

# INDAGINE TRIMESTRALE 'VIAGGI E VACANZE' <sup>1</sup>

## FILES STANDARD 1999

### INDICE

<b>1. PREMESSA .....</b>	<b>2</b>
<b>2. FINALITÀ E CARATTERISTICHE DELL'INDAGINE .....</b>	<b>2</b>
<b>3. DEFINIZIONI .....</b>	<b>3</b>
<b>4. AVVERTENZE PER L'UTILIZZAZIONE DEI FILES.....</b>	<b>4</b>
<b>5. COSTRUZIONE DELLE STIME ED ERRORI DI CAMPIONAMENTO.....</b>	<b>5</b>
<b>6. TRACCIATI .....</b>	<b>7</b>
6.1 - Tracciato record del file 'individui' tipo A (107 bytes, 40.690 records) .....	7
6.2 - Tracciato record del file 'individui' tipo B (107 bytes, 40.690 records).....	8
6.3 - Tracciato record del file 'viaggi' tipo A (98 bytes, 15.222 records).....	9
6.4 - Tracciato record del file 'viaggi' tipo B (98 bytes, 15.222 records).....	10
<b>7. MODALITÀ DELLE VARIABILI.....</b>	<b>11</b>
<b>APPENDICE A - Il questionario C.A.T.I., anno 1999 .....</b>	<b>177</b>
<b>APPENDICE B - Strategia di campionamento e valutazione degli errori campionari.....</b>	<b>18</b>

---

<sup>1</sup> I principali risultati dell'indagine per il 1999 sono stati pubblicati nel volume ISTAT, 'I viaggi in Italia e all'estero nel 1999', Collana Informazioni.

## 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 legge 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 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.

## 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 un'indagine trimestrale telefonica C.A.T.I. (Computer Assisted Telephone Interview) 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: in ciascun periodo vengono rilevate, per tutti i componenti delle famiglie del campione, informazioni sui viaggi conclusi nel trimestre, indipendentemente dalla loro data di inizio.

È la prima volta che vengono raccolte informazioni trimestrali sui viaggi con almeno un pernottamento realizzati dai residenti per ragioni di lavoro o di vacanza e tale peculiarità permette di disporre sistematicamente, insieme ai dati sul movimento dei clienti nelle strutture ricettive<sup>2</sup>, anche di informazioni sul numero e sulle caratteristiche dei viaggi e dei turisti.

Rispetto alle precedenti indagini ISTAT sulle vacanze degli italiani, esperienza iniziata nel 1959, proseguita con cadenza pluriennale sino al 1993<sup>3</sup> e successivamente con cadenza annuale nell'indagine 'Aspetti della vita quotidiana', l'indagine trimestrale telefonica rileva, per la prima volta, anche informazioni sulle vacanze di breve durata (1-3 notti) e sul turismo per motivi di lavoro con carattere di continuità<sup>4</sup>.

---

<sup>2</sup> 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).

<sup>3</sup> 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.).

<sup>4</sup> Con riferimento alle vacanze di 'breve' durata ed ai viaggi di lavoro sono tuttora in atto alcune sperimentazioni tese a migliorare la qualità delle stime. La tecnica di indagine adottata, che rileva i viaggi 'ex post', è caratterizzata da specifiche problematiche che sono alla base dell'attuale sperimentazione. Tra queste, un primo aspetto riguarda la scelta del periodo di riferimento per cui l'intervistato è chiamato a ricordare il numero dei viaggi effettuati; un secondo

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.

### 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, nonché dai luoghi abitualmente frequentati. 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 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.
- **Il motivo del viaggio di vacanza, destinazione, mezzo di trasporto e 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.

I dati individuali fanno riferimento alle caratteristiche della persona 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;

---

aspetto riguarda le difficoltà che possono presentarsi da parte dell'intervistato nel distinguere i flussi turistici da quelli non turistici, in base ai quesiti ed alle definizioni adottate.

- **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, inabile al lavoro, benestante, detenuto, ecc.).
- **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.
  - **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 esterni all'ISTAT due tipi di file:

- file 'individui'
- file 'viaggi'.

Il file 'individui' è composto da un record per ciascun individuo; il record contiene informazioni sulla famiglia (coll. 1-59), informazioni strutturali sull'individuo (coll. 65-85) ed informazioni sui viaggi effettuati dall'individuo (coll. 91-107).

Il file 'viaggi' è composto da un record per ciascun viaggio realizzato da un singolo individuo; il record contiene informazioni dettagliate sul viaggio.

I files sono stati costruiti mettendo in sequenza i quattro files trimestrali. In ogni record è specificato il trimestre di rilevazione.

Per un corretto utilizzo dei files si precisa che:

- la famiglia è identificata attraverso il trimestre di rilevazione (TRIM) e il numero progressivo della famiglia nel trimestre (IDFAM1 nei file di tipo A e IDFAM2 nei file di tipo B);
- l'individuo è identificato attraverso il trimestre di rilevazione (TRIM), il numero progressivo della famiglia nel trimestre (IDFAM1 o IDFAM2) e il codice identificativo dell'individuo all'interno di ogni famiglia (IDIND);
- il viaggio è identificato attraverso il trimestre di rilevazione (TRIM), il numero progressivo della famiglia nel trimestre (IDFAM1 o IDFAM2), il codice identificativo dell'individuo all'interno di ogni famiglia (IDIND) e il numero progressivo del viaggio all'interno della famiglia (IDVIA).

A seconda della selezione che si opera nel file 'individui' è possibile effettuare elaborazioni sulle seguenti unità di analisi: individui e famiglie. Volendo elaborare statistiche sulle famiglie, occorre selezionare un solo individuo per famiglia, l'individuo identificato dal codice IDIND=01.

Nel file 'viaggi', ogni record rappresenta un viaggio di un singolo individuo. Pertanto, nel caso in cui due o più componenti della famiglia abbiano effettuato un viaggio insieme (con stessa data di inizio, stessa durata, stessa destinazione ed uguali caratteristiche del viaggio), sono presenti tanti record quanti sono i partecipanti al viaggio. In ogni record è comunque riportato il numero dei partecipanti allo stesso viaggio (NPART) ed un codice (INSIEME) che permette di identificare i viaggi

effettuati insieme da più componenti della famiglia (numero progressivo nella famiglia, che identifica i viaggi fatti insieme da componenti della famiglia stessa).

