Istat
Istituto Nazionale di Statistica

Indagine Viaggi e Vacanze

Periodo di riferimento: anno 2015

Aspetti metodologici dell'indagine

Anno 2017

INDICE

1. Introduzione 3
2. La popolazione di riferimento 4
3. Il disegno campionario 4
4. La rilevazione e il trattamento dei dati 9
5. La metodologia di calcolo dei pesi e degli errori campionari 10
6. La diffusione dei risultati dell'indagine 32
7. Glossario 33
8. Riferimenti bibliografici 34
9. Contatti 35

1. Introduzione

L'indagine ha la finalità di ottenere informazioni sui movimenti turistici della popolazione (domanda turistica). Le stime prodotte riguardano il numero di turisti, di viaggi, di pernottamenti in viaggio e di escursioni sul territorio nazionale o all'estero.

L'indagine viene svolta in conformità alle definizioni concettuali e metodologiche espresse dal Regolamento Europeo per le Statistiche del Turismo 692/2011, in vigore dal 1 gennaio 2012.

Il turismo è definito come l'insieme delle attività e dei servizi riguardanti le persone che si spostano al di fuori del loro 'ambiente abituale', per vacanza o per motivi di lavoro. Rientrano pertanto nei flussi turistici tutti gli spostamenti non abituali, con pernottamento (viaggi) o senza (escursioni). L'individuazione dell'ambiente abituale di una persona permette di distinguere correttamente il fenomeno turistico dalla mobilità, che non rientra nel campo di osservazione della domanda turistica.

Ad esempio, i viaggi e le escursioni abituali, quelli cioè effettuati settimanalmente nella stessa località, diversa dal luogo in cui si vive, sono comunque assimilabili all'ambiente abituale e non rientrano nei flussi turistici; si presuppone, infatti, che tali spostamenti siano riconducibili alla vita quotidiana e alle abitudini dell'individuo. Sono altresì esclusi dalla definizione di 'turista' le persone che si spostano giornalmente o settimanalmente per lavoro, per studio o per motivi personali, quando cioè lo spostamento rientra nell'ambito di attività di routine.

I viaggi turistici (non abituali) sono classificati, secondo gli standard internazionali, distinguendo i viaggi per motivi di lavoro da quelli per motivi di vacanza e le vacanze 'brevi' (da 1 a 3 notti) da quelle 'lunghe' (più di 3 notti). Tra le vacanze rientrano i viaggi per svago, piacere, relax, per visitare parenti o amici, per trattamenti di salute o per motivi religiosi.

In questa nota sono riportati in sintesi gli aspetti salienti dell'indagine. Per maggiori informazioni sulla rilevazione, è possibile consultare la pagina informativa dell'indagine sul sito dell'Istituto http://www.istat.it/it/archivio/123949

2. La popolazione di riferimento

La popolazione di riferimento è costituita dalle famiglie residenti e dagli individui che le compongono.

Per famiglia si intende la famiglia di fatto, definita come l'insieme delle persone coabitanti, legate da vincoli di matrimonio o parentela, affinità, adozione, tutela o affettivi e che compartecipano alla spesa familiare e/o condividono il reddito familiare.

Sono quindi escluse dalla popolazione di riferimento le persone che vivono permanentemente nelle comunità (caserme, ospedali, brefotrofi, istituti religiosi, convitti, eccetera) e quelle presenti ma non residenti sul territorio nazionale.

3. Il disegno campionario

Le informazioni sui viaggi e sulle escursioni sono rilevate attraverso un focus incluso nell'indagine sulle Spese delle Famiglie. 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 cinque grandi ripartizioni geografiche (Nord-ovest : Piemonte, Valle d'Aosta, Lombardia, Liguria; Nord-est: Trentino-Alto Adige, Veneto, Friuli-Venezia Giulia, Emilia-Romagna; Centro: Toscana, Umbria, Marche, Lazio; Sud: Abruzzo, Molise, Campania, Puglia, Basilicata, Calabria; Isole: Sicilia, Sardegna);
- la tipologia comunale, ottenuta suddividendo i comuni italiani in tre classi, formate in base a caratteristiche socio-economiche e demografiche (sulla base della popolazione residente per l'anno di riferimento), che consta delle seguenti modalità:
    - T1: comuni centro dell'area metropolitana: Torino, Milano, Venezia, Genova, Bologna, Firenze, Roma, Napoli, Bari, Palermo, Catania e Cagliari;
    - T2: comuni che gravitano intorno al centro dell'area metropolitana e comuni con più di 50.000 abitanti,
    - T3: rimanenti comuni.

Il disegno d'indagine è stato progettato, inoltre, per fornire stime a livello regionale sull'intero anno. Tuttavia, poiché per alcune regioni l'esiguità del fenomeno non permette di ottenere la robustezza delle stime, non sono diffuse stime per tale dominio territoriale.
La base di campionamento adottata, ossia la lista di selezione delle unità campionarie, è costituita dalle LAC, un archivio unico delle famiglie residenti nei comuni italiani, proveniente dalle liste anagrafiche comunali.

3.1. Descrizione generale

Il tipo di disegno adottato è a due stadi di selezione (comuni, famiglie). Ciò è determinato dal fatto che l'indagine utilizza l'intervista diretta con tecnica CAPI, e pertanto è necessario, per ragioni sia di costo e sia soprattutto organizzative, che il campione sia concentrato in un numero limitato di comuni.

Le unità di primo stadio, i comuni, sono stratificati all'interno di ciascuna regione geografica secondo le due variabili tipologia comunale e dimensione demografica. La prima è definita sulla base dei domini illustrati nel precedente paragrafo, la seconda variabile di stratificazione – espressa in termini di popolazione residente – viene utilizzata per costruire una soglia in base alla quale suddividere i comuni in due sottoinsiemi: (i) gli Auto Rappresentativi o AR, sono i comuni con popolazione al di sopra della soglia e costituiscono ognuno uno strato a sé; (ii) i Non Auto Rappresentativi o NAR, sono i restanti comuni. Per i comuni AR si adotta un disegno ad uno stadio stratificato, mentre per i NAR si usa un disegno a due stadi con stratificazione delle unità di primo stadio.

Da ogni comune campione si estrae un campione di famiglie anagrafiche che costituiscono le unità finali di campionamento.

Ogni trimestre si intervistano circa 7.000 famiglie, per un totale di 28.000 famiglie all'anno. Il disegno di campionamento, definito con riferimento a un generico trimestre dell'anno, viene replicato in modo identico per i quattro trimestri e viene anche effettuata una stratificazione mensile del campione trimestrale. Di conseguenza, la dimensione temporale può essere considerata un'ulteriore variabile di stratificazione del campione.

Essendo l'indagine basata su una rilevazione trimestrale, il disegno di campionamento viene definito con riferimento al generico trimestre di un anno e il campione di comuni risultante è coinvolto quattro volte in modo identico nella rilevazione nell'anno di riferimento. Inoltre, viene attuata una stratificazione mensile suddividendo il campione di ciascuno strato della popolazione nei tre mesi che compongono ciascun trimestre. In

particolare, la partecipazione dei comuni inclusi nel campione nell'arco dell'anno di riferimento è la seguente:

- ogni comune AR è coinvolto nell'indagine tutti i mesi (e il numero di famiglie campione è suddiviso nei mesi);
- ciascuno dei tre comuni campione di ogni strato NAR partecipa all'indagine quattro mesi nell'anno a distanza di tre mesi secondo lo schema seguente:

comuni	**mese di rilevazione**			
1	gennaio	aprile	luglio	ottobre
2	febbraio	maggio	agosto	novembre
3	marzo	giugno	settembre	dicembre

3.2. Stratificazione e selezione del campione

L'obiettivo della stratificazione è quello di formare gruppi (o strati) di unità caratterizzate, relativamente alle variabili oggetto d'indagine, da massima omogeneità interna agli strati e massima eterogeneità fra gli strati. Il raggiungimento di tale obiettivo si traduce in termini statistici in un guadagno nella precisione delle stime, ossia in una riduzione dell'errore campionario a parità di numerosità campionaria.

Nell'indagine in esame, i comuni vengono stratificati in base alla loro dimensione demografica e nel rispetto delle seguenti condizioni:

- auto-ponderazione del campione al livello di regione;
- definizione di un numero minimo di famiglie da intervistare per comune;
- stratificazione dei comuni sulla base dell'ampiezza demografica;
- formazione di strati di comuni di ampiezza approssimativamente costante in termini di popolazione residente.

Il procedimento di stratificazione è attuato all'interno di ogni dominio territoriale individuato, per ciascuna regione geografica, dalle tre tipologie comunali considerate.

