

***Indagine multiscopo sulle famiglie
Viaggi e vacanze
Anno 2011***

Manuale per l'utente

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 condizioni economiche delle famiglie
U.O. Viaggi e vacanze
viale Oceano Pacifico, 171 – 00144
Roma
tel: 06.4673.4655
fax: 06.4673.3908
e-mail: maditorr@istat.it

INDICE

1. PREMESSA.....	3
2. L'INDAGINE TRIMESTRALE “VIAGGI E VACANZE”	3
3. I DATI CAMPIONARI E IL LORO UTILIZZO	4
APPENDICE A - IL QUESTIONARIO	7
APPENDICE B - ALCUNE DEFINIZIONI.....	8
APPENDICE C - STRATEGIA DI CAMPIONAMENTO E VALUTAZIONE DEGLI ERRORI CAMPIONARI	9

1. 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 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 informazioni che, pur essendo state oggetto di indagine, non sono risultate essere attendibili dal punto di vista campionario e quindi statisticamente non analizzabili.

2. L'INDAGINE TRIMESTRALE "VIAGGI E VACANZE"

A partire dal 1997, l'Istat conduce l'indagine trimestrale telefonica CATI¹ "Viaggi e vacanze" su un campione nazionale annuo di circa 14.000 famiglie (circa 3.500 per trimestre per un totale annuo di circa 40.000 individui). L'indagine, in conformità al Regolamento europeo 692/2011 (che ha sostituito dal 1 gennaio 2012 la Direttiva 95/57/CE), ha l'obiettivo di quantificare e analizzare i flussi turistici dei residenti in Italia, sia all'interno del Paese che all'estero, oltre che di fornire informazioni circa le modalità di effettuazione dei viaggi e le caratteristiche socio-demografiche dei turisti².

Le informazioni vengono rilevate per tutti i componenti della famiglia. L'intervista viene eseguita intervistando, nella famiglia, una persona di almeno 18 anni che fornisce notizie sia per sé che per ciascun altro componente.

I periodi di osservazione sono i trimestri gennaio-marzo, aprile-giugno, luglio-settembre e ottobre-dicembre: per ciascun periodo vengono rilevate, per tutti i componenti delle famiglie del campione, informazioni sui viaggi conclusi nel trimestre, indipendentemente dalla loro data di inizio.

Rispetto alle precedenti indagini dell'Istat sulle vacanze degli italiani, esperienza iniziata nel 1959, proseguita con cadenza pluriennale sino al 1993³ e successivamente con cadenza annuale nell'indagine "Aspetti della vita quotidiana", l'indagine trimestrale "Viaggi e vacanze" rileva, per la prima volta e con carattere di continuità, anche informazioni sulle vacanze di breve durata (1-3 notti) e sul turismo per motivi di lavoro.

L'indagine costituisce, infatti, la prima esperienza nella raccolta di informazioni trimestrali sui viaggi con almeno un pernottamento realizzati dai residenti per ragioni di lavoro o di vacanza e tale peculiarità permette di disporre sistematicamente, in aggiunta ai dati sul movimento dei clienti nelle strutture ricettive⁴, anche di informazioni sul numero e sulle caratteristiche dei viaggi e dei turisti.

Tali informazioni sono ormai ritenute necessarie a comprendere più a fondo la complessità del fenomeno turismo; l'indagine arricchisce così il quadro informativo finora esistente a livello nazionale.

¹ Computer Assisted Telephone Interview.

² La metodologia e l'organizzazione dell'indagine sono disponibili all'indirizzo www.istat.it/dati/catalogo/20060615_00/ e all'indirizzo http://www3.istat.it/dati/catalogo/20030717_01/volmetodviaggi vacanze.pdf.

³ Le indagini sulle vacanze degli italiani condotte dall'Istat fino al 1982, e in particolare con riferimento agli anni 1959, 1965, 1968, 1972, 1975, 1978 e 1982 hanno avuto per oggetto le vacanze di almeno 3 notti consecutive di durata. Per le successive indagini, la definizione di vacanza è stata estesa a periodi di almeno 4 notti consecutive. Dal 1982 al 1993 tali indagini hanno avuto per oggetto, con riferimento però al solo mese di ottobre, anche le brevi vacanze (di durata non superiore a 3 notti consecutive, ma comunque con un pernottamento fuori dal luogo di residenza abituale) ed i viaggi per altri motivi (di almeno una notte di durata), cioè effettuati per motivi diversi da quelli della vacanza (affari, lavoro temporaneo fuori sede, convegni, salute, studio, pellegrinaggi, ecc.).

⁴ I dati si riferiscono alle statistiche correnti Istat dell'attività degli esercizi ricettivi alberghieri e delle altre strutture ricettive (consistenza degli esercizi e movimento dei clienti).

I risultati dell'indagine sono pubblicati annualmente nel volume "I viaggi in Italia e all'estero" della Collana Informazioni Istat. A partire dal 2007, la pubblicazione dei volumi è sostituita dal rilascio di tavole di dati⁵.

La rilevazione utilizza le definizioni standard indicate nella metodologia adottata per l'attuazione del Regolamento europeo 692/2011 sulle Statistiche del Turismo. In particolare, i flussi turistici vengono individuati in quegli spostamenti con pernottamento effettuati fuori dal luogo in cui si vive. In tal senso, per una corretta interpretazione dei dati, assume notevole importanza il significato delle definizioni adottate:

- **viaggio**: spostamento realizzato, per turismo di vacanza o per ragioni di lavoro (purché lo scopo non sia quello di esercitare un'attività alle dipendenze di un soggetto residente nel luogo di destinazione), fuori dal luogo dove si vive e che comporta almeno un pernottamento nel luogo visitato. Sono esclusi i viaggi e gli spostamenti effettuati nelle località frequentate abitualmente tutte le settimane con soste di uno o più pernottamenti, nonché i viaggi di durata superiore ad un anno. In questi casi, infatti, il viaggio non costituisce flusso turistico poiché la località visitata viene associata al luogo dove si vive;
- **turista**: un individuo viene considerato "turista" se ha effettuato uno o più viaggi nel trimestre di riferimento dell'indagine.

I viaggi sono distinti per motivo, secondo la seguente tipologia:

- **viaggio per motivi di lavoro o professionali**: Il viaggio svolto per motivi prevalenti di lavoro quali missioni, partecipazione a congressi, riunioni d'affari o esercizio di attività di rappresentanza, docenza o altre attività professionali. Sono esclusi gli impieghi presso il luogo di destinazione (lavoro stagionale, supplenze, altri lavori temporanei);
- **viaggio di vacanza**: viaggio svolto per motivi prevalenti di relax, piacere, svago o riposo, per visite a parenti o amici, per motivi religiosi/pellegrinaggio, per cure termali o trattamenti di salute.

In relazione alla durata, le vacanze sono distinte in:

- **vacanza "breve"**: quando la durata del soggiorno è inferiore a 4 notti;
- **vacanza "lunga"**: quando la durata del soggiorno è di 4 o più notti.

Le informazioni **destinazione, mezzo di trasporto, tipo di alloggio, motivo della vacanza, tipo della vacanza di piacere/svago, motivo del viaggio di lavoro** sono rilevate sulla base del concetto della "prevalenza". In particolare, la destinazione del viaggio ed il tipo di alloggio vengono associate rispettivamente alla località ed al tipo di alloggio in cui è stato trascorso il maggior numero di notti, mentre il mezzo di trasporto viene indicato come il mezzo con cui è stata coperta la maggiore distanza.

3. I DATI CAMPIONARI E IL LORO UTILIZZO

Vengono messi a disposizione degli utenti i seguenti file di dati elementari:

- **un file degli individui** (tipo A o tipo B)
lunghezza record: 107 bytes
numero record: 32368
- **un file dei viaggi** (tipo A o tipo B)
lunghezza record: 132 bytes
numero record: 8694

Per garantire la riservatezza delle informazioni a tutela della privacy e per limiti connessi alla numerosità campionaria, non è possibile fornire contemporaneamente i dati sulle suddivisioni territoriali di residenza: regione e tipologia socio-demografica del comune. Per questo motivo, del file individui e del file viaggi sono predisposte due versioni (tipo A e tipo B):

- la prima (**tipo A**) con i codici della regione (REG) e della ripartizione geografica (RIP);
- la seconda (**tipo B**) con i codici della tipologia socio-demografica del comune (DOM) e della ripartizione geografica (RIP).

⁵ Il volume e le tavole sono consultabili sul sito Istat <http://www.istat.it>.

Le due versioni non possono essere ricongiunte in quanto ogni singola famiglia ha nei due file un identificativo familiare (IDFAM) diverso.

I file sono riferiti all'intero anno e sono ottenuti mettendo in sequenza i quattro file trimestrali; in ogni record è specificato il trimestre di riferimento (TRIM).

Il **file individui** è composto da un record per ciascun individuo componente la famiglia. Ogni record "individuo" contiene una prima parte (coll. 1-44) di informazioni generali sulla famiglia e sull'individuo (codici identificativi e caratteristiche territoriali), una seconda parte (coll. 45-64) contenente il numero di viaggi effettuati dalla famiglia e una terza parte (coll. 65-107) di informazioni sull'individuo (caratteristiche socio-demografiche e numero di viaggi effettuati dall'individuo). Ciò significa che nel file individui sono rappresentati tanto gli individui che hanno viaggiato (turisti), quanto quelli che non hanno viaggiato: quindi è possibile, con riferimento ad ogni trimestre, individuare la quota di popolazione che ha viaggiato e la quota di popolazione che non ha viaggiato.

Il **file viaggi** è composto da un record per ciascun viaggio realizzato da un individuo. Si fa presente che dal 2011, recepite le raccomandazioni internazionali IRTS (International Recommendations for Tourism Statistics) 2008, sono stati esclusi i record corrispondenti a viaggi di lavoro, il cui motivo principale riguarda l'esercizio di un'attività alle dipendenze di un soggetto residente nel luogo di destinazione (come lavori stagionali, supplenze o altri lavori a tempo determinato). Si avverte, pertanto, di tenerne conto nel caso di eventuali confronti con dati di anni precedenti, riguardanti i viaggi totali e/o di lavoro (e relative durate in notti). Ogni record "viaggio" contiene una prima parte di informazioni generali sulla famiglia e sull'individuo (che corrisponde alle coll. 1-31 del file individui) e una seconda parte (coll. 32-132) di informazioni dettagliate sul viaggio (durata, destinazione, mezzo di trasporto, alloggio, ecc.).

