglianza delle stime campionarie di alcune variabili ausiliarie con i rispettivi totali noti.

Nella fase di calibrazione dei pesi diretti degli alunni stranieri, poiché nell'Anagrafe degli studenti del MIUR per 13 scuole non erano presenti tutte le informazioni ausiliarie di cui si voleva tener conto, è stata adottata una procedura che prevede due passi di calibrazione.

Questa scelta è stata determinata dal fatto che in una situazione in cui sussiste una incompletezza di informazioni per alcune unità dell'archivio di riferimento, i vincoli di calibrazione non possono essere soddisfatti contemporaneamente perché riportano ad un ammontare di popolazione diverso.

Per la definizione dei due passi di calibrazione sono stati individuati e distinti i totali che portano allo stesso ammontare di popolazione.

Nel primo passo si è tenuto conto di tre vincoli, ovvero dei totali noti per:

- Regione (21 modalità, con Trento e Bolzano distinte nella regione Trentino Alto Adige) e Sesso (2 modalità);
- Regione e 4 cittadinanze (le prime tre cittadinanze più frequenti a livello regionale e una che ingloba tutte le altre cittadinanze);
- 11 Cittadinanze (le prime 10 più frequenti e una che ingloba tutte le altre cittadinanze a livello nazionale).

Nel secondo passo di calibrazione si è tenuto conto di un solo vincolo, ovvero di un totale noto dato dall'incrocio delle modalità delle seguenti variabili ausiliarie: Regione (21 modalità, con Trento e Bolzano distinte nella regione Trentino Alto Adige); Tipologia di comune (2 modalità: comuni del Centro-Nord con più di 250.000 abitanti e il comune di Napoli del Sud Italia; altri comuni); Tipologia di scuola (3 modalità: medie, licei, istituti tecnico/professionali).

Il primo passo della procedura ha condotto alla determinazione di un peso corretto su totali di alunni stranieri per i quali erano note le informazioni sul sesso e la cittadinanza. I pesi finali ottenuti con il secondo passo di calibrazione riportano al numero totale di alunni stranieri presenti nell'universo delle scuole che comprende anche le 13 scuole con informazioni mancanti sul sesso e la cittadinanza degli alunni stranieri.

Il peso finale \tilde{w}_{hik} assegnato al k-mo alunno straniero della i-ma scuola campione appartenente allo strato h è espresso da:

$$\tilde{w}_{hik} = w_{hik} f_k' f_k''$$

dove f_k' e f_k'' rappresentano rispettivamente il primo e il secondo fattore correttivo attribuito al k-mo alunno straniero attraverso i due passi della procedura di calibrazione.

Per quanto riguarda gli alunni italiani, il calcolo dei coefficienti di riporto all'universo è stato effettuato, a partire dal peso diretto ad essi associato, tenendo conto di un solo vincolo di

calibrazione, ovvero un totale noto dato dall'incrocio delle modalità delle variabili ausiliarie: Regione (21 modalità, con Trento e Bolzano distinte nella regione Trentino Alto Adige); Tipologia di Comune (2 modalità: comuni del Centro-Nord con più di 250.000 abitanti e il comune di Napoli del Sud Italia; altri comuni); Tipologia di scuola (2 modalità: Medie, Superiori).

Il peso finale \tilde{w}_{hijk^*} assegnato al k^* -mo alunno italiano della j -ma classe della i -ma scuola campione appartenente allo strato h è espresso da:

$$\tilde{w}_{hijk^*} = w_{hijk^*} f_{k^*}.$$

La procedura di calibrazione è stata realizzata utilizzando il software generalizzato Regenesees messo a punto dall'Istat (Zardetto, 2015³).

