

File Standard

Indagine multiscopo sulle famiglie Viaggi e vacanze Anno 2004

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: 06.4673.3283
fax: 06.4673.3477
e-mail: richieste.dati@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: perez@istat.it

INDICE

1. PREMESSA.....	4
2. L'INDAGINE TRIMESTRALE “VIAGGI E VACANZE”	4
3. I DATI CAMPIONARI E IL LORO UTILIZZO.....	5
4. IL TRACCIATO DEL FILE INDIVIDUI	8
5. IL TRACCIATO DEL FILE VIAGGI	13
APPENDICE A - IL QUESTIONARIO.....	20
APPENDICE B - ALCUNE DEFINIZIONI	21
APPENDICE C - STRATEGIA DI CAMPIONAMENTO E VALUTAZIONE DEGLI ERRORI CAMPIONARI...	22

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: 37.719
- **un file dei viaggi** (tipo A o tipo B)
lunghezza record: 132 bytes
numero record: 14.569

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) **notte**: 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, Sud);
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.

4. IL TRACCIATO DEL FILE INDIVIDUI

IDENTIFICATIVI FAMIGLIA E INDIVIDUO

Coll. 1-2	TR	TIPO RECORD "IA"=record individuale di tipo A "IB"=record individuale di tipo B
Coll. 3-4	TRIM	TRIMESTRE DI RIFERIMENTO 01=primo trimestre (gennaio-marzo 2004) 02=secondo trimestre (aprile-giugno 2004) 03=terzo trimestre (luglio-settembre 2004) 04=quarto trimestre (ottobre-dicembre 2004)
Coll. 5-8	ANNRIF	ANNO DI RIFERIMENTO = 2004
Coll. 9-12	IDFAM	PROGRESSIVO (tipo A o tipo B) DELLA FAMIGLIA NEL TRIMESTRE
Coll. 13-14	IDIND	NUMERO D'ORDINE DELL'INDIVIDUO ENTRO LA FAMIGLIA
Coll. 15-22	COE	COEFFICIENTE DI RIPORTO ALL'UNIVERSO (coll. 15-19 valori interi, col. 20 punto separatore, coll.21-22 valori decimali)
Col. 23	-	RISERVATO ISTAT
Col. 24	RIP	RIPARTIZIONE GEOGRAFICA DI RESIDENZA 1=Italia nord-occidentale 2=Italia nord-orientale 3=Italia centrale 4=Italia meridionale 5=Italia insulare
Coll. 25-27	REG	REGIONE DI RESIDENZA (<i>se record di tipo A</i>) 010=Piemonte 020=Valle d'Aosta 030=Lombardia 041=Bolzano- <i>Bozen</i> 042=Trento 050=Veneto 060=Friuli-Venezia Giulia 070=Liguria 080=Emilia-Romagna 090=Toscana 100=Umbria 110=Marche 120=Lazio 130=Abruzzo 140=Molise 150=Campania 160=Puglia 170=Basilicata 180=Calabria 190=Sicilia 200=Sardegna

Col. 28	DOM	TIPOLOGIA SOCIO-DEMOGRAFICA DEL COMUNE DI RESIDENZA <i>(se record di tipo B)</i> 1=Comune centro dell'area metropolitana 2=Comune periferia dell'area metropolitana 3=Comune con oltre 50.000 abitanti 4=Comune da 10.001 a 50.000 abitanti 5=Comune da 2.001 a 10.000 abitanti 6=Comune aventi fino a 2.000 abitanti
Coll. 29-31	-	RISERVATO ISTAT
Coll. 32-33	NCOMP	NUMERO COMPONENTI DELLA FAMIGLIA
Coll. 34-44	-	RISERVATO ISTAT

VIAGGI DELLA FAMIGLIA NEL TRIMESTRE

Col. 45	FVIA	PRESENZA DI VIAGGI IN FAMIGLIA 0=nessun viaggio 1=almeno un viaggio
Coll. 46-47	FNVIA	NUMERO DI VIAGGI EFFETTUATI IN FAMIGLIA
Col. 48	FLAV	PRESENZA DI VIAGGI DI LAVORO IN FAMIGLIA 0=nessun viaggio 1=almeno un viaggio
Coll. 49-50	FNLAV	NUMERO DI VIAGGI DI LAVORO EFFETTUATI IN FAMIGLIA
Col. 51	FVAC	PRESENZA DI VIAGGI DI VACANZA IN FAMIGLIA 0=nessun viaggio 1=almeno un viaggio
Coll. 52-53	FNVAC	NUMERO DI VIAGGI DI VACANZA EFFETTUATI IN FAMIGLIA
Col. 54	FBREVE	PRESENZA DI VACANZE "BREVI" (1-3 NOTTI) IN FAMIGLIA 0=nessun viaggio 1=almeno un viaggio
Coll. 55-56	FNBREVE	NUMERO DI VACANZE "BREVI" EFFETTUATE IN FAMIGLIA
Col. 57	FLUNGA	PRESENZA DI VACANZE "LUNGHE" (4 O PIÙ NOTTI) IN FAMIGLIA 0=nessun viaggio 1=almeno un viaggio
Coll. 58-59	FNLUNGA	NUMERO DI VACANZE "LUNGHE" EFFETTUATE IN FAMIGLIA
Coll. 60-64	-	RISERVATO ISTAT

