

# LA NUOVA INDAGINE SULLA DOMANDA TURISTICA





## LA NUOVA INDAGINE SULLA DOMANDA TURISTICA

ISBN 978-88-458-1964-3

© 2018

Istituto nazionale di statistica  
Via Cesare Balbo, 16 - Roma



Salvo diversa indicazione, tutti i contenuti pubblicati sono soggetti alla licenza Creative Commons - Attribuzione - versione 3.0. <https://creativecommons.org/licenses/by/3.0/it/>

È dunque possibile riprodurre, distribuire, trasmettere e adattare liberamente dati e analisi dell'Istituto nazionale di statistica, anche a scopi commerciali, a condizione che venga citata la fonte.

Immagini, loghi (compreso il logo dell'Istat), marchi registrati e altri contenuti di proprietà di terzi appartengono ai rispettivi proprietari e non possono essere riprodotti senza il loro consenso.



# INDICE

	Pag.
<b>Premessa</b>	5
<b>1. L'adozione del Regolamento</b>	7
1.1 Dalla Direttiva 95/57/CE al nuovo Regolamento per le Statistiche sul turismo 692/2011	7
1.2 L'ampliamento del contenuto informativo	8
1.3 Le sperimentazioni delle nuove variabili nel periodo 2011-2012	11
1.4 Le nuove modalità di trasmissione dei dati a Eurostat	15
1.4.1 <i>Struttura e codificazione dei file</i>	15
1.4.2 <i>Il processo di validazione dei dati</i>	17
1.5 Il report sulla qualità e sui metadati	18
<b>2. L'integrazione dell'indagine viaggi e vacanze nella nuova indagine sulle spese</b>	21
2.1 I motivi dell'integrazione	21
2.2 Il percorso verso l'integrazione	23
2.3 L'indagine pilota	26
<b>3. La nuova rilevazione delle informazioni sulla domanda turistica</b>	29
3.1 Il disegno di campionamento e la selezione del campione	30
3.1.1 <i>La popolazione di riferimento, i domini di stima e lo schema generale del disegno di campionamento</i>	30
3.1.2 <i>Stratificazione e selezione del campione</i>	32
3.1.3 <i>L'utilizzo delle liste anagrafiche comunali per la selezione del campione</i>	34
3.2 La struttura del questionario	36
3.2.1 <i>I tempi della rilevazione</i>	36
3.2.2 <i>Dove si collocano i quesiti sui viaggi</i>	37
3.2.3 <i>Il promemoria e l'intervista proxy</i>	39
3.2.4 <i>Il flusso dei quesiti sui viaggi nell'intervista iniziale</i>	41
3.2.5 <i>Il flusso dei quesiti sui viaggi nell'intervista finale</i>	42
3.3 La rete di rilevazione e il monitoraggio dell'indagine	48
3.3.1 <i>La formazione dei rilevatori</i>	48
3.3.2 <i>Gli indicatori di performance della rete di rilevazione</i>	51
3.3.3 <i>Il sistema degli esiti familiari</i>	53
3.4 Il trattamento dei dati	56
3.4.1 <i>L'acquisizione controllata dei dati e il piano delle incompatibilità</i>	57

3.4.2	<i>I controlli mensili</i>	59
3.4.3	<i>I controlli e il trattamento trimestrali</i>	60
3.4.4	<i>La correzione delle variabili strutturali dell'individuo</i>	63
3.4.5	<i>Il trattamento annuale</i>	66
3.4.6	<i>Le variabili quantitative: da RIDA a IVEware per la correzione delle spese del viaggio e delle escursioni</i>	67
3.5	<i>La costruzione delle stime</i>	70
3.5.1	<i>Le stime della domanda turistica</i>	70
3.5.2	<i>Il sistema di ponderazione e il trattamento degli outliers</i>	71
3.5.3	<i>La probabilità di inclusione e il peso diretto</i>	72
3.5.4	<i>La correzione per mancata risposta</i>	73
3.5.5	<i>La calibrazione a fonti esterne</i>	73
3.5.6	<i>Gli errori di campionamento</i>	75
3.6	<i>La diffusione delle stime della domanda turistica</i>	78
3.7	<i>CATI e CAPI a confronto: un quadro di sintesi</i>	80
<b>4.</b>	<b>La ricostruzione delle serie storiche</b>	83
4.1	<i>Il periodo di sovrapposizione e i risultati della nuova indagine</i>	83
4.2	<i>Le serie da ricostruire</i>	85
4.3	<i>La metodologia</i>	85
4.4	<i>I risultati della ricostruzione</i>	90
4.5	<i>La diffusione delle serie ricostruite</i>	93
<b>5.</b>	<b>Prospettive future</b>	95
	<b>Glossario</b>	99
	<b>Riferimenti bibliografici</b>	103

## PREMESSA<sup>1</sup>

L'Istat rileva informazioni sulla domanda turistica sin dal 1959. Nell'arco di oltre cinquant'anni, profonde trasformazioni hanno modificato gli stili e i comportamenti di viaggio, ma anche le modalità e le tecniche che l'Istat ha utilizzato per testimoniarli.

La rilevanza economica e sociale del fenomeno turistico per il nostro Paese ha stimolato una grande attenzione da parte della statistica ufficiale, con un forte dinamismo innovativo dei sistemi di rilevazione e l'utilizzo di una molteplicità di strumenti e opportunità di raccolta e analisi dei dati.

L'evoluzione delle statistiche ufficiali sulla domanda turistica si è snodata in quattro grandi fasi. Nel periodo iniziale (1959-1982), le informazioni sulla domanda turistica sono rilevate prevalentemente mediante indagini speciali abbinate alle rilevazioni sulle Forze di Lavoro; in una seconda fase, dal 1993 al 1996, il forte interesse verso gli aspetti più strettamente "sociali" del fenomeno turistico conduce alla rilevazione di informazioni sul turismo nell'ambito del nascente sistema di Indagini Multiscopo sulle famiglie, in particolare nell'indagine "Aspetti della vita quotidiana".

Successivamente, con l'entrata in vigore della Direttiva Europea 95/57/Ce<sup>2</sup>, in Italia viene attuata una riorganizzazione del sistema statistico di raccolta dei dati sul turismo, con lo scopo di rendere sistematica l'acquisizione delle informazioni e al contempo di fornire stime tempestive sull'andamento del fenomeno. Nel 1997 viene quindi avviata, sempre nel sistema di indagini Multiscopo sulle famiglie, l'indagine trimestrale tematica "Viaggi, vacanze e vita quotidiana" (Istat, 2003), per la quale si sceglie la tecnica CATI (Computer Assisted Telephone Interview), sulla base di esigenze di economicità, tempestività e monitoraggio della qualità dei dati. Tale indagine, dal 1997 al 2013, ha fornito, in ottemperanza alla Direttiva e per la diffusione nazionale, stime su viaggi e pernottamenti realizzati in Italia e all'estero dai residenti, nonché sul numero di turisti. Inoltre, ha consentito di comprendere stili e comportamenti di viaggio descrivendo qualitativamente le caratteristiche della domanda turistica e la loro evoluzione nel tempo.

La quarta fase di trasformazione della rilevazione sulla domanda turistica vede, contemporaneamente, a livello comunitario l'abolizione della Direttiva 95/57, sostituita dal 1° gennaio 2012 dal nuovo Regolamento delle Statistiche del Turismo 692/2011<sup>3</sup> e, a livello nazionale, la cessazione dell'indagine CATI "Viaggi, vacanze e vita quotidiana" e la sua integrazione nella nuova indagine CAPI sulle Spese delle famiglie (dal 1° gennaio 2014).

Il presente volume tratterà nel dettaglio questi ultimi due rilevanti aspetti di cambiamento, descrivendo come essi siano stati affrontati introducendo profonde innovazioni nell'indagine, senza venir meno agli obiettivi di soddisfare le crescenti esigenze informative nazionali ed europee, nel rispetto dei consolidati livelli di qualità ormai raggiunti.

<sup>1</sup> Il volume è stato curato da Mascia Di Torrice, che ha redatto anche la premessa.

<sup>2</sup> Direttiva 95/57/EC del 23 Novembre 1995 relativa alla raccolta di dati statistici nel settore del turismo of 23 <http://eur-lex.europa.eu/legal-content/IT/TXT/PDF/?uri=CELEX:31995L0057&from=EN>.

<sup>3</sup> Regolamento (UE) N. 692/2011 del Parlamento Europeo e del Consiglio del 6 luglio 2011 relativo alle statistiche europee sul turismo e che abroga la direttiva 95/57/CE del Consiglio <http://eur-lex.europa.eu/legal-content/IT/TXT/PDF/?uri=CELEX:32011R0692&from=EN>.



## 1. L'ADOZIONE DEL REGOLAMENTO<sup>1</sup>

### 1.1 Dalla Direttiva 95/57/CE al nuovo Regolamento per le Statistiche sul turismo 692/2011

Il settore del turismo rappresenta una parte importante dell'economia in molti paesi europei, con un enorme potenziale in termini di crescita e di occupazione<sup>2</sup>, ma anche di integrazione sociale e culturale. Pertanto, le statistiche del turismo sono state oggetto, in ambito comunitario, di una profonda riflessione sui fabbisogni informativi che ha portato all'abrogazione della Direttiva 95/57/Ce e all'entrata in vigore del Regolamento 692/2011 per le Statistiche sul turismo (dal 1° gennaio 2012). Il superamento della Direttiva costituisce un importante passo in avanti verso un sistema ulteriormente armonizzato di statistiche del turismo tra i Paesi europei e verso l'approfondimento della conoscenza del settore.

Il Regolamento, richiamandosi al quadro concettuale e metodologico vigente per le statistiche sul turismo (UNWTO, 2010), ha l'obiettivo di fornire informazioni complete ed esaustive sul turismo in Europa e sul turismo degli europei. Infatti, in riferimento ai domini del fenomeno, schematizzati nella Fig. 1.1, il Regolamento copre due grandi aree:

- il turismo interno ad ogni Paese, sia effettuato dai residenti (*domestic*), sia da stranieri (*inbound*);
- il turismo in uscita effettuato dai residenti (*outbound*).

Nel sistema italiano delle statistiche ufficiali, il turismo domestic e quello outbound dei residenti vengono indagati rilevando informazioni sulla domanda turistica presso le famiglie (*demand side statistics*).

Il turismo inbound degli stranieri viene rilevato tramite l'indagine censuaria sul movimento dei clienti presso le strutture ricettive collettive (*supply side statistics*)<sup>3</sup>, che fornisce anche dati sul turismo dei residenti presso le medesime strutture<sup>4</sup>.

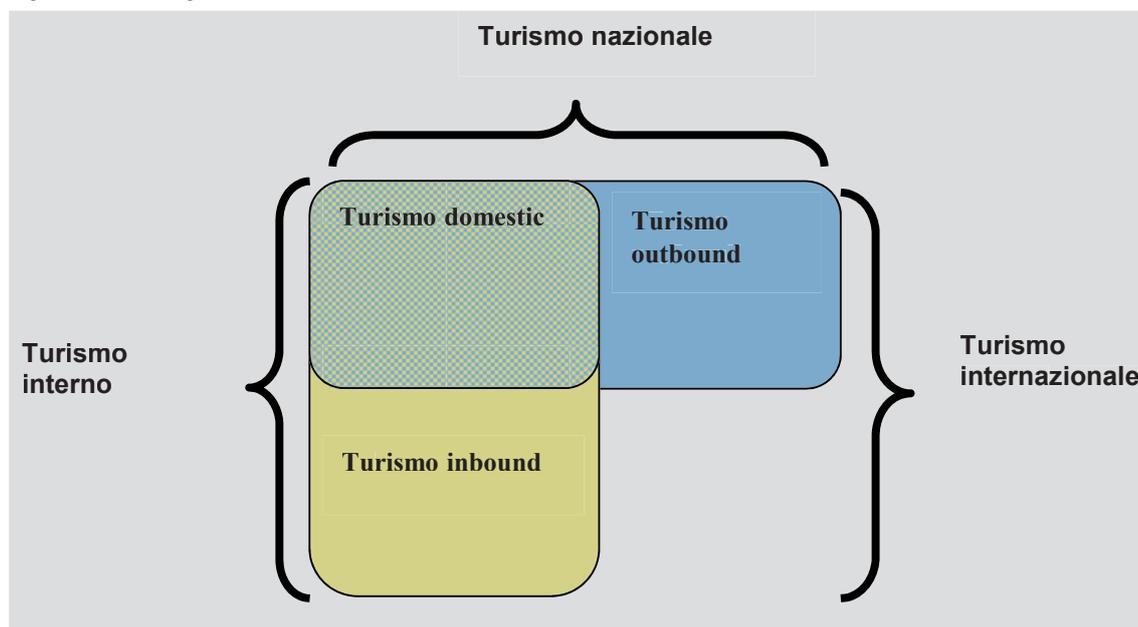
<sup>1</sup> I paragrafi 1.1, 1.2 e 1.3 sono stati curati e redatti da M. Di Torrice. Il paragrafo 1.4 è stato curato e redatto da M. Sabato. Il paragrafo 1.5 è stato curato e redatto da B. Dattilo.

<sup>2</sup> Nel 2017, il numero di notti trascorse in strutture ricettive turistiche nell'Unione europea (UE) ha raggiunto i 3,2 miliardi, in aumento del 5,1% rispetto al 2016. Per maggiori informazioni si veda <http://ec.europa.eu/eurostat/web/tourism/overview>.

<sup>3</sup> Indagine "Movimento dei clienti negli esercizi ricettivi" (<http://www.istat.it/it/archivio/15073>) e Indagine "Capacità degli esercizi ricettivi" (<http://www.istat.it/it/archivio/14517>).

<sup>4</sup> Il turismo inbound degli stranieri e outbound dei residenti vengono anche rilevati dalla Banca d'Italia, tramite l'indagine campionaria sul turismo internazionale, con l'obiettivo primario della compilazione della voce "Viaggi" della bilancia dei pagamenti dell'Italia. Per maggiori informazioni si veda <http://www.bancaditalia.it/statistiche/tematiche/rapporti-estero/turismo-internazionale/index.html>.

Figura 1.1 - Tipologie di flussi turistici



## 1.2 L'ampliamento del contenuto informativo

Il Regolamento 692/2011 si prefigge l'esplicito obiettivo di istituire un quadro comune ai fini dello sviluppo, della produzione e della divulgazione sistematici di statistiche europee sul turismo. Gli Stati membri, a tale scopo, si impegnano a raccogliere, compilare, elaborare e trasmettere statistiche armonizzate sull'offerta e sulla domanda turistica<sup>5</sup>.

E' evidente la volontà di continuare a tracciare un percorso comune in ambito europeo, nel solco di quanto già realizzato con la Direttiva 95/57, riconoscendo l'importanza di un settore considerato strategico. Allo stesso tempo, nelle premesse, viene evidenziata l'esigenza di un aggiornamento delle statistiche sul turismo, per poter rappresentare adeguatamente i mutamenti delle abitudini dei turisti.

Per quanto riguarda le statistiche *demand-side*, molte sono le novità introdotte dal Regolamento rispetto a quanto previsto dalla precedente Direttiva, sia relative alle informazioni richieste, sia alla loro modalità di trasmissione. Inoltre, viene prevista, per la prima volta, la compilazione di un *Quality Report* annuale (cfr. par. 1.5).

La prima novità si riferisce ad un'analisi maggiormente approfondita della partecipazione annuale al turismo, cioè la quantificazione del flusso dei turisti di 15 anni o più per vacanza durante l'anno, distinta per le diverse tipologie di viaggio<sup>6</sup>. Questo aspetto, precedentemente indagato solo in riferimento alle vacanze lunghe, viene esteso anche alla rilevazione del numero di persone che hanno effettuato vacanze brevi durante l'anno. A complemento di tali informazioni, è richiesta anche la trasmissione dell'ammontare di persone che non abbiano effettuato vacanze nell'anno di riferimento, insieme ai motivi della mancata partenza (da trasmettere ogni tre anni).

Tutte le informazioni sulle persone sono articolate in disaggregazioni socio-demografiche, alcune di carattere obbligatorio (sesso e classe di età), altre facoltative (livello di istruzione, condizione lavorativa, reddito familiare).

<sup>5</sup> Articolo 1 del Regolamento 692/2011.

<sup>6</sup> Come noto, nella metodologia i viaggi per vacanza vengono distinti, secondo la durata, in viaggi "brevi" (da una a tre notti) e viaggi "lunghi" (più di tre notti).

## 1. L'adozione del regolamento

Un'altra importante innovazione investe il set di informazioni riguardanti il singolo viaggio, notevolmente ampliato e richiesto, per la prima volta, sotto forma di microdati (cfr. par. 1.4). Il viaggio è così descritto in tutte le sue caratteristiche qualitative, alcune delle quali già trasmesse (ma solo a livello aggregato), altre completamente nuove. Nello schema riportato nel Prospetto 1.1, vengono riportate tutte le informazioni richieste per ogni viaggio, unitamente alla periodicità di trasmissione e al carattere di novità rispetto alla Direttiva.

**Prospetto 1.1 - Quadro sinottico delle informazioni da trasmettere sui viaggi in base al Regolamento Europeo 692/2011**

VARIABILE	Periodicità di trasmissione		Novità rispetto alla Direttiva 95/57/Ce
	Annuale	Triennale	
Numero progressivo del viaggio	X		X
Mese di inizio viaggio	X		
Durata del viaggio in notti	X		
Numero di notti spese sul territorio nazionale ( <i>solo per i viaggi all'estero</i> )		X	X
Principale Paese di destinazione	X		
Motivo principale del viaggio	X		
Tipo di destinazione		X	X
Partecipazione di bambini		X	X
Mezzo principale di trasporto	X		
Tipo di alloggio prevalente	X		
Prenotazione del mezzo di trasporto tramite agenzia/tour operator		X	
Prenotazione dell'alloggio tramite agenzia/tour operator		X	
Prenotazione diretta del viaggio		X	
Prenotazione del viaggio: pacchetto viaggio		X	
Prenotazione del mezzo di trasporto tramite internet		X	X
Prenotazione dell'alloggio tramite internet		X	X
Spesa pro capite per il trasporto	X		X
Spesa pro capite per l'alloggio	X		X
Spesa pro capite per pasti ( <i>opzionale</i> )	X		X
Spesa pro capite per altro (inclusi beni durevoli e di valore)	X		X
Spesa pro capite per beni durevoli e di valore	X		X
Sesso	X		
Età	X		
Paese di residenza	X		
Livello di istruzione ( <i>opzionale</i> )	X		X
Condizione lavorativa ( <i>opzionale</i> )	X		X
Reddito familiare in quartili ( <i>opzionale</i> )	X		X
Coefficiente di riporto	X		X

Viene data maggiore importanza, rispetto al passato, alle informazioni sulle spese del viaggio<sup>7</sup>, che devono essere disaggregate in spese per trasporto, per alloggio, per pasti

<sup>7</sup> La Direttiva richiedeva la trasmissione della spesa complessiva, distinta per alcune caratteristiche del viaggio e del turista.

(opzionale) e per tutto il resto degli acquisti finalizzati al viaggio. Inoltre, la rilevazione separata delle eventuali spese sostenute per l'acquisto di beni durevoli o di valore<sup>8</sup> consente di individuare e isolare somme, che, per loro natura, potrebbero essere particolarmente elevate e quindi distorcere il calcolo dell'ammontare di spesa riferita al viaggio.

Sempre nell'ambito delle informazioni aventi carattere di novità, sono particolarmente rilevanti quelle relative alla prenotazione del mezzo di trasporto e dell'alloggio tramite internet, inserite proprio in considerazione dell'importanza crescente assunta nell'ultimo periodo da questo tipo di organizzazione del viaggio.

L'obiettivo dell'armonizzazione ha invece ispirato la rivisitazione di alcune scelte riguardanti variabili già trasmesse nel passato: è il caso del tipo di alloggio, che, superando la tradizionale contrapposizione tra "Alloggio collettivo" e "Alloggio privato", è ora distinto in "Alloggio in locazione" e "Alloggio non in locazione" (Eurostat, 2014), nell'ottica di una maggiore corrispondenza con la classificazione NACE<sup>9</sup> utilizzata per le statistiche *supply-side*.

Rispetto ai nuovi fabbisogni informativi del settore turistico, il Regolamento ha ampliato il campo d'osservazione, prevedendo la trasmissione di informazioni sulle escursioni, cioè le visite che si svolgono in giornata (permanendo almeno tre ore nel luogo di destinazione, ma senza pernottarvi). Si tratta del riconoscimento esplicito dell'importanza di un fenomeno che contribuisce notevolmente al volume d'affari prodotto dal turismo in molti Paesi e che deve essere tenuto in considerazione anche nell'analisi dell'impatto sul territorio, in termini di pressione ambientale e di mobilità<sup>10</sup>. La conoscenza dell'aspetto economico è importante anche nel caso delle escursioni, poiché, più in generale, le spese turistiche sono

**Prospetto 1.2 - Quadro sinottico delle informazioni da trasmettere sulle escursioni in base al Regolamento Europeo 692/2011**

Variabili	Disaggregazioni (opzionale)	Disaggregazioni socio-demografiche (opzionale)	Periodicità di trasmissione	
			Annuale	Triennale
Numero escursioni verso l'estero per motivi personali	Paese di destinazione	Sesso, classe di età, livello di istruzione, condizione lavorativa, reddito familiare	X	
Numero escursioni verso l'estero per motivi di lavoro			X	
Spese sostenute durante le escursioni verso l'estero per motivi personali	Paese di destinazione Categoria di spesa (trasporto, pasti, altri acquisti)		X	
Spese sostenute durante le escursioni verso l'estero per motivi di lavoro			X	
Numero escursioni domestiche per motivi personali	Paese di destinazione Categoria di spesa (trasporto, pasti, altri acquisti)	Sesso, classe di età, livello di istruzione, condizione lavorativa, reddito familiare		X
Numero escursioni domestiche per motivi di lavoro				X
Spese sostenute durante le escursioni domestiche per motivi personali				X
Spese sostenute durante le escursioni domestiche per motivi di lavoro				X

<sup>8</sup> Si intende, per esempio: macchine, computer, oggetti d'arte, gioielli, elettrodomestici.

<sup>9</sup> Per la classificazione NACE rev. 2 si veda il Regolamento (CE) n. 1893/2006 del Parlamento europeo e del Consiglio che definisce la classificazione statistica delle attività economiche NACE Revisione 2 <http://eur-lex.europa.eu/legal-content/IT/TXT/PDF/?uri=CELEX:32006R1893&from=en>.

<sup>10</sup> In Italia, nel 2016, il numero di escursioni è stato pari a circa 74 milioni, rispetto a circa 66 milioni di viaggi con pernottamento (Fonte Indagine CAPI "Viaggi e Vacanze" <http://dati.istat.it/>).

## 1. L'adozione del regolamento

rilevanti sia nel contesto della Bilancia dei Pagamenti (per quelle sostenute all'estero), sia per la compilazione del Conto Satellite del Turismo<sup>11</sup>. Ciononostante, l'esigenza di ridurre il fastidio statistico per i rispondenti ha condotto alla scelta di mantenere opzionale, per le escursioni, la disaggregazione della spesa complessiva nelle diverse voci che la compongono (trasporto, pasti, eccetera). Inoltre, in considerazione di una certa stabilità nel tempo del numero di escursioni svolte sul territorio nazionale, le informazioni relative al flusso e alle spese di questo tipo di visite sono richieste solo con cadenza triennale (Prospetto 1.2).

L'implementazione delle modalità di raccolta e di trasmissione delle informazioni richieste dal Regolamento verranno trattate nei successivi paragrafi.

### 1.3 Le sperimentazioni delle nuove variabili nel periodo 2011-2012

Le innovazioni introdotte nel questionario per l'adempimento alle richieste espresse dal Regolamento hanno riguardato i seguenti aspetti: la partecipazione al turismo nell'anno, le spese per viaggio e le escursioni.

La rilevazione della partecipazione al turismo nell'anno ha la finalità di fornire stime sull'ammontare di turisti in vacanza, distinguendo tra vacanze brevi e lunghe, nell'arco di un intero anno.

Come noto, l'indagine CATI "Viaggi e vacanze" nasce nel 1997 per fornire stime di viaggi, pernottamenti e turisti con riferimento trimestrale. Inizialmente, i quesiti sull'effettuazione di vacanze lunghe con riferimento temporale annuale furono introdotti nella rilevazione del 4° trimestre (che si svolgeva a gennaio dell'anno successivo), solo per permettere analisi comparative con le stime prodotte dall'indagine "Aspetti della vita quotidiana" (AVQ). Quest'ultima indagine, infatti, ha costituito fin dal 1993 la fonte ufficiale per la stime dei turisti per vacanza lunga nell'anno per la diffusione nazionale, divenendo, nel 1997, anche la fonte ufficiale per la trasmissione a Eurostat (Istat, 2003).

Nel 2012, l'entrata in vigore del Regolamento ha modificato la tempistica di trasmissione delle stime definitive sul turismo, anticipando di sei mesi la scadenza per l'invio delle stime annuali, portandola al 30 giugno dell'anno successivo a quello di riferimento. Pertanto, i tempi di realizzazione della rilevazione "Aspetti della vita quotidiana", non erano più compatibili con la scadenza per la trasmissione dei dati a Eurostat.

La raccolta dati dell'indagine AVQ avviene, infatti, nel mese di marzo e, trattandosi di un'indagine PAPI (*Paper and Pencil Interview*), è necessario un tempo di lavorazione adeguato per il controllo e la correzione.

Inoltre, come già detto, il Regolamento richiede anche la stima del numero di persone che hanno effettuato vacanze brevi durante l'anno, informazione non rilevata affatto dall'indagine AVQ.

Nel 2009, pertanto, è iniziata una sperimentazione finalizzata alla diffusione dei dati sui turisti annuali da parte dell'indagine Viaggi e Vacanze, in previsione dell'entrata in vigore del Regolamento.

A tale scopo, i quesiti del 4° trimestre che rilevano i turisti nell'anno precedente (t) sono stati replicati anche nel primo trimestre dell'anno t+1 (rilevazione del mese di aprile), rendendo di fatto maggiormente sovrapponibili i periodi di riferimento tra le indagini

<sup>11</sup> L'Istat ha pubblicato il Conto Satellite del turismo riferito agli anni 2010 e 2015. Per maggiori informazioni si veda <http://www.istat.it/it/archivio/66113>, <http://www.istat.it/it/archivio/207454>.

Viaggi e Vacanze e AVQ, per poter effettuare una comparazione. È stato innanzitutto testato se vi fosse una differenza significativa tra le stime ottenute nei due punti di osservazione nell'ambito dell'indagine Viaggi e Vacanze. La sperimentazione si è conclusa nel 2011 e ha dato risultati positivi (Tav. 1.1) poiché l'evidenza ha mostrato assenza di variazioni significative nelle risposte fornite. Quindi si è scelto di continuare a rilevare queste informazioni, all'interno dell'indagine Viaggi e Vacanze, nell'intervista del 4° trimestre dello stesso anno di riferimento dei dati, anche allo scopo di contenere l'effetto memoria.

E' noto che le stime dei turisti annuali prodotte dall'indagine Viaggi e Vacanze sono solitamente inferiori rispetto a quelle prodotte da AVQ, che infatti ha continuato ad essere la fonte ufficiale per la produzione di queste stime durante il periodo di sperimentazione sopra descritto. Tuttavia, nel periodo 2009-2012, l'avvio di un'attenta attività di miglioramento operata su più fronti (dal questionario alla formazione dei rilevatori) ha permesso di produrre stime sempre più attendibili fino all'allineamento con quelle di AVQ, nel 2012 (Tav. 1.2).

Ciò ha consentito la diffusione delle stime dei turisti per vacanza nell'anno (breve e lunga) prodotte dall'indagine Viaggi e Vacanze a partire dall'anno di riferimento 2012 (per Eurostat), in sostituzione di quelle derivanti dall'indagine "Aspetti della vita quotidiana".

**Tavola 1.1 - Turisti annuali per vacanza lunga. Anni 2009-2012** (composizioni percentuali)

Anno T	% turisti annuali per vacanza lunga anno T rilevati dal campione 4° trimestre anno T (gennaio)	% turisti annuali per vacanza lunga anno T rilevati dal campione 1° trimestre anno T+1 (aprile)
2009	45,9	47,8
2010	42,2	42,9
2011	41,4	40,4
2012	43,1	40,3

Fonte: Indagine CATI "Viaggi e vacanze"

**Tavola 1.2 - Turisti annuali di 15 anni e più per vacanza lunga. Anni 2009-2012** (composizioni percentuali)

Anno T	% (campione 4° trimestre anno T, V&V)	% (AVQ, anno T+1)
2009	43,3	48,7
2010	40,6	47,7
2011	39,0	44,3
2012	43,1	41,2

Fonte: Indagine CATI "Viaggi e vacanze", Indagine "Aspetti della vita quotidiana"

Il secondo set di quesiti che è stato modificato per adempire al Regolamento è quello relativo alle spese del viaggio. Fino al secondo trimestre 2011 l'indagine rilevava, per ogni viaggio, due grandi voci di spesa, riferite ai mezzi di trasporto utilizzati e a tutto il resto (pasti, attività ricreative, culturali e sportive, acquisti durante il viaggio, inclusa la spesa per alloggio).

Il Regolamento delle statistiche del Turismo ha richiesto la trasmissione separata del costo dell'alloggio a partire dal 2012, quindi tale voce è stata inserita nel questionario già dal 3° trimestre 2011. Al rispondente viene quindi chiesto di scorporare le categorie di spesa, identificando separatamente quella per trasporto, per alloggio e per altro.

## 1. L'adozione del regolamento

La terza grande innovazione è relativa al set dei quesiti riguardante le escursioni. La definizione di escursione è stata formalizzata, a livello europeo, durante la discussione delle linee guida attuative del Regolamento<sup>12</sup>.

In Italia, nell'indagine sulla domanda turistica, erano già presenti alcuni quesiti sperimentali per rilevare informazioni su questo fenomeno, che costituisce un segmento rilevante nel quadro del settore turistico<sup>13</sup>. La rilevazione era circoscritta alle visite a città o località italiane diverse dal comune di residenza aventi carattere di non abitualità<sup>14</sup>. Si chiedeva di specificare il luogo di destinazione, il numero di escursioni e il motivo prevalente di ogni escursione. Un ampio set di motivazioni era già codificato e comprendeva lo svago, la visita a parenti o amici, i motivi religiosi, la formazione, i trattamenti di salute e le cure mediche, lo shopping, i motivi di lavoro e la possibilità di inserire un campo di testo per altri motivi, ove non fossero già codificati. Precedenti analisi del suddetto testo avevano condotto all'inserimento di un'ulteriore motivazione dell'escursione, cioè l'accompagnare un familiare o un amico.