Nel caso in cui due o più componenti della famiglia abbiano effettuato un viaggio insieme, sono presenti tanti record "viaggio" quanti sono i partecipanti al viaggio. In ogni record "viaggio" è presente il numero dei partecipanti della famiglia allo stesso viaggio (NPART) ed una variabile (INSIEME) che permette di individuare i record "viaggio" effettuati insieme (si tratta di un numero progressivo entro la famiglia che si ripete per tutti i viaggi fatti insieme da componenti della famiglia stessa). Per distinguere i viaggi di lavoro dai viaggi di vacanza è presente una variabile (TIPOVIA) che individua la tipologia del viaggio. Nel caso di più componenti familiari partecipanti allo stesso viaggio, coincidono tutte le caratteristiche del viaggio (la data di inizio, la durata, il mezzo, la destinazione, ecc.), ad eccezione della tipologia del viaggio (TIPOVIA): è infatti ammesso che alcuni dei partecipanti abbiano viaggiato per motivi di lavoro e altri per motivi personali (vacanza, ecc.).

Per una corretta utilizzazione dei dati, si rammenta che l'indagine rileva i viaggi che si concludono in ciascun trimestre di riferimento, indipendentemente dalla data di inizio.

Per un corretto utilizzo dei file si precisa che è possibile effettuare elaborazioni sulle seguenti unità di analisi:

- a) **famiglie**: volendo analizzare le famiglie occorre selezionare solo il primo componente di ciascuna, utilizzando il numero d'ordine all'interno della famiglia (IDIND=01). Ogni famiglia è individuata da: trimestre di riferimento (TRIM), numero progressivo della famiglia all'interno del trimestre (IDFAM);
- b) **individui**: ogni individuo è identificato da: trimestre di riferimento (TRIM), numero progressivo della famiglia (IDFAM) all'interno del trimestre, numero d'ordine dell'individuo all'interno della famiglia (IDIND);
- c) **viaggi**: ogni viaggio è individuato da: trimestre di riferimento (TRIM), numero progressivo della famiglia all'interno del trimestre (IDFAM), numero progressivo del viaggio all'interno della famiglia (IDVIA);
- d) **notti**: le notti trascorse in viaggio sono espresse dalla durata del viaggio (DURATA) presente nel file viaggi.

Le informazioni riportate nei file sono di carattere campionario. In particolare, ciascun sotto-file trimestrale si riferisce ad un campione di circa 3.500 famiglie che viene selezionato in modo indipendente dai campioni delle altre tre rilevazioni.

Dato il carattere campionario dell'indagine, per ottenere stime relative all'intera popolazione di riferimento (le famiglie residenti in Italia e gli individui che le compongono, al netto dei membri permanenti delle convivenze), è necessario procedere alla ponderazione dei dati elementari, attraverso i coefficienti di riporto all'universo (contenuti, in ciascun record, nella variabile COE). I valori contenuti in questa variabile, calcolati separatamente per ciascun trimestre, rappresentano il peso che ogni unità campionaria ha nella popolazione e nel trimestre di riferimento. Il COE è espresso con due valori decimali separati dal punto (es. "04782.72"). Per l'elaborazione dei dati si consiglia di utilizzare il

COE arrotondato all'unità. I coefficienti di riporto all'universo sono stati determinati in modo da poter essere utilizzati indifferentemente per costruire stime trimestrali relative alle famiglie, agli individui, ai viaggi e alle notti.

Ad esempio, se in un record individuale del primo trimestre la variabile COE vale "04782.72", significa che l'individuo rappresenta 4.782,72 individui nella popolazione residente in Italia nel primo trimestre dell'anno. Ponderando quindi per questo valore i dati presenti nel suo record, le caratteristiche relative a tale individuo saranno stimate per 4.782,72 individui. Di conseguenza, ponderando i dati di tutti gli individui campionati in un trimestre, si otterranno stime riferite all'intera popolazione residente in Italia in quel trimestre. Analogamente, ponderando i dati contenuti nel file viaggi sarà possibile ottenere stime sui viaggi effettuati dai residenti in Italia con riferimento a ciascun trimestre di indagine.

Per quanto appena detto, è importante sottolineare che, mentre i viaggi e le notti rilevati in ciascun trimestre possono essere sommati per ottenere la stima annuale, non è corretto sommare i file trimestrali degli individui, per non quadruplicare la popolazione. Per evitare ciò, qualora si volessero comunque utilizzare dati sugli individui rilevati in tutti e quattro i trimestri, sarà necessario far riferimento ad un valore medio di popolazione, utilizzando il coefficiente COE diviso per 4. Analogamente, i dati trimestrali sul numero dei turisti (cioè il numero di individui che hanno fatto almeno un viaggio in un dato trimestre) non sono sommabili, in quanto una stessa persona può essere stata turista in trimestri diversi. Un'eventuale somma dei turisti trimestrali comporterebbe, pertanto, una sovrastima del numero annuale dei turisti.

L'indagine ha la finalità di fornire stime di parametri di diversa natura con riferimenti temporali e territoriali differenti. In particolare, il disegno campionario consente di ottenere stime trimestrali con riferimento soltanto all'intero territorio nazionale, mentre per le stime annuali sono previsti i seguenti dettagli territoriali:

1. l'intero territorio nazionale;
2. le tre grandi ripartizioni geografiche (Nord, Centro, Mezzogiorno);
3. le regioni geografiche (solamente per alcune stime più importanti);
4. le sei tipologie socio-demografiche dei comuni.

Trattandosi di un'indagine di natura campionaria, l'utilizzatore dei dati dovrà valutare l'attendibilità delle stime ottenibili, sfruttando tutte le informazioni e gli strumenti di calcolo messi a disposizione, per procedere ad una corretta interpretazione dei dati. A tale scopo, si riportano, in Appendice C, le informazioni relative alla strategia di campionamento e alla valutazione degli errori di stima, con esempi di calcolo.

APPENDICE A - IL QUESTIONARIO

Il questionario, somministrato con tecnica CATI, è strutturato in Sezioni in modo da garantire continuità di contenuto fra le domande e da facilitare la concentrazione del rispondente, al quale viene chiesto di rispondere su un certo argomento fino a che questo non sia stato del tutto esaurito e solo successivamente di passare ad un altro. All'interno del questionario sono quindi individuabili una prima fase di presentazione dell'intervista, una seconda fase in cui vengono raccolti i dati individuali ed una terza fase in cui vengono raccolti i dati sui viaggi.

La fase di presentazione rappresenta una componente cruciale dell'intervista. L'intervistatore si presenta specificando che l'indagine è effettuata per conto dell'Istat, riassume il contenuto dell'indagine e ricorda che le notizie raccolte sono coperte da segreto d'ufficio. Verifica, inoltre, che la famiglia abbia ricevuto la lettera firmata dal Presidente dell'Istat che avvertiva dell'intervista telefonica e spiegava dettagliatamente i motivi dell'indagine. Successivamente si passa a valutare l'eleggibilità della famiglia contattata. Essendo il target dell'indagine costituito dalle famiglie, si verifica in primo luogo che l'intestatario del contratto telefonico sia effettivamente una famiglia e non un'impresa, un'azienda, un ufficio o un gruppo di individui. Per evitare, inoltre, ridondanze all'interno del campione si escludono anche le dimore non abituali della famiglia, come ad esempio le seconde case.

Una volta accertata l'eleggibilità della famiglia, viene intervistata una persona, di almeno 18 anni (esclusi domestici, amici o visitatori casuali), che fornirà informazioni per sé e per ciascuno degli altri componenti della famiglia. Quindi, per ogni famiglia del campione, viene realizzata l'intervista "diretta" per un solo componente (persona di riferimento), mentre per ciascun altro componente l'intervista è di tipo "proxy".

Il questionario si compone delle seguenti sezioni:

- "Sezione generale", dove vengono rilevate le informazioni strutturali di ciascun componente della famiglia;
- "Scheda riepilogativa viaggi", dove viene riportato il numero dei viaggi conclusi da ciascun componente nel trimestre di riferimento;
- "Scheda viaggio di lavoro", dove vengono rilevate le informazioni per ciascun viaggio effettuato dai componenti della famiglia per motivi professionali;
- "Scheda viaggio di vacanza", dove vengono rilevate le informazioni per ciascun viaggio effettuato dai componenti della famiglia per motivi di vacanza/personali.

Di seguito è riportato l'elenco dei quesiti posti all'intervistato in ciascuna sezione.

APPENDICE B - ALCUNE DEFINIZIONI

Le **ripartizioni geografiche** sono una suddivisione amministrativa del territorio italiano e sono così definite:

- **Nord**:: Piemonte, Valle d'Aosta, Lombardia, Liguria, Trentino-Alto Adige, Veneto, Friuli-Venezia Giulia, Emilia-Romagna
- **Centro**: Toscana, Umbria, Marche, Lazio
- **Mezzogiorno**: Abruzzo, Molise, Campania, Puglia, Basilicata, Calabria, Sicilia, Sardegna.

La **tipologia socio-demografica del comune (dominio)** è una suddivisione in classi dei comuni italiani, in base all'ampiezza demografica e alle caratteristiche socio-economiche:

- **comuni centro dell'area metropolitana**: Torino, Milano, Venezia, Genova, Bologna, Firenze, Roma, Napoli, Bari, Palermo, Catania, Cagliari;
- **comuni che gravitano intorno al centro dell'area metropolitana** come definiti sulla base dei risultati del Censimento della Popolazione;
- **altri comuni** suddivisi per dimensione demografica (oltre i 50.000 abitanti, da 10.001 a 50.000 abitanti, da 2.001 a 10.000 abitanti, fino a 2.000 abitanti).

Le caratteristiche socio-demografiche dell'individuo (stato civile, titolo di studio, condizione professionale, posizione nella professione, attività economica) fanno riferimento alla situazione al momento dell'intervista e sono quelle standard dell'Istat. In particolare:

- **l'età** è espressa in anni compiuti;
- **la condizione professionale** è quella dichiarata come unica o prevalente;
- **occupato** è chi svolge un'occupazione in proprio o alle dipendenze da cui trae un profitto o una retribuzione (utile, onorario, salario, stipendio, ecc.) o collabora con un familiare che svolge attività lavorativa in conto proprio, senza avere un regolare contratto di lavoro (coadiuvante);
- **in cerca di nuova occupazione** è chi avendo perduto una precedente occupazione è alla ricerca attiva di una nuova occupazione ed è in grado di accettarla se gli viene offerta;
- **in cerca di prima occupazione** è chi avendo concluso o sospeso o abbandonato un ciclo di studi e chi non avendo mai esercitato un'attività lavorativa è alla ricerca attiva di un'occupazione ed è in grado di accettarla se gli viene offerta;
- **casalinga** è chi si dedica prevalentemente alle cure della propria famiglia o della propria casa;
- **studente** è chi si dedica prevalentemente allo studio;
- **ritirato dal lavoro** è chi ha cessato un'attività lavorativa per raggiunti limiti di età, invalidità o altra causa; la figura di ritirato dal lavoro non coincide necessariamente con quella di pensionato in quanto non sempre il ritirato dal lavoro gode di una pensione;
- **in altra condizione** è chi si trova in condizione diversa da quelle sopra elencate, e dalla condizione di inabile al lavoro o militare in servizio di leva/servizio civile (ad es. benestante, pensionato per motivi diversi dall'attività lavorativa, detenuto agli arresti domiciliari, ecc.).