6. Valutazione del livello di precisione delle stime

Le principali statistiche di interesse per valutare la variabilità campionaria delle stime prodotte da un'indagine sono l'errore di campionamento assoluto e l'errore di campionamento relativo, definite dalle seguenti espressioni:

$$\begin{aligned}\hat{\sigma}(\hat{Y}) &= \sqrt{\hat{V}(\hat{Y})}, \\ \hat{\varepsilon}(\hat{Y}) &= \frac{\hat{\sigma}(\hat{Y})}{\hat{Y}}.\end{aligned}$$

Come è stato descritto nel paragrafo precedente, le stime prodotte dall'indagine sono state ottenute mediante uno stimatore di calibrazione in due passi sulla base di una funzione di distanza di tipo lineare. Poiché lo stimatore adottato non è funzione lineare dei dati campionari non è possibile ottenere una espressione analitica per la stima della varianza. Pertanto si è utilizzato il metodo proposto da Woodruff che, ricorrendo all'espressione linearizzata in serie di Taylor, consente di ottenere la varianza di ogni stimatore non lineare calcolando la varianza dell'espressione linearizzata ottenuta. Tale metodologia di stima della varianza è implementata nel software generalizzato ReGenesees, che è stato utilizzato per la stima della varianza delle stime.

Poiché le stime prodotte dall'indagine in oggetto in numero molto elevato, si è fatto ricorso ad una procedura in cui gli errori campionari delle stime sono presentati in forma sintetica. Tale procedura utilizza il metodo dei modelli regressivi che si basa sulla determinazione di

³ Zardetto Diego, 2015. ReGenesees: an Advanced R System for Calibration, Estimation and Sampling Error Assessment in Complex Sample Surveys. Journal of Official Statistics. Volume 31, Issue 2, Pages 177–203, ISSN (Online) 2001-7367, June 2015

una funzione matematica che mette in relazione ciascuna stima con il proprio errore campionario relativo stimato.

Il modello utilizzato per le stime di frequenze assolute e relative è:

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

dove i parametri a e b sono stimati con il metodo dei minimi quadrati. I modelli regressivi del tipo descritto, che permettono la presentazione sintetica degli errori di campionamento, sono stati ottenuti sempre tramite il software Regenesees.

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

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

Relativamente alle stime riferite agli alunni stranieri sono stati utilizzati diverse tipologie di modelli: un modello per tipologia di scuola (3 modalità) definito a livello nazionale; un modello per tipologia di scuola (3 modalità) definito a livello di ripartizione geografica; un modello per tipologia di scuola (2 modalità) definito a livello di regione geografica; un modello per i grandi comuni.

Di seguito sono riportati, per i domini di stima definiti, i prospetti con i valori interpolati degli errori campionari delle stime riferite al totale degli alunni stranieri, ottenuti utilizzando i corrispondenti modelli. Ad esempio nel caso di celle con 250 alunni stranieri nei Licei si ha un errore di poco inferiore al 20%.

Prospetto 3 - Valori interpolati degli errori campionari delle stime riferite al totale degli alunni stranieri a livello Italia e tre tipologie di scuole - Anno 2015

Dominio di stima Nazionale e tipologia di scuola		250	500	1.000	2.500	5.000	10.000	25.000	50.000
Italia		36,38	24,27	16,19	9,48	6,32	4,22	2,47	1,65
	Istituti ^a	34,47	23,02	15,37	9,01	6,02	4,02	2,36	1,57
	Licei	19,97	13,36	8,94	5,25	3,51	2,35	1,38	0,92
	Medie ^b	29,28	20,16	13,88	8,48	5,84	4,02	2,45	1,69

Note: (a) – Si tratta delle scuole secondarie di secondo grado diverse dai licei
(b) – Scuole secondarie di primo grado

Prospetto 4 - Valori interpolati degli errori campionari delle stime riferite al totale degli alunni stranieri per ripartizione geografica e tre tipologie di scuole - Anno 2015