Il trimestre di rilevazione (TRIM), il progressivo della famiglia nel trimestre (IDFAM) e il progressivo dell'individuo all'interno della famiglia (IDIND), presenti sia nel file 'individui' che nel file 'viaggi', permettono l'aggancio dei record riconducibili ad uno stesso individuo.

Per distinguere i viaggi di lavoro dai viaggi di vacanza è presente un codice (TIPOVIA) che individua la tipologia del viaggio.

Per una corretta utilizzazione dei dati, si rammenta che l'indagine rileva i viaggi che si concludono in ciascun trimestre di rilevazione, indipendentemente dalla data di inizio.

#### FILES DI DATI DISPONIBILI

##### Dati per regione (tipo A):

IND99A.DAT - file 'individui' (IA)	lunghezza record 107 bytes	numero record 40.690
VIA99A.DAT - file 'viaggi' (VA)	lunghezza record 98 bytes	numero record 15.222

##### Dati per tipologia socio-demografica del Comune (tipo B):

IND99B.DAT - file 'individui' (IB)	lunghezza record 107 bytes	numero record 40.690
VIA99B.DAT - file 'viaggi' (VB)	lunghezza record 98 bytes	numero record 15.222

## 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, calcolati 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 il valore di COE è pari a 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 relative alle famiglie, agli individui e ai viaggi.

Si fa presente che, mentre i viaggi di ciascuna rilevazione trimestrale possono essere sommati per ottenere una stima annuale, 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 per parametri di diversa natura con riferimenti temporali e territoriali differenti. In particolare, il disegno campionario consente stime trimestrali riferite all'intero territorio nazionale, mentre prevede stime relative all'intero anno con i seguenti riferimenti 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.

Per garantire la riservatezza e per limiti connessi alla numerosità campionaria, non è possibile fornire contemporaneamente i dati sulle suddivisioni territoriali suindicate. Per questo motivo, sia del file 'individui' che del file 'viaggi' vengono fornite due versioni (tipo A e tipo B), una con i codici di regione e l'altra con i codici di tipologia socio-demografica del Comune. Nelle due versioni A e B, una stessa famiglia è individuata da un diverso 'numero progressivo della famiglia nel trimestre'.

Nel diffondere i risultati di un'indagine campionaria occorre fornire agli utilizzatori le informazioni necessarie per valutare l'attendibilità delle stime ottenibili. 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. TRACCIATI

### 6.1 - Tracciato record del file 'individui' tipo A (107 bytes, 40.690 records)

COLONNE	BYTES	VARIABILE	DESCRIZIONE
<b>IDENTIFICATIVI</b>			
1-2	(2)	TR	Tipo record = 'IA'
3-4	(2)	TRIM	Trimestre di rilevazione
5-8	(4)	ANNRIL	Anno di rilevazione = '1999'
9-12	(4)	IDFAM1	Progressivo tipo A della famiglia nel trimestre
13-14	(2)	IDIND	Numero identificativo dell'individuo all'interno della famiglia
15-22	(8)	COE	Coefficiente di riporto all'universo
23	(1)	RIP	Ripartizione geografica di residenza
24-26	(3)	REG	Regione di residenza
27-31	(5)	-	Blank
<b>INFORMAZIONI SULLA FAMIGLIA</b>			
32-33	(2)	NCOMP	Numero componenti della famiglia
34-44	(11)	-	Blank
45	(1)	FVIA	Presenza di viaggi nella famiglia nel trimestre
46-47	(2)	FNVIA	Numero di viaggi effettuati nella famiglia nel trimestre
48	(1)	FLAV	Presenza di viaggi di lavoro nella famiglia nel trimestre
49-50	(2)	FNLAV	Numero di viaggi di lavoro effettuati nella famiglia nel trimestre
51	(1)	FVAC	Presenza di viaggi di vacanza nella famiglia nel trimestre
52-53	(2)	FNVAC	Numero di viaggi di vacanza effettuati nella famiglia nel trimestre
54	(1)	FBREVE	Presenza di vacanze 'brevi' (1-3 notti) nella famiglia nel trimestre
55-56	(2)	FNBREVE	Numero di vacanze 'brevi' effettuate nella famiglia nel trimestre
57	(1)	FLUNGA	Presenza di vacanze 'lunghe' (4 notti e più) nella famiglia nel trimestre
58-59	(2)	FNLUNGA	Numero di vacanze 'lunghe' effettuate nella famiglia nel trimestre
60-64	(5)	-	Blank
<b>INFORMAZIONI STRUTTURALI SULL'INDIVIDUO</b>			
65-66	(2)	RELPAR	Relazione di parentela con l'individuo che ha effettuato l'intervista
67	(1)	SESSO	Sesso
68-70	(3)	ETA	Età
71-76	(6)	-	Blank
77	(1)	STACIV	Stato civile
78-79	(2)	ISTR	Titolo di studio ( <i>se età uguale o maggiore di 6 anni</i> )
80-81	(2)	COND	Condizione professionale ( <i>se età uguale o maggiore di 14 anni</i> )
82-83	(2)	POSIZ	Posizione nella professione ( <i>se occupato</i> )
84	(1)	ATECO	Settore di attività economica ( <i>se occupato</i> )
85	(1)	CITT	Cittadinanza
86-90	(5)	-	Blank
<b>INFORMAZIONI SUI VIAGGI DELL'INDIVIDUO</b>			
91	(1)	IVIA	Presenza di viaggi dell'individuo nel trimestre
92-93	(2)	INVIA	Numero di viaggi effettuati dall'individuo nel trimestre
94	(1)	ILAV	Presenza di viaggi di lavoro dell'individuo nel trimestre
95-96	(2)	INLAV	Numero di viaggi di lavoro effettuati dall'individuo nel trimestre
97	(1)	IVAC	Presenza di viaggi di vacanza dell'individuo nel trimestre
98-99	(2)	INVAC	Numero di viaggi di vacanza effettuati dall'individuo nel trimestre
100	(1)	IBREVE	Presenza di vacanze 'brevi' (1-3 notti) dell'individuo nel trimestre
101-102	(2)	INBREVE	Numero di viaggi di vacanza 'breve' effettuati dall'individuo nel trimestre
103	(1)	ILUNGA	Presenza di vacanze 'lunghe' (4 notti e più) dell'individuo nel trimestre
104-105	(2)	INLUNGA	Numero di viaggi di vacanza 'lunga' effettuati dall'individuo nel trimestre
106-107	(2)	NOVAC	Motivo per cui l'individuo non ha effettuato vacanze nel trimestre

