

File Standard

***Indagine Multiscopo sulle Famiglie
Viaggi e vacanze
Anno 2000***

Manuale utente con tracciato record

Per informazioni sull'acquisto rivolgersi a:
ISTAT - Servizio Biblioteca e Servizio 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: 06.5952.4841, 4840, 4844
fax: 06.5952.4596
e-mail: perez@istat.it

INDICE

1. PREMESSA	4
2. FINALITÀ E CARATTERISTICHE DELL'INDAGINE	4
3. DEFINIZIONI	5
4. AVVERTENZE PER L'UTILIZZAZIONE DEI FILES.....	6
5. COSTRUZIONE DELLE STIME ED ERRORI DI CAMPIONAMENTO	7
6. TRACCIATO DEL FILE DEGLI INDIVIDUI (tipo A e B)	9
7. TRACCIATO DEL FILE DEI VIAGGI (tipo A e B)	14
APPENDICE A - Il questionario CATI, anno 2000	19
APPENDICE B - Strategia di campionamento e valutazione degli errori campionari.....	200

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 l'eventuale 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. FINALITÀ E CARATTERISTICHE DELL'INDAGINE

Il crescente interesse sul turismo, sia in ambito nazionale che 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.

3. DEFINIZIONI

La rilevazione ha utilizzato 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, fatta eccezione per quelle destinazioni abitualmente frequentate. 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. Vengono 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**: persona che ha effettuato uno o più viaggi nel trimestre di riferimento dell'indagine.

I viaggi sono distinti secondo le seguenti tipologie:

- **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.

Il motivo del viaggio di vacanza, la destinazione, il mezzo di trasporto e il tipo di alloggio sono informazioni che vengono 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 dove si è trascorso il maggior numero di notti, mentre il mezzo di trasporto viene individuato nel mezzo con cui è stata coperta la maggiore distanza.

Le caratteristiche individuali fanno riferimento all'epoca dell'intervista. In particolare:

- **l'età** è espressa in anni compiuti;
- **la condizione** è quella dichiarata come unica o prevalente dalle persone di 15 anni e più. Si precisa, inoltre il significato dei seguenti termini:
 - **occupato**: chi possiede un'occupazione in proprio o alle dipendenze da cui trae un profitto o una retribuzione (utile, onorario, stipendio, salario) o chi collabora con un familiare che svolge un'attività lavorativa in conto proprio senza avere un regolare contratto di lavoro (coadiuvante);
 - **persona in cerca di occupazione**: chi ha perduto una precedente occupazione alle dipendenze o non ha mai esercitato un'attività lavorativa ed è alla ricerca attiva di un'occupazione che è in grado di accettare se gli venisse offerta;
 - **casalinga**: chi si dedica prevalentemente alle faccende domestiche;

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 con i dati del 2000 è consultabile anche sul sito ISTAT all'URL <http://www.istat.it/Aproserv/noved/viaggi2001/index.html>.

- **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 del ritirato dal lavoro non coincide necessariamente con quella del 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 (militare di leva, inabile al lavoro, benestante, ecc.).

Le caratteristiche territoriali della famiglia sono così definite:

- **le ripartizioni geografiche** costituiscono una suddivisione geografica del territorio e sono così articolate:
 - **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**: I Comuni italiani sono suddivisi nelle seguenti classi:
 - **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 Popolazione del 1991;
 - **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);

4. AVVERTENZE PER L'UTILIZZAZIONE DEI FILES

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: 40.690
- **un file dei viaggi** (tipo A o tipo B)
lunghezza record: 120 bytes
numero records: 15.222

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.

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.

I files dell'anno 2000 (individui e viaggi) sono ottenuti mettendo in sequenza i quattro files trimestrali (individui e viaggi); in ogni record è specificato il trimestre di riferimento.

Per un agevole e corretto utilizzo dei files si precisa che, a seconda della selezione che si opera sui files, è possibile effettuare elaborazioni sulle seguenti unità di analisi:

-) **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). Il totale delle famiglie è 14.012.
-) **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). Il totale degli individui è 40.880.
-) **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). Il totale dei viaggi è 14.948.

5. COSTRUZIONE DELLE STIME ED ERRORI DI CAMPIONAMENTO

Le informazioni riportate nei files sono di carattere campionario. In particolare, ciascun sub-file trimestrale si riferisce ad un campione di circa 3500 famiglie che viene selezionato in modo indipendente dai campioni delle altre tre rilevazioni.

