

I viaggi in Italia e all'estero nel 2007

Nota Metodologica

Cenni sull'indagine Viaggi e vacanze

L'indagine Viaggi e vacanze sulla domanda turistica è condotta dall'Istat dal 1997 con cadenza trimestrale e fa parte del Sistema delle Indagini Multiscopo sulle famiglie.

L'indagine è di tipo campionario (cfr. Strategia di campionamento e valutazione degli errori campionari). La popolazione di interesse dell'indagine – ossia l'insieme delle unità statistiche oggetto di investigazione – è costituita dalle famiglie residenti in Italia e dagli individui che le compongono (cittadini italiani e stranieri residenti in famiglia, di qualunque età). Le informazioni sono raccolte tramite intervista telefonica assistita da computer (CATI).

L'indagine si presenta come uno strumento idoneo ad analizzare le diverse tipologie di domanda turistica, i comportamenti turistici e le modalità di effettuazione dei viaggi (con pernottamento) attraverso la raccolta di informazioni sulla destinazione, la forma di organizzazione scelta, il tipo di alloggio utilizzato, il mezzo di trasporto, la durata del soggiorno, il periodo dell'anno in cui si è viaggiato, ecc. Inoltre, è la fonte privilegiata per conoscere la domanda turistica 'sommersa', vale a dire quella che si realizza negli alloggi privati a pagamento, nelle seconde case di proprietà oppure tramite l'ospitalità di parenti o amici.

Gli spostamenti turistici sono classificati, secondo gli standard internazionali, distinguendo i viaggi per motivi di lavoro da quelli per motivi di vacanza e le vacanze 'brevi' da quelle 'lunghe'. Tra le vacanze rientrano i viaggi per svago, piacere, relax, per visitare parenti o amici, per trattamenti di salute o per motivi religiosi.

Oltre ai viaggi, costituiscono oggetto di analisi anche gli individui che hanno viaggiato e le notti trascorse in viaggio.

Avvertenze

Per una corretta interpretazione dei dati riportati nelle tavole si suggerisce la lettura della sezione "Definizioni e classificazioni" con particolare riferimento alla definizione di viaggio adottata e che riguarda, più in particolare, gli spostamenti turistici con pernottamento effettuati fuori dall' "ambiente abituale", vale a dire fuori dal luogo in cui si vive, nonché dai luoghi frequentati settimanalmente con soste di almeno un pernottamento. Inoltre, poiché la stagionalità del turismo e le caratteristiche che contraddistinguono le diverse tipologie di viaggio fanno sì che alcuni aspetti del viaggiare siano, in termini quantitativi, particolarmente rari in particolari periodi dell'anno, è opportuno porre particolare cura ed attenzione nell'analisi e valutazione delle stime. A tal fine è bene sempre far riferimento agli errori campionari (sezione "Strategia di campionamento e valutazione degli errori campionari").

Si rammenta che l'indagine rileva i flussi turistici che si concludono in ciascun trimestre di rilevazione anche se iniziati precedentemente. Per quanto concerne i prospetti e le tavole si precisa, pertanto, che i dati trimestrali fanno riferimento ai viaggi conclusi nel trimestre, indipendentemente dalla data di inizio.

I dati sui viaggi, nonché sulle notti, si riferiscono a stime trimestrali e a stime annuali; queste ultime sono ottenute come somma dei valori trimestrali.

I dati sulle persone che viaggiano, diversamente dai viaggi e dalle notti, si riferiscono soltanto a stime trimestrali o, in alcuni casi, a valori medi trimestrali ottenuti sulla base delle stime dei quattro trimestri.

Si avverte che i dati trimestrali sul numero di persone che viaggiano non sono cumulabili, in quanto una persona può viaggiare in trimestri diversi: l'eventuale somma dei dati trimestrali comporterebbe pertanto una sovrastima del numero di persone che hanno effettuato viaggi nell'anno. Per analogo motivo, si segnala che anche i totali sulle persone che hanno viaggiato per tipo di viaggio o per destinazione non corrispondono alla somma dei parziali, per la presenza di persone che risultano viaggiare contemporaneamente per diverse tipologie di viaggio o per diverse destinazioni. Ad esempio, un individuo che nel corso del trimestre luglio-settembre ha effettuato sia vacanze brevi che vacanze lunghe viene conteggiato tanto tra i turisti per vacanza breve quanto tra i turisti per vacanza lunga, ma una sola volta come turista per vacanza. Analogamente, un individuo che ha realizzato

sia vacanze brevi in Italia che vacanze brevi all'estero risulta presente sia tra i turisti per vacanza breve in Italia sia tra quelli per vacanza breve all'estero, pur essendo conteggiato una sola volta nel totale dei turisti per vacanza breve.

Si fa notare che i prospetti riguardano sia i confronti temporali che i dati relativi all'anno di riferimento dell'indagine, mentre le tavole di approfondimento soltanto i dati relativi all'anno di riferimento.

Si fa presente che i dati assoluti sono espressi in migliaia; inoltre, i totali possono non quadrare per effetto di arrotondamento. Le percentuali e i quozienti, essendo calcolati sui dati assoluti prima dell'arrotondamento, possono non coincidere con i risultati che si ottengono elaborando i dati espressi in migliaia.

Si segnala che il segno (-) indica che il fenomeno esiste e viene rilevato, ma i casi non si sono verificati nel campione; (..) indica che la numerosità non raggiunge la metà della cifra dell'ordine minimo considerato oppure che l'esiguità del fenomeno rende i valori calcolati non significativi.

Si ricorda che i risultati riguardano la popolazione residente al netto dei membri permanenti delle convivenze (ospedali, carceri, comunità religiose, eccetera), poiché trattasi di un'indagine campionaria effettuata presso le famiglie, le quali rappresentano le unità di rilevazione.