## 6.2 - Tracciato record del file 'individui' tipo B (107 bytes, 40.690 records)

COLONNE	BYTES	VARIABILE	DESCRIZIONE
<b>IDENTIFICATIVI</b>			
1-2	(2)	TR	Tipo record = 'IB'
3-4	(2)	TRIM	Trimestre di rilevazione
5-8	(4)	ANNRIL	Anno di rilevazione = '1999'
9-12	(4)	IDFAM2	Progressivo tipo B della famiglia nel trimestre
13-14	(2)	IDIND	Numero identificativo dell'individuo all'interno della famiglia
15-22	(8)	COE	Coefficiente di riporto all'universo
23	(1)	RIP	Ripartizione geografica di residenza
24	(1)	DOM	Tipologia socio-demografica del Comune di residenza
25-31	(7)	-	Blank
<b>INFORMAZIONI SULLA FAMIGLIA</b>			
32-33	(2)	NCOMP	Numero componenti della famiglia
34-44	(11)	-	Blank
45	(1)	FVIA	Presenza di viaggi nella famiglia nel trimestre
46-47	(2)	FNVIA	Numero di viaggi effettuati nella famiglia nel trimestre
48	(1)	FLAV	Presenza di viaggi di lavoro nella famiglia nel trimestre
49-50	(2)	FNLAV	Numero di viaggi di lavoro effettuati nella famiglia nel trimestre
51	(1)	FVAC	Presenza di viaggi di vacanza nella famiglia nel trimestre
52-53	(2)	FNVAC	Numero di viaggi di vacanza effettuati nella famiglia nel trimestre
54	(1)	FBREVE	Presenza di vacanze 'brevi' (1-3 notti) nella famiglia nel trimestre
55-56	(2)	FNBREVE	Numero di vacanze 'brevi' effettuate nella famiglia nel trimestre
57	(1)	FLUNGA	Presenza di vacanze 'lunghe' (4 notti e più) nella famiglia nel trimestre
58-59	(2)	FNLUNGA	Numero di vacanze 'lunghe' effettuate nella famiglia nel trimestre
60-64	(5)	-	Blank
<b>INFORMAZIONI STRUTTURALI SULL'INDIVIDUO</b>			
65-66	(2)	RELPAR	Relazione di parentela con l'individuo che ha effettuato l'intervista
67	(1)	SESSO	Sesso
68-70	(3)	ETA	Età
71-76	(6)	-	Blank
77	(1)	STACIV	Stato civile
78-79	(2)	ISTR	Titolo di studio <i>(se età uguale o maggiore di 6 anni)</i>
80-81	(2)	COND	Condizione professionale <i>(se età uguale o maggiore di 14 anni)</i>
82-83	(2)	POSIZ	Posizione nella professione <i>(se occupato)</i>
84	(1)	ATECO	Settore di attività economica <i>(se occupato)</i>
85	(1)	CITT	Cittadinanza
86-90	(5)	-	Blank
<b>INFORMAZIONI SUI VIAGGI DELL'INDIVIDUO</b>			
91	(1)	IVIA	Presenza di viaggi dell'individuo nel trimestre
92-93	(2)	NVIA	Numero di viaggi effettuati dall'individuo nel trimestre
94	(1)	ILAV	Presenza di viaggi di lavoro dell'individuo nel trimestre
95-96	(2)	NLAV	Numero di viaggi di lavoro effettuati dall'individuo nel trimestre
97	(1)	IVAC	Presenza di viaggi di vacanza dell'individuo nel trimestre
98-99	(2)	NVAC	Numero di viaggi di vacanza effettuati dall'individuo nel trimestre
100	(1)	IBREVE	Presenza di vacanze 'brevi' (1-3 notti) dell'individuo nel trimestre
101-102	(2)	NBREVE	Numero di viaggi di vacanza 'breve' effettuati dall'individuo nel trimestre
103	(1)	ILUNGA	Presenza di vacanze 'lunghe' (4 notti e più) dell'individuo nel trimestre
104-105	(2)	NLUNGA	Numero di viaggi di vacanza 'lunga' effettuati dall'individuo nel trimestre
106-107	(2)	NOVAC	Motivo per cui l'individuo non ha effettuato vacanze nel trimestre



### 6.3 - Tracciato record del file 'viaggi' tipo A (98 bytes, 15.222 records)

COLONNE	BYTES	VARIABILE	DESCRIZIONE
<b>IDENTIFICATIVI</b>			
1-2	(2)	TR	Tipo record = 'VA'
3-4	(2)	TRIM	Trimestre di rilevazione
5-8	(4)	ANNRIL	Anno di rilevazione = '1999'
9-12	(4)	IDFAM1	Progressivo tipo A della famiglia nel trimestre
13-14	(2)	IDIND	Numero identificativo dell'individuo all'interno della famiglia
15-22	(8)	COE	Coefficiente di riporto all'universo
23	(1)	RIP	Ripartizione geografica di residenza
24-26	(3)	REG	Regione di residenza
27-31	(5)	-	Blank
<b>INFORMAZIONI SUI VIAGGI</b>			
32-34	(3)	IDVIA	Numero progressivo del viaggio all'interno della famiglia
35-38	(4)	TIPOVIA	Tipologia del viaggio
39-43	(5)	-	Blank
44-45	(2)	INSIEME	Numero progressivo nella famiglia: identifica i viaggi fatti insieme da componenti della famiglia
46-47	(2)	NPART	Numero di partecipanti allo stesso viaggio
48-69	(22)	REL1-REL11	Relazione di parentela dei partecipanti al viaggio, rispetto all'intervistato. Undici campi da 2 bytes
70-74	(5)	-	Blank
75-82	(8)	INIZIO	Data di inizio del viaggio (ggmmaaaa)
83-85	(3)	DURATA	Durata del viaggio (numero di notti)
86-87	(2)	MOTIVO	Motivo prevalente del viaggio di vacanza ( <i>se viaggio di vacanza</i> )
88-89	(2)	TIPO	Tipo della vacanza di piacere/svago ( <i>se il motivo del viaggio di vacanza è piacere/svago</i> )
90-91	(2)	ORGAN	Organizzazione del viaggio
92-93	(2)	MEZZO	Mezzo prevalente utilizzato
94-95	(2)	ALLOG	Tipo di alloggio prevalente
96-98	(3)	DEST	Regione (o Stato estero) di destinazione principale