Dato il carattere campionario dell'indagine, per poter ottenere stime relative all'intera popolazione di riferimento è necessario procedere alla ponderazione dei dati contenuti nei files, attraverso i coefficienti di riporto all'universo (rappresentati, per ciascun record, dai valori contenuti nella variabile COE). I valori contenuti in questa variabile, ottenuti separatamente per ciascun trimestre, rappresentano il peso che ogni unità campionaria di analisi ha nella popolazione e nel trimestre di riferimento. Ad esempio, se in un record individuale del primo trimestre la variabile COE vale '04782.72', significa che l'individuo rappresenta 4782,72 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 4782,72 individui. Di conseguenza, ponderando i dati riferiti a tutti gli individui campionati in uno specifico trimestre si otterranno stime riferite all'intera popolazione residente in Italia nello stesso 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.

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.

È importante sottolineare che, mentre i viaggi di ciascuna rilevazione trimestrale possono essere sommati per ottenere la stima annuale, l'operazione di somma non è corretta se effettuata sui files trimestrali degli individui. Infatti i dati sugli individui (turisti e non) riferiti a ciascun trimestre e ponderati con il valore di COE riproducono l'ammontare di tutta la popolazione residente in Italia; un'eventuale somma di questi dati sui quattro trimestri comporterebbe pertanto una quadruplicazione

della popolazione. Per evitare ciò, qualora si volessero utilizzare dati sugli individui riferiti a tutte e quattro le rilevazioni, sarà necessario utilizzare un valore medio di popolazione, ricavabile dividendo COE per 4. Analogamente, i dati trimestrali sul numero dei turisti non sono cumulabili, in quanto una stessa persona può essere turista in trimestri diversi. Una eventuale somma dei dati 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.

Essendo l'indagine, come già detto, 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. Si riportano più avanti (appendice B) le informazioni relative alla strategia di campionamento e alla valutazione degli errori di stima, con esempi di calcolo.

6. TRACCIATO DEL FILE DEGLI INDIVIDUI (tipo A e B)

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 2000) 02=secondo trimestre (aprile-giugno 2000) 03=terzo trimestre (luglio-settembre 2000) 04=quarto trimestre (ottobre-dicembre 2000)
Coll. 5-8	ANNRIF	ANNO DI RIFERIMENTO = 2000
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
Coll. 32-33	NCOMP	NUMERO COMPONENTI DELLA FAMIGLIA
Coll. 34-44	-	RISERVATO ISTAT
INFORMAZIONI SUI VIAGGI FATTI DALLA FAMIGLIA		
Col. 45	FVIA	PRESENZA DI VIAGGI IN FAMIGLIA NEL TRIMESTRE 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 NEL TRIMESTRE 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 NEL TRIMESTRE 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 NEL TRIMESTRE 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 NOTTI E PIÙ) IN FAMIGLIA NEL TRIMESTRE 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 LA PERSONA DI RIFERIMENTO 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 13 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 (se occupato: COND=01) 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 (se occupato: COND=01) 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
INFORMAZIONI SUI VIAGGI FATTI DALL'INDIVIDUO		
Col. 91	IVIA	PRESENZA DI VIAGGI DELL'INDIVIDUO NEL TRIMESTRE 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 NEL TRIMESTRE 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 NEL TRIMESTRE 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 NEL TRIMESTRE 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 NOTTI E PIÙ) DELL'INDIVIDUO NEL TRIMESTRE 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 EFFETTUATO VACANZE NEL TRIMESTRE *(se non ha effettuato vacanze: 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

