

File Standard

**L'indagine campionaria sulle nascite
e le madri – Anno 2012**
Indagine retrospettiva

Manuale per l'utente



Istituto Nazionale di Statistica

I file standard vengono rilasciati per finalità di studio e ricerca. Per ottenere tali file è necessario registrarsi al Contact centre. Una volta effettuata la registrazione, la richiesta deve essere formulata selezionando nel Contact centre l'area "Collezioni campionarie di dati elementari (file standard) e compilando un modulo on-line.

Per informazioni sull'indagine rivolgersi a:
Istat - Servizio 'Registro della popolazione, statistiche demografiche e condizioni di vita'
Viale Liegi, 13 – 00198
Roma
tel: 06.4673.7322
fax: 0646737621
e-mail: cicastag@istat.it

Il manuale è curato da Cinzia Castagnaro.

La premessa e il paragrafo 1.1 sono curati da Cinzia Castagnaro e Sabrina Prati.

I paragrafi 1.2 e 1.3 sono stati curati da Claudia Iaccarino.

Il capitolo 2 è stato curato da Claudia De Vitiis, Francesca Inglese, Marco Terribili, Nicoletta Cibella e Adriano Pareto.

I programmi per la correzione dei dati e la creazione del file standard sono stati progettati e realizzati da Claudia Iaccarino.

La rilevazione dell'indagine è stata curata da Cinzia Castagnaro e Sabrina Prati.

Indice

Premessa.....	5
1. L'indagine Campionaria sulle Nascite: caratteristiche e contenuti.....	6
1.1 Le informazioni statistiche sulle nascite: nuove rilevazioni per nuove esigenze informative.....	6
1.2 Informazioni errate o incompatibili.....	8
1.3 Individuazione e correzione degli errori.....	9
2. Strategia di campionamento e livello di precisione delle stime.....	10
2.1 Obiettivi dell'indagine.....	10
2.2 Disegno di campionamento	10
2.3 La procedura di stima e calcolo degli errori campionari	10
2.3.1 Presentazione sintetica degli errori campionari	12

Premessa

Il decreto legislativo n. 322 del 6/9/1989 regola la diffusione delle informazioni statistiche prodotte nell'ambito del Sistema Statistico Nazionale al fine di garantire la riservatezza dei rispondenti. In particolare, per la diffusione di dati elementari, l'articolo 10, comma 2, dispone quanto segue: "Sono distribuite altresì, ove disponibili, su richiesta motivata e previa autorizzazione del Presidente dell'Istat, collezioni campionarie di dati elementari, resi anonimi e privi di ogni riferimento che ne permetta il collegamento con singole persone fisiche e giuridiche".

Nell'osservanza di tale disposizione e del d. lgs del 30/06/2003 n. 196 (Codice in materia di protezione dei dati personali) l'Istat ha adottato misure e tecniche che rendono impossibile, o altamente improbabile, il collegamento dei dati rilasciati con l'unità statistica a cui si riferiscono. Per tale motivo vengono apportate alcune modifiche sui file originali delle indagini, nell'intento di garantire la massima protezione ai dati, contenendo al minimo la perdita di informazioni. Le metodologie applicate si concretizzano nell'accorpamento e/o riclassificazione di modalità di variabili e nella mancata fornitura di variabili che permetterebbero l'identificazione dei rispondenti o di variabili di lavorazione riservate Istat.

Nelle prossime pagine viene riportata una breve descrizione delle caratteristiche dell'Indagine e del disegno di campionamento e la metodologia adottata per la protezione dei dati.

Per permettere l'elaborazione dei dati e l'interpretazione dei risultati, negli allegati a seguire sono riportati il tracciato record, il questionario e le principali classificazioni e definizioni adottate.

1. L'indagine Campionaria sulle Nascite: caratteristiche e contenuti

1.1 *Le informazioni statistiche sulle nascite: nuove rilevazioni per nuove esigenze informative*

Il sistema di raccolta e produzione dei dati statistici sulle nascite è stato, negli ultimi anni, fortemente modificato e rinnovato. Il processo di cambiamento, che si inquadra nella strategia dell'Istituto Nazionale di Statistica di osservare gli eventi e i comportamenti demografici in una prospettiva conoscitiva, è stato indirettamente accelerato dalla necessità di adeguare i flussi informativi alle nuove norme in materia di denuncia di nascita entrate in vigore tra il 1997 e il 1999.

Per oltre 70 anni l'Istat ha diffuso le principali informazioni statistiche sulle nascite e i parti attraverso i dati provenienti dalla rilevazione delle nascite di fonte Stato Civile, con un dettaglio informativo molto ricco ai fini della descrizione dei fenomeni. Sulla base di questa rilevazione, corrente ed esaustiva, è stato possibile fornire al paese con regolarità e accuratezza le informazioni relative alle modificazioni dei comportamenti riproduttivi avvenute nel nostro paese.

La rilevazione delle nascite ha consentito infatti per lungo tempo di monitorare con continuità e precisione la forte riduzione della fecondità, soprattutto per i figli successivi al primo, l'incremento dell'infertilità e il fortissimo innalzarsi dell'età media alla nascita del primogenito, con i conseguenti crescenti rischi non solo di infertilità, ma anche di gravidanze a maggior rischio di complicanze, particolarmente per le primipare. Essa ha inoltre garantito al paese un'informazione strutturale puntuale su alcuni fenomeni di grande rilevanza bio-demografica e socio-sanitaria, quali la natimortalità, i parti plurimi, le caratteristiche del parto rispetto alle principali caratteristiche demografiche dei genitori.

