

File Standard

Indagine multiscopo sulle famiglie Viaggi e vacanze Anno 2008

Manuale per l'utente

Per informazioni sull'acquisto rivolgersi a:
Istat – Servizio Sistema informativo
diffusione e servizi all'utenza
via Cesare Balbo, 16 – 00184 Roma
tel: 06.4673.4126
fax: 06.46678200
<https://contact.istat.it/>

Per informazioni sull'indagine rivolgersi a:
Istat - Servizio struttura e dinamica sociale
U.O. Mobilità, viaggi e vacanze
via Adolfo Ravà, 150 – 00142 Roma
tel: 06.4673.4655, 4616
fax: 06.4673.4596
e-mail: maditorr@stat.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 alla Direttiva 95/57/CE attuata dal Consiglio dell'Unione Europea il 23/11/1995, 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 turisti.

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 1993² 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 ricettive³, 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 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.).

³ 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 Istat⁴.

La rilevazione utilizza le definizioni standard indicate nella metodologia adottata per l'attuazione della Direttiva 95/57/CE 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, 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**: viaggio svolto per motivi prevalenti di lavoro temporaneo fuori sede o più in generale per motivi professionali (missione, partecipazione a congressi, convegni ecc.). In questi viaggi sono inclusi anche quelli effettuati per formazione/aggiornamento professionale o per sostenere concorsi;
- **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: 35870
- **un file dei viaggi** (tipo A o tipo B)
lunghezza record: 132 bytes
numero record: 15786

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).

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

⁴ Il volume è consultabile anche sul sito Istat all'URL <http://www.istat.it>, scegliendolo nel [Catalogo](#), sotto la voce [Dati e prodotti](#).

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. 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 vincolata⁵, 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 le regioni un campione di uguale numerosità. È necessario quindi ricorrere a un procedimento complesso articolato in più fasi.

⁵ 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.

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 2008

REGIONI	Indirizzi campione	Tasso di sostituzione totale (a)	Tasso di sostituzione al netto dei mancati contatti (b)
Piemonte	266	22,2	13,9
Valle d'Aosta	93	37,6	21,5
Lombardia	329	18,5	12,8
Trentino-Alto Adige	206	28,2	15,0
<i>Bolzano-Bozen</i>	99	29,3	18,2
<i>Trento</i>	107	27,1	12,1
Veneto	201	19,9	16,4
Friuli-Venezia Giulia	139	27,3	20,1
Liguria	152	27,0	11,8
Emilia-Romagna	197	21,3	14,2
Toscana	213	23,0	15,0
Umbria	103	29,1	19,4
Marche	141	19,1	11,3
Lazio	220	29,5	18,6
Abruzzo	144	20,8	13,9
Molise	95	22,1	16,8
Campania	196	32,7	25,5
Puglia	187	24,6	19,8
Basilicata	99	25,3	19,2
Calabria	158	29,1	23,4
Sicilia	213	33,8	28,6
Sardegna	150	31,3	17,3
Italia	3502	25,6	17,5

(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 (secondo 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 2008

TIPOLOGIA DEL COMUNE	Indirizzi campione	Tasso di sostituzione totale	Tasso di sostituzione al netto dei mancati contatti
Comuni centro dell'area metropolitana	420	26,7	17,9
Comuni che gravitano intorno al centro dell'area metropolitana	320	24,4	18,8
Comuni aventi fino a 2.000 abitanti	296	30,1	18,9
Comuni con 2.001-10.000 abitanti	922	25,3	16,3
Comuni con 10.001-50.000 abitanti	911	25,4	17,8
Comuni con oltre 50.000 abitanti	633	24,2	17,2
Italia	3502	25,6	17,5

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:

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

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 noti⁶ sono:
 - popolazione residente per le 6 tipologie comunali;
 - popolazione residente per sesso nelle 20 regioni italiane;
 - popolazione residente per 8 classi di età⁷;
 - 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 ampiezza⁸

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:

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

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:

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

6 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".

7 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.

