

File Standard

***Indagine Multiscopo sulle Famiglie
Viaggi e vacanze
Anno 2002***

Manuale per l'utente

Per informazioni sull'acquisto rivolgersi a:
Istat - Servizio Biblioteca e servizi all'utenza
via Cesare Balbo, 16 – 00184 Roma
tel: 06.4673.5106
fax: 06.4673.5198
e-mail: diffdati@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: 06.5952.4655, 4673
fax: 06.5952.4596
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.....	18
APPENDICE B - ALCUNE DEFINIZIONI	19
APPENDICE C - STRATEGIA DI CAMPIONAMENTO ED ERRORI CAMPIONARI.....	20

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 Decreto Legislativo e della Legge n. 675 del 31/12/1996 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 **motivo del viaggio, destinazione, mezzo di trasporto e tipo di alloggio** 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: 39.192
- **un file dei viaggi** (tipo A o tipo B)
lunghezza record: 120 bytes
numero record: 15.433

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).

⁴ Il volume è consultabile anche sul sito Istat all'URL <http://www.istat.it>, selezionando 'Società' e l'area tematica 'Comportamenti sociali'.

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

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

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

Il **file viaggi** è composto da un record per ciascun viaggio realizzato da un individuo. 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. 32-120) 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).

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. 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 e ai viaggi.

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 tale 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 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 2002) 02=secondo trimestre (aprile-giugno 2002) 03=terzo trimestre (luglio-settembre 2002) 04=quarto trimestre (ottobre-dicembre 2002)
Coll. 5-8	ANNRIF	ANNO DI RIFERIMENTO = 2002
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
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 2002) 02=secondo trimestre (aprile-giugno 2002) 03=terzo trimestre (luglio-settembre 2002) 04=quarto trimestre (ottobre-dicembre 2002)
Coll. 5-8	ANNRIF	ANNO DI RIFERIMENTO = 2002
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
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 "L102"=viaggio di lavoro concluso nel primo trimestre 2002 "L202"=viaggio di lavoro concluso nel secondo trimestre 2002 "L302"=viaggio di lavoro concluso nel terzo trimestre 2002 "L402"=viaggio di lavoro concluso nel quarto trimestre 2002 "V102"=viaggio di vacanza concluso nel primo trimestre 2002 "V202"=viaggio di vacanza concluso nel secondo trimestre 2002 "V302"=viaggio di vacanza concluso nel terzo trimestre 2002 "V402"=viaggio di vacanza concluso nel quarto trimestre 2002
-------------	---------	---

Coll. 39-40	MOTLAV	MOTIVO DEL VIAGGIO DI LAVORO (se viaggio di lavoro: TIPOVIA="L102", "L202", "L302", "L402") 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
-------------	--------	--

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= "V102", "V202", "V302", "V402") 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	ORGAN	ORGANIZZAZIONE DEL VIAGGIO 01=prenotazione diretta solo alloggio 02=prenotazione diretta solo trasporto 03=prenotazione diretta alloggio e trasporto 04=prenotazione in agenzia solo alloggio 05=prenotazione in agenzia solo trasporto 06=prenotazione in agenzia alloggio e trasporto 07=nessuna prenotazione 08=non sa / non risponde

Coll. 92-93	MEZZO	<p>MEZZO DI TRASPORTO UTILIZZATO</p> <p>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</p>
Coll. 94-95	ALLOG	<p>TIPO DI ALLOGGIO PREVALENTE</p> <p>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</p>
Coll. 96-98	DEST	<p>DESTINAZIONE PRINCIPALE (REGIONE O PAESE ESTERO)</p> <p>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</p>

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

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)

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 ED 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 degli abbonati Telecom al telefono; 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 abbonate alla Telecom. 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.