7. TRACCIATO DEL FILE DEI VIAGGI (tipo A e B)

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 2000) 02=secondo trimestre (aprile-giugno 2000) 03=terzo trimestre (luglio-settembre 2000) 04=quarto trimestre (ottobre-dicembre 2000)
Coll. 5-8	ANNRIF	ANNO DI RIFERIMENTO = 2000
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 "L100"=viaggio di lavoro concluso nel primo trimestre 2000 "L200"=viaggio di lavoro concluso nel secondo trimestre 2000 "L300"=viaggio di lavoro concluso nel terzo trimestre 2000 "L400"=viaggio di lavoro concluso nel quarto trimestre 2000 "V100"=viaggio di vacanza concluso nel primo trimestre 2000 "V200"=viaggio di vacanza concluso nel secondo trimestre 2000 "V300"=viaggio di vacanza concluso nel terzo trimestre 2000 "V400"=viaggio di vacanza concluso nel quarto trimestre 2000
Coll. 39-40	MOTLAV	MOTIVO DEL VIAGGIO DI LAVORO (se viaggio di lavoro: TIPOVIA="L100", "L200", "L300", "L400") 01=partecipazione a congresso, convegno, conferenza, seminario 02=partecipazione a riunione d'affari 03=esibizione, fiera, mostra, esposizione (visita e/o partecipazione professionale) 04=missione di lavoro o militare (amministrazione pubblica e/o privata, organismo internazionale, altra istituzione) 05=viaggio o meeting d'azienda per la familiarizzazione dei dipendenti e/o 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 il trasporto di persone e/o merci 12=altro motivo
Coll. 41-43	-	RISERVATO ISTAT
Coll. 44-45	INSIEME	VARIABILE CHE INDIVIDUA I VIAGGI FATTI INSIEME DAI COMPONENTI FAMILIARI (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 VARIABILI 'RELPAR' (RELAZIONI DI PARENTELA RISPETTO PERSONA DI RIFERIMENTO) RELATIVE AI COMPONENTI FAMILIARI PARTECIPANTI AL VIAGGIO (sequenza di 11 campi da 2 bytes, serrati a sinistra) 01=persona di riferimento (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. 70	ALPARTN	PARTECIPAZIONE AL VIAGGIO DI PERSONE CHE NON VIVONO IN FAMIGLIA 1=no 2=sì, solo parenti 3=sì, solo amici/conoscenti 4=sì, 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 sa, non risponde
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 PREVALENTE DEL VIAGGIO DI VACANZA (<i>se viaggio di vacanza: TIPOVIA= "V100", "V200", "V300", "V400"</i>) 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 (<i>se il motivo della vacanza è piacere/svago: MOTIVO=01</i>) 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 PREVALENTE 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	REGIONE O PAESE ESTERO DI DESTINAZIONE PRINCIPALE 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=Germaia
 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 DELLE VARIABILI 'IDIND' (NUMERO D'ORDINE ENTRO LA FAMIGLIA) RELATIVE AI COMPONENTI FAMILIARI PARTECIPANTI AL VIAGGIO (sequenza di 11 campi da 2 bytes, serrati a sinistra)

APPENDICE A - Il questionario CATI, anno 2000

Il questionario è strutturato in Sezioni in modo da garantire continuità di contenuto fra le domande e da favorire 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, definendo il contenuto stesso dell'indagine e ricordando che le notizie raccolte sono coperte da segreto d'Ufficio. Verifica, inoltre, che la famiglia abbia ricevuto la lettera del 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 la famiglia il target dell'indagine, si verifica in primo luogo che l'intestatario del contratto telefonico sia effettivamente una famiglia e non un'impresa, un'azienda, un ufficio, un gruppo di individui. Per evitare, inoltre, possibili 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 di età (escludendo 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 registrate 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 registrate le informazioni per ciascun viaggio effettuato dai componenti della famiglia per motivi professionali;
- 'Scheda viaggio di vacanza', dove vengono registrate le informazioni per ciascun viaggio effettuato dai componenti della famiglia per motivi di vacanza/personali.

Si fa presente che può non esserci perfetta corrispondenza tra le modalità delle informazioni indicate nel questionario e quelle fornite nei files standard. Per evitare di incorrere in equivoci, nell'utilizzo del file standard è opportuno far riferimento alle modalità indicate nei tracciati e non a quelle usate nel questionario.

APPENDICE B - Strategia di campionamento e valutazione degli errori campionari

0. - 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 dell'ultimo Censimento Generale della Popolazione (*"I sistemi locali del lavoro 1991"*, ISTAT, Argomenti n.10 - 1997).

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 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 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 vincolata⁵ che tengono conto di totali noti (desunti da fonti esterne all'indagine) della popolazione relativi a variabili ausiliarie correlate a quelle d'interesse.

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

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.

⁵ 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)

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.

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 B.1 e B.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 B.1 - Distribuzione e tassi di sostituzione del campione per regione*

REGIONI	INDIRIZZI CAMPIONE	TASSO DI SOSTITUZIONE TOTALE (a)	TASSO DI SOSTITUZIONE AL NETTO DEI MANCATI CONTATTI (b)
		(%)	(%)
Piemonte	265	43,5	35,9
Valle d'Aosta	92	35,6	28,7
Lombardia	330	37,9	33,0
Trentino-Alto Adige	206	29,0	23,5
Veneto	201	30,8	27,6
Friuli-Venezia Giulia	139	37,4	33,1
Liguria	152	44,1	32,2
Emilia-Romagna	197	43,2	38,1
Toscana	213	37,3	31,1
Umbria	103	42,4	35,3
Marche	141	35,5	29,1
Lazio	219	39,7	31,8
Abruzzo	144	48,9	39,7
Molise	96	50,0	38,7
Campania	196	49,5	40,0
Puglia	187	37,8	32,4
Basilicata	99	42,7	36,5
Calabria	158	42,8	36,8
Sicilia	213	46,8	36,1
Sardegna	150	45,6	34,0
ITALIA	3501	40,6	33,5

* Dato della rilevazione del quarto trimestre.

