

File Standard

Indagine Multiscopo sulle Famiglie Viaggi e vacanze Anno 2001

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, 4616
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. TRACCIATO DEL FILE INDIVIDUI	8
5. TRACCIATO DEL FILE VIAGGI	12
APPENDICE A - IL QUESTIONARIO	17
APPENDICE B - ALCUNE DEFINIZIONI.....	18
APPENDICE C - STRATEGIA DI CAMPIONAMENTO ED ERRORI CAMPIONARI	19

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 files 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'

Il crescente interesse sul turismo, sia in ambito nazionale sia internazionale, ha reso necessario integrare il Sistema di informazioni statistiche sul turismo in Italia attraverso la raccolta sistematica di dati dal lato della 'domanda turistica'. Ciò in conformità a quanto previsto dalla Direttiva 95/57/CE, attuata dal Consiglio dell'Unione Europea il 23/11/1995 e finalizzata alla raccolta e diffusione dei dati statistici armonizzati sull'offerta e sulla domanda nel settore del turismo nell'ambito dell'Unione Europea.

In tale ottica, 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) con 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

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.).

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.

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 files di dati elementari:

- **un file degli individui** (tipo A o tipo B)
lunghezza record: 107 bytes
numero records: 39.410
- **un file dei viaggi** (tipo A o tipo B)
lunghezza record: 120 bytes
numero records: 14.686

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).

⁴ Il volume è consultabile anche sul sito ISTAT all'URL <http://www.istat.it>, nell'area tematica 'Famiglia e Società'.

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 di Regione (REG) e di Ripartizione geografica (RIP);
- la seconda (**tipo B**) con i codici di Tipologia socio-demografica del Comune (DOM) e di Ripartizione geografica (RIP).

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

I files sono riferiti all'intero anno e sono ottenuti mettendo in sequenza i quattro files 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 gli 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 files 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 files 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 files 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. 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 2001) 02=secondo trimestre (aprile-giugno 2001) 03=terzo trimestre (luglio-settembre 2001) 04=quarto trimestre (ottobre-dicembre 2001)
Coll. 5-8	ANNRIF	ANNO DI RIFERIMENTO = 2001
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 (<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 99=non indica

5. 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 2001) 02=secondo trimestre (aprile-giugno 2001) 03=terzo trimestre (luglio-settembre 2001) 04=quarto trimestre (ottobre-dicembre 2001)
Coll. 5-8	ANNRIF	ANNO DI RIFERIMENTO = 2001
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 <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

INFORMAZIONI SUL VIAGGIO

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

Coll. 35-38 TIPOVIA TIPOLOGIA DEL VIAGGIO
 "L101"=viaggio di lavoro concluso nel primo trimestre 2001
 "L201"=viaggio di lavoro concluso nel secondo trimestre 2001
 "L301"=viaggio di lavoro concluso nel terzo trimestre 2001
 "L401"=viaggio di lavoro concluso nel quarto trimestre 2001
 "V101"=viaggio di vacanza concluso nel primo trimestre 2001
 "V201"=viaggio di vacanza concluso nel secondo trimestre 2001
 "V301"=viaggio di vacanza concluso nel terzo trimestre 2001
 "V401"=viaggio di vacanza concluso nel quarto trimestre 2001

Coll. 39-40 MOTLAV MOTIVO DEL VIAGGIO DI LAVORO
 (se *viaggio di lavoro*: TIPOVIA="L101", "L201", "L301", "L401")
 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	MOTIVO	MOTIVO DEL VIAGGIO DI VACANZA (se <i>viaggio di vacanza</i> : TIPOVIA= "V101", "V201", "V301", "V401") 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 <i>il motivo della vacanza è piacere/svago</i> : MOTIVO=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	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

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 e Liechtenstein
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

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 o alla caratteristica sociale-economica:

- **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. Si precisa inoltre il significato dei seguenti termini:
- **occupato:** 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:** 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:** 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:** chi si dedica prevalentemente alle cure della propria famiglia o della propria casa;
- **studente:** chi si dedica prevalentemente allo studio;
- **ritirato dal lavoro:** 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:** 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 indagine – 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:
 - comuni centro dell'area metropolitana: Torino, Milano, Venezia, Genova, Bologna, Firenze, Roma, Napoli, Bari, Palermo, Catania e Cagliari;
 - comuni che gravitano intorno al centro dell'area metropolitana;
 - B) *area non metropolitana* suddivisa in:
 - comuni aventi fino a 2.000 abitanti;
 - comuni con 2.001-10.000 abitanti;
 - comuni con 10.001-50.000 abitanti;
 - comuni con oltre 50.000 abitanti.

