

FILE DATI

***Indagine “Inclusione sociale delle
persone con limitazioni funzionali”
Anno 2011***

Manuale utente

INDAGINE INCLUSIONE SOCIALE DELLE PERSONE CON LIMITAZIONI FUNZIONALI
ANNO 2011
DOCUMENTAZIONE TECNICA E DESCRIZIONE DEL FILE

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 30/6/2003 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 sono state apportate alcune modifiche sui file originali delle indagini, nell'intento di garantire la massima protezione ai dati contenendo al minimo l'eventuale perdita di informazioni.

Le metodologie applicate si concretizzano nell'accorpamento e/o riclassificazione di modalità di variabili e nell'oscuramento di variabili. In quest'ultimo caso nei campi del tracciato record è riportata la dicitura **"RISERVATO ISTAT"**.

Va considerato, inoltre, che la stessa dicitura è stata utilizzata anche per quelle variabili non attendibili dal punto di vista campionario e quindi non analizzabili statisticamente.

FINALITÀ E CARATTERISTICHE DELL'INDAGINE

L'indagine sull'inclusione sociale delle persone con limitazioni funzionali è stata realizzata nel 2011 con la finalità di studiare l'integrazione sociale delle persone con problemi di salute e limitazioni funzionali, vale a dire difficoltà nel vedere, sentire, camminare, comunicare, svolgere le attività essenziali della vita quotidiana, ecc. Più in particolare, interesse dell'indagine è stato quello di studiare l'integrazione sociale di queste persone nei contesti scolastici e lavorativi, nelle reti di relazioni sociali e nelle attività del tempo libero e i fattori che ne ostacolano la piena partecipazione (limitazioni nella mobilità, barriere nell'accesso al lavoro, mancanza di adeguati sostegni, ecc).

Le varie aree tematiche che si susseguono nel questionario, coerentemente con gli obiettivi di intervento individuati con la normativa vigente e con gli strumenti internazionali per lo studio del fenomeno della disabilità, sono quindi le condizioni di salute, le limitazioni nelle attività, il contesto di vita delle persone (scuola, lavoro, interazioni sociali, mobilità e tempo libero), il ricorso ad aiuti e assistenza. La presenza di restrizioni alle opportunità di partecipazione nei diversi contesti di vita è studiata dal punto di vista di ciò che le persone desiderano, di ciò che vorrebbero fare e non possono fare per ragioni di diverso tipo, tra le quali le loro condizioni di salute, la presenza di barriere, altri fattori ambientali o personali.

Il collettivo di riferimento è costituito dalle persone intervistate in occasione dell'indagine "Condizioni di salute e ricorso ai servizi sanitari" condotta nel 2004-2005 che avevano un'età compresa tra i 6 e gli 80 anni al momento dell'intervista.

Tra di esse sono stati selezionati due gruppi di popolazione distinti:

1. *persone con limitazioni funzionali gravi*, vale a dire che, nel 2004-2005, anche con l'aiuto di ausili ed apparecchi sanitari, hanno riferito il massimo grado di difficoltà in almeno una delle funzioni della mobilità e della locomozione, delle funzioni essenziali della vita quotidiana (lavarsi, vestirsi, spogliarsi, mangiare, sedersi e alzarsi dal letto o dalla sedia), delle funzioni della comunicazione (vedere, sentire, parlare). Nelle indagini sulla salute condotte dal 1990 al 2005 questo gruppo è stato identificato come quello delle persone con disabilità.
2. *persone con limitazioni funzionali lievi*, vale a dire che, nel 2004-2005, anche con l'aiuto di ausili ed apparecchi sanitari, hanno riferito un livello minore di difficoltà in almeno una delle funzioni sopra indicate; Non sono stati selezionati i minori di 15 anni che hanno riferito soltanto qualche difficoltà in una o più funzioni della vita quotidiana: lavarsi, vestirsi, spogliarsi, mangiare, alzarsi o sedersi dal letto o dalla sedia. Ciò in quanto si è ritenuto che, per i bambini tra i 6 e i 14 anni, una lieve difficoltà in questo tipo di compiti, in assenza di limitazioni in altre aree funzionali, non indicasse necessariamente una condizione di riduzione delle capacità funzionali.

Non sono incluse le persone che durante l'indagine o nel corso dei controlli preliminari alle interviste sono risultate decedute, istituzionalizzate, trasferite all'estero o che nel 2011 hanno riferito limitazioni molto lievi che sono state considerate non rilevanti per gli obiettivi dell'indagine.

Sono state intervistate 3.121 persone, di età compresa tra gli 11 e gli 87 anni nel 2011, che rappresentano una popolazione di 3 milioni 947 mila individui.

La rilevazione è stata condotta mediante tecnica CATI (Computer Assisted Telephone Interview); le interviste telefoniche sono state somministrate direttamente al componente o ai componenti della famiglia selezionati dal campione dell'indagine "Condizioni di salute e ricorso ai servizi sanitari 2004-2005" con limitazioni funzionali se in grado di rispondere, oppure a familiari per i bambini di età inferiore ai 14 anni, nei casi in cui la persona da intervistare non era in grado di rispondere all'intervista (per particolari situazioni di disagio, malattia) o era assente dall'abitazione per tutto il periodo di tempo previsto per la rilevazione.