- () Il tasso di sostituzione totale è calcolato come rapporto percentuale tra il numero di famiglie sostituite ed il numero teorico di famiglie campione. Esso include 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).
- (a) 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 B.2 - Distribuzione e tassi di sostituzione del campione per tipologia di comune*

TIPOLOGIA DEL COMUNE	INDIRIZZI CAMPIONE	TASSO DI SOSTITUZIONE TOTALE*	TASSO DI SOSTITUZIONE AL NETTO DEI MANCATI CONTATTI*
		(%)	(%)
Comuni centro dell'area metropolitana	440	38,8	31,4
Comuni che gravitano intorno al centro dell'area metropolitana	316	39,0	34,1
Comuni aventi fino a 2.000 abitanti	618	47,0	33,7
Comuni con 2.001-10.000 abitanti	868	42,9	36,3
Comuni con 10.001-50.000 abitanti	950	39,0	31,5
Comuni con oltre 50.000 abitanti	309	38,3	33,0
ITALIA	3501	40,6	33,5

* Cfr. note al prospetto B.1.

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 noti sono: la distribuzione della popolazione regionale per sesso; la distribuzione della popolazione nelle aree A_1 , A_2 , B_1 , B_2 , B_3 , B_4 definite nel paragrafo 1; la distribuzione della popolazione per classi di età a livello di ripartizione; la distribuzione delle famiglie per numero di

componenti⁶. Indicando, quindi, con ${}_kX$ il totale noto della k-ma variabile ($k=1\dots K$) ausiliaria e con ${}_kX_{thjp}$ il valore assunto dalla k-ma variabile ausiliaria per l'individuo rispondente $thjp$, la condizione sopra descritta è espressa dalla seguente uguaglianza:

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

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

- 0) viene dapprima calcolato il peso base (o peso diretto), ottenuto come reciproco della probabilità di inclusione di ogni unità campionaria;
- 1) viene calcolato poi il fattore correttivo per mancata risposta totale, ottenuto come l'inverso del tasso di risposta in ciascuna regione;
- 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 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, ${}_dY$, 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 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, ${}_dY$ 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:

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

⁶ La distribuzione delle famiglie per numero di componenti è desunta dai dati dell'indagine multiscopo "Aspetti della vita quotidiana" dell'anno 1999.

5. - Valutazione del livello di precisione delle stime

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

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

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

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

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

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

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

dove

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

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

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

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

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

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

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

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

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 \hat{Y}_t corrisponde un errore campionario relativo $\hat{\varepsilon}(\hat{Y}_t)$, 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:

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

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 B.3.1 (per i trimestri) e B.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 B.4.1 (per i trimestri) e B.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 delle tavole 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 nelle tavole permettono di calcolare l'errore relativo di una generica stima di frequenza assoluta mediante due procedimenti che risultano di facile applicazione, anche se conducono a risultati meno precisi di quelli ottenibili mediante l'applicazione diretta dell'espressione (12).

Il 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 nella tavola (C.4.1 o C.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:

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

dove i parametri a, b e c vengono stimati utilizzando il metodo dei minimi quadrati, adattando il modello (13) 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 (13) 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 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 B.5 e B.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, ${}_d\hat{Y}_t$, 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 B.6 e B.8 (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 B.3.1 - Valori dei coefficienti a , b e di R^2 dei modelli per l'interpolazione degli errori campionari relativi alle stime di frequenze assolute sugli individui per aree territoriali – Trimestri

AREE TERRITORIALI	TRIMESTRI 1, 2 e 4			TRIMESTRE 3		
	a	b	R^2	a	b	R^2
Nord	11,79265	-1,25146	93,7	11,03917	-1,19616	91,6
Centro	10,83994	-1,19959	92,0	11,03315	-1,20555	91,2
Sud	10,73358	-1,19921	92,2	10,51299	-1,17389	92,1
ITALIA	11,94248	-1,25397	93,9	11,33211	-1,20826	92,3

Prospetto B.3.2 - Valori dei coefficienti a , b e di R^2 dei modelli per l'interpolazione degli errori campionari relativi alle stime di frequenze assolute sugli individui per aree territoriali – Anno

AREE TERRITORIALI	a	b	R^2
Nord	12,065225	-1,2438976	93,3
Centro	11,537211	-1,2248743	93,1
Sud	11,716247	-1,2402313	94,0
ITALIA	12,254514	-1,2484688	93,5