CARATTERISTICHE SOCIO-DEMOGRAFICHE DELL'INDIVIDUO

Coll. 65-66	RELPAR	RELAZIONE DI PARENTELA CON L'INTERVISTATO 01=persona di riferimento (l'intervistato) 02=coniuge 03=convivente 04=genitore 05=suocero/suocera 06=figlio nato dall'ultimo matrimonio/convivenza 07=figlio nato da precedente matrimonio/convivenza 08=genero/nuora 09=convivente del figlio 10=nipote di nonno 11=nipote di zio 12=fratello/sorella 13=fratello/sorella del coniuge/convivente 14=moglie/marito del fratello/sorella 15=convivente del fratello/sorella 16=nonno/a 17=cugino/a 18=zio/a 19=altro parente 20=amico
Col. 67	SESSO	SESSO 1=maschio 2=femmina
Coll. 68-70	ETA	ETÀ (in anni compiuti)
Coll. 71-76	-	RISERVATO ISTAT
Col. 77	STACIV	STATO CIVILE 1=celibe/nubile 2=coniugato/a coabitante con il coniuge 3=separato/a di fatto 4=separato/a legalmente 5=divorziato/a 6=vedovo/a
Coll. 78-79	ISTR	TITOLO DI STUDIO <i>(se età maggiore di 5 anni)</i> 01=dottorato di ricerca o specializzazione post-laurea 02=laurea 03=diploma universitario 04=diploma o qualifica scuola media superiore (4-5 anni) 05=diploma o qualifica scuola media superiore (2-3 anni) 06=licenza di scuola media inferiore 07=licenza elementare 08=nessun titolo - sa leggere e scrivere 09=nessun titolo - non sa leggere e/o scrivere

Coll. 80-81	COND	CONDIZIONE PROFESSIONALE <i>(se età maggiore di 14 anni)</i> 01=occupato 02=in cerca di nuova occupazione 03=in cerca di prima occupazione 04=in servizio di leva 05=casalinga 06=studente 07=inabile al lavoro 08=persona ritirata dal lavoro 09=in altra condizione
Coll. 82-83	POSIZ	POSIZIONE NELLA PROFESSIONE <i>(se occupato: COND=01)</i> 01=dirigente 02=direttivo,quadro 03=impiegato,intermedio 04=capo operaio, operaio subalterno 05=apprendista 06=lavorante a domicilio per conto imprese 07=imprenditore 08=libero professionista 09=lavoratore in proprio 10=socio di cooperativa 11=coadiuvante
Col. 84	ATECO	SETTORE DI ATTIVITÀ ECONOMICA <i>(se occupato: COND=01)</i> 1=agricoltura, caccia e pesca 2=industria, estrazione, manifatture, energia 3=costruzioni 4=commercio all'ingrosso, al dettaglio, alberghi e ristoranti 5=trasporti, magazzinaggio, comunicazioni 6=intermediazione, noleggio ed altre attività professionali 7=pubblica amministrazione e difesa 8=istruzione, sanità ed altri servizi sociali 9=altri servizi
Coll. 85-90	-	RISERVATO ISTAT

VIAGGI DELL'INDIVIDUO NEL TRIMESTRE

Col. 91	IVIA	PRESENZA DI VIAGGI DELL'INDIVIDUO 0=nessun viaggio 1=almeno un viaggio
Coll. 92-93	INVIA	NUMERO DI VIAGGI EFFETTUATI DALL'INDIVIDUO
Col. 94	ILAV	PRESENZA DI VIAGGI DI LAVORO DELL'INDIVIDUO 0=nessun viaggio 1=almeno un viaggio
Coll. 95-96	INLAV	NUMERO DI VIAGGI DI LAVORO EFFETTUATI DALL'INDIVIDUO
Col. 97	IVAC	PRESENZA DI VIAGGI DI VACANZA DELL'INDIVIDUO 0=nessun viaggio 1=almeno un viaggio
Coll. 98-99	INVAC	NUMERO DI VIAGGI DI VACANZA EFFETTUATI DALL'INDIVIDUO

Col. 100	IBREVE	PRESENZA DI VACANZE "BREVI" (1-3 NOTTI) DELL'INDIVIDUO 0=nessun viaggio 1=almeno un viaggio
Coll. 101-102	INBREVE	NUMERO DI VIAGGI DI VACANZA "BREVE" EFFETTUATI DALL'INDIVIDUO
Col. 103	ILUNGA	PRESENZA DI VACANZE "LUNGHE" (4 O PIÙ NOTTI) DELL'INDIVIDUO 0=nessun viaggio 1=almeno un viaggio
Coll. 104-105	INLUNGA	NUMERO DI VIAGGI DI VACANZA "LUNGA" EFFETTUATI DALL'INDIVIDUO
Coll. 106-107	NOVAC	MOTIVO PER CUI L'INDIVIDUO NON HA FATTO VACANZE (se non ha effettuato vacanze nel trimestre: IVAC=0) 01=ragioni economiche 02=motivi di lavoro e di studio 03=mancanza di abitudine 04=residente già in località di villeggiatura 05=motivi di famiglia 06=motivi di salute 07=altri motivi 08=paura di atti terroristici 99=non indica

5. IL TRACCIATO DEL FILE VIAGGI

IDENTIFICATIVI FAMIGLIA E INDIVIDUO

Coll. 1-2	TR	TIPO RECORD "VA"=record viaggi di tipo A "VB"=record viaggi di tipo B
Coll. 3-4	TRIM	TRIMESTRE DI RIFERIMENTO 01=primo trimestre (gennaio-marzo 2004) 02=secondo trimestre (aprile-giugno 2004) 03=terzo trimestre (luglio-settembre 2004) 04=quarto trimestre (ottobre-dicembre 2004)
Coll. 5-8	ANNRIF	ANNO DI RIFERIMENTO = 2004
Coll. 9-12	IDFAM	PROGRESSIVO (tipo A o tipo B) DELLA FAMIGLIA NEL TRIMESTRE
Coll. 13-14	IDIND	NUMERO D'ORDINE DELL'INDIVIDUO ENTRO LA FAMIGLIA
Coll. 15-22	COE	COEFFICIENTE DI RIPORTO ALL'UNIVERSO (coll. 15-19 valori interi, col. 20 punto separatore, coll.21-22 valori decimali)
Col. 23	-	RISERVATO ISTAT
Col. 24	RIP	RIPARTIZIONE GEOGRAFICA DI RESIDENZA 1=Italia nord-occidentale 2=Italia nord-orientale 3=Italia centrale 4=Italia meridionale 5=Italia insulare
Coll. 25-27	REG	REGIONE DI RESIDENZA (<i>se record di tipo A</i>) 010=Piemonte 020=Valle d'Aosta 030=Lombardia 041=Bolzano-Bozen 042=Trento 050=Veneto 060=Friuli-Venezia Giulia 070=Liguria 080=Emilia-Romagna 090=Toscana 100=Umbria 110=Marche 120=Lazio 130=Abruzzo 140=Molise 150=Campania 160=Puglia 170=Basilicata 180=Calabria 190=Sicilia 200=Sardegna