Con riferimento al generico incrocio di regione e tipologia comunale, indicato nel seguito con il termine *dominio*, si denoti con: h $(h = 1, \ldots\ldots, H)$, l'indice di strato di comuni; i $(i = 1, \ldots\ldots, N)$, l'indice di comune; M_h, il numero di famiglie residenti nello strato h; M_{hi} il numero di famiglie residenti nel comune i dello strato h P_h, il numero di individui residenti

nello strato h; P_{hi} il numero di individui residenti nel comune i dello strato h; n, il numero di comuni campione in ogni strato.

Per la definizione del campione di comuni e di famiglie relativo ad un trimestre sono state dapprima effettuate le seguenti scelte:

(a) definizione del numero complessivo di famiglie campione a livello nazionale;

(b) definizione del numero m_r di famiglie campione per ciascuna regione;

(c) distribuzione delle numerosità regionali tra i domini definiti dalla tipologia comunale in misura proporzionale alla popolazione residente;

(d) scelta del numero minimo di famiglie, $\overline{m}_r$, da intervistare in ciascun comune campione.

Dalla scelta di $\overline{m}_r$ e m_r dipende la suddivisione dei comuni in AR e NAR e la formazione degli strati attraverso i seguenti passi:

(1) calcolo della frazione di campionamento regionale $f_r = m_r / M_r$, essendo M_r il numero di famiglie residenti nella regione r ;

(2) determinazione del valore della *soglia*[1] λ_r, mediante la relazione

$$\lambda_r = \frac{\overline{m}_r \delta_r}{f_r},$$

in cui δ_r è il numero medio di componenti per famiglia a livello regionale; risulta evidente da tale espressione che la soglia per la definizione dei comuni AR cresce al crescere di $\overline{m}_r$;

(3) suddivisione dei comuni in AR e NAR sulla base della soglia □r ;

(4) ordinamento decrescente dei comuni NAR all'interno di ogni dominio in funzione della loro dimensione demografica;

(5) suddivisione dei comuni NAR in strati la cui dimensione è approssimativamente uguale al prodotto $n \times \lambda_r$;

(6) selezione di n comuni campione da ciascuno strato h (h=1,…,H) con probabilità proporzionale all'ampiezza; per il generico comune c tale probabilità è espressa dalla formula:

$$z_{hi} = n \cdot P_{hi} / P_h \ ;$$

[1] Dal momento che il campione di famiglie di ogni regione è distribuito in modo proporzionale tra i domini e, di conseguenza, la frazione di campionamento è costante all'interno della regione, la soglia risulta anch'essa definita a livello regionale, così come la suddivisione dei comuni in AR e NAR.

(7) definizione del numero m_{hi} di famiglie da intervistare in ogni comune; dalla condizione di autoponderazione a livello regionale

$$\frac{n \cdot P_{hi}}{P_h} \frac{m_{hi}}{M_{hi}} = f_r ,$$

in cui il primo membro rappresenta la probabilità d'inclusione delle famiglie del comune c dello strato h (essendo le due frazioni rispettivamente la probabilità d'inclusione di primo e di secondo stadio), si ottiene

$$m_{hi} = \frac{f_r}{n} \frac{P_h \cdot M_{hi}}{P_{hi}} .$$

Al fine di raggiungere la numerosità campionaria desiderata in termini di comuni nel rispetto della condizione di auto-ponderazione a livello regionale, il numero minimo di interviste per comune è stato posto pari a 14 per le regioni con popolazione superiore a 1.500.000 di abitanti e 12 per le regioni più piccole.

Relativamente alla quantità m, per consentire la stratificazione mensile del campione di comuni, si è scelta una numerosità per ogni strato NAR pari a 3.

Nella tabella seguente vengono riportate le numerosità campionarie in termini di comuni.

	Comuni campione		
	AR	NAR	TOTALI
Mese	54	147	201
Trimestre	54	441	495
Anno	54	441	495

L'estrazione delle famiglie campione viene effettuata, mediante selezione casuale, dalla lista delle anagrafi comunali; per ciascun comune campione vengono selezionate quindi complessivamente $4m_{hi}$ famiglie che vengono ripartite nell'anno come segue:

- il campione di famiglie di ciascun comune AR viene suddiviso in 12 campioni mensili;
- il campione di famiglie di ciascun comune NAR viene suddiviso in 4 campioni mensili;
- ciascun campione mensile viene suddiviso in due gruppi, uno per ciascun dei due periodi di rilevazione di 14 giorni di ogni mese.

4. La rilevazione e il trattamento dei dati

Le informazioni sono rilevate attraverso un apposito focus incluso nell'intervista dell'indagine sulle Spese delle Famiglie. L'intervista è di tipo diretto, condotta mediante tecnica Capi (intervista faccia a faccia assistita da computer). Ogni individuo della famiglia viene intervistato sui viaggi e sulle escursioni effettuate nel periodo di riferimento. L'indagine è continua su tutti i mesi dell'anno, consentendo così di cogliere la stagionalità del fenomeno del turismo. Le famiglie del campione annuale sono suddivise in dodici sotto-campioni, ciascuno dei quali partecipa alla rilevazione in uno specifico mese di riferimento. Ogni famiglia riceve tre visite del rilevatore, secondo un preciso calendario. I quesiti sulla domanda turistica sono somministrati durante la prima e la terza visita (intervista iniziale e finale). La rilevazione dei viaggi e delle escursioni fa riferimento al mese, tuttavia nella rilevazione sono inseriti anche quesiti aventi periodi di riferimento diversi. In particolare, il numero di viaggiatori per vacanza viene rilevato anche con riferimento all'ultimo anno, così da poter soddisfare le richieste incluse nel nuovo Regolamento Europeo.

L'utilizzo della tecnica Capi consente di anticipare alla fase di acquisizione dell'informazione una parte considerevole dei controlli, ottenendo dati grezzi meno affetti da errore; le risposte fornite sono sottoposte ad un piano di verifica implementato nel questionario elettronico e quindi risolte sin dal momento della rilevazione. Tale piano si basa sul controllo dei domini delle variabili, dei percorsi del questionario e delle incongruenze logiche tra le informazioni raccolte.

Il questionario elettronico prevede sia controlli di tipo hard, sia controlli di tipo soft. Nel primo caso, a fronte di un'informazione incoerente, la registrazione del dato errato è inibita e quindi l'intervista non può proseguire finché esso non viene corretto; nel secondo caso, la registrazione del dato errato viene consentita previa segnalazione dell'errore.

Generalmente, un controllo di tipo hard è attivato se l'informazione risulta assolutamente inaccettabile, un controllo di tipo soft quando essa è altamente improbabile.

La strategia di controllo e correzione successiva cerca di salvaguardare il più possibile l'informazione rilevata, partendo dal presupposto che l'insieme dei controlli inseriti in fase di acquisizione fornisca dati grezzi di elevata qualità.

5. La metodologia di calcolo dei pesi e degli errori campionari

Dato il carattere campionario dell'indagine, per ottenere stime per l'intera popolazione di riferimento (le famiglie residenti) è necessario utilizzare un opportuno coefficiente di riporto all'universo. Il valore di tale coefficiente indica il numero di unità nella popolazione rappresentate dall'unità campionaria. L'indagine produce stime trimestrali e stime annuali.

Il calcolo del coefficiente di riporto all'universo utilizza una procedura generalizzata basata sull'uso di una famiglia di stimatori noti in letteratura come *calibration estimator* (stimatori di ponderazione vincolata). La metodologia alla base di tali stimatori consente la determinazione di un unico coefficiente di riporto all'universo in grado di produrre stime coerenti a totali noti, desunti da fonti esterne, che siano correlati alle principali variabili oggetto di indagine.

La famiglia di stimatori di ponderazione vincolata coincide asintoticamente con lo stimatore di regressione generalizzato e, per campioni sufficientemente grandi, ha approssimativamente le stesse proprietà, cioè si tratta di stimatori corretti, consistenti e con la stessa varianza campionaria[2].

La strategia adottata per la costruzione dei coefficienti di riporto all'universo è simile a quella utilizzata dalle varie indagini campionarie sulle famiglie dell'Istituto e si sviluppa in tre fasi:

- la determinazione della probabilità di inclusione di ogni unità statistica e del relativo peso diretto, pari all'inverso della probabilità di inclusione;
- calcolo dei coefficienti di correzione per mancata risposta totale;
- determinazione dei coefficienti di riporto all'universo finali vincolati ai totali noti desunti da fonti esterne all'indagine.

5.1 La probabilità di inclusione e il peso diretto

Il principio su cui è basato ogni metodo di stima campionaria è che le unità appartenenti al campione rappresentino anche le unità della popolazione non incluse nel campione stesso. Per questo motivo a ogni unità campionaria è possibile attribuire un peso, il coefficiente di riporto all'universo, che indica quante unità della popolazione essa rappresenta.