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

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:

- sovracopertura, quando alcuni elementi dell'insieme L non appartengono all'insieme U;
- sottocopertura, nel caso in cui alcuni elementi di U non sono contenuti in L;
- duplicazione di alcune unità, se alcuni elementi di U sono presenti più volte in L;
- 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*

REGIONI	Indirizzi campione	Tasso di sostituzione totale	Tasso di sostituzione al netto dei mancati contatti
Piemonte	265	21,4	12,4
Valle d'Aosta	92	44,0	26,0
Lombardia	330	19,7	14,2
Trentino-Alto Adige	206	18,0	10,2
Veneto	201	14,9	10,0
Friuli-Venezia Giulia	139	16,5	6,5
Liguria	152	29,6	15,8
Emilia-Romagna	197	17,8	12,7
Toscana	213	23,9	13,1
Umbria	103	22,3	15,5
Marche	141	18,4	9,9
Lazio	219	23,7	17,4
Abruzzo	144	27,1	13,9
Molise	96	24,2	18,9
Campania	196	26,3	20,7
Puglia	187	21,9	16,6
Basilicata	99	22,0	12,0
Calabria	158	31,0	19,6
Sicilia	213	32,4	21,8
Sardegna	150	20,7	13,3
Italia	3.501	23,2	14,8

* dato della rilevazione del quarto trimestre

- (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

TIPOLOGIA DEL COMUNE	Indirizzi campione	Tasso di sostituzione totale	Tasso di sostituzione al netto dei mancati contatti
Comuni centro dell'area metropolitana	440	20,7	14,5
Comuni che gravitano intorno al centro dell'area metropolitana	316	23,1	16,5
Comuni aventi fino a 2.000 abitanti	618	35,3	19,2
Comuni con 2.001-10.000 abitanti	868	24,4	15,0
Comuni con 10.001-50.000 abitanti	950	22,4	15,1
Comuni con oltre 50.000 abitanti	309	17,9	11,2
Italia	3.501	23,2	14,8

* dato della rilevazione del quarto trimestre

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 p ;
- ogni indirizzo selezionato dall'Archivio entra a far parte dell'*elenco base* e per ognuno di essi si selezionano i tre indirizzi successivi, che entrano a far parte dell'*elenco sostitutivo*; in tal modo le unità sostitutive sono le più vicine possibile dal punto di vista territoriale alle rispettive unità del campione base e quindi, presumibilmente, sono anche *simili* dal punto di vista delle caratteristiche socio-economiche.

4. Procedimento per il calcolo delle stime

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

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

Il principio su cui è basato ogni metodo di stima campionaria è che le unità appartenenti al campione rappresentino anche le unità della popolazione che non sono incluse nel campione.

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

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

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

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

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

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

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

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

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

- correggere l'effetto distorsivo dovuto agli errori di lista e al fenomeno della mancata risposta totale;
- tenere conto della conoscenza di alcuni totali noti sulla popolazione oggetto di studio, nel senso che le stime campionarie di tali totali devono coincidere con i rispettivi valori noti. Nell'indagine in oggetto i totali noti⁷ sono:
 - popolazione residente per sesso nelle 20 regioni italiane

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".

- 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:

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Gli errori campionari consentono di valutare il grado di precisione delle stime; inoltre, l'errore assoluto permette di costruire l'intervallo di confidenza che con una certa fiducia P contiene il

parametro di interesse. Con riferimento ad una generica stima \hat{Y} tale intervallo assume la seguente forma:

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

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

6. Presentazione sintetica degli errori campionari

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

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

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

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

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

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

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

I prospetti 3 (per i trimestri) e 4 (per la media dei 4 trimestri) riportano i valori dei coefficienti a e b e dell'indice di determinazione R^2 delle funzioni utilizzate per l'interpolazione degli errori campionari delle stime di frequenze riferite agli individui, per area territoriale.