APPENDICE C - STRATEGIA DI CAMPIONAMENTO E VALUTAZIONE DEGLI ERRORI CAMPIONARI

1. Introduzione

La popolazione di interesse dell'indagine – ossia l'insieme delle unità statistiche oggetto di investigazione – è costituita dalle famiglie residenti in Italia e dagli individui che le compongono, al netto dei membri permanenti delle convivenze. Per famiglia si intende la famiglia di fatto, ossia un insieme di persone coabitanti e legate da vincoli di matrimonio, parentela, affinità, adozione, tutela o affettivi.

L'indagine viene svolta mediante intervista telefonica e utilizza come lista di selezione l'archivio ufficiale nazionale degli abbonati alle utenze private di telefonia fissa; le unità di campionamento sono, pertanto, i numeri telefonici appartenenti a detto archivio.

Nel corso di un anno sono previste quattro rilevazioni, nei mesi di aprile, luglio, ottobre e gennaio. Ciascuna rilevazione ha come periodo di riferimento il trimestre immediatamente precedente. Ad esempio, la rilevazione di aprile si riferisce al primo trimestre dell'anno.

Ciascuna delle quattro rilevazioni trimestrali utilizza un campione di circa 3.500 famiglie che viene selezionato in modo indipendente dai campioni delle altre tre rilevazioni.

L'indagine ha la finalità di fornire stime di parametri di diversa natura (totali, medie, rapporti, frequenze assolute e relative) con diversi riferimenti temporali e territoriali. Le stime relative ai singoli trimestri si riferiscono all'intero territorio nazionale, mentre i domini territoriali di riferimento per le stime relative all'intero anno sono:

- l'intero territorio nazionale;
- le tre grandi ripartizioni geografiche (Nord, Centro e Mezzogiorno);
- le regioni geografiche (solamente per alcune stime più importanti);
- la tipologia comunale, ottenuta suddividendo i comuni italiani in sei classi formate in base a caratteristiche socio-economiche e demografiche (sulla base della popolazione residente per l'anno di riferimento):

A, area metropolitana suddivisa in :

- A1, comuni centro dell'area metropolitana: Torino, Milano, Venezia, Genova, Bologna, Firenze, Roma, Napoli, Bari, Palermo, Catania e Cagliari;
- A2, comuni che gravitano intorno al centro dell'area metropolitana;

B, area non metropolitana suddivisa in:

- B1, comuni aventi fino a 2.000 abitanti;
- B2, comuni con 2.001-10.000 abitanti;
- B3, comuni con 10.001-50.000 abitanti;
- B4, comuni con oltre 50.000 abitanti.

2. Caratteristiche della lista di selezione

La base di campionamento adottata, ossia la lista di selezione delle unità campionarie, è l'archivio informatizzato ufficiale degli abbonati alle utenze private di telefonia fissa. Tale scelta è motivata dal fatto che le informazioni dell'archivio in oggetto sono contenute in un file che viene costantemente aggiornato sulle variazioni degli intestatari e degli indirizzi telefonici; esso è, inoltre, di agevole utilizzo per la selezione delle unità campionarie in quanto si presta facilmente alla scelta di diversi criteri di ordinamento.

Le informazioni relative a ciascun indirizzo, utilizzabili per la stratificazione delle unità della popolazione di riferimento, sono essenzialmente di tipo territoriale; esse sono: la provincia, il comune, la sezione di censimento, la via, il numero civico e infine l'ampiezza del comune di appartenenza definita sia in termini demografici sia in termini di numero di indirizzi.

I principali problemi che derivano dall'utilizzo della lista telefonica come base di campionamento per le indagini sulle famiglie, sono legati alla non perfetta coincidenza tra la popolazione oggetto d'indagine e l'insieme delle unità contenute nella lista; indicando con L ed U rispettivamente la lista e la popolazione d'interesse, le possibili situazioni in cui gli insiemi L e U differiscono sono in generale le seguenti:

- I. sovracopertura, quando alcuni elementi dell'insieme L non appartengono all'insieme U ;
- II. sottocopertura, nel caso in cui alcuni elementi di U non sono contenuti in L ;
- III. duplicazione di alcune unità, se alcuni elementi di U sono presenti più volte in L ;
- IV. grappoli di unità, quando alcuni elementi dell'insieme L contengono grappoli di elementi dell'insieme U .

La lista telefonica utilizzata per l'indagine presenta in misura diversa tutte le situazioni elencate; i fenomeni che appaiono più rilevanti sono, tuttavia, quelli della sovracopertura e della sottocopertura. Rientrano, infatti, nel caso della sovracopertura tutti i numeri telefonici che corrispondono a seconde case oppure ad attività professionali o a pubblici esercizi, ecc.; mentre rientrano nel caso della sottocopertura tutte le famiglie non abbonate al telefono o intestatarie di numeri riservati. Rientrano, poi, nel caso III le famiglie intestatarie di più di un numero riferito all'abitazione principale e nel caso IV i numeri telefonici a cui corrispondono più famiglie.

L'effetto di ciascuna delle situazioni appena elencate è quello di modificare in modo non controllato le probabilità di inclusione delle unità della popolazione rispetto alle probabilità di inclusione teoriche assegnate alle stesse unità dal disegno di campionamento adottato. Ciò può essere causa di distorsioni nelle stime prodotte dall'indagine.

In particolare la sottocopertura determina una distorsione delle stime prodotte tanto maggiore quanto più, relativamente alle variabili d'interesse dell'indagine, le unità di U che appartengono a L sono differenti dalle unità di U non presenti in L . È possibile in parte attenuare tale effetto distorsivo ricorrendo a opportune procedure di stima, note in letteratura come stimatori di ponderazione vincolata⁶, che tengono conto di totali noti (desunti da fonti esterne all'indagine) della popolazione relativi a variabili ausiliarie correlate a quelle d'interesse.

Un effetto dovuto al fenomeno della sovracopertura, è il fatto che la numerosità realizzata del campione sia minore di quella definita nella fase di progettazione del disegno e ciò comporta un aumento di variabilità delle stime. Si tratta allora di decidere se accettare una dimensione campionaria inferiore a quella prevista o, alternativamente, ricorrere a un sovra-dimensionamento del campione o a una lista di unità sostitutive. Nelle indagini di tipo telefonico, in cui la rilevazione è affidata ad una ditta esterna si ricorre in genere alla sostituzione in quanto è necessario garantire che tale ditta svolga un numero prefissato di interviste utili. Al fine di evidenziare il fenomeno della sostituzione, nei prospetti 1 e 2 vengono presentati i tassi di sostituzione, rispettivamente per regione e per tipologia comunale, osservati nel trimestre di indagine ottobre-dicembre.

3 Disegno di campionamento

3.1 Descrizione generale

Il disegno di campionamento è un disegno stratificato a grappoli, in cui ciascun grappolo è un indirizzo telefonico presente nell'archivio di selezione. Tutti gli individui appartenenti ai grappoli campione, cioè appartenenti alle famiglie a cui corrispondono i numeri selezionati, vengono inclusi nel campione.

Gli indirizzi telefonici sono stati stratificati, all'interno di ogni regione geografica, secondo la tipologia socio-demografica dei comuni (descritta nel paragrafo 1) di appartenenza degli indirizzi. Poiché tale informazione non è presente nell'archivio di selezione, è stato necessario integrare l'archivio attribuendo ad ogni record il relativo codice di tipologia.

La determinazione del numero totale di unità campionarie e la sua allocazione tra gli strati, è in genere, per un'indagine ad obiettivi plurimi come quella in esame, un'operazione complessa. È poco realistico, infatti, pensare di poter definire un campione che assicuri prefissati livelli di precisione per tutte le stime d'interesse, considerando anche il fatto che le stime vengono prodotte con diversi riferimenti temporali e territoriali. L'allocazione ottimale delle unità del campione con riferimento ad un dato tipo di dominio può risultare contrastante con l'allocazione ottimale con riferimento ad un altro tipo di dominio territoriale. Infatti, per quanto riguarda le stime riferite all'intero territorio nazionale l'allocazione ottimale risulta vicina a quella proporzionale tra le diverse regioni; per quanto riguarda, invece, le stime riferite alle regioni, l'allocazione ottimale risulta prossima a quella che assegna a tutte

⁶ Noti nella letteratura in lingua anglosassone sul tema come Deville J.C. e Särndal C.E. "Calibration Estimators in Survey Sampling", *Journal of the American Statistical Association*, 87 (1992): 1013-1020.

le regioni un campione di uguale numerosità. È necessario quindi ricorrere a un procedimento complesso articolato in più fasi.

Dapprima, mediando tra esigenze operative e di costo ed esigenze relative all'attendibilità delle principali stime di interesse, si è definito un numero complessivo di indirizzi campione n pari a circa 3.500. Successivamente, sulla base di valutazioni dell'errore di campionamento atteso delle principali stime a livello regionale e nazionale, è stata determinata l'allocazione del campione tra le regioni; si è ottenuta in tal modo un'allocazione di compromesso tra l'allocazione uniforme e quella proporzionale al peso demografico di ciascuna regione. Infine, le numerosità campionarie regionali sono state ripartite tra le diverse tipologie di comune in modo proporzionale alla popolazione residente.

Nei prospetti 1 e 2 sono riportati, rispettivamente per regione e per tipologia di comune, le numerosità campionarie teoriche e i tassi di sostituzione riferiti alla rilevazione del quarto trimestre d'indagine; nell'ambito di ciascuna regione il tasso di sostituzione è calcolato come rapporto percentuale tra il numero di indirizzi sostituiti ed il numero teorico di indirizzi campione.

Prospetto 1 - Distribuzione regionale del campione e tassi di sostituzione - Quarto trimestre 2011

REGIONI	Indirizzi campione	Tasso di sostituzione totale (a)	Tasso di sostituzione al netto dei mancati contatti (b)
Piemonte	266	22,2	10,2
Valle d'Aosta	92	22,8	13,0
Lombardia	330	15,2	6,1
Trentino-Alto Adige	207	25,6	7,7
<i>Bolzano-Bozen</i>	100	27,0	9,0
<i>Trento</i>	107	24,3	6,5
Veneto	202	13,9	7,4
Friuli-Venezia Giulia	139	16,5	6,5
Liguria	152	27,0	10,5
Emilia-Romagna	198	22,7	12,6
Toscana	213	16,9	8,0
Umbria	104	23,1	10,6
Marche	141	20,6	6,4
Lazio	219	28,3	17,8
Abruzzo	143	25,9	13,3
Molise	96	22,9	17,7
Campania	197	22,3	17,8
Puglia	187	19,8	11,8
Basilicata	99	20,2	11,1
Calabria	158	17,7	13,9
Sicilia	212	26,4	13,2
Sardegna	151	29,1	15,2
Italia	3506	21,6	11,2

(a) Il tasso di sostituzione totale è calcolato come rapporto percentuale tra il numero di famiglie sostituite ed il numero teorico di famiglie campione. Esso esclude le sostituzioni rese necessarie dagli 'errori' di lista, molto elevati in alcune zone soprattutto per la presenza di numerici telefonici riferiti ad abitazioni non principali (seconde case).