AVVERTENZE PER L'UTILIZZAZIONE DEL FILE

Per gli utenti esterni all'Istat vengono messi a disposizione dei file con le seguenti caratteristiche:

lunghezza record:	3.213
numero record individuali:	3.121
(uno per ciascuna persona intervistata)	

Ogni record contiene una prima parte di informazioni sull'individuo, una seconda parte comprendente caratteristiche strutturali relative a eventuali altri componenti della famiglia di fatto di appartenenza e, infine, variabili create (cioè non rilevate direttamente). L'unità di analisi, per le elaborazioni, è costituita dagli individui: ogni componente è individuato da un numero progressivo univoco.

La variabile "identificativo univoco di famiglia" è stata rilasciata dato che alcune unità del campione possono appartenere alla stessa famiglia di fatto.

Il numero totale di appartenenti al campione è pari al numero di record: 3.121.

COSTRUZIONE DELLE STIME ED ERRORI DI CAMPIONAMENTO

Le informazioni riportate nei file si riferiscono al collettivo selezionato. Per ottenere stime relative all'intera popolazione oggetto d'indagine è necessario moltiplicare per il coefficiente di riporto all'universo.

Ad ogni unità rispondente è stato assegnato un peso ottenuto come prodotto del peso campionario derivante dall'indagine "Condizioni di salute e ricorso ai servizi sanitari" per un fattore di correzione che tiene conto del mancato contatto e della mancata risposta.

Nel diffondere i risultati di un'indagine campionaria occorre fornire agli utilizzatori le informazioni necessarie per valutare l'attendibilità delle stime ottenibili. Ad ogni stima corrisponde un errore campionario relativo; ciò significa che per consentire un uso corretto delle stime sarebbe necessario fornire per ogni stima il corrispondente errore campionario relativo. Tuttavia, non è possibile pubblicare tutti gli errori di campionamento delle stime fornite e non sarebbero comunque disponibili gli errori delle stime non pubblicate che l'utente può ricavare in modo autonomo. Per questo si ricorre ad una presentazione sintetica degli errori tramite il metodo dei modelli regressivi. Questo metodo si basa sulla determinazione di una funzione matematica che mette in relazione ciascuna stima con il proprio errore di campionamento.

Si riporta in allegato l'appendice contenente le informazioni relative al campionamento e al calcolo degli errori di stima da cui è possibile effettuare il calcolo degli errori campionari.

Glossario

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

- **l'età** è dagli 11 agli 87 anni;
- **il titolo di studio** è quello più elevato conseguito;
- **I quesiti utilizzati per rilevare le limitazioni funzionali:**

I quesiti sono stati predisposti da un gruppo di lavoro dell'OECD sulla base della classificazione ICIDH (International Classification of Impairments, Disabilities and Handicaps) dell'Organizzazione Mondiale della Sanità, che consente di studiare specifiche dimensioni: la dimensione fisica, riferibile alle funzioni del movimento e della locomozione; la sfera di autonomia nelle funzioni quotidiane che si riferisce alle attività di cura della persona; la dimensione della comunicazione che riguarda le funzioni della vista, dell'udito e della parola.

Ad ogni dimensione corrisponde una batteria di quesiti con modalità di risposta che identificano diversi gradi di difficoltà (da una totale autonomia alla presenza di qualche difficoltà, ad un livello di difficoltà maggiore fino ad arrivare all'incapacità di adempiere la funzione senza l'aiuto di altre persone).

In particolare, per ciascuna area funzionale sono state individuate le principali attività e funzioni, è stato rilevato il livello di difficoltà e, in base adesso, è stato assegnato il livello di gravità delle limitazioni funzionali:

Area funzionale	Attività	Modalità Risposta		
		Assenza di limitazioni	Limitazioni lievi	Limitazioni gravi
Movimento e locomozione	<i>Distanza più lunga che può percorrere camminando da solo senza fermarsi e senza stancarsi troppo</i>	–200 metri o più	–Più di qualche passo ma meno di 200 metri	–Qualche passo soltanto –Non in grado di camminare
	<i>Scendere o salire da solo una rampa di scale senza fermarsi</i>	–Senza difficoltà –Qualche difficoltà	–Molta difficoltà	–Non in grado
	<i>Chinarsi per raccogliere una scarpa da terra</i>	–Senza difficoltà –Qualche difficoltà	–Molta difficoltà	–Non in grado
Funzioni della vita quotidiana¹	<i>Mettersi a letto e alzarsi dal letto da solo</i>	–Senza difficoltà	–Qualche difficoltà	–Solo con l'aiuto di qualcuno
	<i>Sedersi e alzarsi da una sedia da solo</i>	–Senza difficoltà	–Qualche difficoltà	–Solo con l'aiuto di qualcuno
	<i>Vestirsi e spogliarsi da solo</i>	–Senza difficoltà	–Qualche difficoltà	–Solo con l'aiuto di qualcuno
	<i>Farsi il bagno o la doccia da solo</i>	–Senza difficoltà	–Qualche difficoltà	–Solo con l'aiuto di qualcuno
	<i>Lavarsi le mani e il viso</i>	–Senza difficoltà	–Qualche difficoltà	–Solo con l'aiuto

¹ Per misurare il grado di autonomia delle persone con limitazioni funzionali nelle attività di base della vita quotidiana si utilizza una scala di valutazione proposta negli anni Sessanta. La scala ADL (*Activities of Daily Living*) misura le differenti abilità della persona nel prendersi cura di sé.