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

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 Istat, "I sistemi locali del lavoro: 1991", *Argomenti* 10/1997.

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 della duplicazione le famiglie intestatarie di più di un numero riferito all'abitazione principale e nel caso dei grappoli i numeri telefonici a cui corrispondono più famiglie.

L'effetto di ciascuna delle situazioni appena elencate è quella 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. E' 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 gennaio-marzo.

3 - Disegno di campionamento

Il disegno di campionamento è un 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. E' 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. E' necessario quindi ricorrere a un procedimento complesso articolato in più fasi.

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)

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 3500. 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 riportate, 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 (dato del quarto trimestre).

REGIONI	INDIRIZZI CAMPIONE	TASSO DI SOSTITUZIONE TOTALE	TASSO DI SOSTITUZIONE AL NETTO DEI MANCATI CONTATTI
Piemonte	265	22,3	15,5
Valle d'Aosta	92	44,0	28,6
Lombardia	330	24,2	18,7
Trentino-Alto Adige	206	24,8	15,5
Veneto	201	24,9	15,9
Friuli-Venezia Giulia	139	28,8	21,6
Liguria	152	32,2	23,0
Emilia-Romagna	197	22,3	17,8
Toscana	213	25,8	20,2
Umbria	103	28,2	18,4
Marche	141	21,3	15,6
Lazio	219	32,9	22,8
Abruzzo	144	29,0	20,0
Molise	96	33,3	28,1
Campania	196	26,4	20,3
Puglia	187	23,0	16,6
Basilicata	99	35,4	28,3
Calabria	158	34,2	26,6
Sicilia	213	34,1	26,5
Sardegna	150	21,9	17,9
ITALIA	3501	27,5	20,2

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

REGIONI	INDIRIZZI CAMPIONE	TASSO DI SOSTITUZIONE TOTALE	TASSO DI SOSTITUZIONE AL NETTO DEI MANCATI CONTATTI
Comuni centro dell'area metropolitana	440	25,0	17,5
Comuni che gravitano intorno al centro dell'area metropolitana	316	24,9	19,9
Comuni aventi fino a 2.000 abitanti	618	35,8	24,8
Comuni con 2.001-10.000 abitanti	868	31,6	22,8
Comuni con 10.001-50.000 abitanti	950	25,3	18,7
Comuni con oltre 50.000 abitanti	309	23,0	18,0

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:

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

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

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

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

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

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

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

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".

- famiglie unipersonali (single) di età inferiore a 65 anni, per sesso
- famiglie unipersonali (single) di 65 anni e più, per sesso.

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

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

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) viene calcolato poi il fattore correttivo per mancata risposta totale, ottenuto come l'inverso del tasso di risposta in ciascuna regione;
- 3) 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;
- 4) il peso finale è dato dal prodotto del peso base per i fattori correttivi sopra indicati.

Il fattore correttivo del punto 3) 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 turisti, 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) in parola - si riferiscono, invece, al numero medio annuo di turisti, di viaggi e di notti. 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:

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

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 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 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:

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

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

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

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

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

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

6 - Presentazione sintetica degli errori campionari

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

delle stime fornite. Inoltre, non sarebbero comunque disponibili gli errori delle stime non pubblicate, che l'utente può ricavare in modo autonomo.