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

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

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:

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

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 $\hat{V}\text{ar}(d\hat{Y}_t)$ 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:

$$\hat{\sigma}(d\hat{Y}_t) = \sqrt{\hat{V}\text{ar}(d\hat{Y}_t)} . \quad (5)$$

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

$$\hat{\varepsilon}(d\hat{Y}_t) = \frac{\sqrt{\hat{V}\text{ar}(d\hat{Y}_t)}}{d\hat{Y}_t} . \quad (6)$$

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

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

dove

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

in cui \underline{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 \underline{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

$$\hat{\sigma}(d\hat{Y}) = \sqrt{\sum_{t=1}^4 \hat{V}\text{ar}(d\hat{Y}_t)} , \quad \hat{\varepsilon}(d\hat{Y}) = \frac{\hat{\sigma}(d\hat{Y})}{d\hat{Y}} \quad (8)$$

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

$$\hat{\sigma}(d\hat{Y}) = \frac{1}{4} \sqrt{\sum_{t=1}^4 \hat{V}\text{ar}(d\hat{Y}_t)} , \quad \hat{\varepsilon}(d\hat{Y}) = \frac{\hat{\sigma}(d\hat{Y})}{d\hat{Y}} . \quad (9)$$

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 $\hat{\epsilon}({}_d\hat{Y}_t)$, 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:

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

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 $\hat{\epsilon}({}_d\hat{Y}_t^k)$. 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 $\hat{\epsilon}({}_d\hat{Y}_t)$ 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 difformi 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, ${}_d\hat{Y}_t$, si perviene alla seguente equazione di secondo grado:

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

la cui radice positiva è espressa dalla seguente formula:

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

Nei prospetti 8, 10 e 12 (per le stime relative alle notti e ai viaggi) vengono riportati i valori delle stime ${}_d\hat{Y}_t^*$ ottenuti sulla base della relazione (15) in corrispondenza di alcuni valori tipici prefissati dell'errore relativo percentuale, indicati come $\varepsilon^* = \varepsilon({}_d\hat{Y}_t)$; le stime con valori superiori a ${}_d\hat{Y}_t^*$ presentano valori dell'errore relativo

inferiori a ϵ^* , mentre le stime che assumono valori inferiori a ${}_d\hat{Y}_t^*$ 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 2008

AREE TERRITORIALI	Trimestri 1,2,e 4			Trimestre 3		
	a	b	R^2	a	b	R^2
Nord	9,57097	-1,04303	91,2	10,96668	-1,14412	93,9
Centro	10,83387	-1,15943	93,8	10,30078	-1,10503	90,7
Mezzogiorno	9,55616	-1,07240	91,1	10,03622	-1,08737	91,9
Italia	11,00738	-1,14963	93,0	11,23708	-1,15500	94,2

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 2008

AREE TERRITORIALI	a	b	R^2
Nord	9,43612	-1,12787	93,0
Centro	9,69583	-1,16485	93,6
Mezzogiorno	8,67177	-1,08197	91,3
Italia	10,07038	-1,16719	94,5

Prospetto 5 - Valori interpolati degli errori relativi percentuali delle stime di frequenze assolute riferite agli individui per aree territoriali - Trimestri – Anno 2008