Le nuove indicazioni europee hanno esteso il campo d'osservazione alle città o località estere, inoltre hanno stabilito, per tutte le escursioni, una durata minima di tre ore nel luogo di destinazione (esclusi gli spostamenti di andata e ritorno). Pertanto, in considerazione della necessità di sperimentare il nuovo set di quesiti sulle escursioni per un numero adeguato di trimestri prima della trasmissione ad Eurostat<sup>15</sup>, il questionario è stato modificato già a partire dal primo trimestre del 2012. In una prima fase di sperimentazione, è stata testata la modalità di raccolta più semplice, che dava luogo al minor impatto in termini di modifica del questionario esistente. Ad ogni componente della famiglia veniva chiesto se avesse effettuato escursioni, e una serie di quesiti successivi approfondiva ognuna delle escursioni dichiarate. Una volta terminata la descrizione analitica di ogni evento, si passava al familiare successivo. Lo schema di somministrazione appena descritto però non si è rivelato efficace per garantire che le informazioni fornite dai diversi membri della famiglia fossero coerenti tra loro, nel caso di escursione effettuata insieme.

In particolare, il quesito volto a rilevare se altri familiari avessero partecipato (Box 1.1), che aveva lo scopo di replicare le informazioni descrittive dell'escursione per tutti i partecipanti, limitando il fastidio statistico, si è rivelato controproducente.

### Box 1.1 - Quesiti sulla partecipazione all'escursione

HANNO PARTECIPATO ALTRI COMPONENTI DELLA FAMIGLIA A QUESTA VISITA?	
• Sì	1
• No	2
QUALI? (COMPONENTI PARTECIPANTI ALLA VISITA) (MULTIPLA)	
• Nome Componente n. 1	
• Nome Componente n. 2	
• .....	

12 Date la rilevanza e la novità del tema, Eurostat ha costituito un'apposita task-force sulle escursioni, alla quale l'Italia ha costantemente partecipato.

13 La diffusione delle informazioni sulle escursioni non ha avuto carattere di continuità. Nel 2009 è stata diffusa la Statistica Report "Le escursioni per motivi personali in Italia" <http://www.istat.it/it/archivio/136516> (riferita al triennio 2005-2007). Per la costruzione del Conto Satellite del turismo, inoltre, sono stati utilizzati i dati sulle escursioni svolte dai residenti sul territorio nazionale.

14 Erano, pertanto, escluse le escursioni all'estero e quelle in generale legate al pendolarismo (per lavoro, per studio, per visite mediche, per sport, per acquisti).

15 La prima trasmissione obbligatoria delle informazioni riguardanti le escursioni era fissata al 2015 per quelle *outbound* e al 2018 per le *domestic*.

Infatti, il rispondente, dopo aver indicato mediante il suddetto quesito tutti i partecipanti, tendeva poi a sottostimare, per le stesse persone, il numero complessivo delle escursioni, probabilmente considerando ridondante questa informazione.

La scelta definitiva, attuata già durante la rilevazione CATI, è stata quella di mutuare, nella sezione escursioni, l'efficace sequenza di quesiti già utilizzata per la raccolta delle informazioni sui viaggi, proponendo una sezione riepilogativa iniziale, che permettesse di quantificare l'ammontare complessivo delle escursioni effettuate da ogni partecipante, prima di passare alla descrizione di dettaglio.

Pertanto, dal 2013, si rileva dapprima il numero totale di escursioni effettuate da ogni membro della famiglia nel periodo di riferimento: ciò consente di individuare ed escludere quelle "abituale"<sup>16</sup> e di predisporre al contempo una scheda atta a descrivere ogni altra escursione (non abituale). La scheda, oltre a razionalizzare il flusso del questionario, ha il valore aggiunto di poter raccogliere un set di informazioni più ampio, mantenendo inalterato il meccanismo di replica automatica delle informazioni per gli altri partecipanti, attuato mediante il quesito sopra descritto (Box 1.1).

Ai quesiti già presenti, si sono aggiunti quelli sul mezzo di trasporto e sulla spesa complessiva. L'opportunità di chiedere le spese analitiche è stata subito scartata per non recare ulteriore fastidio ai rispondenti, considerando che il Regolamento non ne chiede la trasmissione. È noto, però, che solitamente il rispondente ricorda con maggior precisione le spese sostenute se esse sono domandate voce per voce e non a livello complessivo (Crossley, Winter, 2013); in considerazione di ciò, si è optato per una soluzione di compromesso, che facilitasse il ricordo delle attività a pagamento svolte durante l'escursione, senza però chiederne le spese. Il quesito, riportato di seguito, è posto immediatamente prima della domanda riguardante la spesa complessiva (Box 1.2).

#### Box 1.2 - Quesito sulle attività a pagamento svolte durante l'escursione

ORA LE LEGGERÒ ALCUNE ATTIVITÀ, CHE RICHIEDONO DELLE SPESE. MI PUÒ INDICARE, PER OGNUNA DI ESSE, SE È STATA EFFETTUATA DURANTE LA VISITA?			
a.	Utilizzare mezzi di trasporto (anche propri)	1 sì	2 no  3 non sa
b.	Consumare pasti/bevande o altro presso ristorante/bar	1 sì	2 no  3 non sa
c.	Effettuare acquisti	1 sì	2 no  3 non sa
d.	Recarsi al cinema/museo/mostra/teatro/concerto/parco divertimenti	1 sì	2 no  3 non sa
e.	Assistere ad una manifestazione sportiva o praticare sport a pagamento	1 sì	2 no  3 non sa
f.	Altre attività a pagamento	1 sì	2 no  3 non sa

**Tavola 1.3 - Record escursioni con spesa totale = 0 o non indicata. Anni 2013-2015** (composizioni percentuali sui dati non pesati)

Anno	Record escursioni con spesa totale=0 o non indicata (%)	Record escursioni con spesa totale=0 o non indicata in presenza di almeno una attività a pagamento (%)
2013	8,2	1,8
2014	7,4	0,5
2015	6,3	0,9

Fonte: Indagine CAPI "Spese delle famiglie"

<sup>16</sup> Le escursioni "abituale" sono quelle effettuate tutte le settimane nella stessa località, fuori dal Comune nel quale si vive.

## 1. L'adozione del regolamento

Ciò ha consentito di contenere, in generale, la percentuale di escursioni per le quali la spesa totale era pari a zero o, addirittura, non indicata. Inoltre, l'incidenza di tali casi è risultata minima in presenza di almeno una delle attività a pagamento (Tav. 1.3).

### 1.4 Le nuove modalità di trasmissione dei dati a Eurostat

Il nuovo Regolamento 692/2011, oltre ad imporre agli Stati membri una maggiore tempestività per l'invio annuale a Eurostat dei dati definitivi sulla domanda turistica, basa la trasmissione dei macrodati sulla partecipazione al turismo, dei microdati sui viaggi e dei metadati, su *transmission format, processi e tool* di validazione del tutto nuovi rispetto al passato.

Le norme di scambio per l'invio dei file sono specificate da Eurostat e richiedono l'utilizzo, come unico punto di accesso, di eDAMIS (*electronic Data files Administration and Management Information System*)<sup>17</sup>, l'ambiente web integrato che implementa la Single Entry Point (SEP) policy dell'Ufficio Statistico dell'Unione Europea.

#### 1.4.1 Struttura e codificazione dei file

La precedente Direttiva Europea 97/57/CE, in riferimento alla domanda turistica richiedeva esclusivamente la compilazione, su base trimestrale e annuale, di un set di tavole di dati aggregati sul volume dei movimenti turistici (turisti, viaggi e pernottamenti) per destinazione e tipologia di viaggio, sulle caratteristiche del viaggio (durata, organizzazione, trasporto, alloggio), sul profilo del turista (sesso ed età) e sulla spesa turistica. I microdati sui viaggi e i dati aggregati sulle visite in giornata non costituivano oggetto di trasmissione a Eurostat.

Prospetto 1.3 - Dati aggregati: eDAMIS dataset

DATASET	VARIABILE	DESCRIZIONE
TOUR_PARTIC_A (Partecipazione al turismo per motivi personali)	dem_ptour_pers	Numero di residenti di 15 anni e più che hanno effettuato viaggi per motivi personali durante l'anno di riferimento
	dem_nptour_pers	Numero di residenti di 15 anni e più che non hanno effettuato viaggi per motivi personali durante l'anno di riferimento
TOUR_SDVOUT_Q (Visite in giornata verso l'estero)	dem_sdvout_pers	Numero di visite in giornata outbound per motivi personali
	dem_sdvout_prof	Numero di visite in giornata outbound per motivi di lavoro
	dem_ex_sdvout_pers	Spesa sostenuta per visite in giornata outbound per motivi personali
	dem_ex_sdvout_prof	Spesa sostenuta per visite in giornata outbound per motivi professionali
TOUR_SDVDOM_Q (Visite nazionali in giornata)	dem_sdvdom_pers	Numero di visite nazionali in giornata per motivi personali
	dem_sdvdom_prof	Numero di visite nazionali in giornata per motivi di lavoro
	dem_ex_sdvdom_pers	Spesa sostenuta per visite nazionali in giornata per motivi personali
	dem_ex_sdvdom_prof	Spesa sostenuta per visite nazionali in giornata per motivi professionali

<sup>17</sup> Per maggiori informazioni su eDAMIS si veda <http://ec.europa.eu/idabc/en/document/3537/5927.html>.

Secondo il nuovo Regolamento, invece, i dati aggregati riguardanti la partecipazione al turismo e le visite effettuate in giornata devono essere inviati a Eurostat, popolando tre distinti dataset: 1) **TOUR\_PARTIC\_A** (*Partecipazione al turismo per motivi personali*), 2) **TOUR\_SDVOUT\_Q** (*Visite in giornata verso l'estero*), 3) **TOUR\_SDVDOM\_Q** (*Visite nazionali in giornata*) (Prospetto 1.3).

Per ogni serie di dati, il Regolamento esplicita le variabili da inviare, le loro disaggregazioni e la periodicità della trasmissione, il formato (Excel o CSV), la struttura e i codici identificativi da applicare. Ogni dataset si articola in sei campi: 1) *Variable*, in cui è riportato il codice identificativo della stessa; 2) *Breakdown*, in cui è riportato il codice di identificazione della categoria di disaggregazione o della combinazione di categorie di disaggregazione della variabile osservata; 3) *Unit* con il codice identificativo dell'unità di misura ("NBR" per i valori numerici, "EUR" per i dati sulle spese); 4) *Value* contenente il valore assunto dalla variabile osservata secondo le disaggregazioni richieste; 5) *Flag* contenente i simboli che segnalano se il dato è "non affidabile e non pubblicabile a livello nazionale, ma può essere usato se aggregato a livello europeo" ("u"), "confidenziale" ("c") o "stimato" ("e"); 6) *Comment* che riporta ogni breve commento o nota riferiti a un particolare valore<sup>18</sup> (Prospetto 1.4).

**Prospetto 1.4 - Partecipazione al turismo per motivi personali: esempio di eDAMIS dataset**

Variable	Breakdown	Unit	Value	Flag	Comment
dem_ptour_pers	IT_DOM_NGE1_F	NBR	6978586		
dem_ptour_pers	IT_DOT_NGE4_Y15TO24	NBR	339568	u	

Le osservazioni e le note riguardanti non il singolo valore, ma le variabili e le loro disaggregazioni, oppure l'intero dataset, devono essere inviate in uno *sheet* a parte (se il dataset è trasmesso in Excel) oppure in un *Csv* separato (se il dataset è trasmesso in formato *Csv*). Si tratta di note metodologiche o spiegazioni aggiuntive utili per una migliore comprensione dei dati trasmessi.

Anche per i microdati sui viaggi (TOUR\_TRIPS\_A), Eurostat definisce le modalità di trasmissione, il formato (Excel o *Cvs*) e il tracciato record. Il dataset dei microdati, infatti, contiene, oltre al codice identificativo (ID) di ciascun record viaggio, 33 variabili riportanti informazioni sulle caratteristiche del viaggio (mese di partenza, durata, destinazione, mezzo di trasporto, alloggio, ecc.) e sul profilo del turista (sesso, età, situazione lavorativa, ecc.), nonché il coefficiente di riporto alla popolazione. Ulteriori variabili sono riservate all'eventuale inserimento del flag "i" associato a ciascun valore, per indicare che il dato è imputato<sup>19</sup>. Anche in questo caso, è possibile completare la trasmissione dei microdati con l'invio di note esplicative riguardanti un record viaggio, una variabile, l'associazione tra un record viaggio e una variabile o l'intero dataset.

18 Allegato II del Regolamento di esecuzione (UE) n. 1051/2011 della Commissione del 20 ottobre 2011 recante disposizioni di attuazione del Regolamento (UE) n. 692/2011 del Parlamento europeo e del Consiglio relativo alle statistiche europee sul turismo per quanto riguarda la struttura delle relazioni sulla qualità e la trasmissione dei dati <http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32011R1051>.

19 Il tracciato record del file dei microdati è illustrato nell'Allegato III del Regolamento di esecuzione (UE) n. 1051/2011 della Commissione del 20 ottobre 2011 recante disposizioni di attuazione del Regolamento (UE) n. 692/2011 del Parlamento europeo e del Consiglio relativo alle statistiche europee sul turismo per quanto riguarda la struttura delle relazioni sulla qualità e la trasmissione dei dati. <http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32011R1051>.

### 1.4.2 Il processo di validazione dei dati

Prima dell'entrata in vigore del nuovo Regolamento, il processo di validazione dei file dati trasmessi a Eurostat era interamente a carico di quest'ultimo, che effettuava tutti i controlli previsti prima della diffusione delle stime. Dal 2012, microdati e macrodati sono pre-validati dagli Stati membri al fine di garantirne la completezza, l'aderenza ai format internazionali di trasmissione, nonché la coerenza interna. A tal fine Eurostat ha implementato e messo a disposizione degli Stati membri il tool EDIT/EBB (*Editing Building Block*), che consente di effettuare controlli intra e inter-record, nonché tra dataset diversi e che rappresenta il primo esempio di servizio di validazione condiviso a livello di Sistema Statistico Europeo<sup>20</sup>. La finalità è quella di ridurre il numero degli scambi tra Eurostat e gli Stati membri e, conseguentemente, migliorare la puntualità dei dati diffusi dall'Istituto europeo attraverso il suo database.

Il processo di validazione dei dati aggregati e dei microdati prevede una fase di pre-validazione e cinque step di controllo (Prospetto 1.5): per ciascun livello, Eurostat ha definito specifiche regole di validazione con le condizioni da rispettare, le soglie di tolleranza e i livelli di *severity* (da *warning* a *error*), nonché la responsabilità dei controlli (Eurostat e/o Stati membri).

Nella fase di pre-validazione, gli Stati membri hanno il compito di verificare che siano rispettati gli standard fissati per il formato dei file da trasmettere, la struttura dei file stessi con i relativi campi e il formato dei dati ivi contenuti (posizione, tipo, lunghezza).

Il primo livello di validazione prevede controlli intra-dataset: gli Stati membri verificano la completezza e la coerenza interna di ciascun dataset di macrodati e microdati. Per i dati aggregati, si tratta di effettuare controlli sulla completezza e correttezza delle code list utilizzate per le diverse combinazioni dataset/variabile/disaggregazione, nonché sulla coerenza interna tra i valori delle variabili e le loro rispettive disaggregazioni (territoriali, per genere, per classe di età, per tipologia di vacanza, per destinazione). Per quanto riguarda i microdati, oltre a verificarne la completezza, gli Stati membri effettuano controlli sui *range* di ogni singola variabile (*validation on single position*) e sulla congruenza interna tra più variabili (*validation on multiple position*). Eurostat interviene in questa fase verificando insieme agli Stati membri la plausibilità e l'attendibilità dei record (in termini di numerosità) per ogni variabile dei dataset dei macrodati. Inoltre, Eurostat si occupa direttamente del controllo della plausibilità dei microdati, esaminando la frequenza dei valori di alcune variabili all'interno di un intervallo che tenga conto del coefficiente di riporto all'universo. Infine, l'Istituto di statistica europeo completa i controlli di primo livello, individuando gli *outliers* per le variabili sulle spese turistiche.

La seconda e terza fase del processo di validazione, che vedono impegnati insieme Eurostat e Stati membri per i macrodati e solamente Eurostat per i microdati, sono dedicate ai controlli inter-dataset, come i confronti in serie storica dei dati trasmessi, il controllo di coerenza tra i dati aggregati sulla partecipazione al turismo e le informazioni sui viaggi desumibili dai microdati e, in ultimo, i confronti con i microdati dell'offerta turistica.

Durante la quarta fase del processo di validazione, Eurostat verifica la plausibilità e la coerenza tra due domini disponibili nello stesso Istituto. Ad esempio, per i macrodati, si procede con il confronto tra le informazioni sulla partecipazione e sulla non partecipazione al turismo e quelle sulla popolazione; per i microdati, Eurostat si riserva di comparare le

<sup>20</sup> Per maggiori informazioni sul tool EDIT/EBB si veda [https://webgate.ec.europa.eu/fpfis/mwikis/essvalidserv/index.php/Introduction\\_to\\_EDIT](https://webgate.ec.europa.eu/fpfis/mwikis/essvalidserv/index.php/Introduction_to_EDIT).

stime sulle spese turistiche con quelle rilasciate dall'Indagine sulle spese delle famiglie, laddove essa venga condotta.

Infine, l'ultimo livello di validazione coinvolge sia gli Stati membri, chiamati a validare i propri dati confrontandoli con altre rilevanti statistiche ufficiali e non ufficiali, sia Eurostat, che è tenuto a effettuare comparazioni con le statistiche pubblicate da altre organizzazioni internazionali (OECD, UNWTO, ecc.).

#### Prospetto 1.5 - Il processo di validazione dei dati trasmessi a Eurostat

LIVELLO DI VALIDAZIONE	TIPO DI VALIDAZIONE		RESPONSABILITA'
	Dati aggregati (DA)	Microdati (MI)	
Pre-validazione	Formato e struttura file	Formato e struttura file	Stati Membri
	Formato dati	Formato dati	Stati Membri
Livello 1	Completezza e correttezza delle code list	Completezza	Stati Membri
	Coerenza intra-dataset tra variabili e rispettive disaggregazioni	Coerenza intra-dataset: range delle variabili ( <i>validation on single position</i> ) e controlli tra variabili ( <i>validation on multiple position</i> )	Stati Membri
	Plausibilità record	Plausibilità record	Eurostat e Stati Membri (DA); Eurostat (MI)
	-	Individuazione <i>outliers</i>	Eurostat
Livello 2 e 3	Controlli inter-dataset	Controlli inter-dataset	Eurostat e Stati Membri (DA); Eurostat (MI)
Livello 4	Plausibilità e coerenza tra due domini disponibili nello stesso Istituto	Plausibilità e coerenza tra due domini disponibili nello stesso Istituto	Eurostat
Livello 5	Plausibilità e coerenza tra dati disponibili nello stesso Istituto e informazioni disponibili al di fuori dell'Istituto	Plausibilità e coerenza tra dati disponibili nello stesso Istituto e informazioni disponibili al di fuori dell'Istituto	Eurostat e Stati Membri

### 1.5 Il report sulla qualità e sui metadati

Il Regolamento Europeo per le statistiche sul turismo ha introdotto un ulteriore importante elemento di novità, includendo, tra i prodotti da inviare ogni anno, anche i report sui metadati d'indagine e sulla qualità di processo e di prodotto<sup>21</sup>. Misurare la qualità del dato e conoscere i metadati d'indagine mediante report standardizzati per ogni paese dell'Unione Europea è utile a ogni singolo paese per monitorare l'andamento dell'indagine stessa e consente di effettuare comparazioni, trasversali e longitudinali, tra i Paesi dell'Unione.

Questi report, dal 2012 ad oggi, hanno subito un lungo processo di revisione da parte di Eurostat per accogliere tutti i cambiamenti sperimentati in materia di standard per la compilazione dei *Quality&Metadata* reports a livello europeo, fino a giungere alla forma definitiva, introdotta a partire dall'anno di riferimento 2016. La modifica principale, rispetto a quanto stabilito nel 2012, quando il Regolamento entrò in vigore, è consistita nell'unione dei due report inizialmente concordati (sui metadati e sulla qualità) in un solo output che ogni paese deve inviare annualmente, denominato *Quality report*.

<sup>21</sup> Allegato II del Regolamento di esecuzione (UE) n. 1051/2011 della Commissione del 20 ottobre 2011 recante disposizioni di attuazione del Regolamento (UE) n. 692/2011 del Parlamento europeo e del Consiglio relativo alle statistiche europee sul turismo per quanto riguarda la struttura delle relazioni sulla qualità e la trasmissione dei dati <http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32011R1051>.

In esso sono incluse, innanzitutto, alcune informazioni in merito all'aderenza di ciascuna indagine ai principi del Codice delle statistiche europee. In secondo luogo, vi sono informazioni comuni a tutti i paesi dell'Unione Europea (precompilate da Eurostat): la pertinenza dei dati rilevati, il periodo di riferimento dei dati e alcune informazioni di tipo statistico, per esempio le unità di analisi e la popolazione di riferimento, rispondenti alle specifiche richieste del Regolamento. I singoli paesi, invece, documentano le informazioni sulle misure adottate a livello nazionale per garantire adeguati standard di qualità dei processi e dei prodotti statistici, nonché informazioni sui metadati d'indagine specifici di ogni Paese.

Vengono indicate, dunque, le iniziative in materia di tutela della riservatezza dei dati; i tempi di diffusione dei risultati; l'accessibilità (principali modalità di diffusione) e la chiarezza dei dati (documentazione sulle metodologie); le misure di gestione della qualità del dato; la pertinenza, a livello nazionale, dei dati raccolti; l'accuratezza, in termini di misura dell'errore campionario (per alcuni aggregati principali) e dell'errore non campionario (rispetto al numero di unità rispondenti e non rispondenti, eleggibili e non eleggibili, i rifiuti e ad alcuni tassi come quelli di mancata risposta totale e parziale); le azioni adottate per mantenere un buon livello di accuratezza; la tempestività (rispetto al periodo di riferimento) e la puntualità (rispetto al termine del periodo di rilevazione); la coerenza e la comparabilità; il processo statistico (lista di riferimento, campione lordo e netto, disegno campionario). Completano le informazioni alcuni documenti allegati al quality report, come il questionario, in lingua originale e in inglese.

Il *quality report* è inviato annualmente, entro il 30 settembre dell'anno successivo a quello di riferimento, mediante un tool specifico, detto *ESS Metadata Handler*, direttamente sul sito di Eurostat, cui possono accedere solo i referenti dei singoli paesi. Dopo essere stato validato, è pubblicato sul sito come scheda informativa, accanto alle tavole di dati<sup>22</sup>, con accesso libero per ogni utente che desideri consultare queste informazioni<sup>23</sup>.

<sup>22</sup> Il database Eurostat è accessibile all'indirizzo <http://ec.europa.eu/eurostat/data/database>.

<sup>23</sup> La scheda informativa è disponibile all'indirizzo [http://ec.europa.eu/eurostat/cache/metadata/en/tour\\_dem\\_esms.htm](http://ec.europa.eu/eurostat/cache/metadata/en/tour_dem_esms.htm).



## 2. L'INTEGRAZIONE DELL'INDAGINE VIAGGI E VACANZE NELLA NUOVA INDAGINE SULLE SPESE<sup>1</sup>

### 2.1 I motivi dell'integrazione

Le stime ufficiali sulla domanda turistica dei residenti, dal 1° gennaio 2014, sono fornite dalla nuova indagine continua sulle Spese delle famiglie, condotta con tecnica CAPI. Tra esse, per la prima volta nel 2014, sono diffuse a livello nazionale le stime sui turisti nell'anno e sulle escursioni.

L'intero questionario dell'indagine Viaggi e vacanze è divenuto un focus incluso nella nuova indagine sulle Spese: pertanto la precedente rilevazione Viaggi e vacanze, condotta trimestralmente dal 1997 al 2013, con tecnica CATI, è cessata.

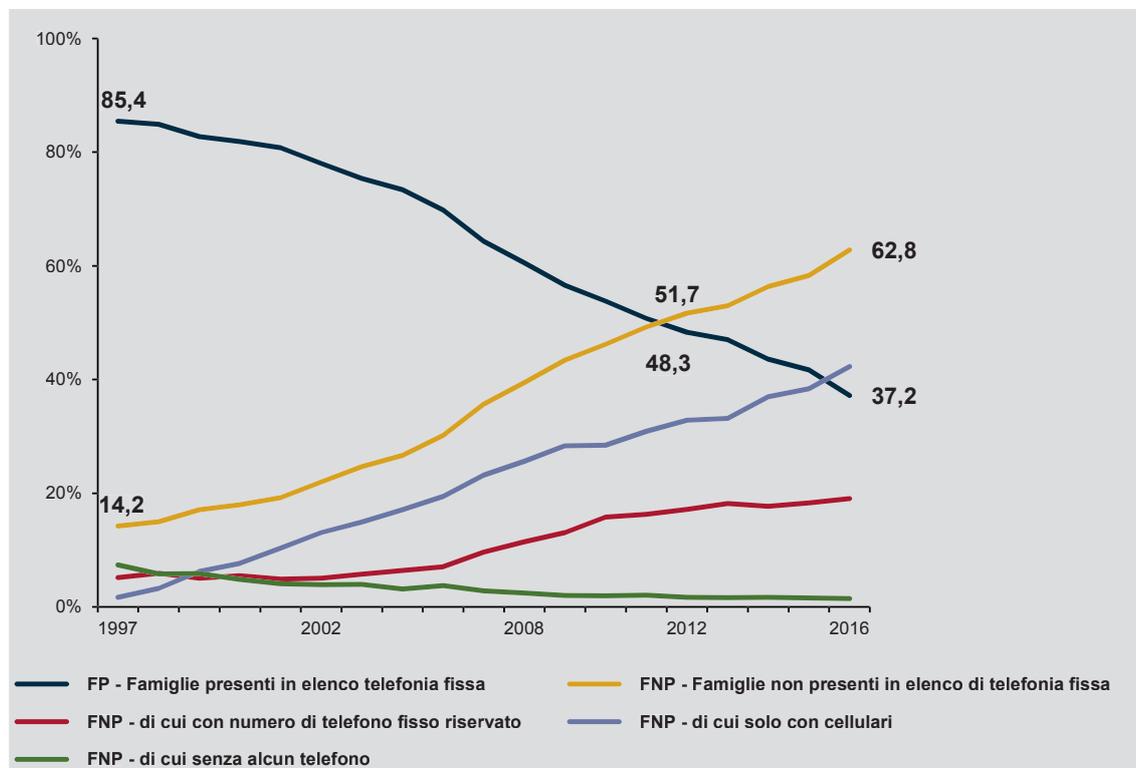
La scelta di trasformare l'intero questionario dell'indagine Viaggi e vacanze in un focus incluso nella nuova indagine sulle Spese si inserisce nell'ambito del progetto di modernizzazione delle indagini sulle famiglie (Istat, 2014), rispondendo a due ordini di motivazioni: la necessità di superare il limite della tecnica CATI, che permette di intervistare solo le famiglie in possesso di telefono fisso non riservato e la maggior economicità insita nell'unificazione delle due indagini.

In merito alla prima motivazione, è noto che il numero di famiglie presenti nell'elenco degli abbonati di telefonia fissa è diminuito progressivamente dalla fine degli Anni 90 fino ad oggi, in favore dell'aumento del numero di famiglie presenti in elenco con numero riservato e delle famiglie in possesso del solo telefono cellulare (Grafico 2.1). Nel 2012, per la prima volta, le famiglie non presenti negli elenchi hanno superato quelle raggiungibili telefonicamente, arrivando alla quota del 51,7%. Il divario si è progressivamente allargato negli anni successivi e, nel 2016, si stima che appena un terzo delle famiglie (37,2%) abbia un telefono fisso con numero non riservato.

Diversi approfondimenti (Istat, 2009a e 2009b), avviati già a partire dal 2009, hanno evidenziato la diversità delle caratteristiche socio-economiche e demografiche della sottopopolazione non presente nell'elenco telefonico rispetto alla sottopopolazione raggiungibile con un'indagine CATI. Tali differenze possono causare stime distorte della domanda turistica, poiché i comportamenti turistici dei due segmenti risultano essere diversi, con una maggior propensione a fare vacanze tra le famiglie presenti negli elenchi telefonici. Il campione estratto dalle liste di abbonati agli elenchi di telefonia fissa, per stimare la domanda turistica dei residenti con tecnica CATI, presentava quindi un problema di sottocopertura. Per risolverlo, è stata dapprima condotta una sperimentazione (De Vitiis C., Righi P., 2012) per modificare il sistema di calibrazione, allo scopo di compensare le distorsioni, ma tale operazione non si è dimostrata sufficiente. La riprogettazione dell'indagine sulle Spese delle Famiglie, basata su un campione estratto da liste anagrafiche e quindi non affetto da problemi di sottocopertura, ha offerto l'opportunità di superare le problematiche della rilevazione Viaggi e vacanze.

<sup>1</sup> Il capitolo è stato curato e redatto da M. Di Torrice.

Grafico 2.1 - Famiglie per modalità di copertura telefonica – Anni 1997-2016 (valori percentuali)



Fonte: Indagine "Aspetti della vita quotidiana"

Peraltro, l'inclusione di un modulo per la rilevazione della domanda turistica nelle indagini sulle spese delle famiglie era stata già discussa e valutata positivamente a livello internazionale. Il World Tourism Organization (UNWTO) e l'International Labour Organization (ILO) (UNWTO, 2003) avevano individuato, nel turismo, una potenziale area di interesse nel framework delle indagini sul reddito e sulle spese delle famiglie. In particolare, quest'ultima era ritenuta la più idonea a rilevare le spese per turismo, essendo un'indagine appositamente disegnata per rilevare le spese (UNWTO, 2005), con il vantaggio di essere condotta nella maggioranza dei Paesi.