Al fine di permettere comunque una valutazione della variabilità campionaria di tutte le stime d'interesse, si ricorre a una presentazione sintetica degli errori relativi basata su modelli regressivi; ossia fondata sulla determinazione di una funzione matematica che mette in relazione ciascuna stima con il proprio errore di campionamento. L'approccio utilizzato per la costruzione dei modelli è differente a seconda che la variabile oggetto di stima sia qualitativa ovvero quantitativa. Infatti, mentre per le stime di frequenze assolute (o relative) riferite alle modalità di variabili qualitative, è possibile utilizzare modelli che hanno un fondamento teorico, secondo cui gli errori relativi delle stime di frequenze assolute sono funzione decrescente dei valori delle stime stesse; per le stime di totali di variabili quantitative, invece, 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. E' stato inoltre stimato un modello analogo al (12) con riferimento all'anno.

I prospetti 3.1 (per i trimestri) e 3.2 (per l'anno) 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 4.1 (per i trimestri) e 4.2 (per l'anno) 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 primo 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 (4.1 o 4.2) 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 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 difformi nel terzo trimestre rispetto agli altri tre.

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

I prospetti 5. e 7 (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 6 e 8 (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 3.1 - 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	10,90801	-1,17561	90,9	11,12555	-1,18390	91,9
Centro	11,34863	-1,23061	93,8	10,31165	-1,14465	91,2
Sud	11,55579	-1,24263	93,6	11,24473	-1,20493	92,3
ITALIA	12,15213	-1,25639	94,3	11,52058	-1,20456	92,5

Prospetto 3.2 - 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 – Anno

AREE TERRITORIALI	a	b	R^2
Nord	10,67732	-1,23975	91,1
Centro	10,34927	-1,23925	91,3
Sud	10,68124	-1,25312	93,6
ITALIA	11,06774	-1,25644	92,2

Prospetto 4.1 - 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
20000	69,3	65,7	68,7	86,5	74,1	59,9	70,9	81,5
30000	54,6	51,2	53,4	67,0	58,3	47,5	55,5	63,9
40000	46,1	42,9	44,7	56,0	49,2	40,3	46,7	53,7
50000	40,4	37,4	38,9	48,6	43,1	35,5	40,8	46,9
60000	36,3	33,4	34,7	43,4	38,7	32,0	36,6	42,1
70000	33,2	30,4	31,5	39,4	35,3	29,3	33,3	38,3
80000	30,7	28,0	29,0	36,2	32,6	27,1	30,7	35,4
90000	28,6	26,1	27,0	33,6	30,4	25,3	28,6	32,9
100000	26,9	24,4	25,3	31,5	28,6	23,9	26,9	30,9
200000	17,9	15,9	16,4	20,4	19,0	16,0	17,7	20,4
300000	14,1	12,4	12,8	15,8	14,9	12,7	13,9	16,0
400000	11,9	10,4	10,7	13,2	12,6	10,8	11,7	13,4
500000	10,4	9,1	9,3	11,4	11,0	9,5	10,2	11,7
750000	8,2	7,1	7,2	8,9	8,7	7,5	8,0	9,2
1000000	6,9	5,9	6,0	7,4	7,3	6,4	6,7	7,7
2000000	4,6	3,9	3,9	4,8	4,9	4,3	4,4	5,1
3000000	3,6	3,0	3,1	3,7	3,8	3,4	3,5	4,0
4000000	3,1	2,5	2,6	3,1	3,2	2,9	2,9	3,4
5000000	2,7	2,2	2,2	2,7	2,8	2,5	2,5	2,9
7500000	2,1	1,7	1,7	2,1	2,2	2,0	2,0	2,3
10000000	1,8	1,4	1,4	1,7	1,9	1,7	1,7	1,9
15000000	1,4	1,1	1,1	1,4	1,5	1,4	1,3	1,5
20000000	1,2	0,9	0,9	1,1	1,2	1,1	1,1	1,3
25000000	1,0	0,8	0,8	1,0	1,1	1,0	1,0	1,1

Prospetto 4.2 - Valori interpolati degli errori relativi percentuali delle stime di frequenze assolute riferite agli individui per aree territoriali- Anno