	<i>da solo</i>			di qualcuno
	<i>Mangiare da solo, anche tagliando il cibo da solo</i>	–Senza difficoltà	–Qualche difficoltà	–Solo con l'aiuto di qualcuno
Comunicazione	<i>Sentire abbastanza per seguire una trasmissione televisiva ad un volume che non disturbi gli altri considerando l'uso eventuale di apparecchi acustici</i>	–Si	– No	No
	<i>Sentire una trasmissione televisiva alzando il volume</i>	-	–Si	–No
	<i>Vedere abbastanza da riconoscere un amico a 4 metri di distanza (dall'altro lato della strada), considerando l'uso eventuale di occhiali o lenti a contatto</i>	–Si	– No	–No
	<i>Vedere abbastanza da riconoscere un amico a 1 metro di distanza (alla distanza di un braccio)?</i>	-	–Si	–No
	<i>Parlare senza difficoltà</i>	–Senza difficoltà –Qualche difficoltà	–Molta difficoltà	–Non in grado

- **Indicatore di gravità delle limitazioni nelle funzioni della vita quotidiana (ADL)**

Per valutare il livello di compromissione dell'autonomia nelle attività essenziali della vita quotidiana, è stato costruito un indicatore che sintetizza il livello di difficoltà riferito per ciascuna di esse (fare il bagno o la doccia, vestirsi e spogliarsi, sedere e alzarsi da una sedia, mettersi ed alzarsi dal letto, lavarsi le mani ed il viso, mangiare anche tagliando il cibo).

L'indicatore è articolato in tre modalità :

- 1) Nessuna o qualche difficoltà nelle attività della vita quotidiana
- 2) Non in grado di svolgere 1-2 attività della vita quotidiana
- 3) Non in grado di svolgere 3 o più attività della vita quotidiana

- **Indicatore di limitazioni nelle attività domestiche (IADL)**

Per valutare il livello di autonomia in funzioni più complesse che richiedono competenza nell'uso di strumenti, è stata utilizzata la scala delle IADL (Instrumental Activities of Daily Living)² mediante la quale sono rilevate le limitazioni in alcune delle attività domestiche che vengono svolte abitualmente: usare il telefono, fare la spesa, preparare i pasti,

² Lawton M.P. e Brody E.M., Gerontologist, 9:179-186, 1969; Source: Lawton, M.P., and Brody, E.M. "Assessment of older people: Self-maintaining and instrumental activities of daily living." Gerontologist 9:179-186, (1969).

prendere le medicine, fare lavori di casa leggeri, fare lavori di casa pesanti, occuparsi delle risorse economiche e delle attività quotidiane di gestione amministrativa. La versione, somministrata alle persone di 18 anni e più, è quella proposta per l'indagine European Health and Social Integration Survey (EHSIS).

Operativamente è stato posto un primo quesito per rilevare l'eventuale presenza e il livello di difficoltà nello svolgere ciascuna attività facendo riferimento a ciò che gli intervistati effettivamente fanno (performance) e non a ciò che sono in grado o non in grado di fare (capacità). Agli intervistati che avevano riferito delle difficoltà è stata posta una seconda domanda sulla causa di queste difficoltà per distinguere coloro che hanno difficoltà nelle attività domestiche per problemi di salute dalle persone che hanno difficoltà a svolgerle per altri motivi. Infatti le persone potrebbero avere difficoltà a svolgere alcune di queste attività perché ad esempio, all'interno della famiglia c'è una divisione dei compiti oppure perché hanno scelto di farle svolgere a un collaboratore domestico. E' stato quindi possibile individuare le persone che hanno una riduzione di autonomia in questo tipo di attività per problemi di salute o limitazioni funzionali.

Per avere una misura sintetica del livello di compromissione dell'autonomia nelle attività domestiche, è stato costruito un indicatore che è articolato in quattro modalità:

- 1) Nessuna difficoltà/difficoltà non per motivi di salute nelle attività domestiche
- 2) Qualche difficoltà per motivi di salute in almeno un'attività domestica
- 3) Molta difficoltà/non è in grado per motivi di salute in una/due attività domestiche
- 4) Molta difficoltà/non è in grado per motivi di salute in tre o più attività domestiche

• **Indicatore di restrizione alla partecipazione sociale**

L'indagine ha offerto la possibilità di utilizzare alcuni nuovi quesiti proposti in ambito internazionale per lo studio della disabilità secondo il framework concettuale della nuova classificazione ICF. In particolare sono stati selezionati alcuni quesiti tratti dall' European Health and Social Integration Survey (EHSIS)³ che ha l'obiettivo di misurare lo stato di salute e l'integrazione sociale a livello europeo.

L'inserimento di questi quesiti ha consentito di costruire un indicatore che tenga conto sia dei problemi di salute e delle limitazioni nelle attività, che delle restrizioni nella partecipazione sociale rilevata per alcune delle aree di vita previste nel questionario EHSIS: istruzione e formazione, lavoro, mobilità (uscire di casa e utilizzare il veicolo), uso di internet, tempo libero (attività fisica, partecipazione a funzioni religiose, cinema e teatro, musei, mostre e siti) e viaggi.

