

I VIAGGI IN ITALIA E ALL'ESTERO NEL 1998

INDAGINE TRIMESTRALE SUI VIAGGI DELLA POPOLAZIONE RESIDENTE IN ITALIA- ANNO 1998

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.

FINALITA' 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 ricettive¹, 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 1993² e successivamente con cadenza annuale attraverso l'indagine 'Aspetti della vita quotidiana', l'indagine trimestrale telefonica rileva, per la prima

¹ 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).

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.).

volta, anche informazioni sulle vacanze di breve durata (1-3 notti) e sul turismo per motivi di lavoro con carattere di continuità³.

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.

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. 98 98
- **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:

³ 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 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.

- **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;
- **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.
- **tipo di 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);

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 strutturali sulla famiglia di appartenenza (coll. 1-33), informazioni sui viaggi effettuati dai componenti della famiglia (coll. 45-59), 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 analizzare le informazioni relative alla famiglia occorre selezionare un solo componente di ciascuna (ad esempio l'individuo 01), utilizzando il codice identificativo dell'individuo all'interno della famiglia (IDIND=01).

Nel file 'viaggi', ogni record rappresenta un viaggio di un singolo individuo. 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 pertanto 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 dei viaggi nella famiglia che si ripete per tutti i viaggi fatti insieme).

Il trimestre di rilevazione, il progressivo della famiglia nel trimestre e il progressivo dell'individuo all'interno della famiglia, 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 (TIPOVGG) che individua la tipologia del viaggio.

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

FILES DISPONIBILI

Dati per regione (tipo A):

IND98A.DAT - file 'individui' (IA)	lunghezza record 107 bytes	numero record 41.251
VIA98A.DAT - file 'viaggi' (VA)	lunghezza record 98 bytes	numero record 15.858

Dati per tipo di Comune (tipo B):

IND98B.DAT - file 'individui' (IB)	lunghezza record 107 bytes	numero record 41.251
VIA98B.DAT - file 'viaggi' (VB)	lunghezza record 98 bytes	numero record 15.858

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 di parametri di diversa natura 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:

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. sei aree basate sulla tipologia socio-demografica dei Comuni.

Per garantire la riservatezza e per limiti connessi alla numerosità campionaria non è possibile fornire contemporaneamente i dati sulle tre le 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 area 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 riporta più avanti l'appendice A del volume ISTAT 'I viaggi in Italia e all'estero nel 1998'⁴, contenente le informazioni relative al campionamento e al calcolo degli errori di stima, da cui è possibile individuare gli esempi di calcolo degli errori campionari. In essa sono accluse le tavole per il calcolo degli errori relativi ai dati contenuti nei files standard, per stime sugli individui e sui viaggi.

⁴ ISTAT 'I viaggi in Italia e all'estero nel 1998', Collana *Informazioni* - 1999.

I VIAGGI IN ITALIA E ALL'ESTERO NEL 1998
TRACCIATO RECORD DEL FILE 'INDIVIDUI' TIPO A
lrec=107 bytes, nrec=41.251

COLONNE	BYTES	VARIABILE	CONTENUTO DEL CAMPO
1-2	(2)	TR	tipo record IA=record individuale di tipo A
3-4	(2)	TRIM	trimestre di rilevazione
5-8	(4)	ANNRIL	anno di rilevazione = 1998
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
INFORMAZIONI SULLA FAMIGLIA			
32-33	(2)	NCOMP	numero componenti della famiglia
34-44	(11)	-	blank
45	(1)	FVIA	presenza/assenza di viaggi nella famiglia nel trimestre 0=nessun viaggio 1=almeno un viaggio
46-47	(2)	FNVIA	numero di viaggi effettuati nella famiglia nel trimestre
48	(1)	FLAV	presenza/assenza di viaggi di lavoro nella famiglia nel trimestre 0=nessun viaggio di lavoro 1=almeno un viaggio di lavoro
49-50	(2)	FNLAV	numero di viaggi di lavoro effettuati nella famiglia nel trimestre
51	(1)	FVAC	presenza/assenza di viaggi di vacanza nella famiglia nel trimestre 0=nessun viaggio di vacanza 1=almeno un viaggio di vacanza
52-53	(2)	FNVAC	numero di viaggi di vacanza effettuati nella famiglia nel trimestre
54	(1)	FBREVE	presenza/assenza di vacanze 'brevi' (1-3 notti) nella famiglia nel trimestre 0=nessuna vacanza 'breve' 1=almeno una vacanza 'breve'
55-56	(2)	FNBREVE	numero di vacanze 'brevi' effettuate nella famiglia nel trimestre
57	(1)	FLUNGA	presenza/assenza di vacanze 'lunghe' (4 notti e più) nella famiglia nel trimestre 0=nessuna vacanza 'lunga' 1=almeno una vacanza 'lunga'
58-59	(2)	FNLUNGA	numero di vacanze 'lunghe' effettuate nella famiglia nel trimestre
60-64	(5)	-	blank

INFORMAZIONI STRUTTURALI SULL'INDIVIDUO

65-66	(2)	RELPAR	relazione di parentela con l'individuo che ha effettuato l'intervista
67	(1)	SESSO	sex
68-70	(3)	ETA	age
71-76	(6)	-	blank
77	(1)	STACIV	state civil
78-79	(2)	ISTR	title of study (<i>se età uguale o maggiore di 6 anni</i>)
80-81	(2)	COND	professional condition (<i>se età uguale o maggiore di 14 anni</i>)
82-83	(2)	POSIZ	position in the profession (<i>se occupato</i>)
84	(1)	ATECO	sector of economic activity (<i>se occupato</i>)
85	(1)	CITT	citizenship
86-90	(5)	-	blank