STIME	AREE TERRITORIALI			
	Nord	Centro	Sud	ITALIA
20000	44,9	38,2	42,1	50,3
30000	34,9	29,7	32,7	39,0
40000	29,2	24,9	27,3	32,5
50000	25,5	21,7	23,7	28,3
60000	22,7	19,3	21,2	25,2
70000	20,7	17,6	19,2	22,9
80000	19,0	16,2	17,7	21,0
90000	17,7	15,1	16,4	19,5
100000	16,6	14,1	15,4	18,3
200000	10,8	9,2	10,0	11,8
300000	8,4	7,1	7,7	9,2
400000	7,0	6,0	6,4	7,7
500000	6,1	5,2	5,6	6,7
750000	4,8	4,0	4,3	5,2
1000000	4,0	3,4	3,6	4,3
2000000	2,6	2,2	2,4	2,8
3000000	2,0	1,7	1,8	2,2
4000000	1,7	1,4	1,5	1,8
5000000	1,5	1,2	1,3	1,6
7500000	1,1	1,0	1,0	1,2
10000000	1,0	0,8	0,9	1,0
15000000	0,7	0,6	0,7	0,8
20000000	0,6	0,5	0,6	0,7
25000000	0,5	0,5	0,5	0,6

Prospetto 5 - 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	R²
		gennaio-marzo		
Nord	60591,340	0,303162	-0,000000027441	89,9
Centro	32959,451	0,338903	-0,000000068143	87,9
Sud	68292,967	0,383171	-0,000000042500	82,7
ITALIA	134392,192	0,201611	-0,000000008544	86,7
		luglio-settembre		
Nord	62748,623	0,287782	-0,000000032686	87,2
Centro	43495,965	0,514109	-0,000000058986	91,6
Sud	88965,452	0,343657	-0,000000028472	90,0
ITALIA	124446,217	0,225183	-0,000000008282	92,4
		Anno		
Nord	181771,175	0,153764	-0,000000003711	89,9
Centro	75527,137	0,242248	-0,000000011495	92,8
Sud	155632,721	0,191158	-0,000000004859	93,6
ITALIA	274232,025	0,109146	-0,000000001100	93,4
NOTTI – vacanze				
AREE TERRITORIALI	a	b	c	R²
		gennaio-marzo		
Nord	203173,403	0,136341	-0,000000001928	81,3
Centro	81939,917	0,177030	-0,000000006590	93,0
Sud	89811,610	0,161849	-0,000000005184	91,9
ITALIA	259206,455	0,091029	-0,000000000695	85,6
		luglio-settembre		
Nord	629336,388	0,066876	-0,000000000152	88,4
Centro	356547,319	0,111396	-0,000000000684	88,3
Sud	327011,406	0,118382	-0,000000000516	92,8
ITALIA	843097,512	0,051421	-0,000000000062	90,6
		Anno		
Nord	749305,951	0,053546	-0,000000000081	90,2
Centro	405521,900	0,087216	-0,000000000361	89,6
Sud	384483,439	0,086715	-0,000000000251	90,6
ITALIA	984579,113	0,040107	-0,000000000032	91,2
NOTTI – totali				
AREE TERRITORIALI	a	b	c	R²
		gennaio-marzo		
Nord	238799,216	0,121218	-0,000000001407	84,2
Centro	103110,231	0,158014	-0,000000004578	93,3
Sud	116797,263	0,161378	-0,000000004332	90,8
ITALIA	319102,666	0,081602	-0,000000000505	88,3
		luglio-settembre		
Nord	637591,165	0,065226	-0,000000000139	89,8
Centro	385781,769	0,111408	-0,000000000640	90,5
Sud	354647,212	0,112373	-0,000000000444	93,4
ITALIA	885880,482	0,050041	-0,000000000056	91,9
		Anno		
Nord	789544,268	0,051610	-0,000000000074	90,8
Centro	434393,757	0,084257	-0,000000000310	90,9
Sud	436527,312	0,081590	-0,000000000219	90,4
ITALIA	1074889,921	0,038171	-0,000000000028	91,9