Per confronti con i dati delle precedenti indagini "Viaggi e vacanze", si deve tener conto di alcune modifiche introdotte nel corso dei primi anni di indagine al fine di migliorare la qualità delle stime. Ci si riferisce, in particolare, alle modifiche apportate alla scheda riepilogativa dei viaggi nel questionario del 1998 che hanno contribuito in maniera sostanziale a migliorare le stime sui flussi turistici, soprattutto per quanto riguarda le vacanze brevi ed i viaggi di lavoro.¹

Per i confronti storici con i dati anteriori al 1997, ricavati dalle precedenti indagini Istat sulle vacanze dei residenti in Italia e dalle indagini "Aspetti della vita quotidiana", si deve tener conto, oltre che degli usuali accorgimenti nel paragonare fonti diverse, di problematiche che emergono nel confrontare dati ricavati da indagini che hanno periodi di riferimento diversi, dal momento che in passato il periodo di riferimento era l'anno solare. Si può ipotizzare che più ampio è il periodo a cui ci si riferisce e maggiore è la difficoltà per l'intervistato a ricordare gli eventi accaduti in quel lasso di tempo. Si osservano, infatti, differenze tra i dati dell'indagine trimestrale e quelli relativi alle precedenti indagini sulle vacanze, basate su periodi di riferimento annuali. Analizzando le distribuzioni mensili dei viaggi risulta che, nel caso delle stime ottenute con periodi di riferimento annuali, vi è una maggiore concentrazione delle vacanze nei mesi ad esse abitualmente dedicati (luglio e agosto) rispetto alle informazioni rilevate trimestralmente. Quando il periodo di riferimento dell'indagine è annuale si può verificare, infatti, che l'intervistato ricordi principalmente i viaggi effettuati nel periodo dell'anno in cui è solito trascorrere le proprie vacanze, dimenticando più facilmente quelli effettuati negli altri periodi ed in particolare quelli più remoti rispetto al momento della rilevazione.

Principali definizioni e classificazioni

L'indagine "Viaggi e vacanze" utilizza le definizioni standard indicate nella metodologia adottata per l'attuazione della Direttiva 95/57/CE sulle Statistiche del Turismo.²

I flussi turistici vengono individuati in quegli spostamenti con pernottamento effettuati fuori dall' "ambiente abituale" ovvero fuori dal luogo in cui si vive, nonché dai luoghi frequentati settimanalmente con soste di almeno un pernottamento.

In tal senso, per una corretta interpretazione dei risultati, assume notevole importanza il significato delle definizioni adottate. In particolare:

- **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 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;

¹ Per approfondimenti si suggerisce la lettura dell'appendice B del volume: Istat. *I viaggi in Italia e all'estero nel 1998* Roma: Istat, 1999. (Informazioni n. 25); oppure del volume: Istat. *Metodologia e organizzazione dell'indagine multiscope sulla domanda turistica 'Viaggi e vacanze'*. Roma: Istat, 2003. (Metodi e Norme n. 17).

² Eurostat. Community methodology on Tourism statistics. Bruxelles: Eurostat, 1998.

- **turista**
persona che ha effettuato uno o più viaggi nel trimestre di riferimento dell'indagine.

I viaggi sono, inoltre, distinti secondo le seguenti tipologie:

- **viaggio per motivi di lavoro o professionali**
viaggio svolto per motivi prevalenti di lavoro temporaneo fuori sede o più in generale per motivi professionali (missioni, partecipazione a congressi, convegni, eccetera): in questi viaggi vengono 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;
nella presentazione dei risultati, il soggiorno di vacanza è stato distinto, in relazione alla durata, in:
 - *vacanza breve* : quando la durata del soggiorno è inferiore a 4 pernottamenti;
 - *vacanza lunga* : quando la durata del soggiorno è di 4 o più notti.

Sono state adottate le seguenti definizioni:

- **destinazione, mezzo di trasporto, tipo di alloggio, motivo della vacanza, tipo della vacanza di piacere/svago, motivo del viaggio di lavoro**
sono informazioni che vengono rilevate sulla base del concetto di "prevalenza": in particolare, la destinazione del viaggio ed il tipo di alloggio vengono associate rispettivamente alla località ed al tipo di alloggio in cui 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;
- **organizzazione**
 - per *prenotazione diretta* si intende la prenotazione dell'alloggio e/o del trasporto direttamente presso la struttura ricettiva o la compagnia di viaggio;
 - per *prenotazione presso agenzia* si intende la prenotazione dell'alloggio e/o del trasporto tramite agenzia di viaggi o tour operator;
 - La prenotazione tramite internet riguarda l'utilizzo di internet per prenotare direttamente o presso agenzia l'alloggio e/o il trasporto;
- **estero**
 - *Europa* include i paesi dell'Unione europea e gli altri paesi europei;
 - *Unione europea* comprende: Austria, Belgio, Danimarca, Finlandia, Francia, Germania, Grecia, Irlanda, Lussemburgo, Olanda, Portogallo, Regno Unito, Spagna, Svezia, Cipro, Estonia, Latvia (Lettonia), Lituania, Malta, Polonia, Repubblica Ceca, Slovenia, Slovacchia, Ungheria, Bulgaria, Romania;
 - *altri paesi europei* comprende: Norvegia, Russia, Svizzera, Liechtenstein, Turchia e altri paesi europei altrove non classificati;
 - *paesi extra-europei* include tutti i paesi non menzionati tra quelli dell'Europa;
- **ripartizione geografica**
le ripartizioni geografiche sono suddivisioni del territorio italiano derivate dalle regioni amministrative e così definite:
 - *Nord*: Piemonte, Valle d'Aosta/Vallée d'Aoste, Lombardia, Liguria, Trentino-Alto Adige, Veneto, Friuli-Venezia Giulia, Emilia-Romagna
 - *Centro*: Toscana, Umbria, Marche, Lazio
 - *Mezzogiorno*: Abruzzo, Molise, Campania, Puglia, Basilicata, Calabria, Sicilia, Sardegna;

- **tipo di comune**

i comuni italiani sono suddivisi nelle seguenti tipologie:

- *comuni centro dell'area metropolitana*: Torino, Milano, Venezia, Genova, Bologna, Firenze, Roma, Napoli, Bari, Palermo, Catania, Cagliari;
- *comuni periferia dell'area metropolitana*: comuni che gravitano intorno al centro dell'area metropolitana, come definiti sulla base dei risultati del Censimento Popolazione;
- *altri comuni* suddivisi in base alla dimensione demografica (oltre 50.000 abitanti, da 10.001 a 50.000 abitanti, da 2.001 a 10.000 abitanti, fino a 2.000 abitanti).

I dati relativi alle caratteristiche delle persone fanno riferimento al momento 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:
 - *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 prima occupazione*: chi non ha mai esercitato un'attività lavorativa ed è alla ricerca attiva di un'occupazione che è in grado di accettare se gli viene offerta;
 - *persona in cerca di nuova occupazione*: chi ha perduto una precedente occupazione alle dipendenze ed è alla ricerca attiva di un'occupazione che è in grado di accettare se gli viene 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 e non sempre chi riceve una pensione la riceve per motivi di lavoro;
 - *in altra condizione*: chi si trova in condizione diversa da quelle sopra elencate (militare di leva, inabile al lavoro, benestante, detenuto eccetera).

