Istat
Istituto Nazionale di Statistica

Indagine Viaggi e Vacanze

Periodo di riferimento: anno 2020

Aspetti metodologici dell'indagine

Anno 2021

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 35
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 e di pernottamenti in viaggio, le spese (per il viaggio e giornaliere) e il numero 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[1].

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

[1] E dalle successive modifiche introdotte dal Regolamento Delegato (UE) 2019/1681 del 1 agosto 2019

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 8.100 famiglie, per un totale di 32.500 famiglie all'anno[2]. 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

[2] Nel triennio 2020-2022

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, pari a 19.500 famiglie annue;

(b) definizione del numero m_r di famiglie campione per ciascuna regione, definito mediante una procedura di allocazione ottima multivariata del campione (Bethel, 1989[3]; Falorsi et al., 1998[4]) che consente di tenere sotto controllo gli errori campionari attesi dei principali capitoli di spesa (di interesse dell'indagine sulla spese);

(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*[5] 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 $n \times \lambda_r$;

[3] Bethel J. (1989). Sample Allocation in Multivariate Surveys, Survey Methodology, 15, 47-57.

[4] Falorsi P.D., Ballin M., De Vitiis C., Scepi G. (1998) "Principi e metodi del software generalizzato per la definizione del disegno di campionamento nelle indagini sulle imprese condotte dall'ISTAT", Statistica Applicata Vol. 10, n.2.

[5] 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.

(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:

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

(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 famiglie e di comuni nel rispetto della condizione di auto-ponderazione a livello regionale e degli errori campionari attesi per le stime di spesa, il numero minimo di interviste per comune è stato posto pari a 8.

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	65	159	224
Trimestre	65	477	542
Anno	65	477	542

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)[6]. 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

[6] Per una parte del mese di marzo e per tutto il secondo trimestre del 2020, a causa dell'emergenza sanitaria dovuta al diffondersi del Covid-19, la tecnica CAPI non è stata più utilizzata e si è fatto ricorso alla tecnica CATI (Computer Assisted Telephone Interview). Nei trimestri successivi si è tornati all'utilizzo congiunto di entrambe le tecniche (CAPI e CATI) con prevalente ricorso alla CATI. L'impossibilità di recarsi presso le famiglie durante il lockdown ha comportato la sospensione dell'indagine nel mese di aprile. Non è stato possibile adottare tecniche di imputazione dei dati mancanti riferiti a viaggi ed escursioni di aprile, a causa della drastica rarefazione dei casi osservati anche negli altri mesi. Tuttavia, l'assenza di viaggi ed escursioni in aprile può ritenersi una stima verosimile del fenomeno, stante il periodo di lockdown.

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[7].
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;

[7] 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.

- 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.
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:

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

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):

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

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:

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

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[8]. 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:

[8] 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 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;
- 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 2020, il file contiene n. 58.622 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 2020, il file contiene n. 3.366 records ed è di lunghezza pari a 141.

Il file escursioni è composto da un record per ciascuna escursione realizzata da un individuo nel mese di riferimento. Per l'anno 2020, il file contiene n. 3.659 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).

- spese per viaggio e spese giornaliere (in euro): sono informazioni presenti nel file viaggi.

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:

- ammontare dei viaggi, delle notti e delle escursioni nel mese, nel trimestre e nell'anno; spese medie per viaggi 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.
- spese medie giornaliere nel trimestre e nell'anno; a questo fine si deve utilizzare il coefficiente denominato COEV moltiplicato per la DURATA del viaggio.