#### 6.4 - Tracciato record del file 'viaggi' tipo B (98 bytes, 15.222 records)

COLONNE	BYTES	VARIABILE	DESCRIZIONE
<b>IDENTIFICATIVI</b>			
1-2	(2)	TR	Tipo record = 'VB'
3-4	(2)	TRIM	Trimestre di rilevazione
5-8	(4)	ANNRIL	Anno di rilevazione = '1999'
9-12	(4)	IDFAM2	Progressivo tipo B della famiglia nel trimestre
13-14	(2)	IDIND	Numero identificativo dell'individuo all'interno della famiglia
15-22	(8)	COE	Coefficiente di riporto all'universo
23	(1)	RIP	Ripartizione geografica di residenza
24	(1)	DOM	Tipologia socio-demografica del Comune di residenza
25-31	(7)	-	Blank
<b>INFORMAZIONI SUI VIAGGI</b>			
32-34	(3)	IDVIA	Numero progressivo del viaggio all'interno della famiglia
35-38	(4)	TIPOVIA	Tipologia del viaggio
39-43	(5)	-	Blank
44-45	(2)	INSIEME	Numero progressivo nella famiglia: identifica i viaggi fatti insieme da componenti della famiglia
46-47	(2)	NPART	Numero di partecipanti allo stesso viaggio
48-69	(22)	REL1-REL11	Relazione di parentela dei partecipanti al viaggio, rispetto all'intervistato. Undici campi da 2 bytes
70-74	(5)	-	Blank
75-82	(8)	INIZIO	Data di inizio del viaggio (ggmmaaaa)
83-85	(3)	DURATA	Durata del viaggio (numero di notti)
86-87	(2)	MOTIVO	Motivo prevalente del viaggio di vacanza (se viaggio di vacanza)
88-89	(2)	TIPO	Tipo della vacanza di piacere/svago (se il motivo del viaggio di vacanza è piacere/svago)
90-91	(2)	ORGAN	Organizzazione del viaggio
92-93	(2)	MEZZO	Mezzo prevalente utilizzato
94-95	(2)	ALLOG	Tipo di alloggio prevalente
96-98	(3)	DEST	Regione (o Stato estero) di destinazione principale

## 7. MODALITA' DELLE VARIABILI

TR (tipo di record)

IA=record individuale tipo A (contiene REG=regione di residenza)

IB=record individuale tipo B (contiene DOM=tipologia socio-demografica del Comune di residenza)

VA=record viaggi tipo A (contiene REG=regione di residenza)

VB=record viaggi tipo B (contiene DOM= tipologia socio-demografica del Comune di residenza)

TRIM (trimestre di rilevazione)

01=primo trimestre (gennaio-marzo)

02=secondo trimestre (aprile-giugno)

03=terzo trimestre (luglio-settembre)

04=quarto trimestre (ottobre-dicembre)

RIP (ripartizione geografica di residenza)

1=Italia nord-occidentale

2=Italia nord-orientale

3=Italia centrale

4=Italia meridionale

5=Italia insulare

DOM (tipologia socio-demografica del Comune)

1=Comune centro dell'area metropolitana

2=Comune periferia dell'area metropolitana

3=Comune aventi fino a 2.000 abitanti

4=Comune da 2.001 a 10.000 abitanti

5=Comune da 10.001 a 50.000 abitanti

6=Comune con oltre 50.000 abitanti

REG (regione di residenza)

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

FVIA, FLAV, FVAC, FBREVE, FLUNGA (presenza di viaggi nella famiglia nel trimestre)

0=nessun viaggio

1=almeno un viaggio

IVIA, ILAV, IVAC, IBREVE, ILUNGA (presenza di viaggi dell'individuo nel trimestre)

0=nessun viaggio

1=almeno un viaggio

RELPAR (relazione di parentela con l'individuo che ha effettuato l'intervista)

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

SESSO

1=maschio

2=femmina

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

ISTR (titolo di studio massimo conseguito)

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

COND (condizione professionale)

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

POSIZ (posizione nella professione)

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

ATECO (settore di attività economica)

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

CITT (cittadinanza)

1=italiana  
2=altra  
3=apolide

NOVAC (motivo per cui l'individuo non ha effettuato vacanze)

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

TIPOVIA (tipologia del viaggio)

L199=viaggio di lavoro concluso nel primo trimestre 1999  
L299=viaggio di lavoro concluso nel secondo trimestre 1999  
L399=viaggio di lavoro concluso nel terzo trimestre 1999  
L499=viaggio di lavoro concluso nel quarto trimestre 1999  
V199=viaggio di vacanza concluso nel primo trimestre 1999  
V299=viaggio di vacanza concluso nel secondo trimestre 1999  
V399=viaggio di vacanza concluso nel terzo trimestre 1999  
V499=viaggio di vacanza concluso nel quarto trimestre 1999

REL1-REL11 (relazione di parentela dei familiari partecipanti al viaggio, rispetto all'intervistato)

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

**MOTIVO** (motivo prevalente del viaggio di vacanza)

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

**TIPO** (tipo prevalente della vacanza di piacere/svago)

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

**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

**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

ALLOG (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

DEST (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=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



## APPENDICE A - Il questionario C.A.T.I., anno 1999

Il questionario 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, 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 informazioni rilevate tramite questionario e quelle fornite nei files standard, come descritte negli appositi tracciati.**

## APPENDICE B - Strategia di campionamento e valutazione degli errori campionari

### 1. Introduzione

La popolazione di interesse dell'indagine in oggetto – 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 circa 3.500 famiglie che viene selezionato in modo indipendente dai campioni delle altre tre rilevazioni.