Tuttavia, la realizzazione effettiva di questa potenzialità comportava una serie di difficoltà:

- le rilevazioni del fenomeno turistico devono necessariamente riferirsi alle definizioni concettuali e metodologiche delle International Recommendations on Tourism Statistics (IRTS) (UNWTO, 2010) e della metodologia comunitaria, ove il "Turista" viene distinto dal più ampio insieme dei "Visitatori" applicando il concetto di "Ambiente abituale", generalmente non definito nell'ambito delle HBS<sup>2</sup> (*Households Budget Surveys*);
- l'approccio per la rilevazione delle spese generalmente utilizzato nelle HBS non con-

2 Secondo le raccomandazioni internazionali (IRTS 2008), soltanto le visite effettuate al di fuori dell'ambiente abituale devono rientrare nella definizione di "turismo". L'individuazione dell'ambiente abituale permette di escludere dalla definizione di "visitatori" le persone che si spostano giornalmente o settimanalmente per lavoro, per studio, o per motivi personali, quando il viaggio rientra nell'ambito di attività di routine. Ciò per evitare che, in fase di misurazione statistica, si considerino come turistici degli eventi che sono connessi alla mobilità degli individui, ma rientrano nel campo di osservazione di altri fenomeni. Eurostat raccomanda di identificare la "fuoriuscita" dall'ambiente abituale adottando i seguenti criteri: attraversamento di confini amministrativi (il comune di residenza), durata, frequenza, scopo della visita (Eurostat, 2014).

sente di distinguere le spese effettuate durante un viaggio dagli altri tipi di spese della famiglia, poiché non viene richiesto di specificare dove<sup>3</sup> queste spese siano state effettuate;

- la rilevazione delle spese a livello familiare non consente di distinguere chi abbia effettuato il viaggio e quindi sostenuto la relativa spesa, tra i singoli componenti della famiglia;
- le spese connesse con un viaggio di lavoro, con l'eccezione degli esborsi a titolo personale, non devono essere rilevate tra le spese per consumi della famiglia, mentre sono oggetto di rilevazione nell'ambito della domanda turistica (anche se sostenute dal datore di lavoro<sup>4</sup>).

In definitiva, il problema da affrontare è come ricondurre delle spese per l'acquisto di singoli prodotti in spese da riferire all'evento "Viaggio", secondo le definizioni in uso per il turismo. In considerazione di tali limiti, l'ILO ipotizzò la possibilità di estendere lo scopo delle HBS per rilevare il luogo di acquisto per le spese effettuate in viaggio (ILO, 2003). Conseguentemente, UNWTO propose di includere un "tourism module" nelle indagini sulle spese/consumi delle famiglie (Di Torrice et al., 2012).

### 2.2 Il percorso verso l'integrazione

La scelta italiana di applicare quanto suggerito dall'UNWTO, innestando l'intero questionario utilizzato per la rilevazione della domanda turistica nella nuova indagine sulle Spese delle Famiglie, ha permesso di superare le problematiche precedentemente elencate. Inoltre, la rilevazione delle spese delle famiglie sia nello svolgimento della loro vita quotidiana, sia durante l'evento eccezionale "viaggio", apre nuove opportunità di ricerca sul rapporto tra spese destinate al turismo e altre spese familiari (cfr. Cap. 5).

Tuttavia, la realizzazione operativa dell'integrazione ha richiesto un notevole sforzo progettuale e di implementazione. Infatti, a fronte di una serie di ricadute positive evidenti e prevedibili sulle stime della domanda turistica, la presenza di criticità altrettanto importanti ha reso necessario adottare degli accorgimenti in diverse fasi del processo.

Tra i potenziali vantaggi dell'integrazione, oltre al già discusso superamento dei problemi di copertura delle liste telefoniche, vanno citati:

- la maggiore dimensione campionaria teorica dell'indagine sulle Spese delle Famiglie (28.000 famiglie annuali<sup>5</sup>, rispetto alle 14.000 famiglie dell'indagine CATI Viaggi e vacanze);
- i domini di stima dell'indagine sulle spese, maggiormente dettagliati a livello annuale sulle ripartizioni geografiche (5, rispetto a 3 ripartizioni geografiche dell'indagine CATI) e disegnati per fornire anche stime a livello regionale;
- la continuità della rilevazione sul campo dell'indagine sulle Spese (ogni mese), che può ridurre l'effetto ricordo sugli eventi più brevi (vacanze fino a tre giorni, visite in giornata) e su quelli più rari (viaggi di lavoro);
- la prospettiva di ridurre le risposte proxy, che erano indispensabili nell'intervista telefonica, ma non necessarie nel caso della CAPI, in cui possono comunque essere supportate dalla presenza fisica degli altri componenti della famiglia, laddove possi-

3 Se nell'ambito territoriale del Comune di residenza della famiglia o fuori dai confini comunali.

4 Secondo le IRTS 2008, i viaggi per motivi professionali vanno inclusi nella definizione di spostamento turistico.

5 28.000 famiglie è la numerosità per gli anni 2012-2016, successivamente divenuta 19.500 per gli anni 2017-2019.

- bile;
- i vantaggi di ordine gestionale, legati sia alla riduzione del tempo necessario per le attività amministrative e organizzative (in precedenza effettuate separatamente per due indagini e successivamente gestibili da risorse Istat in modo unitario), sia alla semplificazione di alcune attività comuni alle due indagini.

Diverse criticità sono state altrettanto chiare sin dall'inizio e hanno richiesto interventi mirati di gestione.

Dal punto di vista della raccolta dei dati, ci si è a lungo soffermati sulla necessità di contenere il "fastidio statistico" sui rispondenti. Il fabbisogno informativo soddisfatto dall'indagine CATI Viaggi e Vacanze non era comprimibile, poiché essa deve rispondere alle esigenze conoscitive espresse dal Regolamento Europeo e a quelle degli utenti nazionali. Sebbene la durata media dell'intervista telefonica CATI non fosse elevata (circa 10 minuti), nel momento in cui essa si doveva aggiungere a un'intervista già lunga, come quella sulle Spese delle Famiglie, ciò poteva costituire un problema da tenere sotto controllo. Per farvi fronte, il questionario dell'indagine CATI Viaggi e vacanze, già semplice poiché somministrato telefonicamente, è stato ulteriormente rivisitato per minimizzare il numero di quesiti necessari a reperire le informazioni di interesse. Inoltre, è stata prevista la possibilità che il rispondente al Focus viaggi potesse cambiare, in corso d'intervista, rispetto alle precedenti sezioni (purché fosse sempre maggiorenne).

Un'altra possibile criticità legata alla raccolta dei dati è stata fin da subito identificata nella ridondanza delle informazioni richieste alle famiglie. Come già accennato nel precedente paragrafo, l'indagine sulle Spese rileva, per i propri fini, alcune spese: 1) potenzialmente, ma non certamente turistiche, come per esempio l'acquisto di biglietti aerei (potrebbero, infatti, essere effettuate nell'ambito di spostamenti per pendolarismo); 2) univocamente identificabili come turistiche, come spese per alberghi, campeggi, eccetera; 3) "ibride", la cui "turisticità" dipende dal momento e dal luogo di effettuazione (spese per il ristorante durante un viaggio vs. spese per il ristorante per il pranzo della domenica nel proprio quartiere) e nelle quali la quota destinata al turismo non è scorporabile dal resto.

La soluzione di trasferire l'intero questionario della domanda turistica nella nuova indagine sulle spese ha risolto il problema di contestualizzare le spese nell'evento turistico, ma ha lasciato immutato il problema della ridondanza. Di segno opposto, è stata la necessità di rilevare le spese per i viaggi di lavoro, che non sono osservate nell'ambito della rilevazione sulle Spese, non essendo consumi per soddisfare i bisogni della famiglia.

Un'altra questione che si è mostrata particolarmente complessa da dirimere è stata la scelta del periodo di riferimento del Focus viaggi, ovvero se mantenere o meno il trimestre come periodo di riferimento dei quesiti sui viaggi e le escursioni, così come era nell'indagine CATI. Sono state pertanto discusse<sup>6</sup> diverse possibilità, tutte nell'ambito di uno schema di rilevazione in cui le famiglie campione sono intervistate ogni mese dell'anno, così come prevede l'indagine sulle Spese delle Famiglie, pur tenendo presente che le stime sulla domanda turistica dovevano rimanere le tradizionali stime trimestrali. In particolare, sono state esaminate due ipotesi: 1) rilevare gli spostamenti turistici conclusi nei tre mesi precedenti l'intervista, chiedendo al rispondente di riportare l'indicazione del mese di conclusione del viaggio; 2) rilevare gli spostamenti turistici conclusi nel solo mese precedente

<sup>6</sup> La discussione è avvenuta nell'ambito di un Gruppo di Lavoro ISTAT interdipartimentale costituito nel 2011 per "Valutare soluzioni per il miglioramento delle stime dell'indagine campionaria sulle famiglie Viaggi, vacanze e vita quotidiana nella transizione da tecnica telefonica ad altra tecnica assistita da computer e delle ulteriori stime derivanti dall'indagine secondo quanto previsto dal Regolamento delle Statistiche del Turismo".

l'intervista.

La prima ipotesi avrebbe consentito di basare le stime trimestrali su un campione più ampio (di numerosità tripla) rispetto alla seconda. Ogni mese, infatti, le famiglie del campione avrebbero risposto per i tre mesi precedenti (periodo mobile), e quindi ogni mese dell'anno sarebbe stato rilevato tre volte, con conseguente "irrobustimento" delle stime.

Tuttavia, adottando quest'ipotesi, per costruire stime trimestrali sarebbe stato necessario "ricomporre" il trimestre dalle risposte fornite da famiglie campionate anche in trimestri diversi, introducendo una notevole complessità di calcolo (anche sui coefficienti di riporto all'universo) e incidendo negativamente sulla tempestività: infatti, la produzione delle stime di un trimestre sarebbe stata possibile solo aspettando la fine del trimestre successivo. Infine, il mantenimento di un periodo di riferimento trimestrale non avrebbe contribuito alla riduzione del probabile effetto ricordo sulla rilevazione di eventi lontani nel tempo (soprattutto per escursioni e vacanze brevi).

Una variante della prima ipotesi consisteva nel mantenere il periodo di riferimento trimestrale, inserendo il Focus viaggi solo nei quattro punti di rilevazione coincidenti con quelli dell'indagine CATI Viaggi e vacanze, cioè aprile, luglio, ottobre e gennaio. Ma quest'ipotesi è stata fin da subito abbandonata, perché avrebbe comportato la somministrazione del Focus solo ad un terzo del campione trimestrale (2300 famiglie circa), valore inferiore a quello dell'indagine CATI.

La seconda alternativa, che consisteva nell'adozione di un periodo di riferimento mensile, presentava, invece, diversi vantaggi:

- la possibilità di adottare un unico sistema di calibrazione, sia per le stime dei consumi, sia per quelle della domanda turistica, ovvero il coefficiente familiare calcolato sui quattro campioni trimestrali indipendenti, senza sovrapposizioni tra trimestri;
- la tempestività delle stime, poiché il trimestre sarebbe stato completato nel mese immediatamente successivo alla fine del trimestre stesso, così come già avveniva nell'indagine CATI;
- la riduzione dell'effetto ricordo;
- la raccolta più puntuale delle informazioni sui viaggi e sulle escursioni, poiché le famiglie sarebbero state avvertite a ridosso della rilevazione della necessità di riferire i propri spostamenti turistici;
- la riduzione del numero complessivo di eventi da riportare, con conseguente diminuzione della durata complessiva dell'intervista e del burden statistico sulle famiglie rispondenti. Inoltre, il minor numero di eventi da registrare avrebbe semplificato la gestione del software CAPI, consentendo di superare i limiti di performance riscontrati quando si doveva trattare un elevato numero di schede viaggi. Questo aspetto era molto importante, anche in vista della necessità di aumentare il numero di schede/eventi per rilevare il dettaglio delle escursioni, così come richiesto dall'imminente Regolamento delle Statistiche del Turismo.

L'unico svantaggio consisteva nel fatto che le stime si sarebbero basate su un *subset* di famiglie rispondenti meno ampio, di numerosità pari a un terzo rispetto alla prima ipotesi (circa 2300 unità). Ogni mese dell'anno, infatti, sarebbe stato rilevato una sola volta. Tuttavia, rispetto alla precedente indagine CATI Viaggi e vacanze, la numerosità teorica del campione trimestrale risultava comunque raddoppiata (da 3500 famiglie a 7000).

E' stato stabilito, pertanto, di scegliere la seconda ipotesi, che presentava una maggio-

re semplicità sia dal punto di vista metodologico, sia implementativo<sup>7</sup>.

I problemi legati al momento della raccolta dei dati hanno riguardato anche la scelta della collocazione più opportuna del Focus viaggi all'interno del nuovo questionario. Diversamente dalle sezioni relative alle spese per consumi, i cui quesiti si rivolgono al complesso della famiglia, nel Focus viaggi la prospettiva è individuale, dovendo rilevare gli eventi turistici di ogni componente della famiglia. Inoltre, le sezioni riguardanti le spese per consumi restringono progressivamente il periodo di riferimento, passando da quesiti con riferimento annuale a quesiti trimestrali, per arrivare, infine, a quelli mensili. Conseguentemente, stabilito che il periodo di riferimento del focus dovesse essere quello mensile, la naturale collocazione delle sezioni sui viaggi, anche in virtù della loro natura di focus, è stata individuata nell'intervista finale, al termine dei quesiti sulle spese con periodo di riferimento mensile (cfr. Par. 3.2.2).

### 2.3 L'indagine pilota

La soluzione individuata è stata sperimentata mediante una rilevazione pilota nei mesi di aprile, maggio e giugno 2011, su un campione di circa 1000 famiglie dislocate in 32 Comuni.

L'obiettivo era quello di testare il nuovo impianto dell'indagine sulle Spese, soprattutto rispetto alla tecnica CAPI, agli strumenti di rilevazione (diario settimanale auto compilato per 14 giorni, questionario elettronico), alla suddivisione in fasi (visita iniziale, visita intermedia, visita finale), all'organizzazione del lavoro sul campo, alla valutazione degli effetti di burden sui rispondenti e sui rilevatori, all'introduzione del Focus viaggi. Non rientrava tra gli obiettivi la produzione di stime, data l'esiguità del campione e la limitatezza del periodo di rilevazione.

La rilevazione pilota ha dato anche l'opportunità di consolidare tutto ciò che attiene alla sfera organizzativa della nuova indagine, permettendo di mettere a punto gli strumenti definitivi per la conduzione dell'indagine sulle Spese delle Famiglie. In particolare, sono stati sottoposti a verifica:

- l'efficacia del modulo formativo, in termini di durata, sequenza degli interventi e contenuti;
- la gestione del numero verde, compreso il modulo formativo per gli operatori;
- l'ampiezza dei periodi di contatto e di somministrazione dell'intervista;
- la formulazione della lettera alle famiglie;
- l'efficienza del sistema di collaborazione e di comunicazione instaurato con i Comuni;
- la completezza e la comprensibilità del manuale di istruzioni per i rilevatori;
- il corretto funzionamento degli strumenti tecnologici di ausilio alla gestione del lavoro sul campo (portale web per le comunicazioni da e verso i Comuni e alla società esterna).

Al termine dell'indagine pilota, è stato organizzato un *debriefing* conclusivo con i rilevatori, anche al fine di discutere insieme le risposte alle problematiche riscontrate durante il lavoro sul campo. Tutti gli aspetti organizzativi hanno trovato, in questo confronto, delle

<sup>7</sup> Il periodo di riferimento mensile consente di ottenere l'ammontare trimestrale dei viaggi, delle notti e delle escursioni per somma degli eventi mensili. In più, permette di calcolare medie mensili su persone, viaggi ed escursioni riferite ad un trimestre o a un anno (cfr. par. 3.5.1). La stima dei turisti nel trimestre, calcolata al fine di mantenere la serie storica precedente, si è ottenuta inserendo quesiti retrospettivi sui mesi precedenti.

soluzioni più efficaci o una prospettiva di miglioramento.

Particolare attenzione è stata data all'analisi dei tassi di caduta, soprattutto relativamente alla motivazione del rifiuto, che ha evidenziato una forte criticità del contatto iniziale con la famiglia, dovuta alla tematica e alla complessità dell'indagine. Si è confermato, anche durante la discussione avuta coi rilevatori, quanto il primo contatto con la famiglia sia uno snodo molto critico e quanto sia fondamentale un corretto approccio da parte del rilevatore, da realizzarsi anche attraverso la comunicazione trasparente dell'impegno richiesto alla famiglia. Peraltro, si è anche constatato che la quasi totalità delle famiglie che avevano effettuato l'intervista iniziale accettavano senza particolari resistenze di proseguire con la visita intermedia e l'intervista finale.

Rispetto a quest'ultima, che presentava il problema dell'eccessiva lunghezza, è stato osservato, in considerazione del contenuto numero di interruzioni, che le famiglie sembravano essere in grado di sostenerla senza difficoltà.



### 3. LA NUOVA RILEVAZIONE DELLE INFORMAZIONI SULLA DOMANDA TURISTICA<sup>1</sup>

Un'ampia trattazione della nuova indagine sulle spese delle famiglie è fornita nel volume metodologico a essa dedicato (Istat, 2015), al quale si rimanda per tutti gli approfondimenti necessari. Nell'ambito dell'indagine sulle spese delle famiglie, la nuova modalità di rilevazione delle informazioni sulla domanda turistica verrà illustrata nei successivi paragrafi, dedicando particolare attenzione alle molteplici differenze con la precedente indagine CATI.

La novità più rilevante riguarda l'estrazione del campione, come già detto, infatti, l'indagine Viaggi e Vacanze veniva svolta mediante intervista telefonica e utilizzava come lista di selezione l'archivio ufficiale nazionale degli abbonati alle utenze private di telefonia fissa; le unità di campionamento erano, pertanto, i numeri telefonici appartenenti a detto archivio. Nel corso di un anno erano previste quattro rilevazioni, nei mesi di aprile, luglio, ottobre e gennaio, su quattro campioni indipendenti di circa 3.500 famiglie. La nuova indagine sulle spese, invece, ha un disegno campionario a due stadi con stratificazione delle unità di primo stadio, è continua nel corso dell'anno con la stessa numerosità campionaria, fissata a 19.500 famiglie<sup>2</sup>. Il campione annuale viene suddiviso in dodici sotto-campioni mensili di 1.625 famiglie, che vengono estratte dalle Liste Anagrafiche comunali (cfr. par. 3.1).

Lo strumento di rilevazione, pur rimanendo un questionario elettronico, è stato ristrutturato per essere agevolmente inserito nel flusso dell'intervista iniziale e finale dell'indagine sulle spese. Ciò ha comportato l'eliminazione di intere sezioni (i quesiti della scheda generale, per esempio) e la necessità di spostarne altre, per migliorare la fluidità complessiva. Inoltre, un grande elemento di innovazione ha riguardato il riferimento temporale dei quesiti sui viaggi e sulle escursioni, passato da trimestrale a mensile (cfr. par. 3.2).

Il cambiamento nella tecnica di rilevazione, da CATI a CAPI, ha determinato la necessità di sostituire la rete di rilevazione, pur rimanendo nell'ambito dell'affidamento ad una rete di rilevazione professionale. Ciò ha comportato una rivisitazione delle fasi relative alla formazione e al monitoraggio dell'attività dei rilevatori (cfr. par. 3.3).

L'ampio set di controlli già implementato nel questionario elettronico della vecchia indagine per verificare la congruità e la correttezza delle risposte fornite direttamente in fase di rilevazione è stato mantenuto, così come le successive procedure di controllo e correzione dei dati. Tuttavia, il piano delle incompatibilità è stato rivisto in considerazione delle nuove variabili introdotte nel questionario (per es. le escursioni), mentre le procedure di controllo e correzione sono state innovate introducendo l'utilizzo di software generalizzati (procedura famiglie, Scia) per il trattamento di alcune variabili qualitative e cambiando il software utilizzato per il trattamento delle variabili quantitative (cfr. par. 3.4).

<sup>1</sup> L'introduzione al capitolo 3, i paragrafi 3.4.4, 3.5.1 e 3.7 sono stati curati e redatti da M. Di Torrice. Il paragrafo 3.1.1 è stato curato e redatto da C. De Vitiis, il paragrafo 3.1.2 è stato curato e redatto da A. Guandalini e il paragrafo 3.5.6 è stato curato e redatto da M. D. Terribili. I paragrafi 3.3.2 e 3.3.3 sono stati curati e redatti da M. Morricone. Il paragrafo 3.1.3 è stato curato e redatto da C. Ceccarelli. I paragrafi 3.2, 3.4.6 e 3.6 sono stati curati e redatti da B. Dattilo. Le introduzioni ai paragrafi 3.3 e 3.4, i paragrafi 3.3.1, 3.4.1, 3.4.2, 3.4.3, 3.4.5 sono stati curati e redatti da M. Sabato. I paragrafi 3.5.2, 3.5.3, 3.5.4, 3.5.5 sono stati curati e redatti da A. Cutillo.

<sup>2</sup> Per il periodo 2017-2019.

Il nuovo disegno di campionamento ha consentito di migliorare il sistema di ponderazione con l'inserimento di un più ampio numero di vincoli e l'introduzione di metodi per il trattamento degli outliers (cfr. par. 3.5).

Le innovazioni introdotte nella diffusione dei dati, riconducibili solo in parte all'innesto dell'indagine Viaggi e vacanze nella nuova indagine sulle spese delle famiglie, sono state realizzate anche a seguito dell'evoluzione delle strategie di diffusione introdotte in Istituto (par. 3.6).

Una sintesi di tutte le innovazioni introdotte nell'indagine viene presentata nel par. 3.7.

### 3.1 Il disegno di campionamento e la selezione del campione

#### 3.1.1 La popolazione di riferimento, i domini di stima e lo schema generale del disegno di campionamento

La popolazione di interesse, sia dell'indagine CATI Viaggi e vacanze, sia dell'indagine sulle Spese delle famiglie, è costituita dalle famiglie<sup>3</sup> residenti in Italia e dagli individui che le compongono, al netto dei membri permanenti delle convivenze.

Nella precedente indagine CATI Viaggi e vacanze la lista di selezione delle famiglie<sup>4</sup> era l'archivio ufficiale informatizzato degli abbonati alle utenze private di telefonia fissa, costantemente aggiornato sulle variazioni degli intestatari e degli indirizzi telefonici. Le informazioni relative ad ogni indirizzo (provincia, comune, sezione di censimento, via, numero civico, ampiezza del comune di appartenenza) erano utilizzate per la stratificazione delle unità della popolazione di riferimento. Gli indirizzi telefonici erano stratificati, all'interno di ogni regione geografica, secondo la tipologia socio-demografica dei comuni<sup>5</sup> di appartenenza degli indirizzi. Il disegno di campionamento era un disegno stratificato a grappoli, in cui ciascun grappolo era un indirizzo telefonico presente nell'archivio di selezione. Tutti gli individui appartenenti ai grappoli campione, cioè alle famiglie corrispondenti ai numeri selezionati, venivano inclusi nel campione. L'estrazione degli indirizzi campione da ciascuno strato, ottenuto come incrocio della regione e della tipologia comunale, era effettuata con probabilità uguali e senza re-immissione mediante tecnica di selezione sistematica<sup>6</sup>. Ogni indirizzo selezionato dall'archivio entrava a far parte dell'elenco base e per ognuno di essi si selezionavano i tre indirizzi successivi, che entravano a far parte dell'elenco sostitutivo; in tal modo le unità sostitutive erano le più vicine possibile dal punto di vista territoriale alle rispettive unità del campione base e quindi, presumibilmente, erano anche simili dal pun-

3 Per famiglia si intende la famiglia di fatto, ossia un insieme di persone coabitanti e legate da vincoli di matrimonio, parentela, affinità, adozione, tutela o affettivi.

4 Si intende la lista di selezione delle unità campionarie.

5 I comuni italiani erano suddivisi in sei classi, formate in base a caratteristiche socio-economiche e demografiche: comuni centro dell'area metropolitana (Torino, Milano, Venezia, Genova, Bologna, Firenze, Roma, Napoli, Bari, Palermo, Catania e Cagliari); comuni che gravitano intorno al centro dell'area metropolitana; comuni aventi fino a 2.000 abitanti; comuni con 2.001-10.000 abitanti; comuni con 10.001-50.000 abitanti; comuni con oltre 50.000 abitanti.

6 La selezione del campione di indirizzi avveniva nel seguente modo: per ciascuno strato si ordinavano gli indirizzi dell'Archivio Telefonico delle Famiglie per ampiezza demografica del comune, strada, numero civico e piano; si definiva il passo di estrazione  $p$  come rapporto tra il numero di indirizzi in archivio e il numero di indirizzi da estrarre; si selezionava un numero casuale compreso tra 1 e  $p$  e si procedeva, poi, alla selezione degli indirizzi campione sulla base del passo di estrazione a partire dal punto di partenza casuale.

to di vista delle caratteristiche socio-economiche. Il campione nazionale prevedeva circa 3.500 famiglie per trimestre (per un totale annuo di circa 14.000 famiglie), che venivano intervistate nei mesi di aprile, luglio, ottobre e gennaio.

Nell'indagine sulle Spese delle famiglie, sia la lista di selezione delle unità campionarie sia la loro numerosità sono diverse rispetto all'indagine CATI Viaggi e vacanze. La prima, infatti, è data dalle unità appartenenti all'archivio unico delle famiglie residenti nei comuni italiani, proveniente dalle liste anagrafiche comunali (LAC) (cfr. par. 3.1.3), la seconda è stata ampliata per cui, in ogni trimestre, si intervistano circa 4.875 famiglie, per un totale di 19.500 famiglie all'anno<sup>7</sup>.

L'indagine CAPI, come la precedente versione CATI, ha mantenuto la finalità di fornire informazioni sui movimenti turistici della popolazione e, pertanto, di fornire stime di parametri di diversa natura (totali, medie, rapporti, frequenze assolute e relative) con più riferimenti temporali e territoriali sul numero di turisti, di viaggi, di pernottamenti in viaggio e di escursioni sul territorio nazionale o all'estero.

In continuità rispetto al passato, 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 cambiati<sup>8</sup>:

- 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.

Il tipo di disegno adottato nell'indagine sulle Spese delle famiglie 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, 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. Quest'ultima - espressa in termini di popolazione residente - viene utilizzata per suddividere i comuni, all'interno di ogni incrocio definito da regione e tipologia comunale, tra: (i) *Auto Rappresentativi* (AR), comuni con popolazione al di sopra di una soglia opportunamente definita (cfr. par 3.1.2.), ognuno dei quali rappresenta uno strato; ii) *Non Auto Rappresentativi* (NAR),

<sup>7</sup> Per il triennio 2017-2019.

<sup>8</sup> Nella CATI le ripartizioni territoriali erano tre: Nord, Centro e Mezzogiorno; le tipologie comunali erano sei: comuni centro dell'area metropolitana (Torino, Milano, Venezia, Genova, Bologna, Firenze, Roma, Napoli, Bari, Palermo, Catania e Cagliari), comuni che gravitano intorno al centro dell'area metropolitana, comuni aventi fino a 2.000 abitanti, comuni con 2.001-10.000 abitanti, comuni con 10.001-50.000 abitanti, comuni con oltre 50.000 abitanti.

comuni con popolazione al di sotto della suddetta soglia. Per le famiglie appartenenti ai comuni AR si adotta, pertanto, un disegno ad uno stadio stratificato, mentre per l'insieme dei comuni NAR si usa un *disegno a due stadi con stratificazione delle unità di primo stadio*. Da ogni comune campione, sia esso AR o NAR, si estrae un numero predeterminato di famiglie anagrafiche che costituiscono le unità finali di campionamento.

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 (Prospetto 3.1):

**Prospetto 3.1 - Partecipazione dei comuni NAR all'indagine sulle Spese delle famiglie 2017-2019**

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

### 3.1.2 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:  $h$  ( $h=1, K, H$ ) l'indice di strato di comuni;  $i$

### 3. La nuova rilevazione delle informazioni sulla domanda turistica

( $i=1, K, N$ ) 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:

- definizione del numero complessivo di famiglie campione a livello nazionale, pari a 19.500 famiglie annue;
- definizione del numero  $m_r$  di famiglie campione per ciascuna regione, definito mediante una procedura di allocazione ottima multivariata del campione (Bethel, 1989; Falorsi et al., 1998) che consente di tenere sotto controllo gli errori campionari attesi dei principali capitoli di spesa (di interesse dell'indagine sulle spese);
- distribuzione delle numerosità regionali tra i domini definiti dalla tipologia comunale in misura proporzionale alla popolazione residente;
- scelta del numero minimo di famiglie,  $\bar{m}_r$ , da intervistare in ciascun comune campione.

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

- calcolo della frazione di campionamento regionale  $f_r = m_r / M_r$ , essendo  $M_r$  il numero di famiglie residenti nella regione  $r$ ;
- determinazione del valore della soglia<sup>9</sup>  $\lambda_r$ , mediante la relazione

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

in cui  $\lambda_r$  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  $\bar{m}_r$ ;

- suddivisione dei comuni in AR e NAR sulla base della soglia  $\lambda_r$ ;
- ordinamento decrescente dei comuni NAR all'interno di ogni dominio in funzione della loro dimensione demografica;
- suddivisione dei comuni NAR in strati la cui dimensione è approssimativamente uguale al prodotto  $n \times \lambda_r$ ;
- selezione di  $n$  comuni campione da ciascuno strato  $h$  ( $h=1, \dots, H$ ) con probabilità proporzionale all'ampiezza; per il generico comune  $c$  tale probabilità è espressa dalla formula:

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

- 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'in-

<sup>9</sup> 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.

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

clusione di primo e di secondo stadio), si ottiene

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.

Nel Prospetto 3.2 vengono riportate le numerosità campionarie in termini di comuni.

**Prospetto 3.2 - Numero di comuni del campione indagine sulle Spese delle famiglie 2017-2019**

	Comuni campione		
	AR	NAR	TOTALI
Mese	47	148	195
Trimestre	47	444	491
Anno	47	444	491

L'estrazione delle famiglie campione viene effettuata, mediante selezione casuale, dalla lista delle anagrafi comunali (cfr. paragrafo 3.1.3); 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;

Sia nei comuni AR sia nei NAR, il campione mensile di famiglie viene ulteriormente ripartito in due sottoinsiemi, distinguendo le famiglie che compileranno il diario nella prima metà del mese da quelle che lo dovranno compilare nella seconda.