INFORMAZIONI SUI VIAGGI DELL'INDIVIDUO

91	(1)	IVIA	presence/absence of trips of the individual in the trimester 0=no trip 1=at least one trip
92-93	(2)	NVIA	number of trips made by the individual in the trimester
94	(1)	ILAV	presence/absence of work trips of the individual in the trimester 0=no work trip 1=at least one work trip
95-96	(2)	NLAV	number of work trips made by the individual in the trimester
97	(1)	IVAC	presence/absence of vacation trips of the individual in the trimester 0=no vacation trip 1=at least one vacation trip
98-99	(2)	NVAC	number of vacation trips made by the individual in the trimester
100	(1)	IBREVE	presence/absence of 'short' vacations (1-3 nights) of the individual in the trimester 0=no 'short' vacation 1=at least one 'short' vacation
101-102	(2)	NBREVE	number of 'short' vacation trips made by the individual in the trimester
103	(1)	ILUNGA	presence/absence of 'long' vacations (4 nights and more) of the individual in the trimester 0=no 'long' vacation 1=at least one 'long' vacation
104-105	(2)	NLUNGA	number of 'long' vacation trips made by the individual in the trimester
106-107	(2)	NOVAC	reason for which the individual did not have a vacation in the trimester

I VIAGGI IN ITALIA E ALL'ESTERO NEL 1998
TRACCIATO RECORD DEL FILE 'INDIVIDUI' TIPO B
lrec=107, nrec=41.251

COLONNE	BYTES	VARIABILE	CONTENUTO DEL CAMPO
1-2	(2)	TR	tipo record IB=record individuale di tipo B
3-4	(2)	TRIM	trimestre di rilevazione
5-8	(4)	ANNRIL	anno di rilevazione = 1998
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	area socio-demografica del Comune (dominio)
25-31	(7)	-	blank
INFORMAZIONI SULLA FAMIGLIA			
32-33	(2)	NCOMP	numero componenti della famiglia
34-44	(11)	-	blank
45	(1)	FVIA	presenza/assenza di viaggi nella famiglia nel trimestre 0=nessun viaggio 1=almeno un viaggio
46-47	(2)	FNVIA	numero di viaggi effettuati nella famiglia nel trimestre
48	(1)	FLAV	presenza/assenza di viaggi di lavoro nella famiglia nel trimestre 0=nessun viaggio di lavoro 1=almeno un viaggio di lavoro
49-50	(2)	FNLAV	numero di viaggi di lavoro effettuati nella famiglia nel trimestre
51	(1)	FVAC	presenza/assenza di viaggi di vacanza nella famiglia nel trimestre 0=nessun viaggio di vacanza 1=almeno un viaggio di vacanza
52-53	(2)	FNVAC	numero di viaggi di vacanza effettuati nella famiglia nel trimestre
54	(1)	FBREVE	presenza/assenza di vacanze 'brevi' (1-3 notti) nella famiglia nel trimestre 0=nessuna vacanza 'breve' 1=almeno una vacanza 'breve'
55-56	(2)	FNBREVE	numero di vacanze 'brevi' effettuate nella famiglia nel trimestre
57	(1)	FLUNGA	presenza/assenza di vacanze 'lunghe' (4 notti e più) nella famiglia nel trimestre 0=nessuna vacanza 'lunga' 1=almeno una vacanza 'lunga'
58-59	(2)	FNLUNGA	numero di vacanze 'lunghe' effettuate nella famiglia nel trimestre
60-64	(5)	-	blank

INFORMAZIONI STRUTTURALI SULL'INDIVIDUO

65-66	(2)	RELPAR	relazione di parentela con l'individuo che ha effettuato l'intervista
67	(1)	SESSO	sex
68-70	(3)	ETA	age
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

INFORMAZIONI SUI VIAGGI DELL'INDIVIDUO

91	(1)	IVIA	presenza/assenza di viaggi dell'individuo nel trimestre 0=nessun viaggio 1=almeno un viaggio
92-93	(2)	NVIA	numero di viaggi effettuati dall'individuo nel trimestre
94	(1)	ILAV	presenza/assenza di viaggi di lavoro dell'individuo nel trimestre 0=nessun viaggio di lavoro 1=almeno un viaggio di lavoro
95-96	(2)	NLAV	numero di viaggi di lavoro effettuati dall'individuo nel trimestre
97	(1)	IVAC	presenza/assenza di viaggi di vacanza dell'individuo nel trimestre 0=nessun viaggio di vacanza 1=almeno un viaggio di vacanza
98-99	(2)	NVAC	numero di viaggi di vacanza effettuati dall'individuo nel trimestre
100	(1)	IBREVE	presenza/assenza di vacanze 'brevi' (1-3 notti) dell'individuo nel trimestre 0=nessuna vacanza 'breve' 1=almeno una vacanza 'breve'
101-102	(2)	NBREVE	numero di viaggi di vacanza 'breve' effettuati dall'individuo nel trimestre
103	(1)	ILUNGA	presenza/assenza di vacanze 'lunghe' (4 notti e più) dell'individuo nel trimestre 0=nessuna vacanza 'lunga' 1=almeno una vacanza 'lunga'
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