Prospetto 6 - 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	9458975	7690585	5952361	4275753	2742419	1544667	860157
Centro	4350808	3638812	2936835	2253068	1605843	1037218	618794
Sud	8039118	6895801	5765049	4655054	3582004	2579815	1716583
ITALIA	18590795	13093922	8005426	4061413	2041457	1233743	862809
luglio-settembre							
Nord	7529686	6061725	4629957	3272285	2079214	1211200	729193
Centro	7960752	7123972	6290036	5460194	4636525	3822723	3025863
Sud	10608410	8908506	7233610	5603194	4059200	2693425	1659782
ITALIA	21840754	16051572	10508105	5684300	2657607	1435307	938581
Anno							
Nord	29613619	17315556	7523953	3139983	1768230	1206082	910606
Centro	17108920	12885007	8774118	4991740	2248215	1076947	655149
Sud	30111741	20333954	11303130	4822112	2233098	1348685	952065
ITALIA	58039838	20479132	5804802	2915420	1918173	1425155	1132718
NOTTI – vacanze							
AREE TERRITORIALI	5%	10%	15%	20%	25%	30%	35%
gennaio-marzo							
Nord	47031867	23363782	7317670	2931391	1736434	1223804	942903
Centro	19900656	12670090	6129974	2190561	1027567	644108	465469
Sud	22352333	13240117	5459423	1875952	964173	634970	471221
ITALIA	64752830	13904944	4188575	2343616	1619057	1235314	998232
luglio-settembre							
Nord	140814949	17584354	7469310	4702265	3426945	2694855	2220190
Centro	95286228	32646842	8079965	3906331	2540586	1877677	1487963
Sud	137193201	48664944	9016412	3910005	2460823	1791434	1407447
ITALIA	128958453	16988493	8507187	5661089	4240063	3388815	2822053
Anno							
Nord	120540032	15700352	7718519	5101933	3808179	3037318	2525820
Centro	113166903	20206443	6235713	3555159	2477570	1899677	1539922
Sud	156227060	20786118	5935842	3368836	2346228	1798873	1458308
ITALIA	79099735	16295763	8935960	6150094	4687479	3786618	3176106
NOTTI – totali							
AREE TERRITORIALI	5%	10%	15%	20%	25%	30%	35%
gennaio-marzo							
Nord	53771775	22591988	6334892	2882707	1818166	1321944	1037167
Centro	24513584	14252961	5701248	2013699	1064537	709950	530365
Sud	26718699	15867131	6669002	2385684	1242459	821468	610648
ITALIA	71392505	12826398	4514811	2664866	1884279	1456201	1186267
luglio-settembre							
Nord	141848977	17158262	7430526	4707954	3441742	2711413	2236496
Centro	101894120	35038540	8732282	4225646	2748705	2031590	1609961
Sud	145990833	45453173	8560844	3967514	2555811	1881797	1488318
ITALIA	126251494	17393719	8818971	5894545	4424844	3541303	2951718
Anno							
Nord	114962779	15929475	7976922	5306731	3973882	3175649	2644284
Centro	121833771	19834840	6413241	3716052	2608147	2007678	1631529
Sud	156900173	19285450	6255708	3661761	2583366	1994667	1624191
ITALIA	76882776	17250112	9588871	6634510	5070924	4103508	3445980