I mutamenti normativi riguardanti la dichiarazione di nascita hanno imposto la soppressione, a partire dal 1° gennaio 1999, della rilevazione individuale delle nascite di fonte Stato Civile. Ne è seguita una vera e propria azione di rigenerazione di tutta la strumentazione logica e metodologica finora utilizzata per la produzione delle statistiche sulle nascite.

Da una rilevazione sulle nascite si è passati ad un sistema di rilevazioni che consente non solo di colmare il debito informativo creatosi, ma anche di ampliare considerevolmente la produzione di informazioni rilevanti per la comprensione dei fenomeni oggetto di osservazione, venendo così incontro alle mutate esigenze della domanda informativa. Si fa sempre più pressante, infatti, l'esigenza di approfondire le determinanti e le dinamiche che influiscono sulle scelte di maternità e di paternità, così come l'esigenza di analizzare i contesti di vita familiari e sociali in cui tali determinanti svolgono la loro azione.

Il compito di soddisfare le esigenze informative che si sono venute a creare sono affidate all'Indagine Campionaria sulle Nascite, che rappresenta un'assoluta novità nel settore delle statistiche demografiche e la cui prima edizione è stata effettuata nel 2002 con tecnica CATI (i principali risultati sono pubblicati nel volume *"Avere un figlio in Italia Approfondimenti tematici dall'Indagine campionaria sulle nascite Anno 2002 – Settore Informazioni"*, http://www3.istat.it/dati/catalogo/20061220_00/)¹; una seconda edizione è stata realizzata nel 2005 sempre con tecnica CATI e una terza edizione, rinnovata nella tecnica e nei contenuti, nel 2012.

¹ Per maggiori informazioni sugli aspetti metodologici è possibile consultare il volume *"Indagine Campionaria sulle Nascite: obiettivi, metodologia e organizzazione, Anno 2002 - Settore Metodi e Norme"* http://www.istat.it/dati/catalogo/20060317_00/.

Nella terza edizione, infatti, si è scelto di introdurre delle importanti novità al processo produttivo realizzando, oltre all'Indagine trasversale che garantiva la comparabilità col passato, un'indagine sulle madri straniere, a integrazione dell'indagine trasversale, e una retrospettiva.

I contenuti dell'Indagine, nella terza edizione, sono stati arricchiti anche grazie all'apporto dell'Isfol che, nell'ambito di una Convenzione Istat-Isfol del 2008, progetto "Maternità e partecipazione femminile al mercato del lavoro", ha collaborato alla riprogettazione dell'indagine anche in un'ottica retrospettiva, in modo da poter analizzare sia i comportamenti riproduttivi delle donne con almeno un figlio, sia l'interazione maternità-lavoro nel medio-lungo periodo.

In particolare, dunque, la terza edizione dell'Indagine è stata realizzata mediante 3 questionari distinti: uno *trasversale* (**CATI Long Form**), con tecnica CATI, somministrato alle madri dei nati iscritti in anagrafe tra luglio 2009 e giugno 2010 (circa 17.800 interviste), uno sulle *madri straniere* (circa 1.600 interviste), con tecnica PAPI, somministrato alle madri dei nati stranieri iscritti in anagrafe tra luglio 2009 e giugno 2010 e, infine, uno *retrospettivo* (circa 12.500 interviste), con tecnica CATI, somministrato ad un campione di madri già intervistate nel 2005.

L'Indagine trasversale e quella delle madri straniere sono state 'unite' (**CATI+PAPI STRANIERE**) ed è stato costruito un peso unico rappresentativo del totale dei nati iscritti in Anagrafe nel periodo di riferimento; questa unione consente analisi rappresentative per macro-area di cittadinanza.

I principali risultati sono disponibili nella pubblicazione: *Avere figli in Italia negli anni 2000 - Approfondimenti dalle indagini campionarie sulle nascite e sulle madri*, <http://www.istat.it/it/archivio/147180>

Nell'indagine retrospettiva le madri sono state ricontattate dopo circa sette anni al fine di aggiornare le loro variabili socio-demografiche e quelle dei membri della loro famiglia. Particolare attenzione è stata dedicata al monitoraggio del comportamento riproduttivo e alle eventuali variazioni intercorse nella condizione professionale a partire dalla gravidanza del primo figlio.

Il questionario è articolato in 4 sezioni. Di seguito verranno illustrati brevemente i contenuti di ciascuna di queste.

Sezione 1

La prima sezione è volta all'acquisizione delle principali variabili di interesse demografico della madre e del suo attuale partner.

Sono inoltre presenti dei quesiti sull'unione (sia essa matrimonio o convivenza) della madre attualmente in corso mentre verranno approfonditi attraverso l'utilizzo di quesiti retrospettivi gli aspetti relativi alle eventuali unioni precedenti della donna (convivenze, matrimoni e loro scioglimenti). Una sotto-sezione è dedicata alla storia riproduttiva: verranno qui rilevate le principali informazioni demografiche su tutti i figli conviventi (inclusi i figli adottati, affidati o esclusivamente del partner) e anche su quelli non conviventi della donna. La sezione si chiude con i quesiti relativi ai progetti riproduttivi di breve e lungo periodo.

Tale sezione è articolata, infatti, nelle seguenti sotto-sezioni:

- ☐ La madre e il contesto familiare;
- ☐ Il partner;
- ☐ Storia delle unioni;
- ☐ Storia riproduttiva, figli conviventi e non conviventi e le intenzioni di fecondità.

Sezione 2

In questa sezione si vogliono acquisire informazioni sull'attuale condizione professionale della madre e sulle iniziative di formazione professionale cui la donna ha partecipato negli ultimi sette anni. Il questionario, poi, si articola in più percorsi per tener conto delle specificità emerse in termini di condizione professionale attuale. Così come nella sezione precedente, verranno utilizzati dei quesiti

retrospettivi per ricostruire l'intera carriera lavorativa dell'intervistata a partire dalla sua condizione professionale precedente la prima gravidanza.