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

La prima colonna dei prospetti riporta K valori crescenti delle stime ${}_d\hat{Y}_t^k$ ($k=1, \dots, K$), la seconda riporta i rispettivi errori relativi interpolati $\hat{\varepsilon}({}_d\hat{Y}_t^k)$. Le informazioni contenute nei prospetti permettono di calcolare l'errore relativo di 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 $\hat{\varepsilon}(\hat{Y}_t)$ 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 fuori casa per viaggi di lavoro;
- numero di notti fuori casa per viaggi di vacanza;
- numero totale di notti fuori casa;
- 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, ${}_d\hat{Y}_t$, 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 e 10 (per le stime relative alle notti e ai viaggi rispettivamente) 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: 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	11,53557	-1,21828	92,4	11,53792	-1,21067	92,2
Centro	11,01043	-1,20084	93,5	10,19051	-1,13059	90,4
Sud	10,80231	-1,18807	92,1	11,11226	-1,19508	91,3
Italia	11,78546	-1,22896	93,5	11,74898	-1,21707	92,5

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	9,88711	-1,19273	92,5
Centro	10,08666	-1,22747	93,5
Sud	10,20464	-1,23108	93,1
Italia	10,28524	-1,21304	93,2

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				Aeree territoriali			
	Nord	Centro	Sud	Italia	Nord	Centro	Sud	Italia
20.000	76,7	64,3	61,8	82,5	79,8	60,5	69,7	85,9
30.000	59,9	50,4	48,5	64,3	62,4	48,1	54,7	67,1
40.000	50,3	42,4	40,9	53,9	52,4	40,9	46,0	56,3
50.000	43,9	37,1	35,8	47,0	45,8	36,0	40,3	49,2
60.000	39,3	33,3	32,2	42,0	41,0	32,5	36,1	44,0
70.000	35,8	30,3	29,3	38,2	37,4	29,8	33,0	40,1
80.000	33,0	28,0	27,1	35,2	34,5	27,6	30,4	36,9
90.000	30,7	26,1	25,3	32,7	32,1	25,8	28,4	34,4
100.000	28,8	24,5	23,7	30,7	30,1	24,3	26,6	32,3
200.000	18,9	16,1	15,7	20,0	19,8	16,5	17,6	21,2
300.000	14,7	12,7	12,4	15,6	15,5	13,1	13,8	16,5
400.000	12,4	10,6	10,4	13,1	13,0	11,1	11,6	13,9
500.000	10,8	9,3	9,1	11,4	11,4	9,8	10,2	12,1
750.000	8,4	7,3	7,2	8,9	8,9	7,8	8,0	9,5
1.000.000	7,1	6,1	6,0	7,5	7,5	6,6	6,7	7,9
2.000.000	4,6	4,1	4,0	4,9	4,9	4,5	4,4	5,2
3.000.000	3,6	3,2	3,1	3,8	3,8	3,6	3,5	4,1
4.000.000	3,0	2,7	2,7	3,2	3,2	3,0	2,9	3,4
5.000.000	2,7	2,3	2,3	2,8	2,8	2,7	2,6	3,0
7.500.000	2,1	1,8	1,8	2,2	2,2	2,1	2,0	2,3
10.000.000	1,7	1,5	1,5	1,8	1,9	1,8	1,7	2,0
15.000.000	1,4	1,2	1,2	1,4	1,5	1,4	1,3	1,5
20.000.000	1,1	1,0	1,0	1,2	1,2	1,2	1,1	1,3
25.000.000	1,0	0,9	0,9	1,0	1,1	1,1	1,0	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	38,2	35,5	37,0	42,1
30.000	30,0	27,7	28,8	33,0
40.000	25,3	23,2	24,2	27,7
50.000	22,1	20,2	21,1	24,2
60.000	19,8	18,1	18,8	21,6
70.000	18,1	16,5	17,1	19,7
80.000	16,7	15,2	15,8	18,2
90.000	15,6	14,1	14,7	16,9
100.000	14,6	13,2	13,7	15,9
200.000	9,7	8,6	9,0	10,4
300.000	7,6	6,7	7,0	8,2
400.000	6,4	5,7	5,9	6,8
500.000	5,6	4,9	5,1	6,0
750.000	4,4	3,8	4,0	4,7
1.000.000	3,7	3,2	3,3	3,9
2.000.000	2,5	2,1	2,2	2,6
3.000.000	1,9	1,6	1,7	2,0
4.000.000	1,6	1,4	1,4	1,7
5.000.000	1,4	1,2	1,2	1,5
7.500.000	1,1	0,9	1,0	1,2
10.000.000	0,9	0,8	0,8	1,0
15.000.000	0,7	0,6	0,6	0,8
20.000.000	0,6	0,5	0,5	0,6
25.000.000	0,5	0,4	0,5	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	31.132,133	0,381388	-0,000000031088	97,9
Centro	17.012,517	0,523357	-0,000000087283	89,3
Sud	41.205,239	0,518130	-0,000000056503	79,2
Italia	123.240,591	0,259609	-0,000000009906	87,3
		Luglio-settembre		
Nord	65.696,624	0,388913	-0,000000027059	92,0
Centro	63.588,941	0,368091	-0,000000072671	89,1
Sud	13.012,295	0,466576	-0,000000026508	92,5
Italia	114.389,705	0,261591	-0,000000007471	89,6
		Anno		
Nord	152.844,021	0,190819	-0,000000004547	92,0
Centro	97.628,082	0,215924	-0,000000011802	89,9
Sud	183.985,900	0,301631	-0,000000005750	88,0
Italia	398.030,665	0,145967	-0,000000001258	84,0
Notti -vacanze				
AREE TERRITORIALI	a	b	c	R ²
		Gennaio-marzo		
Nord	167.854,829	0,133126	-0,000000002061	89,0
Centro	93.944,877	0,269016	-0,000000008080	75,5
Sud	151.820,635	0,197487	-0,000000003272	88,1
Italia	310.961,156	0,106110	-0,000000000751	84,6
		Luglio-settembre		
Nord	756.034,530	0,060254	-0,000000000124	86,9
Centro	374.340,998	0,093669	-0,000000000493	89,6
Sud	367.413,584	0,111569	-0,000000000402	92,9
Italia	929.730,593	0,047108	-0,000000000050	90,0
		Anno		
Nord	846.036,264	0,052353	-0,000000000074	89,7
Centro	428.312,093	0,081880	-0,000000000304	89,5
Sud	511.720,307	0,095416	-0,000000000270	89,7
Italia	1.140.490,339	0,040426	-0,000000000030	90,3
Notti - totali				
AREE TERRITORIALI	a	b	c	R ²
		Gennaio-marzo		
Nord	196.807,919	0,127285	-0,000000001705	91,7
Centro	136.201,633	0,240119	-0,000000006296	78,6
Sud	196.890,437	0,191565	-0,000000003265	88,5
Italia	401.348,855	0,096169	-0,000000000568	87,7
		Luglio-settembre		
Nord	775.892,298	0,059891	-0,000000000118	88,7
Centro	379.206,792	0,090900	-0,000000000447	90,7
Sud	418.004,034	0,109955	-0,000000000388	93,4
Italia	982.111,966	0,046448	-0,000000000047	91,3
		Anno		
Nord	899.839,664	0,050580	-0,000000000068	90,5
Centro	456.474,863	0,078189	-0,000000000273	90,0
Sud	730.519,545	0,091015	-0,000000000234	89,2
Italia	1.375.891,138	0,038612	-0,000000000027	90,9