Prospetto B.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	74,0	59,4	56,5	78,8	66,8	63,6	57,3	72,8
30000	57,4	46,6	44,3	61,1	52,4	49,8	45,2	57,0
40000	48,0	39,2	37,3	51,0	44,1	41,9	38,2	47,9
50000	41,7	34,3	32,6	44,4	38,6	36,6	33,5	41,9
60000	37,2	30,8	29,2	39,6	34,6	32,8	30,1	37,5
70000	33,8	28,0	26,6	35,9	31,6	29,9	27,5	34,2
80000	31,1	25,9	24,6	33,0	29,2	27,6	25,4	31,5
90000	28,9	24,1	22,9	30,7	27,2	25,7	23,7	29,4
100000	27,0	22,6	21,5	28,7	25,5	24,1	22,3	27,5
200000	17,5	14,9	14,2	18,6	16,9	15,9	14,8	18,1
300000	13,6	11,7	11,1	14,4	13,2	12,4	11,7	14,2
400000	11,4	9,9	9,4	12,0	11,1	10,4	9,9	11,9
500000	9,9	8,6	8,2	10,5	9,7	9,1	8,7	10,4
750000	7,7	6,8	6,4	8,1	7,6	7,2	6,8	8,2
1000000	6,4	5,7	5,4	6,8	6,4	6,0	5,8	6,9
2000000	4,1	3,8	3,6	4,4	4,3	4,0	3,8	4,5
3000000	3,2	2,9	2,8	3,4	3,3	3,1	3,0	3,5
4000000	2,7	2,5	2,4	2,8	2,8	2,6	2,6	3,0
5000000	2,3	2,2	2,1	2,5	2,5	2,3	2,2	2,6
7500000	1,8	1,7	1,6	1,9	1,9	1,8	1,8	2,0
10000000	1,5	1,4	1,4	1,6	1,6	1,5	1,5	1,7
15000000	1,2	1,1	1,1	1,2	1,3	1,2	1,2	1,3
20000000	1,0	0,9	0,9	1,0	1,1	1,0	1,0	1,1
25000000	0,9	0,8	0,8	0,9	0,9	0,9	0,9	1,0

Prospetto B.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	88,1	74,3	75,3	94,7
30000	68,5	58,0	58,6	73,5
40000	57,2	48,6	49,0	61,4
50000	49,8	42,4	42,7	53,4
60000	44,5	37,9	38,1	47,7
70000	40,4	34,5	34,6	43,3
80000	37,2	31,8	31,9	39,8
90000	34,6	29,6	29,6	37,0
100000	32,4	27,7	27,8	34,7
200000	21,0	18,1	18,1	22,5
300000	16,3	14,2	14,1	17,5
400000	13,7	11,9	11,8	14,6
500000	11,9	10,4	10,2	12,7
750000	9,2	8,1	8,0	9,9
1000000	7,7	6,8	6,7	8,2
2000000	5,0	4,4	4,3	5,3
3000000	3,9	3,5	3,4	4,1
4000000	3,3	2,9	2,8	3,5
5000000	2,8	2,5	2,5	3,0
7500000	2,2	2,0	1,9	2,3
10000000	1,8	1,7	1,6	2,0
15000000	1,4	1,3	1,2	1,5
20000000	1,2	1,1	1,0	1,3
25000000	1,0	0,9	0,9	1,1

Prospetto B.5 - Valori dei coefficienti a, b, c e di R² dei modelli per l'interpolazione degli errori campionari delle stime del numero totale di notti per tipologia del viaggio, riferimento temporale e area territoriale

NOTTI – lavoro				
AREE TERRITORIALI	a	b	c	R ²
		Trimestri 1, 2 e 4		
Nord	17589,226	0,446675	-0,000000045020	88,6
Centro	21729,245	0,356668	-0,000000059487	92,8
Sud	3851,230	0,589446	-0,000000112948	87,5
ITALIA	89466,796	0,247022	-0,000000011875	86,6
		Trimestre 3		
Nord	-15132,887	0,505686	-0,000000035219	92,2
Centro	10795,180	0,715393	-0,000000157674	90,0
Sud	-17628,230	0,604683	-0,000000029379	85,1
ITALIA	84475,834	0,606493	-0,000000025567	93,0
		Anno		
Nord	127512,309	0,192632	-0,000000004121	91,3
Centro	102238,610	0,252970	-0,000000013452	87,9
Sud	117345,294	0,205562	-0,000000002566	90,2
ITALIA	262807,712	0,127242	-0,000000001086	89,1
NOTTI – vacanza				
AREE TERRITORIALI	a	b	c	R ²
		Trimestri 1, 2 e 4		
Nord	160879,121	0,136659	-0,000000001795	88,0
Centro	83851,441	0,180386	-0,000000008074	86,9
Sud	82221,337	0,146195	-0,000000004796	92,6
ITALIA	214703,974	0,090606	-0,000000000681	88,7
		Trimestre 3		
Nord	605396,078	0,066756	-0,000000000152	87,2
Centro	381969,791	0,099506	-0,000000000628	83,5
Sud	305482,971	0,106310	-0,000000000443	91,6
ITALIA	870051,184	0,048409	-0,000000000056	88,8
		Anno		
Nord	4254488,070	0,019067	-0,000000000002	81,4
Centro	1528664,047	0,029003	-0,000000000014	82,9
Sud	3000322,466	0,019430	-0,000000000003	77,1
ITALIA	5898185,637	0,012911	-0,000000000001	79,0
NOTTI – totali				
AREE TERRITORIALI	a	b	c	R ²
		Trimestri 1, 2 e 4		
Nord	187822,169	0,131346	-0,000000001593	90,2
Centro	101650,478	0,163236	-0,000000005912	87,8
Sud	104947,679	0,145671	-0,000000004269	92,4
ITALIA	267946,558	0,085316	-0,000000000572	91,0
		Trimestre 3		
Nord	653601,821	0,065654	-0,000000000141	88,2
Centro	418309,856	0,096773	-0,000000000548	83,9
Sud	342218,264	0,112053	-0,000000000490	90,1
ITALIA	966885,887	0,048446	-0,000000000055	88,9
		Anno		
Nord	867844,829	0,052148	-0,000000000074	89,5
Centro	459495,510	0,080892	-0,000000000311	85,6
Sud	448506,451	0,081463	-0,000000000231	88,8
ITALIA	1156172,631	0,038222	-0,000000000029	89,6