La procedura di assegnazione dei periodi di compilazione del diario prevede che per ciascun dominio, definito dall'incrocio della regione e della tipologia comunale, l'assegnazione avvenga in maniera casuale.

### 3.1.3 L'utilizzo delle liste anagrafiche comunali per la selezione del campione

Nell'ambito delle operazioni censuarie, l'ISTAT ha raccolto le liste anagrafiche di tutti i comuni italiani a partire dal 31 dicembre 2010.

Per ogni occasione è stato creato un archivio che contiene le informazioni relative a tutti gli individui residenti sul territorio di ciascun comune italiano, distinti per famiglia o convivenza (Liste Anagrafiche Comunali-LAC). Tali informazioni sono state raccolte tramite un'applicazione *web*, messa a disposizione dall'Istituto, che ha garantito il controllo e la qualità dei dati acquisiti, nel rispetto dei più elevati standard di sicurezza, e ha consentito un efficace supporto centralizzato e un monitoraggio in "tempo reale" dello stato delle operazioni (Ceccarelli *et al.* 2013).

Oltre ad essere uno degli strumenti utilizzati nell'ambito del Censimento generale della popolazione e delle abitazioni del 2011, tale archivio rappresenta un elemento cardine di innovazione, contribuendo alla standardizzazione dei processi e all'armonizzazione delle

metodologie su cui si basano le indagini sociali e socioeconomiche dell'Istat.

Dal 2011, il processo di estrazione dei campioni è svolto interamente all'interno dell'Istat, tramite l'impiego della LAC che garantisce, oltre alla standardizzazione delle procedure, una notevole tempestività nell'estrazione dei nominativi delle famiglie campione, grazie all'azzeramento dei tempi prima necessari per tutte le attività di competenza dei singoli comuni coinvolti. Tale procedura di estrazione, che agisce uniformemente su un *database* centralizzato, è stata anche in grado di attenuare in modo significativo i possibili effetti distorsivi dovuti all'inclusione o esclusione non casuale di famiglie residenti in alcune zone del territorio. Inoltre, è possibile definire, sia a priori sia in corso d'opera, strategie di sovra-campionamento, soprattutto nel caso di particolari sottogruppi di popolazione dove il fenomeno delle mancate risposte totali risulta più diffuso; si pensi, ad esempio, alla popolazione straniera che, a causa sia della lingua sia della più elevata mobilità sul territorio, presenta tassi di mancata partecipazione più elevati rispetto al resto della popolazione residente.

In particolare, nel caso dell'Indagine CAPI sulle spese che prevede il ricorso alle sostituzioni per le mancate risposte, l'uso delle LAC ha consentito di rimpiazzare gli elenchi sostitutivi con le sestine<sup>10</sup>, in analogia con quanto fatto per altre indagini (come già accennato, ad ogni famiglia base vengono associate 5 famiglie sostitutive con caratteristiche simili), rendendo molto più mirata la scelta della famiglia sostitutiva e standardizzandone le modalità di selezione.

Infine, a ogni estrazione, il campione di famiglie viene confrontato con i campioni già estratti in altre indagini e con quello dell'Indagine sulle spese relativo all'anno precedente.

In tal modo, si evita che la famiglia partecipi a più indagini nel corso dell'anno o all'indagine per due volte consecutive, riducendo così la mancata partecipazione e la molestia statistica.

L'esperienza già maturata, tuttavia, evidenzia due criticità nell'uso delle LAC. La prima è la complessità intrinseca delle procedure e delle elaborazioni necessarie affinché l'archivio possa essere utilizzato, soprattutto per quanto riguarda l'intero processo di acquisizione e aggiornamento della LAC. La seconda, parzialmente legata alla prima, è la discrepanza temporale fra la data di acquisizione delle LAC e quella dell'estrazione: più elevata è la discrepanza, tanto maggiore è la probabilità di estrarre famiglie che hanno cambiato abitazione e che, per questo, risulteranno irreperibili.

Le LAC rappresentano anche un possibile strumento per il trattamento della mancata risposta totale, fenomeno sempre più diffuso nelle rilevazioni statistiche del quale si deve tenere conto, anche da un punto di vista teorico, a partire dalla definizione della strategia di indagine (Särndal e Lundström, 2005).

La complessità delle indagini, soprattutto in campo socio-economico, la concomitanza di differenti tecniche di rilevazione, l'utilizzo di diverse reti di rilevazione impongono risposte metodologiche articolate, che prevedono la disponibilità di informazioni anche sui non rispondenti: lo sfruttamento di fonti aggiuntive può limitare gli effetti distorsivi dei meccanismi di mancata risposta totale sulle stime finali. Con l'uso degli stimatori post-stratificati e soprattutto degli stimatori calibrati (Deville e Särndal, 1992), l'impiego di informazioni ausiliarie da fonti esterne per la costruzione delle stime da indagini campionarie è ormai prassi consolidata, anche e soprattutto nel trattamento delle mancate risposte totali (Kott 2006; Ceccarelli e Cutillo, 2007; Ceccarelli *et al.*, 2008).

<sup>10</sup> Per il triennio 2017-2019.

È noto che i passaggi che dal campione teorico portano alla determinazione delle stime finali o, meglio, alla determinazione dei coefficienti di riporto all'universo, possono essere riassunti in due macro fasi ben distinte:

- una volta raccolte le informazioni e determinato il coefficiente di riporto all'universo da disegno, si procede ad adeguare il campione realizzato a quello teorico; si tiene cioè conto della mancata risposta totale in modo da ripristinare sia la numerosità campionaria di partenza sia la struttura del campione inizialmente estratto;
- grazie all'uso di stimatori, più o meno complessi, che sfruttano informazioni ausiliarie, si arriva alla determinazione dei coefficienti di riporto all'universo finali che eliminano anche la distorsione dovuta alla sotto copertura delle liste di campionamento.

Nel contesto della prima fase si inseriscono le sperimentazioni sull'utilizzo delle LAC per la determinazione dei coefficienti correttivi della mancata risposta totale.

### 3.2 La struttura del questionario

Come già detto (cfr. par. 2.1), l'integrazione dell'indagine Viaggi e Vacanze all'interno dell'indagine sulle Spese delle famiglie è stata realizzata mediante la progettazione di un modulo aggiuntivo, da inserire nel nuovo questionario CAPI dell'indagine sulle Spese delle famiglie.

Il questionario di Viaggi e vacanze diventa, quindi, un approfondimento tematico dei comportamenti di spesa delle famiglie italiane relativamente all'evento turistico, inteso sia come viaggio con pernottamento, sia come visita in giornata (definita, nell'ambito delle statistiche sul turismo, "escursione"). I quesiti della parte dedicata a Viaggi e vacanze si inseriscono in diversi momenti del complesso ciclo di rilevazione dell'indagine sulle Spese delle famiglie, di seguito sinteticamente illustrato.

#### 3.2.1 I tempi della rilevazione

La rilevazione, continua, è condotta ogni mese dell'anno. Ogni famiglia riceve tre visite consecutive da parte del rilevatore per completare l'intero ciclo d'intervista, che avviene mediante la somministrazione di questionari CAPI e l'autocompilazione di un diario cartaceo. Ogni sottocampione mensile di famiglie partecipa alla rilevazione in uno specifico mese di riferimento, che comporta il coinvolgimento delle famiglie campionate in più mesi di calendario. La prima visita (visita iniziale) si svolge all'interno di un intervallo temporale di 20 giorni, che termina il giorno che precede l'inizio del periodo assegnato alla famiglia per compilare il diario; in occasione di tale visita, il rilevatore somministra l'intervista inizia-

Figura 3.1 - Il ciclo di rilevazione (per un generico mese M di riferimento)

		Inizio ciclo	Mese M (Mese di riferimento)																				Mese M+1							Fine ciclo
		12	...	31	1	...	5	...	11	...	15	16	17	18	...	22	...	31	1	...	5	6	7	...	10	11	...	15	16	17
GRUPPO 1	Primo contatto e intervista iniziale	Compilazione del diario																		Controllo e ritiro del diario							Chiusura data entry diari			
		Visita intermedia																		Intervista finale										
GRUPPO 2		Primo contatto e intervista iniziale																		Controllo e ritiro del diario							Chiusura data entry diari			
		Visita intermedia																		Intervista finale										
		Inizio data entry diari																		Inizio data entry diari										

le e consegna il diario cartaceo. Al fine di coprire tutti i giorni di calendario, le famiglie del sottocampione mensile sono suddivise in due gruppi: le famiglie del Gruppo 1, che compilano il diario per 14 giorni nella prima metà del mese di riferimento, e quelle del Gruppo 2, che lo compilano per 14 giorni nella seconda metà (Figura 3.1).

La seconda visita (visita intermedia) è effettuata tra il quarto e l'undicesimo giorno di compilazione del diario con l'obiettivo di verificare la compilazione dello stesso, fornire spiegazioni o aiuto e, se necessario, sollecitare le famiglie "pigre", minimizzando così eventuali interruzioni d'intervista. Questa seconda intervista può essere condotta anche per telefono, somministrando lo stesso questionario.

La terza visita (visita finale) ricade nei primi 15 giorni del mese successivo a quello di riferimento per tutte le famiglie<sup>11</sup>, sia per quelle del Gruppo 1, sia per quelle del Gruppo 2. Durante la visita, il rilevatore effettua un nuovo *check* del diario, recuperando insieme alla famiglia le informazioni mancanti o incomplete e, subito dopo, somministra l'intervista finale.

Al termine del ciclo di visite i rilevatori effettuano la trasmissione dei dati d'intervista e il data entry del diario cartaceo su supporto informatico.

#### 3.2.2 Dove si collocano i quesiti sui viaggi

Nella riprogettazione del questionario di Viaggi e vacanze, particolare attenzione è stata posta nella valutazione della posizione migliore in cui inserire i quesiti rispetto a quelli sulle Spese delle famiglie e, di conseguenza, nei cambiamenti indispensabili per la loro integrazione (passaggi da una sezione all'altra).

La presenza simultanea delle due indagini in un unico impianto di rilevazione e la suddivisione dello stesso in tre occasioni temporali distinte, ha fatto optare, inizialmente, per una suddivisione del vecchio questionario di Viaggi e vacanze in due parti.

La scheda generale dell'intervista iniziale del questionario dell'indagine sulle spese, essendo già finalizzata alla raccolta di informazioni socio-economiche per ciascun membro della famiglia, ben si prestava ad accogliere i quesiti introduttivi di contatto e quelli sulle informazioni strutturali<sup>12</sup> del questionario di Viaggi e vacanze<sup>13</sup>.

I quesiti sui viaggi e sulle escursioni, invece, sono stati posti dopo le sezioni dedicate alle spese familiari riferite all'ultimo mese, a chiusura dell'intervista finale, replicando la struttura del questionario CATI, sia in termini di contenuto sia in termini di flusso (cfr. Par. 2.2). La scelta di una sequenzialità tra le sezioni sulle spese e quelle sui viaggi aveva l'obiettivo di semplificare il passaggio dall'una all'altra per gli intervistati e per i rilevatori. Infatti, va sottolineato che l'oggetto di analisi rispetto al quale riportare le informazioni cambia:

11 Con l'eccezione dei mesi di luglio e dicembre, per i quali la terza visita ricade rispettivamente nei primi 25 giorni e 20 giorni del mese successivo.

12 Informazioni individuali (sesso, età, titolo di studio, stato civile, condizione lavorativa, cittadinanza, etc.) e familiari (numero dei componenti della famiglia).

13 Nella precedente rilevazione CATI, invece, tali informazioni erano suddivise in due momenti diversi della stessa intervista, uno all'inizio e uno alla fine, dopo la parte sui viaggi e sulle escursioni. Tale scelta era stata dettata dalla necessità di rendere più fluida l'intervista, contenere il numero di cadute iniziali e mantenere elevata la qualità delle informazioni sulle variabili tematiche. Infatti, essendo l'intervista svolta in unico contatto, suddividere la raccolta delle informazioni strutturali in due momenti della stessa intervista distanti tra di loro, alleggeriva il carico sul rispondente, posponendo una parte dei quesiti a un momento in cui la fiducia e livello di disponibilità erano senz'altro più elevati. Quindi, inizialmente, erano raccolte le sole informazioni necessarie per stabilire un contatto con la famiglia, identificare correttamente ciascun componente e permettere al software di strutturare correttamente l'intervista. In seguito, a chiusura d'intervista, erano rilevate le altre informazioni strutturali.

nelle sezioni sulle spese le unità di analisi sono l'abitazione e la famiglia, mentre nelle sezioni sui Viaggi sono il viaggio o l'escursione. Inoltre, la collocazione delle sezioni sui viaggi dopo l'ultima sezione sulle spese evita all'intervistato un eccessivo impegno mnemonico, poiché il periodo di riferimento dei quesiti è lo stesso (il mese).

Tale scelta, mantenendo il più possibile invariati i due questionari, ha permesso di garantire la continuità delle serie storiche, consentendo la ricostruzione dei principali aggregati. Le modifiche sulle singole domande, dovute alle richieste del Regolamento europeo e ad alcuni miglioramenti nel *wording* dei testi, sono state riportate per tutto il periodo di sovrapposizione (6 trimestri, dal 3° 2012 al 4° 2013), in entrambe le versioni, CATI e CAPI, del questionario di Viaggi e vacanze.

L'unica differenza di rilievo, rispetto al questionario CATI, riguarda il diverso collocamento dei quesiti sui turisti per vacanza nell'anno (ammontare delle persone che effettuano almeno una vacanza breve o lunga nell'anno t). In virtù del diverso timing della rilevazione CAPI, è stato necessario somministrarli al campione di famiglie del primo trimestre dell'anno t+1, invece che a quello del 4° trimestre dell'anno t (come era per la CATI). Il motivo di questa scelta risiede nel fatto che il focus Viaggi e vacanze è inserito nell'intervista finale, e nel caso del 4° trimestre questa è realizzata per due terzi del campione quando l'anno non è ancora terminato, cioè all'inizio dei mesi di novembre e dicembre, e ciò non consentiva di avere l'informazione sui turisti annuali per l'intero anno precedente, ma solo su 10 e 11 mesi rispettivamente. Per l'anno di riferimento 2012, durante il periodo di sovrapposizione delle due indagini CATI e CAPI, si è sperimentata, quindi, l'introduzione della sezione sui turisti annuali nel campione del primo trimestre 2013 in entrambe le indagini<sup>14</sup> (quindi, nel caso dell'indagine sulle spese, durante le interviste finali di febbraio, marzo e aprile dell'anno t+1).

Tuttavia, gli stessi quesiti non hanno dato i risultati attesi. Il confronto delle stime CATI-CAPI ha evidenziato una forte sottostima per tutte le tipologie di turisti, sia per vacanza breve, sia per vacanza lunga, da parte dell'indagine sulle spese.

La spiegazione più verosimile per queste differenze va ricercata nella posizione dei quesiti all'interno del questionario. La sezione sulla partecipazione al turismo nell'anno appena trascorso, infatti, si trova nell'ultima parte dell'intervista finale CAPI, dopo una serie molto lunga di quesiti sulle spese, con riferimenti temporali diversi e dopo l'approfondimento dei viaggi effettuati nel mese di riferimento. Potrebbe essersi verificato un effetto stanchezza, dipeso probabilmente dall'eccessivo carico sui rispondenti dovuto alla somministrazione congiunta dei quesiti sulle spese e sul turismo.

Per l'anno di riferimento 2013, è stato quindi deciso di ripensare la posizione della sezione sui turisti annuali anticipandola all'intervista iniziale, sempre del primo trimestre. La parte del questionario idonea ad accogliere queste domande individuali si colloca alla fine della scheda generale e prima dei quesiti familiari dell'indagine Spese. Per testare e controllare l'effetto ricordo, i quesiti sono stati inseriti anche nell'intervista iniziale del secondo trimestre. Tuttavia, questo cambiamento comporta che il primo sotto campione mensile, intervistato a dicembre dell'anno precedente, ha come periodo di riferimento gli 11 mesi precedenti il momento dell'intervista e non l'intero anno. Inoltre, lo spostamento della sezione sui turisti annuali dall'intervista finale a quella iniziale, comporta un leggero aggravio del *burden* statistico sulle famiglie perché nell'intervista finale i quesiti possono essere rivolti solo a chi non ha viaggiato nel mese (chi viaggia nel mese di riferimento è automaticamente

<sup>14</sup> La somministrazione di tali quesiti nel primo trimestre, nell'indagine CATI "Viaggi e vacanze", era stata già sperimentata, con risultati positivi, per gli anni 2009-2011 (cfr. par. 1.3).

### 3. La nuova rilevazione delle informazioni sulla domanda turistica

**Tavola 3.1 - Turisti annuali per vacanza lunga e breve. Confronto Cati-Capi con intervalli di confidenza. Valori in migliaia. Anno 2013**

Stima	CATI		Stima	CAPI	
	Intervallo di confidenza 95%			Intervallo di confidenza 95%	
	Estremo inferiore	Estremo superiore		Estremo inferiore	Estremo superiore
	Turisti per Vacanza lunga				
21.916	21.182	22.650	19.244	18.667	19.821
	Turisti per Vacanza breve				
11.546	10.990	12.102	10.720	10.269	11.171

Fonte: Indagine CATI "Viaggi e vacanze", Indagine CAPI "Spese delle famiglie"

anche turista nell'anno), mentre in quella iniziale i quesiti devono essere somministrati a tutti i membri della famiglia.

Aver anticipato tali quesiti all'intervista iniziale si è rivelata la scelta migliore, poiché ha permesso sia di superare l'effetto stanchezza, come dimostra la notevole diminuzione del numero di risposte mancanti, sia di avere stime più accurate, del tutto comparabili con quelle rilevate con la parallela rilevazione CATI (Tavola 3.1). La differenza osservata tra le stime dei turisti per vacanza lunga nelle due rilevazioni, CATI e CAPI, è attribuibile, invece, ai problemi di sottocopertura che già dalle ultime edizioni avevano afflitto la lista degli abbonati di telefonia fissa da cui si estraggono i nominativi, rendendo il campione di famiglie non più adeguatamente rappresentativo del fenomeno (Istat, 2014).

A partire dal periodo di riferimento 2013, i turisti nell'anno sono rilevati, quindi, nelle interviste iniziali del primo e del secondo trimestre dell'anno successivo a quello di riferimento: a dicembre dell'anno  $t$ , a gennaio, febbraio, marzo, aprile e maggio dell'anno  $t+1$ . Per le sole interviste di dicembre, il periodo di riferimento dei quesiti è 1° gennaio-30 novembre dell'anno in corso; in tutti gli altri casi è 1° gennaio-31 dicembre dell'anno precedente.

#### 3.2.3 Il promemoria e l'intervista proxy

Al termine della prima visita, il rilevatore consegna alla famiglia, insieme al diario delle spese, un promemoria (Figura 3.2) per ricordare che ci sarà un'altra visita all'inizio del mese successivo a quello di compilazione del diario. Durante la visita finale la famiglia sarà intervistata su alcune specifiche tipologie di spese sostenute e su eventuali viaggi ed escursioni conclusi nel mese di compilazione del diario.

Il promemoria è uno strumento da sempre utilizzato per l'indagine Viaggi e Vacanze, in passato inviato per posta insieme alla lettera alle famiglie (informativa) pochi giorni prima dell'inizio della rilevazione. In occasione della nuova indagine CAPI, il promemoria è stato riprogettato nella sua veste grafica e arricchito nei contenuti, allo scopo sia di favorirne la compilazione da parte delle famiglie, sia di includere uno schema aggiuntivo per le escursioni che nella vecchia indagine non era previsto. Per facilitare la raccolta delle informazioni, nel promemoria sono riportate le definizioni di viaggio, escursione e spesa per viaggio. Si chiede a tutti i membri della famiglia di riportare i viaggi e/o le escursioni che avranno termine nel corso del mese successivo e di annotare le caratteristiche principali di tali eventi (incluse le spese complessive). Tale promemoria, se compilato, potrà essere di aiuto durante l'intervista finale.

La possibilità di consegnare direttamente il promemoria alla famiglia, invece di essere spedito per posta e la presenza dell'intervistatore, garantiscono una maggior attenzione agli

Figura 3.2 - Il promemoria per l'intervista finale su viaggi ed escursioni

### Provate a compilare i due schemi, inserendo....

...i **viaggi** fuori dal Comune dove vivete, effettuati per motivi personali o di lavoro, conclusi nello stesso mese di compilazione del diario. I motivi personali possono essere: vacanza, visita a parenti o amici, motivi religiosi, trattamenti di salute, cure termali. Escludete, per il momento, i viaggi fatti nello stesso luogo tutte le settimane.



Riportare la **spesa totale** di tutti i familiari partecipanti al viaggio. Includere tutti i **costi del viaggio** (alloggio, trasporto, pasti, altro) + le spese sostenute **prima della partenza** (per es. manutenzione della macchina, acquisto di creme solari o attrezzature, assicurazione) + le **spese al rientro** (per es. stampa delle foto).

Viaggio n.	Nomi di chi ha viaggiato	Destinazione	Motivo (lavoro/vacanza/altro)	Data Partenza	Durata (in notti)	Spesa totale per il viaggio (€ euro)
1	Laura, Mario	Cortina d'Ampezzo	Laura in missione di lavoro, Mario è andato a sciare	12/3/2015	6	1380
2	Mario	Francia	Lavoro (fiara)	21/3/2015	3	963

... **gli spostamenti in giornata**, ovvero le visite fuori dal Comune dove vivete, senza pernottamento, effettuate per motivi personali o di lavoro, svolte nello stesso mese di compilazione del diario. Le visite devono iniziare e terminare nella vostra abitazione e avere una durata di almeno 3 ore nel luogo visitato (spostamenti esclusi).

Visita n.	Nomi di chi ha partecipato	Destinazione	Motivo (lavoro/vacanza/altro)	Spesa totale per la visita (€ euro)
1	Laura, Mario	Como	Sono andata a trovare la zia, Mario mi ha accompagnato	50
2	Laura	Monza	Shopping all'outlet	250

*Grazie per la collaborazione!*

eventi turistici. Con un confronto diretto, ad esempio, è possibile rassicurare la famiglia, ma anche sciogliere dubbi e perplessità, chiarendo, se richiesto, concetti e definizioni necessari a rispondere correttamente ai quesiti sul turismo somministrati nel corso dell'intervista finale.

Inoltre, l'organizzazione della rilevazione CAPI in tre fasi, favorisce la funzione del promemoria come "memory jogger" per le informazioni su viaggi ed escursioni. Infatti, le famiglie apprendono nel corso dell'intervista iniziale che saranno chiamate a rispondere successivamente, durante l'intervista finale, su eventi che si devono ancora verificare,

### 3. La nuova rilevazione delle informazioni sulla domanda turistica

come viaggi ed escursioni, ma sui quali potranno focalizzare la loro attenzione nei giorni successivi la visita del rilevatore, eventualmente prendendone nota nel promemoria.

Per quanto riguarda la modalità di conduzione dell'intervista, così come accadeva per la CATI, anche nella CAPI un membro della famiglia maggiorenne risponde per sé e, in modalità *proxy*, per tutti gli altri componenti della famiglia, riportando le caratteristiche dei viaggi e delle escursioni effettuati da ciascun componente. Tuttavia, in sede d'intervista iniziale si raccomanda che tutti i familiari, se possibile, siano presenti al momento dell'intervista finale perché possono partecipare fornendo personalmente le informazioni sui loro viaggi e sulle loro escursioni. Tale forma partecipativa è certamente più gestibile nel caso di un'intervista a domicilio, rispetto a una telefonica.

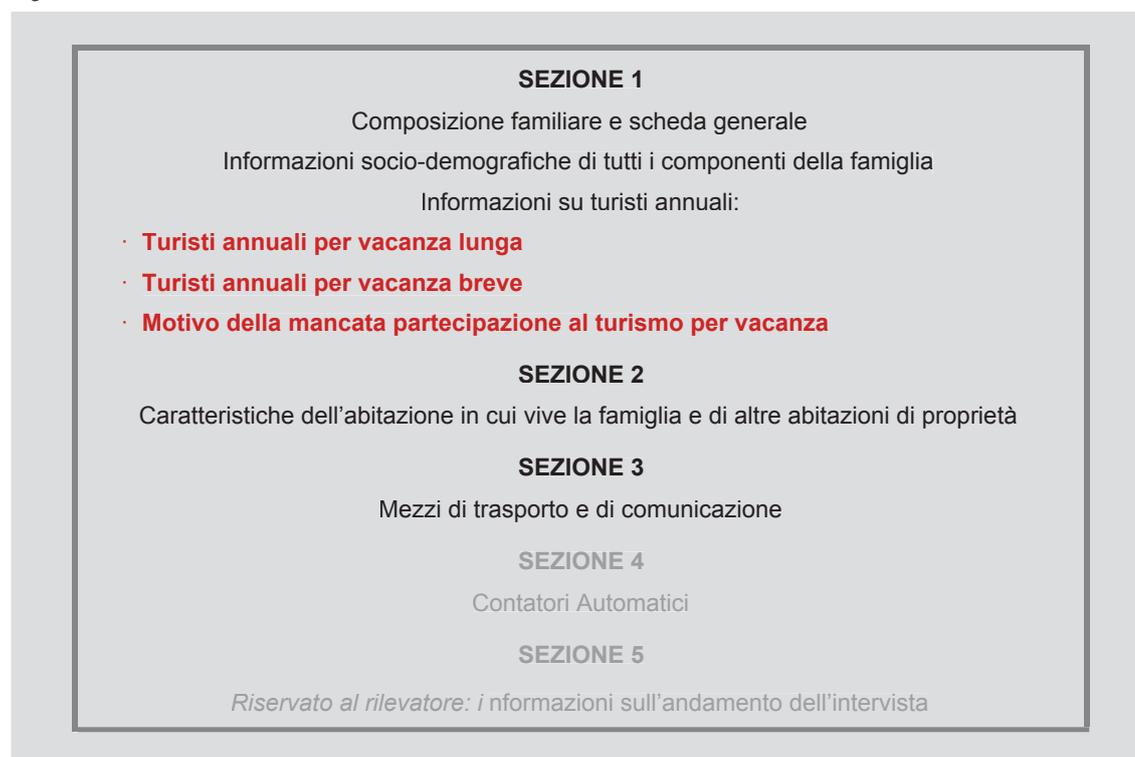
I quesiti vengono somministrati alla persona di riferimento della famiglia, che può essere:

- un membro della famiglia maggiorenne;
- un'altra persona maggiorenne che vive in famiglia legata da vincoli economici (es. collaboratore domestico convivente, baby-sitter convivente, affittuario o simile);
- un parente o altra persona maggiorenne che non vive in famiglia ma che sia in grado di rispondere in modo esaustivo a tutte le domande<sup>15</sup>.

#### 3.2.4 Il flusso dei quesiti sui viaggi nell'intervista iniziale

L'intervista iniziale è strutturata in più sezioni e, come già detto (cfr. par. 3.2.2), i quesiti sui turisti nell'anno si collocano alla fine della prima sezione, nella quale sono definite la composizione familiare e le caratteristiche principali dei membri della famiglia (Figura 3.3).

Figura 3.3 - Flusso dell'intervista iniziale



<sup>15</sup> Questi ultimi due casi non erano previsti nell'indagine CATI "Viaggi e vacanze".

Diversamente dalla rilevazione Viaggi e vacanze, l'indagine sulle spese prevede che la famiglia di fatto sia composta da tutti i membri che coabitano, includendo anche persone legate da soli vincoli economici (ad esempio personale di servizio convivente o affittuari di porzioni di abitazione), prima escluse dalla rilevazione CATI. Per questi componenti, sono riportate solo le principali caratteristiche demografiche, perché la loro inclusione nella famiglia di fatto è finalizzata alla corretta identificazione dell'unità economica di riferimento per la rilevazione della spesa per consumi e non alla stima dei flussi turistici. Pertanto, i quesiti sui viaggi e sulle escursioni non vengono loro somministrati.

Per tutti gli altri componenti della famiglia, il set d'informazioni socio-demografiche include l'età, il sesso, la cittadinanza, lo stato civile, il titolo di studio e la situazione lavorativa. In più, rispetto al passato, vi sono il paese di nascita e la professione svolta, per la cui codifica è utilizzato un motore di ricerca specifico che si basa sulla Nomenclatura e Classificazione delle Unità Professionali (CP2011<sup>16</sup>).

Come per la rilevazione CATI, l'età e il sesso, assieme al numero di componenti, guidano i percorsi e il *wording* di tutto il questionario. Per questo motivo sono previsti diversi quesiti di controllo al fine di accertarne la correttezza insieme alla famiglia.

I quesiti sui turisti nell'anno sono somministrati a chiusura della scheda generale, prima di passare alle sezioni successive, nelle quali sono raccolte informazioni a carattere familiare. Le domande, rivolte a tutti i membri della famiglia, sono uguali a quelle della CATI e rilevano le seguenti informazioni:

- se la persona abbia svolto almeno un viaggio di vacanza lunga (di durata pari o superiore a 4 notti) e quale sia stata la destinazione principale (Italia e/o estero);
- se la persona abbia svolto almeno un viaggio di vacanza breve (di durata da 1 a 3 notti) e quale sia stata la destinazione principale (Italia e/o estero);
- i motivi della non vacanza, nel caso non siano state effettuate vacanze nel periodo di riferimento (solo per i componenti di 15 anni o più).

Le altre sezioni che completano l'intervista iniziale riguardano le caratteristiche dell'abitazione in cui vive la famiglia e quelle di altre abitazioni di proprietà, informazioni sui mezzi di trasporto e di comunicazione<sup>17</sup>.

### 3.2.5 Il flusso dei quesiti sui viaggi nell'intervista finale

I quesiti sui viaggi e sulle escursioni nell'intervista finale sono articolati in più sezioni (Figura 3.4). Il *core* del questionario è rimasto invariato rispetto alla versione CATI, prevedendo anch'esso la compilazione di una "scheda riepilogativa" dei viaggi/escursioni effettuati da ciascun membro della famiglia nel periodo di riferimento, da cui dipende l'apertura delle "Schede viaggio" e/o delle "Schede escursioni". In ogni scheda sono descritte le caratteristiche principali dell'evento.

Rispetto al questionario CATI, tuttavia, è stato necessario apportare alcune modifiche, dovute alla diversa organizzazione dell'indagine e all'affiancamento alle tematiche sulle spese delle famiglie.