[2] La metodologia è illustrata da Deville, J.C. e Särndal, C.E. in *Calibration Estimators in Survey Sampling*, Journal of the American Statistical Association, Vol. 87, n.418, 1992.

Senza perdere di generalità, definiamo la seguente simbologia:

U popolazione di riferimento oggetto di indagine;
yk valore della variabile Y assunto dalla k-esima osservazione della popolazione;
yj valore della variabile Y assunto dalla j-esima osservazione della popolazione;
πj probabilità, assegnata dal disegno di campionamento, che l'unità j-esima sia inclusa nel campione S;

Il totale di una generica variabile Y, calcolato sull'intera popolazione, assume la seguente forma:

$$Y = \sum_{k \in U} y_k \qquad (1)$$

Il disegno di campionamento assegna le probabilità di inclusione ad ogni unità del campione in modo tale che

$$\hat{Y} = \sum_{j \in s} y_j \frac{1}{\pi_j} \qquad (2)$$

sia uno stimatore corretto della (1).
Nel disegno di campionamento dell'indagine, la probabilità di inclusione di un generico individuo è data dalla probabilità di estrazione del suo comune di residenza (direttamente proporzionale all'ampiezza demografica dei comuni all'interno dello strato) e dalla probabilità di estrazione della sua famiglia tra le famiglie eleggibili del comune. Per la natura del disegno campionario, la probabilità di inclusione, e dunque il peso base, è uguale per tutte le famiglie appartenenti allo stesso comune.

Per una generica famiglia eleggibile *j*, nel comune *i* dello strato *h*, il peso diretto d_{hij}, inverso della probabilità di inclusione π_{hij}, assume la seguente forma:

$$d_{hij} = \frac{1}{\pi_{hij}} = \frac{P_h}{P_{hi}} \frac{M_{hi}}{m_{hi}} \qquad (3)$$

h denota l'indice di strato;
i è l'indice di comune;

j denota l'indice della famiglia;

P_h indica il totale della popolazione residente nello strato *h*;

P_{hi} indica il totale della popolazione residente nel comune *i* dello strato *h*;

M_{hi} indica il totale di famiglie eleggibili nel comune *i* dello strato *h*;

m_{hi} indica il numero di famiglie campione nel comune *i* dello strato *h.*

I pesi base vengono di volta in volta calcolati sui dati anagrafici più aggiornati per popolazione e famiglie residenti.

5.2 La correzione per mancata risposta

Nel corso della fase di raccolta delle informazioni, come sempre accade nelle indagini, alcune unità campionarie non partecipano (per irreperibilità, rifiuto, interruzione definitiva o altri motivi). Tale fenomeno, noto come mancata risposta totale, implica che la numerosità campionaria teorica e quella effettivamente raggiunta siano diverse. Nella fase di calcolo dei coefficienti di riporto all'universo si tiene conto di tale fenomeno e viene introdotto un correttore per mancata risposta totale: sotto l'ipotesi che in determinati gruppi (ottenuti come aggregazioni di strati appartenenti alla stessa regione e tipologia comunale) il comportamento dei rispondenti sia simile a quello dei non rispondenti, il correttore assume la forma dell'inverso del tasso di risposta (δ_g):

$$\frac{1}{\delta_g} = \frac{m_g}{m_g^r} \qquad (4)$$

in cui m_g^r rappresenta il numero di famiglie rispondenti nel determinato gruppo *g.*

Il coefficiente di riporto all'universo, corretto per mancata risposta, per una generica famiglia eleggibile *j*, nel comune *i* dello strato *h*, del gruppo *g* (dato come intersezione tra regione e tipologia comunale), risulta pertanto essere:

$$k_{ghij} = d_{hij}\frac{1}{\delta_g} = \frac{P_h}{P_{hi}}\frac{M_{hi}}{m_{hi}}\frac{m_g}{m_g^r} \qquad (5)$$

5.3 La calibrazione a fonti esterne

Per il calcolo dei coefficienti di riporto all'universo finali si adottano gli stimatori di ponderazione vincolata (*calibration estimator*). La metodologia si basa sull'utilizzo di informazioni ausiliarie, sintetizzate in totali noti, che, essendo correlate con le variabili

principali oggetto di indagine, hanno la funzione di aumentare l'accuratezza delle stime. I pesi finali si ottengono risolvendo un problema di minimo vincolato, in cui la funzione da minimizzare è una funzione di distanza tra i pesi diretti corretti per la mancata risposta (**k**) e i pesi finali (**w**) delle famiglie del campione rispondente (S^r); i vincoli sono proprio le condizioni di uguaglianza tra le stime campionarie delle variabili ausiliarie e i rispettivi totali noti desunti da fonti esterne all'indagine[3]. Formalmente:

$$\begin{cases} Min\left\{\sum_{j\in S^r} dist\,(k_j, w_j)\right. \\ \sum_{j\in S^r} \boldsymbol{x}_j * w_j = t \end{cases}$$

dove t è il vettore dei totali noti e x_j è il vettore delle variabili ausiliarie osservate sulla j-esima unità campionaria appartenente al campione rispondente (S^r). La funzione di distanza utilizzata è la logaritmica troncata.

I totali noti, introdotti come vincoli nel calcolo dei pesi finali, consentono di migliorare l'accuratezza delle stime: quanto più le variabili ausiliarie considerate sono correlate con le variabili oggetto d'indagine, tanto più si riduce la distorsione delle stime. Nello specifico dell'Indagine sulle Spese delle Famiglie, le stime campionarie sono state vincolate ai seguenti totali noti:

- Popolazione residente per ripartizione, sesso e classi di età (0-14, 15-29, 30-44, 45-59, 60-74, 75 e più);
- Popolazione residente per regione;
- Famiglie residenti per regione;
- Popolazione residente per ripartizione e tipologia di comune (Comuni metropolitani, comuni periferia dei comuni metropolitani e comune oltre 50.000 abitanti, altri comuni);
- Famiglie residenti per ripartizione e tipologia di comune (Comuni metropolitani, comuni della cintura metropolitana e comuni con più di 50.000 abitanti, altri comuni);
- Popolazione straniera residente per ripartizione e sesso;

[3] La calibrazione è una calibrazione integrata, vale a dire che i coefficienti di riporto all'universo sono stati determinati in modo da poter essere utilizzati, indifferentemente, per la costruzione delle stime relative alle famiglie e agli individui. Tutti gli individui di una stessa famiglia hanno il medesimo coefficiente di riporto della famiglia stessa.

- Popolazione di 15 anni e più per ripartizione, condizione e posizione nella professione (Dirigenti, quadri e impiegati; operai e assimilati; imprenditori e liberi professionisti; lavoratori in proprio e altri autonomi; disoccupati; inattivi);
- Popolazione per ripartizione e mese;
- Famiglie per ripartizione e mese.

I vincoli imposti sono gli stessi sia per le stime trimestrali che per le stime annuali. Si è adottata una procedura di individuazione delle osservazioni anomale, vale a dire osservazioni con livelli e comportamenti di spesa che, seppur veri, sono molto lontani o dissimili dal collettivo socio-economico al quale appartengono. Tali unità campionarie rappresentano quindi un numero ristretto di famiglie dell'universo ed è quindi opportuno che abbiano associato un coefficiente di riporto abbastanza basso per non influenzare la stima dal punto di vista sia della distorsione sia della varianza. I gruppi socio-economici all'interno dei quali individuare comportamenti di spesa anomali sono stati definiti dall'incrocio tra ripartizione di appartenenza, tipologia di comune e numerosità familiare; all'interno di ogni gruppo, le famiglie anomale sono individuate tramite una *cluster analysis* che considera tutti i dati di spesa rilevati dall'indagine, sia a livello complessivo che rispetto alla loro composizione tra le diverse voci o capitoli.
Alle famiglie individuate come anomale è attribuito, nella fase precedente alla calibrazione finale, un peso corretto per la mancata risposta più basso rispetto a quello che avrebbero avuto tramite la procedura generalizzata. Questa procedura permette di mantenere nel campione le osservazioni anomale e rende possibile l'uso degli stimatori classici.
Per non avere famiglie con coefficienti troppo elevati è stata adottata una procedura di calibrazione ricorsiva, individuando le famiglie con valori dei coefficienti esterni a un intervallo di accettazione (inferiori al quinto percentile o superiori al 95esimo) e imponendo a queste valori dei coefficienti pari ai limiti del suddetto intervallo. La procedura di calibrazione viene nuovamente eseguita in modo da rispettare i vincoli imposti.