Col. 28	DOM	TIPOLOGIA SOCIO-DEMOGRAFICA DEL COMUNE DI RESIDENZA (se record di tipo B) 1=Comune centro dell'area metropolitana 2=Comune periferia dell'area metropolitana 3=Comune con oltre 50.000 abitanti 4=Comune da 10.001 a 50.000 abitanti 5=Comune da 2.001 a 10.000 abitanti 6=Comune aventi fino a 2.000 abitanti
---------	-----	---

Coll. 29-31	-	RISERVATO ISTAT
-------------	---	-----------------

INFORMAZIONI SUL VIAGGIO

Coll. 32-34	IDVIA	NUMERO D'ORDINE DEL VIAGGIO ALL'INTERNO DELLA FAMIGLIA
-------------	-------	--

Coll. 35-38	TIPOVIA	TIPOLOGIA DEL VIAGGIO "L104"=viaggio di lavoro concluso nel primo trimestre 2004 "L204"=viaggio di lavoro concluso nel secondo trimestre 2004 "L304"=viaggio di lavoro concluso nel terzo trimestre 2004 "L404"=viaggio di lavoro concluso nel quarto trimestre 2004 "V104"=viaggio di vacanza concluso nel primo trimestre 2004 "V204"=viaggio di vacanza concluso nel secondo trimestre 2004 "V304"=viaggio di vacanza concluso nel terzo trimestre 2004 "V404"=viaggio di vacanza concluso nel quarto trimestre 2004
-------------	---------	---

Coll. 39-40	MOTLAV	MOTIVO DEL VIAGGIO DI LAVORO (se viaggio di lavoro: TIPOVIA= "L104", "L204", "L304", "L404") 01=partecipazione a congresso, convegno, conferenza, seminario 02=partecipazione a riunione d'affari 03=esibizione, fiera, mostra, esposizione (visita/partecipazione professionale) 04=missione di lavoro o militare (amministrazione pubblica o privata, organismo internazionale, altra istituzione) 05=viaggio o meeting d'azienda per la familiarizzazione dei dipendenti/soci (inclusi viaggi di incentivazione) 06=attività di rappresentanza, vendita, acquisto, pubblicità, installazione, collaudo, manutenzione, riparazione, ecc. 07=attività culturale, artistica, religiosa, sportiva 08=attività di docenza (didattica, consulenza, istruzioni) 09=partecipazione (allievo) a corso di lingua straniera 10=partecipazione (allievo) a corso di aggiornamento professionale (esclusa lingua straniera) 11=attività del personale viaggiante a bordo di mezzi per trasporto di persone e/o merci 12=altro motivo 13=attività nel settore dell'edilizia e dell'artigianato 14=attività di controllo ed ispezione
-------------	--------	--

Coll. 41-43	-	RISERVATO ISTAT
-------------	---	-----------------

Coll. 44-45	INSIEME	CODICE CHE INDIVIDUA I VIAGGI FATTI INSIEME (se più componenti familiari hanno viaggiato insieme, la variabile ha lo stesso valore in tutti i record viaggi considerati)
-------------	---------	---

Coll. 46-47	NPART	NUMERO DI COMPONENTI FAMILIARI PARTECIPANTI AL VIAGGIO
-------------	-------	--

Coll. 48-69	REL1-REL11	SEQUENZA DELLE RELAZIONI DI PARENTELA (RISPETTO PERSONA DI RIFERIMENTO) DEI COMPONENTI FAMILIARI PARTECIPANTI AL VIAGGIO (sequenza di 11 campi da 2 bytes, serrati a sinistra) - Codici di RELPAR.
-------------	------------	--

Col. 70	ALPARTN	PARTECIPAZIONE AL VIAGGIO DI PERSONE CHE NON VIVONO IN FAMIGLIA
---------	---------	---

1=no
 2=si, solo parenti
 3=si, solo amici/conoscenti
 4=si, sia parenti che amici/conoscenti

Coll. 71-73	NPARTN	NUMERO DI PERSONE CHE NON VIVONO IN FAMIGLIA PARTECIPANTI AL VIAGGIO 001-100= da 1 a 100 persone 101= oltre 100 persone 999= non indica
Col. 74	-	RISERVATO ISTAT
Coll. 75-82	INIZIO	DATA DI INIZIO DEL VIAGGIO (ggmmaaaa)
Coll. 83-85	DURATA	DURATA DEL VIAGGIO (numero di notti)
Coll. 86-87	MOTVAC	MOTIVO DEL VIAGGIO DI VACANZA (se <i>viaggio di vacanza</i> : TIPOVIA= "V104", "V204", "V304", "V404") 01=piacere, svago 02=visita a parenti e/o amici 03=motivi religiosi, pellegrinaggio 04=trattamenti di salute, cure termali 05=non sa / non risponde
Coll. 88-89	TIPO	TIPO DELLA VACANZA DI PIACERE/SVAGO (se il <i>motivo della vacanza</i> è <i>piacere/svago</i> : MOTVAC=01) 01=mare 02=crociera 03=lago 04=montagna 05=campagna, collina 06=agriturismo 07=giro turistico 08=visita a città e località d'arte 09=vacanza studio 10=vacanza sport 11=altro tipo di vacanza
Coll. 90-91	-	RISERVATO ISTAT
Coll. 92-93	MEZZO	MEZZO DI TRASPORTO UTILIZZATO 01=aereo 02=treno 03=nave, battello, motoscafo 04=auto a noleggio 05=auto propria, di parenti o di amici 06=pullman turistico 07=pullman di linea 08=camper, autocaravan 09=moto, motoscooter 10=altro
Coll. 94-95	ALLOG	TIPO DI ALLOGGIO PREVALENTE 01=albergo/motel/pensione 02=residenza per cure fisiche ed estetiche 03=campo lavoro e vacanza 04=mezzo pubblico di trasporto (cuccetta, vagone letto) 05=centro congressi e conferenze 06=villaggio vacanza 07=campeggio 08=marina (barca in porto, darsena) 09=istituto religioso