Dominio di stima Nazionale e tipologia di scuola		100	250	500	1.000	2.500	5.000	10.000	25.000
Nord-Ovest		68,27	40,05	26,75	17,87	10,48	7,00	4,68	2,74
	Istituti ^a	60,17	35,38	23,67	15,84	9,31	6,23	4,17	2,45
	Licei	30,30	17,97	12,10	8,15	4,83	3,26	2,19	1,30
	Medie ^b	55,49	34,32	23,87	16,59	10,26	7,14	4,96	3,07
Nord-Est		47,60	28,24	19,03	12,82	7,60	5,12	3,45	2,05
	Istituti ^a	48,76	28,98	19,55	13,19	7,84	5,29	3,57	2,12
	Licei	28,52	16,72	11,16	7,45	4,37	2,92	1,95	1,14
	Medie ^b	32,40	20,66	14,70	10,46	6,67	4,75	3,38	2,16
Centro		54,21	31,43	20,81	13,78	7,99	5,29	3,50	2,03
	Istituti ^a	55,07	31,94	21,15	14,01	8,13	5,38	3,56	2,07
	Licei	34,15	20,19	13,56	9,11	5,39	3,62	2,43	1,44
	Medie ^b	41,03	24,73	16,86	11,49	6,93	4,72	3,22	1,94
Sud e isole		28,63	16,60	10,99	7,28	4,22	2,80	1,85	1,07
	Istituti ^a	27,78	16,04	10,59	6,99	4,03	2,66	1,76	1,01
	Licei	23,86	13,93	9,27	6,17	3,60	2,40	1,59	0,93
	Medie ^b	20,27	12,57	8,76	6,10	3,79	2,64	1,84	1,14

Note: (a) – Si tratta delle scuole secondarie di secondo grado diverse dai licei
(b) – Scuole secondarie di primo grado

Prospetto 5 - Valori interpolati degli errori campionari delle stime riferite al totale degli alunni stranieri per regione geografica di residenza e due tipologie di scuole - Anno 2015

Dominio di stima Regione e tipologia di scuola		100	250	500	1.000	2.500	5.000	10.000	25.000
Piemonte		42,39	24,94	16,69	11,17	6,57	4,40	2,95	1,73
	Medie	38,97	23,03	15,47	10,39	6,14	4,13	2,77	1,64
	Superiori	39,50	23,53	15,90	10,74	6,40	4,32	2,92	1,74
Valle d'Aosta		11,60	6,64	4,36	2,86	1,64	1,07	0,70	0,40
	Medie	9,93	5,62	3,66	2,38	1,35	0,88	0,57	0,32
	Superiori	11,82	6,85	4,53	2,99	1,73	1,15	0,76	0,44
Lombardia		71,73	42,43	28,52	19,17	11,34	7,62	5,12	3,03
	Medie	62,32	38,62	26,89	18,73	11,61	8,08	5,63	3,49
	Superiori	60,00	35,44	23,80	15,98	9,44	6,34	4,26	2,51
Veneto		47,06	27,54	18,36	12,25	7,17	4,78	3,19	1,86
	Medie	37,21	22,84	15,78	10,91	6,69	4,63	3,20	1,96
	Superiori	44,60	26,19	17,51	11,70	6,87	4,59	3,07	1,80
Friuli-Venezia Giulia		20,46	12,22	8,28	5,61	3,35	2,27	1,54	0,92
	Medie	18,89	11,47	7,87	5,40	3,28	2,25	1,54	0,94
	Superiori	20,15	11,96	8,07	5,44	3,23	2,18	1,47	0,87
Liguria		23,65	14,09	9,52	6,43	3,83	2,59	1,75	1,04
	Medie	22,33	13,33	9,03	6,11	3,65	2,47	1,67	1,00
	Superiori	21,90	13,31	9,13	6,26	3,81	2,61	1,79	1,09
Emilia-Romagna		47,69	28,66	19,50	13,27	7,97	5,43	3,69	2,22
	Medie	37,73	23,83	16,83	11,89	7,51	5,30	3,75	2,37
	Superiori	47,57	28,54	19,39	13,17	7,90	5,37	3,65	2,19
Toscana		53,06	30,63	20,22	13,34	7,70	5,08	3,35	1,94
	Medie	34,78	21,34	14,74	10,19	6,25	4,32	2,99	1,83
	Superiori	50,95	30,21	20,34	13,70	8,12	5,47	3,68	2,18
Umbria		24,25	14,38	9,68	6,52	3,87	2,60	1,75	1,04
	Medie	21,18	12,69	8,62	5,85	3,51	2,38	1,62	0,97
	Superiori	24,34	14,49	9,79	6,61	3,94	2,66	1,80	1,07
Marche		31,29	18,09	11,96	7,90	4,57	3,02	2,00	1,15
	Medie	24,22	14,53	9,88	6,71	4,03	2,74	1,86	1,12
	Superiori	32,12	18,52	12,22	8,06	4,65	3,06	2,02	1,17