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 viaggi

VIAGGI – lavoro				
AREE TERRITORIALI	a	b	c	R²
		gennaio-marzo		
Nord	15266,011	0,128288	-0,000000004527	96,8
Centro	10110,391	0,170408	-0,000000013385	98,3
Sud	9891,129	0,157318	-0,000000006930	98,7
ITALIA	18168,989	0,088183	-0,000000001384	97,9
		luglio-settembre		
Nord	12740,898	0,155166	-0,000000004611	99,4
Centro	8469,308	0,265123	-0,000000018728	98,8
Sud	9764,621	0,190054	-0,000000008311	97,8
ITALIA	17339,084	0,113984	-0,000000001615	99,1
		Anno		
Nord	26919,902	0,082830	-0,000000000679	98,3
Centro	18839,602	0,116136	-0,000000002212	98,0
Sud	20771,202	0,107550	-0,000000001204	96,1
ITALIA	39116,547	0,057312	-0,000000000217	98,0
VIAGGI – vacanza				
AREE TERRITORIALI	a	b	c	R²
		gennaio-marzo		
Nord	28902,882	0,100187	-0,000000004204	92,9
Centro	16872,380	0,159951	-0,000000015078	92,3
Sud	14899,555	0,175887	-0,000000008812	89,9
ITALIA	38113,802	0,076239	-0,000000001403	92,1
		luglio-settembre		
Nord	35544,063	0,103978	-0,000000005921	93,3
Centro	22812,094	0,138484	-0,000000015553	92,2
Sud	26646,132	0,106338	-0,000000005766	95,4
ITALIA	52004,152	0,065302	-0,000000001481	95,3
		Anno		
Nord	68092,478	0,053580	-0,000000000603	93,1
Centro	40585,277	0,075683	-0,000000001852	93,9
Sud	41161,866	0,067852	-0,000000000854	94,4
ITALIA	95857,328	0,035822	-0,000000000171	93,1
VIAGGI - totali				
AREE TERRITORIALI	a	b	c	R²
		gennaio-marzo		
Nord	25294,812	0,100065	-0,000000004469	92,7
Centro	15337,233	0,156114	-0,000000015954	92,8
Sud	13988,718	0,161084	-0,000000008408	89,4
ITALIA	32168,040	0,074491	-0,000000001452	91,9
		luglio-settembre		
Nord	29638,179	0,105379	-0,000000006259	90,6
Centro	19094,947	0,160328	-0,000000021324	90,9
Sud	22694,398	0,109487	-0,000000006286	95,1
ITALIA	40182,065	0,072008	-0,000000001807	95,2
		Anno		
Nord	40550,314	0,063762	-0,000000000794	87,9
Centro	27322,110	0,089897	-0,000000002484	88,6
Sud	29160,642	0,076515	-0,000000001037	87,6
ITALIA	54109,190	0,042682	-0,000000000224	87,4

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 viaggi

VIAGGI – lavoro							
AREE TERRITORIALI	5%	10%	15%	20%	25%	30%	35%
gennaio-marzo							
Nord	17485463	6748147	622363	210094	124848	88698	68759
Centro	9079050	5400169	1918463	300751	124425	77399	56062
Sud	15576829	8439673	1834088	223620	105883	69091	51240
ITALIA	28052643	1330215	292005	162162	112173	85729	69370
luglio-settembre							
Nord	22927160	12190171	2314278	276328	133483	87724	65293
Centro	11525618	8867650	6219633	3602718	1188092	217442	97678
Sud	16920603	10942572	5051805	639745	159368	88224	60857
ITALIA	39892401	9760054	471463	200823	127286	93138	73429
Anno							
Nord	49164421	1481103	399162	229446	160928	123910	100734
Centro	30182558	8319122	537469	223330	140412	102339	80497
Sud	48166163	8340878	482706	224022	145635	107858	85636
ITALIA	38361836	912105	421608	274026	202958	161157	133633
VIAGGI – vacanze							
AREE TERRITORIALI	5%	10%	15%	20%	25%	30%	35%
gennaio-marzo							
Nord	12488487	2644373	554297	286122	191893	144212	115474
Centro	7442663	4240074	1438121	369809	181833	118952	88163
Sud	14403554	8803999	3430600	519335	196449	119042	85206
ITALIA	20056467	1475516	511741	306896	218959	170151	139124
luglio-settembre							
Nord	9732656	2808848	707855	362080	241059	180344	143976
Centro	5936251	2968481	896198	341371	199039	139367	107009
Sud	10222632	2768466	567723	279678	184118	137032	109076
ITALIA	13029331	1413499	607542	384455	280930	221270	182491
Anno							
Nord	14004750	1439962	703119	464163	346300	276141	229609
Centro	15296614	1498070	538875	324894	232252	180659	147803
Sud	22999031	1239581	498488	310859	225742	177194	145823
ITALIA	6284824	1487720	838490	583507	447399	362766	305054
VIAGGI – totali							
AREE TERRITORIALI	5%	10%	15%	20%	25%	30%	35%
gennaio-marzo							
Nord	11685848	2386248	485461	250311	167865	126159	101023
Centro	6792695	3772051	1190631	313703	159061	105362	78596
Sud	13335710	7486795	2107573	335185	155052	100093	73805
ITALIA	18088521	1181579	422583	255546	183008	142516	116687
luglio-settembre							
Nord	9353911	2647782	611733	306997	203151	151548	120786
Centro	5341542	3116457	1218960	396724	203130	133973	99559
Sud	9831239	2799156	518467	246510	160360	118658	94127
ITALIA	13790727	1322567	509201	312563	225237	175998	144409
Anno							
Nord	19905020	1092840	468196	297129	217532	171551	141611
Centro	16716277	1856655	446353	246776	170204	129842	104938
Sud	26624899	1180190	394629	235682	167919	130403	106583
ITALIA	6212664	940562	503666	343780	260923	210243	176047