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

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

dove Z_{hij} è la variabile linearizzata espressa come $Z_{hij} = Y_{hij} - \mathbf{X}'_{\mathbf{hij}}\beta$, essendo $\mathbf{X}_{\mathbf{hij}} = \left({}_1X_{hij},\ldots,{}_kX_{hij},\ldots{}_KX_{hij}\right)'$ 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

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

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, $\hat{\mathrm{V}}\mathrm{ar}\left(\hat{Z}_h\right)$, della stima $\hat{Z}_h$ è differente a seconda che lo strato sia AR oppure NAR. Possiamo, quindi scomporre come segue

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

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

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

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

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

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

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, $\hat{\mathrm{V}}\mathrm{ar}\left(\hat{Y}_d\right)$, 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:

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

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

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

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:

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

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 2020

AREE TERRITORIALI	Trimestri 1,2,e 4			Trimestre 3		
	a	b	R^2	a	b	R^2
Nord-Ovest	10,58275	-1,16442	0,95671	9,14451	-1,05913	0,94114
Nord-Est	8,42031	-1,00383	0,93068	8,76008	-1,05201	0,91683
Centro	8,81893	-1,04194	0,94515	9,38004	-1,10077	0,95301
Sud	8,55501	-1,00598	0,91688	9,38526	-1,08919	0,94744
Isole	9,28948	-1,06562	0,90703	9,28705	-1,08982	0,93936
Italia	8,90194	-1,02932	0,95412	9,47391	-1,08246	0,96189

Prospetto 3 - Valori interpolati degli errori relativi percentuali delle stime di frequenze assolute riferite agli individui per aree territoriali- Trimestri – Anno 2020

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	62,2	46,7	47,2	49,5	53,2	52,4	51,1	43,6	46,7	49,6	47,1	53,6
30000	49,1	38,1	38,2	40,3	42,8	42,5	41,2	35,3	37,4	39,8	37,8	43,1
40000	41,6	33,0	32,9	34,9	36,7	36,7	35,4	30,3	31,9	34,0	32,3	36,9
50000	36,5	29,5	29,3	31,2	32,6	32,7	31,4	26,9	28,2	30,1	28,6	32,7
60000	32,8	26,9	26,7	28,5	29,6	29,8	28,5	24,5	25,5	27,3	25,9	29,6
70000	30,0	24,9	24,6	26,3	27,3	27,5	26,3	22,6	23,5	25,1	23,8	27,2
80000	27,8	23,3	22,9	24,6	25,4	25,7	24,5	21,0	21,8	23,3	22,1	25,3
90000	25,9	22,0	21,6	23,2	23,9	24,2	23,0	19,8	20,4	21,9	20,8	23,8
100000	24,4	20,8	20,4	22,0	22,5	22,9	21,8	18,7	19,3	20,7	19,6	22,4
200000	16,3	14,7	14,2	15,5	15,6	16,0	15,1	13,0	13,2	14,2	13,4	15,4
300000	12,9	12,0	11,5	12,7	12,6	13,0	12,2	10,5	10,5	11,4	10,8	12,4
400000	10,9	10,4	9,9	11,0	10,8	11,2	10,4	9,0	9,0	9,7	9,2	10,6
500000	9,6	9,3	8,8	9,8	9,6	10,0	9,3	8,0	7,9	8,6	8,2	9,4
750000	7,5	7,6	7,1	8,0	7,7	8,1	7,5	6,5	6,4	6,9	6,5	7,5
1000000	6,4	6,6	6,2	6,9	6,6	7,0	6,4	5,6	5,4	5,9	5,6	6,5
2000000	4,3	4,6	4,3	4,9	4,6	4,9	4,5	3,9	3,7	4,0	3,8	4,4
3000000	3,4	3,8	3,5	4,0	3,7	4,0	3,6	3,1	3,0	3,2	3,1	3,6
4000000	2,8	3,3	3,0	3,4	3,2	3,4	3,1	2,7	2,5	2,8	2,6	3,0
5000000	2,5	2,9	2,7	3,1	2,8	3,1	2,7	2,4	2,2	2,5	2,3	2,7
7500000	2,0	2,4	2,2	2,5	2,3	2,5	2,2	1,9	1,8	2,0	1,9	2,2
10000000	1,7	2,1	1,9	2,2	1,9	2,1	1,9	1,7	1,5	1,7	1,6	1,9