10=altra struttura collettiva
 11=stanza in affitto
 12=abitazione in affitto
 13=abitazione in proprietà o multiproprietà
 14=abitazione di parenti o amici gratuita
 15=barca in sito non organizzato
 16=altro tipo di sistemazione privata
 17=agriturismo

Coll. 96-98 DEST

DESTINAZIONE PRINCIPALE (REGIONE O PAESE ESTERO)

001=Piemonte
 002=Valle d'Aosta
 003=Lombardia
 004=Trentino-Alto Adige
 005=Veneto
 006=Friuli-Venezia Giulia
 007=Liguria
 008=Emilia-Romagna
 009=Toscana
 010=Umbria
 011=Marche
 012=Lazio
 013=Abruzzo
 014=Molise
 015=Campania
 016=Puglia
 017=Basilicata
 018=Calabria
 019=Sicilia
 020=Sardegna
 021=Austria
 022=Belgio
 023=Danimarca
 024=Finlandia
 025=Francia
 026=Germania
 027=Grecia
 028=Irlanda
 029=Islanda
 030=Lussemburgo
 031=Norvegia
 032=Olanda
 033=Polonia
 034=Portogallo
 035=Regno Unito
 036=Repubblica Ceca
 037=Repubblica Slovacca
 038=Spagna
 039=Svezia
 040=Svizzera
 041=Turchia
 042=Ungheria
 043=Altri Paesi Europei
 044=Stati Uniti d'America
 045=Canada
 046=America centro-meridionale
 047=Egitto
 048=Tunisia-Marocco-Algeria
 049=Altri Paesi africani
 050=Giappone
 051=Altri Paesi asiatici
 052=Australia
 053=Nuova Zelanda

054=Resto del mondo
 055=Liechtenstein
 056=Altri Paesi Oceanici
 057=Argentina
 058=Brasile
 059=Bulgaria
 060=Cina
 061=Cipro
 062=Corea del Sud
 063=Estonia
 064=Lettonia
 065=Lituania
 066=Malta
 067=Romania
 068=Russia
 069=Slovenia
 070=Sudafrica

Coll. 99-120	ORD1-ORD11	SEQUENZA DEGLI IDENTIFICATIVI INDIVIDUALI (NUMERO D'ORDINE ENTRO LA FAMIGLIA) DEI COMPONENTI PARTECIPANTI AL VIAGGIO (sequenza di 11 campi da 2 bytes, serrati a sinistra)
--------------	------------	--

Coll. 121-123	DEST_PR	PRINCIPALE PROVINCIA ITALIANA DI DESTINAZIONE <i>(se la destinazione principale è una regione italiana: DEST=001-020)</i> 001=Torino 002=Vercelli 003=Novara 004=Cuneo 005=Asti 006=Alessandria 007=Aosta 008=Imperia 009=Savona 010=Genova 011=La Spezia 012=Varese 013=Como 014=Sondrio 015=Milano 016=Bergamo 017=Brescia 018=Pavia 019=Cremona 020=Mantova 021=Bolzano 022=Trento 023=Verona 024=Vicenza 025=Belluno 026=Treviso 027=Venezia 028=Padova 029=Rovigo 030=Udine 031=Gorizia 032=Trieste 033=Piacenza 034=Parma 035=Reggio Emilia 036=Modena 037=Bologna 038=Ferrara 039=Ravenna
---------------	---------	--

040=Forlì-Cesena
041=Pesaro e Urbino
042=Ancona
043=Macerata
044=Ascoli Piceno
045=Massa-Carrara
046=Lucca
047=Pistoia
048=Firenze
049=Livorno
050=Pisa
051=Arezzo
052=Siena
053=Grosseto
054=Perugia
055=Terni
056=Viterbo
057=Rieti
058=Roma
059=Latina
060=Frosinone
061=Caserta
062=Benevento
063=Napoli
064=Avellino
065=Salerno
066=L'Aquila
067=Teramo
068=Pescara
069=Chieti
070=Campobasso
071=Foggia
072=Bari
073=Taranto
074=Brindisi
075=Lecce
076=Potenza
077=Matera
078=Cosenza
079=Catanzaro
080=Reggio di Calabria
081=Trapani
082=Palermo
083=Messina
084=Agrigento
085=Caltanissetta
086=Enna
087=Catania
088=Ragusa
089=Siracusa
090=Sassari
091=Nuoro
092=Cagliari
093=Pordenone
094=Isernia
095=Oristano
096=Biella
097=Lecco
098=Lodi
099=Rimini
100=Prato
101=Crotone
102=Vibo Valentia

103=Verbano-Cusio-Ossola
999=Non sa, non ricorda

Coll. 124-128	-	RISERVATO ISTAT
Coll. 129	ORGALL	ORGANIZZAZIONE DELL'ALLOGGIO 1=prenotazione diretta 2=prenotazione in agenzia 3=nessuna prenotazione 4=non sa
Coll. 130	IORGALL	PRENOTAZIONE DELL'ALLOGGIO TRAMITE INTERNET (se è stato prenotato l'alloggio: ORGALL=1, 2) 1=sì 2=no 3=non sa
Coll. 131	ORGTRA	ORGANIZZAZIONE DEL TRASPORTO 1=prenotazione diretta 2=prenotazione in agenzia 3=nessuna prenotazione 4=non sa
Coll. 132	IORGTRA	PRENOTAZIONE DEL TRASPORTO TRAMITE INTERNET (se è stato prenotato il trasporto: ORGTRA=1,2) 1=sì 2=no 3=non sa

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.

APPENDICE B - ALCUNE DEFINIZIONI

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

- **Nord-ovest:** Piemonte, Valle d'Aosta, Lombardia, Liguria
- **Nord-est:** Trentino-Alto Adige, Veneto, Friuli-Venezia Giulia, Emilia-Romagna
- **Centro:** Toscana, Umbria, Marche, Lazio
- **Sud:** Abruzzo, Molise, Campania, Puglia, Basilicata, Calabria
- **Isole:** 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 Telecom degli utenti 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 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 Sud);
- le regioni geografiche (solamente per alcune stime più importanti);
- sei aree basate sulla tipologia socio-demografica dei comuni, così definite:

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.