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	10.752.756	9.160.613	7.575.164	6.001.499	4.451.281	2.956.676	1.625.657
Centro	5.458.925	4.890.231	4.322.618	3.756.566	3.192.880	2.633.014	2.079.855
Sud	8.372.097	7.497.357	6.625.259	5.756.958	4.894.380	4.040.946	3.203.239
Italia	21.731.847	16.850.338	12.093.401	7.644.698	4.045.322	2.035.256	1.204.432
Luglio-settembre							
Nord	12.716.063	10.900.036	9.096.366	7.313.591	5.569.697	3.907.321	2.435.143
Centro	4.568.670	3.912.744	3.268.769	2.643.991	2.051.532	1.514.674	1.068.144
Sud	15.746.361	13.864.360	11.983.684	10.105.067	8.229.903	6.361.192	4.506.714
Italia	28.851.550	22.314.633	15.899.122	9.805.241	4.764.732	2.111.193	1.176.829
Anno							
Nord	32.020.648	21.535.147	11.821.196	4.875.564	2.208.079	1.326.627	935.210
Centro	14.624.028	10.602.202	6.801.744	3.628.746	1.774.450	1.016.221	686.651
Sud	44.477.604	35.953.244	27.530.585	19.328.807	11.710.703	5.800.019	2.842.934
Italia	80.229.524	43.768.754	16.256.694	6.409.834	3.663.682	2.531.708	1.927.895
Notti – vacanze							
AREE TERRITORIALI	5%	10%	15%	20%	25%	30%	35%
Gennaio-marzo							
Nord	42.253.195	20.117.621	5.815.671	2.341.066	1.401.552	993.678	768.361
Centro	27.528.280	21.459.598	15.480.727	9.735.789	4.783.860	1.994.566	1.050.032
Sud	46.083.394	31.278.484	17.209.714	6.438.643	2.501.293	1.416.913	975.065
Italia	79.859.294	24.811.665	6.386.691	3.228.551	2.137.248	1.593.955	1.270.036
Luglio-settembre							
Nord	129.848.683	18.011.198	8.328.482	5.384.383	3.974.154	3.148.365	2.606.400
Centro	96.401.609	21.868.350	6.298.057	3.464.837	2.376.719	1.806.474	1.456.301
Sud	159.053.757	47.903.312	8.758.647	4.079.231	2.633.999	1.941.820	1.536.985
Italia	110.775.519	17.296.906	8.996.888	6.068.985	4.577.255	3.673.738	3.067.965
Anno							
Nord	123.814.904	17.290.924	8.607.966	5.713.738	4.273.694	3.412.815	2.840.408
Centro	117.056.093	18.129.867	6.120.614	3.592.885	2.536.037	1.958.312	1.594.583
Sud	178.682.468	35.854.736	8.976.058	4.832.569	3.291.367	2.493.062	2.005.754
Italia	92.236.197	18.961.565	10.378.685	7.137.448	5.437.685	4.391.458	3.682.742
Notti – totali							
AREE TERRITORIALI	5%	10%	15%	20%	25%	30%	35%
Gennaio-marzo							
Nord	47.748.506	21.398.540	5.980.152	2.553.681	1.569.560	1.126.961	877.780
Centro	30.898.381	23.189.207	15.692.916	8.824.091	3.932.209	1.896.412	1.162.150
Sud	44.713.017	30.054.851	16.407.880	6.580.828	2.899.652	1.726.048	1.212.429
Italia	89.156.982	23.416.831	6.946.323	3.786.917	2.584.353	1.958.336	1.575.609
Luglio-settembre							
Nord	133.374.878	18.355.181	8.515.790	5.512.225	4.071.029	3.226.308	2.671.586
Centro	99.931.302	20.670.859	6.131.832	3.427.615	2.367.693	1.806.540	1.459.876
Sud	161.062.188	48.038.949	9.553.289	4.552.777	2.960.482	2.189.704	1.736.480
Italia	111.855.972	18.054.931	9.443.994	6.383.558	4.819.540	3.870.655	3.233.791
Anno							
Nord	119.024.509	17.770.883	8.995.206	6.005.696	4.505.316	3.604.162	3.003.213
Centro	117.580.372	17.220.042	6.210.093	3.716.460	2.645.723	2.052.759	1.676.560
Sud	191.405.152	39.869.042	11.829.253	6.609.084	4.564.216	3.481.977	2.813.547
Italia	98.134.270	22.197.711	12.315.710	8.513.315	6.503.475	5.260.948	4.416.894