I VIAGGI IN ITALIA E ALL'ESTERO NEL 1998
TRACCIATO RECORD DEL FILE 'VIAGGI' TIPO A
lrec=98 bytes, nrec=15.858

COLONNE	BYTES	VARIABILE	CONTENUTO DEL CAMPO
1-2	(2)	TR	tipo record VA=record viaggi di tipo A
3-4	(2)	TRIM	trimestre di rilevazione
5-8	(4)	ANNRIL	anno di rilevazione = 1998
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
INFORMAZIONI SUI VIAGGI			
32-34	(3)	IDVIA	numero progressivo del viaggio all'interno della famiglia
35-38	(4)	TIPOVGG	tipologia del viaggio
39-43	(5)	-	blank
44-45	(2)	INSIEME	numero progressivo che individua i viaggi fatti insieme dai familiari
46-47	(2)	NPART	numero di componenti della famiglia partecipanti allo stesso viaggio
48-69	(22)	REL1-REL11	undici campi da 2 bytes. Relazione di parentela dei familiari partecipanti al viaggio, rispetto all'intervistato
70-74	(5)	-	blank
75-76	(2)	GGINIZ	giorno di inizio del viaggio
77-78	(2)	MMINIZ	mese di inizio del viaggio
79-82	(4)	AAINIZ	anno di inizio del viaggio
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_RE	regione/Stato estero di destinazione principale

INDAGINE 'VIAGGI E VACANZE' - ANNO 1998
TRACCIATO RECORD DEL FILE 'VIAGGI' TIPO B
Irec=98, nrec=15.858

COLONNE	BYTES	VARIABILE	CONTENUTO DEL CAMPO
1-2	(2)	TR	tipo record VB=record viaggi di tipo B
3-4	(2)	TRIM	trimestre di rilevazione
5-8	(4)	ANNRIL	anno di rilevazione = 1998
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	area socio-demografica del Comune (dominio)
25-31	(7)	-	blank
INFORMAZIONI SUI VIAGGI			
32-34	(3)	IDVIA	numero progressivo del viaggio all'interno della famiglia
35-38	(4)	TIPOVGG	tipologia del viaggio
39-43	(5)	-	blank
44-45	(2)	INSIEME	numero progressivo che individua i viaggi fatti insieme dai familiari
46-47	(2)	NPART	numero di componenti della famiglia partecipanti allo stesso viaggio
48-69	(22)	REL1-REL11	undici campi da 2 bytes. Relazione di parentela dei familiari partecipanti al viaggio, rispetto all'intervistato
70-74	(5)	-	blank
75-76	(2)	GGINIZ	giorno di inizio del viaggio
77-78	(2)	MMINIZ	mese di inizio del viaggio
79-82	(4)	AAINIZ	anno di inizio del viaggio
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_RE	regione/Stato estero di destinazione principale

MODALITA' DELLE VARIABILI (in ordine alfabetico)

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

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

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

DEST_RE regione/Stato 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=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

DOM (tipologia socio-demografica del Comune)

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

ISTR titolo di studio

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

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

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

NOVAC motivo per cui l'individuo non ha effettuato vacanze

01=ragioni economiche
02=motivi di lavoro e di studio
03=manca di abitudine
04=residente già in località di villeggiatura
05=motivi di famiglia
06=motivi di salute

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

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

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

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

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

RIP ripartizione geografica di residenza

1=Italia nord-occidentale
2=Italia nord-orientale
3=Italia centrale
4=Italia meridionale
5=Italia insulare

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

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

TIPOVGG tipologia del viaggio

L198=viaggio di lavoro nel primo trimestre
L298=viaggio di lavoro nel secondo trimestre
L398=viaggio di lavoro nel terzo trimestre
L498=viaggio di lavoro nel quarto trimestre
V198=viaggio di vacanza nel primo trimestre
V298=viaggio di vacanza nel secondo trimestre
V398=viaggio di vacanza nel terzo trimestre
V498=viaggio di vacanza nel quarto trimestre

TR tipo di record

IA=record individuale tipo A (con REG=regione di residenza)
IB=record individuale tipo B (con DOM=area socio-demografica del Comune di residenza)
VA=record viaggi tipo A (con REG=regione di residenza)

VB=record viaggi tipo B (con DOM=area 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)

APPENDICE A

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 e dagli individui che le compongono, al netto dei membri permanenti delle convivenze. Per famiglia si intende la *famiglia di fatto*, ovvero 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.

2. Caratteristiche della lista di selezione

La *base di campionamento* adottata, ovvero 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, l'ampiezza del comune di appartenenza, in termini demografici e 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) *uplicazione 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 o ad attività professionali oppure 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 le unità di U che appartengono ad L sono differenti dalle unità di U non presenti in L relativamente alle variabili d'interesse dell'indagine. E' possibile in parte attenuare tale effetto distorsivo ricorrendo a opportune procedure di stima, note in letteratura come stimatori di *ponderazione vincolata*, che tengono conto di totali noti (sulla base di fonti esterne all'indagine) sulla popolazione relativi a variabili ausiliarie correlate a quelle d'interesse.

Un ulteriore effetto dovuto alla *sovracopertura* è il fatto che la numerosità effettiva 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.

3. Descrizione del disegno di campionamento

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, che è stata definita dall'ISTAT sulla base delle caratteristiche socio-demografiche dei comuni italiani. 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 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 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 definita la numerosità n complessiva del campione.