5.4 I dati campionari e il loro utilizzo

Il file dei microdati per la ricerca si compone di tre diversi archivi: individui, viaggi ed escursioni. I file sono riferiti all'intero anno e sono ottenuti mettendo in sequenza i quattro file trimestrali.
Il file individui è composto da un record per ciascun individuo componente la famiglia intervistata, include quindi sia gli individui che hanno viaggiato (turisti), sia quelli che non

hanno viaggiato. Per l'anno 2015, il file contiene n.35.574 records ed è di lunghezza pari a 107.

Il file viaggi è composto da un record per ciascun viaggio realizzato da un individuo. Per distinguere i viaggi di lavoro dai viaggi di vacanza è presente una variabile (TIPOVGG) che individua la tipologia del viaggio. Per una corretta utilizzazione dei dati, si rammenta che l'indagine rileva i viaggi che si concludono in ciascun mese di riferimento, indipendentemente dalla data di inizio. Per l'anno 2015, il file contiene n. 3.026 records ed è di lunghezza pari a 124.

Il file escursioni è composto da un record per ciascuna escursione realizzata da un individuo nel mese di riferimento. Per l'anno 2015, il file contiene n. 3.890 records ed è di lunghezza pari a 91.

Per un corretto utilizzo dei file, si precisa che è possibile effettuare elaborazioni sulle seguenti unità di analisi:

- individui: ogni individuo è identificato da: anno di riferimento (ANNRIF), mese di riferimento (MESE), numero progressivo univoco (PROGIND);
- viaggi: ogni viaggio è individuato da: anno di riferimento (ANNRIF), mese di riferimento (MESE), numero progressivo univoco dell'individuo che ha viaggiato (PROGIND); numero progressivo del viaggio tra quelli dello stesso individuo (PROGVIA);
- notti: le notti trascorse in viaggio sono espresse dalla durata del viaggio (DURATA) presente nel file viaggi.
- escursioni: ogni escursione è individuata da: anno di riferimento (ANNRIF), mese di riferimento (MESE), numero progressivo univoco dell'individuo che ha effettuato l'escursione (PROGIND); numero progressivo dell'escursione tra quelle dello stesso individuo (PROGESC).

Il coefficiente di riporto all'universo è stato determinato in modo da poter essere utilizzato indifferentemente per costruire stime relative agli individui, ai viaggi, alle notti e alle escursioni.

Il coefficiente può avere fino a sette cifre decimali separati dal punto (es. "37377.259718"). Per l'elaborazione dei dati si consiglia di utilizzarlo arrotondato all'unità.

Ad esempio, se in un record individuale del primo trimestre dell'anno il coefficiente di riporto all'universo vale "37377.259718", significa che l'individuo rappresenta 37.377,259718 individui nella popolazione residente in Italia in quel trimestre. Ponderando

quindi per questo valore i dati presenti nel suo record, le caratteristiche relative a tale individuo saranno stimate per 37.377,259718 individui.

Per quanto appena detto, è importante sottolineare che, mentre i viaggi, le notti e le escursioni rilevati in ciascun trimestre possono essere sommati per ottenere le stime annuali, non è corretto sommare i record trimestrali degli individui, per non quadruplicare la popolazione. Inoltre, poiché, una stessa persona può essere stata turista in mesi diversi, anche un'eventuale somma dei turisti mensili comporterebbe una sovrastima del numero dei turisti nel trimestre o nell'anno. Pertanto, per gli individui (turisti e non turisti), non è possibile ottenere stime di ammontare, ma solo medie.

Le stime ottenibili dal file dei microdati per la ricerca sono quindi di due tipologie:

- ammontare dei viaggi ed escursioni nel mese, nel trimestre e nell'anno. A questo fine si deve utilizzare il coefficiente denominato COEV;
- medie mensili (turisti, non turisti, viaggi ed escursioni); in tal caso si deve utilizzare il coefficiente denominato W.

Per completezza, si precisa che le stime dell'ammontare dei turisti e dei non turisti nell'anno e dei turisti nel trimestre sono fornite esclusivamente nel datawarehouse dell'Istituto I.Stat (vedere Par. 6) e pertanto non sono ottenibili a partire dal presente file.

Trattandosi di un'indagine di natura campionaria, l'utilizzatore dei dati dovrà valutare l'attendibilità delle stime ottenibili, sfruttando tutte le informazioni e gli strumenti di calcolo messi a disposizione, per procedere ad una corretta interpretazione dei dati. A tale scopo, si riportano, di seguito, le informazioni relative alla valutazione degli errori di stima, con esempi di calcolo.

5.5 Valutazione del livello di precisione delle stime

Le principali statistiche di interesse per valutare la variabilità campionaria delle stime prodotte da un'indagine sono l'errore di campionamento assoluto e l'errore di campionamento relativo. Indicando con $\hat{V}ar(\hat{Y}_d)$ la stima della varianza della generica stima $\hat{Y}_d$, la stima dell'errore di campionamento assoluto di $\hat{Y}_d$ si può ottenere mediante la seguente espressione

$$\hat{\sigma}(\hat{Y}_d) = \sqrt{\hat{V}ar(\hat{Y}_d)}\,; \qquad (1)$$

la stima dell'errore di campionamento relativo di $\hat{Y}_d$ è invece definita dall'espressione

$$\hat{\varepsilon}(\hat{Y}_d) = \frac{\hat{\sigma}(\hat{Y}_d)}{\hat{Y}_d}\,. \qquad (2)$$

Come è stato descritto nel paragrafo 5, le stime prodotte dall'indagine sono state ottenute mediante uno stimatore di ponderazione vincolata definito in base ad una funzione di distanza di tipo logaritmico troncato. Poiché, lo stimatore adottato non è funzione lineare dei dati campionari, per la stima della varianza $\hat{V}ar(\hat{Y}_d)$ si è utilizzato il metodo proposto da Woodruff; in base a tale metodo, che ricorre all'espressione linearizzata in serie di Taylor, è possibile ricavare la varianza di ogni stimatore non lineare (funzione regolare di totali) calcolando la varianza dell'espressione linearizzata ottenuta. In particolare, per la definizione dell'espressione linearizzata dello stimatore ci si è riferiti allo stimatore di regressione generalizzata, sfruttando la convergenza asintotica di tutti gli stimatori di ponderazione vincolata a tale stimatore, poiché nel caso di stimatori di ponderazione vincolata che utilizzano funzioni distanza differenti dalla distanza euclidea (che conduce allo stimatore di regressione generalizzata) non è possibile derivare l'espressione linearizzata dello stimatore. L'espressione linearizzata dello stimatore è data, quindi, da

$$\hat{Y}_d \cong \hat{Z}_d = \sum_{h=1}^{H_d} \hat{Z}_h \quad , \quad \text{essendo} \quad \hat{Z}_h = \sum_{i=1}^{n_h} \sum_{j=1}^{m_{hi}} Z_{hij} W_{hij} \tag{3}$$

dove Z_{hij} è la variabile linearizzata espressa come $Z_{hij} = Y_{hij} - \mathbf{X}_{\mathbf{hij}}^{'}\beta$, essendo $\mathbf{X}_{\mathbf{hij}} = \left({}_1X_{hij},\cdots,{}_kX_{hij},\cdots,{}_KX_{hij}\right)'$ il vettore contenente i valori delle variabili ausiliarie, osservati per la generica famiglia hij e $\hat{\beta}$, il vettore dei coefficienti di regressione del modello lineare che lega la variabile di interesse y alle K variabili ausiliarie x. In base alla (3), si ha, quindi, che la stima della varianza della generica stima $\hat{Y}_d$ è ottenuta mediante la seguente relazione

$$\hat{V}ar\left(\hat{Y}_d\right) \cong \hat{V}ar\left(\hat{Z}_d\right) = \sum_{h=1}^{H_d} \hat{V}ar\left(\hat{Z}_h\right). \tag{4}$$

Dalla (4) risulta che la stima della varianza della stima $\hat{Y}_d$ viene calcolata come somma della stima delle varianze dei singoli strati, AR e NAR, appartenenti al dominio d. La formula di calcolo della varianza, $\hat{V}ar\left(\hat{Z}_h\right)$, della stima $\hat{Z}_h$ è differente a seconda che lo strato sia AR oppure NAR. Possiamo, quindi scomporre come segue

$$\hat{V}ar\left(\hat{Y}_d\right) \cong \hat{V}ar\left(\hat{Z}_d\right) = \sum_{h=1}^{H_{AR}} \hat{V}ar\left(\hat{Z}_h\right) + \sum_{h=1}^{H_{NAR}} \hat{V}ar\left(\hat{Z}_h\right) , \tag{5}$$

in cui H_{AR} e H_{NAR} indicano rispettivamente il numero di strati AR e NAR appartenenti al dominio d.