Prospetto B.6 - Valori dei totali corrispondenti ad alcuni valori tipici degli errori relativi percentuali per la stima del numero di notti per tipologia del viaggio, riferimento temporale e area territoriale

NOTTI – lavoro							
AREE TERRITORIALI	5%	10%	15%	20%	25%	30%	35%
Trimestri 1, 2 e 4							
Nord	8855139	7750817	6648566	5549594	4456259	3373781	2316059
Centro	5225148	4397777	3576331	2765742	1977833	1245825	663026
Sud	4783166	4341212	3899421	3457858	3016620	2575874	2135921
ITALIA	17034119	12962395	9006987	5364345	2622316	1306275	795770
Trimestre 3							
Nord	12905282	11481473	10056490	8629746	7200180	5765652	4321057
Centro	4236217	3920409	3604828	3289538	2974632	2660242	2346572
Sud	18848589	17143512	15437734	13730992	12022873	10312686	8599182
ITALIA	21916707	19975721	18037856	16104215	14176479	12257336	10351336
Anno							
Nord	35479963	23777280	12767482	4739829	1949643	1137919	793781
Centro	15576613	12004812	8544255	5356622	2869472	1516286	932998
Sud	61377610	42227398	23594588	7933222	2327815	1203236	801031
ITALIA	74381349	32525952	8278022	3435877	2101781	1506972	1173081
NOTTI – vacanza							
AREE TERRITORIALI	5%	10%	15%	20%	25%	30%	35%
Trimestri 1, 2 e 4							
Nord	50061917	24133409	6453862	2379410	1388867	974488	749367
Centro	16769028	10908728	5613723	2229339	1071394	670658	483251
Sud	20879443	11167619	3762824	1362628	765032	525955	399672
ITALIA	64486131	12149729	3476306	1939251	1339337	1021962	825923
Trimestre 3							
Nord	138872044	16903775	7178423	4520197	3294762	2591171	2134919
Centro	85962678	24279811	6962166	3714735	2511792	1893913	1519083
Sud	132425742	34347475	6556468	3211819	2112235	1571527	1250728
ITALIA	110820923	16563955	8523845	5727237	4310713	3455514	2883313
Anno							
Nord	136449794	52506344	32479026	23508623	18420374	15142655	12855133
Centro	69603660	21441562	12615651	8933237	6914124	5639260	4761257
Sud	97141964	37182809	22965474	16610849	13010229	10692333	9075391
ITALIA	158627463	67694936	43016553	31522972	24875984	20543936	17496882
NOTTI – totalii							
AREE TERRITORIALI	5%	10%	15%	20%	25%	30%	35%
Trimestri 1, 2 e 4							
Nord	53262947	24485455	6480819	2581135	1550644	1102174	853680
Centro	20013380	12115744	5414517	2073545	1090539	720796	535205
Sud	23460437	12643685	4477198	1703662	967623	667695	508225
ITALIA	68615490	12329629	4000921	2309793	1617946	1243976	1010122
Trimestre 3							
Nord	143400074	17738619	7651224	4840480	3535956	2784383	2296001
Centro	93461574	24833857	7308693	3968668	2703840	2047033	1646050
Sud	131810181	41420095	8158164	3810240	2459298	1812257	1433976
ITALIA	119597473	18396452	9472681	6365225	4790944	3840454	3204488
Anno							
Nord	123822921	17654355	8810292	5852554	4379166	3497813	2911573
Centro	112575152	18489225	6461270	3819748	2703742	2090920	1704123
Sud	149162668	19469403	6405623	3756167	2651529	2047876	1667789
ITALIA	81864465	18555410	10316233	7137658	5455346	4414489	3707062