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*. Infine, le numerosità campionarie regionali sono state ripartite tra le diverse tipologie di comune in modo proporzionale alla popolazione residente.

Nel prospetto A.1 sono riportate le numerosità campionarie per regione.

Prospetto A.1 - Distribuzione regionale del campione

REGIONI	INDIRIZZI CAMPIONE
Piemonte	265
Valle d'Aosta	92
Lombardia	330
Trentino-Alto Adige	206
Veneto	201
Friuli-Venezia Giulia	139
Liguria	152
Emilia-Romagna	197
Toscana	213
Umbria	103
Marche	141
Lazio	219
Abruzzo	144
Molise	96
Campania	196
Puglia	187
Basilicata	99
Calabria	158
Sicilia	213
Sardegna	150
ITALIA	3500

3.1. 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-economiche⁵.

⁵ Il tasso di sostituzione, calcolato rapportando il numero di interviste ottenute utilizzando gli indirizzi dell'elenco sostitutivo al totale delle interviste realizzate, è risultato pari al 37,0%, al 38,1%, al 36,2% e al 36,3%, rispettivamente per ciascuno dei quattro trimestri di indagine del 1998.

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, 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 ad ogni unità campionaria un peso che indica il numero di unità della popolazione rappresentate dall'unità medesima. Se, ad esempio, ad 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,...,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 ad un generico dominio d (ad esempio una regione geografica) e relativamente ad un generico trimestre t, il totale della variabile y oggetto di indagine, 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 ad 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 A₁, A₂, B₁, B₂, B₃, B₄ definite nel primo paragrafo; la distribuzione della popolazione per classi di età; la distribuzione delle famiglie per ampiezza.

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, ed 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 numero totale dei viaggi ed al numero totale delle notti nell'anno. Le medie annue di dati trimestrali - ottenute come media aritmetica dei totali (1) in parola - si riferiscono, invece, al numero medio annuo di *turisti*. Con riferimento al generico dominio d , le stime dei suddetti parametri, ${}_dY$ e ${}_d\bar{Y}$, si ottengono rispettivamente mediante le seguenti espressioni:

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

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

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 ed 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 calcolata come somma della stima della varianza dei singoli strati appartenenti al dominio d ; in simboli:

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

dove

$$\hat{Y}_{thj} = Y_{thj} W_{thj} = \sum_{p=1}^{P_{hj}} W_{hjp} Y_{hjp} \quad \text{e} \quad \hat{\bar{Y}}_{th} = \frac{1}{m_h} \sum_{j=1}^{m_h} \hat{Y}_{thj}.$$

Passiamo, adesso, 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:

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

Un'espressione approssimata delle (8) e (9) si può ottenere utilizzando ipotesi semplificatrici di costanza delle varianze $\hat{Var}_{(d)\hat{Y}_t}$ e dei totali ${}_d\hat{Y}_t$ tra i quattro trimestri dell'anno, che in formule sono espresse come

$${}_d\hat{Y}_t = {}_d\hat{Y}_{\cdot} \quad , \quad \hat{Var}_{(d)\hat{Y}_t} = \hat{Var}_{(d)\hat{Y}_{\cdot}} \quad \text{per } t = 1, 2, 3, 4 \quad .$$

Su tale base le (8) e le (9) diventano:

$$\hat{\sigma}_{(d)\hat{Y}} = 2\sqrt{\hat{Var}_{(d)\hat{Y}_{\cdot}}} \quad \hat{\varepsilon}_{(d)\hat{Y}} = \frac{1}{2} \frac{\sqrt{\hat{Var}_{(d)\hat{Y}_{\cdot}}}}{{}_d\hat{Y}_{\cdot}} = \frac{1}{2} \hat{\varepsilon}_{(d)\hat{Y}_{\cdot}} \quad (10)$$

$$\hat{\sigma}_{(d)\hat{Y}} = \frac{1}{2} \sqrt{\hat{Var}_{(d)\hat{Y}_{\cdot}}} \quad \hat{\varepsilon}_{(d)\hat{Y}} = \frac{1}{2} \frac{\sqrt{\hat{Var}_{(d)\hat{Y}_{\cdot}}}}{{}_d\hat{Y}_{\cdot}} = \frac{1}{2} \hat{\varepsilon}_{(d)\hat{Y}_{\cdot}} \quad (11)$$

Gli errori campionari espressi nelle espressioni (5), (6), (8) e (9) consentono di valutare il grado di precisione delle stime; inoltre, l'errore assoluto permette di costruire l'intervallo di confidenza, che, con una certa probabilità, contiene il parametro d'interesse. Con riferimento ad una generica stima \hat{Y} tale intervallo assume la seguente forma:

$$\Pr\{\hat{Y} - k\hat{\sigma}(\hat{Y}) \leq Y \leq \hat{Y} + k\hat{\sigma}(\hat{Y})\} = P \quad (12)$$

Nella (12) il valore di k dipende dal valore fissato per la probabilità P ; ad esempio, per $P=0,95$ si ha $k=2$.

6. Presentazione sintetica degli errori campionari

Ad ogni stima ${}_d\hat{Y}_t$ corrisponde un errore campionario relativo $\hat{\varepsilon}_{(d)\hat{Y}_t}$; ciò significa che per consentire un uso corretto delle stime sarebbe necessario pubblicare per ogni stima il corrispondente errore di campionamento relativo. Questo tuttavia non è possibile 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.