Negli strati AR (in cui ciascun comune fa strato a sé e $N_h = n_h = 1$, l'indice i di comune diviene superfluo e viene omesso) la varianza è stimata mediante la seguente espressione

$$\sum_{h=1}^{H_{AR}} \hat{V}ar\left(\hat{Z}_h\right) = \sum_{h=1}^{H_{AR}} M_h \frac{(M_h - m_h)}{m_h(m_h - 1)} \sum_{j=1}^{m_h} \left(Z_{hj} - \bar{Z}_h\right)^2 , \qquad (6)$$

dove si è posto $M_h = M_{hi}$, $m_h = m_{hi}$, $Z_{hj} = Z_{hij}$ e $\bar{Z}_h = \frac{1}{m_h} \sum_{j=1}^{m_h} Z_{hj}$.

Per l'insieme degli strati NAR la varianza viene stimata invece mediante la formula seguente

$$\sum_{h=1}^{H_{NAR}} \hat{V}ar\left(\hat{Z}_h\right) = \sum_{h=1}^{H_{NAR}} \frac{n_h}{n_h - 1} \sum_{i=1}^{n_h} \left(\hat{Z}_{hi} - \frac{\hat{Z}_h}{n_h} \right)^2 \qquad (7)$$

dove le quantità sono espresse come

$$\hat{Z}_{hi} = \sum_{j=1}^{m_{hi}} Z_{hij} W_{hij} \qquad \text{e} \qquad \hat{Z}_h = \sum_{i=1}^{n_h} \sum_{j=1}^{m_{hi}} Z_{hij} W_{hij} .$$

Utilizzando le espressioni (6) e (7) è possibile, infine, calcolare la varianza di campionamento, $\hat{V}ar\left(\hat{Y}_d\right)$, in base alla (5) e calcolare, quindi, in base alla (1) ed alla (2) rispettivamente l'errore di campionamento assoluto e l'errore di campionamento relativo.

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

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

in cui si è indicato con $\hat{V}ar(_d\hat{Y}_t)$ la stima della varianza della stima $_d\hat{Y}_t$, riferita al dominio d e al trimestre di indagine t, ottenuta secondo il procedimento sopra descritto.

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

$$\hat{\varepsilon}(_d\hat{R}) = \sqrt{\hat{\varepsilon}^2(_d\hat{Y}_1) - \hat{\varepsilon}^2(_d\hat{Y}_2)} . \qquad (9)$$

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

$$\left\{\hat{Y} - k\hat{\sigma}(\hat{Y}), \hat{Y} + k\hat{\sigma}(\hat{Y})\right\} \qquad (10)$$

Nella (10) il valore di k dipende dal valore fissato per P; ad esempio, per P=0,95 si ha k=2.

5.6 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.

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

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

$$\log \hat{\varepsilon}^2({}_d\hat{Y}_t) = a + b \log({}_d\hat{Y}_t) \qquad (11)$$

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.

Il prospetto 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.

Inoltre, allo scopo di facilitare il calcolo degli errori campionari, nel prospetto 3 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,…,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 (11).

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 4 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:

$$\hat{\varepsilon}({}_d\hat{Y}_t) = \hat{\varepsilon}({}_d\hat{Y}_t^{k-1}) + \frac{\hat{\varepsilon}({}_d\hat{Y}_t^{k}) - \hat{\varepsilon}({}_d\hat{Y}_t^{k-1})}{{}_d\hat{Y}_t^{k} - {}_d\hat{Y}_t^{k-1}}({}_d\hat{Y}_t - {}_d\hat{Y}_t^{k-1}) \qquad (12)$$

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.

5.6.2 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, è del tipo seguente:

$$\hat{\sigma}({}_d\hat{Y}) = a + b\ {}_d\hat{Y} + c\ {}_d\hat{Y}^{\,2} \qquad (13)$$

dove i parametri a, b e c sono stimati utilizzando il metodo dei minimi quadrati, adattando il modello (13) a una nuvola di punti costituita da un consistente numero di coppie $(\hat{\sigma}({}_d\hat{Y}), {}_d\hat{Y})$. In considerazione del fatto che il modello (13) è di tipo empirico, l'insieme delle stime ${}_d\hat{Y}$ 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 ai trimestri e/o all'intero anno a seconda del dominio territoriale - 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;
- numero totale di escursioni.

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. Fanno eccezione le stime delle escursioni per cui è stato studiato un modello di interpolazione per ogni trimestre, data la difformità del fenomeno nei 4 diversi periodi dell'anno.

I prospetti 4, 6 (per le stime relative alle notti e ai viaggi) 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 (13) e dividendo ambo i membri del modello per il valore della stima, ${}_d\hat{Y}$, si perviene alla seguente equazione di secondo grado:

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

la cui radice positiva è espressa dalla seguente formula:

$${}_d\hat{Y} = \frac{-[b - \varepsilon({}_d\hat{Y})] - \sqrt{[b - \varepsilon({}_d\hat{Y})]^2 - 4ac}}{2c} \qquad (14)$$

Nei prospetti 5 ed 7 (per le stime relative alle notti e ai viaggi) vengono riportati i valori delle stime ${}_d\hat{Y}$ ottenuti sulla base della relazione (14) in corrispondenza di alcuni valori tipici prefissati dell'errore relativo percentuale, indicati come $\varepsilon^* = \varepsilon({}_d\hat{Y})$; le stime con valori superiori a ${}_d\hat{Y}$ presentano valori dell'errore relativo inferiori a ε^*, mentre le stime che assumono valori inferiori a ${}_d\hat{Y}$ 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.

I prospetti 6 e 7 riportano rispettivamente 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 dei viaggi ed i valori delle stime ${}_d\hat{Y}^*$ ottenuti in corrispondenza di alcuni valori tipici prefissati dell'errore relativo percentuale. I viaggi vengono distinti in viaggi di lavoro, di vacanza e totali, facendo inoltre riferimento alla tipologia comunale, espressa nelle 3 seguenti modalità: comuni centro dell'area metropolitana, comuni periferici della stessa o con comunque più di 50.000 abitanti e comuni al di sotto di quest'ultima soglia.

I prospetti 8 e 9 riportano invece 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 delle escursioni totali, con riferimento ad ogni trimestre e all'intero anno, e i valori delle stime ${}_d\hat{Y}$ ottenuti in corrispondenza di alcuni valori tipici prefissati dell'errore relativo percentuale, analogamente a quanto visto nei prospetti 6 e 7.

Prospetto 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 - Trimestri – Anno 2015

AREE TERRITORIALI	Trimestri 1,2,e 4			Trimestre 3		
	a	b	R^2	a	b	R^2
Nord-Ovest	10,27981	-1,07492	0,91704	10,60305	-1,08611	0,92105
Nord-Est	11,10075	-1,18965	0,94554	11,07804	-1,17597	0,93607
Centro	11,04105	-1,17474	0,93737	10,81805	-1,12020	0,91155
Sud	10,81937	-1,15939	0,95151	8,56189	-0,98571	0,91332
Isole	10,33363	-1,10748	0,92326	10,08340	-1,09035	0,92914
Italia	11,14037	-1,14704	0,95096	10,77109	-1,11052	0,94945

Prospetto 3 - Valori interpolati degli errori relativi percentuali delle stime di frequenze assolute riferite agli individui per aree territoriali - Trimestri – Anno 2015

STIME	Trimestri 1,2 e 4						Trimestre 3					
	Aree territoriali						Aree territoriali					
	Nord-Ovest	Nord-Est	Centro	Sud	Isole	Italia	Nord-Ovest	Nord-Est	Centro	Sud	Isole	Italia
20000	83,3	71,1	74,3	71,8	72,8	89,6	92,6	75,3	87,1	54,9	69,9	89,3
30000	67,0	55,9	58,6	56,8	58,2	71,0	74,3	59,3	69,4	44,9	56,1	71,3
40000	57,4	47,1	49,5	48,0	49,6	60,2	63,6	50,1	59,1	39,0	47,9	60,8
50000	50,9	41,2	43,4	42,2	43,8	53,0	56,3	43,9	52,1	34,9	42,4	53,7
60000	46,1	37,0	39,0	38,0	39,6	47,7	51,0	39,5	47,1	31,9	38,4	48,5
70000	42,5	33,8	35,6	34,7	36,4	43,7	46,9	36,0	43,2	29,6	35,3	44,5
80000	39,5	31,2	32,9	32,1	33,8	40,5	43,6	33,3	40,1	27,7	32,9	41,3
90000	37,1	29,1	30,7	30,0	31,7	37,8	40,9	31,1	37,5	26,1	30,8	38,7
100000	35,1	27,3	28,9	28,2	29,9	35,6	38,7	29,2	35,4	24,8	29,1	36,5
200000	24,2	18,1	19,2	18,9	20,3	23,9	26,5	19,4	24,0	17,6	19,9	24,9
300000	19,4	14,2	15,2	14,9	16,3	19,0	21,3	15,3	19,1	14,4	16,0	19,8
400000	16,6	12,0	12,8	12,6	13,9	16,1	18,2	12,9	16,3	12,5	13,7	16,9
500000	14,8	10,5	11,2	11,1	12,3	14,1	16,1	11,3	14,4	11,2	12,1	14,9
750000	11,9	8,2	8,8	8,8	9,8	11,2	12,9	8,9	11,4	9,2	9,7	11,9
1000000	10,2	6,9	7,5	7,4	8,3	9,5	11,1	7,5	9,7	8,0	8,3	10,2
2000000	7,0	4,6	5,0	5,0	5,7	6,4	7,6	5,0	6,6	5,7	5,7	6,9
3000000	5,6	3,6	3,9	3,9	4,5	5,1	6,1	4,0	5,3	4,6	4,6	5,5
4000000	4,8	3,0	3,3	3,3	3,9	4,3	5,2	3,3	4,5	4,0	3,9	4,7
5000000	4,3	2,7	2,9	2,9	3,4	3,8	4,6	2,9	4,0	3,6	3,4	4,2
7500000	3,4	2,1	2,3	2,3	2,7	3,0	3,7	2,3	3,2	3,0	2,8	3,3
10000000	3,0	1,8	1,9	2,0	2,3	2,5	3,2	1,9	2,7	2,6	2,4	2,8