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	12.683,725	0,153990	-0,000000005265	98,3
Centro	11.190,704	0,222341	-0,000000016301	99,0
Sud	9.517,533	0,154712	-0,000000006967	97,7
Italia	17.020,042	0,103914	-0,000000001596	98,6
		Luglio-settembre		
Nord	16.139,457	0,225292	-0,000000006537	97,8
Centro	11.202,182	0,258390	-0,000000018686	96,1
Sud	11.262,601	0,182058	-0,000000007552	98,5
Italia	22.331,070	0,139208	-0,000000001928	97,7
		Anno		
Nord	30.685,568	0,096568	-0,000000000782	97,3
Centro	20.161,685	0,126000	-0,000000002416	98,2
Sud	21.380,589	0,108227	-0,000000001200	95,5
Italia	42.574,974	0,063712	-0,000000000240	97,5
Viaggi – vacanza				
AREE TERRITORIALI	a	b	c	R ²
		Gennaio-marzo		
Nord	29.059,916	0,106433	-0,000000004334	92,2
Centro	17.562,483	0,200399	-0,000000019436	89,2
Sud	19.633,705	0,147202	-0,000000007370	92,3
Italia	45.598,272	0,078657	-0,000000001435	91,3
		Luglio-settembre		
Nord	44.460,598	0,091151	-0,000000004365	91,2
Centro	20.192,487	0,128349	-0,000000014271	95,2
Sud	26.399,215	0,111132	-0,000000006284	94,1
Italia	57.242,150	0,063185	-0,000000001379	93,1
		Anno		
Nord	78.435,920	0,047511	-0,000000000503	92,7
Centro	43.288,386	0,076778	-0,000000001898	92,6
Sud	45.903,801	0,064500	-0,000000000814	93,7
Italia	108.407,073	0,033155	-0,000000000154	92,6
Viaggi – totali				
AREE TERRITORIALI	a	b	c	R ²
		Gennaio-marzo		
Nord	25.020,033	0,107541	-0,000000004715	92,0
Centro	16.874,410	0,198138	-0,000000020807	89,1
Sud	16.519,604	0,141714	-0,000000007340	92,4
Italia	36.988,618	0,078600	-0,000000001523	90,9
		Luglio-settembre		
Nord	44.375,001	0,087434	-0,000000004531	80,4
Centro	18.690,667	0,146370	-0,000000019458	91,6
Sud	23.301,731	0,114665	-0,000000006961	94,1
Italia	45.825,916	0,073983	-0,000000001934	91,3
		Anno		
Nord	45.600,749	0,059783	-0,000000000742	87,0
Centro	29.528,791	0,089968	-0,000000002492	86,6
Sud	31.622,961	0,073100	-0,000000000995	87,7
Italia	60.922,000	0,040536	-0,000000000213	86,6