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

Per tali motivi si ricorre, in genere, ad una presentazione sintetica degli errori relativi basata sul *metodo dei modelli regressivi* 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 di questi modelli è diverso a seconda che si tratti di variabili qualitative o quantitative. Infatti, per quanto riguarda le stime di frequenze assolute (o relative) riferite alle modalità di variabili

qualitative, è possibile utilizzare modelli che hanno un fondamento teorico e secondo cui gli errori relativi delle stime di frequenze assolute sono funzione decrescente dei valori delle stime stesse. Per quanto riguarda, invece, le stime di totali di variabili quantitative, si tratta di un problema di notevole complessità, per il fatto che non è stata elaborata un'adeguata base teorica per l'interpolazione degli errori campionari delle stime in questione. L'approccio qui seguito per risolvere il problema in oggetto è di tipo empirico ed è fondato sull'evidenza empirica che l'errore assoluto di un totale è una funzione crescente del totale stesso.

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:

$$\log \hat{\varepsilon}^2({}_d\hat{Y}_t) = a + b \log({}_d\hat{Y}_t) \quad (13)$$

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 ai quattro trimestri sono sostanzialmente uguali tra loro e pertanto, per semplicità, vengono presentati solamente i risultati relativi al primo trimestre.

Il prospetto A.2 riporta 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. Le informazioni contenute nel prospetto A.2 permettono di calcolare l'errore relativo di una qualsiasi stima di frequenza assoluta (e relativa).

Allo scopo di facilitare il calcolo degli errori campionari, nel prospetto A.3 sono riportati i valori interpolati degli errori campionari relativi percentuali di alcuni valori tipici assunti dalle stime di frequenze assolute per area territoriale.

Il prospetto A.3 è costituito da due colonne, la prima riporta K valori crescenti delle stime ${}_d\hat{Y}_t^k$ ($k=1, \dots, K$) e la seconda riporta i rispettivi errori relativi interpolati $\hat{\varepsilon}({}_d\hat{Y}_t^k)$. Le informazioni contenute in tale prospetto permettono, quindi, 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'espressione (13).

Il primo metodo consiste nel cercare nella prima colonna del prospetto il livello di stima che più si avvicina alla stima di interesse ${}_d\hat{Y}_t$; l'errore relativo corrispondente $\hat{\varepsilon}({}_d\hat{Y}_t)$ si trova sulla stessa riga della seconda colonna.

Nel secondo metodo, l'errore campionario della stima ${}_d\hat{Y}_t$, si ricava mediante la seguente espressione:

$$\hat{\varepsilon}({}_d\hat{Y}_t) = \hat{\varepsilon}({}_d\hat{Y}_t^{k-1}) + \frac{\hat{\varepsilon}({}_d\hat{Y}_t^k) - \hat{\varepsilon}({}_d\hat{Y}_t^{k-1})}{{}_d\hat{Y}_t^k - {}_d\hat{Y}_t^{k-1}} ({}_d\hat{Y}_t - {}_d\hat{Y}_t^{k-1}) \quad (14)$$

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 ${}_d\hat{Y}_t$ e $\hat{\varepsilon}({}_d\hat{Y}_t^{k-1})$ e $\hat{\varepsilon}({}_d\hat{Y}_t^k)$ i corrispondenti errori relativi letti sul prospetto.

A partire dall'errore relativo $\hat{\varepsilon}({}_d\hat{Y}_t)$ riferito ad una stima trimestrale, è possibile, utilizzando la (11), ottenere l'errore relativo di una stima annua del tipo (4) come

$$\hat{\varepsilon}({}_d\hat{Y}) = \frac{1}{2} \hat{\varepsilon}({}_d\hat{Y}_t).$$

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:

$$\hat{\sigma}(\hat{Y}_t) = a + b \hat{Y}_t + c \hat{Y}_t^2 \quad (15)$$

dove i parametri a , b e c vengono stimati utilizzando il metodo dei minimi quadrati, adattando il modello (15) ad 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 (15) 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 1998 - 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 complessivo di notti fuori casa per viaggi;
- numero di viaggi di lavoro;
- numero di viaggi di vacanza;
- numero complessivo di viaggi.

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 e al terzo trimestre.

I prospetti A.4 e A.6 riportano i valori dei coefficienti a , b , c e dell'indice di determinazione R^2 delle funzioni utilizzate 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 e al terzo trimestre. Il prospetto A.4 fa riferimento alle stime del numero di notti, mentre il prospetto A.6 alle stime del numero di viaggi. Come viene illustrato nel paragrafo seguente, sulla base dei dati riportati in tali prospetti, è possibile, quindi, calcolare facilmente l'errore di campionamento assoluto e relativo di una stima a livello trimestrale.