La sezione 2 è articolata nelle seguenti sotto-sezioni:

- ❑ La formazione e il lavoro attuale della madre;
- ❑ Soddisfazione sul lavoro e intenzioni lavorative future;
- ❑ Intenzioni lavorative future (solo per le donne attualmente non occupate);
- ❑ Storia lavorativa (solo per le donne, attualmente occupate o non occupate, che hanno lavorato per almeno un anno nel corso della loro vita).

Sezione 3

Questa sezione è dedicata alla rilevazione delle informazioni sull'organizzazione familiare messa in atto per provvedere alle esigenze di accudimento dei figli e alla cura della casa. I quesiti riguardano i seguenti temi: la divisione del lavoro familiare e gli aiuti.

Sezione 4

Tale sezione serve a contestualizzare le informazioni socio-demografiche fin qui acquisite, rilevando le principali caratteristiche dell'abitazione e del contesto socio-economico. Sono stati introdotti anche dei quesiti sulla soddisfazione della donna rispetto ai diversi aspetti della sua vita quotidiana (situazione economica, salute, relazione con amici, ...).

Questa sezione contiene una sola sotto-sezione: l'abitazione e il contesto socio-economico.

1.2 Informazioni errate o incompatibili

Uno degli aspetti principali nell'espletamento di un'indagine campionaria è quello che riguarda la qualità dei dati dal punto di vista della correttezza e della coerenza delle informazioni raccolte.

Varie sono le possibili cause che introducono errori durante l'intervista e altrettanto varie sono le strategie che permettono di limitarne l'introduzione. Ad esempio gli errori possono derivare dalla reticenza o dalla mancanza di interesse e/o di attenzione dei soggetti intervistati; per motivare le madri in merito alla rilevanza della loro collaborazione all'indagine e rammentare l'obbligo di risposta è stata loro inviata una lettera a firma del Presidente dell'Istat che preannunciava l'intervista e illustrava i contenuti e gli scopi dell'indagine.

Anche la difficoltà nel ricordare può essere causa di risposte errate o mancate risposte; per questo motivo si è cercato di limitare, per quanto possibile, i quesiti relativi a eventi lontani nel tempo. Ancora, fonte di errore può essere anche l'operato delle intervistatrici, che possono registrare valori non corretti o, nel caso di questionari cartacei, possono gestire i percorsi in maniera errata.

Gli errori che si riscontrano nei dati di un'indagine possono essere sia casuali sia sistematici; quelli casuali non portano a distorsioni nelle stime finali. Gli errori sistematici, invece, tendono a concentrarsi solo in alcune variabili o modalità di risposta, hanno sempre lo stesso segno e ogni ripetizione dell'indagine ne è affetta. Questi errori, quindi, causano distorsioni nei risultati finali e devono, dunque, essere attentamente tenuti sotto controllo grazie a indicatori di monitoraggio appositamente costruiti.

Il controllo e la correzione degli errori non campionari, che possono presentarsi in ogni fase del processo produttivo, richiede una strategia complessa. In questo ambito, poiché ogni metodologia di correzione a posteriori risolve solo parzialmente e a volte in modo non del tutto soddisfacente il problema, è fondamentale la prevenzione e la correzione degli errori contestualmente all'acquisizione dei dati; l'utilizzo del Cati, dunque, è di notevole supporto perché consente di:

- evitare errori e incompatibilità fra variabili gestendo in maniera automatica la navigazione condizionata all'interno del questionario;

- stabilire il range ammesso per ciascuna variabile, evitando che gli intervistatori inseriscano valori non ammessi e, quindi, errati;
- effettuare controlli di coerenza fra le risposte inserite durante l'intervista;
- utilizzare una codifica assistita, supportando gli intervistatori nel momento in cui debbano, per esempio, inserire un codice comunale (è sufficiente scrivere qualche lettera del nome del Comune per avere già a disposizione il relativo codice comunale da inserire) o un titolo di studio.

Grazie a questi controlli le interviste Cati presentano una elevata qualità in termini di correttezza e coerenza delle risposte fornite dalle intervistate.

Nel dettaglio, le correzioni hanno riguardato sia le eventuali incoerenze delle singole variabili o tra variabili logicamente collegate, sia la gestione delle mancate risposte, rappresentate dall'assenza di risposta ad uno o più quesiti nell'ambito di un'intervista effettuata.

Gli approcci per la correzione di questo tipo di errori sono stati sia di tipo probabilistico che di tipo deterministico: in generale, l'approccio probabilistico prevede la definizione delle condizioni di errore e la correzione avviene a seguito dell'applicazione di un algoritmo probabilistico; l'approccio deterministico prevede invece che, a priori, vengano stabilite le condizioni di errore e le azioni da intraprendere per ciascuna di esse. Le regole impiegate nell'approccio deterministico sono del tipo:

SE (condizione di errore) ALLORA (azione di correzione).

È stato quindi necessario stabilire il valore "corretto" da assegnare alla variabile per la quale si è verificata la condizione di errore.

1.3 Individuazione e correzione degli errori

Le interviste sono state condotte mediante tecnica CATI (il questionario elettronico è stato sviluppato in-house mediante l'utilizzo del software Blaise); in questo modo è stato possibile prevedere svincoli e filtri automatici, come anche controlli in fase di inserimento delle risposte e verifiche di compatibilità fra variabili. L'utilizzo di un questionario elettronico ha indubbiamente permesso di avere a disposizione dati di partenza di buona qualità; ciononostante sono stati effettuati controlli sulla correttezza e compatibilità delle informazioni raccolte.