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 2020

NOTTI – LAVORO				
AREE TERRITORIALI	a	b	c	R^2
		Gennaio-marzo		
Italia	59.401,72	0,282276991	-0,0000000462	0,87
		Luglio-settembre		
Italia	41.928,17	0,526801509	-0,0000001248	0,87
		Anno		
Nord-Ovest	65.264,58	0,492112267	-0,0000000808	0,89
Nord-Est	11.337,52	0,551150293	-0,0000002141	0,95
Centro	33.795,79	0,422023517	-0,0000000973	0,92
Sud	11.831,40	0,615433403	-0,0000002697	0,88
Isole	5.187,95	0,820184346	-0,0000007920	0,96
Italia	94.183,11	0,245947307	-0,0000000155	0,89
NOTTI – VACANZE				
AREE TERRITORIALI	a	b	c	R^2
		Gennaio-marzo		
Italia	142.740,21	0,195458765	-0,0000000025	0,91
		Luglio-settembre		
Italia	474.840,79	0,0706661351	-0,0000000002	0,94
		Anno		
Nord-Ovest	346.021,11	0,165193795	-0,0000000006	0,95
Nord-Est	241.683,25	0,152324376	-0,0000000007	0,95
Centro	227.072,09	0,154129380	-0,0000000011	0,95
Sud	125.210,07	0,287204321	-0,0000000048	0,92
Isole	142.520,55	0,270623531	-0,0000000094	0,93
Italia	543.389,90	0,086375769	-0,0000000001	0,96
NOTTI – TOTALI				
AREE TERRITORIALI	a	b	c	R^2
		Gennaio-marzo		
Italia	142.833,50	0,185866534	-0,0000000022	0,90
		Luglio-settembre		
Italia	514.105,67	0,077281538	-0,0000000002	0,92
		Anno		
Nord-Ovest	376.777,85	0,162394928	-0,0000000005	0,95
Nord-Est	219.130,05	0,157597977	-0,0000000007	0,95
Centro	223.598,47	0,149815942	-0,0000000009	0,94
Sud	113.791,84	0,293430777	-0,0000000054	0,90
Isole	139.957,71	0,284629636	-0,0000000110	0,89
Italia	564.025,67	0,086329921	-0,0000000001	0,95

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 2020

NOTTI – LAVORO							
AREE TERRITORIALI	1%	5%	10%	15%	20%	25%	30%
			Gennaio-marzo				
Italia	**6.104.478**	**5.271.885**	**4.248.325**	**3.258.002**	**2.332.320**	**1.535.882**	**958.238**
			Luglio-settembre				
Italia	**4.222.109**	**3.907.883**	**3.516.667**	**3.127.766**	**2.742.096**	**2.361.090**	**1.987.097**
			Anno				
Nord-Ovest	6.101.739	5.617.882	5.016.062	4.418.760	3.827.909	3.246.627	2.680.155
Nord-Est	2.547.771	2.362.613	2.131.558	1.901.086	1.671.429	1.442.961	1.216.315
Centro	4.313.207	3.910.575	3.409.959	2.913.653	2.424.069	1.945.638	1.487.022
Sud	2.264.108	2.117.147	1.933.730	1.750.719	1.568.252	1.386.535	1.205.892
Isole	1.029.300	979.122	916.450	853.844	791.321	728.900	666.610
Italia	**15.659.614**	**13.145.290**	**10.052.014**	**7.071.615**	**4.368.883**	**2.341.238**	**1.276.582**

NOTTI – VACANZE							
AREE TERRITORIALI	1%	5%	10%	15%	20%	25%	30%
			Gennaio-marzo				
Italia	**75.680.428**	**59.725.436**	**40.003.256**	**21.096.861**	**6.731.488**	**2.363.556**	**1.323.894**
			Luglio-settembre				
Italia	**348.579.632**	**135.788.644**	**14.849.763**	**5.907.082**	**3.653.073**	**2.640.883**	**2.067.206**
			Anno				
Nord-Ovest	278.573.323	208.092.476	121.178.742	41.798.107	8.715.758	3.975.486	2.539.930
Nord-Est	198.350.169	143.717.467	76.659.248	19.950.998	4.729.755	2.430.575	1.623.663
Centro	137.832.831	100.589.332	55.077.732	16.734.225	4.486.225	2.309.676	1.539.484
Sud	57.810.624	49.607.458	39.396.236	29.276.956	19.382.157	10.231.208	3.935.619
Isole	28.226.595	24.069.418	18.928.014	13.904.734	9.157.060	5.138.167	2.631.997
Italia	**657.438.071**	**324.048.140**	**31.390.892**	**8.410.070**	**4.758.936**	**3.313.085**	**2.540.125**