Strategia di campionamento e valutazione degli errori campionari

1. Introduzione

La popolazione di interesse dell'indagine – ossia l'insieme delle unità statistiche oggetto di investigazione – è costituita dalle famiglie residenti in Italia e dagli individui che le compongono, al netto dei membri permanenti delle convivenze. Per famiglia si intende la famiglia di fatto, ossia un insieme di persone coabitanti e legate da vincoli di matrimonio, parentela, affinità, adozione, tutela o affettivi.

L'indagine viene svolta mediante intervista telefonica assistita da computer (CATI) e utilizza come lista di selezione l'archivio ufficiale nazionale degli abbonati alle utenze private di telefonia fissa; 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 Mezzogiorno);
- le regioni geografiche (solamente per alcune stime più importanti);
- sei aree basate sulla tipologia socio-demografica dei comuni, così definite:

A, *area metropolitana* suddivisa in :

A₁, comuni centro dell'area metropolitana: Torino, Milano, Venezia, Genova, Bologna, Firenze, Roma, Napoli, Bari, Palermo, Catania e Cagliari;

A₂, comuni che gravitano intorno al centro dell'area metropolitana;

B, area non metropolitana suddivisa in:

B₁, comuni aventi fino a 2.000 abitanti;

B₂, comuni con 2.001-10.000 abitanti;

B₃, comuni con 10.001-50.000 abitanti;

B₄, comuni con oltre 50.000 abitanti.

La suddivisione dei comuni nelle aree A e B è stata effettuata utilizzando i dati del Censimento della Popolazione³.

2. Caratteristiche della lista di selezione

La base di campionamento adottata, ossia la lista di selezione delle unità campionarie, è l'archivio informatizzato ufficiale degli abbonati alle utenze private di telefonia fissa. Tale scelta è motivata dal fatto che le informazioni dell'archivio in oggetto sono contenute in un file che viene costantemente aggiornato sulle variazioni degli intestatari e degli indirizzi telefonici; esso è, inoltre, di agevole utilizzo per la selezione delle unità campionarie in quanto si presta facilmente alla scelta di diversi criteri di ordinamento.

Le informazioni relative a ciascun indirizzo, utilizzabili per la stratificazione delle unità della popolazione di riferimento, sono essenzialmente di tipo territoriale; esse sono: la provincia, il comune, la sezione di censimento, la via, il numero civico e infine l'ampiezza del comune di appartenenza definita sia in termini demografici sia in termini di numero di indirizzi.

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

- 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 misura diversa tutte le situazioni elencate; i fenomeni che appaiono più rilevanti sono, tuttavia, quelli della sovracopertura e della sottocopertura. Rientrano, infatti, nel caso della sovracopertura tutti i numeri telefonici che corrispondono a seconde case oppure ad attività professionali o a pubblici esercizi, ecc.; mentre rientrano nel caso della sottocopertura tutte le famiglie non abbonate al telefono o intestatarie di numeri riservati. Rientrano, poi, nel caso III le famiglie intestatarie di più di un numero riferito all'abitazione principale e nel caso IV i numeri telefonici a cui corrispondono più famiglie.

L'effetto di ciascuna delle situazioni appena elencate è quello di modificare in modo non controllato le probabilità di inclusione delle unità della popolazione rispetto alle probabilità di inclusione teoriche assegnate alle stesse unità dal disegno di campionamento adottato. Ciò può essere causa di distorsioni nelle stime prodotte dall'indagine.

In particolare la sottocopertura determina una distorsione delle stime prodotte tanto maggiore quanto più, relativamente alle variabili d'interesse dell'indagine, le unità di U che appartengono a L sono differenti dalle unità di U non presenti in L. È possibile in parte attenuare tale effetto distorsivo ricorrendo a opportune procedure di stima, note in letteratura come stimatori di ponderazione vincolata⁴, 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,

³ Cfr. il volume Istat *I sistemi locali del lavoro 1991*, Roma: Istat, 1997. (Argomenti n. 10).

⁴ Noti nella letteratura in lingua anglosassone sul tema come Deville J.C. e Särndal C.E. "Calibration Estimators in Survey Sampling", *Journal of the American Statistical Association*, 87 (1992): 1013-1020.

alternativamente, ricorrere a un sovra-dimensionamento del campione o a una lista di unità sostitutive. Nelle indagini di tipo telefonico, in cui la rilevazione è affidata ad una ditta esterna si ricorre in genere alla sostituzione in quanto è necessario garantire che tale ditta svolga un numero prefissato di interviste utili. Al fine di evidenziare il fenomeno della sostituzione, nei prospetti 1 e 2 vengono presentati i tassi di sostituzione, rispettivamente per regione e per tipologia comunale, osservati nel trimestre di indagine ottobre-dicembre.

3. Disegno di campionamento

3.1 Descrizione generale

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

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

La determinazione del numero totale di unità campionarie e la sua allocazione tra gli strati, è in genere, per un'indagine ad obiettivi plurimi come quella in esame, un'operazione complessa. È poco realistico, infatti, pensare di poter definire un campione che assicuri prefissati livelli di precisione per tutte le stime d'interesse, considerando anche il fatto che le stime vengono prodotte con diversi riferimenti temporali e territoriali. L'allocazione ottimale delle unità del campione con riferimento ad un dato tipo di dominio può risultare contrastante con l'allocazione ottimale con riferimento ad un altro tipo di dominio territoriale. Infatti, per quanto riguarda le stime riferite all'intero territorio nazionale l'allocazione ottimale risulta vicina a quella proporzionale tra le diverse regioni; per quanto riguarda, invece, le stime riferite alle regioni, l'allocazione ottimale risulta prossima a quella che assegna a tutte le regioni un campione di uguale numerosità. È necessario quindi ricorrere a un procedimento complesso articolato in più fasi.