I controlli hanno riguardato gli errori di percorso (se, per esempio, la madre è single non deve rispondere ai quesiti sul partner), gli errori dovuti a valori fuori dominio (se per un quesito sono previsti valori che vanno da 1 a 5, l'operatore che effettua l'intervista non potrà registrare il valore 6) e le incompatibilità (ad es. la madre non può indicare un anno di acquisizione del titolo di studio precedente la sua data di nascita).

L'aver sviluppato internamente all'Istat il questionario elettronico ha permesso, come si è detto, di prevedere numerosi controlli di *range* e di coerenza; è stato inoltre possibile verificare immediatamente la veridicità delle risposte fornite. Ne deriva che i dati dell'indagine Cati hanno richiesto un numero molto contenuto di interventi di correzione ex-post. Questi interventi si sono limitati essenzialmente ai casi in cui si è scelto di non inserire regole su alcuni quesiti e di riservarsi il controllo e la correzione alla fase successiva alle interviste. Questa scelta è stata dettata dalla necessità di non appesantire il questionario rendendone difficoltosa la navigazione durante l'intervista.

In media sono stati riscontrati errori sul 6% per cento dei record. Quasi la totalità è costituito da errori di percorso (90%) a cui seguono gli errori di incoerenza (11%). I controlli di *range*, già inseriti alla fonte, non hanno evidenziato la necessità di apportare ulteriori correzioni.

2. Strategia di campionamento e livello di precisione delle stime

2.1 Obiettivi dell'indagine

La popolazione di interesse dell'indagine – ossia l'insieme delle unità statistiche relativamente alle quali si intende investigare – è costituita dalle Madri di nati iscritti in Anagrafe nel 2003 intervistate nel 2005.

I parametri di popolazione oggetto di stima sono costituiti da totali e frequenze relative riferite anche a particolari sottoclassi della popolazione.

I domini di studio, ossia gli ambiti rispetto ai quali sono riferite le stime di detti parametri di popolazione, sono le tre classi di età della madre (≤ 29 ; 30-34; ≥ 35), i due ordini di nascita (primogenito, altro) e, da un punto di vista territoriale, le tre ripartizioni geografiche (Nord, Centro, Sud e Isole). Le stime dell'indagine, pertanto, sono prodotte con riferimento a tali domini o ad incroci e aggregazioni ottenibili a partire da questi.

2.2 Disegno di campionamento

L'indagine in oggetto si configura come un'indagine retrospettiva basata su un campione di madri che hanno preso parte all'indagine campionaria sulle nascite, che hanno cioè avuto un figlio nel 2003 e che sono state re-intervistate nel 2005 (49.385). La dimensione effettiva del campione è stata di 26.250 unità, una numerosità raggiunta per sovradimensionamento del campione teorico iniziale costituito da 12.465 unità. La numerosità finale del campione è stata ottenuta tenendo conto dei tassi di risposta rilevati all'indagine del 2005.

Le unità della popolazione di interesse dell'indagine sono state stratificate tenendo conto dell'incrocio delle modalità di tre variabili misurate con riferimento al bambino per il quale la madre era stata intervistata nel 2005: la classe di età della madre (≤ 29 ; 30-34; ≥ 35), l'ordine di nascita (primogenito, altro) e la ripartizione geografica (Nord, Centro, Sud e Isole). Complessivamente è stato definito un numero di strati pari a 18.

Per l'allocatione del campione negli strati è stata utilizzata una metodologia basata su una generalizzazione del metodo di allocatione multivariata di Bethel al caso di più tipologie di domini di stima. Per tale studio sono state utilizzate alcune variabili target dell'indagine effettuata nel 2005 e gli errori di campionamento attesi delle stime ad esse afferenti sono stati definiti per i domini di stima sopra descritti.

Una volta definite le numerosità campionarie teoriche negli strati, la selezione delle unità campionarie è stata effettuata senza reimmissione e con probabilità uguali.

2.3 La procedura di stima e calcolo degli errori campionari

Le stime dei parametri di popolazione sono ottenute mediante uno stimatore di ponderazione vincolata (o di calibrazione). Supposto che il parametro oggetto di stima sia il totale della variabile Y in un generico dominio d , espresso dalla relazione

$$Y_d = \sum_{h=1}^{H_d} \sum_{i=1}^{N_h} Y_{hi} \delta_{hid} ,$$

in cui δ_{hid} è una variabile indicatrice che assume valore 1 se l'unità i dello strato h appartiene al dominio di interesse d , una stima del totale Y_d è data dalla seguente espressione:

$$\hat{Y}_d = \sum_{h=1}^{H_d} \sum_{i=1}^{n_h} Y_{hi} \delta_{hid} \tilde{w}_{hi} ,$$

dove \tilde{w}_{hi} rappresenta il peso finale da attribuire all'individuo campione i dello strato h .

Il peso finale delle unità campionarie è ottenuto per mezzo di una procedura complessa che si sviluppa in più fasi volte alla:

- (i) definizione del peso diretto associato alle unità campionarie;

- (ii) determinazione di un fattore correttivo che attenua l'effetto distorsivo dovuto al fenomeno della mancata risposta totale;
- (iii) costruzione di un fattore di ponderazione che tiene conto della conoscenza di alcuni totali noti sulla popolazione oggetto di studio.

Il peso diretto della i -ma unità campionaria selezionata nel 2012 nello strato h ($h=1,\dots,H$) è dato da:

$$w_{hi} = w_i^* \left(\frac{n_h}{n_h^*} \right)^{-1},$$

dove: w_i^* è il peso campionario associato alla i -ma unità derivante dal disegno adottato nell'indagine del 2005; $\frac{n_h}{n_h^*}$ è la probabilità di inclusione delle unità campionarie nello strato h (campione definito per il 2012); n_h e n_h^* rappresentano rispettivamente il numero di unità campionarie (campione 2012) e il numero di unità rispondenti all'indagine del 2005 nello strato h .