Prospetto B.7 - Valori dei coefficienti a, b, c e di R² dei modelli per l'interpolazione degli errori campionari delle stime del numero totale di viaggi per tipologia del viaggio, riferimento temporale e area territoriale

VIAGGI – lavoro				
AREE TERRITORIALI	a	b	c	R ²
		Trimestri 1, 2 e 4		
Nord	14635,353	0,137955	-0,000000004730	97,7
Centro	10732,404	0,159390	-0,000000012941	97,2
Sud	8725,108	0,175440	-0,000000007703	97,8
ITALIA	18013,878	0,095310	-0,000000001487	98,0
		Trimestre 3		
Nord	10077,727	0,124090	-0,000000003452	99,8
Centro	9758,249	0,219505	-0,000000015258	99,2
Sud	8631,403	0,200090	-0,000000008534	97,5
ITALIA	14828,341	0,102806	-0,000000001405	99,5
		Anno		
Nord	22062,012	0,070640	-0,000000000575	99,1
Centro	15527,513	0,103172	-0,000000001943	98,9
Sud	16270,080	0,094725	-0,000000001051	97,6
ITALIA	29985,663	0,050220	-0,000000000189	98,9
VIAGGI – vacanza				
AREE TERRITORIALI	a	b	c	R ²
		Trimestri 1, 2 e 4		
Nord	24834,963	0,104474	-0,000000004234	92,6
Centro	21695,177	0,149049	-0,000000014194	89,4
Sud	20215,756	0,136634	-0,000000006754	92,6
ITALIA	39481,353	0,072293	-0,000000001305	92,2
		Trimestre 3		
Nord	30550,384	0,100115	-0,000000005626	84,9
Centro	19026,852	0,148109	-0,000000018840	92,2
Sud	24868,068	0,095212	-0,000000005335	94,2
ITALIA	44261,692	0,067317	-0,000000001628	93,9
		Anno		
Nord	65073,495	0,042379	-0,000000000463	90,3
Centro	44932,575	0,064333	-0,000000001617	90,1
Sud	43055,209	0,053621	-0,000000000679	92,6
ITALIA	97018,797	0,028609	-0,000000000136	90,4
VIAGGI – totali				
AREE TERRITORIALI	a	b	c	R ²
		Trimestri 1, 2 e 4		
Nord	23381,932	0,104187	-0,000000004625	92,7
Centro	18360,379	0,147508	-0,000000015051	90,2
Sud	17605,278	0,134832	-0,000000006929	92,9
ITALIA	34391,174	0,071913	-0,000000001389	92,3
		Trimestre 3		
Nord	27413,889	0,092988	-0,000000005128	82,5
Centro	17928,435	0,148843	-0,000000019815	90,4
Sud	21826,575	0,099850	-0,000000006053	94,0
ITALIA	34364,090	0,071702	-0,000000001868	93,3
		Anno		
Nord	35770,643	0,058324	-0,000000000718	88,6
Centro	27717,331	0,086106	-0,000000002420	88,3
Sud	27971,161	0,064690	-0,000000000875	88,2
ITALIA	50603,012	0,038749	-0,000000000203	87,8

Prospetto B.8 - Valori dei totali corrispondenti ad alcuni valori tipici degli errori relativi percentuali per la stima del numero di viaggi per tipologia del viaggio, riferimento temporale e area territoriale