La suddivisione dei comuni nelle aree A e B è stata effettuata utilizzando i dati del Censimento della Popolazione⁵.

2. Caratteristiche della lista di selezione

La base di campionamento adottata, ossia la lista di selezione delle unità campionarie, è l'archivio informatizzato ufficiale delle famiglie incluse nell'archivio Telecom degli utenti 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

⁵ Cfr. il volume 'I sistemi locali del lavoro 1991', Istat, Argomenti n. 10 - 1997.

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.

⁶ Noti nella letteratura in lingua anglosassone sul tema come *Calibration Estimators* (Deville e Särndal "Calibration Estimators in Survey Sampling", 1992, Journal of the American Statistical Association)

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.

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 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 2004

REGIONI	Indirizzi campione	Tasso di sostituzione totale (a)	Tasso di sostituzione al netto dei mancati contatti (b)
Piemonte	265	21,1	11,7
Valle d'Aosta	92	31,5	23,9
Lombardia	330	19,4	13,0
Trentino-Alto Adige	206	16,0	7,8
Veneto	201	17,9	12,4
Friuli-Venezia Giulia	139	18,7	13,7
Liguria	152	30,3	17,8
Emilia-Romagna	197	22,3	11,7
Toscana	213	19,7	12,7
Umbria	103	21,4	11,7
Marche	141	16,3	11,3
Lazio	219	24,7	10,5
Abruzzo	144	23,6	16,0
Molise	96	21,9	15,6
Campania	196	26,0	18,9
Puglia	187	26,2	20,9
Basilicata	99	23,2	9,1
Calabria	158	28,5	18,4
Sicilia	213	26,3	17,4
Sardegna	150	21,3	10,0
Italia	3.501	22,5	13,9

- (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 2004

TIPOLOGIA DEL COMUNE	Indirizzi campione	Tasso di sostituzione totale	Tasso di sostituzione al netto dei mancati contatti
Comuni centro dell'area metropolitana	441	21,4	12,3
Comuni che gravitano intorno al centro dell'area metropolitana	316	25,6	15,8
Comuni aventi fino a 2.000 abitanti	309	20,6	14,1
Comuni con 2.001-10.000 abitanti	950	22,2	14,5
Comuni con 10.001-50.000 abitanti	868	21,5	12,6
Comuni con oltre 50.000 abitanti	617	28,1	16,5
Italia	3.501	22,5	13,9

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 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 sesso nelle 20 regioni italiane
 - popolazione residente per sesso nelle aggregazioni di comuni basate sulla tipologia socio-demografica
 - popolazione per ripartizione territoriale e per sesso
 - famiglie per ampiezza
 - famiglie unipersonali (single) di età inferiore a 65 anni, per sesso
 - famiglie unipersonali (single) di 65 anni e più, per sesso.

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:

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

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

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, ${}_d Y$, 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. 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, ${}_d Y$ 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:

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

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

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

dove

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

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:

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

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

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

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 ad 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 un 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:

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

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 $(\hat{\sigma}(\hat{Y}_t), \hat{Y}_t)$. In considerazione del fatto che il modello (14) è di tipo empirico, l'insieme delle stime \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 e 9 (per le stime relative alle notti e ai viaggi rispettivamente) 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 e 10 (per le stime relative alle notti e ai viaggi rispettivamente) 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: 5, 10, 15, 20, 25, 30 e 35%. 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

AREE TERRITORIALI	Trimestri 1,2,e 4			Trimestre 3		
	a	b	R^2	a	b	R^2
Nord	11,98947	-1,24131	92,7	11,98947	-1,24131	92,7
Centro	10,99685	-1,19553	92,1	10,99685	-1,19553	92,1
Sud	11,01204	-1,19591	91,7	11,01204	-1,19591	91,7
Italia	12,22875	-1,25171	93,2	12,22875	-1,25171	93,2

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

AREE TERRITORIALI	a	b	R^2
Nord	10,33272	-1,21709	93,9
Centro	10,21026	-1,23217	93,7
Sud	10,34692	-1,23609	93,8
Italia	10,96260	-1,25345	94,4

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

STIME	Trimestri 1,2 e 4				Trimestre 3			
	Aree territoriali				Aree territoriali			
	Nord	Centro	Sud	Italia	Nord	Centro	Sud	Italia
20.000	85,9	65,6	66,0	92,0	85,9	65,6	66,0	92,0
30.000	66,8	51,5	51,8	71,4	66,8	51,5	51,8	71,4
40.000	55,9	43,3	43,6	59,6	55,9	43,3	43,6	59,6
50.000	48,6	37,9	38,1	51,8	48,6	37,9	38,1	51,8
60.000	43,4	34,0	34,2	46,2	43,4	34,0	34,2	46,2
70.000	39,5	31,0	31,2	42,0	39,5	31,0	31,2	42,0
80.000	36,3	28,6	28,8	38,6	36,3	28,6	28,8	38,6
90.000	33,8	26,7	26,8	35,9	33,8	26,7	26,8	35,9
100.000	31,6	25,1	25,2	33,6	31,6	25,1	25,2	33,6
200.000	20,6	16,6	16,7	21,8	20,6	16,6	16,7	21,8
300.000	16,0	13,0	13,1	16,9	16,0	13,0	13,1	16,9
400.000	13,4	10,9	11,0	14,1	13,4	10,9	11,0	14,1
500.000	11,7	9,6	9,6	12,3	11,7	9,6	9,6	12,3
750.000	9,1	7,5	7,6	9,5	9,1	7,5	7,6	9,5
1.000.000	7,6	6,3	6,4	7,9	7,6	6,3	6,4	7,9
2.000.000	4,9	4,2	4,2	5,2	4,9	4,2	4,2	5,2
3.000.000	3,8	3,3	3,3	4,0	3,8	3,3	3,3	4,0
4.000.000	3,2	2,8	2,8	3,3	3,2	2,8	2,8	3,3
5.000.000	2,8	2,4	2,4	2,9	2,8	2,4	2,4	2,9
7.500.000	2,2	1,9	1,9	2,3	2,2	1,9	1,9	2,3
10.000.000	1,8	1,6	1,6	1,9	1,8	1,6	1,6	1,9
15.000.000	1,4	1,3	1,3	1,5	1,4	1,3	1,3	1,5
20.000.000	-	-	-	1,2	-	-	-	1,2
25.000.000	-	-	-	1,1	-	-	-	1,1