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	19.871.381	10.483.761	1.976.562	267.486	131.165	86.598	64.598
Centro	10.637.058	7.595.569	4.587.501	1.760.478	337.471	139.986	86.701
Sud	15.119.927	8.023.188	1.554.855	203.766	99.162	65.303	48.651
Italia	34.099.259	4.714.950	364.705	176.615	116.359	86.738	69.132
Luglio-settembre							
Nord	26.905.683	19.293.539	11.727.741	4.426.604	567.887	212.098	128.552
Centro	11.205.700	8.546.539	5.902.160	3.306.112	1.030.646	242.752	119.374
Sud	17.570.335	11.000.665	4.570.990	515.759	162.822	94.916	66.862
Italia	46.516.251	20.889.491	1.607.545	363.153	200.857	138.651	105.836
Anno							
Nord	60.194.011	4.442.687	569.540	296.010	199.791	150.752	121.035
Centro	31.724.870	11.489.763	779.003	270.076	162.083	115.686	89.920
Sud	48.893.252	8.866.236	504.512	232.266	150.616	111.411	88.394
Italia	60.150.263	1.164.288	492.729	312.218	228.476	180.149	148.695
Viaggi – vacanze							
AREE TERRITORIALI	5%	10%	15%	20%	25%	30%	35%
Gennaio-marzo							
Nord	13.518.033	3.436.005	627.811	306.235	201.192	149.627	119.058
Centro	7.853.219	5.334.982	2.904.198	960.893	315.154	170.645	115.657
Sud	13.387.775	6.796.522	1.453.340	354.337	188.447	127.708	96.476
Italia	21.445.258	1.894.974	631.131	374.125	265.533	205.733	167.898
Luglio-settembre							
Nord	10.407.453	2.335.100	717.345	401.984	277.773	211.946	171.268
Centro	5.736.760	2.542.911	652.233	267.559	162.875	116.508	90.572
Sud	10.141.720	3.118.446	617.545	291.071	188.495	139.132	110.199
Italia	12.802.533	1.473.521	652.592	416.641	305.721	241.377	199.387
Anno							
Nord	10.253.869	1.473.518	762.457	513.501	386.987	310.459	259.190
Centro	15.575.407	1.643.420	582.405	349.425	249.221	193.607	158.263
Sud	20.546.224	1.256.819	534.167	338.086	247.191	194.789	160.710
Italia	6.094.970	1.615.737	926.650	649.357	499.751	406.159	342.088
Viaggi – totali							
AREE TERRITORIALI	5%	10%	15%	20%	25%	30%	35%
Gennaio-marzo							
Nord	12.623.179	3.237.981	555.063	266.973	174.621	129.591	102.987
Centro	7.231.689	4.882.628	2.622.733	856.912	291.321	160.404	109.475
Sud	12.672.947	6.054.947	1.038.454	273.972	151.010	103.865	79.092
Italia	19.995.770	1.556.124	512.447	303.528	215.391	166.875	136.184
Luglio-settembre							
Nord	9.312.832	2.036.211	676.138	388.147	270.920	207.837	168.515
Centro	5.139.628	2.734.377	891.232	312.972	174.633	119.841	90.996
Sud	9.637.053	3.164.542	590.707	267.235	170.679	125.139	98.727
Italia	14.083.224	1.576.614	593.866	361.642	259.609	202.404	165.833
Anno							
Nord	16.828.482	1.111.090	503.373	324.658	239.507	189.721	157.063
Centro	16.747.510	1.974.835	482.235	266.755	183.991	140.358	113.435
Sud	2.450.7143	1.128.446	409.054	248.710	178.582	139.284	114.157
Italia	5.705.600	1.020.798	555.950	381.849	290.762	234.755	196.837