Nel passaggio dalle sezioni sulle spese alle sezioni sui viaggi ed escursioni il rilevatore elegge la persona di riferimento rispetto alla quale saranno personalizzate tutte le domande, anche quelle degli altri membri della famiglia (Box 2.1) :

<sup>16</sup> Per maggiori informazioni si veda <http://professioni.istat.it/sistemainformativoprofessionioni/cp2011/>.

<sup>17</sup> Per approfondimenti consultare Istat, 2015.

#### Box 2.1 - Script di passaggio dalle sezioni su Spese delle famiglie a viaggi

##### SCRIPT RV1

Passiamo ora a parlare dei viaggi effettuati dalla sua famiglia nel periodo compreso tra il “...” e il “...”.

RVb Indicare chi risponde alle domande sui viaggi della famiglia:

Poiché una stessa famiglia è intervistata più volte nel tempo (visita iniziale, visita intermedia e visita finale), può accadere che uno o più componenti registrati nella scheda generale durante l'intervista iniziale non siano più presenti in famiglia al momento dell'intervista finale per trasferimento definitivo o decesso, oppure che si sia aggiunto qualcuno

Figura 3.4 - Flusso dell'intervista finale

**SEZIONE A**  
I beni e i servizi autoprodotti

**SEZIONE B**  
Spese per l'abitazione in cui vive la famiglia

**SEZIONE C**  
Spese per altre abitazioni

**SEZIONE D**  
Spese effettuate negli ultimi 12 mesi

**SEZIONE E**  
Spese effettuate negli ultimi 3 mesi

**SEZIONE F**  
Spese effettuate nell'ultimo mese

**SEZIONE RV**  
Passaggio a viaggi e vacanze

**SEZIONE G**  
Scheda riepilogativa viaggi

**SEZIONE H**  
Scheda viaggio di lavoro

**SEZIONE I**  
Scheda viaggio personale

**SEZIONE J**  
Turisti trimestrali

**SEZIONE L**  
Scheda riepilogativa escursioni

**Sezione M**  
Scheda escursione

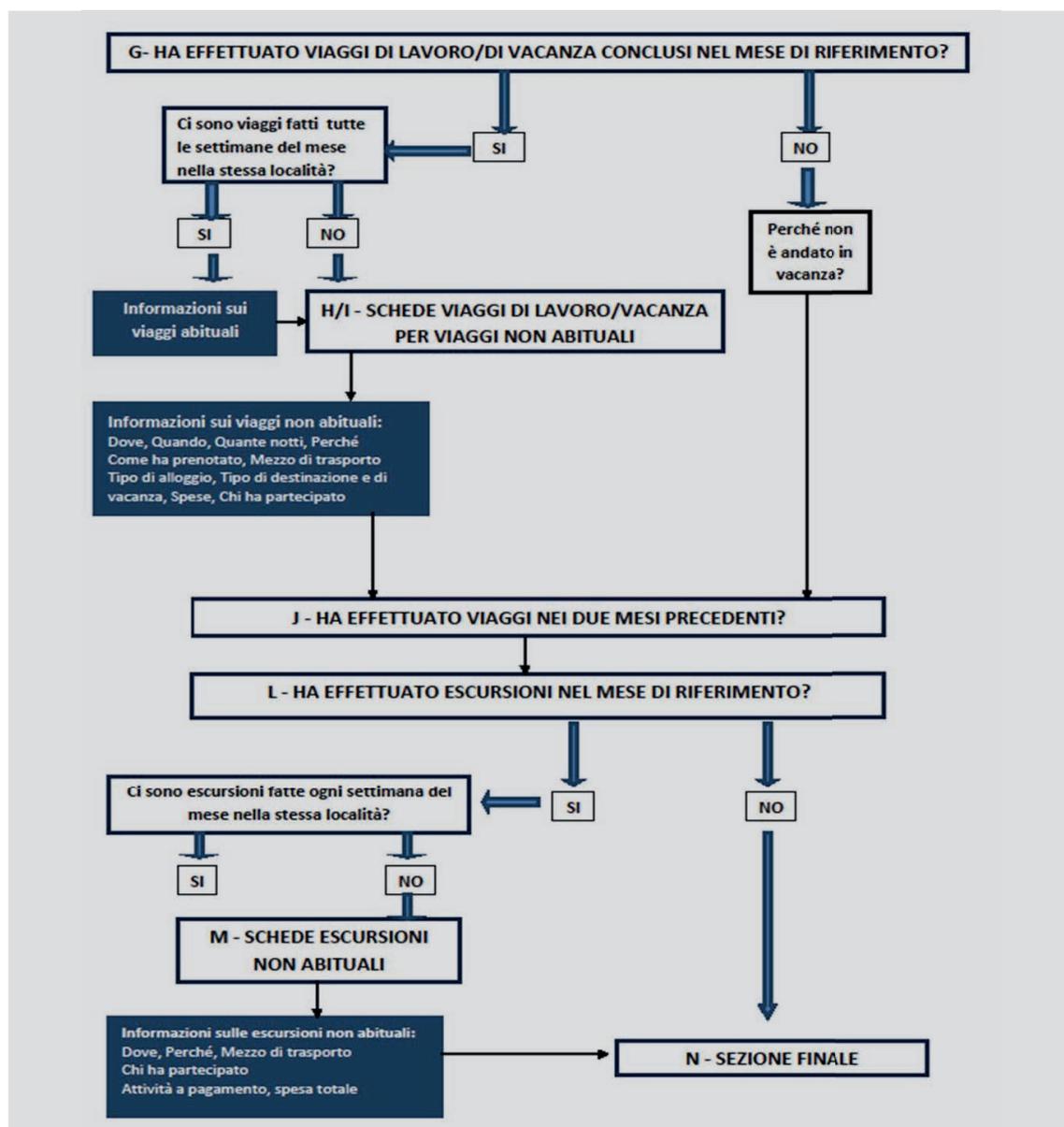
**SEZIONE N**  
Sezione finale

**SEZIONE O**  
Sezione chiusura intervista (*riservato al rilevatore*)

per matrimonio/convivenza o nascita. In questi casi, diversamente dalla rilevazione CATI che fotografava la famiglia di fatto una sola volta al momento dell'intervista, è necessario annotare tali cambiamenti. L'assenza dei componenti che non fanno più parte della famiglia è gestita direttamente dal software inibendo la somministrazione dei quesiti riferiti a tali componenti. Invece, gli eventuali membri della famiglia sopraggiunti successivamente all'intervista iniziale sono stati conteggiati durante una sperimentazione condotta fino al 2016, senza che venissero loro somministrati quesiti<sup>18</sup>. A fronte di una quota trascurabile di persone aggiunte al momento dell'intervista finale, nel 2016 è stato deciso di non inserire nuovi quesiti per intervistarle, allo scopo di evitare ulteriore *burden* sulle famiglie in assenza di un effettivo beneficio in termini di completezza delle interviste raccolte.

Dopo l'aggiornamento della famiglia di fatto, vengono somministrate le domande sui viaggi e sulle escursioni.

Figura 3.5 - Flow-chart delle sezioni del questionario su viaggi e vacanze per l'intervista finale



18 Dal 2017 questa informazione non è più rilevata.

Per ciascun membro della famiglia, come nella versione CATI, prima di entrare nel dettaglio della descrizione di ogni viaggio, è prevista la compilazione della “Scheda Ri-pilogativa viaggi” che riassume il numero di viaggi con pernottamento svolti per motivi di lavoro e per motivi personali, conclusi nel mese di riferimento. (Sezione G). Nel caso degli individui di età inferiore a 15 anni i quesiti sui viaggi effettuati per motivi di lavoro non sono proposti. In questa sezione sono registrati tutti i viaggi, sia quelli svolti abitualmente, sia quelli non abituali<sup>19</sup>. Il flusso del software CAPI guida il rilevatore nel percorso d'intervista tra i viaggi abituali e non abituali (Figura 3.5).

Per i viaggi abituali, di lavoro e personali, sono rilevate poche informazioni relative al complesso dei viaggi, così come descritto in Figura 3.6. Per ciascun viaggio non abituale, invece, si apre una “Scheda viaggio” per descrivere dettagliatamente tutte le principali caratteristiche del viaggio, prima per quelli di lavoro (sezione H), poi per quelli personali (sezione I), con uguali quesiti per entrambe le tipologie. L'unica eccezione è la domanda sulla motivazione del viaggio, che è necessariamente diversa.

Figura 3.6 - Informazioni rilevate per i viaggi abituali di lavoro personali



Le informazioni rilevate per ciascun viaggio non abituale sono le seguenti<sup>20</sup>:

- **destinazione principale** (Comune, Provincia, Regione italiana o Stato estero), definita in base al maggior numero di notti ivi trascorso;
- **data di inizio del viaggio**;
- **durata** (in notti);
- **motivo** in assenza del quale il viaggio non avrebbe avuto luogo;
- **tipologia del luogo visitato** (solo per le vacanze);
- **organizzazione**, intesa come presenza o assenza di prenotazione dell'alloggio e/o del trasporto e, in caso di risposta affermativa, quale tipologia di prenotazione è stata scelta, diretta o in agenzia;
- **principale mezzo di trasporto utilizzato**, corrispondente al mezzo con cui si è co-

<sup>19</sup> Per la definizione di ambiente abituale si veda Eurostat, 2014.

<sup>20</sup> Per maggiori informazioni si rimanda al Sistema di Qualità delle indagini dell'Istituto <http://siqual.istat.it/SIQual/>.

- perta la maggiore distanza;
- **principale tipo di alloggio**, individuato dal maggior numero di notti trascorse in quella tipologia;
  - presenza di **destinazioni secondarie in Italia** (al massimo 3) con dettagli su destinazione e durata;
  - presenza di **destinazioni secondarie all'estero**, senza dettagli su destinazione e durata;
  - **condivisione di uno stesso viaggio con uno o più componenti della famiglia**, nel caso in cui tutte le caratteristiche del viaggio siano uguali per ogni partecipante. Per ognuno dei partecipanti (denominati "aggregati") allo stesso viaggio di uno dei membri della famiglia (denominato "titolare") non ci sarà bisogno di somministrare nuovamente tutti i quesiti, poiché il software replica le informazioni già inserite creando delle schede viaggio precompilate. L'unica informazione che potrebbe essere diversa tra i partecipanti al viaggio è il motivo del viaggio, che viene quindi chiesto individualmente;
  - **spese del viaggio**: complessive, nonché distinte tra spesa per il trasporto, per l'alloggio e per altro, come pasti, souvenir, acquisti vari considerati nel loro insieme. Tra le spese sono rilevate a parte, se presenti, quelle per agenzia o per un alloggio in affitto. Infine, è rilevata anche la spesa per l'acquisto di beni durevoli.

Riguardo le spese del viaggio, è importante sottolineare che la spesa può essere sostenuta sia dalla famiglia che viaggia, sia da persone che non fanno parte della famiglia intervistata e che possono non aver viaggiato con la famiglia stessa. Questo è il caso delle vacanze ricevute in regalo, o dei viaggi d'affari pagati dal datore di lavoro. In questi casi, si chiede alle famiglie che hanno viaggiato di provare a ricostruire la spesa del viaggio pur non avendola sostenuta, cercando di fornire una stima del valore reale pagato da altri. Nel caso di alcune tipologie di viaggi, infatti, tale valore è noto e può essere registrato.

Terminata la descrizione dei viaggi non abituali, nella versione CAPI è stata aggiunta una nuova sezione (Sezione J) che è funzionale alla costruzione della stima del numero di turisti trimestrali per vacanza lunga, per vacanza breve e per lavoro. Infatti, sia la periodicità della rilevazione, sia il diverso periodo di riferimento dei viaggi (entrambi mensili e non più trimestrali), non consentono di calcolare il numero di turisti nel trimestre come somma dei turisti mensili<sup>21</sup>. Per ottenerlo, così come tradizionalmente diffuso dall'indagine CATI, è stato necessario aggiungere dei quesiti retrospettivi a due e tre mesi di distanza rispetto al mese d'intervista (Sezione J). Poiché ogni mese le risposte sui viaggi conclusi nei vari periodi di riferimento (individuati con le schede H/I e con la sezione J) sono fornite da un terzo del campione, è necessario ripetere l'intervista per tre mesi successivi per avere un campione completo idoneo a calcolare la stima dei turisti di uno specifico mese. In altre parole, ogni mese è stimato con le risposte provenienti da tre sotto campioni mensili diversi, intervistati in tre mesi successivi.

Per esempio, per costruire le stime dei turisti trimestrali del trimestre estivo (luglio-agosto-settembre), occorrono le informazioni delle interviste finali di agosto e settembre dalle quali si ricavano i turisti nei mesi di luglio e agosto (a partire dai viaggi riportati per quei mesi) e le informazioni delle interviste finali di ottobre, novembre e dicembre che rilevano, tramite i quesiti retrospettivi della sezione J, i turisti nei mesi di luglio, agosto e settembre (Figura 3.7). Per esempio, il sottocampione che risponde all'intervista finale a

<sup>21</sup> Per poter ottenere i turisti trimestrali come somma di turisti mensili occorrerebbero uno schema di campionamento e un sistema di ponderazione mensili.

### 3. La nuova rilevazione delle informazioni sulla domanda turistica

ottobre, riporta i viaggi conclusi a settembre nelle Sezioni H e I, da cui si può contare una parte di turisti nel mese di settembre, e sui viaggi conclusi a luglio e agosto nella sezione J, da cui deriva un'altra parte dei turisti per ciascuno dei quei due mesi. Ovviamente, chi nell'intervista finale a ottobre dichiara di aver viaggiato sia settembre sia a luglio e/o agosto, verrà conteggiato una sola volta come turista per il trimestre luglio-settembre. Ripetendo gli stessi quesiti anche per le interviste finali di novembre e dicembre e sommando i risultati ottenuti nelle tre occasioni d'intervista, è possibile ricostruire i turisti trimestrali.

In più, la nuova indagine CAPI, rispetto alla CATI, consente di stimare il numero medio di turisti mensili in ciascun trimestre.

Figura 3.7 - Costruzione dei turisti trimestrali del periodo Luglio-Settembre

MAG	GIU	LUG	AGO	SET	OTT	NOV	DIC
Turisti del mese-2 (sez.J) Int.Fin AGO	Turisti del mese-1 (sez.J) Int.Fin AGO	Turisti del mese Int.Fin AGO	Int.Fin.				
	Turisti del mese-2 (Sez.J) Int.Fin. SET	Turisti del mese-1 (Sez.J) Int.Fin. SET	Turisti del mese (sez.H/I) Int.Fin. SET	Int.Fin.			
		Turisti del mese-2 (Sez.J) Int.Fin. OTT	Turisti del mese-1 (Sez.J) Int.Fin. OTT	Turisti del mese (sez.H/I) Int.Fin.OTT	Int.Fin.		
			Turisti del mese-2 (Sez.J) Int.Fin.NOV	Turisti del mese-1 (Sez.J) Int.Fin.NOV	Turisti del mese (sez.H/I) Int.Fin NOV	Int.Fin.	
				Turisti del mese-2 (Sez.J) Int.Fin. DIC	Turisti del mese-1 (sez.J) Int.Fin DIC	Turisti del mese (sez.H/I) Int.Fin. DIC	Int.Fin.

Conclusa la Sezione J sui turisti trimestrali, il focus viaggi prevede due sezioni sulle visite in giornata (escursioni) effettuate nel mese di riferimento da ogni componente della famiglia, per motivi di lavoro e personali. Le escursioni per motivi di lavoro, come per i viaggi, interessano solo le persone di età pari a 15 anni o più. La definizione di escursione (cfr. par.1.2), analoga a quella di viaggio, comporta la distinzione tra quelle abituali, soltanto conteggiate, e quelle non abituali, per le quali sono acquisite le principali caratteristiche.

Al pari dei viaggi, anche per le escursioni si somministra prima una "scheda riepilogativa" (Scheda L) e poi, per ogni escursione non abituale dichiarata, si apre una "scheda escursione" (Scheda M) dove descriverne le principali caratteristiche:

- la destinazione (Comune, Provincia, Regione italiana o Stato estero);
- il motivo principale;
- l'eventuale partecipazione di altri componenti della famiglia;
- le attività svolte durante l'escursione che abbiano comportato delle spese (utilizzo mezzi di trasporto, consumo cibi e bevande, acquisti vari, biglietti di ingresso per cinema, musei, parchi di divertimento, ecc.);
- la spesa complessiva sostenuta per l'escursione, considerando tutti i familiari par-

tecipanti.

Anche per le escursioni, nel caso di più familiari partecipanti non c'è bisogno di descrivere la stessa visita con una nuova scheda, perché questa verrà replicata in automatico dal software. L'unica informazione somministrata individualmente e non riprodotta dal software è quella sul motivo, che potrebbe essere diverso per ogni partecipante.

Il questionario si chiude con la Sezione N, che include dei quesiti familiari sul reddito puntuale, sulla valutazione soggettiva della condizione economica della famiglia e su come questa sia cambiata negli ultimi 12 mesi.

### 3.3 La rete di rilevazione e il monitoraggio dell'indagine

Caratteristica comune alla vecchia e alla nuova indagine è il ricorso a una rete di rilevatori professionisti facente capo a una società esterna, cui competono, per questo aspetto, le attività di selezione e reclutamento, l'assegnazione dei carichi di lavoro individuali e la gestione del rapporto di lavoro. Tuttavia, diversa è l'organizzazione della rete poiché differenti sono le modalità di raccolta dei dati tra tecnica d'indagine CATI e CAPI. Nell'indagine telefonica Viaggi e Vacanze, la rete era centralizzata e il lavoro degli intervistatori si svolgeva in un'unica sede. Ciò garantiva un monitoraggio simultaneo dell'operato dei rilevatori da parte dei supervisor, che erano in grado di fornire supporto direttamente nel corso dello svolgimento delle interviste. La centralizzazione della rete consentiva anche di schedulare, ogni anno, più interventi formativi in aula, realizzando briefing teorici e pratici all'inizio di ogni periodo di rilevazione e, laddove necessario, anche *debriefing* nel corso di ogni trimestre di indagine (Istat, 2003).

Nell'indagine CAPI sulle Spese delle Famiglie, invece, la rete degli intervistatori è dislocata su tutto il territorio nazionale e impegnata in una rilevazione continua, che richiede una organizzazione costante e capillare del lavoro sul campo. In questo nuovo assetto, i comuni hanno assunto un ruolo fondamentale di affiancamento alla rete, di supporto e di facilitazione nella ricerca delle famiglie.

Nella nuova cornice organizzativa dell'indagine sulle Spese delle Famiglie, la distribuzione sul territorio della rete dei rilevatori, il passaggio alla tecnica CAPI, l'articolazione in più fasi della raccolta dei dati e l'integrazione delle rilevazioni sulla domanda turistica e sulle spese per i consumi hanno richiesto la rimodulazione dell'offerta formativa per i rilevatori e la riprogettazione del sistema di monitoraggio.

Le pagine successive sono dedicate all'approfondimento delle novità introdotte in tali ambiti.

#### 3.3.1 La formazione dei rilevatori

La qualità del dato grezzo raccolto dipende fortemente dalla capacità del singolo intervistatore di padroneggiare i contenuti dell'indagine, gli aspetti giuridici e metodologici, gli strumenti informatici, nonché dalla sua abilità nel governare il processo comunicativo per la "conquista della famiglia", contenendone i rifiuti e le interruzioni. Da questo punto di vista, nelle indagini sulle famiglie, la formazione degli intervistatori rappresenta, da sempre, un grosso investimento in termini di tempo ed energie per tutti gli attori coinvolti (ISTAT, la società esterna che gestisce la rete dei rilevatori sul territorio e i rilevatori stessi). Infatti, ol-

tre ad essere uno strumento di prevenzione volto a minimizzare i potenziali errori imputabili agli intervistatori, le sessioni formative costituiscono un'insostituibile occasione per creare quel clima di fiducia e di collaborazione necessario a sviluppare un senso di partecipazione a un progetto unico e condiviso.

La formazione specifica del Focus “Viaggi e Vacanze” è parte integrante del processo formativo dell'Indagine sulle Spese delle Famiglie e, pertanto, ne condivide obiettivi, metodologie, aspetti organizzativi e strumenti. La proposta formativa per l'indagine sulle Spese delle Famiglie si è concretizzata in un'offerta unica centralizzata, con sessioni in aula svolte per tutti gli intervistatori, corredata da ulteriori strumenti e attività di supporto e di assistenza (manuale, *debriefing*, formazione a distanza, test di apprendimento, ecc.), di cui si parlerà più avanti. In questo paragrafo non ci si soffermerà sull'organizzazione, i programmi, le attività e i “numeri” della formazione della nuova indagine CAPI sulle Spese delle Famiglie<sup>22</sup>, bensì sugli elementi di continuità e di novità rispetto alla formazione dei rilevatori precedentemente svolta per l'indagine CATI “Viaggi e Vacanze”.

L'esperienza maturata nel periodo di conduzione della CATI ha rappresentato il punto di partenza per la riformulazione della proposta formativa sulla domanda turistica, tenendo conto della nuova veste di “Viaggi e Vacanze” come focus *embedded* nell'indagine sulle Spese delle Famiglie. L'intervento formativo tematico si è tradotto, pertanto, nella progettazione di un modulo specifico, somministrato durante l'ultima giornata di formazione. Tale modulo si inserisce in un programma che ripercorre in sequenza temporale le diverse fasi e operazioni dell'intera rilevazione, che di fatto costituisce un unico processo di raccolta dei dati. Il programma formativo, pertanto, include sia moduli tematici sulle spese per i consumi e sul turismo, sia moduli “trasversali” cioè quello sulla gestione dei contatti con le famiglie (Scheda contatti) e quello sulla raccolta delle informazioni socio-anagrafiche di ciascun membro della famiglia (Scheda Generale). Completano la proposta formativa il modulo tecnico sugli strumenti informatici e un modulo sulla comunicazione e il contenimento dei rifiuti. L'attenzione alle strategie da attivare per limitare i rifiuti e le mancate risposte totali e parziali era già fortemente presente durante le sessioni formative dei rilevatori dell'indagine CATI. Tuttavia, la complessità e l'articolazione dell'indagine sulle Spese delle Famiglie e la necessità di potenziare le capacità comunicative dei rilevatori impegnati in interviste faccia a faccia, hanno reso necessario il disegno di un vero e proprio modulo ad hoc sulle tecniche di somministrazione e sulle strategie comunicative, basato su un opportuno bilanciamento tra richiami teorici ed esercitazioni condotte sotto forma di *role playing*.

Per quel che riguarda il metodo formativo, la modalità detta a “doppio binario” tra teoria e pratica, introdotta per la nuova Indagine sulle Spese (Istat, 2015), era già implementata nella formazione dei rilevatori dell'indagine CATI, in cui l'alternanza tra descrizione teorica di definizioni, concetti e classificazioni da una parte e sessioni pratiche di esercitazione basate su simulazioni di interviste complete, dall'altra, era stata applicata con successo.

Una novità rispetto alla formazione della rete CATI, invece, ha riguardato l'utilizzo di schede di autovalutazione dell'apprendimento, proposte a conclusione del corso in aula come strumento per consolidare le conoscenze acquisite durante le giornate formative e per tarare meglio i successivi interventi formativi.

Dal punto di vista dei contenuti, il materiale formativo già utilizzato per l'indagine CATI ha sicuramente rappresentato un ottimo punto di partenza per strutturare l'intervento formativo sul focus “Viaggi e Vacanze”. Tuttavia, esso è stato ampiamente rivisto, eliminando

<sup>22</sup> Per approfondimenti si rimanda al paragrafo 2.3.3.2 del volume Istat, 2015.

le ridondanze su tematiche affrontate in altri moduli e focalizzando le problematiche legate al passaggio alle sezioni sui viaggi e le vacanze nell'intervista finale. Oltre a introdurre i principali concetti e definizioni sul turismo e a illustrare la struttura e i contenuti dei relativi quesiti, il modulo formativo sui viaggi pone l'accento su come cambia la prospettiva di rilevazione delle informazioni in questa parte dell'intervista, che da familiare diventa individuale. Nello specifico, l'attenzione è rivolta alle spese, a come si rilevano nel Focus e alle principali differenze rispetto ai quesiti sulle spese per consumi delle sezioni precedenti. La simulazione di interviste complete, iniziata durante i precedenti moduli tematici, si conclude durante l'ultima giornata formativa incentrata sul Focus viaggi e certamente aiuta a mettere in luce gli snodi significativi della rilevazione.

Rispetto al passato, quando le attività di training erano rivolte a intervistatori telefonici impegnati in una indagine *stand alone*, il programma formativo progettato per l'indagine sulle Spese delle Famiglie è stato caratterizzato da una notevole compressione dei tempi dedicati alla formazione in aula specifica sulle sezioni dei viaggi e delle escursioni. Tuttavia, nuovi strumenti a supporto dei rilevatori sono stati implementati a complemento della formazione in aula. Tra essi, va citato il sistema degli "*help in linea*", inserito nel questionario elettronico CAPI, che consente al rilevatore di accedere direttamente, in fase di intervista, a spiegazioni veloci ed efficaci per chiarire il significato dei quesiti più complessi e/o delle modalità di risposta previste, con evidenti ricadute positive sull'andamento di tutta l'intervista. Inoltre, come in passato, i rilevatori hanno a disposizione il manuale ISTAT "Istruzioni per gli intervistatori", che è il principale strumento di consultazione a supporto dell'attività sul campo. In esso, infatti, sono maggiormente approfonditi i concetti espressi durante la formazione, sono sottolineati gli aspetti più critici della rilevazione e, infine, si trovano dettagli e chiarimenti su definizioni e classificazioni, per permettere al rilevatore di fronteggiare ogni situazione, dalla più semplice alla più complessa o anomala che si possa presentare in corso di intervista.

Come nell'indagine CATI, anche nella nuova indagine CAPI, i *debriefing* costituiscono ulteriori momenti per proporre richiami formativi, ma anche per monitorare e valutare l'impianto di indagine con tutti gli attori coinvolti nel processo. Si tratta degli incontri periodici con gli intervistatori, svolti di norma a metà e alla fine di ogni anno di rilevazione. Essi sono, in primo luogo, occasioni di confronto che consentono ai rilevatori di approfondire la conoscenza dei processi di indagine, con interventi di *refreshing* su particolari concetti o definizioni eventualmente non compresi appieno in sede di formazione iniziale. Inoltre, durante il *debriefing* si comunicano e si spiegano ai rilevatori le novità eventualmente introdotte nel questionario e/o nel processo di indagine. Infine, il *debriefing* rappresenta un momento di riflessione per ricevere dalla rete un feedback approfondito sull'andamento della rilevazione, per cogliere impressioni, problemi e criticità in merito al funzionamento del questionario elettronico, alla comprensibilità dei quesiti e alla disponibilità delle famiglie. Il *debriefing*, in definitiva, costituisce l'occasione in cui tutti gli attori coinvolti nella rilevazione (ISTAT, società esterna e rete di rilevatori) interagiscono in modo costruttivo per rimuovere qualunque elemento di disturbo al buon esito dell'indagine, arginando distorsioni che possono ripercuotersi sulla qualità del dato finale<sup>23</sup>.

23 I *debriefing* si possono svolgere sia attraverso incontri in aula, sia attraverso la realizzazione di attività formative e di valutazione dell'indagine "a distanza". Nel primo caso, viene favorito il confronto diretto e approfondito con la rete; nel secondo caso, ai rilevatori viene offerta la possibilità di una riflessione autonoma e meditata sull'andamento della rilevazione.

### 3. La nuova rilevazione delle informazioni sulla domanda turistica

Un'ulteriore novità rispetto al passato, è rappresentata dall'accesso alla piattaforma ISTAT per la formazione statistica<sup>24</sup>, dove i rilevatori possono: 1) reperire tutti i materiali, come ad esempio le presentazioni e i questionari di valutazione, utilizzati durante la formazione in aula e i *debriefing*; 2) effettuare test di apprendimento; 3) scaricare materiali utili per la conduzione dell'intervista (la guida per gli intervistatori, i questionari, le lettere informative per le famiglie, le circolari per i comuni); 4) acquisire informazioni sui principali utilizzi dei dati rilevati.

La piattaforma è anche il luogo virtuale in cui viene erogata la Formazione a distanza (FaD), con il coinvolgimento degli Uffici Territoriali dell'ISTAT per far fronte al fisiologico turnover degli intervistatori.

**Prospetto 3.3 - Strumenti formativi utilizzati nell'indagine CATI Viaggi e Vacanze e nell'indagine CAPI sulle Spese delle Famiglie**

STRUMENTI FORMATIVI	INDAGINE CATI	INDAGINE CAPI
Formazione centralizzata in aula	x	x
FaD (Formazione a Distanza)		x
Questionario elettronico – Help in linea		x
Autovalutazione		x
Manuale intervistatori	x	x
Debriefing in aula	x	x
Debriefing a distanza		x
Piattaforma e-learning		x

#### 3.3.2 Gli indicatori di performance della rete di rilevazione

Il monitoraggio della qualità di un'indagine, come noto (Istat, 2007), è un'attività fondamentale per controllarne l'andamento, allo scopo di prevenire gli errori non campionari.

Il perseguimento della qualità è sotteso a tutte le attività del processo e si realizza in diversi momenti e con diversi strumenti. In questo paragrafo si vuole focalizzare l'attenzione sul monitoraggio durante la fase di rilevazione, aspetto centrale soprattutto per le indagini continue, come l'indagine sulle Spese delle famiglie. Tale attività si realizza mediante il controllo di una serie di indicatori dai quali si può avere l'immediata percezione del lavoro dei rilevatori sul campo. Questi indicatori (cosiddetti "diretti") sono generati dalle attività che i rilevatori svolgono ogni giorno, registrate dagli stessi rilevatori tramite il software a loro disposizione (la Scheda Contatti), che produce un archivio incrementale degli esiti dei contatti sulle famiglie del campione.

Anche l'indagine CATI sulla domanda turistica disponeva di un sistema organizzato di controllo basato principalmente sul monitoraggio quotidiano della rilevazione (Istat, 2003).