Per ciascuna delle aree individuate è stato costruito un indicatore a 4 livelli:

- 1) Barriere all'integrazione sociale che i rispondenti riferiscono essere connesse solo a problemi/condizioni di salute o a limitazioni funzionali (vedere, sentire, camminare, ecc.);
- 2) Barriere all'integrazione sociale che i rispondenti riferiscono essere connesse a problemi/condizioni di salute o a limitazioni funzionali (vedere, sentire, camminare, ecc.) e ad altri fattori personali o ambientali;
- 3) Barriere all'integrazione sociale che i rispondenti riferiscono essere connesse solo a fattori personali o ambientali;
- 4) Nessuna barriera alla partecipazione o mancanza di interesse a partecipare.

L'indicatore è stato poi sintetizzato in 2 livelli:

- 1) Barriere all'integrazione sociale che i rispondenti riferiscono essere connesse solo a problemi/condizioni di salute o a limitazioni funzionali (vedere, sentire, camminare, ecc.) e barriere all'integrazione sociale che i rispondenti

³ Il primo questionario dell'European Health and Social Integration Survey, denominato *European Disability Social Integration Module (EDSIM)* è stato progettato nel 2008 dall'Università di Leicester per rilevare la disabilità secondo l'approccio ICF. L'attuale versione (giugno 2011) sarà utilizzata per una indagine che sarà condotta nel 2012 in tutti i paesi dell'Unione Europea. Per evidenti ragioni di sfasatura temporale delle fasi di progettazione dell'indagine europea e di quella condotta in Italia, i quesiti inseriti nell'indagine sull'inclusione sociale delle persone con limitazioni funzionali sono stati selezionati da una versione precedente del questionario e non sono quindi perfettamente allineati con la versione diffusa da Eurostat a giugno 2011.

riferiscono essere connesse a problemi/condizioni di salute o a limitazioni funzionali (vedere, sentire, camminare, ecc.) in associazione ad altri fattori personali o ambientali;

- 2) Barriere all'integrazione sociale che i rispondenti riferiscono essere connesse solo a fattori personali o ambientali
- Nessuna barriera alla partecipazione o mancanza di interesse a partecipare.

Per la costruzione dell'indicatore di accessibilità agli edifici l'attenzione è focalizzata sulle limitazioni associate alla mancanza di assistenza o supporti per la mobilità. Anche questo indicatore è articolato in 4 livelli:

- 3) Barriere all'accessibilità per mancanza di assistenza da parte di una persona o per mancanza di supporti per la mobilità e non per altri motivi;
- 4) Barriere all'accessibilità per mancanza di assistenza da parte di una persona o per mancanza di supporti per la mobilità e per altri motivi;
- 5) Barriere all'accessibilità solo per altri motivi;
- 6) Nessuna barriera all'accessibilità.

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

Italia nord-occidentale comprende: Piemonte, Valle d'Aosta, Lombardia, Liguria.

Italia nord-orientale comprende: Trentino-Alto Adige, Veneto, Friuli-Venezia Giulia, Emilia Romagna.

Italia centrale comprende: Toscana, Umbria, Marche, Lazio.

Italia meridionale comprende: Abruzzo, Molise, Campania, Puglia, Basilicata, Calabria.

Italia insulare comprende: Sicilia, Sardegna.

Disegno di campionamento e livello di precisione delle stime

1. Il disegno campionario

L'indagine Istat "Inclusione sociale delle persone con limitazioni funzionali" è stata condotta nel 2011 su un sottocampione dell'indagine "Condizioni di salute e ricorso ai servizi sanitari" relativa al biennio 2004-2005. L'indagine ha come obiettivo l'acquisizione di informazioni sul livello di integrazione sociale delle persone con limitazioni funzionali nel contesto scolastico, lavorativo e in generale della vita sociale.

Il collettivo di riferimento dell'indagine è costituito dagli individui di età compresa tra i 6 e gli 80 anni nel 2004-2005 che, in occasione dell'indagine sulla salute, avevano dichiarato di avere difficoltà nell'espletamento di almeno una funzione tra quelle attinenti la locomozione, le abituali attività della vita quotidiana (vestirsi, mangiare, lavarsi, sedersi e alzarsi da una sedia, ecc.) e la comunicazione (vedere, sentire, parlare). Sulla base della gravità delle limitazioni riferite dagli intervistati la popolazione di interesse dell'indagine è composta da persone con limitazioni funzionali gravi⁴ e da persone con limitazioni funzionali lievi.

Dalla popolazione eleggibile dell'indagine sono stati esclusi gli individui che in occasione dei controlli preliminari, o nel momento in cui sono state effettuate le interviste nel 2011, sono risultati deceduti, istituzionalizzati, trasferiti all'estero o non più colpiti da limitazioni funzionali.

La rilevazione è stata pertanto condotta su un campione di circa 9.000 individui (2.744 persone con limitazioni funzionali gravi e 6.293 persone con limitazioni funzionali lievi) e mediante tecnica CATI (Computer Assisted Telephone Interview); le interviste telefoniche sono state somministrate direttamente alla persona con limitazioni funzionali se in grado di rispondere, oppure a familiari per i bambini di età inferiore ai 14 anni, nei casi in cui la persona da intervistare non era in grado di rispondere all'intervista (per particolari situazioni di disagio, malattia) o era assente dall'abitazione per tutto il periodo di tempo previsto per la rilevazione.