Dapprima, mediando tra esigenze operative e di costo ed esigenze relative all'attendibilità delle principali stime di interesse, si è definito un numero complessivo di indirizzi campione n pari a circa 3.500. Successivamente, sulla base di valutazioni dell'errore di campionamento atteso delle principali stime a livello regionale e nazionale, è stata determinata l'allocazione del campione tra le regioni; si è ottenuta in tal modo un'allocazione di compromesso tra l'allocazione uniforme e quella proporzionale al peso demografico di ciascuna regione. Infine, le numerosità campionarie regionali sono state ripartite tra le diverse tipologie di comune in modo proporzionale alla popolazione residente.

Nei prospetti 1 e 2 sono riportati, rispettivamente per regione e per tipologia di comune, le numerosità campionarie teoriche e i tassi di sostituzione riferiti alla rilevazione del quarto trimestre d'indagine; nell'ambito di ciascuna regione il tasso di sostituzione è calcolato come rapporto percentuale tra il numero di indirizzi sostituiti ed il numero teorico di indirizzi campione.

Prospetto 1 - Distribuzione regionale del campione e tassi di sostituzione - Quarto trimestre 2007

REGIONI	Indirizzi campione	Tasso di sostituzione totale (a)	Tasso di sostituzione al netto dei mancati contatti (b)
Piemonte	265	26,8	23,8
Valle d'Aosta	92	41,3	30,4
Lombardia	330	27,6	23,0
Trentino-Alto Adige	206	30,6	21,4
<i>Bolzano-Bozen</i>	99	19,2	29,3
<i>Trento</i>	107	23,4	31,8
Veneto	201	26,9	20,9
Friuli-Venezia Giulia	139	29,5	23,7
Liguria	152	42,1	34,2
Emilia-Romagna	197	31,0	24,9
Toscana	213	29,6	26,3
Umbria	103	22,3	18,4
Marche	141	18,4	15,6
Lazio	219	30,6	22,4
Abruzzo	144	18,8	14,6
Molise	96	22,9	18,8
Campania	196	36,2	35,7
Puglia	187	37,4	35,3
Basilicata	99	34,3	28,3
Calabria	158	31,6	26,6
Sicilia	213	37,1	30,5
Sardegna	150	30,0	25,3
Italia	3.501	30,3	25,2

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

(b) Il tasso di sostituzione al netto dei mancati contatti è calcolato come rapporto percentuale tra il numero di famiglie sostituite, depurate dalle sostituzioni per mancato contatto, ed il numero teorico di famiglie campione. Per mancato contatto con la famiglia ci si riferisce ai casi di 7 tentativi di contatto telefonico con esito 'libero', effettuati in giorni e orari diversi.

Prospetto 2 - Distribuzione del campione per tipologia di comune e tassi di sostituzione - Quarto trimestre 2007

TIPOLOGIA DEL COMUNE	Indirizzi campione	Tasso di sostituzione totale	Tasso di sostituzione al netto dei mancati contatti
Comuni centro dell'area metropolitana	441	29,3	25,6
Comuni che gravitano intorno al centro dell'area metropolitana	316	28,2	26,3
Comuni aventi fino a 2.000 abitanti	310	41,6	31,9
Comuni con 2.001-10.000 abitanti	950	31,5	25,5
Comuni con 10.001-50.000 abitanti	868	27,5	23,5
Comuni con oltre 50.000 abitanti	616	28,4	22,7
Italia	3.501	30,3	25,2

3.2. Selezione del campione

L'estrazione degli indirizzi campione da ciascuno strato, ottenuto come incrocio della regione e della tipologia comunale, è stata effettuata con probabilità uguali e senza reimmissione mediante tecnica di selezione sistematica. La selezione del campione di indirizzi avviene nel seguente modo:

- per ciascuno strato si ordinano gli indirizzi dell'Archivio Telefonico delle Famiglie per ampiezza demografica del comune, strada, numero civico e piano;
- si definisce il passo di estrazione p come rapporto tra il numero di indirizzi in archivio e il numero di indirizzi da estrarre; si seleziona un numero casuale compreso tra 1 e p e si procede, poi, alla selezione degli indirizzi campione sulla base del passo di estrazione a partire dal punto di partenza casuale;
- ogni indirizzo selezionato dall'Archivio entra a far parte dell'*elenco base* e per ognuno di essi si selezionano i tre indirizzi successivi, che entrano a far parte dell'*elenco sostitutivo*; in tal modo le unità sostitutive sono le più vicine possibile dal punto di vista territoriale alle rispettive unità del campione base e quindi, presumibilmente, sono anche *simili* dal punto di vista delle caratteristiche socio-economiche.

4. Procedimento per il calcolo delle stime

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

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

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

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

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

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

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

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

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

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

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

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

- correggere l'effetto distorsivo dovuto agli errori di lista e al fenomeno della mancata risposta totale;
- tenere conto della conoscenza di alcuni totali noti sulla popolazione oggetto di studio, nel senso che le stime campionarie di tali totali devono coincidere con i rispettivi valori noti. Nell'indagine in oggetto i totali noti⁵ sono:
 - popolazione residente per sesso nelle 20 regioni italiane
 - popolazione residente per sesso nelle aggregazioni di comuni basate sulla tipologia socio demografica
 - popolazione per ripartizione territoriale e per sesso
 - famiglie per ampiezza
 - famiglie unipersonali (single) di età inferiore a 65 anni, per sesso
 - famiglie unipersonali (single) di 65 anni e più, per sesso.

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

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

⁵ I primi tre totali sono desunti da fonti demografiche (anagrafiche), mentre i totali riferiti alle famiglie derivano da stime dell'indagine annuale multiscopo "Aspetti della vita quotidiana".

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

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

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

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

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

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

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

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

5. Valutazione del livello di precisione delle stime

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

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

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

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

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

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

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

dove

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

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

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

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

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

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

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

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

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

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

6. Presentazione sintetica degli errori campionari

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

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

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

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

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

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

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

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

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

La prima colonna dei prospetti riporta K valori crescenti delle stime ${}_d\hat{Y}_t^k$ ($k=1, \dots, K$), la seconda riporta i rispettivi errori relativi interpolati $\hat{\varepsilon}({}_d\hat{Y}_t^k)$. Le informazioni contenute nei prospetti permettono di calcolare l'errore relativo di una generica stima di frequenza assoluta mediante due procedimenti che risultano di facile applicazione, anche se conducono a risultati meno precisi di quelli ottenibili mediante l'applicazione diretta dell'espressione (12).

Il 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 (5 o 6) che più si avvicina al valore della stima ${}_d\hat{Y}_t$.