NOTTI – TOTALI							
AREE TERRITORIALI	1%	5%	10%	15%	20%	25%	30%
			Gennaio-marzo				
Italia	**80.609.366**	**62.687.904**	**40.562.707**	**19.585.067**	**5.459.191**	**2.078.661**	**1.222.600**
			Luglio-settembre				
Italia	**329.402.976**	**147.245.672**	**19.228.252**	**6.931.719**	**4.159.840**	**2.965.909**	**2.303.342**
			Anno				
Nord-Ovest	320.018.278	237.517.337	135.801.827	43.767.705	8.988.352	4.204.052	2.712.452
Nord-Est	200.521.601	147.113.374	81.309.975	23.060.979	4.770.014	2.327.994	1.526.676
Centro	158.741.298	114.393.697	60.171.780	15.750.364	4.150.219	2.189.313	1.475.926
Sud	53.368.623	45.957.318	36.729.124	27.576.829	18.604.294	10.201.402	4.038.376
Isole	25.358.118	21.818.503	17.438.509	13.149.625	9.058.851	5.456.347	2.930.988
Italia	**699.540.218**	**344.327.589**	**32.657.210**	**8.726.678**	**4.938.297**	**3.438.150**	**2.636.117**

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 2020

VIAGGI – LAVORO				
AREE TERRITORIALI	a	b	c	R^2
		Gennaio-marzo		
Italia	10.280,52	0,309189792	-0,0000001701	0,97
		Luglio-settembre		
Italia	8.908,25	0,387872457	-0,0000002907	0,95
		Anno		
Nord-Ovest	5.416,956	0,441568836	-0,0000003401	0,96
Nord-Est	4.183,974	0,475684290	-0,0000004704	0,95
Centro	4.970,298	0,385150474	-0,0000003128	0,96
Sud	2.654,691	0,518550294	-0,0000007833	0,95
Isole	258,533	0,766765162	-0,0000028061	0,99
Italia	17.782,131	0,200970900	-0,0000000434	0,97
VIAGGI – VACANZA				
AREE TERRITORIALI	a	b	c	R^2
		Gennaio-marzo		
Italia	45.237,78	0,132723195	-0,0000000104	0,92
		Luglio-settembre		
Italia	81.507,34	0,075716834	-0,0000000021	0,96
		Anno		
Nord-Ovest	49.833,406	0,149578802	-0,0000000047	0,97
Nord-Est	46.408,282	0,135704932	-0,0000000048	0,97
Centro	35.667,007	0,134652108	-0,0000000070	0,97
Sud	18.362,443	0,269234111	-0,0000000466	0,96
Isole	11.957,356	0,387961336	-0,0000000949	0,98
Italia	103.836,258	0,074954031	-0,0000000008	0,98
VIAGGI – TOTALI				
AREE TERRITORIALI	a	b	c	R^2
		Gennaio-marzo		
Italia	44.484,59	0,124773807	-0,0000000082	0,93
		Luglio-settembre		
Italia	79.311,19	0,079068008	-0,0000000019	0,97
		Anno		
Nord-Ovest	52.004,363	0,137662413	-0,0000000040	0,97
Nord-Est	50.412,660	0,126004203	-0,0000000040	0,98
Centro	40.592,170	0,117902547	-0,0000000056	0,97
Sud	22.138,331	0,225215969	-0,0000000341	0,96
Isole	13.652,150	0,342664645	-0,0000000813	0,96
Italia	**110.270,782**	**0,067655578**	**-0,0000000007**	**0,98**