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

STIME	Aree territoriali			
	Nord	Centro	Sud	Italia
20.000	42,3	36,9	38,8	48,4
30.000	33,1	28,8	30,2	37,5
40.000	27,7	24,1	25,3	31,4
50.000	24,2	21,0	22,0	27,3
60.000	21,7	18,8	19,7	24,3
70.000	19,7	17,1	17,9	22,1
80.000	18,2	15,7	16,5	20,3
90.000	16,9	14,6	15,3	18,9
100.000	15,9	13,7	14,3	17,7
200.000	10,4	8,9	9,3	11,4
300.000	8,1	7,0	7,3	8,9
400.000	6,8	5,8	6,1	7,4
500.000	6,0	5,1	5,3	6,4
750.000	4,7	4,0	4,1	5,0
1.000.000	3,9	3,3	3,5	4,2
2.000.000	2,6	2,2	2,3	2,7
3.000.000	2,0	1,7	1,8	2,1
4.000.000	1,7	1,4	1,5	1,7
5.000.000	1,5	1,2	1,3	1,5
7.500.000	1,1	1,0	1,0	1,2
10.000.000	1,0	0,8	0,8	1,0
15.000.000	0,8	0,6	0,6	0,8
20.000.000	-	-	-	0,6
25.000.000	-	-	-	0,6

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, e aree territoriali - Stime del numero di notti

Notti – lavoro				
AREE TERRITORIALI	a	b	c	R ²
		Gennaio-marzo		
Nord	56.610,573	0,355707	-0,000000039924	91,3
Centro	36.020,288	0,529707	-0,000000132999	92,3
Sud	35.478,520	0,438933	-0,000000069206	81,8
Italia	114.976,526	0,240254	-0,000000013201	89,5
		Luglio-settembre		
Nord	59.930,134	0,309470	-0,000000036449	92,4
Centro	29.164,840	0,585171	-0,000000080152	81,2
Sud	54.446,833	0,431160	-0,000000033065	86,8
Italia	140.821,713	0,248141	-0,000000008982	84,1
		Anno		
Nord	142.578,221	0,162548	-0,000000003856	92,0
Centro	126.701,851	0,265236	-0,000000014651	80,3
Sud	174.251,455	0,199355	-0,000000004692	90,5
Italia	298.462,781	0,114603	-0,000000001141	90,7
Notti – vacanze				
AREE TERRITORIALI	a	b	c	R ²
		Gennaio-marzo		
Nord	204.616,342	0,131352	-0,000000001615	91,0
Centro	107.031,580	0,180941	-0,000000006677	90,9
Sud	80.685,135	0,200203	-0,000000005641	92,9
Italia	263.584,742	0,094622	-0,000000000662	92,2
		Luglio-settembre		
Nord	740.291,268	0,064186	-0,000000000156	83,9
Centro	364.370,055	0,123892	-0,000000000764	88,5
Sud	272.799,396	0,121618	-0,000000000579	91,6
Italia	863.944,270	0,052520	-0,000000000068	87,7
		Anno		
Nord	831.468,922	0,055544	-0,000000000084	90,1
Centro	404.000,378	0,101108	-0,000000000409	88,5
Sud	340.841,517	0,093681	-0,000000000302	90,7
Italia	1.012.842,373	0,043184	-0,000000000036	90,7
Notti - totali				
AREE TERRITORIALI	a	b	c	R ²
		Gennaio-marzo		
Nord	236.671,085	0,124428	-0,000000001332	92,2
Centro	135.565,604	0,175045	-0,000000005585	92,1
Sud	109.168,025	0,187065	-0,000000004577	93,2
Italia	311.981,899	0,089589	-0,000000000543	93,4
		Luglio-settembre		
Nord	748.251,993	0,062663	-0,000000000141	85,9
Centro	406.986,436	0,118852	-0,000000000685	89,3
Sud	325.941,193	0,118561	-0,000000000532	91,4
Italia	933.319,853	0,050818	-0,000000000061	89,0
		Anno		
Nord	854.225,70	0,053048	-0,000000000076	90,8
Centro	445.091,10	0,093595	-0,000000000351	88,7
Sud	440.276,02	0,085920	-0,000000000239	90,2
Italia	1.118.484,78	0,040281	-0,000000000030	91,2