Poiché durante la rilevazione parte delle unità campionarie non hanno fornito una risposta, nella fase di stima il peso base associato alle unità rispondenti è stato modificato per tener conto dei non rispondenti.

Per la correzione della mancata risposta totale si è fatto riferimento a metodi di correzione dei pesi campionari, detti tecniche di riponderazione, che sono fondati sull'uso di informazioni ausiliarie legate alle probabilità di risposta e note sia per le unità rispondenti che per le unità non rispondenti. Tali tecniche assumono l'esistenza di un legame funzionale tra la propensione delle unità del campione ad essere rispondenti e le caratteristiche da esse possedute.

Nel caso specifico, per la correzione dei pesi campionari delle unità rispondenti, è stato prima stimato un modello di risposta non parametrico, basato su un algoritmo di classificazione ad albero (CART), e successivamente sono stati calcolati i fattori correttivi della mancata risposta totale come inverso del tasso di risposta osservato all'interno dei nodi terminali (celle) individuati attraverso l'albero di classificazione risultato ottimale.

Le variabili ausiliarie utilizzate per lo studio del modello si riferiscono a caratteristiche della madre (età, stato civile, titolo di studio, condizione professionale, titolo di godimento dell'abitazione e cittadinanza), alla tipologia familiare (numero di componenti della famiglia) e, infine, alla ripartizione geografica dell'Anagrafe in cui è iscritto il bambino oggetto della rilevazione nel 2005.

I nodi terminali, o celle, dell'albero di classificazione stimato sono risultati pari a 5. Tali celle sono costituite sia da unità rispondenti che da unità non rispondenti: la prima cella identifica le madri con età inferiore a 30, la seconda le madri con età superiore a 30 che vivono in affitto/subaffitto, la terza le madri con età superiore a 30 che non vivono in affitto/subaffitto e che hanno conseguito un titolo di studio medio/alto, la quarta le madri con età superiore a 30 che non vivono in affitto/subaffitto, che hanno conseguito un titolo di studio basso e che vivono nelle ripartizioni geografiche Nord-Ovest, Sud e Isole, infine, la quinta identifica le madri con età superiore a 30 che non vivono in affitto/subaffitto, che hanno conseguito un titolo di studio basso e che vivono nel Nord-Est e nel Centro Italia.

Il fattore correttivo della mancata risposta totale è dunque dato per ogni cella g ($g=1,\dots,G$) da

$$f_g = \left(\frac{n_g^r}{n_g} \right)^{-1},$$

dove n_g^r e n_g rappresentano rispettivamente il numero di rispondenti e il numero di unità campionarie nella g -ma cella.

Il peso base corretto per la mancata risposta totale è ottenuto moltiplicando il peso diretto w_{hi} con il fattore correttivo f_g .

I coefficienti di riporto finali \tilde{w}_{hi} sono ottenuti tramite una procedura di calibrazione, applicata al peso base corretto, che consente di soddisfare la condizione di uguaglianza tra i totali noti di alcune variabili ausiliarie e le corrispondenti stime campionarie. Le variabili ausiliarie di cui si è tenuto conto in questa fase sono le stesse

considerate nella stratificazione della popolazione di interesse (classe d'età, ordine di nascita e ripartizione geografica). La procedura di calibrazione è stata realizzata utilizzando il software generalizzato Regenesees messo a punto dall'Istat (Zardetto, 2015²).

Per la valutazione della precisione delle stime è stato determinato l'errore di campionamento relativo delle stime di interesse dell'indagine nei domini di stima definiti per diversi livelli territoriali (nazionale, ripartizionale), per classe di età della madre, ordine di nascita e per gli incroci ordine di nascita e ripartizione geografica e, infine, classe di età e ordine di nascita. Per la determinazione degli errori campionari, relativi a totali, stime di frequenze relative, stime di frequenze relative per sottoclassi della popolazione, è stato utilizzato il software generalizzato Regenesees.

L'errore di campionamento relativo per la generica stima \hat{Y}_d nel dominio d assume la seguente espressione

$$\hat{\varepsilon}(\hat{Y}_d) = \frac{\hat{\sigma}(\hat{Y}_d)}{\hat{Y}_d},$$

dove $\hat{\sigma}(\hat{Y}_d) = \sqrt{\hat{V}(\hat{Y}_d)}$ è l'errore di campionamento assoluto di \hat{Y}_d .

2.3.1 Presentazione sintetica degli errori campionari

Ad ogni stima \hat{Y}_d è associato un errore campionario relativo $\hat{\varepsilon}(\hat{Y}_d)$ e quindi, per consentire un uso corretto delle stime fornite dall'indagine, sarebbe necessario presentare, per ogni stima pubblicata, anche il corrispondente errore di campionamento relativo. Tuttavia, non è possibile soddisfare questa esigenza di informazione, sia per motivi di tempi e di costi di elaborazione, sia perché le tavole della pubblicazione risulterebbero eccessivamente appesantite e di non agevole consultazione per l'utente finale. Inoltre, non sarebbero in ogni caso disponibili gli errori delle stime non pubblicate, che l'utente può ricavare in modo autonomo.

Per questi motivi, generalmente, si ricorre ad una presentazione sintetica degli errori relativi, basata sul metodo dei modelli regressivi. Tale metodo si basa sulla determinazione di una funzione matematica che mette in relazione ciascuna stima con il proprio errore relativo.

Il modello utilizzato per le stime di frequenze assolute o relative è il seguente:

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

dove i parametri a e b vengono stimati mediante il metodo dei minimi quadrati.