L'esperienza maturata nel periodo di conduzione della CATI e nell'ambito di altre indagini sulle famiglie presenti in Istituto ha costituito il punto di partenza per la progettazione

<sup>24</sup> La piattaforma e-learning accessibile all'indirizzo <https://formazione.istat.it> è un ambiente di apprendimento articolato destinato al personale ISTAT, agli operatori del Sistema statistico nazionale e delle reti di rilevazione e agli studenti delle scuole primarie, secondarie e dell'Università. La piattaforma consente non solo di accedere a materiali relativi ai corsi d'aula, ad attività formative in modalità e-learning, video, tutorial e *webinar*, ma anche di utilizzare strumenti di apprendimento cooperativo come chat, forum ed esercitazioni on line.

di un analogo sistema di controllo della qualità nell'indagine sulle Spese delle famiglie, tenendo conto delle differenze rispetto alla CATI. Tali differenze, dovute sia al diverso grado di coinvolgimento dei rilevatori nelle operazioni di pianificazione del lavoro sul campo tra le due tecniche di rilevazione, sia all'organizzazione complessiva dell'attività di rilevazione per le due indagini, si riflettono sui rispettivi set di indicatori per il monitoraggio.

E' evidente che il ricorso a una società esterna incaricata di effettuare le interviste per conto dell'Istat, caratteristica comune alla vecchia e alla nuova indagine, renda indispensabile il controllo dell'attività sul campo, affinché anche i soggetti esterni all'Istituto garantiscano gli standard di qualità che ci si attende. Tuttavia, le diverse modalità di raccolta dei dati tra CATI e CAPI hanno conseguenze dirette sul tipo di monitoraggio che è possibile o necessario fare. Come mostrato nel Prospetto 3.4, un rilevatore CAPI, rispetto ad un rilevatore CATI, ha una responsabilità individuale maggiore della gestione del proprio lavoro, sotto diversi aspetti.

**Prospetto 3.4 - Attività del processo di raccolta dati con tecnica CATI e CAPI**

ATTIVITÀ	INDAGINE	
	CATI (Viaggi e vacanze)	CAPI (Indagine sulle spese)
Somministrazione dell'intervista	Unica sede centralizzata (che rende possibile anche un monitoraggio di sala)	Sul territorio
Assegnazione delle famiglie da intervistare	Automatica, effettuata dal software di gestione delle telefonate	Preassegnazione al rilevatore che opera nella provincia. Il rilevatore procede alla personale pianificazione del proprio lavoro e a contattare i nominativi assegnatigli tramite l'utilizzo dell'Agenda.
Registrazione degli esiti dei tentativi di contatto	In parte automatica (effettuata dal software di gestione delle telefonate) e in parte manuale effettuata dal rilevatore tramite appositi quesiti del questionario CATI	Manuale, effettuata dal rilevatore tramite la scheda contatti CAPI
Modalità di effettuazione del primo contatto con la famiglia	Al telefono	Al telefono (quando è disponibile almeno un numero) o a domicilio
Rapporto con i Comuni	Non necessario	A volte indispensabile

Pertanto, la valutazione della performance dei rilevatori CAPI si deve basare su indicatori differenti rispetto a quelli della CATI. Nello specifico, nel caso dell'indagine CAPI, è possibile calcolare indicatori che si basano sull'ammontare delle famiglie assegnate ad ogni rilevatore e misurarne la produttività (per esempio, il tasso di completezza e il tasso di pigrizia), mentre in un'indagine CATI tale misurazione è ottenibile solo valutando gli scostamenti rispetto alla media giornaliera realizzata da tutti i rilevatori attivi in sala, poiché non è prevista un'assegnazione iniziale delle interviste che ogni rilevatore deve svolgere.

Inoltre, l'organizzazione in tre fasi dell'indagine sulle Spese delle famiglie rende necessario monitorare ogni fase di rilevazione con un adeguato set di indicatori. Tra essi, alcuni si ripetono in ciascuna delle tre fasi, altri sono specifici di una fase. Nell'indagine CATI Viaggi e vacanze, ad esempio, l'intervista si considerava completa se l'unico questionario veniva somministrato nella sua interezza, mentre l'indagine sulle Spese delle famiglie prevede il completamento dell'intervista iniziale, della visita intermedia, dell'intervista finale e la compilazione del diario delle spese.

A scopo illustrativo, nel Prospetto 3.5 si riportano le modalità di calcolo dei principali indicatori di performance della rete nell'indagine CAPI e la eventuale corrispondenza con

### 3. La nuova rilevazione delle informazioni sulla domanda turistica

**Prospetto 3.5 - Principali indicatori di performance della rete CATI e CAPI**

INDICATORI	DESCRIZIONE	
	CATI	CAPI
Tasso di Assegnazione	Non calcolato	(Famiglie assegnate/ famiglie teoriche)*100
Tasso di Pigrizia	Non calcolato	(Sestine non toccate/ Sestine assegnate)*100
Tasso di Realizzazione	% interviste complete sul totale delle toccate (Fuori Target + Complete + Interrotte definitivamente + Rifiuti + Sospese con almeno un tentativo di contatto)	(Sestine complete/ Sestine contattate, con almeno un contatto)*100
Tasso di completezza 1	Numero medio di interviste complete quotidiane	(Famiglie con intervista iniziale completa/ Famiglie/ Sestine assegnate)*100
Tasso di completezza 2		(Famiglie con intervista intermedia completa/ Famiglie con intervista iniziale completa)*100
Tasso di completezza 3		(Famiglie con intervista finale completa/ Famiglie con intervista intermedia completa)*100
% Quartine bucate 1	(Quartine bucate/Quartine teoriche )*100	(Sestine bucate in Fase 1/ Sestine toccate)*100
% Quartine bucate 2		(Sestine bucate in Fase 2/ Sestine toccate)*100
% Quartine bucate 3		(Sestine bucate in Fase 3/ Sestine toccate)*100
Tasso di fedeltà delle famiglie base	(Interviste complete delle famiglie sostituite/interviste complete totali)*100	(Sestine complete BASE/ Sestine complete TUTTE)*100

l'indagine CATI<sup>25</sup>.

Gli indicatori di performance della rete di rilevazione e gli indicatori sugli esiti familiari vengono forniti giornalmente ai referenti Istat tramite un portale web con accesso dedicato e sono elaborati per ciascun ciclo mensile di rilevazione, nonché a livello di trimestre e di anno.

#### 3.3.3 Il sistema degli esiti familiari

Tutti gli indicatori di monitoraggio, sia quelli di performance della rete<sup>27</sup>, sia quelli sui motivi di caduta delle famiglie, vengono costruiti a partire dagli esiti finali, che distinguono le famiglie con intervista completa da quelle cadute per i diversi motivi. Gli esiti definitivi familiari, in entrambe le indagini, vengono generati al momento della chiusura dei contatti con la famiglia campione e vengono determinati dalle risposte, codificate dai rilevatori durante la gestione del contatto, e dalle regole connesse, rispettivamente, al questionario per intervista CATI e al questionario di Scheda contatti CAPI. Infatti, in genere in un'indagine telefonica la gestione del contatto è semplificata e mirata ad ottenere nel più breve tempo possibile informazioni circa l'eleggibilità e la reperibilità della famiglia. Pertanto, i quesiti che svolgono tale funzione sono stati collocati direttamente nel questionario elettronico per intervista; inoltre, viene sfruttata la potenzialità del sistema CATI, che consente di salvare in automatico le principali caratteristiche di tutti i tentativi di contatto effettuati (data, ora, numero chiamato e sua validità, esito del tentativo, eccetera). Al contrario, in un'indagine basata su modalità faccia-a-faccia e dal disegno complesso, come quella sulle spese, è necessario standardizzare l'enorme varietà delle casistiche di contatto (telefonico o a domicilio, primo contatto per intervista iniziale o contatti per fasi successive) e creare uno

<sup>25</sup> Per una disamina degli ulteriori indicatori tipici dell'indagine "Viaggi e vacanze" si rimanda a Istat, 2003.

<sup>26</sup> Il tasso di fedeltà è il complemento del tasso di sostituzione, utilizzato per il monitoraggio delle interviste CATI.

<sup>27</sup> Con l'unica eccezione del tasso di assegnazione.

strumento in grado di supportare il rilevatore nella gestione del contatto attraverso percorsi guidati. Per questo si è reso necessario un vero e proprio questionario, la Scheda Contatti, che consente il puntuale riconoscimento dell'eleggibilità della famiglia, l'applicazione delle regole per la sostituzione della famiglia in caso di caduta, la registrazione degli esiti (provvisori e definitivi) e degli stati di lavorazione (da contattare, tentativi in corso, contatto concluso), nonché l'accesso al questionario per intervista CAPI e al modulo di acquisizione del Diario, secondo la specifica fase di rilevazione. Tuttavia, il sistema CAPI, nei tratti appena delineati, non garantisce il salvataggio automatico di tutti i tentativi di contatto compiuti dai rilevatori, ai quali viene chiesto di effettuare le operazioni di registrazione anche dei tentativi con esito non positivo, al fine di stimare correttamente il loro impegno.

Tali premesse spiegano almeno in parte la minore articolazione delle modalità degli esiti definitivi dell'indagine Viaggi e vacanze CATI, rispetto a quelli dell'indagine sulle spese CAPI.

Nel prospetto 3.6 si riportano gli esiti (e i rispettivi codici numerici in uso) delle due indagini a confronto; poiché non è possibile una perfetta corrispondenza tra gli esiti analitici, il confronto viene proposto in base alle categorie di esiti sintetici dell'indagine CATI.

Nell'indagine CATI Viaggi e vacanze, l'eventualità di un'interruzione si presenta con caratteristiche classiche, ovvero durante l'intervista telefonica (o durante uno degli appuntamenti successivi per somministrare l'intervista telefonica). Nell'indagine sulle Spese, invece, l'interruzione può avvenire durante la somministrazione dell'intervista (iniziale o finale), ma anche tra una fase e un'altra della rilevazione (ad esempio, se la famiglia ha rilasciato l'intervista iniziale e ricevuto il diario da compilare e rifiuta di procedere con la visita intermedia).

Un aspetto molto importante da tenere sotto controllo, nell'indagine telefonica, sono gli appuntamenti. Tendenzialmente, infatti, il contatto al telefono può indurre più facilmente al rifiuto (e all'interruzione), che a volte può essere mascherato dalla richiesta, da parte della famiglia, di essere richiamata in un altro momento. Inoltre, vi sono dei casi in cui chi risponde alla chiamata non concede all'intervistatore neppure il tempo di fornire spiegazioni (riattacca immediatamente). In questi casi, i rilevatori sono istruiti a fissare degli appuntamenti fittizi, per tentare di contattare la famiglia in un altro momento, auspicando una maggiore disponibilità al tentativo successivo. Per questo, tra gli esiti analitici dell'indagine CATI, alcune modalità di esiti di rifiuto e di contatore esaurito evidenziano se la caduta si è verificata in relazione ad appuntamenti reali o fittizi.

Dal confronto tra esiti analitici relativi alla categoria dei Fuori target, si conferma la maggiore capacità della tecnica faccia-a-faccia di individuare e accertare le diverse tipologie di unità non eleggibili. Inoltre, è ben visibile anche l'effetto del tipo di campionamento, tale per cui i numeri di telefono inesistenti o errati vengono considerati, nell'ambito dell'indagine CATI, come fuori target (e non irreperibili) in quanto la lista di estrazione non è la Lista Anagrafica Comunale, bensì la lista degli abbonati alla telefonia fissa. Parallelamente, i motivi di irreperibilità nell'indagine CAPI evidenziano la bontà della lista di estrazione (in che misura, cioè, il nominativo o l'indirizzo della famiglia è corretto); all'interno di questa categoria, l'assenza prolungata costituisce per la CAPI un motivo di caduta con dignità propria, in quanto viene assegnato se la famiglia risulta assente anche solo in uno dei tre intervalli temporali prefissati.

Infine, nell'indagine CAPI una specifica categoria di cadute da monitorare è quella denominata "Fine periodo", in quanto può evidenziare un problema di carichi di lavoro inadeguati. Tra queste famiglie vi sono quelle per le quali non sono stati fatti tentativi di contatto

### 3. La nuova rilevazione delle informazioni sulla domanda turistica

**Prospetto 3.6 - Tabella di corrispondenza degli esiti di caduta Viaggi e vacanze CATI e Spese CAPI**

MOTIVI DI CADUTA CATI		MOTIVI DI CADUTA CAPI
Esiti sintetici CATI	Esiti analitici CATI	Esiti analitici CAPI
Interruzioni definitive	Interruzione definitiva (03)	Interruzione definitiva (002) Intervista iniziale completa (101) Visita intermedia completa (201)
Rifiuti	Rifiuto prima quesito ricezione lettera (04) Rifiuto dopo quesito ricezione lettera (05) Rifiuto prima del quesito sulla ricezione lettera conseguente al terzo appuntamento fittizio (29) Rifiuto dopo il quesito sulla ricezione lettera conseguente al terzo appuntamento fittizio (30)	Rifiuto pre-individuazione PR (005) Rifiuto post-individuazione PR (006) Malato grave (023)
Caduta per contatore esaurito con contatto	Caduta per contatore esaurito per non risponde/fax/segreteria/appuntamento reale con contatto (09) Caduta per contatore esaurito per appuntamenti (10)	Contatore esaurito per eccesso di tentativi con almeno un contatto per 1° appuntamento/intervista iniziale (011) Contatore esaurito per eccesso di tentativi dopo 1° appuntamento/inizio intervista iniziale (013) Contatore esaurito per eccesso di tentativi dopo 1° appuntamento/intervista iniziale (escluso caso precedente) (014)
Caduta per contatore esaurito senza contatto	Caduta per contatore esaurito per non risponde/fax/segreteria senza contatto (08)	Contatore esaurito per eccesso di altri tentativi senza contatto per 1° appuntamento/intervista iniziale (012) Contatore esaurito per 1° appuntamento/intervista iniziale per tutti non risponde/utente non raggiungibile/nessuno in casa (007) Contatore esaurito per 1° appuntamento/intervista iniziale per tutti occupato (008) Contatore esaurito per 1° appuntamento/intervista iniziale per tutti segreteria/fax (009) Contatore esaurito per 1° appuntamento/intervista iniziale per combinazione di soli non risponde/non raggiungibile/occupato/segreteria/fax/nessuno (010)
Fuori target	Fuori target: numero di telefono errato/inesistente (13) Fuori target: gruppo di individui (14) Fuori target: seconda casa (15) Fuori target: ufficio, negozio, impresa (16)	Famiglia deceduta (021) Seconda casa (016) La famiglia vive in un altro comune italiano (018) La famiglia vive all'estero (019) La famiglia vive permanentemente in istituzione/convivenza (022) Negozio, impresa, azienda, ufficio (036) Insieme di individui (ISF vive altrove) (037)
Irreperibilità	Irreperibilità della famiglia (18)	Famiglia vive nel comune ma non si sa dove (017) Irreperibilità per nome/indirizzo/ sconosciuto (020) Assenza prolungata (015)
-assente	Alle famiglie che, al momento della chiusura del lavoro sul campo, restano sospese (con tentativi in corso), viene attribuito d'ufficio un codice di Fine Periodo.	Fine periodo per INI: nuovo (nessuna azione) (024) Fine periodo per INI: meno tentativi di quelli richiesti (025) Fine periodo per INI: più tentativi di quelli richiesti, tutti non risponde/non raggiungibile/nessuno (026) Fine periodo per INI: più tentativi di quelli richiesti, tutti occupati (027) Fine periodo per INI: più tentativi di quelli richiesti, tutti segreteria/fax (028) Fine periodo per INI: più tentativi di quelli richiesti, combinazione di non risponde/non raggiungibile/occupato/segreteria (029) Fine periodo per INI: più tentativi con almeno un contatto (030) Fine periodo per INI: più tentativi senza un contatto (031) Fine periodo per INI: iniziati tentativi senza contatto Comune (032) Fine periodo per INI: preso il 1° appuntamento (033) Fine periodo per INI: intervista iniziata (034) Fine periodo per IFIN: intervista non iniziata (038) Fine periodo per IFIN: intervista iniziata (039) Fine periodo: manca codifica 99999 (035)

(nominativi “vergini”) e quelle con tentativi in corso ma che sfiorano le ben delimitate finestre temporali prefissate per lo svolgimento di ciascuna fase. In particolare, è importante tenere sotto controllo tutte le cadute per fine periodo in fase di primo contatto e intervista iniziale (Cfr. Par. 3.2.1), in quanto non consentono di sfruttare il meccanismo delle sostituzioni e determinano quindi una perdita del campione effettivo.

Fatte salve le dovute differenze, entrambi i sistemi di esiti rappresentano la base fondamentale per il calcolo di un ampio set di indicatori, fortemente specializzato in base alle caratteristiche metodologiche delle due indagini. Ciascuno dei due sistemi di esiti consente, inoltre, di ricostruire i principali aggregati del Sistema Informativo sulla Documentazione delle Indagini (SIDI), garantendo la confrontabilità degli indicatori di qualità tra le indagini<sup>28</sup>.

### 3.4 Il trattamento dei dati

I benefici derivanti dall'utilizzo della tecnica *computer assisted* sono stati ampiamente sperimentati nell'indagine CATI Viaggi e Vacanze che prevedeva già l'implementazione di un piano di controllo delle incompatibilità tra variabili attivo in fase di rilevazione. Infatti, il questionario elettronico incorporava domande di controllo, verifiche del rispetto dei *range* e regole di coerenza tra le risposte fornite, che all'occorrenza attivavano diverse segnalazioni di errori e incongruenze causati dal rispondente o dal rilevatore. Ciò ha permesso di migliorare la qualità del dato “grezzo” e di ridurre l'impatto dei controlli e delle correzioni sui file dati.

Il Focus Viaggi ha ereditato il precedente piano delle incompatibilità, ampliato e migliorato in considerazione dei nuovi contenuti della rilevazione e della sua collocazione in qualità di focus dell'indagine sulle Spese delle Famiglie. L'esperienza di oltre un quindicennio di rilevazione CATI ha consentito, infatti, di beneficiare di indicazioni utili a valutare la validità e la pertinenza delle regole predisposte e la loro aderenza alla realtà indagata e ai suoi cambiamenti. Nuove regole sono state introdotte, alcuni controlli sono stati riprogettati in *soft check*, altri da *warning* sono diventati “bloccanti”. Inoltre, è stata migliorata l'efficacia della messaggistica abbinata alla segnalazione dell'errore, per aiutare l'intervistatore a riconoscere l'origine dell'errore stesso e a intervenire per ripristinare l'ammissibilità delle risposte fornite dall'intervistato. Per le specifiche del piano dei controlli integrato nel software CAPI si veda il paragrafo 3.4.1.

Come per l'indagine CATI, l'implementazione di alcuni controlli sui dati durante la somministrazione del questionario non annulla del tutto la necessità di una fase di individuazione degli errori (incoerenze e mancate risposte parziali), che prevede l'adozione di un piano di *check* dei dati articolato (distribuzioni di frequenza, analisi delle note dei rilevatori, analisi delle coerenze intra e inter-record, intra e inter dataset) basate su procedure più o meno complesse (cfr. parr. da 3.4.2 a 3.4.6). Alla fase di individuazione degli errori, segue quella delle correzioni con l'imputazione dei dati mancanti o incoerenti/anomali. L'obiettivo è quello di sanare sia gli errori di tipo sistematico, sia quelli di tipo casuale: generalmente, i primi vengono corretti con procedure deterministiche che applicano correzioni precise a fronte di condizioni di errore identificate nel set di record coinvolti. Per gli errori non sistematici o casuali, si usa, invece, un approccio di tipo probabilistico, che non prevede regole di cor-

<sup>28</sup> Nella progettazione degli indicatori di monitoraggio nelle indagini Istat si fa riferimento agli standard nazionali (al sistema Sidi, strumento generalizzato per il controllo della qualità delle indagini Istat) ed internazionali (Aapor, ovvero American association for public opinion research).

rezione predeterminate, ma utilizza degli algoritmi per l'imputazione dei valori mancanti o errati. Nell'indagine CATI "Viaggi e Vacanze" veniva utilizzata per lo più la prima tipologia di correzione, limitando le procedure probabilistiche alle variabili quantitative sulle spese, di cui si parlerà più avanti (cfr. par. 3.4.6). Nella nuova indagine, invece, le procedure probabilistiche e la Procedura Famiglie vengono usate anche per la correzione delle informazioni sull'individuo (cfr. par. 3.4.4), mentre si continua ad utilizzare un approccio deterministico per le variabili qualitative riguardanti i viaggi e le escursioni (cfr. par. 3.4.3).

In definitiva, la qualità dell'informazione statistica rilasciata dall'indagine è garantita da un sistema di controllo complesso, che comprende sia il piano delle incompatibilità integrato nel questionario elettronico, che permette un'acquisizione controllata delle informazioni, sia il piano di *check* e correzione dei dati a posteriori.

#### 3.4.1 L'acquisizione controllata dei dati e il piano delle incompatibilità

Durante la fase di riprogettazione del questionario elettronico per le sezioni del Focus Viaggi, sono state revisionate e aggiornate le regole formali e quelle sostanziali che consentono di minimizzare i possibili errori in fase di rilevazione/acquisizione delle informazioni presso le famiglie. Come è noto, le regole formali derivano dai percorsi strutturali del questionario e quindi sono strettamente collegate alle norme di compilazione e ai filtri in esso presenti (Istat, 2004). Esse permettono, dunque, di evitare errori di rotta in base ai percorsi filtrati del questionario. Le regole sostanziali, invece, dipendono da considerazioni di tipo statistico-matematico o da conoscenze a priori del fenomeno oggetto di rilevazione. In quanto tali, esse sono soggette a continue revisioni, non solo perché i contenuti del questionario possono variare, ma anche perché il fenomeno oggetto di studio è in continuo mutamento e situazioni prima inammissibili e trattate come anomale, diventano possibili e da non sanare (Istat, 2006a).

Alla luce di tali considerazioni, dei cambiamenti introdotti nella struttura e nei contenuti del questionario e della modifica del periodo di riferimento dei dati su viaggi ed escursioni (da trimestrale a mensile), per la nuova indagine CAPI sono stati revisionati i due principali strumenti che accompagnano il questionario elettronico: il *piano dei controlli di range delle variabili* e il *piano delle incompatibilità*.

I controlli di *range* sul dominio delle variabili sono distribuiti in quasi tutte le sezioni del Focus Viaggi e Vacanze e sono previsti principalmente per evitare errori di digitazione, inserendo un codice corrispondente ad una modalità che nella realtà non esiste per le variabili qualitative (ad esempio il "Mese di partenza del viaggio"='01-12'), oppure un valore non ammissibile per le variabili quantitative. In alcuni casi, il dominio di alcune variabili dipende dalle definizioni adottate; è il caso ad esempio del "Numero dei viaggi di vacanza o di lavoro abituali" che non può essere minore di quattro nel mese di riferimento, oppure della "Durata del viaggio" che non può essere inferiore a un pernottamento (altrimenti lo spostamento non può essere definito "Viaggio", bensì "Escursione") e non deve superare le 365 notti (oltre le quali, il luogo di destinazione del viaggio diventa il nuovo ambiente abituale).

Per quanto riguarda il piano di incompatibilità, come nella vecchia indagine, esso prevede che i controlli possano essere di due tipi: 1) *hard*, definiti anche "bloccanti" che inibiscono la registrazione del dato errato e richiedono necessariamente una correzione per poter proseguire l'intervista; 2) *soft*, definiti anche *warning*, di fronte ai quali l'intervistatore è chiamato a dare conferma o meno del dato registrato.

La gran parte dei controlli del Focus Viaggi e Vacanze sono *hard*, sebbene sia previsto

un ristretto numero di *soft check* che segnalano situazioni caratterizzate non da una vera e propria incoerenza, ma da risposte tra loro poco plausibili, per cui è richiesta una conferma dal rilevatore. Si attivano, ad esempio, messaggi di conferma quando l'intervistato dichiara una durata del viaggi di lavoro superiore a venti notti, oppure quando indica come tipo di viaggio una vacanza al mare, ma la destinazione principale è una località geografica senza mare, o ancora quando afferma di aver svolto almeno un'attività che richiede delle spese durante un'escursione, ma allo stesso tempo di non aver effettuato alcuna spesa.

In fase di somministrazione del questionario elettronico, le regole del piano di incompatibilità del Focus Viaggi e Vacanze permettono di garantire la coerenza tra:

- *le risposte fornite a due o più quesiti diversi*: ad esempio, la località di destinazione del viaggio deve essere diversa dal comune di residenza, oppure la destinazione principale del viaggio non deve coincidere con le destinazioni secondarie nel caso di un viaggio con più tappe;
- *la classe e il valore assoluto*: ad esempio, per le variabili quantitative di tipo monetario, come la spesa turistica, prima della domanda in cui si richiede di quantificare in valore assoluto la spesa puntuale, viene richiesto di indicare la classe di appartenenza della variabile. In questo caso il controllo agisce impedendo di digitare nella domanda in valore assoluto un valore non compatibile con la classe digitata precedentemente;
- *unità diverse*: un esempio di questo tipo di controllo è quello che opera sulla data di inizio e sulla durata di viaggi realizzati da uno stesso individuo, al fine di evitare che tra gli spostamenti turistici della stessa persona vi siano periodi di sovrapposizione.

**Prospetto 3.7 - I controlli interni al Focus "Viaggi e Vacanze": regole formali, *range* e piano delle incompatibilità.**

STRUMENTI	TIPO DI CONTROLLO	SEVERITY	ESEMPI
Regole formali	Percorsi filtrati del questionario (filtri e salti)	-	Filtro età >=15 anni per i rispondenti della sezione riepilogativa viaggi di lavoro
Controllo <i>range</i> variabili	Valori fuori dominio	Hard	Durata del viaggio >365 notti
Piano delle incompatibilità	Incoerenza tra risposte fornite a due o più quesiti	Hard e Soft	Destinazione del viaggio=Comune di residenza
	Incoerenza tra classe e valore assoluto	Hard	Valore minimo classe di spesa per alloggio > valore assoluto spesa per alloggio > Valore massimo classe di spesa per alloggio
	Incoerenza tra unità diverse	Hard e Soft	Data di inizio del viaggio inclusa tra la data di inizio e la data di fine dei viaggi precedentemente descritti per uno stesso individuo

La gestione automatizzata del questionario con l'attivazione delle regole di coerenza in fase di rilevazione non esaurisce le attività di individuazione e correzione di tutti i possibili errori presenti sui file dati. In primo luogo, sebbene il questionario elettronico incorpori una quantità rilevante di controlli e indirizzi il flusso delle risposte verso i percorsi prestabiliti, non può essere appesantito con vincoli eccessivi. Questi, infatti, finirebbero per rendere poco fluido il software CAPI e rallenterebbero l'intervista, esponendo il rilevatore al rischio di un rifiuto o di una scarsa disponibilità del rispondente. In altri casi, la risoluzione di alcune problematiche durante la somministrazione del questionario non è vincolante per il proseguimento dell'intervista, anzi, potrebbe essere controproducente, soprattutto quando

per sanare l'incongruenza il rilevatore sia costretto a tornare indietro di parecchi quesiti.

Pertanto, per l'implementazione del piano delle incompatibilità si è perseguito il corretto equilibrio tra costi e benefici, rimandando al momento successivo all'intervista la correzione di alcune incoerenze.

Le procedure di controllo e correzione *ex post* si articolano in diverse fasi:

- i controlli mensili;
- i controlli e il trattamento trimestrali;
- il trattamento annuale.

#### 3.4.2 I controlli mensili

Come nell'indagine CATI, anche nella nuova indagine l'organizzazione dell'archivio dei microdati si basa sulle relazioni esistenti tra le unità "famiglia", "individuo", "evento" (viaggio e/o escursione). La differenza rispetto al passato consiste nel fatto che tale archivio viene popolato non più trimestralmente, ma mensilmente, da quattro file:

- File *Famiglie*, con le informazioni generali su ciascuna famiglia (una per ogni record);
- File *Individui*, con informazioni su ciascun individuo (uno per ogni record) membro della famiglia, che abbia viaggiato (turista) o meno;
- File *Viaggi*, con notizie su ciascun viaggio (uno per ogni record) realizzato da ogni membro della famiglia;
- File *Escursioni*, con dettagli su ciascuna visita in giornata (una per ogni record) realizzata da ogni membro della famiglia.

In questo modo i quattro file sono controllabili ed elaborabili in maniera indipendente, ma allo stesso tempo possono essere trattati contemporaneamente attraverso chiavi identificative, ciascuna delle quali distingue una unità statistica dall'altra ("Famiglia", "Individuo", "Viaggio", "Escursione") e ne permette l'assegnazione all'unità di ordine superiore (l'individuo alla famiglia, il viaggio o l'escursione all'individuo e quindi alla famiglia)<sup>29</sup>.

I primi controlli sui quattro file vengono effettuati mensilmente, immediatamente dopo l'invio dei dati grezzi da parte della società incaricata di effettuare le interviste. Si tratta di controlli "di primo livello", finalizzati a verificare la coerenza dei file scaricati rispetto ai tracciati record forniti. In questa fase vengono prodotte distribuzioni di frequenza semplici su ciascuno dei quattro *file* al fine di controllare prontamente: 1) il rispetto delle lunghezze record; 2) la presenza di tutte le variabili; 3) l'eventuale scarico di valori *missing*, laddove non siano ammissibili; 4) la presenza di codici non ammessi. Distribuzioni doppie di frequenza, inoltre, vengono usate per fare una prima valutazione circa il corretto funzionamento delle variabili filtro per il controllo dei percorsi del questionario.

Le distribuzioni di frequenza, unitamente all'analisi delle note testuali inserite dall'intervistatore in ogni fase della rilevazione e controllate ad ogni invio mensile, sono degli ottimi strumenti di monitoraggio e di supporto, anche per la segnalazione di errati funzionamenti del software CAPI e/o delle procedure di scarico dei dati.

Data la complessità del questionario elettronico, il controllo mensile sulla correttezza dei vincoli è un'attività che ha richiesto molte energie nella fase di impianto della nuova indagine CAPI. Inoltre, essa continua ad essere cruciale ogni volta che cambia la società esterna responsabile della rilevazione sul campo e di conseguenza gli strumenti hardware e software per realizzare le interviste, nonché ad ogni nuova modifica introdotta nel que-

<sup>29</sup> Per approfondimenti sulle motivazioni, i vantaggi e i limiti di tale scelta si veda Istat, 2003, pp. 85—87.

stionario stesso.