(b) Il tasso di sostituzione al netto dei mancati contatti è calcolato come rapporto percentuale tra il numero di famiglie sostituite, depurate dalle sostituzioni per mancato contatto, ed il numero teorico di famiglie campione. Per mancato contatto con la famiglia ci si riferisce ai casi di 7 tentativi di contatto telefonico con esito 'libero', effettuati in giorni e orari diversi.

Prospetto 2 - Distribuzione del campione per tipologia di comune e tassi di sostituzione - Quarto trimestre 2011

TIPOLOGIA DEL COMUNE	Indirizzi campione	Tasso di sostituzione totale	Tasso di sostituzione al netto dei mancati contatti
Comuni centro dell'area metropolitana	413	23,2	9,0
Comuni che gravitano intorno al centro dell'area metropolitana	323	24,8	17,3
Comuni aventi fino a 2.000 abitanti	287	25,4	12,5
Comuni con 2.001-10.000 abitanti	919	22,0	11,0
Comuni con 10.001-50.000 abitanti	926	19,4	10,2
Comuni con oltre 50.000 abitanti	638	20,1	10,8
Italia	3506	21,6	11,2

3.2. Selezione del campione

L'estrazione degli indirizzi campione da ciascuno strato, ottenuto come incrocio della regione e della tipologia comunale, è stata effettuata con probabilità uguali e senza reimmissione mediante tecnica di selezione sistematica. La selezione del campione di indirizzi avviene nel seguente modo:

- per ciascuno strato si ordinano gli indirizzi dell'Archivio Telefonico delle Famiglie per ampiezza demografica del comune, strada, numero civico e piano;
- si definisce il passo di estrazione p come rapporto tra il numero di indirizzi in archivio e il numero di indirizzi da estrarre; si seleziona un numero casuale compreso tra 1 e p e si procede, poi, alla selezione degli indirizzi campione sulla base del passo di estrazione a partire dal punto di partenza casuale;
- ogni indirizzo selezionato dall'Archivio entra a far parte dell'elenco base e per ognuno di essi si selezionano i tre indirizzi successivi, che entrano a far parte dell'elenco sostitutivo; in tal modo le unità sostitutive sono le più vicine possibile dal punto di vista territoriale alle rispettive unità del campione base e quindi, presumibilmente, sono anche simili dal punto di vista delle caratteristiche socio-economiche.

4 Procedimento per il calcolo delle stime

Le stime prodotte dall'indagine sono stime del numero di viaggi e del numero di notti trascorse fuori casa per tipologia di viaggio, nonché del numero e delle caratteristiche di individui che effettuano tali viaggi.

Tali stime sono ottenute mediante uno stimatore di ponderazione vincolata (cfr. par. 2), che è il metodo di stima standard per la maggior parte delle indagini Istat sulle imprese e sulle famiglie.

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.

Questo principio viene realizzato attribuendo a ogni unità campionaria un peso che indica il numero di unità della popolazione rappresentate dall'unità medesima. Se, ad esempio, a un'unità campionaria viene attribuito un peso pari a 100, vuol dire che questa unità rappresenta se stessa ed altre 99 unità della popolazione che non sono state incluse nel campione.

Al fine di rendere più chiara la successiva esposizione, introduciamo la seguente simbologia: d , indice di livello territoriale di riferimento delle stime; t , indice di trimestre ($t=1, \dots, 4$); j , indice di famiglia; p , indice del componente della famiglia; h , indice di strato; y , generica variabile oggetto di indagine; Y_{thjp} valore di y osservato sul componente p della famiglia j dello strato h rilevata nel trimestre t ; P_{hj} , numero di componenti della famiglia j dello strato h ; M_h , numero di famiglie residenti nello strato h ; m_h , campione di famiglie nello strato h ; H_d , numero di strati nel dominio d .

Ipotizziamo di voler stimare, con riferimento a un generico dominio d (ad esempio una regione geografica) e relativamente a un generico trimestre t , il totale della variabile y , espresso dalla seguente relazione:

$${}_dY_t = \sum_{h=1}^{H_d} \sum_{j=1}^{M_h} \sum_{p=1}^{P_{hj}} Y_{thjp} \quad (1)$$

Una stima del totale (1) è data dalla seguente espressione:

$${}_d\hat{Y}_t = \sum_{h=1}^{H_d} \hat{Y}_h = \sum_{h=1}^{H_d} \sum_{j=1}^{m_h} \sum_{p=1}^{P_{hj}} Y_{thjp} W_{thj}, \quad (2)$$

in cui W_{thj} è il peso finale da attribuire alla famiglia j dello strato h e a tutti i componenti a essa appartenenti.

Dalla precedente relazione si desume, quindi, che per ottenere la stima del totale (1) occorre moltiplicare il peso finale associato a ciascuna unità campionaria per il valore della variabile y assunto da tale unità ed effettuare, a livello del dominio di interesse, la somma dei prodotti così ottenuti.

Il peso da attribuire alle unità campionarie è ottenuto per mezzo di una procedura complessa che ha le seguenti finalità:

- correggere l'effetto distorsivo dovuto agli errori di lista e al fenomeno della mancata risposta totale;

- tenere conto della conoscenza di alcuni totali noti sulla popolazione oggetto di studio, nel senso che le stime campionarie di tali totali devono coincidere con i rispettivi valori noti. Nell'indagine in oggetto i totali noti⁷ sono:
 - popolazione residente per le 6 tipologie comunali;
 - popolazione residente per sesso nelle 20 regioni italiane;
 - popolazione residente per 8 classi di età⁸;
 - famiglie unipersonali di età inferiore a 65 anni, per sesso;
 - famiglie unipersonali di 65 anni e più, per sesso;
 - famiglie con più di un componente per classi di ampiezza⁹

Indicando, quindi, con ${}_kX$ il totale noto della k -ma variabile ($k=1\dots K$) ausiliaria e con ${}_kX_{thjp}$ il valore assunto dalla k -ma variabile ausiliaria per l'individuo rispondente $thjp$, la condizione sopra descritta è espressa dalla seguente uguaglianza:

$${}_kX_t = \hat{X}_t = \sum_{h=1}^H \sum_{j=1}^{m_h} \sum_{p=1}^{P_{hj}} {}_kX_{thjp} W_{thj} \quad k=1\dots K.$$

La procedura per la costruzione dei pesi finali da attribuire alle unità campionarie, è articolata nelle seguenti fasi :

- 1) viene dapprima calcolato il peso base (o peso diretto), ottenuto come reciproco della probabilità di inclusione di ogni unità campionaria;
- 2) si calcola quindi il fattore correttivo che consente di soddisfare la condizione di uguaglianza tra i totali noti della popolazione e le corrispondenti stime campionarie;
- 3) il peso finale è dato dal prodotto del peso base per i fattori correttivi sopra indicati.

Il fattore correttivo del punto 2) è ottenuto mediante la risoluzione di un problema di minimo vincolato, in cui la funzione da minimizzare è la distanza tra i pesi base ed i pesi finali e i vincoli sono definiti dalla condizione che le stime campionarie dei totali di popolazione sopra definiti coincidano con i valori noti degli stessi.

Con riferimento all'intero anno, i parametri d'interesse possono essere sia dei totali annui, ${}_dY$, che delle medie annue di dati trimestrali ${}_d\bar{Y}$. I totali annui - ottenuti come somma dei totali (1), relativi ai quattro trimestri - si riferiscono al totale annuo dei viaggi ed al totale annuo delle notti trascorse fuori casa per viaggi. Le medie annue di dati trimestrali - ottenute come media aritmetica dei totali (1) - si riferiscono, invece, al numero medio di persone che viaggiano nel trimestre. Con riferimento al generico dominio d , le stime dei suddetti parametri, ${}_dY$ e ${}_d\bar{Y}$, si ottengono rispettivamente mediante le seguenti espressioni:

$${}_d\hat{Y} = \sum_{t=1}^4 {}_d\hat{Y}_t, \quad (3)$$

$${}_d\hat{\bar{Y}} = \frac{1}{4} \sum_{t=1}^4 {}_d\hat{Y}_t \quad (4)$$

Alcuni parametri d'interesse, infine, sono dati dal rapporto tra totali; in tal caso la stima può essere ottenuta come rapporto tra i totali stimati:

$${}_d\hat{R} = \frac{{}_d\hat{Y}_1}{{}_d\hat{Y}_2}.$$

⁷ I primi tre totali sono desunti da fonti demografiche (anagrafiche), mentre i totali riferiti alle famiglie derivano da stime dell'indagine annuale multiscope "Aspetti della vita quotidiana".

⁸ Le classi di età considerate sono: 0-5 anni, 6-14 anni, 15-24 anni, 25-34 anni, 35-44 anni, 45-54 anni, 55-64 anni, 65 anni e più.

⁹ Le classi di ampiezza considerate sono: 2 componenti e 3 componenti e più.

5 Valutazione del livello di precisione delle stime

Le principali statistiche di interesse per valutare la variabilità campionaria delle stime prodotte dall'indagine sono l'errore di campionamento assoluto e l'errore di campionamento relativo.

Indicando con $\hat{Var}({}_d\hat{Y}_t)$ la stima della varianza della stima ${}_d\hat{Y}_t$, riferita al dominio d e al trimestre di indagine t , la stima dell'errore di campionamento assoluto di ${}_d\hat{Y}_t$ si può ottenere mediante la seguente espressione:

$$\hat{\sigma}({}_d\hat{Y}_t) = \sqrt{\hat{Var}({}_d\hat{Y}_t)} \quad (5)$$

La stima dell'errore di campionamento relativo di ${}_d\hat{Y}_t$, è invece definita dall'espressione:

$$\hat{\varepsilon}({}_d\hat{Y}_t) = \frac{\sqrt{\hat{Var}({}_d\hat{Y}_t)}}{{}_d\hat{Y}_t} \quad (6)$$

La stima della varianza, $\hat{Var}({}_d\hat{Y}_t)$, viene ottenuta come somma della stima della varianza dei singoli strati appartenenti al dominio d ; in simboli:

$$\hat{Var}({}_d\hat{Y}_t) = \sum_{h=1}^{H_d} \hat{Var}(\hat{Y}_{th}) = \sum_{h=1}^{H_d} M_h^2 \frac{M_h - m_h}{m_h M_h} \sum_{j=1}^{m_h} \frac{(e_{thj} - \bar{e}_{th})^2}{m_h - 1} \quad (7)$$

dove

$$e_{thj} = \sum_{p=1}^{P_{hj}} Y_{thjp} - \hat{\beta} \sum_{p=1}^{P_{hj}} X_{thjp} \quad \bar{e}_{th} = \frac{1}{m_h} \sum_{j=1}^{m_h} e_{thj};$$

in cui X_{thjp} è il vettore delle K variabili ausiliarie utilizzate per la costruzione dello stimatore di ponderazione vincolata e $\hat{\beta}$ è il vettore dei coefficienti di regressione stimati del modello lineare che lega la variabile Y alle variabili ausiliarie X .