Dalla precedente espressione, mediante semplici passaggi, si ricava la formula per il calcolo dell'errore relativo di una stima di frequenza (assoluta o relativa):

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

Una volta calcolato l'errore relativo è possibile costruire l'intervallo di confidenza delle stime al 95% come:

$$\{\hat{Y}_d^* - 1,96 \cdot \hat{\varepsilon}(\hat{Y}_d^*) \cdot \hat{Y}_d^*; \hat{Y}_d^* + 1,96 \cdot \hat{\varepsilon}(\hat{Y}_d^*) \cdot \hat{Y}_d^*\}.$$

Allo scopo di facilitare il calcolo degli errori campionari, nei prospetti dell'appendice A sono riportati, oltre ai parametri dei modelli, gli errori relativi percentuali corrispondenti a valori crescenti di stime di frequenze

² 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

assolute (o relative).

Le informazioni riportate nei prospetti (coefficienti a e b del modello) consentono di calcolare l'errore relativo di una generica stima di frequenza assoluta o relativa. Il metodo descritto per il calcolo dell'errore relativo consiste nell'approssimare l'errore relativo della stima di interesse corrispondente al livello di stima che più vi si avvicina. Un secondo metodo, più preciso del primo, si basa sull'uso di una formula di interpolazione lineare per il calcolo degli errori di stime non comprese tra i valori forniti nei prospetti. In tal caso, l'errore campionario della stima \hat{Y}_d^* , si ricava mediante l'espressione:

$$\hat{\varepsilon}(\hat{Y}_d^*) = \hat{\varepsilon}(\hat{Y}_d^{k-1}) - \frac{\hat{\varepsilon}(\hat{Y}_d^{k-1}) - \hat{\varepsilon}(\hat{Y}_d^k)}{\hat{Y}_d^k - \hat{Y}_d^{k-1}} (\hat{Y}_d^* - \hat{Y}_d^{k-1})$$

dove \hat{Y}_d^{k-1} e \hat{Y}_d^k sono i valori delle stime entro i quali è compresa la stima \hat{Y}_d^* , mentre $\hat{\varepsilon}(\hat{Y}_d^{k-1})$ e $\hat{\varepsilon}(\hat{Y}_d^k)$ sono i corrispondenti errori relativi.

Tavola 1. - Valori dei coefficienti a e b e dell'indice di determinazione R² (%) del modello per l'interpolazione degli errori campionari delle stime riferite ai nati per ripartizione geografica, classe di età della madre e ordine di nascita. Valori assoluti (*)

DOMINIO DI STIMA		a	b	R ²
ITALIA		6,98	-1,27	96,27
RIPARTIZIONI GEOGRAFICHE				
Nord		6,57	-1,24	94,92
Centro		5,88	-1,20	94,01
Sud e Isole		6,35	-1,25	95,88
CLASSI DI ETÀ DELLA MADRE AL PARTO				
Fino a 29		6,37	-1,25	96,27
30-34		7,11	-1,29	95,86
35 e oltre		5,51	-1,16	93,57
ORDINE DI NASCITA				
1° figlio		6,86	-1,26	96,13
2° figlio e oltre		6,09	-1,21	94,50
RIPARTIZIONE GEOGRAFICA E ORDINE DI NASCITA				
Nord	1° figlio	6,26	-1,22	94,17
Centro	1° figlio	6,58	-1,27	94,76
Sud	1° figlio	6,14	-1,25	95,89
Nord	2° figlio e oltre	5,47	-1,15	91,97
Centro	2° figlio e oltre	5,49	-1,21	93,50
Sud	2° figlio e oltre	5,63	-1,20	94,06
ORDINE DI NASCITA E CLASSI DI ETÀ DELLA MADRE AL PARTO				
1° figlio	Fino a 29	7,10	-1,32	0,95
2° figlio e oltre	30-34	5,39	-1,23	0,91
1° figlio	35 e oltre	7,40	-1,32	0,95
2° figlio e oltre	Fino a 29	6,80	-1,31	0,94
1° figlio	30-34	5,81	-1,25	0,94
2° figlio e oltre	35 e oltre	6,36	-1,24	0,93

(*) Le variabili sono riferite al bambino per il quale la madre era stata intervistata nel 2005.

Tavola 2. - Valori dei coefficienti a e b e dell'indice di determinazione R² (%) del modello per l'interpolazione degli errori campionari delle stime riferite ai nati per ripartizione geografica, classe di età della madre e ordine di nascita. Frequenze relative (*)

DOMINIO DI STIMA		a	b	R ²
ITALIA		-9,73	-1,27	96,27
RIPARTIZIONI GEOGRAFICHE				
Nord		-8,70	-1,24	94,92
Centro		-7,90	-1,20	94,01
Sud e Isole		-8,96	-1,25	95,88
CLASSI DI ETÀ DELLA MADRE AL PARTO				
Fino a 29		-8,91	-1,25	96,27
30-34		-8,64	-1,29	95,86
35 e oltre		-8,15	-1,16	93,57
ORDINE DI NASCITA				
1° figlio		-8,97	-1,26	96,13
2° figlio e oltre		-9,03	-1,21	94,50
RIPARTIZIONI GEOGRAFICHE E ORDINE DI NASCITA				
Nord	1° figlio	-8,00	-1,22	94,17
Centro	1° figlio	-7,22	-1,27	94,76
Sud	1° figlio	-8,13	-1,25	95,89
Nord	2° figlio e oltre	-7,84	-1,15	91,97
Centro	2° figlio e oltre	-7,46	-1,21	93,50
Sud	2° figlio e oltre	-8,32	-1,20	94,06
ORDINE DI NASCITA E CLASSI DI ETÀ DELLA MADRE AL PARTO				
1° figlio	Fino a 29	-8,50	-1,32	0,95
2° figlio e oltre	30-34	-8,21	-1,23	0,91
1° figlio	35 e oltre	-7,73	-1,32	0,95
2° figlio e oltre	Fino a 29	-8,29	-1,31	0,94
1° figlio	30-34	-7,45	-1,25	0,94
2° figlio e oltre	35 e oltre	-7,83	-1,24	0,93

(*) Le variabili sono riferite al bambino per il quale la madre era stata intervistata nel 2005.