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:

$$\hat{\sigma}({}_C\hat{Y}_3) = 63.588,941 + 0,368091{}_C\hat{Y}_3 - 0,000000072671({}_C\hat{Y}_3)^2.$$

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 17.296.906 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, pari a 18.127.000.

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 18.127.000 risulta essere compreso tra i valori 15.899.122 e 22.314.633.

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 2002 è stato pari a 10.792.000.

Si cerca il livello di stima che più si avvicina a 10.792.000 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 1,8%.

L'errore assoluto sarà:

$$\sigma(10.792.000) = 1,8/100 \times 10.792.000 = 194.256$$

L'intervallo di confidenza avrà come estremi:

$$10.792.000 - (2 \times 194.256) = 10.403.488$$

$$10.792.000 + (2 \times 194.256) = 11.180.512$$

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 1,8 e 1,4 percentuali.

L'errore corrispondente a 10.792.000 è pari a:

$$\sigma(10.792.000) = 1,8 - ((1,8 - 1,4) / (15.000.000 - 10.000.000)) \times (10.792.000 - 10.000.000) = 1,74\%$$

Il corrispondente errore assoluto è $1,74/100 \times 10.792.000 = 187.781$ e l'intervallo di confidenza avrà come estremi:

$$10.792.000 - (2 \times 187.781) = 10.416.438$$

$$10.792.000 + (2 \times 187.781) = 11.167.562$$

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 10.792.000 si ha:

$$\log \varepsilon^2 = 11,78546 - 1,22896 \times \log(10.792.000) = -8,11671.$$

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

Infatti, da $\log(\varepsilon^2) = -8,11671$ si ha:

$$\varepsilon^2 = \exp(-8,11671)$$

e dunque:

$$\varepsilon = 0,017277$$

L'errore relativo percentuale sarà dunque pari a 1,73 %.