La determinazione numerica dei valori dei coefficienti del modello (15) consente di pervenire ad una presentazione semplificata degli errori di campionamento basata sul fatto che gli errori relativi di un totale diminuiscono al crescere del valore del totale stesso. Utilizzando i valori stimati dei parametri del modello (15) 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:

$$\hat{Y}_t = \frac{-[b - \varepsilon(\hat{Y}_t)] - \sqrt{[b - \varepsilon(\hat{Y}_t)]^2 - 4ac}}{2c} \quad (16)$$

Nei prospetti A.5 e A.7 sono riportati i valori delle stime \hat{Y}_t^* ottenuti in corrispondenza di valori tipici prefissati, indicati come $\varepsilon^* = \varepsilon(\hat{Y}_t^*)$, dell'errore relativo percentuale ottenuti in base alla (16); tutte le stime con valori superiori a \hat{Y}_t^* presentano valori dell'errore relativo inferiori a ε^* , mentre le stime che assumono valori inferiori a \hat{Y}_t^* presentano valori dell'errore relativo superiori a ε^* . Il prospetto A.5 fa riferimento alle stime del numero di notti e il prospetto A.7 alle stime del numero di

viaggi. In tali prospetti, articolati, allo stesso modo dei prospetti A.4 e A.6, per gruppo di stime, trimestre e area territoriale, i valori prefissati di ε^* sono: 5, 10, 15, 20, 25, 30 e 35%. Per meglio chiarire l'utilizzo dei suddetti prospetti, viene riportato un esempio nel paragrafo successivo.

Per quanto riguarda le stime a livello annuo, gli errori relativi si possono calcolare in base alla (8) utilizzando il modello (15), in cui i coefficienti a , b e c relativi al primo trimestre gennaio-marzo vengono utilizzati per calcolare le varianze per i trimestri gennaio-marzo, aprile-giugno, ottobre-dicembre, mentre i coefficienti relativi al trimestre luglio-settembre vengono utilizzati per calcolare la varianza del terzo trimestre stesso.

Prospetto A.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

AREE TERRITORIALI	a	b	R^2
Nord	11,9585	-1,26313	93,63
Centro	11,1221	-1,21954	93,11
Sud	11,3604	-1,24196	93,12
ITALIA	12,2355	-1,27245	93,83

Prospetto A.3 - Valori interpolati degli errori relativi percentuali delle stime di frequenze assolute riferite agli individui per aree territoriali

STIME	AREE TERRITORIALI			
	Nord	Centro	Sud	ITALIA
20000	75,9	62,0	62,5	83,3
30000	58,8	48,4	48,6	64,3
40000	49,0	40,6	40,7	53,6
50000	42,6	35,5	35,4	46,5
60000	37,9	31,7	31,6	41,4
70000	34,4	28,9	28,7	37,5
80000	31,6	26,6	26,4	34,5
90000	29,4	24,8	24,6	32,0
100000	27,5	23,2	23,0	29,9
200000	17,7	15,2	15,0	19,2
300000	13,7	11,9	11,6	14,9
400000	11,4	10,0	9,7	12,4
500000	9,9	8,7	8,5	10,7
750000	7,7	6,8	6,6	8,3
1000000	6,4	5,7	5,5	6,9
2000000	4,1	3,7	3,6	4,4
3000000	3,2	2,9	2,8	3,4
4000000	2,7	2,5	2,3	2,9
5000000	2,3	2,1	2,0	2,5
7500000	1,8	1,7	1,6	1,9
15000000	1,2	1,0	1,2
20000000	1,0	0,9	1,0
25000000	0,8	0,9

Prospetto A.4 - Valori dei coefficienti a, b, c e di R² delle funzioni utilizzate per l'interpolazione degli errori campionari delle stime di totali per gruppi di stime, trimestre e aree territoriali – stime del numero di notti

NOTTI – lavoro				
AREE TERRITORIALI	a	b	c	R²
		gennaio-marzo		
Nord	115623,79	0,32386	-0,000000024118	84,4
Centro	59841,94	0,26698	-0,000000033243	84,2
Sud	80875,89	0,32259	-0,000000023121	92,6
ITALIA	146040,84	0,18793	-0,000000005282	88,8
		luglio-settembre		
Nord	93492,68	0,26535	-0,000000016586	91,8
Centro	81098,01	0,39119	-0,000000035842	78,0
Sud	38401,59	0,48070	-0,000000037753	90,7
ITALIA	129812,25	0,23513	-0,000000006585	90,1
NOTTI – vacanza				
AREE TERRITORIALI	a	b	c	R²
		gennaio-marzo		
Nord	147975,33	0,13764	-0,000000002230	85,2
Centro	139254,90	0,15983	-0,000000004723	84,6
Sud	101878,14	0,18188	-0,000000005702	96,1
ITALIA	228253,94	0,09093	-0,000000000701	90,4
		luglio-settembre		
Nord	446557,14	0,09549	-0,000000002749	95,5
Centro	287697,98	0,14653	-0,000000010607	96,1
Sud	241665,68	0,16430	-0,000000011055	94,0
ITALIA	578432,62	0,07266	-0,000000001152	95,8
NOTTI – totale				
AREE TERRITORIALI	a	b	c	R²
		gennaio-marzo		
Nord	227320,96	0,12876	-0,000000001773	86,9
Centro	130693,88	0,15528	-0,000000004148	88,7
Sud	150424,57	0,16303	-0,000000003728	94,9
ITALIA	332141,90	0,08488	-0,000000000546	91,7
		luglio-settembre		
Nord	483470,14	0,09156	-0,000000000253	96,1
Centro	326147,77	0,13922	-0,000000000943	96,3
Sud	344669,97	0,15431	-0,000000000967	94,8
ITALIA	662950,63	0,06903	-0,000000000104	96,2

Prospetto A.5 - Valori dei totali corrispondenti ad alcuni valori tipici degli errori relativi percentuali per gruppi di stime, trimestre e aree territoriali – stime del numero di notti