Tavola 3. -Valori interpolati degli errori campionari relativi percentuali delle stime riferite ai nati per ripartizione geografica. Valori assoluti (*) ()**

STIMA	RIPARTIZIONE GEOGRAFICA			Italia
	Nord	Centro	Sud e Isole	
500	57,4	45,4	48,8	63,8
1.000	37,4	29,9	31,6	41,1
2.500	21,2	17,3	17,8	23,0
5.000	13,8	11,4	11,5	14,8
10.000	9,0	7,5	7,5	9,6
25.000	5,1	4,3	4,2	5,3
50.000	3,3	2,9	2,7	3,4
75.000	2,6	2,2	2,1	2,7
100.000	2,2	1,9	1,8	2,2
200.000	1,4	1,2	1,1	1,4
300.000				1,1
500.000				0,8

(*) Le variabili sono riferite al bambino per il quale la madre era stata intervistata nel 2005.

(**) Oltre la soglia del 33% le stime sono da considerarsi poco attendibili.

Tavola 4. -Valori interpolati degli errori campionari relativi percentuali delle stime riferite ai nati per ripartizione geografica. Frequenze relative (*) ()**

STIMA	RIPARTIZIONE GEOGRAFICA			Italia
	Nord	Centro	Sud e Isole	
0,005	34,2	46,3	31,3	
0,01	22,3	30,6	20,3	14,3
0,02	14,5	20,2	13,1	9,2
0,05	8,2	11,6	7,4	5,1
0,1	5,4	7,7	4,8	3,3
0,2	3,5	5,1	3,1	2,1
0,3	2,7	4,0	2,4	1,7
0,4	2,3	3,3	2,0	1,4
0,5	2,0	2,9	1,8	1,2
0,6	1,8	2,6	1,6	1,1
0,7	1,6	2,4	1,4	1,0
0,8	1,5	2,2	1,3	0,9
0,9	1,4	2,1	1,2	0,8

(*) Le variabili sono riferite al bambino per il quale la madre era stata intervistata nel 2005.

(**) Oltre la soglia del 33% le stime sono da considerarsi poco attendibili.

Tavola 5. -Valori interpolati degli errori campionari relativi percentuali delle stime riferite ai nati per classe di età della madre. Valori assoluti (*) ()**

STIMA	CLASSE DI ETA'		
	Fino a 29	30-34	35 e oltre
500	49,8	63,2	42,8
1.000	32,3	40,4	28,7
2.500	18,2	22,3	16,8
5.000	11,8	14,3	11,3
10.000	7,7	9,1	7,5
25.000	4,3	5,0	4,4
50.000	2,8	3,2	3,0
75.000	2,2	2,5	2,3
100.000	1,8	2,1	2,0
200.000	1,2	1,3	1,3

(*) Le variabili sono riferite al bambino per il quale la madre era stata intervistata nel 2005.

(**) Oltre la soglia del 33% le stime sono da considerarsi poco attendibili.

Tavola 6. -Valori interpolati degli errori campionari relativi percentuali delle stime riferite ai nati per classe di età della madre. Frequenze relative (*) ()**

STIMA	CLASSE DI ETA'		
	Fino a 29	30-34	35 e oltre
0,005	31,9	40,8	36,7
0,01	20,7	26,1	24,6
0,02	13,4	16,6	16,4
0,05	7,6	9,2	9,7
0,1	4,9	5,9	6,5
0,2	3,2	3,8	4,3
0,3	2,5	2,9	3,4
0,4	2,1	2,4	2,9
0,5	1,8	2,1	2,5
0,6	1,6	1,8	2,3
0,7	1,5	1,7	2,1
0,8	1,3	1,5	1,9
0,9	1,2	1,4	1,8

(*) Le variabili sono riferite al bambino per il quale la madre era stata intervistata nel 2005.

(**) Oltre la soglia del 33% le stime sono da considerarsi poco attendibili.

Tavola 7. -Valori interpolati degli errori campionari relativi percentuali delle stime riferite ai nati per ordine di nascita. Valori assoluti (*) ()**

STIMA	ORDINE DI NASCITA	
	1° figlio	2° figlio e oltre
500	60,5	48,4
1.000	39,0	31,8
2.500	21,9	18,2
5.000	14,1	12,0
10.000	9,1	7,9
25.000	5,1	4,5
50.000	3,3	3,0
75.000	2,5	2,3
100.000	2,1	1,9
200.000	1,4	1,3

(*) Le variabili sono riferite al bambino per il quale la madre era stata intervistata nel 2005.

(**) Oltre la soglia del 33% le stime sono da considerarsi poco attendibili.

Tavola 8. -Valori interpolati degli errori campionari relativi percentuali delle stime riferite ai nati per ordine di nascita. Frequenze relative (*)

STIMA	ORDINE DI NASCITA	
	1° figlio	2° figlio e oltre
0,005	32,1	27,2
0,01	20,7	17,8
0,02	13,4	11,7
0,05	7,5	6,7
0,1	4,8	4,4
0,2	3,1	2,9
0,3	2,4	2,3
0,4	2,0	1,9
0,5	1,7	1,7
0,6	1,6	1,5
0,7	1,4	1,4
0,8	1,3	1,2
0,9	1,2	1,2

(*) Le variabili sono riferite al bambino per il quale la madre era stata intervistata nel 2005.