La modalità telefonica di rilevazione dell'indagine ha portato ad alcune difficoltà di tipo tecnico, infatti l'indagine è stata affetta da un numero particolarmente elevato di non risposte. Il tasso di mancata risposta è stato di circa il 59% per le persone con limitazioni funzionali gravi, e del 68% per persone con limitazioni funzionali lievi. In entrambi i casi la mancata risposta è stata determinata soprattutto dalla circostanza che gli individui sottoposti a rilevazione non sono risultati reperibili: i mancati contatti telefonici ammontano al 47% circa per le persone con limitazioni funzionali gravi e al 51% per le persone con limitazioni funzionali lievi, mentre i rifiuti veri e propri ammontano rispettivamente al 12,4% e al 16,7%.

L'ampio lasso di tempo trascorso tra le due indagini ha reso non semplice contattare gli individui a sette anni di distanza. All'assenza di informazioni relative al recapito telefonico e al rilascio di recapiti telefonici errati da parte di alcune famiglie intervistate nell'indagine Salute, si sono aggiunte altre cause che hanno reso impossibile la reperibilità, come la variazione del numero telefonico precedentemente rilasciato o addirittura la dismissione del telefono fisso da parte di molte famiglie (tendenza questa in aumento nel nostro Paese).

2. Il calcolo dei pesi e la correzione per mancata risposta

2.1. Premessa

Il principio su cui è basato ogni metodo di stima campionaria è che le unità appartenenti al campione rappresentino anche le unità della popolazione che non sono incluse nel campione; tale obiettivo viene conseguito attribuendo alle unità rispondenti un peso campionario (detto anche coefficiente di riporto all'universo), che indica il numero di unità della popolazione che l'unità stessa rappresenta. Per l'indagine sull'inclusione sociale delle persone con limitazioni funzionali,

⁴ Nell'indagine "Inclusione sociale delle persone con limitazioni funzionali" si fa riferimento ad una definizione diversa della disabilità rispetto a quanto avvenuto nelle indagini Istat sulla salute sulla base delle quali sono state fornite le stime sulla consistenza del fenomeno della disabilità. La disabilità è la risultante tra la presenza di limitazioni nelle funzioni sensoriali, motorie o nelle attività della vita quotidiana e le restrizioni alla partecipazione nelle diverse aree di vita (lavoro, scuola, relazioni sociali, tempo libero, ecc.). Sono definite "persone con limitazioni funzionali gravi" le persone che nell'indagine "Condizioni di salute e ricorso ai servizi sanitari" erano state definite "persone con disabilità".

nel calcolo dei pesi si è dovuto tenere conto del fatto che le unità campionarie costituiscono un sottoinsieme del campione dell'indagine "Condizioni di salute e ricorso ai servizi sanitari" e che pertanto già possiedono un coefficiente di riporto all'universo, coerente con gli obiettivi dell'indagine 2004-2005. Se non ci fossero state mancate risposte, questo stesso peso sarebbe stato utilizzato anche per l'indagine sull'inclusione sociale delle persone con limitazioni funzionali senza ulteriori modifiche. In presenza di mancate risposte, tuttavia, è stato necessario modificare tali pesi per far sì che le unità rispondenti all'indagine sull'inclusione sociale rappresentassero anche le unità non rispondenti.

Nel contesto dei metodi di correzione per mancata risposta, una procedura comune è dividere l'intero campione in celle di aggiustamento (si veda in proposito Särndal⁵, 1992) sulla base di variabili ausiliarie, note sia per i rispondenti sia per i non rispondenti e correlate con la probabilità di risposta, per poi attribuire un correttore per mancata risposta alle unità rispondenti utilizzando una stima del tasso di risposta nelle celle. In questo modo è possibile correggere il peso campionario iniziale delle unità rispondenti in modo che esse rappresentino anche le non rispondenti appartenenti alla stessa cella, con lo scopo di ridurre la distorsione causata dalla non risposta.

2.2. La metodologia adottata

Il trattamento della mancata risposta totale per entrambi i sottoinsiemi della popolazione è stato effettuato mediante una procedura sviluppata in due fasi che, tenendo distinte due differenti modalità di mancata risposta, determina due fattori di aggiustamento: il primo atto a correggere i pesi diretti degli individui risultati reperibili, per rappresentare gli irreperibili; il secondo per la correzione dei pesi diretti degli individui rispondenti, per rappresentare i non rispondenti (che hanno espresso un rifiuto) tra i reperibili.

Per individuare le variabili maggiormente esplicative della propensione a rispondere sono stati utilizzati modelli di tipo parametrico, come il logistico, e modelli non parametrici di classificazione, in particolare il metodo noto in letteratura come Classification and Regression Trees (CART) (Breiman, 1984⁶).

Le variabili ausiliarie individuali utilizzate nello studio - variabili rilevate nell'indagine "Condizioni di salute e ricorso ai servizi sanitari" 2004-2005 e disponibili quindi sia per i rispondenti sia per i non rispondenti - sono sia di tipo socio-demografico (sesso, età, stato civile, titolo di studio, numero di componenti della famiglia) sia relative alle difficoltà nella vita quotidiana riferite al momento dell'indagine 2005. Ulteriori variabili considerate sono la ripartizione geografica del proprio domicilio ed una variabile che identifica gli individui che in occasione dell'indagine sulla salute avevano rilasciato un recapito telefonico. Tra i modelli logistici studiati e gli alberi generati con il metodo CART sono stati scelti i migliori (in termini di adattamento i primi e di ottimalità i secondi). Per la ponderazione dei pesi campionari si è proceduto ad una doppia modellizzazione, per stimare, nella prima fase, le probabilità di reperibilità individuali e nella seconda fase le probabilità di risposta individuali.