Prospetto 4 - Valori dei coefficienti a, b, c e di R^2 delle funzioni utilizzate per l'interpolazione degli errori campionari delle stime di totali per gruppi di stime, trimestre e anno, aree territoriali - Stime del numero di notti – Anno 2015

NOTTI – LAVORO				
AREE TERRITORIALI	a	b	c	R^2
		Gennaio-marzo		
Italia	47976,740	0,410345800	-0,000000034	0,91
		Luglio-settembre		
Italia	39.699,060	0,526441500	-0,000000031	0,91
		Anno		
Nord-Ovest	11.652,62	0,80820273	-0,00000020	0,94
Nord-Est	83.466,69	0,31880478	-0,00000002	0,93
Centro	53.829,11	0,45211743	-0,00000005	0,93
Sud	-10.089,62	0,65683235	-0,00000005	0,94
Isole	6.922,99	0,78616520	-0,00000073	0,97
Italia	158.285,23	0,24209930	-0,000000005	0,90
NOTTI – VACANZE				
AREE TERRITORIALI	a	b	c	R^2
		Gennaio-marzo		
Italia	193.936,20	0,197299500	-0,000000003	0,93
		Luglio-settembre		
Italia	706.797,90	0,1183088000	-0,0000000002	0,94
		Anno		
Nord-Ovest	440.705,164	0,22551322	-0,00000000078	0,94
Nord-Est	421.593,106	0,16811490	-0,00000000042	0,91
Centro	243.833,044	0,24404917	-0,00000000289	0,96
Sud	185.350,613	0,24376334	-0,00000000236	0,93
Isole	66.419,580	0,52605977	-0,00000001361	0,95
Italia	929.095,257	0,10581492	-0,00000000012	0,93
NOTTI – TOTALI				
AREE TERRITORIALI	a	b	c	R^2
		Gennaio-marzo		
Italia	269.024,000	0,176710400	-0,0000000020	0,91
		Luglio-settembre		
Italia	845.391,800	0,118879600	-0,00000000025	0,93
		Anno		
Nord-Ovest	418.569,206	0,225177493	-0,00000000074	0,93
Nord-Est	457.086,482	0,160210992	-0,00000000037	0,88
Centro	308.137,350	0,209048001	-0,00000000166	0,95
Sud	166.883,089	0,306087927	-0,00000000526	0,91
Isole	51.291,593	0,507047320	-0,00000001123	0,95
Italia	1.031.484,088	0,099973208	-0,00000000011	0,93

Prospetto 5 - Valori dei totali corrispondenti ad alcuni valori tipici degli errori relativi percentuali per gruppi di stime, trimestre e anno, aree territoriali - Stime del numero di notti – Anno 2015

NOTTI – LAVORO							
AREE TERRITORIALI	1%	5%	10%	15%	20%	25%	30%
			Gennaio-marzo				
Italia	**11.987.512**	**10.814.521**	**9.352.751**	**7.898.427**	**6.456.324**	**5.036.123**	**3.659.989**
			Luglio-settembre				
Italia	**16.529.444**	**15.261.471**	**13.678.150**	**12.097.320**	**10.520.081**	**8.948.285**	**7.385.285**
			Anno				
Nord-Ovest	4.079.615	3.876.665	3.623.098	3.369.691	3.116.483	2.863.526	2.610.893
Nord-Est	14.485.130	12.680.967	10.443.510	8.239.559	6.100.705	4.104.673	2.440.673
Centro	8.429.871	7.689.602	6.767.745	5.851.379	4.943.360	4.048.802	3.177.541
Sud	13.976.594	13.110.286	12.027.192	10.943.801	9.860.014	8.775.684	7.690.580
Isole	1.068.222	1.014.098	946.519	879.041	811.689	744.496	677.506
Italia	**43.422.128**	**36.188.677**	**27.243.623**	**18.536.714**	**10.524.510**	**4.720.701**	**2.256.358**
NOTTI – VACANZE							
AREE TERRITORIALI	1%	5%	10%	15%	20%	25%	30%
			Gennaio-marzo				
Italia	70.092.563	55.607.184	37.775.344	20.869.996	7.973.633	3.164.655	1.802.575
			Luglio-settembre				
Italia	**471.320.787**	**303.201.375**	**106.950.826**	**19.505.610**	**8.448.504**	**5.317.070**	**3.870.892**
			Anno				
Nord-Ovest	279.329.027	228.314.966	164.934.801	102.684.475	45.335.526	12.798.650	5.590.462
Nord-Est	378.548.137	284.330.004	167.904.919	59.820.974	11.482.822	5.019.186	3.164.727
Centro	82.067.005	68.422.370	51.515.104	34.977.722	19.566.586	8.215.709	3.664.731
Sud	100.045.568	83.220.104	62.306.892	41.700.438	22.137.611	7.645.613	2.935.125
Isole	38.044.281	35.116.092	31.458.642	27.805.416	24.158.305	20.520.509	16.897.885
Italia	**797.707.867**	**475.189.362**	**114.547.784**	**19.934.022**	**9.742.063**	**6.409.133**	**4.770.339**
NOTTI – TOTALI							
AREE TERRITORIALI	1%	5%	10%	15%	20%	25%	30%
			Gennaio-marzo				
Italia	**83.871.427**	**64.599.771**	**41.095.474**	**19.867.893**	**7.129.558**	**3.358.837**	**2.108.963**
			Luglio-settembre				
Italia	**448.238.603**	**290.513.111**	**108.059.731**	**22.974.101**	**10.110.077**	**6.370.952**	**4.638.217**
			Anno				
Nord-Ovest	294.317.102	240.398.259	173.372.567	107.445.134	46.454.590	12.344.508	5.316.184
Nord-Est	403.767.208	298.121.717	167.899.567	51.104.089	10.457.543	4.986.858	3.241.655
Centro	121.321.642	97.619.568	68.341.886	40.155.118	16.610.060	6.042.778	3.200.750
Sud	56.811.401	49.296.931	39.948.021	30.688.082	21.621.870	13.080.050	6.238.736
Isole	44.360.429	40.807.761	36.369.380	31.934.780	27.505.769	23.085.523	18.680.139
Italia	**836.476.732**	**478.103.635**	**97.139.779**	**19.767.005**	**10.198.696**	**6.841.317**	**5.142.315**

Prospetto 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 anno, aree territoriali - Stime del numero di viaggi – Anno 2015

VIAGGI – LAVORO				
AREE TERRITORIALI	a	b	c	R^2
		Gennaio-marzo		
Italia	29.618,970	0,256655000	-0,0000000689	0,93
		Luglio-settembre		
Italia	15.736,820	0,439287100	-0,000000130	0,97
		Anno		
Nord-Ovest	3.920,704	0,598846634	-0,0000003083	0,98
Nord-Est	15.218,184	0,335928646	-0,0000000719	0,91
Centro	10.300,625	0,495361743	-0,0000002500	0,95
Sud	5.607,366	0,469369391	-0,0000003643	0,96
Isole	580,999	0,847037029	-0,0000012344	0,98
Italia	**41.066,362**	**0,198324651**	**-0,0000000159**	**0,96**