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 2020

VIAGGI – LAVORO

AREE TERRITORIALI	1%	5%	10%	15%	20%	25%	30%
			Gennaio-marzo				
Italia	**1.792.283**	**1.562.139**	**1.276.892**	**996.330**	**725.125**	**475.093**	**274.305**
			Luglio-settembre				
Italia	**1.323.179**	**1.188.196**	**1.020.419**	**854.244**	**690.719**	**531.945**	**382.448**
			Anno				
Nord-Ovest	1.281.552	1.165.165	1.020.074	875.614	1.281.552	1.165.165	1.020.074
Nord-Est	998.960	914.739	809.699	705.029	998.960	914.739	809.699
Centro	1.212.487	1.086.129	928.754	772.365	1.212.487	1.086.129	928.754
Sud	654.429	603.796	540.619	477.612	654.429	603.796	540.619
Isole	270.032	255.796	238.004	220.217	270.032	255.796	238.004
Italia	**4.492.763**	**3.593.658**	**2.491.685**	**1.456.232**	**4.492.763**	**3.593.658**	**2.491.685**

VIAGGI – VACANZE

AREE TERRITORIALI	1%	5%	10%	15%	20%	25%	30%
			Gennaio-marzo				
Italia	**12.200.268**	**8.496.356**	**4.197.647**	**1.415.903**	**614.288**	**373.414**	**266.052**
			Luglio-settembre				
Italia	**31.991.977**	**14.658.061**	**2.710.812**	**1.064.690**	**648.597**	**465.024**	**362.165**
			Anno				
Nord-Ovest	29.797.456	21.495.567	11.382.407	3.198.182	29.797.456	21.495.567	11.382.407
Nord-Est	26.613.598	18.422.769	8.584.693	1.959.791	26.613.598	18.422.769	8.584.693
Centro	18.013.895	12.449.628	5.803.634	1.411.443	18.013.895	12.449.628	5.803.634
Sud	5.635.859	4.789.399	3.738.996	2.705.734	5.635.859	4.789.399	3.738.996
Isole	4.015.399	3.597.415	3.076.310	2.557.581	4.015.399	3.597.415	3.076.310
Italia	**80.743.504**	**34.126.672**	**3.697.918**	**1.363.317**	**80.743.504**	**34.126.672**	**3.697.918**

VIAGGI – TOTALI

AREE TERRITORIALI	1%	5%	10%	15%	20%	25%	30%
			Gennaio-marzo				
Italia	**14.324.925**	**9.647.033**	**4.275.072**	**1.252.054**	**557.362**	**347.308**	**250.913**
			Luglio-settembre				
Italia	**37.937.446**	**17.864.257**	**2.988.623**	**1.086.892**	**649.294**	**461.654**	**357.897**
			Anno				
Nord-Ovest	32.315.936	22.492.251	10.637.182	2.379.424	32.315.936	22.492.251	10.637.182
Nord-Est	29.303.093	19.559.910	8.034.559	1.646.832	29.303.093	19.559.910	8.034.559
Centro	19.752.507	12.768.910	4.750.832	1.067.147	19.752.507	12.768.910	4.750.832
Sud	6.410.027	5.259.621	3.839.567	2.467.829	6.410.027	5.259.621	3.839.567
Isole	4.132.114	3.645.572	3.039.796	2.438.462	4.132.114	3.645.572	3.039.796
Italia	**88.236.790**	**31.663.730**	**3.198.159**	**1.324.910**	**88.236.790**	**31.663.730**	**3.198.159**

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 2020

TIPOLOGIE COMUNALI	a	b	c	R^2
	VIAGGI – LAVORO			
Comuni centro dell'area metropolitana	2.827,38	0,516815632	-0,000000612	0,97
Comuni periferici dell'area metropolitana o con oltre 50.000 abitanti	5.886,29	0,352475644	-0,000000229	0,96
Comuni fino a 50.000 abitanti	10.565,93	0,313311068	-0,000000148	0,96
	VIAGGI – VACANZE			
Comuni centro dell'area metropolitana	43.891,35	0,120421542	-0,000000008	0,95
Comuni periferici dell'area metropolitana o con oltre 50.000 abitanti	58.500,89	0,117222121	-0,000000003	0,96
Comuni fino a 50.000 abitanti	64.574,73	0,133939804	-0,000000003	0,97
	VIAGGI - TOTALI			
Comuni centro dell'area metropolitana	46.120,98	0,115005824	-0,000000007	0,95
Comuni periferici dell'area metropolitana o con oltre 50.000 abitanti	59.065,72	0,112310647	-0,000000003	0,97
Comuni fino a 50.000 abitanti	65.903,55	0,128424422	-0,000000003	0,98