Per la definizione delle celle di aggiustamento e la determinazione dei fattori correttivi sono stati utilizzati metodi diversi: metodo logistico predetto, metodo logistico osservato e metodo basato sull'algoritmo CART.

L'insieme dei fattori correttivi presi in considerazione sono stati dunque ottenuti come inverso delle probabilità predette di risposta o come inverso del tasso di risposta osservato nelle celle di aggiustamento definite in base ai metodi suddetti. Successivamente gli stessi correttori sono stati posti a confronto, mediante opportuni indicatori, al fine di individuare e scegliere il peso finale corretto avente le performance migliori. Gli indicatori di cui si è tenuto maggiormente conto sono l'indice 1+CV2 (Kish, 1992) che misura la variabilità dei pesi e l'indice di concordanza che misura l'adattamento dei modelli ai dati.

Sia nel caso delle persone con limitazioni funzionali gravi⁷ che delle persone con limitazioni funzionali lievi, i pesi finali aventi migliori performance sono risultati quelli i cui fattori correttivi sono stati determinati nelle celle di aggiustamento definite con il metodo CART, in entrambe le fasi di aggiustamento.

⁵ Särndal C.E., Swensson B., Wretman J. (1992), *Model assisted survey sampling*, Springer Verlag, New York.

⁶ Breiman L., Friedman J.H., Olshen R.A., Stone C.J. (1984), *Classification and Regression Trees*, Wadsworth International Group, Belmont, California.

⁷ Relativamente al campione delle persone con limitazioni funzionali gravi il fattore correttivo risultato migliore nella prima fase di correzione è stato quello calcolato sui quartili della distribuzione delle probabilità individuali.

In particolare, relativamente alle persone con limitazioni funzionali gravi, la variabile ausiliaria risultata significativa nel modello per la stima della probabilità di contatto è la presenza o meno del recapito telefonico, mentre per la stima della probabilità di risposta, è risultato essere significativa la variabile età suddivisa in classi. Per quanto riguarda invece l'insieme delle persone con limitazioni funzionali lievi, la variabile ausiliaria risultata significativa nel modello per la stima della probabilità di contatto è la presenza o meno del recapito telefonico seguita dalla variabile età suddivisa in classi per gli individui con un recapito telefonico, mentre per la stima della probabilità di risposta è risultato essere significativo l'indice di gravità delle limitazioni funzionali e, per gli individui che presentano la modalità più grave di tale indice, il titolo di studio.

Quindi, ad ogni unità rispondente appartenente ai due campioni, è stato assegnato un peso ottenuto come prodotto del peso campionario derivante dall'indagine "Condizioni di salute e ricorso ai servizi sanitari", corretto per mancato contatto e mancata risposta. Infine è stata effettuata una post-stratificazione per sesso e quattro classi di età per fare in modo che i rispondenti con limitazioni funzionali gravi e i rispondenti con limitazioni funzionali lievi riproducessero la medesima struttura per sesso e classi di età (fino ai 34 anni, tra i 35 ed i 59, tra i 60 ed i 74 ed oltre i 75) del campione completo dell'indagine "Condizioni di salute e ricorso ai servizi sanitari".

3. 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 (o coefficiente di variazione). Indicando con $\hat{Var}(\hat{Y}_d)$ la stima della varianza della generica stima \hat{Y}_d , la stima dell'errore di campionamento assoluto di \hat{Y}_d si può ottenere mediante la seguente espressione:

$$\hat{\sigma}(\hat{Y}_d) = \sqrt{\hat{Var}(\hat{Y}_d)} \quad (1)$$

la stima dell'errore di campionamento relativo di \hat{Y}_d è invece definita dall'espressione

$$\hat{\varepsilon}(\hat{Y}_d) = \frac{\hat{\sigma}(\hat{Y}_d)}{\hat{Y}_d} \quad (2)$$

Come è stato descritto nel paragrafo 2, le stime prodotte dall'indagine sono state ottenute mediante una procedura complessa basata su passaggi di correzione per mancato contatto, mancata risposta e post-stratificazione.

Poiché, quindi, lo stimatore adottato non è funzione lineare dei dati campionari e i "totali noti" utilizzati derivano da stime campionarie, per la stima della varianza $\hat{Var}(\hat{Y}_d)$ si è fatto ricorso a un metodo basato sulle replicazioni del campione. In particolare, si è utilizzata una procedura, sviluppata all'interno dell'ISTAT, basata sul metodo dei gruppi casuali (Särndal⁸, 1992; Kott, 2006⁹), che ha consentito di ottenere una valutazione della varianza campionaria che tenga conto sia della correzione per mancata risposta, sia della post-stratificazione.