NOTTI – lavoro							
AREE TERRITORIALI	5%	10%	15%	20%	25%	30%	35%
gennaio-marzo							
Nord	11762576	9772437	7821649	5942350	4203061	2739371	1713688
Centro	6792121	5358926	3972128	2685242	1621175	934016	584169
Sud	12079296	9977753	7907026	5895437	4011540	2421543	1369211
ITALIA	27132252	18168870	9957633	4238353	2009284	1231628	876083
luglio-settembre							
Nord	13404365	10505797	7687873	5055141	2881618	1549272	933660
Centro	9751315	8393829	7050191	5729180	4447931	3242116	2184830
Sud	11496840	10183849	8874191	7569549	6272927	4990210	3734360
ITALIA	28798423	21440331	14306856	7847048	3452201	1705757	1065053
NOTTI - vacanza							
AREE TERRITORIALI	5%	10%	15%	20%	25%	30%	35%
gennaio-marzo							
Nord	40930816	20172851	5833604	2199897	1284249	900273	691789
Centro	24461196	14677540	6569677	2644465	1436310	962273	719412
Sud	23877023	15511718	7863248	2926806	1344301	829298	594017
ITALIA	63497823	12698824	3701486	2065306	1425964	1087790	878960
luglio-settembre							
Nord	174772902	32927587	7879126	4225892	2875440	2177175	1751263
Centro	93894646	49362047	14914503	4903563	2705462	1850942	1403686
Sud	105464887	61706361	22605620	5746697	2724173	1755768	1291448
ITALIA	219569404	19547045	7397574	4523912	3254834	2541078	2083841
NOTTI – totale							
AREE TERRITORIALI	5%	10%	15%	20%	25%	30%	35%
gennaio-marzo							
Nord	47149160	22041656	6820256	2971239	1826204	1309741	1019163
Centro	26565656	15375391	6285420	2391830	1305188	880841	661857
Sud	31600328	19030067	8336464	3099938	1617483	1067234	792032
ITALIA	71594459	14250918	4871420	2837117	1993538	1535144	1247707
luglio-settembre							
Nord	175379733	30120563	7996447	4413031	3036733	2312983	1867316
Centro	98171375	48708830	13741855	4981212	2873826	2004966	1536775
Sud	110763459	61328069	20123056	5925432	3087629	2062985	1541109
ITALIA	213803051	20061128	8103658	5041761	3655674	2866606	2354459

Prospetto A.6 - 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 aree territoriali – stime del numero di viaggi

VIAGGI – lavoro				
AREE TERRITORIALI	a	b	c	R^2
		gennaio-marzo		
Nord	12854,74	0,14881	-0,000000004992	98,7
Centro	12330,21	0,25367	-0,000000018865	94,9
Sud	10596,44	0,14060	-0,000000005729	98,7
ITALIA	18099,66	0,10371	-0,000000001540	97,9
		luglio-settembre		
Nord	11761,64	0,17817	-0,000000004980	98,9
Centro	9136,50	0,20183	-0,000000013096	98,8
Sud	8745,07	0,15912	-0,000000006550	99,0
ITALIA	16676,17	0,10885	-0,000000001443	99,0
VIAGGI – vacanza				
AREE TERRITORIALI	a	b	c	R^2
		gennaio-marzo		
Nord	25134,27	0,10284	-0,000000004260	93,0
Centro	19132,92	0,16399	-0,000000015168	84,2
Sud	19658,20	0,17240	-0,000000008466	90,8
ITALIA	40981,02	0,07751	-0,000000001403	89,7
		luglio-settembre		
Nord	49022,43	0,07511	-0,000000003379	84,5
Centro	19148,75	0,14154	-0,000000016396	95,2
Sud	25492,81	0,10118	-0,000000005553	94,6
ITALIA	49004,03	0,06720	-0,000000001571	94,9
VIAGGI – totale				
AREE TERRITORIALI	a	b	c	R^2
		gennaio-marzo		
Nord	22408,10	0,10610	-0,000000004719	93,4
Centro	19152,31	0,16927	-0,000000017833	82,4
Sud	16787,19	0,16117	-0,000000008378	90,0
ITALIA	35045,87	0,07818	-0,000000001528	89,8
		luglio-settembre		
Nord	42135,64	0,06759	-0,000000002684	81,0
Centro	18889,90	0,13567	-0,000000016812	87,8
Sud	20960,38	0,10432	-0,000000006050	94,7
ITALIA	37135,98	0,07357	-0,000000001900	92,8

Prospetto A.7 - Valori dei totali corrispondenti ad alcuni valori tipici degli errori relativi percentuali per gruppi di stime, trimestre e aree territoriali – stime del numero di viaggi