L'indagine ha la finalità di fornire stime di parametri di diversa natura (totali, medie, rapporti, frequenze assolute e relative) con diversi riferimenti temporali e territoriali. Le stime relative ai singoli trimestri si riferiscono all'intero territorio nazionale, mentre i domini territoriali di riferimento per le stime relative all'intero anno sono:

- l'intero territorio nazionale;
- le tre grandi ripartizioni geografiche (Nord, Centro e 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<sup>5</sup>.

### 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,

---

<sup>5</sup> ISTAT, 'I sistemi locali del lavoro 1991', Argomenti n. 10 – 1997.

la sezione di censimento, la via, il numero civico e infine l'ampiezza del comune di appartenenza definita sia in termini demografici sia in termini di numero di indirizzi.

I principali problemi che derivano dall'utilizzo della lista telefonica come base di campionamento per le indagini sulle famiglie, sono legati alla non perfetta coincidenza tra la popolazione oggetto d'indagine e l'insieme delle unità contenute nella lista; indicando con L ed U rispettivamente la lista e la popolazione d'interesse, le possibili situazioni in cui gli insiemi L e U differiscono sono in generale le seguenti:

- i) *sovracopertura*, quando alcuni elementi dell'insieme L non appartengono all'insieme U;
- ii) *sottocopertura*, nel caso in cui alcuni elementi di U non sono contenuti in L;
- iii) *duplicazione di alcune unità*, se alcuni elementi di U sono presenti più volte in L;
- iv) *grappoli di unità*, quando alcuni elementi dell'insieme L contengono grappoli di elementi dell'insieme U.

La lista telefonica utilizzata per l'indagine presenta in varia misura tutte le situazioni elencate; tuttavia, i fenomeni che appaiono più rilevanti sono 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 vincolata<sup>6</sup>, 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 A.1 e A.2 vengono presentati i tassi di sostituzione, rispettivamente per regione e per tipologia comunale, osservati nel trimestre di indagine ottobre-dicembre.

### 3. Disegno di campionamento

#### 3.1 Descrizione generale

Il disegno di campionamento è un disegno stratificato a grappoli, in cui ciascun grappolo è un indirizzo telefonico presente nell'archivio di selezione. Tutti gli individui appartenenti ai grappoli campione, cioè appartenenti alle famiglie a cui corrispondono i numeri selezionati, vengono inclusi nel campione.

Gli indirizzi telefonici sono stati stratificati, all'interno di ogni regione geografica, secondo la tipologia socio-demografica dei comuni (descritta nel paragrafo 1) di appartenenza degli indirizzi. Poiché tale informazione non è presente nell'archivio di selezione, è stato necessario integrare l'archivio attribuendo ad ogni record il relativo codice di tipologia.

La determinazione del numero totale di unità campionarie e la sua allocazione tra gli strati, è in genere, per un'indagine ad obiettivi plurimi come quella in esame, un'operazione complessa. E' poco realistico, infatti, pensare di poter definire un campione che assicuri prefissati livelli di precisione a

<sup>6</sup> Nella letteratura in lingua anglosassone sul tema sono noti come *Calibration Estimators* (Deville e Särndal "Calibration Estimators in Survey Sampling", 1992, Journal of the American Statistical Association)

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. In particolare, 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 uguale 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 A.1 e A.2 sono riportati, rispettivamente per regione e per tipologia di comune, le numerosità campionarie teoriche e i tassi di sostituzione riferiti alla rilevazione del trimestre ottobre-dicembre.

**Prospetto A.1 - Distribuzione regionale del campione e tassi di sostituzione\***

REGIONI	INDIRIZZI CAMPIONE	TASSO DI SOSTITUZIONE TOTALE(a) (%)	TASSO DI SOSTITUZIONE AL NETTO DEI MANCATI CONTATTI (b) (%)
Piemonte	265	42,0	34,1
Valle d'Aosta	92	39,8	32,6
Lombardia	330	35,6	32,2
Trentino-Alto Adige	206	35,4	25,7
Veneto	201	36,3	29,9
Friuli-Venezia Giulia	139	33,1	28,1
Liguria	152	40,7	32,5
Emilia-Romagna	197	39,7	35,7
Toscana	213	43,9	36,3
Umbria	103	45,6	36,9
Marche	141	34,5	28,1
Lazio	219	45,3	36,6
Abruzzo	144	41,7	33,6
Molise	96	51,1	41,3
Campania	196	40,6	36,3
Puglia	187	45,1	36,3
Basilicata	99	38,8	35,7
Calabria	158	48,0	39,2
Sicilia	213	48,8	43,2
Sardegna	150	34,0	31,3
<b>ITALIA</b>	<b>3501</b>	<b>40,7</b>	<b>34,1</b>

\* Dato della rilevazione del quarto trimestre.

- (a) Il tasso di sostituzione totale è calcolato come rapporto percentuale tra il numero di famiglie sostituite ed il numero teorico di famiglie campione.
- (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.

**Prospetto A.2 - Distribuzione 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	41,1	36,6
Comuni che gravitano intorno al centro dell'area metropolitana	316	37,9	33,1
Comuni aventi fino a 2.000 abitanti	618	46,6	34,6
Comuni con 2.001-10.000 abitanti	868	41,3	34,4
Comuni con 10.001-50.000 abitanti	950	40,4	34,1
Comuni con oltre 50.000 abitanti	309	38,6	32,3

\* Cfr. note al prospetto A.1

**3.2. Selezione del campione**

L'estrazione degli indirizzi campione da ciascuno strato 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 ciascuna regione, con riferimento ad una data tipologia comunale, si ordinano gli indirizzi dell'Archivio Telefonico delle Famiglie per ampiezza demografica del comune, strada, numero civico e piano;
- definito il passo di estrazione regionale, si selezionano 6 punti di partenza casuali, uno per ciascuna tipologia comunale. Si procede, poi, alla selezione degli indirizzi campione sulla base del passo di estrazione regionale a partire dal punto di partenza casuale;
- ogni indirizzo selezionato dall'Archivio entra a far parte dell'elenco base, e per ognuno di essi si selezionano i tre indirizzi successivi, che entrano a far parte dell'elenco sostitutivo; in tal modo le unità sostitutive sono le più vicine possibile dal punto di vista territoriale alle rispettive unità del campione base e quindi, presumibilmente, sono anche simili dal punto di vista delle caratteristiche socio-economiche.