STIME	Trimestri 1,2 e 4				Trimestre 3			
	Aree territoriali				Aree territoriali			
	Nord	Centro	Mezzogiorno	Italia	Nord	Centro	Mezzogiorno	Italia
50.000	42,4	42,5	35,9	48,9	49,4	43,7	42,1	53,3
60.000	38,6	38,2	32,6	44,0	44,5	39,5	38,2	47,9
70.000	35,6	35,0	30,0	40,3	40,7	36,3	35,1	43,9
80.000	33,2	32,4	27,9	37,3	37,7	33,7	32,6	40,6
90.000	31,2	30,2	26,2	34,9	35,3	31,6	30,6	37,9
100.000	29,6	28,4	24,8	32,8	33,2	29,8	28,9	35,7
200.000	20,6	19,0	17,1	22,0	22,3	20,3	19,8	23,9
300.000	16,7	15,0	13,7	17,5	17,7	16,2	15,9	18,9
400.000	14,3	12,7	11,8	14,8	15,0	13,9	13,6	16,0
500.000	12,8	11,2	10,5	13,0	13,2	12,2	12,0	14,1
750.000	10,3	8,8	8,4	10,3	10,5	9,8	9,7	11,1
1.000.000	8,9	7,5	7,2	8,7	8,9	8,4	8,3	9,4
2.000.000	6,2	5,0	5,0	5,9	6,0	5,7	5,7	6,3
3.000.000	5,0	4,0	4,0	4,6	4,7	4,6	4,5	5,0
4.000.000	4,3	3,4	3,4	3,9	4,0	3,9	3,9	4,2
5.000.000	3,8	2,9	3,0	3,5	3,5	3,4	3,4	3,7
7.500.000	3,1	2,3	2,4	2,7	2,8	2,7	2,8	2,9
10.000.000	2,7	2,0	2,1	2,3	2,4	2,3	2,4	2,5
15.000.000	-	-	-	1,8	-	-	-	2,0
20.000.000	-	-	-	1,6	-	-	-	1,7
25.000.000	-	-	-	-	-	-	-	1,5
30.000.000	-	-	-	-	-	-	-	1,3

Prospetto 6 - Valori interpolati degli errori relativi percentuali delle stime di frequenze assolute riferite agli individui per aree territoriali - Media dei 4 trimestri – Anno 2008

STIME	Aree territoriali			
	Nord	Centro	Mezzogiorno	Italia
20.000	42,0	39,8	36,0	47,5
30.000	33,4	31,5	28,9	37,5
40.000	28,4	26,6	24,7	31,7
50.000	25,1	23,4	21,9	27,8
60.000	22,6	21,0	19,9	25,0
70.000	20,7	19,2	18,3	22,9
80.000	19,2	17,8	17,0	21,2
90.000	18,0	16,6	16,0	19,7
100.000	17,0	15,6	15,1	18,6
200.000	11,5	10,4	10,4	12,4
300.000	9,1	8,2	8,3	9,8
400.000	7,8	7,0	7,1	8,3
500.000	6,8	6,1	6,3	7,3
750.000	5,4	4,8	5,1	5,7
1.000.000	4,6	4,1	4,3	4,8
2.000.000	3,1	2,7	3,0	3,2
3.000.000	2,5	2,2	2,4	2,6
4.000.000	2,1	1,8	2,0	2,2
5.000.000	1,9	1,6	1,8	1,9
7.500.000	1,5	1,3	1,5	1,5
10.000.000	1,3	1,1	1,2	1,3
15.000.000	1,0	0,8	1,0	1,0
20.000.000	-	-	-	0,8