Prospetto 5 – segue - Valori interpolati degli errori campionari delle stime riferite al totale degli alunni stranieri per regione geografica di residenza e due tipologie di scuole - Anno 2015

Dominio di stima Regione e tipologia di scuola	100	250	500	1.000	2.500	5.000	10.000	25.000	
Lazio		52,25	30,65	20,47	13,67	8,02	5,36	3,58	2,10
	Medie	46,61	28,06	19,11	13,02	7,84	5,34	3,64	2,19
	Superiori	51,39	29,64	19,54	12,89	7,43	4,90	3,23	1,86
Abruzzo		20,01	11,78	7,89	5,28	3,11	2,08	1,39	0,82
	Medie	17,29	10,30	6,96	4,70	2,80	1,89	1,28	0,76
	Superiori	20,42	11,96	7,99	5,33	3,12	2,08	1,39	0,82
Molise		12,33	7,36	4,99	3,38	2,02	1,37	0,93	0,55
	Medie	10,52	6,13	4,08	2,71	1,58	1,05	0,70	0,41
	Superiori	12,57	7,41	4,97	3,33	1,97	1,32	0,88	0,52
Campania		25,14	14,73	9,82	6,55	3,84	2,56	1,71	1,00
	Medie	18,04	11,20	7,82	5,45	3,39	2,36	1,65	1,02
	Superiori	27,27	15,76	10,41	6,88	3,98	2,63	1,73	1,00
Puglia		24,02	14,21	9,55	6,42	3,80	2,55	1,71	1,01
	Medie	24,51	14,27	9,48	6,30	3,67	2,43	1,62	0,94
	Superiori	22,22	12,99	8,65	5,76	3,37	2,24	1,50	0,87
Basilicata		14,31	8,61	5,86	3,99	2,40	1,64	1,11	0,67
	Medie	12,54	7,40	4,96	3,33	1,97	1,32	0,89	0,52
	Superiori	14,50	8,29	5,43	3,56	2,03	1,33	0,87	0,50
Calabria		21,41	12,46	8,28	5,49	3,20	2,12	1,41	0,82
	Medie	18,79	10,74	7,04	4,61	2,63	1,73	1,13	0,65
	Superiori	21,27	12,45	8,30	5,54	3,24	2,16	1,44	0,84
Sicilia		26,70	15,76	10,58	7,10	4,19	2,81	1,89	1,11
	Medie	24,50	14,65	9,93	6,73	4,02	2,73	1,85	1,10
	Superiori	25,75	15,17	10,17	6,81	4,01	2,69	1,80	1,06
Sardegna		18,19	10,71	7,18	4,81	2,84	1,90	1,27	0,75
	Medie	15,92	9,46	6,38	4,30	2,55	1,72	1,16	0,69
	Superiori	18,48	10,80	7,19	4,79	2,80	1,87	1,24	0,73
Trentino - Alto Adige									
Bolzano		18,91	11,36	7,72	5,25	3,15	2,14	1,46	0,88
	Medie	14,03	8,25	5,53	3,70	2,18	1,46	0,97	0,57
	Superiori	21,20	12,98	8,95	6,18	3,78	2,61	1,80	1,10
Trento		15,10	8,93	6,00	4,03	2,38	1,60	1,08	0,64
	Medie	14,96	8,89	5,99	4,04	2,40	1,62	1,09	0,65
	Superiori	13,50	7,98	5,36	3,60	2,13	1,43	0,96	0,57