**4. Procedimento per il calcolo delle stime**

Le stime prodotte dall'indagine sono stime del numero di viaggi e del numero di notti trascorse fuori casa per tipologia di viaggio, nonché del numero 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 oggetto di indagine, 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 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; la distribuzione delle famiglie per numero di componenti<sup>7</sup>. 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:

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

<sup>7</sup> La distribuzione delle famiglie per numero di componenti è desunta dai dati dell'indagine multiscopo 'Aspetti della vita quotidiana' - Anno 1999.

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.

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

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

dove

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

in cui  $\underline{X}_{thjp}$  è il vettore delle  $K$  variabili ausiliarie (cfr. par. 4) 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

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

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

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

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

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

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

$$\left\{ \hat{Y} - k_p \hat{\sigma}(\hat{Y}), \hat{Y} + k_p \hat{\sigma}(\hat{Y}) \right\} \quad (11)$$

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

## 6. Presentazione sintetica degli errori campionari



Poiché a ciascuna stima  ${}_d\hat{Y}_t$  corrisponde un proprio 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:

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

dove 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 uguali 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 A.3.1 (per i trimestri) e A.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 A.4.1 (per i trimestri) e A.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 (A.4.1 o A.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:

$$\hat{\varepsilon}(\hat{Y}_t) = \hat{\varepsilon}(\hat{Y}_t^{k-1}) + \frac{\hat{\varepsilon}(\hat{Y}_t^k) - \hat{\varepsilon}(\hat{Y}_t^{k-1})}{\hat{Y}_t^k - \hat{Y}_t^{k-1}} (\hat{Y}_t - \hat{Y}_t^{k-1}) \quad (13)$$

dove:  $\hat{Y}_t^{k-1}$  e  $\hat{Y}_t^k$  sono i valori delle stime entro i quali è compresa la stima d'interesse  $\hat{Y}_t$  e  $\hat{\varepsilon}(\hat{Y}_t^{k-1})$  e  $\hat{\varepsilon}(\hat{Y}_t^k)$  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 di totali di variabili quantitative, 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 (13) 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 (13) 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 dell'anno 1999 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 di viaggi di lavoro;
- numero di viaggi di vacanza.

Analogamente al caso precedente, dall'analisi dei coefficienti di regressione ottenuti è emerso che i coefficienti relativi al primo, al secondo e al quarto trimestre sono sostanzialmente uguali tra loro, mentre differiscono significativamente da quelli relativi al terzo trimestre; ciò è spiegabile considerando il fatto che i fenomeni indagati presentano andamenti difforni nel terzo trimestre rispetto agli altri tre.

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

I prospetti A.5 e A.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 quattro 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 A.6 e A.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  $\varepsilon^*$ , mentre le stime che assumono valori inferiori a  $\hat{Y}_t^*$  presentano valori dell'errore relativo superiori a  $\varepsilon^*$ . In tali prospetti, articolati per gruppo di stime, trimestre e area territoriale, i valori prefissati di  $\varepsilon^*$  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 A.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	11,3620	-1,22311	93,92	8,05942	-0,88486	85,56
Centro	11,0790	-1,21703	92,72	8,81895	-0,95965	88,17
Sud	11,0009	-1,22196	92,37	9,65127	-1,05234	84,85
<b>ITALIA</b>	<b>11,8277</b>	<b>-1,24801</b>	<b>94,15</b>	<b>8,64314</b>	<b>-0,93361</b>	<b>88,41</b>

**Prospetto A.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	12,0954	-1,25295	93,96
Centro	11,7844	-1,24724	93,18
Sud	11,4527	-1,23209	92,30
<b>ITALIA</b>	<b>12,4985</b>	<b>-1,27013</b>	<b>94,10</b>

**Prospetto A.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	68,70	61,50	57,70	76,60	70,3	71,0	68,0	74,0
30000	53,60	48,00	45,00	59,50	58,8	58,4	55,0	61,2
40000	45,00	40,30	37,80	49,70	51,8	50,9	47,2	53,5
50000	39,20	35,20	32,90	43,30	46,9	45,7	42,0	48,2
60000	35,10	31,50	29,50	38,60	43,3	41,9	38,2	44,3
70000	31,90	28,70	26,80	35,10	40,4	38,9	35,2	41,2
80000	29,40	26,40	24,70	32,30	38,1	36,5	32,8	38,7
90000	27,40	24,60	23,00	30,00	36,2	34,5	30,8	36,7
100000	25,70	23,10	21,60	28,10	34,5	32,8	29,2	34,9
200000	16,80	15,10	14,10	18,20	25,4	23,5	20,3	25,3
300000	13,10	11,80	11,00	14,10	21,2	19,4	16,4	20,9
400000	11,00	9,90	9,20	11,80	18,7	16,9	14,1	18,3
500000	9,60	8,70	8,10	10,30	16,9	15,2	12,5	16,5
750000	7,50	6,80	6,30	8,00	14,2	12,5	10,1	13,6
1000000	6,30	5,70	5,30	6,70	12,5	10,9	8,7	11,9
2000000	4,10	3,70	3,50	4,30	9,2	7,8	6,0	8,6
3000000	3,20	2,90	2,70	3,40	7,7	6,4	4,9	7,1
4000000	2,70	2,40	2,30	2,80	6,7	5,6	4,2	6,2
5000000	2,30	2,10	2,00	2,40	6,1	5,0	3,7	5,6
7500000	1,80	1,70	1,50	1,90	5,1	4,1	3,0	4,7
15000000	1,20	-	1,00	1,20	3,8	-	2,1	3,4
20000000	1,00	-	0,80	1,00	3,3	-	1,8	2,9
25000000	0,90	-	-	0,90	3,0	-	-	2,7

**Prospetto A.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	85,5	75,3	68,8	96,1
30000	66,3	58,5	53,6	74,3
40000	55,4	48,9	44,9	61,9
50000	48,2	42,5	39,1	53,7
60000	43,0	37,9	34,9	47,8
70000	39,0	34,5	31,8	43,4
80000	35,9	31,7	29,3	39,8
90000	33,3	29,5	27,2	37,0
100000	31,2	27,6	25,5	34,6
200000	20,2	17,9	16,6	22,3
300000	15,7	13,9	13,0	17,2
400000	13,1	11,6	10,9	14,3
500000	11,4	10,1	9,5	12,4
750000	8,8	7,9	7,4	9,6
1000000	7,4	6,6	6,2	8,0
2000000	4,8	4,3	4,0	5,2
3000000	3,7	3,3	3,1	4,0
4000000	3,1	2,8	2,6	3,3
5000000	2,7	2,4	2,3	2,9
7500000	2,1	1,9	1,8	2,2
15000000	1,4	-	1,2	1,4
20000000	1,1	-	1,0	1,2
25000000	1,0	-	-	1,0