Gli errori campionari ottenuti consentono di valutare il grado di precisione delle stime; inoltre, l'errore assoluto permette di costruire un intervallo di confidenza, che, con livello di fiducia P contiene il parametro oggetto di stima. L'intervallo viene espresso come

$$\{\hat{Y}_d - k_p \hat{\sigma}(\hat{Y}_d) \leq Y_d \leq \hat{Y}_d + k_p \hat{\sigma}(\hat{Y}_d)\} \quad (3)$$

Nella (3) il termine k_p dipende dal valore fissato per il livello di fiducia P.

⁸ Särndal C.E., Swensson B., Wretman J. (1992), *Model assisted survey sampling*, Springer Verlag, New York.

⁹ Kott P.S. (2006b), *Using Calibration weighting to adjust for non-response and coverage errors*, Survey Methodology, 133-142

3.1 Presentazione sintetica degli errori

Poiché a ciascuna stima corrisponde un errore campionario relativo, per consentire un uso corretto delle informazioni prodotte dall'indagine sarebbe necessario pubblicare, per ogni stima, anche il corrispondente errore di campionamento relativo. Tuttavia, non è possibile pubblicare tutti gli errori di campionamento delle stime fornite e non sarebbero comunque disponibili gli errori delle stime non pubblicate, che l'utente può ricavare in modo autonomo. Pertanto, si fa comunemente ricorso a una presentazione sintetica degli errori relativi basata su *modelli regressivi*, fondata sulla determinazione di una funzione matematica che mette in relazione ciascuna stima con il proprio errore di campionamento.

Per le stime di frequenze assolute (o relative) riferite alle modalità di variabili qualitative, è possibile utilizzare modelli che hanno un fondamento teorico, secondo cui gli errori relativi delle stime di frequenze assolute sono funzione decrescente dei valori delle stime stesse. Il modello che viene generalmente utilizzato per le stime di frequenze assolute, con riferimento al generico dominio d , è del tipo seguente:

$$\log \hat{\epsilon}^2(\hat{Y}_d) = a + b \log(\hat{Y}_d) \quad (4)$$

in cui i parametri a e b vengono stimati, separatamente per ogni dominio d , utilizzando il metodo dei minimi quadrati.

Il prospetto 1 riporta i valori dei coefficienti a e b e dell'indice di determinazione R^2 delle funzioni utilizzate per l'interpolazione degli errori campionari delle stime di frequenze per totale Italia e le sottoclassi sesso e classi di età. Sulla base delle informazioni contenute in tali prospetti, è possibile calcolare la stima dell'errore di campionamento relativo di una determinata stima \hat{Y}_d mediante la formula:

$$\hat{\epsilon}(\hat{Y}_d) = \sqrt{\exp(a + b \log(\hat{Y}_d))} \quad (5)$$

Il prospetto 2 rende più agevole la valutazione degli errori campionari e presenta la seguente struttura: in fiancata sono elencati i valori crescenti di stima (20.000, 30.000, ..., 1.500.000); le colonne successive contengono gli errori relativi di campionamento, per ciascun dominio di interesse, calcolati mediante l'espressione (5), corrispondenti alle stime tipiche riportate nella prima colonna.

Prospetto 1 – Valori dei coefficienti a , b e dell'indice di determinazione R^2 (%) per l'insieme delle persone con limitazioni gravi e delle persone con limitazioni lievi, per totale Italia, sesso e classi di età

		Persone con limitazioni gravi			Persone con limitazioni lievi		
		A	B	R^2	A	B	R^2
Italia senza sottoclassi		11,07	-1,34	93,24	9,78	-1,17	93,25
sesso	Maschi	11,69	-1,31	89,85	9,70	-1,19	93,66
	femmine	13,63	-1,44	93,33	11,06	-1,29	93,43
classi d'età	<35	11,79	-1,29	82,84	6,71	-0,87	82,16
	35-59	13,85	-1,47	85,29	5,40	-0,77	82,65
	60-74	9,57	-1,08	82,18	6,31	-0,88	87,17
	>74	10,72	-1,23	81,86	9,06	-1,13	92,79

Prospetto 2.a - Valori interpolati degli errori campionari relativi percentuali delle stime di frequenze totale Italia per sesso e classi di età - Persone con limitazioni gravi.

		Italia con sottoclassi						Italia senza sottoclassi
		sesso		classi d'età				
Stime	Maschi	Femmine	<35	35-59	60-74	>74		
20.000	52,6	72,9	61,1	70,2	57,0	48,2	32,8	
30.000	40,4	54,5	47,0	52,1	45,8	37,5	25,0	
40.000	33,4	44,3	39,1	42,2	39,2	31,4	20,6	
50.000	28,9	37,7	33,8	35,8	34,7	27,4	17,7	
60.000	25,6	33,1	30,1	31,3	31,5	24,5	15,7	
70.000	23,2	29,6	27,2	27,9	29,0	22,3	14,1	
80.000	21,2	26,9	25,0	25,3	26,9	20,5	12,9	
90.000	19,7	24,7	23,2	23,2	25,3	19,1	12,0	
100.000	18,3	22,9	21,6	21,5	23,9	17,9	11,1	
200.000	11,6	13,9	13,8	12,9	16,4	11,7	7,0	
300.000	8,9	10,4	10,7	9,6	13,2	9,1	5,3	
400.000	7,4	8,4	8,8	7,8	11,3	7,6	4,4	
500.000	6,4	7,2	7,7	6,6	10,0	6,7	3,8	
750.000	4,9	5,4	5,9	4,9	8,0	5,2	2,9	
1.000.000	-	-	-	-	-	-	2,4	
2.000.000	-	-	-	-	-	-	1,5	