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 2020

TIPOLOGIE COMUNALI	1%	5%	10%	15%	20%	25%	30%
	VIAGGI – LAVORO						
Comuni centro dell'area metropolitana	834.283	769.343	688.291	607.426	526.830	446.646	367.129
Comuni periferici dell'area metropolitana o con oltre 50.000 abitanti	1.512.061	1.339.630	1.125.015	912.075	702.221	498.864	311.558
Comuni fino a 50.000 abitanti	2.080.771	1.815.880	1.487.195	1.163.186	848.550	555.508	315.660
	VIAGGI – VACANZE						
Comuni centro dell'area metropolitana	13.526.486	8.963.022	3.803.458	1.124.592	522.698	331.593	241.680
Comuni periferici dell'area metropolitana o con oltre 50.000 abitanti	34.574.633	22.176.965	7.836.844	1.552.985	688.673	436.080	318.319
Comuni fino a 50.000 abitanti	36.158.903	24.887.454	11.391.468	2.579.850	931.813	547.413	385.748
	VIAGGI – TOTALI						
Comuni centro dell'area metropolitana	14.500.229	9.372.284	3.687.564	1.072.650	518.992	335.433	246.853
Comuni periferici dell'area metropolitana o con oltre 50.000 abitanti	36.356.325	22.706.059	7.182.728	1.415.247	659.404	425.224	313.205
Comuni fino a 50.000 abitanti	39.525.154	26.626.032	11.278.366	2.305.813	887.332	534.928	381.530

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 2020

TRIMESTRI	a	b	c	R^2
	ESCURSIONI			
I trimestre	44.621,20	0,1268669357	-0,0000000030	0,98
II trimestre	39.233,52	0,1707587558	-0,0000000102	0,90
III trimestre	59.027,68	0,0988212789	-0,0000000009	0,98
IV trimestre	42.798,80	0,1548280756	-0,0000000043	0,97
Anno	105.368,95	0,0607008755	-0,0000000003	0,98

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 2020

TRIMESTRI	1%	5%	10%	15%	20%	25%	30%
	ESCURSIONI						
I trimestre	39.972.098	26.610.217	10.537.066	1.601.599	595.810	359.288	256.605
II trimestre	15.996.351	12.151.998	7.451.132	3.226.387	995.751	467.030	296.622
III trimestre	101.995.191	56.883.391	7.561.436	1.131.439	580.481	389.570	293.035
IV trimestre	34.281.748	25.003.136	13.605.489	3.786.028	875.209	440.993	292.307
Anno	158.893.334	41.044.441	2.624.543	1.174.958	755.099	556.099	440.062

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:

$$\hat{\sigma}({}_c\hat{Y}_a) = 33.795{,}79 + 0{,}422023517\,({}_c\hat{Y}_a) - -0{,}0000000973\,({}_c\hat{Y}_a)^2$$

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 14.849.763 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 8.870.815.

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 10.052.014 e 7.071.615.

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 2.421.980.

Si cerca il livello di stima che più si avvicina a 2.421.980 nella prima colonna del prospetto 3, che è pari a 2.000.000. 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 4,9%.

L'errore assoluto sarà:

σ(2.421.980) = 4,9/100×2.421.980=118.677

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

2.421.980 – (2 x 118.677) = 2.184.626

2.421.980 + (2 x 118.677) = 2.659.334

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 4,9 e 4,0 percentuali.

L'errore corrispondente a 2.421.980 è pari a:

σ(2.421.980) = 4,9 – ((4,9 – 4,0) / (3.000.000 – 2.000.000)) x (2.421.980 – 2.000.000) = 4,53%.