VIAGGI – VACANZA				
AREE TERRITORIALI	a	b	c	R^2
		Gennaio-marzo		
Italia	58.995,920	0,160572300000	-0,000000009962	0,96
		Luglio-settembre		
Italia	126.490,673	0,118082060	-0,0000000029	0,94
		Anno		
Nord-Ovest	91684,367	0,205717433	-0,0000000085	0,93
Nord-Est	77363,379	0,116736096	-0,0000000030	0,96
Centro	55254,487	0,172228513	-0,0000000075	0,97
Sud	54599,272	0,193477990	-0,0000000140	0,93
Isole	7203,501	0,515055627	-0,0000000948	0,97
Italia	**190113,130**	**0,077697273**	**-0,0000000007**	**0,96**

VIAGGI – TOTALI				
AREE TERRITORIALI	a	b	c	R^2
		Gennaio-marzo		
Italia	65.870,000	0,140100000	-0,0000000073	0,95
		Luglio-settembre		
Italia	124.609,198	0,114312439	-0,0000000026	0,93
		Anno		
Nord-Ovest	89.370,674	0,196599260	-0,0000000075	0,93
Nord-Est	77.910,850	0,116548400	-0,0000000029	0,96
Centro	56.710,290	0,168135097	-0,0000000068	0,97
Sud	56.529,220	0,168782869	-0,0000000103	0,93
Isole	9.298,217	0,459873537	-0,0000000757	0,98
Italia	191.012,539	0,072243506	-0,0000000006	0,96

Prospetto 7 - Valori dei totali corrispondenti ad alcuni valori tipici degli errori relativi percentuali per gruppi di stime, trimestre e anno, aree territoriali - Stime del numero di viaggi – Anno 2015

VIAGGI – LAVORO							
AREE TERRITORIALI	1%	5%	10%	15%	20%	25%	30%
			Gennaio-marzo				
Italia	**3.695.762**	**3.136.039**	**2.448.898**	**1.788.154**	**1.184.926**	**705.679**	**412.639**
			Luglio-settembre				
Italia	**3.349.664**	**3.044.562**	**2.664.337**	**2.285.963**	**1.910.487**	**1.539.869**	**1.178.166**
			Anno				
Nord-Ovest	1.916.732	1.787.460	1.625.977	1.464.649	1.303.533	1.142.716	982.343
Nord-Est	4.577.477	4.027.682	3.343.305	2.664.302	1.995.772	1.351.212	773.151
Centro	1.962.500	1.804.336	1.607.132	1.410.696	1.215.384	1.021.801	831.050
Sud	1.272.918	1.164.259	1.028.766	893.790	759.598	626.661	495.902
Isole	678.789	646.419	605.962	565.512	525.070	484.639	444.222
Italia	**12.022.268**	**9.569.145**	**6.557.679**	**3.721.826**	**1.552.983**	**660.183**	**381.113**
VIAGGI – VACANZE							
AREE TERRITORIALI	1%	5%	10%	15%	20%	25%	30%
			Gennaio-marzo				
Italia	**15.496.615**	**11.609.367**	**6.934.276**	**3.021.323**	**1.157.677**	**617.261**	**411.056**
			Luglio-settembre				
Italia	**38.801.154**	**25.457.197**	**10.499.618**	**3.099.369**	**1.468.566**	**939.652**	**687.854**
			Anno				
Nord-Ovest	23.515.874	18.915.367	13.268.134	7.926.456	3.640.456	1.587.405	899.583
Nord-Est	35.903.318	23.107.983	8.514.163	1.971.396	899.653	573.054	419.233
Centro	21.864.199	16.660.672	10.297.308	4.558.456	1.432.677	667.324	421.948
Sud	13.412.082	10.627.850	7.225.242	4.068.836	1.756.499	805.472	482.058
Isole	5.339.304	4.918.778	4.393.447	3.868.615	3.344.514	2.821.545	2.300.465
Italia	**97.441.874**	**44.700.299**	**6.968.111**	**2.564.395**	**1.540.577**	**1.098.362**	**852.861**
VIAGGI – TOTALI							
AREE TERRITORIALI	1%	5%	10%	15%	20%	25%	30%
			Gennaio-marzo				
Italia	**18.348.870**	**13.058.499**	**6.827.797**	**2.403.177**	**982.298**	**577.270**	**404.490**
			Luglio-settembre				
Italia	**41.097.038**	**26.427.341**	**10.170.330**	**2.883.237**	**1.394.915**	**902.669**	**664.851**
			Anno				
Nord-Ovest	25.307.770	20.105.115	13.724.809	7.739.698	3.230.128	1.398.450	815.942
Nord-Est	37.504.394	24.093.145	8.777.988	1.987.161	905.169	576.598	421.884
Centro	23.749.951	17.946.059	10.853.878	4.533.806	1.377.329	657.084	420.980
Sud	15.814.300	12.027.744	7.439.837	3.433.293	1.275.660	643.656	417.189
Isole	5.962.351	5.436.041	4.778.768	4.122.479	3.467.723	2.815.545	2.168.188
Italia	**107.753.169**	**44.642.641**	**6.088.326**	**2.412.094**	**1.484.877**	**1.070.742**	**836.843**

Prospetto 8 - 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 e tipologie comunali - Stime del numero di viaggi - Anno 2015

TIPOLOGIE COMUNALI	a	b	c	R^2
	VIAGGI – LAVORO			
Comuni centro dell'area metropolitana	10.837,535	0,411735991	-0,000000113	0,97
Comuni periferici dell'area metropolitana o con oltre 50.000 abitanti	15.240,624	0,370379119	-0,000000104	0,94
Comuni fino a 50.000 abitanti	21.350,089	0,273240545	-0,000000050	0,93
	VIAGGI – VACANZE			
Comuni centro dell'area metropolitana	67.876,117	0,215275924	-0,000000012	0,95
Comuni periferici dell'area metropolitana o con oltre 50.000 abitanti	91.026,033	0,129543896	-0,000000004	0,97
Comuni fino a 50.000 abitanti	108.934,714	0,116228530	-0,000000002	0,96
	VIAGGI - TOTALI			
Comuni centro dell'area metropolitana	71.709,840	0,191640123	-0,000000008	0,95
Comuni periferici dell'area metropolitana o con oltre 50.000 abitanti	89.511,395	0,122784166	-0,000000003	0,97
Comuni fino a 50.000 abitanti	109.937,456	0,107957438	-0,000000002	0,96

Prospetto 9 - Valori dei totali corrispondenti ad alcuni valori tipici degli errori relativi percentuali per gruppi di stime e tipologie comunali - Stime del numero di viaggi – Anno 2015

TIPOLOGIE COMUNALI	1%	5%	10%	15%	20%	25%	30%
	VIAGGI – LAVORO						
Comuni centro dell'area metropolitana	3.588.135	3.236.447	2.797.855	2.360.963	1.926.883	1.497.916	1.079.527
Comuni periferici dell'area metropolitana o con oltre 50.000 abitanti	3.517.469	3.136.761	2.662.874	2.192.491	1.728.273	1.276.177	851.413
Comuni fino a 50.000 abitanti	5.394.921	4.601.078	3.617.031	2.650.928	1.728.213	931.846	439.820
	VIAGGI – VACANZE						
Comuni centro dell'area metropolitana	18.167.839	14.765.763	10.577.835	6.571.724	3.181.993	1.350.481	728.983
Comuni periferici dell'area metropolitana o con oltre 50.000 abitanti	34.038.920	23.244.403	10.616.262	2.936.388	1.216.533	739.382	528.139
Comuni fino a 50.000 abitanti	48.743.779	31.326.083	11.536.641	2.733.366	1.258.321	803.593	588.578
	VIAGGI – TOTALI						
Comuni centro dell'area metropolitana	22.159.103	17.468.881	11.717.432	6.345.470	2.473.233	1.066.229	631.108
Comuni periferici dell'area metropolitana o con oltre 50.000 abitanti	36.077.551	23.950.901	9.947.739	2.534.722	1.108.403	691.606	500.581
Comuni fino a 50.000 abitanti	53.799.737	32.974.353	10.123.463	2.367.170	1.166.921	766.291	569.327

Prospetto 10 - Valori dei coefficienti a, b, c e di R^2 delle funzioni utilizzate per l'interpolazione degli errori campionari delle stime di totali per trimestre e anno - Stime del numero di escursioni - Anno 2015

TRIMESTRI	a	b	c	R^2
	ESCURSIONI			
I trimestre	82.846,713	0,1415980484	-0,0000000013	0,99
II trimestre	76.126,545	0,0899400212	-0,0000000014	0,97
III trimestre	95.576,766	0,1037578145	-0,0000000014	0,97
IV trimestre	73.825,395	0,1335070886	-0,0000000023	0,98
Anno	171.795,313	0,0639421086	-0,0000000021	0,99