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 2008

Notti – lavoro				
AREE TERRITORIALI	a	b	c	R ²
Gennaio-marzo				
Nord	53.228,040	0,412473	-0,000000040499	95,4
Centro	-5.602,746	0,899343	-0,000000241940	94,0
Mezzogiorno	14.515,121	0,507661	-0,000000056884	91,4
Italia	120.706,934	0,282019	-0,000000010948	94,8
Luglio-settembre				
Nord	50.136,041	0,468442	-0,000000028442	88,3
Centro	26.969,029	0,442985	-0,000000088100	97,1
Mezzogiorno	29.971,655	0,487099	-0,000000058677	87,9
Italia	108.093,708	0,296477	-0,000000010291	87,5
Anno				
Nord	173.659,03	0,20451225	-0,000000004248	90,6
Centro	100.163,16	0,32046619	-0,000000012395	90,0
Mezzogiorno	131.926,39	0,28702114	-0,000000006489	90,9
Italia	303.360,33	0,15073356	-0,000000001305	91,7
Notti – vacanze				
AREE TERRITORIALI	a	b	c	R ²
Gennaio-marzo				
Nord	169.276,799	0,134299	-0,000000001476	92,1
Centro	127.000,415	0,163146	-0,000000004303	88,6
Mezzogiorno	97.030,631	0,250145	-0,000000005214	73,2
Italia	307.046,685	0,095631	-0,000000000519	86,2
Luglio-settembre				
Nord	750.113,562	0,065945	-0,000000000156	86,8
Centro	389.378,857	0,133661	-0,000000000790	87,3
Mezzogiorno	342.380,294	0,135135	-0,000000000671	88,8
Italia	1.012.770,581	0,054529	-0,000000000067	89,1
Anno				
Nord	930.616,47	0,055077	-0,000000000079	91,3
Centro	483.981,01	0,106572	-0,000000000396	88,3
Mezzogiorno	507.800,43	0,101978	-0,000000000291	86,7
Italia	1.284.875,31	0,044110	-0,000000000033	91,2
Notti - totali				
AREE TERRITORIALI	a	b	c	R ²
Gennaio-marzo				
Nord	202.546,849	0,126640	-0,000000001222	93,3
Centro	161.038,797	0,171766	-0,000000003978	89,0
Mezzogiorno	91.566,654	0,280308	-0,000000005634	82,5
Italia	360.168,663	0,092930	-0,000000000426	88,2
Luglio-settembre				
Nord	797.757,592	0,065073	-0,000000000146	87,7
Centro	401.868,399	0,129540	-0,000000000716	88,8
Mezzogiorno	376.401,326	0,128174	-0,000000000580	89,5
Italia	1.055.995,436	0,053203	-0,000000000062	90,2
Anno				
Nord	982.169,01	0,054157621	-0,000000000075	91,7
Centro	539.033,29	0,102800977	-0,000000000358	89,3
Mezzogiorno	601.412,75	0,099648709	-0,000000000258	87,5
Italia	1.399.055,30	0,043152149	-0,000000000030	91,7