Il corrispondente errore assoluto è 4,53/100 x 2.421.980 = 109.715 e l'intervallo di confidenza avrà come estremi:

2.421.980 – (2 x 109.715) = 2.202.550

2.421.980 + (2 x 109.715) = 2.641.410

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 2.421.980 si ha:

$\log\varepsilon^2$ = 8,90194 - (1,02932 * log (2.421.980) =

= 8,90194 - (1,02932 * 14,700095) =

= 8,90194 - 15,1311 = -6,29916

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

Infatti, da log (ε^2) = -6,29916 si ha:

ε^2 = *exp*(-6,29916)

e dunque:

ε = 0,0443

L'errore relativo percentuale sarà dunque pari a 4,43%.

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, viaggi --> Viaggi.

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

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.

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

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 i tempi per 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.

Organizzazione: L'organizzazione del viaggio si riferisce alla presenza o meno di una prenotazione dell'alloggio e/o del trasporto ed, eventualmente, di altri servizi acquistati in agenzia o da un tour operator. Se c'è stata una prenotazione dell'alloggio e/o del trasporto, questa può essere avvenuta in modo diretto o tramite l'agenzia/tour operator. Nel caso dell'alloggio, per prenotazione diretta si intende la prenotazione effettuata direttamente presso la struttura ricettiva, recandosi fisicamente sul posto o accedendo mediante internet al sito web della struttura alloggiativa. Nel caso del trasporto, la prenotazione diretta consiste nell'acquisto del servizio direttamente presso uffici fisici o virtuali che operano su internet, come le biglietterie ferroviarie, aeree, ecc. Per

prenotazione tramite agenzia/tour operator si intende sia quella effettuata recandosi presso gli uffici dell'agenzia/tour operator dislocati sul territorio, sia quella effettuata mediante un'agenzia/tour operator virtuale su internet (sito web dell'agenzia/tour operator, i portali quali Booking, Expedia, Tripadvisor, Trivago, Kayak o le piattaforme di prenotazione online utilizzate prevalentemente per gli alloggi privati come Airbnb, HomeAway, Scambiocasa, HomeToGo). Nel caso della prenotazione del mezzo di trasporto mediante i servizi di car rental (Hertz, Avis, ecc.), questi non sono classificati come agenzie/tour operator, pertanto la prenotazione deve essere considerata una prenotazione diretta. Per prenotazione tramite internet si intende l'utilizzo di internet per prenotare direttamente o tramite agenzia/tour operator on line l'alloggio e/o il trasporto.

Spesa del viaggio: include tutte le spese finalizzate al viaggio, sostenute prima della partenza, durante il viaggio o al ritorno, in conseguenza del viaggio stesso. E' composta dalle spese per trasporto, per l'alloggio e dalle restanti spese (pasti, acquisti vari, attività ricreative, culturali e sportive, souvenir etc..).

Spesa media per viaggio: ammontare di spesa del viaggio riferito al singolo partecipante al viaggio

Spesa media giornaliera: rapporto tra spesa media per viaggio, riferita al singolo partecipante, e durata del viaggio (numero di notti)

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

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

Viaggio: spostamento realizzato, per vacanza o per ragioni di lavoro, fuori dal comune in cui 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 (ritenuti abituali secondo la definizione di turismo), 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 in cui si vive. Sono altresì esclusi i viaggi effettuati per coercizione (viaggi per traslocare, viaggi per andare a votare per elezioni/referendum, spostamenti dovuti a calamità naturali, ecc.).

Viaggio abituale: viaggio con almeno un pernottamento fuori dal comune in cui si vive effettuato per vacanza o per lavoro tutte le settimane nella stessa località,

Viaggio di vacanza: viaggio svolto per motivi prevalenti di 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).

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.

Di Torrice M., 2018 (a cura di) "*La nuova indagine sulla domanda turistica*"
ISBN 978-88-458-1964-3, Istat

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

Regolamento Delegato (UE) 2019/1681 della Commissione
https://eur-lex.europa.eu/legal-content/IT/TXT/PDF/?uri=CELEX:32019R1681&from=IT

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