**Prospetto A.5 - Valori dei coefficienti a, b, c e di  $R^2$  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**

<b>NOTTI – lavoro</b>				
<b>AREE TERRITORIALI</b>	<b>a</b>	<b>b</b>	<b>c</b>	<b><math>R^2</math></b>
<b>gennaio-marzo</b>				
Nord	38358,19	0,29020	-0,0000000317300	98,96
Centro	2726,41	1,20247	-0,0000003050400	99,69
Sud	46358,50	0,29504	-0,0000000266100	93,09
ITALIA	142039,17	0,18298	-0,0000000067100	87,02
<b>luglio-settembre</b>				
Nord	103820,23	0,26296	-0,0000000184900	82,85
Centro	35043,37	0,37670	-0,0000000493530	99,73
Sud	88482,70	0,26500	-0,0000000251120	80,73
ITALIA	198029,41	0,12955	-0,0000000026900	88,14
<b>anno</b>				
Nord	131799,5	0,12769	-0,000000002716	91,15
Centro	9264,21	0,56041	-0,000000033961	97,18
Sud	111462	0,14653	-0,000000003427	91,87
ITALIA	331280,8	0,07917	-0,000000000596	89,44
<b>NOTTI vacanze</b>				
<b>AREE TERRITORIALI</b>	<b>a</b>	<b>b</b>	<b>c</b>	<b><math>R^2</math></b>
<b>gennaio-marzo</b>				
Nord	146239,43	0,14343	-0,0000000024420	95,39
Centro	60742,28	0,20511	-0,0000000101060	95,55
Sud	135731,75	0,19107	-0,0000000063240	90,90
ITALIA	236953,29	0,09910	-0,0000000009180	95,15
<b>luglio-settembre</b>				
Nord	525298,14	0,10166	-0,0000000003075	93,82
Centro	180226,14	0,16631	-0,0000000012553	97,35
Sud	163448,07	0,17945	-0,0000000012826	95,69
ITALIA	545144,43	0,07918	-0,0000000001320	95,78
<b>anno</b>				
Nord	592104,6	0,07179	-0,00000000014466	93,68
Centro	240561,6	0,11819	-0,00000000062033	95,19
Sud	367433,3	0,11326	-0,00000000049886	91,84
ITALIA	756340,1	0,05377	-0,0000000005907	94,17

**Prospetto A.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**

<b>NOTTI – lavoro</b>							
<b>AREE TERRITORIALI</b>	<b>5%</b>	<b>10%</b>	<b>15%</b>	<b>20%</b>	<b>25%</b>	<b>30%</b>	<b>35%</b>
<b>gennaio-marzo</b>							
Nord	7726582	6189636	4677007	3218360	1902398	955861	505732
Centro	3780459	3616653	3452857	3289073	3125303	2961548	2797812
Sud	9394021	7560017	5753386	4006427	2414216	1229992	643194
ITALIA	20834216	13890551	7673628	3504244	1796300	1139366	823207
<b>luglio-settembre</b>							
Nord	11986034	9410105	6920587	4620347	2745822	1570961	986185
Centro	6725239	5730458	4743141	3768736	2819094	1923297	1155501
Sud	8955109	7069010	5250560	3574220	2199378	1305411	834992
ITALIA	31881564	15680065	5583199	2560573	1587778	1141249	888660
<b>anno</b>							
Nord	30210845	13729619	3979616	1712543	1052965	755891	588630
Centro	15047427	13577113	12107275	10638110	9169938	7703325	6239357
Sud	29278370	15655055	5219194	1862292	1041325	714867	542842
ITALIA	58452258	11871549	4506258	2705602	1926297	1494137	1219931
<b>NOTTI – vacanze</b>							
<b>AREE TERRITORIALI</b>	<b>5%</b>	<b>10%</b>	<b>15%</b>	<b>20%</b>	<b>25%</b>	<b>30%</b>	<b>35%</b>
<b>gennaio-marzo</b>							
Nord	39765577	20680352	6509388	2347266	1331607	920795	702114
Centro	15730403	10949674	6393321	2717457	1087088	601589	407640
Sud	23230978	15762355	8904624	3980264	1911262	1166983	826832
ITALIA	57940720	15583364	4318865	2300258	1555547	1173170	941172
<b>luglio-settembre</b>							
Nord	177617769	44118654	10204355	5255294	3515557	2637687	2109726
Centro	94179593	55414885	20126425	4571019	2088097	1331444	974651
Sud	102175025	63937602	27581481	5831335	2226635	1336839	951549
ITALIA	238385025	22868144	7590225	4490012	3183507	2465095	2010969
<b>anno</b>							
Nord	174134346	19115417	7467558	4594423	3313598	2590307	2125915
Centro	113346690	39212667	6689730	2877699	1809651	1317228	1034887
Sud	132373283	43509138	8920438	4137573	2661257	1957384	1547011
ITALIA	149480262	16031964	7822153	5161502	3849893	3069421	2551920

**Prospetto A.7 - Valori dei coefficienti a, b, c e di R<sup>2</sup> 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**