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

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

dove: ${}_d\hat{Y}_t^{k-1}$ e ${}_d\hat{Y}_t^k$ sono i valori delle stime entro i quali è compresa la stima d'interesse ${}_d\hat{Y}_t$ e $\hat{\varepsilon}({}_d\hat{Y}_t^{k-1})$ e $\hat{\varepsilon}({}_d\hat{Y}_t^k)$ sono i corrispondenti errori relativi presenti nel prospetto.

6.2. Presentazione sintetica degli errori campionari per stime di totali riferiti ai viaggi e alle notti

Il modello utilizzato per le stime del totale del numero di viaggi e di notti, con riferimento al generico dominio d e al trimestre t , è del tipo seguente:

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

dove i parametri a , b e c vengono stimati utilizzando il metodo dei minimi quadrati, adattando il modello (14) a una nuvola di punti costituita da un consistente numero di coppie $(\hat{\sigma}({}_d\hat{Y}_t), {}_d\hat{Y}_t)$. In considerazione del fatto che il modello (14) è di tipo empirico, l'insieme delle stime ${}_d\hat{Y}_t$ utilizzate per interpolare il modello è stato determinato in modo da includere la maggior parte delle stime pubblicate. Al fine di ottenere un migliore adattamento alla nuvola di punti, è stato interpolato un modello - con riferimento ad ogni trimestre e con riferimento all'intero anno - per ciascuno dei seguenti sottogruppi di stime:

- numero di notti per viaggi di lavoro;
- numero di notti per viaggi di vacanza;
- numero totale di notti in viaggio;
- numero di viaggi di lavoro;
- numero di viaggi di vacanza
- numero totale di viaggi.

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

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

I prospetti 7 e 9 (per le stime relative alle notti e ai viaggi rispettivamente) riportano i valori dei coefficienti a, b, c e dell'indice di determinazione R^2 dei modelli utilizzati per l'interpolazione degli errori campionari delle stime di totali per area territoriale e per ciascuno dei sei tipi di stime sopra elencati, con riferimento al primo trimestre, al terzo trimestre e all'intero anno.

Nel paragrafo successivo verrà illustrato come calcolare l'errore di campionamento assoluto e relativo di una stima a partire dalle informazioni contenute nei prospetti.

Utilizzando i valori stimati dei parametri del modello (14) e dividendo ambo i membri del modello per il valore della stima, ${}_d\hat{Y}_t$, si perviene alla seguente equazione di secondo grado:

$$a + [b - \varepsilon({}_d\hat{Y}_t)]{}_d\hat{Y}_t + c({}_d\hat{Y}_t)^2 = 0$$

la cui radice positiva è espressa dalla seguente formula:

$${}_d\hat{Y}_t = \frac{-[b - \varepsilon({}_d\hat{Y}_t)] - \sqrt{[b - \varepsilon({}_d\hat{Y}_t)]^2 - 4ac}}{2c} \quad (15)$$

Nei prospetti 8 e 10 (per le stime relative alle notti e ai viaggi rispettivamente) vengono riportati i valori delle stime ${}_d\hat{Y}_t^*$ ottenuti sulla base della relazione (15) in corrispondenza di alcuni valori tipici prefissati dell'errore relativo percentuale, indicati come $\varepsilon^* = \varepsilon({}_d\hat{Y}_t^*)$; le stime con valori superiori a ${}_d\hat{Y}_t^*$ presentano valori dell'errore relativo inferiori a ε^* , mentre le stime che assumono valori inferiori a ${}_d\hat{Y}_t^*$ presentano valori dell'errore relativo superiori a ε^* . In tali prospetti, articolati per gruppo di stime, trimestre e area territoriale, i valori prefissati di ε^* sono: 1, 5, 10, 15, 20, 25 e 30%. L'utilizzo di tali prospetti verrà meglio chiarito nel paragrafo seguente attraverso un esempio numerico.

Prospetto 3 - Valori dei coefficienti a, b e di R² delle funzioni utilizzate per l'interpolazione degli errori campionari delle stime di frequenze assolute riferite agli individui per aree territoriali - Trimestri - Anno 2007

AREE TERRITORIALI	Trimestri 1,2,e 4			Trimestre 3		
	a	b	R ²	a	b	R ²
Nord	9,97265	-1,07205	94,1	9,38668	-1,02855	92,3
Centro	10,32616	-1,11707	93,2	9,58235	-1,05521	92,9
Mezzogiorno	9,91768	-1,08503	92,4	8,73257	-0,98431	89,3
Italia	10,92367	-1,13956	95,1	9,48417	-1,03407	92,6

Prospetto 4 - Valori dei coefficienti a, b e di R² delle funzioni utilizzate per l'interpolazione degli errori campionari delle stime di frequenze assolute riferite agli individui per aree territoriali - Media dei 4 trimestri - Anno 2007

AREE TERRITORIALI	a	b	R ²
Nord	9,46406	-1,12852	94,9
Centro	8,74660	-1,09205	94,0
Mezzogiorno	8,88694	-1,09431	92,5
Italia	9,83702	-1,15102	95,3

Prospetto 7 - Valori dei coefficienti a, b, c e di R2 delle funzioni utilizzate per l'interpolazione degli errori campionari delle stime di totali per gruppi di stime, trimestre e anno, e aree territoriali - Stime del numero di notti - Anno 2007Notti – lavoro