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 2008

Notti – lavoro							
AREE TERRITORIALI	1%	5%	10%	15%	20%	25%	30%
Gennaio-marzo							
Nord	10.068.414	9.094.709	7.882.334	6.677.809	5.485.966	4.316.288	3.189.287
Centro	3.669.567	3.503.939	3.296.861	3.089.728	2.882.526	2.675.241	2.467.851
Mezzogiorno	8.777.772	8.077.108	7.201.965	6.327.878	5.455.346	4.585.240	3.719.216
Italia	25.283.548	21.701.743	17.265.076	12.913.081	8.751.839	5.090.676	2.599.348
Luglio-settembre							
Nord	16.227.336	14.831.171	13.089.002	11.351.625	9.621.559	7.903.400	6.206.396
Centro	4.976.237	4.528.291	3.970.254	3.415.245	2.864.924	2.322.349	1.793.663
Mezzogiorno	8.193.344	7.517.248	6.673.707	5.832.616	4.995.169	4.163.467	3.341.516
Italia	28.208.588	24.380.349	19.626.321	14.935.995	10.385.713	6.207.923	3.074.202
Anno							
Nord	46.662.437	37.461.951	26.163.617	15.473.479	6.946.660	2.985.356	1.691.377
Centro	25.366.770	22.185.282	18.230.339	14.317.516	10.489.526	6.862.703	3.785.787
Mezzogiorno	43.164.736	37.077.385	29.512.101	22.039.821	14.786.494	8.188.577	3.618.567
Italia	109.973.203	80.105.226	44.148.742	15.531.553	5.388.536	2.942.234	1.997.464
Notti – vacanze							
AREE TERRITORIALI	1%	5%	10%	15%	20%	25%	30%
Gennaio-marzo							
Nord	85.537.299	59.043.869	27.415.621	6.638.099	2.442.428	1.436.716	1.012.447
Centro	36.402.688	27.373.907	16.467.734	7.171.067	2.635.235	1.369.332	902.395
Mezzogiorno	46.461.768	38.867.792	29.431.026	20.132.782	11.269.520	4.327.961	1.658.576
Italia	168.439.893	94.167.529	20.472.383	5.371.901	2.900.099	1.975.915	1.496.723
Luglio-settembre							
Nord	372.640.613	137.575.925	20.168.923	8.781.433	5.559.719	4.061.555	3.198.069
Centro	159.637.478	110.377.127	52.078.440	14.150.498	5.508.234	3.274.139	2.315.412
Mezzogiorno	189.100.421	130.720.272	60.735.219	14.079.919	5.017.783	2.930.531	2.059.462
Italia	682.663.502	160.698.924	21.582.471	10.529.879	6.939.700	5.171.959	4.121.164
Anno							
Nord	589.990.768	145.231.991	20.010.739	9.725.071	6.399.090	4.765.054	3.794.971
Centro	248.639.619	150.871.845	44.213.344	10.195.993	5.071.201	3.343.505	2.489.433
Mezzogiorno	321.477.909	187.890.864	45.308.088	9.971.728	5.103.155	3.407.742	2.554.771
Italia	1.065.137.241	127.141.718	22.683.976	12.088.267	8.227.778	6.234.323	5.017.933
Notti – totali							
AREE TERRITORIALI	1%	5%	10%	15%	20%	25%	30%
Gennaio-marzo							
Nord	97.178.171	65.271.347	27.774.472	6.476.777	2.644.530	1.616.052	1.158.895
Centro	41.635.843	31.878.522	20.058.206	9.661.368	3.736.536	1.878.913	1.210.373
Mezzogiorno	48.311.934	41.269.948	32.501.949	23.810.246	15.314.650	7.535.823	2.646.295
Italia	198.757.525	108.477.059	21.932.920	6.038.572	3.319.968	2.278.947	1.733.171
Luglio-settembre							
Nord	391.462.424	141.862.207	20.998.821	9.246.572	5.875.188	4.299.322	3.388.636
Centro	170.310.390	115.969.892	52.057.347	13.379.556	5.406.592	3.272.495	2.334.671
Mezzogiorno	207.022.556	139.527.713	59.518.975	12.856.138	5.035.791	3.045.524	2.174.638
Italia	723.103.317	159.227.734	21.930.606	10.834.466	7.171.929	5.356.903	4.274.228
Anno							
Nord	613.675.010	145.987.195	20.726.196	10.167.329	6.711.426	5.005.560	3.990.287
Centro	265.218.135	157.253.581	42.940.310	10.573.477	5.436.922	3.629.928	2.720.033
Mezzogiorno	353.820.435	203.738.122	47.588.890	11.290.682	5.903.414	3.972.946	2.990.269
Italia	1.136.838.580	129.844.101	24.296.461	13.045.728	8.904.535	6.757.017	5.443.530

Prospetto 9 - 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 viaggi – Anno 2008