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

Notti – lavoro							
AREE TERRITORIALI	5%	10%	15%	20%	25%	30%	35%
Gennaio-marzo							
Nord	7.838.079	6.619.024	5.414.317	4.234.881	3.104.432	2.077.753	1.264.388
Centro	3.680.436	3.312.663	2.946.869	2.583.839	2.224.811	1.871.824	1.528.393
Sud	5.709.740	5.000.001	4.294.364	3.595.099	2.906.404	2.236.730	1.604.548
Italia	14.992.904	11.389.155	7.934.546	4.846.423	2.605.070	1.456.011	941.121
Luglio-settembre							
Nord	7.342.620	6.020.040	4.723.251	3.476.337	2.335.575	1.418.739	841.634
Centro	6.731.022	6.112.676	5.495.545	4.880.080	4.266.978	3.657.376	3.053.246
Sud	11.668.740	10.177.243	8.692.697	7.219.184	5.764.564	4.345.655	3.002.911
Italia	22.748.133	17.393.817	12.210.014	7.460.863	3.857.388	2.013.368	1.245.684
Anno							
Nord	30.400.277	18.245.265	7.921.124	2.925.603	1.527.469	1.008.744	749.068
Centro	15.257.367	11.998.643	8.843.159	5.914.681	3.506.294	1.984.639	1.232.288
Sud	32.956.263	22.802.368	13.308.482	6.025.498	2.743.339	1.610.427	1.117.782
Italia	60.892.717	23.786.661	6.897.698	3.345.424	2.164.849	1.594.213	1.260.213
Notti – vacanze							
AREE TERRITORIALI	5%	10%	15%	20%	25%	30%	35%
Gennaio-marzo							
Nord	52.785.731	24.574.536	6.877.326	2.796.692	1.685.887	1.199.499	929.446
Centro	20.396.136	13.324.976	6.942.639	2.823.251	1.368.715	857.717	618.015
Sud	27.155.571	18.536.233	10.290.332	3.800.137	1.398.683	774.581	528.127
Italia	72.909.474	16.305.095	4.516.054	2.463.224	1.684.326	1.278.148	1.029.389
Luglio-settembre							
Nord	128.159.861	19.086.260	8.495.718	5.417.131	3.970.834	3.132.823	2.586.473
Centro	101.381.171	42.482.041	10.641.348	4.577.164	2.840.457	2.050.766	1.602.805
Sud	127.381.678	47.295.655	8.229.972	3.395.243	2.104.928	1.521.786	1.190.894
Italia	132.490.984	17.743.524	8.808.480	5.842.255	4.368.253	3.487.614	2.902.278
Anno							
Nord	138.220.410	18.087.772	8.735.166	5.736.815	4.268.041	3.397.356	2.821.486
Centro	132.514.761	32.825.873	7.759.799	4.018.535	2.693.467	2.022.848	1.618.892
Sud	152.129.340	24.728.590	5.867.485	3.177.186	2.171.322	1.648.041	1.327.681
Italia	98.213.903	17.631.799	9.452.309	6.449.336	4.893.180	3.941.684	3.299.871
Notti – totali							
AREE TERRITORIALI	5%	10%	15%	20%	25%	30%	35%
Gennaio-marzo							
Nord	58.876.400	25.342.682	6.826.803	2.975.627	1.848.491	1.334.487	1.042.782
Centro	23.426.144	15.050.072	7.655.293	3.175.576	1.614.428	1.036.865	756.587
Sud	30.719.735	20.200.925	10.392.176	3.670.931	1.558.055	931.478	657.852
Italia	80.107.121	16.230.928	4.944.634	2.787.442	1.932.257	1.477.097	1.195.060
Luglio-settembre							
Nord	130.172.980	18.713.778	8.451.710	5.418.050	3.982.172	3.146.794	2.600.760
Centro	106.091.659	41.744.780	10.596.324	4.819.251	3.054.512	2.227.930	1.751.622
Sud	133.559.788	47.753.748	8.998.230	3.902.830	2.455.401	1.787.063	1.403.796
Italia	130.573.173	18.549.943	9.356.305	6.240.315	4.679.053	3.742.103	3.117.588
Anno							
Nord	127.940.483	17.687.257	8.750.804	5.795.597	4.329.999	3.455.406	2.874.534
Centro	133.751.932	27.641.762	7.537.633	4.126.846	2.827.827	2.148.553	1.731.789
Sud	161.613.650	22.597.497	6.703.064	3.828.638	2.672.890	2.051.893	1.664.698
Italia	89.888.171	18.554.486	10.165.562	6.993.558	5.329.156	4.304.361	3.610.016

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, e aree territoriali - Stime del numero di viaggi

Viaggi – lavoro				
AREE TERRITORIALI	a	b	c	R ²
		Gennaio-marzo		
Nord	17.501,480	0,209960	-0,000000006797	94,8
Centro	488,442	0,615035	-0,000000050703	90,5
Sud	11.908,816	0,174784	-0,000000007700	98,0
Italia	25.271,382	0,184604	-0,000000002852	87,7
		Luglio-settembre		
Nord	12.853,908	0,239537	-0,000000006827	98,5
Centro	11.009,603	0,287293	-0,000000019362	98,0
Sud	10.677,990	0,166357	-0,000000006417	99,6
Italia	18.997,543	0,143293	-0,000000001911	98,8
		Anno		
Nord	29.258,148	0,105383	-0,000000000826	97,6
Centro	21.364,082	0,190765	-0,000000003654	88,4
Sud	29.633,350	0,109515	-0,000000001192	92,1
Italia	50.894,491	0,073040	-0,000000000269	94,8
Viaggi – vacanza				
AREE TERRITORIALI	a	b	c	R ²
		Gennaio-marzo		
Nord	30.778,361	0,136130	-0,000000005574	93,5
Centro	20.758,795	0,161750	-0,000000012595	94,2
Sud	22.152,957	0,153997	-0,000000007724	91,2
Italia	43.699,345	0,088948	-0,000000001562	93,3
		Luglio-settembre		
Nord	40.045,136	0,108828	-0,000000005710	84,5
Centro	24.400,289	0,159927	-0,000000018664	90,3
Sud	28.752,870	0,144146	-0,000000008541	91,2
Italia	50.114,151	0,082165	-0,000000001961	90,8
		Anno		
Nord	70.375,206	0,060080	-0,000000000661	93,2
Centro	40.603,145	0,089792	-0,000000002101	91,3
Sud	49.992,031	0,077006	-0,000000000980	91,7
Italia	99.822,759	0,040779	-0,000000000192	91,8
Viaggi – totali				
AREE TERRITORIALI	a	b	c	R ²
		Gennaio-marzo		
Nord	29.733,525	0,138305	-0,000000006261	91,2
Centro	21.399,383	0,227476	-0,000000023925	68,4
Sud	20.412,687	0,148773	-0,000000007809	91,3
Italia	45.031,267	0,093295	-0,000000001844	85,4
		Luglio-settembre		
Nord	39.998,987	0,100027	-0,000000005195	78,1
Centro	20.073,867	0,181881	-0,000000024825	88,6
Sud	23.185,257	0,146536	-0,000000009175	91,4
Italia	40.548,117	0,089728	-0,000000002384	89,8
		Anno		
Nord	44688,094	0,072830	-0,000000000911	88,1
Centro	31360,460	0,110335	-0,000000003097	79,2
Sud	35708,409	0,086162	-0,000000001192	84,5
Italia	62675,717	0,048979	-0,000000000260	85,0