Notti – lavoro				
AREE TERRITORIALI	a	b	c	R ²
Gennaio-marzo				
Nord	47.508,66	0,441843	-0,000000032445	96,7
Centro	30.197,84	0,537517	-0,000000144196	80,2
Mezzogiorno	56.647,88	0,446816	-0,000000045119	95,1
Italia	127.847,54	0,270756	-0,000000009855	95,4
Luglio-settembre				
Nord	94.541,37	0,379015	-0,000000032534	87,0
Centro	35.318,78	0,385437	-0,000000093925	94,6
Mezzogiorno	48.678,41	0,494814	-0,000000015700	93,2
Italia	129.049,92	0,306573	-0,000000007029	88,9
Anno				
Nord	197.334,53	0,211327	-0,000000005088	90,5
Centro	111.241,77	0,344455	-0,000000017153	83,4
Mezzogiorno	162.171,73	0,256599	-0,000000004143	88,0
Italia	343.749,00	0,150710	-0,000000001267	89,0
Notti – vacanze				
AREE TERRITORIALI	a	b	c	R ²
Gennaio-marzo				
Nord	228.544,00	0,179132	-0,000000002012	89,0
Centro	103.691,95	0,222134	-0,000000006781	77,3
Mezzogiorno	140.024,26	0,279402	-0,000000007330	86,3
Italia	364.546,00	0,127795	-0,000000000810	90,7
Luglio-settembre				
Nord	640.313,19	0,069669	-0,000000000164	89,2
Centro	346.459,99	0,104853	-0,000000000757	89,3
Mezzogiorno	292.302,12	0,149195	-0,000000000786	85,3
Italia	802.553,77	0,057455	-0,000000000078	88,1
Anno				
Nord	726.792,04	0,060165	-0,000000000092	92,2
Centro	434.478,72	0,086447	-0,000000000381	89,8
Mezzogiorno	443.400,06	0,112495	-0,000000000336	89,3
Italia	1.037.127,39	0,046614	-0,000000000039	91,3
Notti - totali				
AREE TERRITORIALI	a	b	c	R ²
Gennaio-marzo				
Nord	278.583,98	0,164758	-0,000000001573	91,1
Centro	128.659,51	0,203269	-0,000000005581	80,5
Mezzogiorno	177.443,54	0,238916	-0,000000004927	88,7
Italia	426.717,26	0,114876	-0,000000000603	92,5
Luglio-settembre				
Nord	674.724,83	0,068156	-0,000000000150	90,8
Centro	352.078,75	0,101739	-0,000000000674	90,8
Mezzogiorno	395.209,17	0,141954	-0,000000000664	86,7
Italia	922.599,45	0,055390	-0,000000000068	89,2
Anno				
Nord	817.413,23	0,058302	-0,000000000084	92,8
Centro	502.469,40	0,084201	-0,000000000346	90,2
Mezzogiorno	570.992,22	0,104424	-0,000000000272	89,3
Italia	1.232.288,14	0,044395	-0,000000000034	91,8

Prospetto 8 - Valori dei totali corrispondenti ad alcuni valori tipici degli errori relativi percentuali per gruppi di stime, trimestre e anno, e aree territoriali - Stime del numero di notti - Anno 2007

Notti – lavoro							
AREE TERRITORIALI	1%	5%	10%	15%	20%	25%	30%
Gennaio-marzo							
Nord	13.419.171	12.197.244	10.673.310	9.154.989	7.645.494	6.150.956	4.684.410
Centro	3.714.704	3.441.774	3.101.696	2.763.217	2.426.967	2.093.941	1.765.778
Mezzogiorno	9.809.434	8.935.408	7.846.720	6.764.146	5.690.967	4.633.154	3.602.500
Italia	26.940.481	22.965.029	18.045.515	13.233.414	8.675.013	4.805.593	2.411.685
Luglio-settembre							
Nord	11.593.175	10.392.639	8.902.574	7.430.382	5.987.743	4.597.617	3.307.334
Centro	4.089.157	3.673.686	3.158.059	2.648.622	2.149.268	1.667.479	1.218.287
Mezzogiorno	30.980.833	28.441.924	25.270.802	22.103.580	18.942.189	15.790.063	12.653.927
Italia	42.626.383	37.000.767	30.002.631	23.072.562	16.290.016	9.903.070	4.777.960
Anno							
Nord	40.525.144	32.885.983	23.528.198	14.692.633	7.439.397	3.495.276	1.996.659
Centro	19.825.488	17.536.234	14.692.855	11.882.311	9.131.760	6.503.779	4.153.193
Mezzogiorno	60.172.899	50.640.427	38.807.462	27.170.795	16.093.724	7.103.404	2.921.719
Italia	113.406.146	82.734.351	45.915.034	16.750.869	6.036.893	3.321.268	2.259.229
Notti – vacanze							
AREE TERRITORIALI	1%	5%	10%	15%	20%	25%	30%
Gennaio-marzo							
Nord	85.387.754	65.901.299	42.030.443	20.122.842	6.666.560	2.973.819	1.834.807
Centro	31.763.653	25.972.405	18.822.795	11.919.994	5.869.231	2.362.652	1.205.177
Mezzogiorno	37.266.157	31.895.419	25.232.261	18.676.683	12.376.081	6.814.506	3.185.939
Italia	148.520.920	100.562.503	44.457.617	11.551.739	4.791.353	2.926.338	2.096.272
Luglio-settembre							
Nord	375.273.534	146.913.826	19.136.311	7.845.623	4.883.058	3.539.406	2.774.505
Centro	128.802.169	78.276.143	24.831.758	6.880.033	3.541.485	2.357.950	1.763.313
Mezzogiorno	179.179.906	129.091.565	68.058.059	18.779.901	5.316.243	2.836.942	1.919.091
Italia	625.293.300	160.000.283	18.253.379	8.609.640	5.612.965	4.161.135	3.305.380
Anno							
Nord	561.910.694	160.456.781	17.537.808	8.024.622	5.179.911	3.821.493	3.026.880
Centro	206.134.841	106.359.384	20.379.409	6.577.066	3.778.303	2.640.255	2.027.188
Mezzogiorno	309.308.757	192.838.324	59.401.568	10.781.006	4.972.184	3.199.584	2.354.795
Italia	967.176.167	125.387.297	19.159.182	9.994.000	6.749.996	5.094.344	4.090.507
Notti – totali							
AREE TERRITORIALI	1%	5%	10%	15%	20%	25%	30%
Gennaio-marzo							
Nord	100.130.186	75.290.175	45.086.354	18.798.766	6.192.743	3.091.717	2.012.760
Centro	35.286.022	28.280.108	19.676.832	11.542.783	5.103.343	2.183.714	1.241.193
Mezzogiorno	47.226.796	39.262.504	29.420.634	19.861.124	11.133.841	4.981.008	2.429.032
Italia	178.003.166	113.869.965	41.675.015	10.321.191	4.846.607	3.114.706	2.287.996
Luglio-settembre							
Nord	400.108.535	151.222.773	19.417.462	8.123.424	5.088.223	3.699.201	2.904.808
Centro	139.798.837	83.026.812	24.177.671	6.673.219	3.499.099	2.349.622	1.765.241
Mezzogiorno	201.573.160	142.584.815	71.474.129	19.075.360	6.347.463	3.579.037	2.474.856
Italia	684.514.797	162.229.890	20.065.149	9.683.910	6.360.803	4.732.897	3.767.751
Anno							
Nord	593.336.011	160.122.033	18.886.883	8.842.786	5.749.168	4.256.155	3.378.008
Centro	221.022.774	111.831.729	21.592.784	7.352.174	4.284.303	3.011.664	2.319.785
Mezzogiorno	352.649.581	209.824.853	54.624.703	11.709.044	5.875.832	3.893.926	2.907.765
Italia	1.057.021.709	125.429.383	21.872.111	11.625.809	7.905.833	5.987.614	4.818.013