Viaggi – lavoro				
AREE TERRITORIALI	a	b	c	R ²
Gennaio-marzo				
Nord	11.719,780	0,272586	-0,000000082415	95,5
Centro	11.929,892	0,392612	-0,000000184801	92,0
Mezzogiorno	8.630,212	0,329973	-0,000000120087	93,6
Italia	20.772,104	0,184248	-0,000000022654	94,0
Luglio-settembre				
Nord	9.994,002	0,331123	-0,000000117525	91,6
Centro	6.632,827	0,404587	-0,000000193006	97,5
Mezzogiorno	8.639,617	0,312903	-0,000000197456	91,0
Italia	14.800,150	0,212445	-0,000000033842	95,0
Anno				
Nord	24.960,20	0,150360	-0,000000010980	95,1
Centro	20.829,90	0,202748	-0,000000028190	94,0
Mezzogiorno	20.540,58	0,185980	-0,000000021746	88,1
Italia	40.463,13	0,100952	-0,000000003437	93,7
Viaggi – vacanza				
AREE TERRITORIALI	a	b	c	R ²
Gennaio-marzo				
Nord	36.994,083	0,130731	-0,000000005839	95,3
Centro	19.259,709	0,216092	-0,000000025327	88,6
Mezzogiorno	19.597,732	0,197831	-0,000000029778	85,5
Italia	51.415,787	0,097847	-0,000000002614	92,6
Luglio-settembre				
Nord	57.586,731	0,089802	-0,000000002375	90,9
Centro	31.834,124	0,140411	-0,000000009488	93,2
Mezzogiorno	25.643,934	0,159629	-0,000000008955	92,1
Italia	78.234,003	0,069007	-0,000000000985	91,4
Anno				
Nord	91.190,08	0,062443	-0,000000000733	91,0
Centro	51.105,39	0,097465	-0,000000002994	90,0
Mezzogiorno	55.749,29	0,103067	-0,000000002755	86,2
Italia	128.320,82	0,046690	-0,000000000302	88,7
Viaggi – totali				
AREE TERRITORIALI	a	b	c	R ²
Gennaio-marzo				
Nord	34.212,881	0,126218	-0,000000005218	94,7
Centro	23.046,531	0,201188	-0,000000021015	88,0
Mezzogiorno	20.561,384	0,187536	-0,000000023157	85,6
Italia	48.169,508	0,093873	-0,000000002265	92,0
Luglio-settembre				
Nord	47.592,413	0,093152	-0,000000002462	89,7
Centro	29.309,729	0,142845	-0,000000009551	92,2
Mezzogiorno	23.578,266	0,156183	-0,000000008462	92,0
Italia	61.435,373	0,071290	-0,000000001012	90,6
Anno				
Nord	57.331,82	0,075285	-0,000000000790	87,8
Centro	37.081,94	0,113661	-0,000000002913	86,8
Mezzogiorno	37.660,05	0,127977	-0,000000002739	84,3
Italia	75.896,36	0,057587	-0,000000000320	86,1

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 2008

Viaggi – lavoro							
AREE TERRITORIALI	1%	5%	10%	15%	20%	25%	30%
Gennaio-marzo							
Nord	3.230.144	2.752.439	2.159.929	1.577.552	1.020.125	538.244	245.829
Centro	2.101.121	1.888.138	1.623.158	1.360.283	1.100.903	847.844	607.420
Mezzogiorno	2.691.221	2.361.853	1.951.878	1.545.203	1.145.089	760.466	420.503
Italia	7.809.135	6.076.915	3.950.990	1.975.866	671.065	287.450	173.559
Luglio-settembre							
Nord	2.763.151	2.427.060	2.008.912	1.594.474	1.187.321	796.961	452.674
Centro	2.061.106	1.855.704	1.599.609	1.344.624	1.091.491	841.772	599.237
Mezzogiorno	1.562.036	1.363.537	1.117.385	875.011	640.138	422.200	244.384
Italia	6.054.370	4.889.602	3.449.474	2.057.752	870.277	308.393	159.238
Anno							
Nord	12.958.692	9.382.563	5.037.781	1.524.220	456.692	243.947	164.809
Centro	6.943.973	5.551.693	3.837.442	2.206.120	909.721	362.450	202.319
Mezzogiorno	8.207.784	6.400.813	4.179.898	2.103.620	701.601	291.904	174.351
Italia	26.900.789	15.580.503	3.572.502	782.110	402.889	269.799	202.575
Viaggi – vacanze							
AREE TERRITORIALI	1%	5%	10%	15%	20%	25%	30%
Gennaio-marzo							
Nord	20.978.985	14.270.368	6.273.133	1.359.677	511.970	305.601	216.929
Centro	8.229.814	6.672.015	4.744.116	2.874.187	1.245.803	429.937	215.515
Mezzogiorno	6.410.400	5.093.661	3.474.761	1.944.685	775.650	317.954	182.147
Italia	34.176.522	19.319.265	4.042.023	941.438	497.001	335.982	253.510
Luglio-settembre							
Nord	34.309.532	18.099.500	3.224.966	923.017	516.820	357.577	273.122
Centro	13.984.996	9.869.128	4.938.636	1.394.832	495.186	283.527	197.165
Mezzogiorno	16.877.994	12.471.352	7.063.882	2.313.184	564.520	276.204	180.606
Italia	61.221.343	22.788.709	2.348.956	954.853	594.581	431.237	338.198
Anno							
Nord	73.235.000	22.501.486	2.322.742	1.032.568	660.601	485.279	383.413
Centro	29.783.539	16.863.924	3.729.679	924.121	491.370	332.866	251.395
Mezzogiorno	34.365.359	20.258.242	5.088.960	1.114.877	566.025	376.758	281.975
Italia	124.993.123	15.854.080	2.375.142	1.237.622	835.628	630.569	506.271
Viaggi – totali							
AREE TERRITORIALI	1%	5%	10%	15%	20%	25%	30%
Gennaio-marzo							
Nord	22.565.190	15.044.011	6.100.024	1.148.990	449.422	273.250	195.723
Centro	9.216.674	7.343.626	5.032.942	2.824.120	1.075.877	402.429	222.690
Mezzogiorno	7.780.857	6.085.286	4.002.031	2.053.375	710.849	296.565	176.418
Italia	37.588.769	20.408.189	3.453.051	830.388	449.570	307.158	233.091
Luglio-settembre							
Nord	34.337.546	18.568.553	3.220.712	808.858	440.943	301.999	229.458
Centro	14.125.619	10.026.563	5.088.706	1.416.773	475.088	267.163	184.434
Mezzogiorno	17.435.269	12.766.641	7.035.591	2.074.123	491.464	245.870	162.395
Italia	61.529.412	23.601.616	1.998.940	772.842	475.536	343.104	268.298
Anno							
Nord	83.551.263	34.149.466	2.169.393	761.220	458.374	327.660	254.903
Centro	35.942.758	22.423.588	6.614.609	948.347	423.442	270.420	198.387
Mezzogiorno	43.392.215	28.945.751	11.419.092	1.448.955	512.888	306.523	218.167
Italia	150.102.175	31.258.601	1.765.914	818.951	532.295	394.187	312.958