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

Viaggi – lavoro							
AREE TERRITORIALI	5%	10%	15%	20%	25%	30%	35%
Gennaio-marzo							
Nord	29.507.289	23.643.750	16.335.992	9.104.708	2.496.748	408.740	191.603
Centro	11.933.809	11.144.952	10.158.894	9.172.853	8.186.838	7.200.857	6.214.927
Sud	21.472.673	16.300.683	9.868.964	3.643.214	418.729	155.842	94.556
Italia	61.358.112	47.377.292	29.956.753	12.822.661	1.319.078	380.134	217.824
Luglio-settembre							
Nord	27.829.934	20.530.270	13.256.867	6.099.787	805.320	207.719	115.539
Centro	12.301.895	9.731.707	7.170.185	4.631.264	2.186.187	494.226	166.964
Sud	18.224.472	10.499.625	3.087.994	300.204	126.435	79.595	58.028
Italia	49.017.718	23.083.360	1.853.513	331.312	177.470	121.051	91.828
Anno							
Nord	67.575.958	10.043.829	647.987	308.397	202.081	150.241	119.560
Centro	38.674.041	25.072.691	11.657.672	1.464.643	352.983	194.317	133.757
Sud	50.420.216	7.379.434	716.834	326.095	210.563	155.417	123.148
Italia	87.727.896	1.853.469	659.788	400.530	287.479	224.185	183.728
Viaggi – vacanze							
AREE TERRITORIALI	5%	10%	15%	20%	25%	30%	35%
Gennaio-marzo							
Nord	15.800.992	7.243.867	1.414.703	463.167	266.809	186.637	143.376
Centro	9.054.459	5.218.481	1.832.348	469.977	227.819	148.154	109.470
Sud	13.673.793	7.379.434	1.971.941	447.877	226.622	150.531	112.525
Italia	26.006.633	2.825.447	703.118	391.348	270.626	206.738	167.229
Luglio-settembre							
Nord	10.944.192	3.531.917	868.114	427.766	280.480	208.177	165.396
Centro	6.103.839	3.576.322	1.439.844	494.848	257.189	170.332	126.795
Sud	11.320.035	5.753.726	1.523.819	479.612	265.922	182.658	138.876
Italia	17.832.422	2.252.106	723.625	422.322	297.557	229.581	186.853
Anno							
Nord	20.464.996	1.714.262	778.190	501.777	370.075	293.091	242.606
Centro	19.913.525	2.593.488	659.221	365.872	252.604	192.786	155.845
Sud	29.290.951	2.003.061	678.690	405.150	288.509	223.965	183.005
Italia	9.103.376	1.676.497	912.489	626.471	476.908	384.978	322.755
Viaggi – totali							
AREE TERRITORIALI	5%	10%	15%	20%	25%	30%	35%
Gennaio-marzo							
Nord	14.433.868	6.815.235	1.436.977	460.429	262.344	182.595	139.876
Centro	7.536.815	5.491.128	3.494.311	1.680.645	585.696	270.863	169.073
Sud	12.851.282	6.639.094	1.540.090	376.828	198.609	134.052	101.045
Italia	24.475.094	3.447.404	774.619	418.983	286.398	217.431	175.200
Luglio-settembre							
Nord	10.372.500	2.777.415	743.016	392.107	264.288	198.993	159.484
Centro	5.460.435	3.527.515	1.747.060	605.528	268.430	164.275	117.369
Sud	10.756.352	5.528.957	1.412.018	405.446	219.805	149.738	113.373
Italia	17.627.324	2.498.480	655.741	364.832	252.050	192.417	155.570
Anno							
Nord	26.873.667	1.562.834	575.180	350.524	251.907	196.562	161.145
Centro	19.989.518	5.261.820	747.057	345.625	223.433	164.902	130.631
Sud	31.294.770	2.173.549	553.639	312.655	217.605	166.833	135.260
Italia	13.680.151	1.220.833	619.436	414.718	311.662	249.619	208.173

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:

$$\hat{\sigma}({}_c\hat{Y}_3) = 29.164,840 + 0,585171{}_c\hat{Y}_3 - 0,000000080152({}_c\hat{Y}_3)^2.$$

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 al 5, 10, 15, 20, 25, 30 e 35%.

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 17.743.524 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 numero di notti per lavoro riferita al totale Italia nel trimestre luglio-settembre 2004, pari a 14.113.000.

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 14.113.000 risulta essere compreso tra i valori 17.393.817 e 12.210.014.

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 2004 è stato pari a 10.804.000.

Si cerca il livello di stima che più si avvicina a 10.804.000 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 1,9%.

L'errore assoluto sarà:

$$\sigma(10.804.000) = 1,9/100 \times 10.804.000 = 205.276$$

L'intervallo di confidenza avrà come estremi:

$$10.804.000 - (2 \times 205.276) = 10.393.448$$

$$10.804.000 + (2 \times 205.276) = 11.214.552$$

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 1,9 e 1,5 percentuali.

L'errore corrispondente a 10.804.000 è pari a:

$$\sigma(10.804.000) = 1,9 - ((1,9 - 1,5) / (15.000.000 - 10.000.000)) \times (10.804.000 - 10.000.000) = 1,84\%$$

Il corrispondente errore assoluto è $1,84/100 \times 10.804.000 = 198.794$ e l'intervallo di confidenza avrà come estremi:

$$10.804.000 - (2 \times 198.794) = 10.406.413$$

$$10.804.000 + (2 \times 198.794) = 11.201.587$$

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 10.804.000 si ha:

$$\log \varepsilon^2 = 12,22875 - 1,25171 \times \log(10.804.000) = -8,04323$$

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

Infatti, da $\log(\varepsilon^2) = -8,04323$ si ha:

$$\varepsilon^2 = \exp(-8,04323)$$

e dunque:

$$\varepsilon = 0,017924$$

L'errore relativo percentuale sarà dunque pari a 1,79%.