Prospetto 9 - Valori dei coefficienti a, b, c e di R² delle funzioni utilizzate per l'interpolazione degli errori campionari delle stime di totali per gruppi di stime, trimestre e anno, e aree territoriali - Stime del numero di viaggi - Anno 2007

Viaggi – lavoro				
AREE TERRITORIALI	a	b	c	R ²
Gennaio-marzo				
Nord	9.639,85	0,423339	-0,000000100490	94,1
Centro	7.752,46	0,388922	-0,000000248371	83,5
Mezzogiorno	11.651,82	0,339955	-0,000000172258	88,7
Italia	19.958,46	0,258093	-0,000000033660	92,9
Luglio-settembre				
Nord	9.771,51	0,285998	-0,000000118068	94,9
Centro	8.458,95	0,371061	-0,000000232974	95,1
Mezzogiorno	10.382,19	0,362072	-0,000000160098	87,9
Italia	17.487,09	0,196628	-0,000000032088	92,1
Anno				
Nord	22.235,27	0,179457	-0,000000013575	93,1
Centro	15.210,36	0,200086	-0,000000034144	94,1
Mezzogiorno	23.092,94	0,162154	-0,000000017290	90,7
Italia	37.362,34	0,107559	-0,000000003811	94,0
Viaggi – vacanza				
AREE TERRITORIALI	a	b	c	R ²
Gennaio-marzo				
Nord	44.661,26	0,148280	-0,000000006536	87,8
Centro	19.759,43	0,162588	-0,000000016104	93,7
Mezzogiorno	23.033,53	0,126793	-0,000000006455	93,2
Italia	58.726,78	0,088388	-0,000000001694	88,9
Luglio-settembre				
Nord	49.633,80	0,090200	-0,000000002241	94,0
Centro	28.850,45	0,142352	-0,000000009349	93,5
Mezzogiorno	24.720,07	0,166046	-0,000000010514	90,5
Italia	72.557,08	0,071766	-0,000000001036	87,1
Anno				
Nord	83.305,58	0,068393	-0,000000000878	85,1
Centro	41.991,33	0,091822	-0,000000003027	91,3
Mezzogiorno	44.598,36	0,088864	-0,000000002872	87,0
Italia	117.855,46	0,046155	-0,000000000328	85,9
Viaggi – totali				
AREE TERRITORIALI	a	b	c	R ²
Gennaio-marzo				
Nord	40.206,56	0,161233	-0,000000007256	88,0
Centro	20.841,47	0,165798	-0,000000016795	91,8
Mezzogiorno	24.423,91	0,128592	-0,000000006618	91,6
Italia	54.713,45	0,094075	-0,000000001838	88,8
Luglio-settembre				
Nord	46.103,70	0,098484	-0,000000002622	85,5
Centro	27.138,77	0,140753	-0,000000008889	93,3
Mezzogiorno	24.466,13	0,168173	-0,000000010837	90,4
Italia	61.971,12	0,072450	-0,000000001032	87,4
Anno				
Nord	52.492,70	0,087112	-0,000000001092	85,9
Centro	31.719,96	0,107616	-0,000000002998	90,9
Mezzogiorno	39.305,59	0,084406	-0,000000001171	87,3
Italia	72.657,22	0,053673	-0,000000000287	87,4

Prospetto 10 - Valori dei totali corrispondenti ad alcuni valori tipici degli errori relativi percentuali per gruppi di stime, trimestre e anno, e aree territoriali - Stime del numero di viaggi - Anno 2007

Viaggi – lavoro							
AREE TERRITORIALI	1%	5%	10%	15%	20%	25%	30%
Gennaio-marzo							
Nord	4.136.439	3.740.841	3.247.177	2.754.893	2.264.864	1.778.872	1.301.111
Centro	1.545.823	1.387.084	1.189.510	993.379	799.679	610.465	430.523
Mezzogiorno	1.950.152	1.722.526	1.439.970	1.160.996	888.594	629.640	400.741
Italia	7.450.195	6.276.711	4.819.812	3.386.430	2.019.493	899.571	367.673
Luglio-settembre							
Nord	2.372.500	2.039.410	1.626.236	1.219.713	828.293	478.023	234.434
Centro	1.572.871	1.403.956	1.193.890	985.698	780.751	582.015	396.571
Mezzogiorno	2.228.206	1.981.975	1.675.648	1.371.908	1.072.780	782.858	513.902
Italia	5.908.420	4.685.902	3.182.608	1.762.373	687.551	280.382	161.110
Anno							
Nord	12.613.365	9.705.574	6.121.034	2.762.914	730.132	298.103	180.780
Centro	5.646.129	4.494.821	3.076.132	1.725.145	668.704	258.885	145.045
Mezzogiorno	8.949.445	6.686.459	3.934.319	1.559.446	497.233	250.527	164.148
Italia	25.974.597	15.725.463	4.275.843	819.951	397.653	260.484	193.408
Viaggi – vacanze							
AREE TERRITORIALI	1%	5%	10%	15%	20%	25%	30%
Gennaio-marzo							
Nord	21.474.676	15.478.052	8.218.167	2.485.739	785.538	427.327	290.725
Centro	9.602.778	7.162.497	4.179.963	1.565.444	443.493	217.347	141.452
Mezzogiorno	18.287.838	12.188.946	4.881.517	810.013	306.359	185.153	132.330
Italia	47.020.466	24.104.081	3.385.562	929.422	522.033	362.007	276.907
Luglio-settembre							
Nord	36.393.204	19.096.708	3.002.750	805.673	447.944	309.259	235.982
Centro	14.371.661	10.181.450	5.131.504	1.394.639	465.339	262.043	181.061
Mezzogiorno	14.998.908	11.246.639	6.636.172	2.475.830	612.045	284.323	181.943
Italia	60.743.708	23.925.578	2.364.599	916.315	563.254	406.130	317.449
Anno							
Nord	67.873.872	24.767.977	2.466.558	1.009.833	630.334	457.699	359.195
Centro	27.531.096	14.754.628	2.610.989	696.527	384.041	264.133	201.120
Mezzogiorno	28.012.978	14.595.478	2.452.996	706.074	397.218	275.423	210.627
Italia	113.444.118	13.982.091	2.160.379	1.130.881	764.820	577.626	464.003
Viaggi – totali							
AREE TERRITORIALI	1%	5%	10%	15%	20%	25%	30%
Gennaio-marzo							
Nord	21.104.100	15.682.423	9.050.751	3.251.893	889.147	437.310	285.479
Centro	9.408.374	7.070.322	4.212.320	1.679.507	490.987	236.373	152.393
Mezzogiorno	18.122.264	12.177.871	5.050.745	893.864	331.828	199.013	141.714
Italia	46.389.592	25.165.641	4.077.491	948.759	511.983	349.457	265.069
Luglio-settembre							
Nord	34.253.689	19.394.168	3.913.710	857.501	448.943	302.696	228.107
Centro	14.914.815	10.500.741	5.174.833	1.302.955	430.284	243.589	168.828
Mezzogiorno	14.748.409	11.107.596	6.631.071	2.559.074	632.497	288.011	182.844
Italia	61.503.119	24.236.556	2.086.364	790.789	483.962	348.329	272.005
Anno							
Nord	71.261.776	35.331.756	3.203.194	822.933	462.923	321.568	246.263
Centro	32.883.139	19.754.367	4.762.259	712.490	339.607	221.743	164.457
Mezzogiorno	64.078.430	30.489.207	2.167.736	592.948	338.868	236.964	182.133
Italia	153.730.610	23.540.049	1.553.409	752.591	496.058	369.883	294.861