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 Centro 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:

$$\hat{\sigma}({}_C\hat{Y}_3) = 26.969,02 + 0,442985{}_C\hat{Y}_3 - 0,0000000881({}_C\hat{Y}_3)^2.$$

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 21.582.471 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 14.909.000 (si veda il prospetto 1.1).

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 14.909.000 risulta essere compreso tra i valori 14.935.995 e 10.385.713.

Da ciò si può facilmente desumere che il valore dell'errore relativo percentuale corrispondente è compreso tra il 15 ed il 20%.

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

Esempio 1

Dal prospetto 2.1, risulta che il totale delle persone che hanno viaggiato nel trimestre gennaio-marzo è pari a 14.226.000.

Si cerca il livello di stima che più si avvicina a 14.226.000 nella prima colonna del prospetto 5, che è pari a 15.000.000. 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 1,8%.

L'errore assoluto sarà:

$$\sigma(14.226.000) = 1,8/100 \times 14.226.000 = 256.068$$

L'intervallo di confidenza avrà come estremi:

$$14.226.000 - (2 \times 256.068) = 13.713.864$$

$$14.226.000 + (2 \times 256.068) = 14.738.136$$

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 2,3 e 1,8 percentuali.

L'errore corrispondente a 14.226.000 è pari a:

$$\sigma(14.226.000) = 2,3 - ((2,3 - 1,8) / (15.000.000 - 10.000.000)) \times (14.226.000 - 10.000.000) = 1,88\%$$

Il corrispondente errore assoluto è $1,88/100 \times 14.226.000 = 267.449$ e l'intervallo di confidenza avrà come estremi:

$$14.226.000 - (2 \times 267.449) = 13.691.102$$

$$14.226.000 + (2 \times 267.449) = 14.760.898$$

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 14.226.000 si ha:

$$\log \varepsilon^2 = 11,00738 - 1,14963 \times \log(14.226.000) = -7,92769.$$

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

Infatti, da $\log(\varepsilon^2) = -7,92769$ si ha:

$$\varepsilon^2 = \exp(-7,92769)$$

e dunque:

$$\varepsilon = 0,019$$

L'errore relativo percentuale sarà dunque pari a 1,9 %.