Passiamo ora alla descrizione della metodologia per il calcolo degli errori di campionamento delle stime dei parametri riferiti all'intero anno. Per quanto riguarda la stima di un totale annuo (3), l'errore di campionamento assoluto e relativo, tenendo presente l'indipendenza dei quattro campioni, sono espressi rispettivamente dalle seguenti formule

$$\hat{\sigma}({}_d\hat{Y}) = \sqrt{\sum_{t=1}^4 \hat{Var}({}_d\hat{Y}_t)} \quad \hat{\varepsilon}({}_d\hat{Y}) = \frac{\hat{\sigma}({}_d\hat{Y})}{{}_d\hat{Y}} \quad (8)$$

Per quanto riguarda, invece, la stima di una media annua (4), l'errore di campionamento assoluto e relativo sono dati rispettivamente dalle seguenti espressioni:

$$\hat{\sigma}({}_d\hat{\bar{Y}}) = \frac{1}{4} \sqrt{\sum_{t=1}^4 \hat{Var}({}_d\hat{Y}_t)} \quad \hat{\varepsilon}({}_d\hat{\bar{Y}}) = \frac{\hat{\sigma}({}_d\hat{\bar{Y}})}{{}_d\hat{\bar{Y}}} \quad (9)$$

Infine, la stima degli errori relativi di stime di rapporto tra totali, nell'ipotesi semplificatrice che il rapporto sia indipendente dal valore del denominatore, può essere ottenuta come

$$\hat{\varepsilon}({}_d\hat{R}) = \sqrt{\hat{\varepsilon}^2({}_d\hat{Y}_1) - \hat{\varepsilon}^2({}_d\hat{Y}_2)} \quad (10)$$

Gli errori campionari consentono di valutare il grado di precisione delle stime; inoltre, l'errore assoluto permette di costruire l'intervallo di confidenza che con una certa fiducia P contiene il parametro di interesse. Con riferimento ad una generica stima \hat{Y} tale intervallo assume la seguente forma:

$$\{\hat{Y} - k\hat{\sigma}(\hat{Y}), \hat{Y} + k\hat{\sigma}(\hat{Y})\} \quad (11)$$

Nella (11) il valore di k dipende dal valore fissato per P ; ad esempio, per $P=0,95$ si ha $k=2$.

6 Presentazione sintetica degli errori campionari

Poiché a ciascuna stima ${}_d\hat{Y}_t$ corrisponde un errore campionario relativo $\hat{\varepsilon}({}_d\hat{Y}_t)$, per consentire un uso corretto delle informazioni prodotte dall'indagine sarebbe necessario pubblicare per ogni stima anche il corrispondente errore di campionamento relativo. Tuttavia sia per limiti di tempo e di costi di elaborazione, sia perché le tavole di pubblicazione risulterebbero appesantite e di non facile consultazione per l'utente finale, non è possibile pubblicare anche tutti gli errori di campionamento delle stime fornite. Inoltre, non sarebbero comunque disponibili gli errori delle stime non pubblicate, che l'utente può ricavare in modo autonomo.

Al fine di permettere comunque una valutazione della variabilità campionaria di tutte le stime d'interesse, si ricorre a una presentazione sintetica degli errori relativi basata su modelli regressivi; ossia fondata sulla determinazione di una funzione matematica che mette in relazione ciascuna stima con il proprio errore di campionamento. L'approccio utilizzato per la costruzione dei modelli è differente a seconda che la variabile oggetto di stima sia qualitativa ovvero quantitativa. Infatti, mentre 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, per le stime di totali di variabili quantitative il problema è piuttosto complesso, dal momento che non è stata ancora elaborata un'adeguata base teorica per l'interpolazione degli errori campionari delle stime in questione. L'approccio adottato per trattare il caso di variabili quantitative è pertanto di tipo empirico ed è fondato sull'evidenza sperimentale che l'errore assoluto di un totale è una funzione crescente del totale stesso.

Vengono di seguito descritti i modelli adottati per i due differenti gruppi di variabili.

6.1. Presentazione sintetica degli errori campionari per stime di frequenze riferite agli individui

Il modello utilizzato per le stime di frequenze assolute, con riferimento al generico dominio d e al trimestre t , è del tipo seguente:

$$\log \hat{\varepsilon}^2({}_d\hat{Y}_t) = a + b \log({}_d\hat{Y}_t) \quad (12)$$

in cui i parametri a e b vengono stimati utilizzando il metodo dei minimi quadrati. Dall'analisi dei coefficienti di regressione ottenuti è emerso che i coefficienti relativi al primo, secondo e quarto trimestre sono sostanzialmente simili tra loro, mentre differenti risultati si ottengono per il terzo trimestre; pertanto, per semplicità, vengono presentati solamente i risultati relativi al primo e al terzo trimestre. È stato inoltre stimato un modello analogo al (12) con riferimento alla media dei 4 trimestri.

I prospetti 3 (per i trimestri) e 4 (per la media dei 4 trimestri) riportano 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 riferite agli individui, per area territoriale.

Inoltre, allo scopo di facilitare il calcolo degli errori campionari, nei prospetti 5 (per i trimestri) e 6 (per la media dei 4 trimestri) sono riportati, per area territoriale, i valori interpolati degli errori campionari relativi percentuali di alcuni valori tipici assunti dalle stime di frequenze assolute.

La prima colonna dei prospetti riporta K valori crescenti delle stime ${}_d\hat{Y}_t^k$ ($k=1,\dots,K$), la seconda riporta i rispettivi errori relativi interpolati $\hat{\varepsilon}({}_d\hat{Y}_t^k)$. Le informazioni contenute nei prospetti permettono di calcolare l'errore relativo di una generica stima di frequenza assoluta mediante due procedimenti che risultano di facile applicazione, anche se conducono a risultati meno precisi di quelli ottenibili mediante l'applicazione diretta dell'espressione (12).

Il metodo consiste nell'approssimare l'errore relativo $\hat{\varepsilon}({}_d\hat{Y}_t)$ della stima di interesse ${}_d\hat{Y}_t$ con l'errore relativo corrispondente al livello stima, presente nel prospetto (5 o 6) che più si avvicina al valore della stima ${}_d\hat{Y}_t$.

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

$$\hat{\varepsilon}({}_d\hat{Y}_t) = \hat{\varepsilon}({}_d\hat{Y}_t^{k-1}) + \frac{\hat{\varepsilon}({}_d\hat{Y}_t^k) - \hat{\varepsilon}({}_d\hat{Y}_t^{k-1})}{{}_d\hat{Y}_t^k - {}_d\hat{Y}_t^{k-1}} ({}_d\hat{Y}_t - {}_d\hat{Y}_t^{k-1}) \quad (13)$$

dove: ${}_d\hat{Y}_t^{k-1}$ e ${}_d\hat{Y}_t^k$ sono i valori delle stime entro i quali è compresa la stima d'interesse ${}_d\hat{Y}_t$ e $\hat{\varepsilon}({}_d\hat{Y}_t^{k-1})$ e $\hat{\varepsilon}({}_d\hat{Y}_t^k)$ sono i corrispondenti errori relativi presenti nel prospetto.

6.2. Presentazione sintetica degli errori campionari per stime di totali riferiti ai viaggi e alle notti

Il modello utilizzato per le stime del totale del numero di viaggi e di notti, con riferimento al generico dominio d e al trimestre t, è del tipo seguente:

$$\hat{\sigma}({}_d\hat{Y}_t) = a + b {}_d\hat{Y}_t + c {}_d\hat{Y}_t^2 \quad (14)$$

dove i parametri a, b e c vengono stimati utilizzando il metodo dei minimi quadrati, adattando il modello (14) a una nuvola di punti costituita da un consistente numero di coppie $(\hat{\sigma}({}_d\hat{Y}_t), {}_d\hat{Y}_t)$. In considerazione del fatto che il modello (14) è di tipo empirico, l'insieme delle stime ${}_d\hat{Y}_t$ utilizzate per interpolare il modello è stato determinato in modo da includere la maggior parte delle stime pubblicate. Al fine di ottenere un migliore adattamento alla nuvola di punti, è stato interpolato un modello - con riferimento ad ogni trimestre e con riferimento all'intero anno - per ciascuno dei seguenti sottogruppi di stime:

- numero di notti per viaggi di lavoro;
- numero di notti per viaggi di vacanza;
- numero totale di notti in viaggio;
- numero di viaggi di lavoro;
- numero di viaggi di vacanza
- numero totale di viaggi.

Analogamente al caso precedente, dall'analisi dei coefficienti di regressione ottenuti è emerso che i coefficienti relativi al primo, al secondo e al quarto trimestre sono sostanzialmente uguali tra loro, mentre differiscono significativamente da quelli relativi al terzo trimestre; ciò è spiegabile considerando il fatto che i fenomeni indagati presentano andamenti difforni nel terzo trimestre rispetto agli altri tre.

Al fine di non appesantire la trattazione vengono qui, pertanto, riportati i risultati relativi al primo trimestre, al terzo trimestre e all'intero anno.

I prospetti 7, 9 e 11 (per le stime relative alle notti e ai viaggi) riportano i valori dei coefficienti a, b, c e dell'indice di determinazione R^2 dei modelli utilizzati per l'interpolazione degli errori campionari delle stime di totali per area territoriale e per ciascuno dei sei tipi di stime sopra elencati, con riferimento al primo trimestre, al terzo trimestre e all'intero anno.

Nel paragrafo successivo verrà illustrato come calcolare l'errore di campionamento assoluto e relativo di una stima a partire dalle informazioni contenute nei prospetti.