### 3.4.3 I controlli e il trattamento trimestrali

I controlli “di secondo livello”, più approfonditi, vengono effettuati alla fine di ogni trimestre sull’archivio dei microdati. Un’operazione preliminare è rappresentata dal trattamento delle modalità “Altro specificare”, cioè le stringhe di testo, previste per alcuni quesiti. Nel corso del tempo, le descrizioni più ricorrenti sono state ricodificate in nuove modalità di risposta, ampliando così l’elenco delle voci di risposta a uno specifico quesito, con l’obiettivo di prevedere, laddove necessario, la categoria “Altro” solamente per i casi residuali. E’ stato, altresì, verificato che nella maggior parte dei casi, gli “Altro specificare”, erano riposizionabili nelle categorie già previste, quindi, in realtà, si trattava di errori di codifica dovuti a disattenzione o alla mancata comprensione del significato delle alternative di risposta predefinite. Per tutte queste ragioni, l’utilizzo di tale opzione di risposta si è progressivamente ridotta<sup>30</sup> e, attualmente, riguarda solo il Comune, la Provincia o lo Stato estero di destinazione del viaggio e dell’escursione e il motivo principale di quest’ultima. Nel primo caso, si tratta di garantire la possibilità di indicare voci non presenti nella lista delle destinazioni inclusa nel motore di ricerca del questionario elettronico, lista che non è sempre possibile aggiornare in tempo reale. Nel secondo caso, poiché le escursioni sono diventate un fenomeno da rilevare solamente in anni recenti, la lista delle motivazioni precodificate potrebbe non essere esaustiva.

Successivamente al trattamento delle modalità “Altro specificare”, ogni file trimestrale (*Famiglie, Individui, Viaggi ed Escursioni*) viene sottoposto a procedure di controllo intra-dataset (intra e inter record) e inter-dataset. Tali procedure, lanciate all’avvio della fase di correzione, forniscono indicazioni sulla dimensione e sulla tipologia degli errori presenti; esse vengono spesso reiterate, per verificare la progressiva correzione delle variabili coinvolte. Al termine del processo di correzione, infine, le stesse sono indispensabili per valutare la bontà e l’efficacia degli interventi correttivi adottati e per evidenziare le incoerenze residue o eventuali nuove incompatibilità generate da un intervento errato.

Nello specifico, sui quattro file vengono effettuati:

- *controlli di coerenza intra-record*, per individuare incompatibilità tra i valori assunti da variabili diverse nell’ambito di una stessa unità statistica (famiglia, individuo, evento viaggio/escursione): ad esempio, tra regione di residenza e ripartizione territoriale, tra numero totale degli individui coabitanti e numero dei componenti della famiglia di fatto, tra viaggi abituali e non abituali, tra viaggi e pernottamenti, tra motivo del viaggio e tipologia, ecc.;
- *controlli di coerenza inter-record*, per identificare le incompatibilità tra i valori che una variabile assume nell’ambito di un’unità statistica e i valori assunti da variabili ad essa correlate, ma relative ad un’altra unità: un esempio è rappresentato da tutti i controlli mirati a verificare il corretto funzionamento del popolamento, da parte del software CAPI, dei record viaggio o escursione degli individui aggregati a un medesimo spostamento<sup>31</sup>.

30 Ad esempio, dal 2015 si è deciso di eliminare il campo “Specificare” dalla modalità “Altro”, per il motivo di lavoro, perché ritenuto superfluo, indicando esplicitamente e tra parentesi tutti quei motivi che sono risultati effettivamente non collocabili all’interno delle 14 modalità di risposta codificate.

31 In questo caso, infatti le unità statistiche sono rappresentate dai viaggi o dalle escursioni dei diversi componenti della famiglia, per i quali un viaggio/escursione è identificato come “principale” e gli altri come “secondari” perché effettuati da membri aggregati (Cfr. Par. 3.2.5). Le procedure di check, in questo caso, controllano la compatibilità di tutte le variabili “identificative” dell’aggregazione (numero dei partecipanti, progressivi viaggio per titolare e aggregato, relazioni di parentela dei partecipanti, ecc.).

### 3. La nuova rilevazione delle informazioni sulla domanda turistica

- *controlli inter-dataset*, finalizzati a verificare la coerenza tra i contenuti informativi di due o più file diversi. Rientrano in questo ambito le verifiche sulla variabile “Numero dei componenti della famiglia di fatto” presente nel dataset *Famiglie* e numero dei record individuali del dataset *Individui*, oppure sulla coerenza tra alcuni *flag* e contatori “valorizzati” nei file *Famiglie* e *Individui* e il corrispondente numero dei record effettivi presenti sui file *Viaggi* o *Escursioni*.

I controlli sui quattro file riguardano diversi aspetti; in primo luogo, è necessario verificare l'efficacia dei controlli bloccanti (*hard*) previsti nel piano delle incompatibilità. L'apparente ridondanza di questo *step* è motivata dalla necessità di controllare il software nei primi mesi di avvio della nuova indagine CAPI e ricontrollarlo ogni volta che cambia la società esterna incaricata di gestire l'attività di rilevazione e con essa il software utilizzato per il questionario elettronico<sup>32</sup>. Inoltre, sebbene le mancate risposte a uno o più quesiti non siano ammesse dal questionario elettronico, esse possono presentarsi a causa di problemi tecnici occorsi durante l'intervista e non controllabili a priori. Allo stesso modo, eventuali errori di percorso del questionario possono essere imputati a malfunzionamenti di componenti hardware e/o software dello strumento di rilevazione, oppure a imprecisioni nelle procedure di archiviazione dei dati da parte della società esterna. Il verificarsi di questi errori richiede evidentemente correzioni di tipo deterministico.

Sempre in questa fase, si procede alla verifica dei *soft check*, esaminando tutte quelle risposte tra loro poco plausibili, ma confermate dal rilevatore a fronte di un messaggio di *warning*. In questo caso, i record ritenuti critici vengono analizzati singolarmente al fine di prendere decisioni ponderate sulle eventuali correzioni da apportare, anche basandosi sull'esperienza accumulata negli anni e su analisi in serie storica dei fenomeni oggetto di studio.

Particolare attenzione è riservata, poi, alle variabili cosiddette “di servizio” create dal software CAPI successivamente alle interviste e che derivano dalle informazioni raccolte. Rientrano in questo gruppo:

- le *chiavi identificative* dei record dei file *Famiglie*, *Individui*, *Viaggi* ed *Escursioni*;
- i *contatori*, come, ad esempio, il contatore progressivo di tutti i viaggi della famiglia intervistata o il numero dei partecipanti a uno stesso viaggio;
- gli *indicatori di presenza/assenza*, denominate variabili “filtro” o “flag”, come ad esempio il *flag* che indica, per ciascuna famiglia, la presenza o meno di record viaggi non abituali, oppure quello che segnala, per ciascun viaggio, la presenza o meno di destinazioni secondarie oltre a quella principale;
- le *variabili “costruite”*, cioè ottenute per funzione matematica, come la spesa media pro capite per viaggio, o logica, come le variabili qualificanti un viaggio (tipologia del viaggio

**Prospetto 3.8 - Tipologia di controlli trimestrali sui file *Famiglie*, *Individui*, *Viaggi*, *Escursioni* e trattamento degli errori**

TIPO DI CONTROLLO	TIPO DI TRATTAMENTO
Analisi Modalità “Altro specificare”	Ricodifica
Verifica efficacia controlli bloccanti del questionario elettronico	Correzioni deterministiche
Verifica attivazione <i>soft check</i>	Analisi record “critici” e correzioni basate su conoscenza del fenomeno e analisi in serie storica
Verifica costruzione variabili derivate <ul style="list-style-type: none"> <li>• Chiavi identificative</li> <li>• Contatori</li> <li>• Indicatori di presenza/assenza</li> <li>• Variabili costruite per funzione matematica o logica</li> </ul>	Correzioni deterministiche
Coerenza variabili intra-record	Correzioni deterministiche e probabilistiche
Coerenza variabili inter-record	
Coerenza variabili inter-dataset	

<sup>32</sup> Tutti i contratti con le società esterne dal 2013 al 2019 hanno previsto lo sviluppo del questionario elettronico come attività in carico alla società.

– vacanza/lavoro -, se viaggio effettuato in qualità di titolare o aggregato, ecc.), che sono importanti per verificare il corretto funzionamento del meccanismo di replicazione dei record viaggi ed escursioni per gli aggregati a uno stesso spostamento (cfr. par. 3.2.5).

Entriamo ora nel dettaglio di alcune procedure di controllo eseguite sui quattro dataset trimestrali. Per quanto riguarda il file *Famiglie*, il piano di *check*, oltre a permettere di sanare errori di incongruenza su informazioni peculiari della famiglia<sup>33</sup> può essere di ausilio per il monitoraggio dell'andamento della rilevazione e dell'operato dei rilevatori. Così, ad esempio, le informazioni sulla durata media dell'intervista o sulle difficoltà di intervista e sulla disponibilità delle famiglie sono indicatori utili a valutare immediatamente la qualità delle interazioni tra intervistato, intervistatore e strumento di rilevazione, individuando eventuali situazioni anomale e rispettive modalità di risoluzione.

Anche sul file *Individui* vengono effettuati alcuni controlli di qualità, come l'analisi delle variabili "di servizio" che forniscono informazioni sull'ammontare delle schede viaggio/escursione annullate (perché erroneamente dichiarate, oppure volontariamente non descritte) a fronte del numero totale dei viaggi/escursioni dichiarati inizialmente nella scheda riepilogativa. In merito a questo aspetto, l'esperienza pregressa evidenzia che una quota di eventi annullati o non descritti intorno al 10-12% è fisiologica, oltre tale soglia, invece, essa costituisce un problema che potrebbe inficiare le stime sul fenomeno. Quando sono stati rilevati valori oltre la soglia, questi sono stati portati subito all'attenzione dei rilevatori durante i *debriefing* per cercare di comprenderne le ragioni, facendo emergere, ad esempio, problemi di funzionamento del software CAPI, oppure difficoltà nel mantenere la collaborazione della famiglia fino alla fine dell'intervista.

Per il file *Individui*, merita un approfondimento particolare il piano di controlli e di correzione delle variabili strutturali, alcune delle quali (relazione di parentela, stato civile attuale, anno di matrimonio e stato civile precedente al matrimonio), vengono trattate con l'ormai consolidata Procedura Famiglia. Le altre variabili individuali (titolo di studio, condizione professionale, ecc.) sono trattate trimestralmente con procedure di correzione probabilistica utilizzando il software SCIA (Sistema di Controllo e Imputazione Automatica) (paragrafo 3.4.4.).

Una volta trattate le variabili strutturali, che sono di primaria importanza perché, in molti casi, indirizzano i percorsi di compilazione delle varie sezioni del questionario, si procede al controllo delle variabili tematiche presenti nei file *Viaggi* ed *Escursioni*, cioè le variabili specifiche sulla domanda turistica.

I principali controlli sulle variabili tematiche effettuati sul file *Viaggi* riguardano:

- la tipologia del viaggio (di lavoro o vacanza), il motivo prevalente del viaggio, con particolare attenzione all'individuazione dei viaggi di lavoro a fronte dei quali vi è un contratto di lavoro nel luogo di destinazione<sup>34</sup>;
- il tipo di vacanza prevalente, la destinazione prevalente, il tipo di destinazione (mare, crociera, montagna, città, campagna, altro) e il numero di destinazioni secondarie indicate. Oltre ai controlli di coerenza tra tutte le variabili, si tratta anche di valutare di volta in volta l'ammissibilità di viaggi con una sola destinazione geografica e tanti tipi di destinazione indicati (il tipo di destinazione, o luogo visitato, è infatti una

33 Le correzioni trimestrali sul file *Famiglie* permettono ad esempio di ripristinare, laddove necessario, la congruenza tra le informazioni su numero totale dei coabitanti, numero componenti della famiglia di fatto dichiarati in intervista iniziale e numero di componenti non facenti più parte del nucleo familiare al momento dell'intervista finale, nonché sugli indicatori di presenza/assenza di record viaggi/escursioni.

34 Tali viaggi non dovrebbero essere rilevati perché non rientrano nel campo di osservazione degli spostamenti turistici; sebbene non inclusi tra i viaggi di lavoro, essi vengono comunque registrati e trattati nella fase correzione trimestrale al fine di tenere sotto controllo il fenomeno.

variabile *multiple response*);

- il meccanismo di aggregazione dei membri della famiglia a uno stesso viaggio, mediante la creazione di una chiave che consente di controllare tutte le variabili coinvolte;
- la durata del viaggio, con la finalità anche di individuare e trattare i casi *outliers*;
- le date dei viaggi, per verificare l'efficacia delle regole che impediscono periodi di sovrapposizione tra viaggi riferibili a uno stesso individuo e la presenza di viaggi che non ricadono all'interno del periodo di riferimento;
- l'organizzazione del viaggio, l'alloggio e il mezzo di trasporto, al fine di garantire la coerenza tra il tipo di prenotazione del viaggio e alcune tipologie di alloggio (come gli alloggi gratuiti) o di mezzo di trasporto (come l'auto propria o aziendale);
- le spese; si tratta di un controllo finalizzato a sanare errori di range non intercettati dal software CAPI (tra la classe di spesa e il valore puntuale della spesa) o errori di calcolo intervenuti all'atto della creazione del file (ad esempio il calcolo della spesa media pro-capite o della spesa totale come somma delle spese parziali). Controlli più approfonditi e correzioni delle variabili di spesa sono rimandati alla fase del trattamento annuale (paragrafo 3.4.6.).

Analoghe procedure di controllo sono lanciate sul file *Escursioni* per trattare prevalentemente:

- errori sulle variabili nei record escursioni di più partecipanti della stessa famiglia (come per i viaggi, una chiave consente di verificare la coerenza tra tutte le variabili interessate);
- destinazioni all'estero non ammissibili o poco probabili da raggiungere in giornata considerando contemporaneamente luogo di origine, destinazione e tipo di mezzo utilizzato;
- coerenza tra la spesa sostenuta e le attività a pagamento effettuate durante la visita in giornata.

Le incoerenze rilevate sui file degli eventi sono sanate mediante correzioni di tipo deterministico di carattere generalizzato, oppure formulate ad hoc dopo aver analizzato singoli record o gruppi di record incongruenti. In questo caso le correzioni e le imputazioni si basano sulla conoscenza specifica del fenomeno, su analisi in serie storica dei dati e sul confronto con altre fonti statistiche (Prospetto 3.9).

#### 3.4.4 La correzione delle variabili strutturali dell'individuo

**Prospetto 3.9 - Quadro di sintesi dei controlli trimestrali sulle variabili tematiche dei file Viaggi ed Escursioni.**

VIAGGI	ESCURSIONI	TIPO DI TRATTAMENTO
Congruenza tra tipologia del viaggio e motivo del viaggio	Partecipazione di più membri della famiglia a una stessa visita in giornata	Correzioni deterministiche di carattere generalizzato o formulate ad hoc: • conoscenze specifiche sul fenomeno; • analisi in serie storica; • confronto con altre fonti statistiche.
Congruenza tra tipo di vacanza, destinazione principale, tipo di destinazione o luogo visitato e numero di destinazioni secondarie	Ammissibilità di escursioni con destinazioni estere (congruenza tra luogo di origine, destinazione e tipo di mezzo utilizzato)	
Partecipazione di più membri della famiglia a uno stesso viaggio		
Durata del viaggio		
Periodo di inizio e fine dei viaggi		
Organizzazione del viaggio, alloggio, mezzo di trasporto	Controllo di primo livello sulle spese turistiche	
Controllo di primo livello sulle spese turistiche		

L'indagine sulle Spese delle famiglie rileva un set di informazioni sull'individuo più ampio rispetto all'indagine CATI Viaggi e vacanze. La correzione di queste informazioni viene effettuata utilizzando software sviluppati in ISTAT, nello specifico la Procedura Famiglie per correggere le variabili anagrafiche individuali e il software SCIA per le altre variabili riferite all'individuo (Prospetto 3.10).

La Procedura Famiglie corregge le informazioni relative alla relazione di parentela<sup>35</sup>, allo

**Prospetto 3.10 - Informazioni rilevate sull'individuo dall'indagine sulle Spese delle famiglie e dall'indagine CATI Viaggi e vacanze e procedura di correzione utilizzata**

INFORMAZIONI RILEVATE SULL'INDIVIDUO	INDAGINE CAPI SULLE SPESE DELLE FAMIGLIE	INDAGINE CATI VIAGGI E VACANZE (AL 2013)	NUOVA INDAGINE CAPI	
			Correzione Procedura Famiglie	Correzione SCIA
Sesso	x	x	x	-
Età	x	x	-	-
Relazione di parentela con l'intervistato	x	x	x	-
Stato civile	x	x	x	-
Anno di matrimonio	x	x	x	
Titolo di studio più alto conseguito	x	x	-	x
Condizione lavorativa oggettiva	x	x	-	x
Condizione lavorativa soggettiva	x	x	-	x
Posizione nella professione	x	x	-	x
Settore di attività economica	x	x	-	x
Cittadinanza	x	x	-	-
Comune o Stato estero di nascita	x	-	-	-
Corso di studi frequentato	x	-	-	x
Corso pubblico o privato	x	-	-	x
Codifica della professione	x	-	-	-
Orario di lavoro	x	-	-	x
Rapporto di lavoro	x	-	-	x
Fonte principale di reddito	x	-	-	x
% di reddito percepito dal componente (sul totale dei redditi familiari)	x	-	-	x

stato civile al momento dell'intervista, allo stato civile precedente al matrimonio e all'anno di matrimonio. Si tratta di una procedura ormai consolidata<sup>36</sup>, sviluppata internamente all'Istituto, che tratta tutti gli individui della famiglia nel suo complesso ed è in grado, attraverso la ricostruzione delle relazioni esistenti tra i diversi componenti della famiglia, di validare e correggere le informazioni sopra citate. L'approccio della procedura, che supera la mera correzione intra-record per tener conto contemporaneamente di tutte le informazioni utili fornite dagli individui della stessa famiglia, consente una correzione qualitativamente elevata, che non è sempre possibile effettuare in fase di intervista. A corredo delle informazioni rilevate presso la famiglia, sono utilizzate anche le informazioni presenti sulle Liste Anagrafiche Comunali (LAC, vedere par. 3.1.3), cioè il contenuto degli stati di famiglia, che possono rivelarsi fondamentali per dirimere le situazioni più complesse (per esempio, quelle delle

<sup>35</sup> La relazione di parentela, rilevata nell'indagine Viaggi e vacanze in base a 20 modalità, è stata nella nuova indagine semplificata a 17 modalità, accorpando alcuni item (nonno, cugino, zio, altro parente) nel generico item "Altro parente".

<sup>36</sup> Per approfondimenti, si veda il paragrafo *La procedura Famiglie* in Istat, 2006a, pag. 208-217.

### 3. La nuova rilevazione delle informazioni sulla domanda turistica

famiglie numerose o delle famiglie di stranieri).

La procedura consente la possibilità di effettuare una verifica immediata delle correzioni proposte, mediante schermate che visualizzano la situazione familiare precedente e successiva alla correzione (Fig. 3.8).

Nell'esempio, la relazione di parentela del componente numero 4 (di 15 anni) è pale-

**Figura 3.8 - PF: Schermata relativa ad una famiglia errata nell'informazione "relazione di parentela" e proposta di correzione**

The screenshot shows a software interface with a data table and a control panel. The control panel includes dropdown menus for 'Dati2' (set to 'Corrette') and '95169', along with navigation buttons and checkboxes for 'Mantieni Sottoinsieme' and 'Accetta Modifiche'. The table has columns: Rp, Sesso, Età, Stc, Stcp, Stcpm, Amatri, and Messaggi di errore. The data is as follows:

	Rp	Sesso	Età	Stc	Stcp	Stcpm	Amatri	Messaggi di errore
1	01	1	56	2	.	.	1 1989	
2	02	2	51	2	.	.	1 1989	
3	06	2	18	1	.	.	.	
4	04	1	15	1	.	.	CMP41	

Below this table, a second table shows the proposed corrections:

	Rp	Sesso	Età	Stc	Stcp	Stcpm	Amatri	Messaggi di errore
1	01	1	56	2	.	.	1 1989	
2	02	2	51	2	.	.	1 1989	
3	06	2	18	1	.	.	.	06 10
4	06	1	15	1	.	.	.	06 10

The bottom control panel contains buttons for 'Where', 'Informa', 'Report', 'Valida', 'Correggi', 'Regole', 'Moduli', 'Param', and 'Chiudi'.

mente errata poiché è stato dichiarato come genitore (relazione di parentela=04), della Persona di Riferimento, cioè il componente numero 1, mentre è ovviamente suo figlio (relazione di parentela=06). In questo caso la Procedura Famiglie ha permesso di intercettare un errore nell'implementazione del controllo hard da parte nel questionario elettronico. Qualora la correzione proposta non sia ritenuta appropriata, per esempio sulla base di informazioni ausiliarie derivanti da fonti esterne (come le LAC) o dallo stesso file dei microdati, è possibile rifiutarla e correggerla interattivamente tramite le stesse schermate.

La possibilità di intercettare con maggiore precisione tutti gli errori eventualmente presenti nelle informazioni individuali, soprattutto relativamente alla relazione di parentela, permette di ricostruire correttamente la struttura della famiglia. In prospettiva, ciò potrà fornire una solida base informativa per l'analisi delle informazioni sui viaggi in chiave familiare (cfr. Cap. 5).

Le altre variabili qualitative riferite all'individuo sono corrette utilizzando il software generalizzato SCIA, che consente il controllo e la correzione probabilistica delle mancate risposte parziali e delle incoerenze logico/formali utilizzando la metodologia proposta da Fellegi e Holt (Fellegi and Holt, 1976). Sulla base di regole di incompatibilità definite a priori tra le variabili, il software individua i record che presentino delle incoerenze. Tali incoeren-

ze vengono risolte imputando, sulle variabili coinvolte, valori, tra quelli ammissibili, con probabilità condizionata alla distribuzione di frequenza relativa all'intero file dei microdati dell'indagine, che viene in tal modo rispettata. L'algoritmo di correzione si basa sul principio del minimo cambiamento, provvede cioè a eliminare gli errori individuati minimizzando nel contempo il numero di variabili modificate.

Si riporta nel Prospetto 3.11, a titolo di esempio, un estratto delle regole del piano di incompatibilità riferite all'informazione "Titolo di studio" per gli specifici casi della licenza elementare e della licenza media. Come si può vedere, nel questionario elettronico sono implementate due diverse tipologie di controlli in relazione all'ammissibilità della condizione (totalmente inammissibile=hard, possibile=soft). La presenza di tali controlli non esclude però che l'algoritmo di SCIA possa intercettare un errore nel file, se i controlli hard già presenti nel software CAPI non hanno funzionato correttamente.

Tuttavia, è importante evidenziare che il corretto funzionamento del questionario elet-

**Prospetto 3.11 - Regole del piano di incompatibilità di alcune informazioni sull'individuo**

CONTROLLO HARD	CONTROLLO SOFT	REGOLA DI SCIA
Licenza elementare < 9 anni	Licenza elementare < 10 anni	ETACAL (0-8) incompatibile con TITSTU (02)
Licenza media < 12 anni	Licenza media < 13 anni	ETACAL (0-11) Incompatibile con TITSTU (03)

tronico è in grado di minimizzare il numero di errori che si riscontrano nel file dei microdati. Ad esempio, nel 2016, sia il tasso medio annuale di famiglie errate (individuate dalla Procedura Famiglie), sia il tasso medio annuale di individui errati (identificati da SCIA) hanno superato di poco l'1%.

### 3.4.5 Il trattamento annuale

Una volta disponibili i microdati di tutti e quattro i trimestri, i file annuali *Famiglie, Individui, Viaggi ed Escursioni* vengono sottoposti a controlli di coerenza generale, anche per verificare l'effetto complessivo delle correzioni apportate nei singoli trimestri.

Durante il trattamento annuale vengono create nuove variabili, come risultato di aggregazione di più modalità di variabili rilevate, in vista della diffusione dei dati attraverso i canali e secondo le modalità previste. Così, ad esempio, si procede con la creazione di una nuova variabile "destinazione" per una migliore lettura dei dati, in cui i paesi esteri sono aggregati per macro-aree territoriali. In altri casi, le nuove variabili sono il risultato della sintesi di più variabili; un esempio è dato dalla nuova variabile "organizzazione", ottenuta considerando contemporaneamente il tipo di prenotazione per l'alloggio e per il mezzo di trasporto.

Inoltre, sui file annuali degli eventi viaggio ed escursione, si effettuano controlli e correzioni deterministiche e da modello sulle variabili quantitative di spesa, di cui si parlerà approfonditamente nel paragrafo successivo.

Infine, la fase del trattamento annuale costituisce anche il momento in cui il dato "grezzo" viene confrontato con il dato "trattato" al fine di produrre indicatori di qualità. Al netto delle correzioni dovute a errori non prevedibili e/o non contemplati nel piano delle incompatibilità per le ragioni sopra descritte (cfr. par. 3.4.1), il piano dei check sui file grezzi in una indagine CAPI dovrebbe produrre un tasso di errore complessivo vicino allo zero. Tuttavia, ciò accade solo nel momento in cui sia il questionario elettronico sia le procedure appron-

tate dalla società per lo scarico dei dati raggiungono una certa stabilità.

#### 3.4.6 Le variabili quantitative: da RIDA a IVEware per la correzione delle spese del viaggio e delle escursioni

La correzione delle mancate risposte parziali delle variabili di spesa del viaggio e dell'escursione<sup>37</sup>, uniformemente ad altre indagini sulle famiglie, avviene, dal 2012, mediante l'applicazione del modulo *Impute* del software IVEware<sup>38</sup>, che esegue un'imputazione da modello, singola o multipla, dei valori mancanti utilizzando il Metodo di imputazione delle Regressioni Sequenziali (Raghunathan *et al.*, 2001).

In precedenza, per le spese del viaggio<sup>39</sup>, era stato utilizzato il software RIDA<sup>40</sup> (Ricostruzione dell'Informazione con Donazione Automatica) che esegue l'imputazione dei valori mancanti delle variabili mediante un donatore, sulla base della minima distanza tra donatore e ricevente. Il cambiamento di strumento utilizzato è stato indotto da diverse necessità. Innanzitutto, l'esigenza di correggere più variabili rispetto al passato e con maggiore tempestività, in seguito all'entrata in vigore del Regolamento Europeo; in secondo luogo, la necessità di superare uno dei limiti del software RIDA, che risiede nel riutilizzo di uno stesso donatore per l'imputazione, operazione che può diventare problematica per variabili con un numero consistente di casi mancanti e/o in presenza di pochi donatori. Il software IVEware ha permesso di superare entrambi i vincoli, rispondendo efficacemente alle nuove esigenze.

Nel 2012, è stata avviata una sperimentazione per confrontare le due tecniche di imputazione, creando nuove stime a partire dalle variabili trattate con IVEware per gli anni già elaborati con RIDA (2010 e 2011) e producendo contemporaneamente, per il 2012, stime basate su valori imputati da entrambi i metodi. È stata fatta contestualmente una doppia sperimentazione, per definire se adottare una procedura trimestrale, come quella di RIDA, oppure passare a un'unica imputazione annuale. Dai risultati ottenuti, è emerso che l'imputazione effettuata tramite IVEware riproduce stime del tutto comparabili a quelle ottenute dalle variabili elaborate con RIDA, sia applicando un'unica procedura annuale, sia replicando i quattro passaggi trimestrali. Conseguentemente, le imputazioni delle spese mancanti del 2012 e degli anni successivi sono state effettuate col software IVEware, che ha sostituito RIDA ed è stato implementato con un'unica procedura annuale invece che trimestrale. Questo cambiamento, oltre a mantenere efficacia in termini di correzione delle stime, ha permesso di ridurre notevolmente i tempi dedicati alla fase di correzione delle variabili di spesa.

Per ogni variabile da trattare con IVEware è possibile definire il modello di regressione appropriato in base alla natura della variabile e specificare le variabili esplicative rilevanti. La tecnica consiste nell'imputazione multivariata dei valori mancanti, attraverso la generazione di modelli di regressione multipla univariata (approccio "*variable by variable*"). Ciascun modello lavora utilizzando i valori osservati nella distribuzione originaria e la distribuzione che si viene a creare, passo dopo passo, dopo le imputazioni. Per tutte le variabili di spesa del viaggio e dell'escursione, si è utilizzato il modello di regressione lineare normale, mag-

37 A queste variabili si aggiunge anche la correzione del reddito netto medio mensile familiare, effettuata tramite IVEware. Per ulteriori dettagli, si veda Istat, 2015.

38 IVEware è un freeware (<http://www.isr.umich.edu/src/smp/ive/>) sviluppato dal Survey Research Center, Institute for Social Research dell'University of Michigan.

39 La variabile di spesa per le escursioni è stata introdotta nel 2012 e quindi corretta fin dall'inizio con IVEware.

40 RIDA è un software open source con licenza Istat 2014. Per maggiori dettagli si veda il manuale utente scaricabile alla pagina <http://www.istat.it/it/strumenti/metodi-e-strumenti-it/strumenti-di-elaborazione/concordjava>.

giormente efficace per il trattamento delle variabili quantitative monetarie; per ottenere una distribuzione normale, i valori delle variabili sono stati preventivamente trasformati attraverso il logaritmo naturale e, dopo l'imputazione, sono stati ricondotti alla scala originaria.

Le variabili esplicative utilizzate sono quelle che descrivono qualitativamente sia il turista, sia il viaggio (o l'escursione) e che entrano nel modello solo dopo essere state trattate durante i precedenti controlli di tipo qualitativo. Le variabili esplicative scelte coincidono con le variabili di strato e di *matching* adottate nella procedura RIDA, già dimostrate correlate con le variabili di spesa. Le caratteristiche del turista incluse nel modello sono: la condizione lavorativa, la posizione nella professione, il livello d'istruzione, il dominio di appartenenza del comune e la ripartizione geografica di residenza. Viaggi ed escursioni sono stati caratterizzati rispetto alla destinazione, al motivo e al mezzo di trasporto utilizzato. Solo per i viaggi, in aggiunta a queste variabili, sono state prese in considerazione anche il tipo di viaggio (di lavoro o personale), l'alloggio, l'organizzazione e la durata in notti. Per tener conto della stagionalità del fenomeno è stata inclusa anche la variabile che identifica il trimestre di riferimento dei dati.

Per ogni viaggio, costituiscono oggetto di imputazione le variabili quantitative relative alla spesa per il trasporto, alla spesa per l'alloggio, alla spesa per il resto (escluso alloggio e trasporto) e alla spesa totale del viaggio; per le escursioni invece, si imputa la sola spesa complessiva, essendo l'unica informazione rilevata.