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 5 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 7 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 5 e si può utilizzare la seguente espressione:

$$\hat{\sigma}({}_c\hat{Y}_3) = 43495,965 + 0,514109 {}_c\hat{Y}_3 - 0,000000058986({}_c\hat{Y}_3)^2.$$

Esempio 2 - La valutazione degli errori di campionamento delle stime mediante un metodo approssimato si può ottenere utilizzando i dati dei prospetti 6 e 8. Il prospetto 6 fa riferimento alle stime del numero di notti; il prospetto 8 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 6, 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 16.988.493 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 6, si consideri la stima numero di notti per lavoro, riferita al totale Italia nel trimestre luglio-settembre, si supponga sia pari a 14.844.000 (si veda il prospetto 1.1). Dal prospetto 6, considerando le notti-lavoro ed esaminando l'ultima riga corrispondente al totale Italia per il trimestre luglio-settembre, si può osservare che il valore 14.844.000 risulta essere compreso tra i valori 10.508.105 e 16.051.572. 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 - Dal prospetto 2.1, risulta che il totale dei turisti italiani nel trimestre gennaio-marzo è pari a 11.015.000. Si cerca il livello di stima che più si avvicina a 11.015.000 nella prima colonna del prospetto 4.1, che è pari a 10.000.000. Dall'ultima colonna del prospetto, con riferimento all'Italia, si ricava il corrispondente errore relativo percentuale della stima considerata, che è pari a 1,7%.

L'errore assoluto sarà: $\sigma(11.015.000) = 1,7/100 \times 11.015.000 = 187.255$

L'intervallo di confidenza avrà come estremi:

$$11.015.000 - (2 \times 187.255) = 10.640.490$$

$$11.015.000 + (2 \times 187.255) = 11.389.510$$

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 1,7 e 1,4 percentuali. L'errore corrispondente a 11.015.000 è pari a:

$$\sigma(11.015.000) = 1,7 - ((1,7 - 1,4) / (15.000.000 - 10.000.000)) \times (11.015.000 - 10.000.000) = 1,64\%$$

Il corrispondente errore assoluto è $1,64/100 \times 11.015.000 = 180.646$ e l'intervallo di confidenza avrà come estremi:

$$11.015.000 - (2 \times 180.646) = 10.653.708$$

$$11.015.000 + (2 \times 180.646) = 11.376.292$$

Esempio 3 - Il calcolo dell'errore può essere effettuato direttamente tramite la funzione interpolante (12):

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

Per la stima di 11.015.0000 si ha: $\log \varepsilon^2 = 12,15213 - 1,25639 \times \log(11.015.000) = -8,21994$.

Da questa il valore di ε è facilmente desumibile, calcolando la radice quadrata dell'antilogaritmo dell'espressione di sopra. Infatti, da $\log(\varepsilon^2) = -8,21994$ si ha:

$$\varepsilon^2 = \exp(-8,21994), \text{ e dunque:}$$

$$\varepsilon = 0,016408$$

L'errore relativo percentuale sarà dunque pari a 1,64 %.