Utilizzando i valori stimati dei parametri del modello (14) e dividendo ambo i membri del modello per il valore della stima, \hat{Y}_t , si perviene alla seguente equazione di secondo grado:

$$a + [b - \varepsilon(\hat{Y}_t)] \hat{Y}_t + c(\hat{Y}_t)^2 = 0$$

la cui radice positiva è espressa dalla seguente formula:

$$\hat{Y}_t = \frac{-[b - \varepsilon(\hat{Y}_t)] - \sqrt{[b - \varepsilon(\hat{Y}_t)]^2 - 4ac}}{2c} \quad (15)$$

Nei prospetti 8, 10 e 12 (per le stime relative alle notti e ai viaggi) vengono riportati i valori delle stime \hat{Y}_t^* ottenuti sulla base della relazione (15) in corrispondenza di alcuni valori tipici prefissati dell'errore relativo percentuale, indicati come $\varepsilon^* = \varepsilon(\hat{Y}_t^*)$; le stime con valori superiori a \hat{Y}_t^* presentano valori dell'errore relativo inferiori a ε^* , mentre le stime che assumono valori inferiori a \hat{Y}_t^* presentano valori dell'errore relativo superiori a ε^* . In tali prospetti, articolati per gruppo di stime, trimestre e area territoriale, i valori prefissati di ε^* sono: 1, 5, 10, 15, 20, 25 e 30%. L'utilizzo di tali prospetti verrà meglio chiarito nel paragrafo seguente attraverso un esempio numerico.

Prospetto 3 - Valori dei coefficienti a, b e di R^2 delle funzioni utilizzate per l'interpolazione degli errori campionari delle stime di frequenze assolute riferite agli individui per aree territoriali - Trimestri – Anno 2011

AREE TERRITORIALI	Trimestri 1,2,e 4			Trimestre 3		
	a	b	R^2	a	b	R^2
Nord	11,09218	-1,15018	91,74900	9,16540	-0,99115	84,48225
Centro	10,49628	-1,12737	90,15162	9,37362	-1,01946	86,04655
Mezzogiorno	10,87365	-1,16376	91,84051	10,94860	-1,15177	91,03979
Italia	11,55210	-1,18053	93,17313	10,52484	-1,08819	91,99072

Prospetto 4 - Valori dei coefficienti a, b e di R^2 delle funzioni utilizzate per l'interpolazione degli errori campionari delle stime di frequenze assolute riferite agli individui per aree territoriali - Media dei 4 trimestri – Anno 2011

AREE TERRITORIALI	a	b	R^2
Nord	9,34707	-1,09515	90,6
Centro	9,62308	-1,13890	91,7
Mezzogiorno	10,01976	-1,17084	94,2
Italia	10,07632	-1,14653	93,7

Prospetto 5 - Valori interpolati degli errori relativi percentuali delle stime di frequenze assolute riferite agli individui per aree territoriali - Trimestri – Anno 2011

STIME	Trimestri 1,2 e 4				Trimestre 3			
	Aree territoriali				Aree territoriali			
	Nord	Centro	Mezzogiorno	Italia	Nord	Centro	Mezzogiorno	Italia
50.000	50,9	42,7	42,4	54,3	45,9	43,7	46,9	53,5
60.000	45,8	38,5	38,1	48,8	41,9	39,8	42,2	48,5
70.000	41,9	35,3	34,8	44,5	38,8	36,8	38,7	44,6
80.000	38,8	32,8	32,2	41,2	36,3	34,4	35,8	41,5
90.000	36,3	30,7	30,1	38,4	34,3	32,4	33,4	38,9
100.000	34,1	28,9	28,3	36,1	32,5	30,7	31,5	36,7
200.000	22,9	19,5	18,9	24,0	23,1	21,5	21,1	25,2
300.000	18,1	15,6	14,9	18,9	18,9	17,5	16,7	20,2
400.000	15,4	13,2	12,6	15,9	16,4	15,1	14,2	17,3
500.000	13,5	11,7	11,1	14,0	14,7	13,5	12,5	15,3
750.000	10,7	9,3	8,8	11,0	12,0	11,0	9,9	12,3
1.000.000	9,1	7,9	7,4	9,3	10,4	9,5	8,4	10,5
2.000.000	6,1	5,3	5,0	6,2	7,4	6,7	5,6	7,2
3.000.000	4,8	4,2	3,9	4,8	6,0	5,4	4,4	5,8
4.000.000	4,1	3,6	3,3	4,1	5,2	4,7	3,8	4,9
5.000.000	3,6	3,2	2,9	3,6	4,7	4,2	3,3	4,4
7.500.000	2,9	2,5	2,3	2,8	3,8	3,4	2,6	3,5
10.000.000	2,4	2,2	1,9	2,4	3,3	2,9	2,2	3,0
15.000.000	1,9	1,7	1,5	1,9	2,7	2,4	1,8	2,4
20.000.000	1,6	1,5	1,3	1,6	2,4	2,1	1,5	2,1
25.000.000	1,4	1,3	1,1	1,4	2,1	1,8	1,3	1,8

Prospetto 6 - Valori interpolati degli errori relativi percentuali delle stime di frequenze assolute riferite agli individui per aree territoriali - Media dei 4 trimestri – Anno 2011

STIME	Aree territoriali			
	Nord	Centro	Mezzogiorno	Italia
20.000	47,3	43,7	45,5	52,8
30.000	37,9	34,7	35,9	41,8
40.000	32,3	29,4	30,3	35,5
50.000	28,6	25,9	26,6	31,2
60.000	25,9	23,4	23,9	28,1
70.000	23,8	21,4	21,8	25,7
80.000	22,1	19,8	20,2	23,8
90.000	20,7	18,6	18,9	22,3
100.000	19,6	17,5	17,7	21,0
200.000	13,4	11,8	11,8	14,1
300.000	10,7	9,3	9,3	11,2
400.000	9,2	7,9	7,9	9,5
500.000	8,1	7,0	6,9	8,3
750.000	6,5	5,5	5,4	6,6
1.000.000	5,5	4,7	4,6	5,6
2.000.000	3,8	3,2	3,1	3,8
3.000.000	3,0	2,5	2,4	3,0
4.000.000	2,6	2,1	2,0	2,5
5.000.000	2,3	1,9	1,8	2,2
7.500.000	1,8	1,5	1,4	1,8
10.000.000	1,6	1,3	1,2	1,5
15.000.000	1,3	1,0	0,9	1,2
20.000.000	1,1	0,9	0,8	1,0

Prospetto 7 - Valori dei coefficienti a, b, c e di R² delle funzioni utilizzate per l'interpolazione degli errori campionari delle stime di totali per gruppi di stime, trimestre e anno, aree territoriali - Stime del numero di notti – Anno 2011

Notti – lavoro				
AREE TERRITORIALI	a	b	c	R ²
Gennaio-marzo				
Nord	40.020,746	0,342714024	-0,000000069	90,89
Centro	24.738,632	0,445300127	-0,000000114	97,15
Mezzogiorno	40.203,738	0,443859163	-0,000000064	94,91
Italia	67.746,577	0,253623009	-0,000000017	95,77
Luglio-settembre				
Nord	14.893,095	0,602312381	-0,000000151	95,76
Centro	-5.213,418	0,818589898	-0,000000345	85,63
Mezzogiorno	27.552,398	0,479305731	-0,000000173	91,17
Italia	53.765,477	0,366406517	-0,000000045	91,68
Anno				
Nord	80.540,755	0,421870981	-0,000000014	82,27
Centro	52.211,319	0,370558683	-0,000000031	90,99
Mezzogiorno	108.815,070	0,294676560	-0,000000016	90,86
Italia	229.454,349	0,233306144	-0,000000004	81,26
Notti – vacanze				
AREE TERRITORIALI	a	b	c	R ²
Gennaio-marzo				
Nord	106.121,636	0,2198352033	-0,0000000033	89,85
Centro	103.689,334	0,2032784793	-0,0000000068	92,55
Mezzogiorno	68.441,911	0,1886503951	-0,0000000088	93,00
Italia	161.899,051	0,1421541039	-0,0000000013	90,57
Luglio-settembre				
Nord	566.346,282	0,09075603608	-0,00000000026	87,54
Centro	326.268,343	0,14060584285	-0,00000000102	91,17
Mezzogiorno	246.407,157	0,17937663026	-0,00000000115	92,90
Italia	688.714,504	0,07168882646	-0,00000000011	90,94
Anno				
Nord	707.477,147	0,08835731135	-0,00000000016	92,14
Centro	341.595,451	0,13171433452	-0,00000000061	93,00
Mezzogiorno	277.038,982	0,15143922924	-0,00000000063	88,56
Italia	916.615,960	0,06597863105	-0,00000000007	93,47
Notti - totali				
AREE TERRITORIALI	a	b	c	R ²
Gennaio-marzo				
Nord	113.932,669	0,2033139649	-0,0000000027	91,33
Centro	117.290,610	0,1880833494	-0,0000000054	93,68
Mezzogiorno	81.191,275	0,1951027948	-0,0000000079	94,95
Italia	184.461,980	0,1276366816	-0,0000000010	92,20
Luglio-settembre				
Nord	571.343,957	0,08927722189	-0,00000000024	89,12
Centro	336.523,330	0,13679572728	-0,00000000093	92,16
Mezzogiorno	244.010,793	0,17752688049	-0,00000000110	93,61
Italia	696.284,878	0,07048766325	-0,00000000011	92,07
Anno				
Nord	812.677,059	0,08732210973	-0,00000000016	91,78
Centro	361.613,965	0,12678052334	-0,00000000056	93,23
Mezzogiorno	310.253,785	0,14193984156	-0,00000000052	89,16
Italia	1.027.370,008	0,06414322067	-0,00000000006	93,57

Prospetto 8 - Valori dei totali corrispondenti ad alcuni valori tipici degli errori relativi percentuali per gruppi di stime, trimestre e anno, aree territoriali - Stime del numero di notti – Anno 2011