Prospetto 11 - Valori dei totali corrispondenti ad alcuni valori tipici degli errori relativi percentuali per gruppi di stime, trimestri e anno - Stime del numero di escursioni – Anno 2015

TRIMESTRI	1%	5%	10%	15%	20%	25%	30%
	ESCURSIONI						
I trimestre	98.921.897	69.311.395	32.949.483	5.331.358	1.375.207	757.174	520.724
II trimestre	56.839.975	29.721.870	4.582.429	1.231.406	685.576	473.609	361.514
III trimestre	66.639.516	39.334.367	9.600.198	1.949.475	978.864	649.432	485.320
IV trimestre	53.195.536	36.430.656	16.211.074	3.104.333	1.069.859	625.839	440.676
Anno	253.107.318	75.193.392	4.635.811	1.986.382	1.260.145	922.357	727.284

5.6.3. Esempi di calcolo degli errori campionari

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 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 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 riferita al Centro per l'anno, $_C\hat{Y}_a$, si possono leggere i valori dei coefficienti *a*, *b* e *c* dal prospetto 4 e si può utilizzare la seguente espressione:

$$\hat{\sigma}(_C\hat{Y}_a) = 53.829{,}11 + 0{,}45211743_C\hat{Y}_a - 0{,}00000005(_C\hat{Y}_a)^2$$

Esempio 2

La valutazione degli errori di campionamento delle stime mediante un metodo approssimato si può ottenere utilizzando i dati dei prospetti 5 e 7. Il prospetto 5 fa riferimento alle stime del numero di notti; il prospetto 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 all'1, 5, 10, 15, 20, 25 e 30%.

Considerando ad esempio il prospetto 5, 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 106.950.826 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 5, si consideri la stima del numero di notti per lavoro, riferita al totale Italia nell'anno, pari a 23.250.221.

Dal prospetto 5, considerando le notti-lavoro ed esaminando la riga corrispondente al totale Italia per l'anno, si può osservare che il valore di tale stima risulta essere compreso tra i valori 27.243.623 e 18.536.714.

Da ciò si può facilmente desumere che il valore dell'errore relativo percentuale corrispondente è compreso tra il 10% e il 15%.

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

Esempio 1

L'ammontare dei turisti medi mensili nel trimestre gennaio-marzo è pari a 2.973.890.

Si cerca il livello di stima che più si avvicina a 2.973.890 nella prima colonna del prospetto 3, che è pari a 3.000.000. 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 5,1%.

L'errore assoluto sarà:

σ(2.973.890) = 5,1/100×2.973.890 =151.668

L'intervallo di confidenza (per P=0,95 e k=2), avrà come estremi:

2.973.890 – (2x 151.668) = 2.670.554

2.973.890 + (2x 151.668) = 3.277.226

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 2.000.000 e 3.000.000 ai quali corrispondono i valori 6,4 e 5,1 percentuali.

L'errore corrispondente a 2.973.890 è pari a:

σ(2.973.890) = 6,4 – ((6,4 – 5,1) / (3.000.000 – 2.000.000)) x (2.973.890– 2.000.000) = 5,13%

Il corrispondente errore assoluto è 5,13/100 x 2.973.890= 152.560 e l'intervallo di confidenza avrà come estremi:

2.973.890– (2x 152.560)= 2.668.770

2.973.890+ (2x 152.560)= 3.279.010

Esempio 3

Il calcolo dell'errore può essere effettuato direttamente tramite la funzione interpolante (11):

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

Per la stima di 2.973.890 si ha:

$\log\varepsilon^2$ = 11,14037–1,14704 x log (2.973.890) =

=11,14037 -(1,14704 *14,90538)=

=11,14037 - 17,09707= -5,9567

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

Infatti, da log (ε^2) = -5,9567

si ha:

ε^2 = *exp*(-5,9567)

e dunque:

ε = 0,0508

L'errore relativo percentuale sarà dunque pari a 5,1%.

6. La diffusione dei risultati dell'indagine

Le stime provvisorie riferite all'anno precedente sono diffuse nel mese di febbraio, nella Statistica Report "Viaggi e vacanze in Italia e all'estero".

Tutti i risultati definitivi dell'indagine sono diffusi attraverso il datawarehouse dell'Istituto I.Stat (http://dati.istat.it/) nella sezione Cultura, comunicazione, tempo libero, uso del tempo--> Viaggi, vacanze e turismo.

In adempimento alle richieste del Regolamento Europeo per le Statistiche del Turismo 692/2011, entro il 30 giugno di ogni anno sono trasmessi a Eurostat i dati sulla partecipazione al turismo nell'anno precedente dei residenti di 15 anni e più e i microdati sui viaggi effettuati dai residenti di 15 anni e più. Con cadenza triennale, inoltre, sono trasmessi i dati sulle escursioni.
Tutte queste informazioni sono successivamente archiviate nel data base di Eurostat, consultabile al link http://ec.europa.eu/eurostat/web/tourism/.

7. Glossario

Viaggio: spostamento realizzato, per turismo di vacanza o per ragioni di lavoro, fuori dal comune dove si vive e che comporta almeno un pernottamento nel luogo visitato; sono 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 a un anno: in questi casi, infatti, il viaggio non costituisce flusso turistico poiché la località visitata viene associata al luogo dove si vive.

Viaggio di vacanza: viaggio svolto per motivi prevalenti di relax, piacere, svago, o riposo, per visita 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.

Viaggio per motivi di lavoro o professionali: viaggio svolto per motivi prevalenti di lavoro quali missioni, partecipazione a congressi, riunioni d'affari o esercizio di attività di rappresentanza, docenza o altre attività professionali. E' escluso l'esercizio di un'attività alle dipendenze di un soggetto residente nel luogo di destinazione (lavori stagionali, supplenze o altri lavori a tempo determinato).

Viaggio abituale: spostamento realizzato per vacanza o lavoro, con almeno un pernottamento, effettuato tutte le settimane nella stessa località, fuori dal comune dove si vive.

Turista: persona che ha effettuato uno o più viaggi nel periodo di riferimento

Durata media del viaggio: rapporto tra il numero di notti trascorse in viaggio e il numero di viaggi.

Destinazione, mezzo di trasporto, tipo di alloggio, motivo della vacanza e del viaggio di lavoro: sono informazioni che vengono rilevate sulla base del concetto di "prevalenza". In particolare, la destinazione del viaggio e il tipo di alloggio vengono associate rispettivamente alla località e 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.

Tipo di luogo: l'informazione connota le caratteristiche del luogo visitato (città, mare, campagna, montagna, altro), oppure se si è trattato di una crociera.

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 o del trasporto tramite agenzia o tour operator; per prenotazione tramite internet si intende l'utilizzo di internet per prenotare direttamente o presso agenzia l'alloggio e/o il trasporto.

Escursione: visita senza pernottamento effettuata fuori dal Comune dove la famiglia vive abitualmente, diretta in località italiane o estere, con una durata di almeno tre ore nel luogo di destinazione, esclusi gli spostamenti di andata e ritorno. Sono escluse le eventuali escursioni che si effettuano durante i soggiorni di vacanza/di lavoro, poiché il luogo di partenza e di ritorno dell'escursione in questi casi non è il Comune dove vive la persona intervistata, bensì il luogo di destinazione del viaggio personale o di lavoro. Sono altresì esclusi gli spostamenti che hanno carattere di periodicità/regolarità nell'arco del mese di riferimento (per seguire un corso di studi, frequentare una palestra, fare la spesa). Le escursioni possono essere effettuate sia per motivi personali che per motivi di lavoro. Tra i motivi personali, vi sono: piacere, svago, vacanza, visita a parenti o amici, motivi religiosi o di pellegrinaggio, formazione/cultura, cure termali o trattamenti di salute, visite e cure mediche, shopping; accompagnare un familiare/parente/amico.

8. Riferimenti bibliografici

Deville J. C., Särndal C. E., (1992), "*Calibration Estimators in Survey Sampling*", Journal of the American Statistical Association, vol. 87, n.418, 1992.

Grassi D., Pannuzi N., 2015 (a cura di) "*La nuova indagine sulle spese per consumi in Italia*" ISBN 978-88-458-1856-1, Istat

Regolamento per le Statistiche sul Turismo 692/2011
http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2011:192:0017:0032:EN:PDF

9. Contatti

“Indagine su viaggi e vacanze e informazione integrata sulla domanda turistica”

Mascia Di Torrice

e-mail: maditorr@istat.it

Barbara Dattilo

e-mail: dattilo@istat.it

Curatori dei capitoli

Il documento è stato curato da Mascia Di Torrice.

Si devono a:

Andrea Cutillo i paragrafi 5.1, 5.2, 5.3

Claudia De Vitiis, Monica Russo e Marco D. Terribili i paragrafi 3, 5.5, 5.6