Prospetto 6 - Valori interpolati degli errori campionari delle stime riferite al totale degli alunni stranieri per grandi comuni - Anno 2015

Dominio di stima Grandi comuni	100	250	500	1.000	2.500	5.000	10.000	25.000
Torino	28,02	16,64	11,21	7,56	4,49	3,03	2,04	1,21
Milano	40,90	25,16	17,43	12,07	7,42	5,14	3,56	2,19
Venezia	23,45	13,62	9,03	5,99	3,48	2,31	1,53	0,89
Verona	20,34	11,97	8,02	5,37	3,16	2,12	1,42	0,84
Genova	19,27	11,62	7,92	5,40	3,25	2,22	1,51	0,91
Bologna	16,51	9,86	6,67	4,52	2,70	1,82	1,23	0,74
Firenze	23,05	13,90	9,48	6,47	3,90	2,66	1,81	1,09
Roma	46,35	27,74	18,81	12,76	7,64	5,18	3,51	2,10
Napoli	21,93	13,33	9,14	6,27	3,81	2,62	1,79	1,09

7. La diffusione dei risultati dell'indagine

La prima diffusione dei risultati dell'indagine è avvenuta sul sito dell'Istat attraverso la statistica report (<http://www.istat.it/it/archivio/182866>). In occasione dell'uscita del Report è stato organizzato il convegno "L'integrazione delle seconde generazioni. Stranieri oggi, italiani domani?" (Roma, 21 marzo 2016), nel quale sono stati presentati alcuni dei più importanti risultati dell'indagine (<https://www.istat.it/it/archivio/182490>).

I principali risultati dell'indagine sono inoltre stati diffusi in occasioni di numerosi altri convegni e seminari scientifici.

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

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

Nel febbraio 2018 è stato pubblicato un volume metodologico dal titolo "L'indagine sull'integrazione delle seconde generazioni: obiettivi, metodologia e organizzazione", al quale si rimanda per maggiori approfondimenti (<http://www.istat.it/it/archivio/209438>).

8. Riferimenti bibliografici

ISTAT (2018), *L'indagine sull'integrazione delle seconde generazioni: obiettivi, metodologia e organizzazione*, Letture statistiche – Metodi, Istat, Roma.

ISMU-MIUR (2016), *Alunni con cittadinanza non italiana. La scuola multiculturale nei contesti locali. Rapporto nazionale A.s. 2014/2015*, Milano

DE SANTIS G. , STROZZA S. (2017), *Rapporto sulla popolazione. Le molte facce della presenza straniera in Italia*, Il Mulino, Bologna

UNITED NATIONS ECONOMIC COMMISSION FOR EUROPE, CONFERENCE OF EUROPEAN STATISTICIANS, WORK SESSION ON MIGRATION STATISTICS (2017), *Istat experience in surveys on social integration: the first and second generation of migrants*, Ginevra

9. Contatti

**Servizio Registro della popolazione,
statistiche demografiche e condizioni di vita**

Cinzia Conti

Tel. +39 06 4673.7366

Email: ciconti@istat.it

Roberto Petrillo

Tel. +39 06 4673.7560

Email: petrillo@istat.it

Il documento è stato curato da Cinzia Conti (paragrafi 1 e 2), Claudia de Vitiis, Marco Dionisio Terribili, Francesca Inglese e Alessio Guandalini (paragrafi 3, 5 e 6), Roberto Petrillo e Fabio Massimo Rottino (paragrafo 4), Roberto Petrillo (paragrafo 7).