Notti – lavoro							
AREE TERRITORIALI	1%	5%	10%	15%	20%	25%	30%
Gennaio-marzo							
Nord	4.955.591	4.389.089	3.687.253	2.996.558	2.325.523	1.692.123	1.134.219
Centro	3.890.563	3.543.686	3.111.773	2.682.530	2.257.381	1.838.901	1.432.114
Mezzogiorno	6.913.934	6.293.936	5.521.727	4.753.961	3.993.046	3.243.384	2.513.709
Italia	14.729.271	12.406.801	9.537.337	6.744.894	4.150.497	2.115.369	1.055.754
Luglio-settembre							
Nord	3.950.202	3.686.909	3.358.183	3.030.019	2.702.618	2.376.286	2.051.514
Centro	2.336.926	2.220.663	2.075.280	1.929.825	1.784.281	1.638.622	1.492.815
Mezzogiorno	2.775.585	2.549.026	2.267.221	1.987.538	1.710.932	1.438.973	1.174.372
Italia	7.993.589	7.130.984	6.059.683	5.000.414	3.961.819	2.962.011	2.041.527
Anno							
Nord	30.511.171	27.587.456	23.939.815	20.303.778	16.686.687	13.103.456	9.588.899
Centro	11.935.288	10.644.423	9.037.673	7.442.887	5.869.135	4.336.676	2.897.071
Mezzogiorno	18.356.284	15.887.826	12.832.535	9.837.325	6.966.905	4.388.325	2.458.972
Italia	58.864.340	48.711.580	36.180.213	24.054.428	13.149.782	5.845.205	2.940.133
Notti – vacanze							
AREE TERRITORIALI	1%	5%	10%	15%	20%	25%	30%
Gennaio-marzo							
Nord	63.545.761	51.643.582	36.868.724	22.404.744	9.364.238	2.708.568	1.258.078
Centro	28.972.765	23.216.307	16.145.363	9.455.312	4.155.186	1.765.838	1.001.570
Mezzogiorno	20.644.162	16.209.044	10.777.898	5.738.115	2.216.134	978.142	587.349
Italia	106.849.607	75.380.921	37.177.247	8.664.532	2.647.243	1.475.938	1.017.473
Luglio-settembre							
Nord	313.319.850	167.496.175	32.026.917	9.184.368	5.120.977	3.535.782	2.697.468
Centro	130.413.046	92.237.300	46.637.972	13.859.908	5.054.281	2.903.829	2.020.779
Mezzogiorno	149.040.496	114.616.037	72.147.348	32.256.501	8.203.421	3.310.911	2.004.554
Italia	552.313.437	218.055.250	22.321.243	8.684.836	5.342.189	3.852.944	3.012.033
Anno							
Nord	490.502.407	252.960.765	39.244.595	11.149.007	6.279.515	4.357.684	3.334.245
Centro	201.041.978	137.170.342	60.813.365	13.004.027	4.795.696	2.845.855	2.015.043
Mezzogiorno	226.174.620	163.504.360	86.620.412	22.129.165	5.335.256	2.762.022	1.850.283
Italia	852.983.894	286.691.746	25.648.975	10.816.188	6.816.140	4.972.045	3.912.430
Notti – totali							
AREE TERRITORIALI	1%	5%	10%	15%	20%	25%	30%
Gennaio-marzo							
Nord	71.452.967	56.938.080	38.947.060	21.488.481	7.098.703	2.166.225	1.141.609
Centro	33.549.541	26.335.867	17.512.223	9.353.408	3.682.745	1.654.927	999.688
Mezzogiorno	23.727.155	18.805.582	12.769.577	7.113.081	2.903.284	1.252.071	733.280
Italia	125.178.974	83.902.830	34.641.042	6.468.265	2.468.955	1.490.225	1.063.944
Luglio-settembre							
Nord	331.026.140	173.933.332	31.145.697	9.077.026	5.102.588	3.535.807	2.702.870
Centro	139.309.002	97.307.263	47.336.053	13.216.349	4.962.924	2.903.630	2.038.363
Mezzogiorno	153.491.096	117.626.819	73.381.777	31.921.376	7.842.483	3.210.239	1.957.877
Italia	582.163.567	222.848.629	21.875.027	8.657.072	5.352.764	3.869.918	3.029.521
Anno							
Nord	502.113.848	257.469.719	42.111.021	12.569.625	7.141.236	4.971.733	3.810.431
Centro	213.290.449	142.786.154	59.207.875	12.081.782	4.766.416	2.896.884	2.073.814
Mezzogiorno	258.002.851	181.473.662	88.095.412	17.923.708	5.111.440	2.832.798	1.950.464
Italia	894.184.235	286.679.376	27.361.032	11.864.705	7.536.304	5.517.622	4.350.943

Prospetto 9 - Valori dei coefficienti a, b, c e di R^2 delle funzioni utilizzate per l'interpolazione degli errori campionari delle stime di totali per gruppi di stime, trimestre e anno, aree territoriali - Stime del numero di viaggi – Anno 2011

Viaggi – lavoro				
AREE TERRITORIALI	a	b	c	R^2
Gennaio-marzo				
Nord	8.675,895	0,36618332	-0,00000023	96,00
Centro	3.507,021	0,62692363	-0,00000031	94,38
Mezzogiorno	8.083,608	0,44418293	-0,00000014	98,02
Italia	13.487,450	0,29889512	-0,00000005	96,42
Luglio-settembre				
Nord	4.430,795	0,56733703	-0,00000059	97,34
Centro	1.289,158	0,73497391	-0,00000094	95,15
Mezzogiorno	5.550,403	0,50997036	-0,00000041	91,79
Italia	10.880,353	0,35571391	-0,00000013	94,36
Anno				
Nord	24.007,341	0,287007582	-0,000000029	89,68
Centro	16.429,891	0,347041693	-0,000000076	89,95
Mezzogiorno	21.004,086	0,280822997	-0,000000037	93,20
Italia	39.205,852	0,177456642	-0,000000008	91,12
Viaggi – vacanza				
AREE TERRITORIALI	a	b	c	R^2
Gennaio-marzo				
Nord	32.703,978	0,174667399	-0,000000014	91,56
Centro	25.119,341	0,241500135	-0,000000037	91,83
Mezzogiorno	16.722,443	0,228605522	-0,000000047	93,25
Italia	44.686,410	0,122367074	-0,000000005	92,35
Luglio-settembre				
Nord	47.918,506	0,137991113	-0,000000004	84,85
Centro	30.279,486	0,202041295	-0,000000017	92,74
Mezzogiorno	33.303,716	0,149855063	-0,000000011	91,97
Italia	71.358,895	0,095784695	-0,000000002	87,22
Anno				
Nord	35.496,076	0,1444312971	-0,0000000028	80,06
Centro	32.078,902	0,1462089614	-0,0000000068	81,45
Mezzogiorno	28.461,567	0,1203519542	-0,0000000061	78,26
Italia	45.506,019	0,0940938584	-0,0000000010	78,51
Viaggi – totali				
AREE TERRITORIALI	a	b	c	R^2
Gennaio-marzo				
Nord	28.912,534	0,1699600303	-0,0000000133	91,13
Centro	27.164,163	0,2507050310	-0,0000000324	88,92
Mezzogiorno	17.212,483	0,2435304546	-0,0000000359	91,82
Italia	42.100,552	0,1253094354	-0,0000000048	91,20
Luglio-settembre				
Nord	42.065,947	0,1437527082	-0,0000000045	84,69
Centro	29.784,926	0,1999259257	-0,0000000163	92,34
Mezzogiorno	32.563,386	0,1486419348	-0,0000000102	90,98
Italia	62.907,972	0,0984593665	-0,0000000017	86,69
Anno				
Nord	51.364,229	0,1186393477	-0,0000000018	84,32
Centro	38.766,598	0,1586738242	-0,0000000056	88,77
Mezzogiorno	38.741,972	0,1362295184	-0,0000000044	85,83
Italia	72.278,923	0,0810833636	-0,0000000007	85,79

Prospetto 10 - Valori dei totali corrispondenti ad alcuni valori tipici degli errori relativi percentuali per gruppi di stime, trimestre e anno, aree territoriali - Stime del numero di viaggi – Anno 2011

Viaggi – lavoro							
AREE TERRITORIALI	1%	5%	10%	15%	20%	25%	30%
Gennaio-marzo							
Nord	1.602.761	1.428.391	1.211.586	996.802	785.554	581.149	391.565
Centro	1.979.198	1.851.629	1.692.250	1.532.990	1.373.888	1.215.007	1.056.443
Mezzogiorno	3.016.776	2.742.406	2.400.032	2.058.612	1.718.697	1.381.351	1.048.883
Italia	6.026.982	5.206.213	4.184.221	3.170.467	2.175.645	1.237.793	517.095
Luglio-settembre							
Nord	947.197	880.381	797.001	713.824	630.931	548.443	466.564
Centro	775.405	732.823	679.614	626.431	573.278	520.168	467.112
Mezzogiorno	1.242.942	1.145.319	1.023.538	902.135	781.279	661.255	542.578
Italia	2.705.287	2.399.837	2.019.677	1.642.485	1.270.715	910.188	576.857
Anno							
Nord	9.538.072	8.187.342	6.507.055	4.844.147	3.223.078	1.734.958	710.173
Centro	4.494.398	3.973.055	3.324.117	2.680.194	2.045.688	1.431.552	869.758
Mezzogiorno	7.420.946	6.349.180	5.017.097	3.701.543	2.426.499	1.280.632	538.227
Italia	20.434.655	15.678.975	9.826.299	4.390.028	1.205.137	510.653	313.296
Viaggi – vacanze							
AREE TERRITORIALI	1%	5%	10%	15%	20%	25%	30%
Gennaio-marzo							
Nord	11.488.426	8.754.184	5.380.359	2.229.656	515.271	215.340	132.750
Centro	6.250.256	5.191.887	3.877.706	2.586.274	1.371.417	495.559	202.678
Mezzogiorno	4.657.175	3.813.548	2.762.981	1.724.056	739.603	160.177	62.375
Italia	21.471.641	13.985.727	4.997.945	645.037	253.509	155.850	112.328
Luglio-settembre							
Nord	29.465.271	20.533.092	9.753.201	2.206.621	734.497	420.854	293.439
Centro	11.488.500	9.167.355	6.305.179	3.571.447	1.398.333	531.538	294.138
Mezzogiorno	13.033.223	9.460.778	5.154.039	1.739.208	588.645	321.292	218.341
Italia	50.543.167	28.015.375	5.324.715	1.265.270	677.135	460.352	348.404
Anno							
Nord	47.584.791	33.613.228	16.402.399	2.688.091	619.179	333.248	227.227
Centro	20.379.300	14.556.096	7.469.886	1.915.849	557.322	303.088	206.709
Mezzogiorno	18.362.246	11.935.508	4.400.834	821.323	348.072	217.310	157.587
Italia	81.653.977	43.540.789	4.363.203	802.043	427.890	291.317	220.758
Viaggi – totali							
AREE TERRITORIALI	1%	5%	10%	15%	20%	25%	30%
Gennaio-marzo							
Nord	12.225.802	9.269.923	5.654.293	2.407.743	728.136	341.842	217.505
Centro	7.550.420	6.335.502	4.831.454	3.362.096	1.989.148	927.221	429.801
Mezzogiorno	6.576.760	5.477.382	4.113.878	2.777.437	1.526.388	608.114	261.372
Italia	24.501.516	16.307.670	6.628.652	1.351.686	544.693	333.383	239.433
Luglio-settembre							
Nord	29.912.850	21.188.362	10.564.422	2.437.230	707.648	389.474	267.163
Centro	11.784.717	9.375.128	6.403.692	3.568.299	1.348.236	509.993	284.427
Mezzogiorno	13.785.155	9.963.178	5.350.443	1.719.095	569.460	311.480	212.101
Italia	51.662.225	29.159.882	5.592.592	1.174.122	613.109	413.167	311.301
Anno							
Nord	60.607.791	38.730.296	12.578.509	1.507.030	622.706	388.937	282.421
Centro	26.619.281	19.619.095	11.026.702	3.501.174	841.439	413.905	271.367
Mezzogiorno	29.039.066	20.069.061	9.205.479	1.790.602	584.026	336.164	235.080
Italia	108.398.875	49.180.547	3.413.263	1.038.431	605.769	427.182	329.837