7. Esempi di calcolo degli errori campionari

7.1. Esempi di calcolo degli errori campionari per stime di totali riferiti ai viaggi e alle notti

Esempio 1

Come visto nel paragrafo precedente, il prospetto 7 presenta il valore dei coefficienti a, b e c delle funzioni utilizzate per l'interpolazione degli errori con riferimento alle stime del numero di notti, mentre il prospetto 9 presenta gli stessi coefficienti con riferimento alle stime del numero di viaggi.

Sulla base dei dati riportati in tali prospetti, è possibile calcolare l'errore di campionamento assoluto e relativo di una stima a livello trimestrale.

Se ad esempio si volesse calcolare l'errore di campionamento assoluto per una stima appartenente al gruppo del numero di notti per lavoro riferita al Centro per il terzo trimestre, ${}_C\hat{Y}_3$, si possono leggere i valori dei coefficienti a, b e c dal prospetto 7 e si può utilizzare la seguente espressione:

$$\hat{\sigma}({}_C\hat{Y}_3) = 35.318,78 + 0,385437{}_C\hat{Y}_3 - 0,000000093925({}_C\hat{Y}_3)^2.$$

Esempio 2

La valutazione degli errori di campionamento delle stime mediante un metodo approssimato si può ottenere utilizzando i dati dei prospetti 8 e 10. Il prospetto 8 fa riferimento alle stime del numero di notti; il prospetto 10 fa riferimento alle stime del numero di viaggi. Tali prospetti sono articolati per trimestre e area territoriale e presentano gruppi di stime per valori prefissati degli errori di campionamento relativi pari all'1, 5, 10, 15, 20, 25 e 30%.

Considerando ad esempio il prospetto 8, si può osservare che il valore riportato all'incrocio tra la riga riferita alle notti per vacanza del terzo trimestre per l'Italia e la colonna relativa all'errore del 10% indica che tutte le stime superiori a 18.253.379 presentano un errore relativo inferiore al 10%; analogamente si avrà che le stime inferiori a tale valore presenteranno un errore relativo superiore al 10%.

Esempio 3

Per fare un ulteriore esempio e meglio specificare l'utilizzo del suddetto prospetto 8, si consideri la stima numero di notti per lavoro nel trimestre luglio-settembre pari a 18.936.000 (si veda il prospetto 1.1).

Dal prospetto 8, considerando le notti-lavoro ed esaminando l'ultima riga corrispondente al totale Italia per il trimestre luglio-settembre, si può osservare che il valore 18.936.000 risulta essere compreso tra i valori 23.072.562 e 16.290.016.

Da ciò si può facilmente desumere che il valore dell'errore relativo percentuale corrispondente è compreso tra il 15% ed il 20%.

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 delle persone che ha viaggiato nel trimestre gennaio-marzo è pari a 12.007.000.

Si cerca il livello di stima che più si avvicina a 12.007.000 nella prima colonna del prospetto 5, che è pari a 10.000.000. Dalla colonna del prospetto, riferita all'Italia per il primo trimestre, si ricava il corrispondente errore relativo percentuale della stima considerata, che è pari a 2,4%.

L'errore assoluto sarà:

$$\sigma(12.007.000) = 2,4/100 \times 12.007.000 = 288.168$$

L'intervallo di confidenza avrà come estremi:

$$12.007.000 - (2 \times 288.168) = 11.430.664$$

$$12.007.000 + (2 \times 288.168) = 12.583.336$$

Esempio 2

Considerando la stima precedente, si possono ottenere valori più precisi dell'errore di campionamento mediante interpolazione lineare dei due livelli di stima consecutivi tra i quali è compreso il valore della stima considerato.

Tali livelli sono 10.000.000 e 15.000.000 ai quali corrispondono i valori 2,4 e 1,9 percentuali.

L'errore corrispondente a 12.007.000 è pari a:

$$\sigma(12.007.000) = 2,4 - ((2,4 - 1,9) / (15.000.000 - 10.000.000)) \times (12.007.000 - 10.000.000) = 2,20\%$$

Il corrispondente errore assoluto è $2,20/100 \times 12.007.000 = 264.154$ e l'intervallo di confidenza avrà come estremi:

$$12.007.000 - (2 \times 264.154) = 11.478.692$$

$$12.007.000 + (2 \times 264.154) = 12.535.308$$

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 12.007.000 si ha:

$$\log \varepsilon^2 = 10,92367 - 1,13956 \times \log(12.007.000) = -7,6523.$$

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

Infatti, da $\log(\varepsilon^2) = -7,6523$ si ha:

$$\varepsilon^2 = \exp(-7,6523)$$

e dunque:

$$\varepsilon = 0,021793$$

L'errore relativo percentuale sarà dunque pari a 2,18 %.