<b>VIAGGI – lavoro</b>				
<b>AREE TERRITORIALI</b>	<b>a</b>	<b>b</b>	<b>c</b>	<b>R<sup>2</sup></b>
<b>gennaio-marzo</b>				
Nord	13940,00	0,17278	-0,000000005892	97,38
Centro	9555,03	0,22146	-0,000000017570	97,85
Sud	9948,70	0,16943	-0,000000007348	98,75
ITALIA	18972,57	0,11069	-0,000000001708	98,05
<b>luglio-settembre</b>				
Nord	10992,35	0,14064	-0,000000003991	99,70
Centro	10731,76	0,22914	-0,000000015338	99,01
Sud	10593,96	0,16210	-0,000000006305	99,47
ITALIA	17145,17	0,10091	-0,000000001338	99,53
<b>anno</b>				
Nord	23264,35	0,08516	-0,0000000006946	98,32
Centro	20211,15	0,11175	-0,0000000021118	98,47
Sud	19643,31	0,08497	-0,0000000008959	99,11
ITALIA	34599,75	0,05506	-0,0000000002042	98,68
<b>VIAGGI – vacanza</b>				
<b>AREE TERRITORIALI</b>	<b>a</b>	<b>b</b>	<b>c</b>	<b>R<sup>2</sup></b>
<b>gennaio-marzo</b>				
Nord	28204,75	0,10691	-0,000000004150	93,46
Centro	17239,04	0,14823	-0,000000012788	96,45
Sud	16902,55	0,14826	-0,000000006806	96,97
ITALIA	38903,63	0,07383	-0,000000001247	94,75
<b>luglio-settembre</b>				
Nord	35175,80	0,10288	-0,000000005671	90,06
Centro	14788,72	0,16338	-0,000000019335	93,87
Sud	32293,25	0,10717	-0,000000005655	88,43
ITALIA	53401,40	0,06861	-0,000000001560	93,69
<b>anno</b>				
Nord	64412,95	0,048844	-0,0000000005113	93,58
Centro	34760,01	0,073772	-0,0000000017017	95,00
Sud	46007,10	0,063502	-0,0000000007488	93,03
ITALIA	93312,43	0,033250	-0,0000000001481	93,37

**Prospetto A.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%
<b>gennaio-marzo</b>							
Nord	20951349	12540997	4403537	465266	178103	109023	78455
Centro	9814092	6990712	4196743	1568187	284845	118516	73595
Sud	16336281	9590011	3083366	303325	122119	75870	54973
ITALIA	35842698	7701169	472922	211579	135963	100129	79235
<b>luglio-settembre</b>							
Nord	22831734	10446566	859446	182931	100149	68859	52452
Centro	11739092	8501909	5291951	2215649	397996	146786	87816
Sud	17873548	10017064	2572316	267610	119499	76556	56275
ITALIA	38383172	3935845	345997	172624	114880	86068	68806
<b>anno</b>							
Nord	51272302	1466954	357428	202333	141049	108249	87823
Centro	29564181	6942519	513820	227780	145868	107234	84768
Sud	39587234	1218446	300819	170540	118952	91317	74099
ITALIA	30360568	767235	364153	238638	177456	141241	117302
VIAGGI – vacanze							
AREE TERRITORIALI	5%	10%	15%	20%	25%	30%	35%
<b>gennaio-marzo</b>							
Nord	14192133	3569213	617796	298998	195998	145615	115797
Centro	7853081	4100278	1093917	309354	165932	112520	84981
Sud	14607278	7425265	1453253	313735	164328	110840	83548
ITALIA	20622655	1393981	506547	307409	220486	171848	140779
<b>luglio-settembre</b>							
Nord	9948143	2757367	689327	354837	236932	177542	141881
Centro	5991633	3496731	1286530	342064	164678	106638	78605
Sud	10646040	3106291	690952	340800	224107	166656	132578
ITALIA	14319971	1577544	648067	404492	293659	230427	189578
<b>anno</b>							
Nord	10150359	1243688	634732	425523	319954	256332	213808
Centro	15304262	1227536	451451	274360	196870	153473	125741
Sud	20962516	1229523	529460	336432	246446	194415	160517
ITALIA	5320592	1393630	798441	559317	430381	349744	294553



## 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 A.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 A.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 Nord per il terzo trimestre,  ${}_N\hat{Y}_3$ , si possono leggere i valori dei coefficienti a, b e c dal prospetto A.4 e si può utilizzare la seguente espressione:

$$\hat{\sigma}({}_N\hat{Y}_3) = 103820,23 + 0,26296{}_N\hat{Y}_3 - 0,00000001849({}_N\hat{Y}_3)^2.$$

#### Esempio 2

La valutazione degli errori di campionamento delle stime mediante un metodo approssimato si può ottenere utilizzando i dati dei prospetti A.6 e A.8. Il prospetto A.6 fa riferimento alle stime del numero di notti; il prospetto A.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 A.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 22.868144 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 A.6, si consideri la stima numero di notti per lavoro, riferita al totale Italia nel trimestre luglio-settembre, e si supponga sia pari a 14.783.000.

Dal prospetto A.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 14.783.000 risulta essere compreso tra i valori 15.680.065 e 5.583.199.

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 in Italia nel trimestre ottobre-dicembre è pari a 9.520.000.

Si cerca il livello di stima che più si avvicina a 9.520.000 nella prima colonna del prospetto A.4.1, che è pari a 7.500.000. Dall'ultima colonna del prospetto, con riferimento all'Italia, si ricava il corrispondente errore relativo percentuale della stima considerata, che è pari a 1,9%.

L'errore assoluto sarà:

$$\sigma(9.520.000) = 1,9/100 \times 9.520.000 = 180.880$$

L'intervallo di confidenza avrà come estremi:

$$9.520.000 - (2 \times 180.880) = 9.158.240$$

$$9.520.000 + (2 \times 180.880) = 9.881.760$$

### 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 15.000.000 ai quali corrispondono i valori 1,9 e 1,2 percentuali.

L'errore corrispondente a 9.520.000 è pari a:

$$\sigma(9.520.000) = 1,9 - ((1,9 - 1,2) / (15.000.000 - 7.500.000)) \times (9.520.000 - 7.500.000) = 1,7\%$$

Il corrispondente errore assoluto è  $1,7/100 \times 9.520.000 = 161.840$  e l'intervallo di confidenza avrà come estremi:

$$9.520.000 - (2 \times 161.840) = 9.196.320$$

$$9.520.000 + (2 \times 161.840) = 9.843.680$$

### 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 9.520.000 si ha:

$\log \varepsilon^2 = 11,8277 - 1,24801 \times \log (9.520.000) = -8,22645$  Da questa il valore di  $\varepsilon$  è facilmente desumibile, calcolando la radice quadrata dell'antilogaritmo dell'espressione di sopra.

Infatti, da  $\log (\varepsilon^2) = -8,22645$  si ha:

$$\varepsilon^2 = e^{-8,22645}$$

e dunque

$$\varepsilon = 0,016355$$

L'errore relativo percentuale sarà dunque pari a 1,6 %.