Prospetto 11 - Valori dei coefficienti a, b, c e di R² delle funzioni utilizzate per l'interpolazione degli errori campionari delle stime di totali per gruppi di stime e tipologie comunali - Stime del numero di viaggi - Anno 2011

TIPOLOGIE COMUNALI	Viaggi – lavoro			R ²
	a	b	c	
Comuni centro dell'area metropolitana	16.964,973	0,376192450	-0,000000110	92,84
Comuni periferia dell'area metropolitana	17.768,076	0,467178263	-0,000000070	94,06
Comuni con oltre 50.000 abitanti	11.907,084	0,340065592	-0,000000103	88,43
Comuni da 10.001 a 50.000 abitanti	19.416,712	0,379084758	-0,000000082	86,55
Comuni da 2.001 a 10.000 abitanti	12.785,413	0,285571694	-0,000000087	95,58
Comuni aventi fino a 2.000 abitanti	3.861,607	0,637101644	-0,000000932	98,99
TIPOLOGIE COMUNALI	Viaggi – vacanze			R ²
	a	b	c	
Comuni centro dell'area metropolitana	29.459,358	0,2665672246	-0,0000000116	81,91
Comuni periferia dell'area metropolitana	28.877,967	0,1884586111	-0,0000000152	80,14
Comuni con oltre 50.000 abitanti	26.306,059	0,1460885580	-0,0000000094	79,07
Comuni da 10.001 a 50.000 abitanti	30.975,418	0,1331664605	-0,0000000066	76,88
Comuni da 2.001 a 10.000 abitanti	26.422,187	0,1446718303	-0,0000000093	76,35
Comuni aventi fino a 2.000 abitanti	12.341,011	0,3412463265	-0,0000000942	79,63
TIPOLOGIE COMUNALI	Viaggi – totali			R ²
	a	b	c	
Comuni centro dell'area metropolitana	39.075,779	0,2282103074	-0,0000000079	84,05
Comuni periferia dell'area metropolitana	40.701,220	0,2024786569	-0,0000000104	82,05
Comuni con oltre 50.000 abitanti	32.408,901	0,1545620915	-0,0000000070	87,69
Comuni da 10.001 a 50.000 abitanti	38.758,086	0,1423713002	-0,0000000052	85,77
Comuni da 2.001 a 10.000 abitanti	29.413,963	0,1470337051	-0,0000000066	88,77
Comuni aventi fino a 2.000 abitanti	15.411,892	0,3098630491	-0,0000000627	86,38

Prospetto 12 - Valori dei totali corrispondenti ad alcuni valori tipici degli errori relativi percentuali per gruppi di stime e tipologie comunali - Stime del numero di viaggi – Anno 2011

TIPOLOGIE COMUNALI	Viaggi – lavoro					
	1%	5%	10%	15%	20%	30%
Comuni centro dell'area metropolitana	3.360.326	3.003.687	2.559.969	2.119.843	1.685.908	866.818
Comuni periferia dell'area metropolitana	6.533.968	5.969.324	5.264.613	4.561.631	3.861.296	2.477.085
Comuni con oltre 50.000 abitanti	3.228.400	2.846.270	2.370.737	1.899.151	1.435.111	584.580
Comuni da 10.001 a 50.000 abitanti	4.543.251	4.062.659	3.464.275	2.869.968	2.282.717	1.165.138
Comuni da 2.001 a 10.000 abitanti	3.226.371	2.772.186	2.208.671	1.653.981	1.119.482	309.802
Comuni aventi fino a 2.000 abitanti	679.274	636.743	583.662	530.698	477.887	372.981
TIPOLOGIE COMUNALI	Viaggi – vacanze					
	1%	5%	10%	15%	20%	30%
Comuni centro dell'area metropolitana	22.174.188	18.755.759	14.496.380	10.269.244	6.136.311	2.455.860
Comuni periferia dell'area metropolitana	11.908.825	9.319.698	6.133.808	3.137.895	1.050.313	424.723
Comuni con oltre 50.000 abitanti	14.668.676	10.489.263	5.419.555	1.477.727	452.283	247.612
Comuni da 10.001 a 50.000 abitanti	18.828.890	12.908.377	5.808.007	1.237.279	443.925	261.252
Comuni da 2.001 a 10.000 abitanti	14.679.845	10.455.395	5.337.682	1.423.484	444.374	245.535
Comuni aventi fino a 2.000 abitanti	3.553.574	3.133.848	2.611.384	2.092.982	1.582.354	1.089.032
TIPOLOGIE COMUNALI	Viaggi – totali					
	1%	5%	10%	15%	20%	30%
Comuni centro dell'area metropolitana	27.812.224	22.785.848	16.535.900	10.381.287	4.639.249	1.237.948
Comuni periferia dell'area metropolitana	18.760.818	14.958.637	10.259.590	5.741.338	2.103.684	737.676
Comuni con oltre 50.000 abitanti	20.823.567	15.205.063	8.330.176	2.498.625	648.343	331.501
Comuni da 10.001 a 50.000 abitanti	25.809.424	18.218.440	8.999.104	2.095.346	636.126	354.068
Comuni da 2.001 a 10.000 abitanti	20.856.334	14.915.041	7.663.738	1.893.423	521.279	280.590
Comuni aventi fino a 2.000 abitanti	4.835.896	4.205.226	3.420.767	2.644.017	1.883.690	1.166.153

7. Esempi di calcolo degli errori campionari

7.1. Esempi di calcolo degli errori campionari per stime di totali riferiti ai viaggi e alle notti

Esempio 1

Come visto nel paragrafo precedente, il prospetto 7 presenta il valore dei coefficienti a, b e c delle funzioni utilizzate per l'interpolazione degli errori con riferimento alle stime del numero di notti, mentre il prospetto 9 presenta gli stessi coefficienti con riferimento alle stime del numero di viaggi.

Sulla base dei dati riportati in tali prospetti, è possibile calcolare l'errore di campionamento assoluto e relativo di una stima a livello trimestrale.

Se ad esempio si volesse calcolare l'errore di campionamento assoluto per una stima appartenente al gruppo del numero di notti per lavoro riferita al Nord per il terzo trimestre, ${}_C\hat{Y}_3$, si possono leggere i valori dei coefficienti a, b e c dal prospetto 7 e si può utilizzare la seguente espressione:

$$\hat{\sigma}({}_C\hat{Y}_3) = 14.893,095 + 0,602312381{}_C\hat{Y}_3 - 0,000000151({}_C\hat{Y}_3)^2.$$

Esempio 2

La valutazione degli errori di campionamento delle stime mediante un metodo approssimato si può ottenere utilizzando i dati dei prospetti 8 e 10. Il prospetto 8 fa riferimento alle stime del numero di notti; il prospetto 10 fa riferimento alle stime del numero di viaggi. Tali prospetti sono articolati per trimestre e area territoriale e presentano gruppi di stime per valori prefissati degli errori di campionamento relativi pari all'1, 5, 10, 15, 20, 25 e 30%.

Considerando ad esempio il prospetto 8, si può osservare che il valore riportato all'incrocio tra la riga riferita alle notti per vacanza del terzo trimestre per l'Italia e la colonna relativa all'errore del 10% indica che tutte le stime superiori a 22.321.243 presentano un errore relativo inferiore al 10%; analogamente si avrà che le stime inferiori a tale valore presenteranno un errore relativo superiore al 10%.

Esempio 3

Per fare un ulteriore esempio e meglio specificare l'utilizzo del suddetto prospetto 8, si consideri la stima del numero di notti per lavoro, riferita al totale Italia nel trimestre luglio-settembre, pari a 5.651.000.

Dal prospetto 8, considerando le notti-lavoro ed esaminando l'ultima riga corrispondente al totale Italia per il trimestre luglio-settembre, si può osservare che il valore 5.651.000 risulta essere compreso tra i valori 6.059.683 e 5.000.414

Da ciò si può facilmente desumere che il valore dell'errore relativo percentuale corrispondente è compreso tra il 10 ed il 15%.

7.2. Esempi di calcolo degli errori campionari per stime di frequenze riferite agli individui

Esempio 1

Il totale delle persone che hanno viaggiato nel trimestre gennaio-marzo 2011 è pari a 11.173.000.

Si cerca il livello di stima che più si avvicina a 11.173.000 nella prima colonna del prospetto 5, che è pari a 10.000.000. Dalla colonna del prospetto, riferita all'Italia per il primo trimestre, si ricava il corrispondente errore relativo percentuale della stima considerata, che è pari a 2,4%.

L'errore assoluto sarà:

$$\sigma(11.173.000) = 2,4/100 \times 11.173.000 = 268.152$$

L'intervallo di confidenza avrà come estremi:

$$11.173.000 - (2 \times 268.152) = 10.636.696$$

$$11.173.000 + (2 \times 268.152) = 11.709.304$$

Esempio 2

Considerando la stima precedente, si possono ottenere valori più precisi dell'errore di campionamento mediante interpolazione lineare dei due livelli di stima consecutivi tra i quali è compreso il valore della stima considerato.

Tali livelli sono 10.000.000 e 15.000.000 ai quali corrispondono i valori 2,4 e 1,9 percentuali.

L'errore corrispondente a 11.173.000 è pari a:

$$\sigma(11.173.000) = 2,4 - ((2,4 - 1,9) / (15.000.000 - 10.000.000)) \times (11.173.000 - 10.000.000) = 2,28\%$$

Il corrispondente errore assoluto è $2,28/100 \times 11.173.000 = 254.744$ e l'intervallo di confidenza avrà come estremi:

$$11.173.000 - (2 \times 254.744) = 10.663.512$$

$$11.173.000 + (2 \times 254.744) = 11.682.488$$

Esempio 3

Il calcolo dell'errore può essere effettuato direttamente tramite la funzione interpolante (12):

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

Per la stima di 11.173.000 si ha (prospetto 3):

$$\log \varepsilon^2 = 11,55210 - 1,18053 \times \log(11.173.000) = -7,606734$$

Da questa il valore di ε è facilmente desumibile, calcolando la radice quadrata dell'antilogaritmo dell'espressione di sopra.

Infatti, da $\log(\varepsilon^2) = -7,606734$ si ha:

$$\varepsilon^2 = \exp(-7,606734)$$

e dunque:

$$\varepsilon = 0,0222$$

L'errore relativo percentuale sarà dunque pari a 2,22%.