VIAGGI – lavoro							
AREE TERRITORIALI	5%	10%	15%	20%	25%	30%	35%
	gennaio-marzo						
Nord	19922921	10034272	1489932	245253	126249	84786	63793
Centro	10856388	8225236	5611830	3058641	911557	242244	124943
Sud	15930384	7338784	767898	175423	96373	66319	50534
ITALIA	35210417	4838270	886048	187409	123564	92142	73455
	luglio-settembre						
Nord	25828389	15845834	6047184	485100	161925	96163	68314
Centro	11653483	7864368	4126754	908044	180786	91940	61330
Sud	16739302	9171527	2045181	207052	95568	61896	45743
ITALIA	41064516	7644758	399652	182426	118003	87184	69124
VIAGGI - vacanza							
AREE TERRITORIALI	5%	10%	15%	20%	25%	30%	35%
	gennaio-marzo						
Nord	12862460	2785103	509508	255820	169959	127132	101515
Centro	7679421	4499116	1675284	447116	214348	138533	102011
Sud	14616692	8815263	3340904	601335	246688	152518	110110
ITALIA	20998982	1651949	559279	333294	237128	183979	150278
	luglio-settembre						
Nord	9036651	1615332	636323	388442	278803	217275	177945
Centro	5784953	2931886	853068	301977	172075	119368	91204
Sud	9690389	2251501	494379	254337	170218	127765	102222
ITALIA	13294705	1400131	585335	367409	267460	210200	173115
VIAGGI - totale							
AREE TERRITORIALI	5%	10%	15%	20%	25%	30%	35%
	gennaio-marzo						
Nord	12274955	2919255	485136	235843	154933	115242	91711
Centro	6845061	4143564	1709006	486113	225960	143686	104886
Sud	13418601	7566094	2231272	398126	185728	120049	88553
ITALIA	19611892	1457396	483005	286655	203600	157821	128837
	luglio-settembre						
Nord	8418460	1183990	503051	316194	230214	180921	148989
Centro	5307466	2560515	716282	274018	161392	113630	87534
Sud	9349086	2252280	433916	216114	143030	106763	85137
ITALIA	13819579	1286153	480151	292442	210011	163781	134218

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.4 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.6 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* e riferita al Nord per il terzo trimestre, in simboli ${}_N\hat{Y}_3$, si possono leggere i valori dei coefficienti a , b e c dal prospetto A.4 e si può utilizzare la seguente espressione:

$$\hat{\sigma}({}_N\hat{Y}_3) = 93492,68 + 0,26535 \cdot {}_N\hat{Y}_3 - 0,000000016586 ({}_N\hat{Y}_3)^2.$$

Esempio 2

La valutazione degli errori di campionamento delle stime mediante un metodo approssimato si può ottenere utilizzando i dati dei prospetti A.5 e A.7. Il prospetto A.5 fa riferimento alle stime del numero di notti; il prospetto A.7 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.5, si può osservare che il valore riportato all'incrocio tra la riga riferita al totale delle notti del terzo trimestre per l'Italia e la colonna relativa all'errore del 10% indica che tutte le stime superiori a 20.061.128 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.5, si consideri la stima del *numero di notti per lavoro*, riferita al totale Italia per il trimestre luglio-settembre, e pari a 18.930.000 (prospetto 1.1).

Considerando nel prospetto A.5 il gruppo dei valori 'notti-lavoro' ed esaminando l'ultima riga corrispondente al totale Italia per il trimestre luglio-settembre, si osserva che il valore 18.930.000 risulta essere compreso tra i valori 21.440.331 e 14.306.856.

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 8.291.000.

Nella prima colonna del prospetto A.3 il livello di stima che più si avvicina a 8.291.000 è 7.500.000. Dall'ultima colonna del prospetto A.3, con riferimento all'Italia, si ricava il corrispondente errore relativo percentuale della stima considerata, che è pari a 1,9 %.

L'errore assoluto sarà:

$$\sigma(8.291.000)=0,019 \times 8.291.000 = 157.529$$

L'intervallo di confidenza avrà come estremi:

$$8.291.000 - (2 \times 157.529) = 7.975.942$$

$$8.291.000 + (2 \times 157.529) = 8.606.058$$

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 percentuali 1,9 e 1,2.

L'errore corrispondente a 8.291.000 è pari a :

$$\sigma(8.291.000) = 1,9 - ((1,9 - 1,2) / (15.000.000 - 7.500.000)) \times (8.291.000 - 7.500.000) = 1,83 \%$$

Il corrispondente errore assoluto è $0,0183 \times 8.291.000 = 151.725$ e l'intervallo di confidenza avrà come estremi:

$$8.291.000 - (2 \times 151.725) = 7.987.550$$

$$8.291.000 + (2 \times 151.725) = 8.594.450$$

Esempio 3

Il calcolo dell'errore può essere effettuato direttamente tramite la funzione interpolante (13):

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

Per la stima di 8.291.000 si ha:

$$\log \varepsilon^2 = 12,23355 - 1,27245 \times \log (8.291.000) = -8,03745$$

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

Infatti da $\log (\varepsilon^2) = -8,03745$ si ha :

$$\varepsilon^2 = e^{-8,03745}$$

e dunque

$$\varepsilon = 0.017976$$

L'errore relativo percentuale sarà dunque pari a 1,8 %.

APPENDICE B

IL QUESTIONARIO C.A.T.I. ANNO 1998

Il questionario è strutturato a blocchi 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 possono quindi essere individuate 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 intestataria 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 per famiglia, 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 un'intervista diretta ad 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', in cui vengono registrate le informazioni strutturali di ciascun componente della famiglia;
- 'Scheda riepilogativa viaggi', in cui viene riportato il numero dei viaggi conclusi da ciascun componente nel trimestre di riferimento;
- 'Scheda viaggio' in cui vengono registrate le informazioni per ciascun viaggio effettuato dai componenti della famiglia.

Si fa presente che ci possono essere differenze tra le informazioni rilevate secondo quanto dettagliato nel questionario e quelle fornite nei files standard e specificate nei tracciati record.