Tavola 9. -Valori interpolati degli errori campionari relativi percentuali delle stime riferite ai nati per ripartizione geografica e ordine di nascita. Valori assoluti (*) ()**

STIMA	ORDINE DI NASCITA E CLASSI DI ETÀ DELLA MADRE AL PARTO					
	Nord		Centro		Sud e Isole	
	1° figlio	2° figlio e oltre	1° figlio	2° figlio e oltre	1° figlio	2° figlio e oltre
500	52,5	43,2	51,5	36,5	45,0	39,6
1.000	34,5	29,0	33,1	24,0	29,3	26,1
2.500	19,7	17,1	18,5	13,8	16,5	15,0
5.000	13,0	11,5	11,9	9,1	10,7	9,9
10.000	8,5	7,7	7,6	6,0	7,0	6,5
25.000	4,9	4,5	4,3	3,4	3,9	3,8
50.000	3,2	3,1	2,7	2,3	2,6	2,5
75.000	2,5	2,4	2,1	1,8	2,0	1,9
100.000	2,1	2,0	1,8	1,5	1,7	1,6
200.000	1,4	1,4	1,1	1,0	1,1	1,1

(*) Le variabili sono riferite al bambino per il quale la madre era stata intervistata nel 2005.

(**) Oltre la soglia del 33% le stime sono da considerarsi poco attendibili.

Tavola 10. -Valori interpolati degli errori campionari relativi percentuali delle stime riferite ai nati per ripartizione geografica e ordine di nascita. Frequenze relative (*) ()**

STIMA	ORDINE DI NASCITA E CLASSI DI ETÀ DELLA MADRE AL PARTO					
	Nord		Centro		Sud e Isole	
	1° figlio	2° figlio e oltre	1° figlio	2° figlio e oltre	1° figlio	2° figlio e oltre
0,005	45,7	41,7	78,8	58,9	46,4	37,8
0,01	30,0	28,0	50,7	38,7	30,1	24,9
0,02	19,7	18,8	32,6	25,5	19,6	16,4
0,05	11,3	11,1	18,2	14,7	11,1	9,5
0,1	7,4	7,4	11,7	9,6	7,2	6,2
0,2	4,9	5,0	7,5	6,3	4,7	4,1
0,3	3,8	4,0	5,8	5,0	3,6	3,2
0,4	3,2	3,4	4,8	4,2	3,0	2,7
0,5	2,8	2,9	4,2	3,7	2,6	2,4
0,6	2,5	2,7	3,7	3,3	2,4	2,1
0,7	2,3	2,4	3,4	3,0	2,1	1,9
0,8	2,1	2,3	3,1	2,7	2,0	1,8
0,9	1,9	2,1	2,9	2,6	1,8	1,7

(*) Le variabili sono riferite al bambino per il quale la madre era stata intervistata nel 2005.

(**) Oltre la soglia del 33% le stime sono da considerarsi poco attendibili.

Tavola 11. -Valori interpolati degli errori campionari relativi percentuali delle stime riferite ai nati per ordine di nascita e classe di età della madre al parto. Valori assoluti (*) ()**

STIMA	ORDINE DI NASCITA E CLASSE DI ETÀ DELLA MADRE AL PARTO					
	1° figlio			2° figlio e oltre		
	Fino a 29	30-34	35 e oltre	Fino a 29	30-34	35 e oltre
500	58,2	32,4	66,1	51,8	37,3	50,5
1.000	36,9	21,1	41,8	33,0	24,1	32,8
2.500	20,2	12,0	22,8	18,1	13,6	18,6
5.000	12,8	7,9	14,4	11,5	8,8	12,1
10.000	8,1	5,1	9,1	7,3	5,7	7,8
25.000	4,4	2,9	5,0	4,0	3,2	4,4
50.000	2,8	1,9	3,1	2,6	2,1	2,9
75.000	2,2	1,5	2,4	2,0	1,6	2,2
100.000	1,8	1,2	2,0	1,6	1,3	1,9
200.000	1,1	0,8	1,3	1,0	0,9	1,2

(*) Le variabili sono riferite al bambino per il quale la madre era stata intervistata nel 2005.

(**) Oltre la soglia del 33% le stime sono da considerarsi poco attendibili.

Tavola 12 - Valori interpolati degli errori campionari relativi percentuali delle stime riferite ai nati per ordine di nascita e classe di età della madre al parto. Frequenze relative (*) ()**

STIMA	ORDINE DI NASCITA E CLASSE DI ETÀ DELLA MADRE AL PARTO					
	1° figlio			2° figlio e oltre		
	Fino a 29	30-34	35 e oltre	Fino a 29	30-34	35 e oltre
0,005	46,5	42,8	69,8	50,3	66,7	53,6
0,01	29,5	28,0	44,1	32,0	43,2	34,9
0,02	18,7	18,3	27,9	20,3	28,0	22,7
0,05	10,2	10,4	15,2	11,2	15,8	12,8
0,1	6,5	6,8	9,6	7,1	10,2	8,3
0,2	4,1	4,4	6,1	4,5	6,6	5,4
0,3	3,1	3,5	4,6	3,5	5,1	4,2
0,4	2,6	2,9	3,8	2,9	4,3	3,5
0,5	2,2	2,5	3,3	2,5	3,7	3,1
0,6	2,0	2,3	2,9	2,2	3,3	2,7
0,7	1,8	2,0	2,7	2,0	3,0	2,5
0,8	1,6	1,9	2,4	1,8	2,8	2,3
0,9	1,5	1,8	2,2	1,7	2,6	2,1

(*) Le variabili sono riferite al bambino per il quale la madre era stata intervistata nel 2005.

(**) Oltre la soglia del 33% le stime sono da considerarsi poco attendibili.