Prospetto 2.b - Valori interpolati degli errori campionari relativi percentuali delle stime di frequenze totale Italia per sesso e classi di età - Persone con limitazioni lievi.

frequenze totali Italia per sesso e classi d'età - Persone con limitazioni non		Italia con sottoclassi						Italia senza sottoclassi
		sesso		classi d'età				
				<35	35-59	60-74	>74	
Stime	Maschi	Femmine	<35	35-59	60-74	>74		
20.000	35,3	42,4	38,6	32,9	30,0	34,5	40,5	
30.000	27,7	32,7	32,3	28,1	25,1	27,4	32,0	
40.000	23,3	27,1	28,5	25,2	22,1	23,3	27,0	
50.000	20,4	23,5	25,9	23,1	20,1	20,5	23,7	
60.000	18,3	20,9	23,9	21,5	18,5	18,5	21,3	
70.000	16,7	18,9	22,4	20,3	17,3	17,0	19,5	
80.000	15,5	17,3	21,1	19,3	16,3	15,7	18,0	
90.000	14,4	16,1	20,0	18,4	15,5	14,7	16,8	
100.000	13,5	15,0	19,1	17,7	14,8	13,9	15,8	
200.000	9,0	9,6	14,2	13,5	10,9	9,4	10,5	
300.000	7,0	7,4	11,9	11,6	9,1	7,5	8,3	
400.000	5,9	6,1	10,5	10,4	8,0	6,3	7,0	
500.000	5,2	5,3	9,5	9,5	7,3	5,6	6,2	
750.000	4,1	4,1	8,0	8,1	6,1	4,4	4,9	
1.000.000	-	-	-	-	-	-	4,1	
2.000.000	-	-	-	-	-	-	2,7	
3.000.000	-	-	-	-	-	-	2,2	

Le informazioni contenute in tali prospetti permettono di calcolare l'errore relativo di una generica stima (di frequenza assoluta o di un totale) mediante due procedimenti che risultano di facile applicazione, anche se conducono a risultati meno precisi di quelli ottenibili mediante l'espressione (5).

Il primo metodo consiste nell'individuare, nella prima colonna del prospetto, il livello di stima che più si avvicina alla stima di interesse e nel considerare come errore relativo il valore che si trova sulla stessa riga, nella colonna corrispondente al domino di riferimento.

Con il secondo metodo invece, l'errore campionario della stima \hat{Y}_d si ricava per interpolazione mediante la seguente espressione:

$$\hat{\varepsilon}(\hat{Y}_d) = \hat{\varepsilon}(\hat{Y}_d^{k-1}) - \frac{\hat{\varepsilon}(\hat{Y}_d^{k-1}) - \hat{\varepsilon}(\hat{Y}_d^k)}{\hat{Y}_d^k - \hat{Y}_d^{k-1}} (\hat{Y}_d - \hat{Y}_d^{k-1}) \quad (6)$$

dove \hat{Y}_d^{k-1} e \hat{Y}_d^k sono i valori delle stime, riportati nella prima colonna, entro i quali è compresa la stima di interesse \hat{Y}_d , ed $\hat{\varepsilon}(\hat{Y}_d^{k-1})$ e $\hat{\varepsilon}(\hat{Y}_d^k)$ i corrispondenti errori relativi.

Se si vuole calcolare l'errore relativo di una stima riferita a una sottopopolazione (ad esempio la popolazione di coloro che presentano una certa modalità di una variabile di interesse, come il titolo di studio), è necessario ricorrere ad un metodo di approssimazione. Infatti, la stima di una frequenza relativa (o di un qualunque indicatore) riferita al sottogruppo di unità individuate dalla caratteristica c (ad esempio le unità in possesso di diploma), è ottenibile come rapporto tra due quantità entrambe stimate:

$$\hat{R}_d = \frac{\hat{N}_d}{\hat{P}_d},$$

in cui \hat{P}_d è la stima del numero di persone che presentano la caratteristica c nel dominio d, \hat{N}_d è la stima del numero di persone che presentano una particolare modalità di interesse tra coloro che hanno la caratteristica c; \hat{R}_d è l'indicatore definito come rapporto tra \hat{N}_d e \hat{P}_d (per esempio la stima della frequenza relativa dei diplomati occupati).

Una valutazione approssimata¹⁰ dell'errore di \hat{R}_d , valida sotto l'ipotesi di incorrelazione tra \hat{R}_d e \hat{P}_d , si può ottenere come:

$$\hat{\varepsilon}(\hat{R}_d) = \sqrt{\hat{\varepsilon}^2(\hat{N}_d) - \hat{\varepsilon}^2(\hat{P}_d)},$$

in cui $\hat{\varepsilon}(\hat{N}_d)$ e $\hat{\varepsilon}(\hat{P}_d)$ si possono calcolare utilizzando la (5).

¹⁰ Si veda: P.D. Falorsi, S. Falorsi (1996) 'Indagine sulle forze di lavoro: descrizione della strategia di campionamento e valutazione dell'errore campionario dei principali indicatori provinciali del mercato del lavoro', 1996, ISTAT-Documenti).