VIAGGI – lavoro							
AREE TERRITORIALI	5%	10%	15%	20%	25%	30%	35%
Trimestri 1, 2 e 4							
Nord	18759184	8392579	898214	231787	129908	90080	68914
Centro	8550097	4763445	1343077	245129	116507	75798	56092
Sud	16354294	9908213	3616003	322619	115640	69747	49874
ITALIA	30868106	2244314	326486	171651	116322	87950	70700
Trimestre 3							
Nord	21598346	7374544	370650	131967	79864	57225	44579
Centro	11166477	7913068	4691606	1662923	280599	118563	74136
Sud	17644640	11814000	6036995	1010970	168107	85763	57390
ITALIA	37853051	4396670	311316	152230	100644	75157	59966
Anno							
Nord	36938426	740698	277443	170419	122956	96166	78961
Centro	27649064	3758224	327146	159849	105606	78827	62877
Sud	42927548	2157372	292720	154311	104708	79228	63719
ITALIA	13199455	601000	300349	200148	150073	120038	100019
VIAGGI – vacanza							
AREE TERRITORIALI	5%	10%	15%	20%	25%	30%	35%
Trimestri 1, 2 e 4							
Nord	13307996	3007375	520333	257052	169817	126669	100974
Centro	7190838	3852396	1203268	384597	208779	141832	107151
Sud	13055699	5928652	1003560	308865	176468	123119	94465
ITALIA	18703026	1340363	503819	308186	221810	173215	142074
Trimestre 3							
Nord	9479732	2340404	575109	300759	202289	152188	121923
Centro	5394810	2901668	956021	327684	180700	123378	93429
Sud	8992698	1756415	435436	234519	159779	121052	97404
ITALIA	12770116	1273526	529796	332237	241767	189971	156436
Anno							
Nord	6200933	1119259	603087	412347	313205	252479	211470
Centro	11318475	1195041	519412	329901	241499	190413	157150
Sud	11058746	916041	445330	293735	219079	174668	145222
ITALIA	4411933	1355481	798511	565813	438106	357423	301833
VIAGGI – totali							
AREE TERRITORIALI	5%	10%	15%	20%	25%	30%	35%
Trimestri 1, 2 e 4							
Nord	12132927	2746227	486485	241228	159548	119075	94951
Centro	6661629	3504556	1024796	320350	174660	119005	90069
Sud	12447163	5489841	839081	262810	151486	106118	81607
ITALIA	17214701	1158120	437023	267721	192824	150643	123594
Trimestre 3							
Nord	8979083	1727452	461675	253107	173613	131995	106438
Centro	5163606	2789360	922456	312606	171473	116819	88357
Sud	8652150	1886577	414492	215142	144523	108694	87071
ITALIA	13027105	1130042	434385	266809	192346	150338	123377
Anno							
Nord	14930409	845970	389000	252159	186490	147946	122601
Centro	15650974	1567104	426896	242114	168697	129395	104931
Sud	18516909	777189	326781	206443	150835	118817	98008
ITALIA	4182121	823912	454479	313692	239485	193666	162562

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 B.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 B.7 presenta gli stessi coefficienti con riferimento alle stime del numero di viaggi.

Sulla base dei dati riportati in tali tavole, 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 dalla tavola C.5 e si può utilizzare la seguente espressione:

$$\hat{\sigma}(\hat{Y}_3) = 10795,180 + 0,715393 \hat{Y}_3 - 0,000000157674 (\hat{Y}_3)^2.$$

Esempio 2

La valutazione degli errori di campionamento delle stime mediante un metodo approssimato si può ottenere utilizzando i dati dei prospetti B.6 e B.8. Il prospetto B.6 fa riferimento alle stime del numero di notti; il prospetto B.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 B.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 16.563.955 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 B.6, si consideri la stima del *numero di notti per lavoro*, riferita al totale per l'anno 2000, pari a 58.230.000.

Dal prospetto B.6, considerando le notti-lavoro ed esaminando l'ultima riga corrispondente al totale Italia riferito all'anno, si può osservare che il valore 58.230.000 risulta essere compreso tra i valori 32.525.952 e 74.381.349.

Da ciò si può facilmente desumere che il valore dell'errore relativo percentuale corrispondente è compreso tra il 5 ed il 10%.

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

Esempio 1

Ammettiamo, per esempio, che il numero dei turisti in Italia nel trimestre ottobre-dicembre è pari a 8.965.000.

Si cerca il livello di stima che più si avvicina a 8.965.000 nella prima colonna del prospetto B.4.1, che è pari a 7.500.000. Dalla quinta colonna del prospetto, con riferimento all'Italia, si ricava il corrispondente errore relativo percentuale della stima considerata, che è pari a 1,9%.

L'errore assoluto sarà:

$$\sigma(8.965.000) = 1,9/100 \times 8.965.000 = 170.335$$

L'intervallo di confidenza avrà come estremi:

$$8.965.000 - (2 \times 170.335) = 8.624.330$$

$$8.965.000 + (2 \times 170.335) = 9.305.670$$

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 7.500.000 e 10.000.000 ai quali corrispondono i valori 1,9 e 1,6 percentuali.

L'errore corrispondente a 8.965.000 è pari a:

$$\sigma(8.965.000) = 1,9 - ((1,9 - 1,6) / (10.000.000 - 7.500.000)) \times (8.965.000 - 7.500.000) = 1,72\%$$

Il corrispondente errore assoluto è $1,72/100 \times 8.965.000 = 154.575$ e l'intervallo di confidenza avrà come estremi:

$$8.965.000 - (2 \times 154.575) = 8.655.850$$

$$8.965.000 + (2 \times 154.575) = 9.274.150$$

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 8.965.000 si ha:

$\log \varepsilon^2 = 11,942478 - 1,2539673 \times \log (8.965.000) = -8,132082$ Da questa il valore di ε è facilmente desumibile, calcolando la radice quadrata dell'antilogaritmo dell'espressione di sopra.

Infatti, da $\log (\varepsilon^2) = -8,132082$ si ha:

$$\varepsilon^2 = \exp(-8,132082)$$

e dunque

$$\varepsilon = 0,017145$$

L'errore relativo percentuale sarà dunque pari a 1,71 %.