I casi di spesa dichiarata nulla, per una o più di queste variabili, sono accettati, in fase d'intervista, al fine di rendere il flusso del questionario più scorrevole e in seguito eliminati e imputati mediante IVEware durante la fase di controllo e correzione, laddove opportuno. La necessità di imputare le spese nulle deriva dal fatto che le spese di viaggio debbano essere sempre presenti, seppur di importo contenuto, a prescindere da chi le abbia sostenute. Obiettivo di queste domande è, infatti, la stima economica (costo) riferita al complesso dei viaggi (o delle escursioni) generato dai residenti e non la spesa effettivamente sostenuta dalle famiglie per turismo. Questo è il motivo per il quale anche la spesa dei viaggi di lavoro, sebbene sia spesso sostenuta dal datore di lavoro, è stimata mediante le dichiarazioni di spesa fatte dalle famiglie, che di frequente conoscono il costo del proprio viaggio di lavoro, pur non avendone sostenuta materialmente la spesa. Fanno eccezione le spese per alloggio per soggiorni trascorsi in abitazioni proprie ad uso vacanza o presso abitazioni di parenti o amici, che sono necessariamente nulle. Le eventuali spese per manutenzione della propria abitazione di proprietà ad uso vacanza non rientrano tra le spese turistiche (UNWTO, 2010).

La correzione mediante imputazione delle variabili di spesa è preceduta da una fase di controllo e correzione dei valori *outliers*. Si definiscono *outliers* quelle osservazioni che superano i valori minimi e massimi, detti valori "soglia", delle distribuzioni di ciascuna variabile di spesa. Gli *outliers*, per la loro rarità, non possono essere usati all'interno del software IVEware, altrimenti influenzerebbero le imputazioni e conseguentemente le stime; pertanto tali soglie diventano dei vincoli (*bounds*) ai valori da imputare. Gli *outliers* possono avere origini diverse, dovute a un errore d'interpretazione delle domande in fase di rilevazione, o alla presenza di valori veri ma per loro natura rari e quindi non replicabili in fase d'imputazione, per evitare distorsioni. Il controllo dei valori *outliers* viene effettuato valutando la loro attendibilità rispetto alla natura del viaggio, cioè stabilendo se siano plausibili in relazione alle caratteristiche del viaggio stesso (destinazione, mezzo di trasporto utilizzato, durata) e del viaggiatore (condizione professionale, regione di residenza, età). Qualora alcuni di questi valori siano ritenuti plausibili, benché rari, sono mantenuti, negli altri casi sono considerati errori e corretti insieme alle altre mancate risposte parziali (MRP),

mediante IVEware.

La scelta della soglia è frutto di un'attenta valutazione dell'affidabilità della spesa dichiarata per ciascun servizio acquistato per il viaggio. Le soglie di ogni variabile di spesa sono determinate in base alla loro distribuzione di frequenza per gruppi omogenei individuati rispetto alla destinazione, al mezzo di trasporto e al tipo di alloggio utilizzato. Tali gruppi, essendo endogeni al campione, possono variare di anno in anno (o, addirittura, ogni trimestre) tenendo conto implicitamente dei cambiamenti delle abitudini di spesa per i viaggi e dei livelli dei prezzi dei servizi connessi al viaggio. Solitamente, valori accettabili per le spese corrispondono a quelli coincidenti o inclusi tra il 5° e 95° percentile della distribuzione, che diventano le soglie della distribuzione. I percentili cui corrispondono le diverse soglie potrebbero variare ad ogni rilevazione, ma nel tempo è stata osservata una certa stabilità sia dei gruppi omogenei di spesa, sia dei percentili soglia.

Nonostante la loro tipologia - economica - e l'effetto ricordo da cui sono spesso affette, per le variabili di spesa le mancate risposte parziali al momento dell'intervista sono molto contenute. Nel 2016 i tassi di MRP sono compresi tra il 6,4% della spesa per l'alloggio e il 7,6% della spesa totale del viaggio (Tavola 3.2).

Tuttavia, tali valori diventano più elevati in seguito alle regole fissate per la correzione

**Tavola 3.2 - Record viaggi per tipo di spesa e per tipo di risposta; tassi di mancata risposta parziale e tasso di imputazione per le variabili di spesa. Composizioni percentuali su dati non pesati. Anno 2016.**

	SPESA ALLOGGIO	SPESA TRASPORTO	SPESA PER ALTRO	SPESA TOTALE
Record senza spesa	13,6	14,7	32,9	11,7
- missing	6,4	6,7	7,0	7,6
- nulla	7,2	8,0	25,9	4,1
Record con spesa >=5 euro (a)	86,5	85,2	67,2	88,3
Totale record	100,0	100,0	100,0	100,0
Tasso d'imputazione per variabile	15,7	16,7	38,3	55,5

Fonte: Indagine CAPI "Spese delle famiglie"

(a) Spese familiari di viaggio comprese tra zero e 4 euro non sono ammesse già in fase d'intervista.

dei valori mancanti, nulli o outliers. Ad esempio, nel caso della spesa totale, a fronte di una quota di record viaggi con spesa mancante o nulla piuttosto contenuta, pari all'11,7% dei record viaggi, il relativo tasso d'imputazione sale al 55,5%. Ciò dipende dal fatto che il tasso di MRP per la spesa totale assorbe anche i tassi di MRP delle spese che la compongono, qualora queste siano nulle, oppure outliers. Infatti, quando è presente soltanto la spesa totale e almeno due spese parziali sono nulle oppure outliers, l'imputazione di tali spese comporta necessariamente il ricalcolo della spesa totale, come somma delle nuove spese parziali. Diverso il caso in cui almeno due spese parziali siano mancanti e la totale valorizzata, poiché quest'ultima è ritenuta attendibile e l'imputazione delle spese parziali si effettua nel rispetto del valore della spesa totale, senza incidere sul suo tasso di imputazione.

Gli effetti delle imputazioni sono controllati attraverso il calcolo dei principali parametri della distribuzione (media, minimo, massimo, *standard error*) per ciascuna variabile di spesa (Tavola 3.3). Le distribuzioni delle variabili di spesa corrette non differiscono in modo significativo da quelle dei dati grezzi. Le medie delle distribuzioni sono molto simili per le tre spese parziali, meno per la spesa totale la cui differenza è ascrivibile alle considerazioni sopra citate. Per i dati corretti, i valori minimo e massimo sono pari ai valori soglia definiti durante la fase di correzione, ad eccezione degli *outliers* mantenuti perché plausibili.

**Tavola 3.3 - Principali parametri della distribuzione delle variabili di spesa prima e dopo le correzioni. Valori per viaggio (in euro). Dati non pesati. Anno 2016.**

VIAGGI	DATI GREZZI	DATI CORRETTI
SPESA PER TRASPORTO		
N	2.771	3.233
Media	96	97
Dev std	149	137
Minimo	3	3
Massimo	1.800	1.615
SPESA PER ALLOGGIO		
N	1.558	1.923
Media	228	222
Dev std	484	270
Minimo	2	6
Massimo	16.000	3.000
SPESA PER ALTRO		
N	2.235	3.233
Media	108	117
Dev std	128	144
Minimo	3	5
Massimo	1.300	1.948
SPESA TOTALE		
N	2.876	3.233
Media	298	346
Dev std	470	383
Minimo	3	11
Massimo	16.180	4.500
ESCURSIONI	DATI GREZZI	DATI CORRETTI
SPESA TOTALE		
N	3.685	3.884
Media	41	38
Dev std	50	36
Minimo	1	3
Massimo	700	450

Fonte: Indagine CAPI "Spese delle famiglie"

### 3.5 La costruzione delle stime

#### 3.5.1 Le stime della domanda turistica

L'indagine ha la finalità di ottenere informazioni sui movimenti turistici della popolazione e, pertanto, fornisce stime di parametri di diversa natura (totali, medie, rapporti, frequenze assolute e relative) con più riferimenti temporali e territoriali sul numero di turisti, di viaggi, di pernottamenti in viaggio e di escursioni sul territorio nazionale o all'estero.

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, secondo i domini di stima previsti (cfr. par. 3.1.1).

I viaggi, le notti e le escursioni rilevati in ciascun trimestre possono essere sommati per ottenere le stime annuali, mentre in virtù del disegno campionario, in ogni trimestre è rappresentata l'intera popolazione per cui 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 direttamente sono quindi di due tipologie:

- ammontare dei viaggi ed escursioni nel mese, nel trimestre e nell'anno;
- medie mensili (turisti, non turisti, viaggi ed escursioni) riferite ad un trimestre o a un anno.

L'indagine permette di stimare anche l'ammontare dei turisti e dei non turisti nell'anno e dei turisti nel trimestre, tramite quesiti retrospettivi sui periodi precedenti alla rilevazione.

#### 3.5.2 Il sistema di ponderazione e il trattamento degli outliers

Dato il carattere campionario dell'indagine, per poter 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.

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, ovvero si tratta di stimatori corretti, consistenti e con la stessa varianza campionaria (Deville, Särndal, 1992).

La strategia adottata per la costruzione dei coefficienti di riporto all'universo è simile a quella utilizzata dalle varie indagini campionarie sulle famiglie condotte dall'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.

### 3.5.3 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 ad 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;

$y_k$  valore della variabile  $Y$  assunto dalla  $k$ -esima osservazione della popolazione;

$y_j$  valore della variabile  $Y$  assunto dalla  $j$ -esima osservazione della popolazione;

$\pi_j$  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:

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

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

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

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  $\pi_{hij}$ , 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.

### 3.5.4 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 ( $\delta_g$ ):

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

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:

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

### 3.5.5 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 ( $\mathbf{k}$ ) e i pesi finali ( $\mathbf{w}$ ) 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<sup>41</sup>. Formalmente:

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

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.

<sup>41</sup> 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.

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. Nell'indagine CAPI sulle Spese delle Famiglie, le stime campionarie sono state vincolate ai seguenti totali noti:

1. Popolazione residente per ripartizione, sesso e classi di età (0-14, 15-29, 30-44, 45-59, 60-74, 75 e più);
2. Popolazione residente per regione;
3. Popolazione residente per ripartizione e tipologia di comune (Comuni metropolitani, comuni periferia dei comuni metropolitani e comuni oltre 50.000 abitanti, altri comuni);
4. Famiglie residenti per regione;
5. Famiglie residenti per ripartizione e tipologia di comune (Comuni metropolitani, comuni della cintura metropolitana e comuni con più di 50.000 abitanti, altri comuni);
6. Popolazione straniera residente per ripartizione e sesso;
7. 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);
8. Popolazione per ripartizione e mese;
9. Famiglie per ripartizione e mese.

I vincoli imposti sono gli stessi sia per le stime trimestrali che per le stime annuali.

L'integrazione dell'indagine Viaggi e vacanze nella nuova indagine sulle Spese delle Famiglie è stata l'occasione anche per un ripensamento del sistema di ponderazione, che è stato riprogettato e maggiormente sofisticato. Rispetto alla vecchia indagine CATI Viaggi e Vacanze, infatti, solo i vincoli ai punti 1, 2 e 3 sono gli stessi, al netto delle classi di età, che nella CATI erano decennali, e della diversa classificazione delle tipologie comunali (cfr. par. 3.1.1).

I vincoli ai punti 4 e 5 sono stati inseriti per rispettare il nuovo disegno di campionamento e considerare i differenti livelli e stili di spesa associati alle diverse tipologie di comune di residenza.

Il vincolo al punto 6 è stato inserito per tener conto del peso che la popolazione straniera ha assunto sul totale della popolazione residente nel corso del tempo e del fatto che questa sottopopolazione ha stili e livelli di consumo (anche turistici) generalmente differenti<sup>42</sup>.

Il vincolo al punto 7 è stato introdotto in considerazione del fatto che la condizione e la posizione professionale impattano sui livelli di reddito, e quindi di spesa, della popolazione.

Questo totale è desunto dalla distribuzione proveniente dall'indagine Istat sulle forze di lavoro.

I vincoli 8 e 9 permettono di ripartire uniformemente la popolazione e le famiglie nei diversi mesi (tre mesi nel trimestre o dodici nell'anno), al fine di tenere conto del carattere fortemente stagionale di alcune tipologie di spesa e comportamenti di viaggio.

Un'altra importante innovazione, rispetto alla vecchia indagine CATI Viaggi e Vacanze, è stata l'adozione di 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 che della varianza. I gruppi socio-economici all'in-

<sup>42</sup> Senza vincolo, la popolazione straniera risulterebbe ampiamente sottostimata. Per diversi motivi, infatti, le famiglie con stranieri hanno generalmente tassi di risposta più bassi rispetto alle famiglie di soli italiani.

terno 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, includendo anche le stime delle spese rilevate dal focus viaggi e vacanze<sup>43</sup>.

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. Il numero di osservazioni anomale è comunque limitato, in media meno di 50 famiglie a trimestre.

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.

#### 3.5.6 Gli errori di campionamento

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  $\hat{\text{Var}}(\hat{Y}_d)$  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

$$\hat{\sigma}(\hat{Y}_d) = \sqrt{\hat{\text{Var}}(\hat{Y}_d)}; \quad (1)$$

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

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

Come è stato descritto nel par. 3.5.2, 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  $\hat{\text{Var}}(\hat{Y}_d)$  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.

<sup>43</sup> Nello specifico, la spesa media per viaggio riferita al trasporto, la spesa media giornaliera per l'alloggio, le altre spese e il totale del viaggio, l'ammontare di spesa totale per famiglia.

L'espressione linearizzata dello stimatore è data, quindi, da

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

dove  $Z_{hij}$  è la variabile linearizzata espressa come  $Z_{hij} = Y_{hij} - \mathbf{X}'_{hij}\beta$ , essendo  $\mathbf{X}_{hij} = (X_{hij,1}, \dots, X_{hij,K})'$  il vettore contenente i valori delle variabili ausiliarie, osservati per la generica famiglia  $h_{ij}$  e  $\beta$ , 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

$$\hat{\text{var}}(\hat{Y}_d) \cong \hat{\text{var}}(\hat{Z}_d) = \sum_{h=1}^{H_d} \hat{\text{var}}(\hat{Z}_h). \quad (4)$$

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

$$\hat{\text{var}}(\hat{Y}_d) \cong \hat{\text{var}}(\hat{Z}_d) = \sum_{h=1}^{H_{AR}} \hat{\text{var}}(\hat{Z}_h) + \sum_{h=1}^{H_{NAR}} \hat{\text{var}}(\hat{Z}_h), \quad (5)$$

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

$$\sum_{h=1}^{H_{AR}} \hat{\text{var}}(\hat{Z}_h) = \sum_{h=1}^{H_{AR}} M_h \frac{(M_h - m_h)}{m_h(m_h - 1)} \sum_{j=1}^{m_h} (Z_{hj} - \bar{Z}_h)^2, \quad (6)$$

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

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

dove le quantità sono espresse come

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

Utilizzando le espressioni (6) e (7) è possibile, infine, calcolare la varianza di campionamento,  $\hat{\text{var}}(\hat{Y}_d)$ , 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:

$$\hat{\sigma}_{(d\hat{Y})} = \frac{1}{4} \sqrt{\sum_{t=1}^4 \hat{\text{Var}}_{(d\hat{Y}_t)}}, \quad \hat{\varepsilon}_{(d\hat{Y})} = \frac{\hat{\sigma}_{(d\hat{Y})}}{d\hat{Y}}, \quad (8)$$

in cui si è indicato con  $\hat{\text{Var}}_{(d\hat{Y}_t)}$  la stima della varianza della stima  $_{d\hat{Y}_t}$ , riferita al dominio  $d$  e al trimestre di indagine  $t$ , 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

$$\hat{\varepsilon}_{(d\hat{R})} = \sqrt{\hat{\varepsilon}^2_{(d\hat{Y}_1)} - \hat{\varepsilon}^2_{(d\hat{Y}_2)}}. \quad (9)$$

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  $P$  contenga il parametro di interesse. Con riferimento ad una generica stima  $\hat{Y}$  tale intervallo assume la seguente forma:

$$\{\hat{Y} - k\hat{\sigma}(\hat{Y}), \hat{Y} + k\hat{\sigma}(\hat{Y})\} \quad (10)$$

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

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.

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:

$$\log \hat{\varepsilon}^2_{(d\hat{Y}_t)} = a + b \log(d\hat{Y}_t) \quad (11)$$

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 diffusi solamente i risultati relativi al primo e al terzo trimestre<sup>44</sup>.

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:

$$\hat{\sigma}(\hat{Y}_d) = a + b \hat{Y}_d + c \hat{Y}_d^2 \quad (12)$$

dove i parametri  $a$ ,  $b$  e  $c$  vengono stimati utilizzando il metodo dei minimi quadrati, adattando il modello (12) a una nuvola di punti costituita da un consistente numero di coppie  $(\hat{\sigma}(\hat{Y}_d), \hat{Y}_d)$ . In considerazione del fatto che il modello (12) è di tipo empirico, l'insieme delle stime  $\hat{Y}_d$  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. 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.

### 3.6 La diffusione delle stime della domanda turistica

La diffusione dei risultati ha cadenza annuale, ripetuta in più momenti dell'anno, in base alla tipologia di stime pubblicate e alle diverse modalità di rilascio.

Le stime provvisorie sono diffuse nella Statistica Report "Viaggi e vacanze in Italia e all'estero"<sup>45</sup>, nel mese di febbraio, solitamente 40 giorni lavorativi dopo la fine del periodo di riferimento e dopo appena circa 20 giorni dalla chiusura dell'ultimo periodo di rilevazione. Questa diffusione avviene il giorno prima dell'apertura della Borsa Internazionale del Turismo<sup>46</sup> che si tiene ogni anno a Milano e che, a livello nazionale, è il punto di riferimento

44 Per la presentazione sintetica degli errori campionari per le stime di frequenze relative agli individui e per le stime di totali riferite a viaggi, notti ed escursioni si vedano le note metodologiche dell'indagine "Viaggi e vacanze" diffuse con i file micro.STAT e con i file per la ricerca MFR <http://www.istat.it/it/prodotti/microdati>.

45 Si veda <http://www.istat.it/it/archivio/viaggi>.

46 <http://bit.fieramilano.it/>.

per tutti gli operatori del settore. È importante diffondere in modo tempestivo le stime sul turismo per fornire agli addetti ai lavori e, più in generale, a tutti gli utenti, un quadro di riferimento informativo ufficiale.

Le stime definitive dell'indagine sono diffuse sotto forma di tavole multidimensionali sul *datawarehouse* dell'Istituto I.Stat<sup>47</sup> nella sezione "Cultura, comunicazione, viaggi/Viaggi". I.Stat è la banca dati, sempre aggiornata, delle statistiche correntemente prodotte dall'Istituto Nazionale di Statistica. Le statistiche sono ricercabili per tema o per parola chiave. I dati possono essere esportati in diversi formati come .xls, .csv, .sdmx; è inoltre possibile creare tabelle personalizzate scegliendo le variabili, il periodo di riferimento e il tipo di dati.

L'indagine diffonde, a titolo gratuito, anche i file dei microdati per la ricerca (MFR) su individui, viaggi ed escursioni. Questi archivi, dal 2014, hanno sostituito i file standard (diffusi solo per i viaggi e gli individui) che, privi di elementi identificativi diretti e trattati secondo modalità che non ne permettessero il collegamento con i singoli rispondenti, venivano rilasciati, per finalità di studio e ricerca, su richiesta motivata. Il trattamento di protezione della riservatezza applicato ai file standard ne limitava in parte i contenuti informativi rispetto al file di origine. Anche gli archivi per la ricerca, che hanno sostituito i file standard, sono soggetti a particolari trattamenti statistici che limitano l'identificabilità del rispondente, pur mantenendo un elevato livello di dettaglio informativo. Questi possono essere richiesti esclusivamente da soggetti appartenenti a università o istituti o enti di ricerca a cui si applica il Codice di deontologia per i trattamenti statistici effettuati al di fuori del Sistan<sup>48</sup> oppure da altri soggetti in possesso dei requisiti chiesti da Eurostat ai fini della fornitura di file di microdati<sup>49</sup>. Requisito essenziale per il rilascio di questa tipologia di file è la presentazione di un progetto di ricerca attestante che il trattamento dei dati sia finalizzato a effettivi scopi statistici o scientifici, pertinenti ai dati richiesti.

Dal 2014 sono diffusi, ogni anno, anche i file ad uso pubblico (micro.STAT) che consistono in collezioni di dati elementari scaricabili liberamente e gratuitamente dal sito Istat da parte di qualunque utente. Tali file sono estratti dai corrispondenti file per la ricerca, cui vengono applicate ulteriori misure di protezione della riservatezza, che implicano necessariamente una riduzione del contenuto informativo. Ciò si è reso necessario per ridurre maggiormente il rischio di identificazione degli intervistati facenti parte del campione, a fronte della possibilità di utilizzo di questi file da parte di un bacino di utenza maggiore di quello della comunità scientifica.

I dati d'indagine popolano anche il laboratorio ADELE (Laboratorio per l'Analisi dei Dati ELEMENTARI), ove ricercatori e studiosi possono accedere ai file di microdati di tutte le rilevazioni dell'Istat (privi di dati identificativi, sensibili e giudiziari). Il laboratorio è un ambiente "sicuro" in cui ricercatori di università, istituti, enti di ricerca o organismi, cui si applica il Codice di deontologia per i trattamenti statistici effettuati al di fuori del Sistan, possono condurre analisi statistiche che necessitino dell'utilizzo di dati elementari, laddove non siano sufficienti le informazioni già disponibili (*datawarehouse* I.Stat, produzione editoriale, file di microdati, elaborazioni personalizzate).

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, con riferimento all'anno precedente e per i soli residenti di 15 anni e più: 1) le stime sulla partecipazione al turi-

47 I.STAT è raggiungibile all'indirizzo <http://dati.istat.it/>.

48 Allegato A.4 del D.lgs. 30 giugno 2003, n. 196.

49 In adempimento al Regolamento (CE) n. 557/2013.

smo, in termini di turisti nell'anno; 2) le stime sulle escursioni outbound; 3) i microdati sui viaggi. Con cadenza triennale, invece, sono trasmesse le stime sulle escursioni domestiche e alcune variabili di approfondimento sui viaggi nel file dei microdati.

Le informazioni trasmesse popolano il data base di Eurostat, consultabile al link <http://ec.europa.eu/eurostat/web/tourism/>.

Annualmente i dati d'indagine contribuiscono alle pubblicazioni editoriali dell'Istituto (Italia in cifre, Noi Italia e Annuario Statistico Nazionale). Inoltre, con la stessa cadenza, alcuni macroaggregati principali confluiscono in pubblicazioni internazionali sul turismo dell'Ocse<sup>50</sup> e dell'Unwto<sup>51</sup>.

Dal 2015 è pubblicata, a novembre di ogni anno, una Statistica Report sul "Movimento turistico in Italia" riferita all'anno precedente e integrata con i dati dell'indagine Movimento dei clienti negli esercizi ricettivi<sup>52</sup>. Questo comunicato offre un interessante approfondimento sul turismo interno nelle strutture ricettive collettive, tenendo in considerazione sia il punto di vista dell'offerta turistica sia quello della domanda. In tal modo, ad un'analisi quantitativa, si affianca una valutazione di tipo qualitativo del dato sul movimento dei clienti.

Occasionalmente, le stime sul turismo vengono diffuse anche nel "Rapporto Annuale sulla situazione del Paese"<sup>53</sup> che l'Istat presenta ogni anno, a maggio, alla Presidenza del Consiglio dei Ministri.

### 3.7 CATI e CAPI a confronto: un quadro di sintesi

Il Prospetto 3.12 sintetizza le principali differenze tra l'indagine CATI Viaggi e vacanze e l'indagine CAPI sulle Spese delle famiglie, ampiamente discusse nei paragrafi precedenti.

50 Per maggiori informazioni si veda <http://www.oecd.org/cfe/tourism/oecd-tourism-trends-and-policies-20767773.htm>.

51 Per maggiori informazioni si veda <http://statistics.unwto.org/publication/compendium-tourism-statistics-2017-edition>.

52 Per maggiori informazioni si veda <http://www.istat.it/it/archivio/205128>.

53 L'ultimo contributo disponibile è presente nell'edizione del 2017, Capitolo 3 [http://www.istat.it/it/files/2017/05/RA2017\\_cap3.pdf](http://www.istat.it/it/files/2017/05/RA2017_cap3.pdf).

54 Comuni centro dell'area metropolitana (Torino, Milano, Venezia, Genova, Bologna, Firenze, Roma, Napoli, Bari, Palermo, Catania e Cagliari); comuni che gravitano intorno al centro dell'area metropolitana; comuni aventi fino a 2.000 abitanti; comuni con 2.001-10.000 abitanti; comuni con 10.001-50.000 abitanti; comuni con oltre 50.000 abitanti.

55 Area metropolitana, Periferia dell'area metropolitana e Comuni con 50.000 abitanti e più, Comuni con meno di 50.000 abitanti.

56 Per il periodo 2017-2019.

### 3. La nuova rilevazione delle informazioni sulla domanda turistica

**Prospetto 3.12 - Sintesi delle differenze tra Indagine CATI Viaggi e vacanze e indagine CAPI sulle Spese delle famiglie**

	INDAGINE CATI "VIAGGI E VACANZE" 1997-2013	INDAGINE SULLE SPESE DELLE FAMIGLIE 2017
<b>Schema di campionamento</b>	Stratificato a grappoli, in cui ciascun grappolo è un indirizzo telefonico presente nell'archivio di selezione. Stratificazione effettuata, all'interno di ogni regione geografica, secondo la tipologia socio-demografica dei comuni di appartenenza degli indirizzi.	Casuale a due stadi (comuni-famiglie) con stratificazione delle unità di primo stadio
<b>Lista di selezione delle unità campionarie</b>	Archivio informatizzato ufficiale degli abbonati alle utenze private di telefonia fissa	LAC
<b>Domini di studio territoriali</b>	Tre ripartizioni, tipologia del comune (a)	Cinque ripartizioni, tipologia del comune (b), Regione
<b>Domini di studio temporali</b>	Trimestre	Trimestre
<b>Numerosità campionaria teorica</b>	Circa 14.000 famiglie	Circa 19.500 mila famiglie <sup>(c)</sup>
<b>Periodicità della rilevazione</b>	Continua, trimestrale	Continua, mensile
<b>Unità di rilevazione</b>	Famiglia di fatto definita come l'insieme di persone coabitanti e legate da vincoli di matrimonio, parentela, affinità, adozione, tutela o affettivi.	Famiglia di fatto definita come l'insieme delle persone coabitanti, legate da vincoli di matrimonio o parentela, affinità, adozione, tutela o affettivi e che partecipano alla spesa familiare e/o condividono il reddito familiare.
<b>Rispondente</b>	Un componente della famiglia di almeno 18 anni che riporta i propri spostamenti e in modalità proxy gli spostamenti degli altri membri della famiglia	Un componente della famiglia di almeno 18 anni che riporta i propri spostamenti e in modalità proxy gli spostamenti degli altri membri della famiglia; sono ammesse interviste proxy rilasciate da persone esterne alla famiglia
<b>Tecnica di rilevazione</b>	CATI	Mista in modalità CAPI (diario cartaceo autocompilato e due interviste dirette)
<b>Modalità di svolgimento della rilevazione</b>	Telefonica	Per ciascuna famiglia selezionata sono previste tre visite a domicilio da parte del rilevatore: <i>prima visita</i> - il rilevatore effettua un'intervista iniziale in cui sono presenti i quesiti sui turisti nell'anno e consegna il diario alla famiglia; <i>seconda visita</i> - a metà del periodo di compilazione del diario, viene effettuata una visita per verificare come procede la compilazione; <i>terza visita</i> - nei primi giorni del mese successivo a quello di compilazione del diario, si effettua la visita finale per ritirare il diario e per svolgere una seconda intervista (intervista finale), nella quale è presente il Focus viaggi.
<b>Rete di rilevazione</b>	Costituita dai rilevatori, selezionati e coordinati dalla società incaricata e formati da Istat	Costituita dai rilevatori, selezionati e coordinati dalla società incaricata e formati da Istat
<b>Periodo di riferimento delle variabili</b>	Trimestre, anno	Mese, trimestre, anno
<b>Tecnica di correzione dei dati</b>	Deterministica e da donatore (RIDA)	Deterministica, probabilistica e da modello (Procedura Famiglie, SCIA, IVEware)
<b>Sistema di ponderazione</b>	<ol style="list-style-type: none"> <li>1. Popolazione residente per le 6 tipologie comunali;</li> <li>2. Popolazione residente per sesso nelle 20 regioni italiane;</li> <li>3. Popolazione residente per classi di età;</li> <li>4. Famiglie unipersonali di età inferiore a 65 anni, per sesso;</li> <li>5. Famiglie unipersonali di 65 anni e più, per sesso;</li> <li>6. Famiglie con più di un componente per classi di ampiezza</li> </ol>	<ol style="list-style-type: none"> <li>1. Popolazione residente per ripartizione, sesso e classi di età;</li> <li>2. Popolazione residente per regione;</li> <li>3. Popolazione residente per ripartizione e tipologia di comune</li> <li>4. Famiglie residenti per regione;</li> <li>5. Famiglie residenti per ripartizione e tipologia di comune</li> <li>6. Popolazione straniera residente per ripartizione e sesso</li> <li>7. Popolazione di 15 anni e più per ripartizione, condizione e posizione nella professione</li> <li>8. Popolazione per ripartizione e mese;</li> <li>9. Famiglie per ripartizione e mese.</li> </ol>

(a) Comuni centro dell'area metropolitana (Torino, Milano, Venezia, Genova, Bologna, Firenze, Roma, Napoli, Bari, Palermo, Catania e Cagliari); comuni che gravitano intorno al centro dell'area metropolitana; comuni aventi fino a 2.000 abitanti; comuni con 2.001-10.000 abitanti; comuni con 10.001-50.000 abitanti; comuni con oltre 50.000 abitanti.

(b) Area metropolitana, Periferia dell'area metropolitana e Comuni con 50.000 abitanti e più, Comuni con meno di 50.000 abitanti.

(c) Per il periodo 2017-2019.



## 4. LA RICOSTRUZIONE DELLE SERIE STORICHE<sup>1</sup>

### 4.1 Il periodo di sovrapposizione e i risultati della nuova indagine

Le molteplici innovazioni sin qui descritte hanno avuto come conseguenza il verificarsi di un break di serie delle principali stime. Tuttavia, la programmazione di un periodo di sovrapposizione dell'indagine CAPI sulle Spese delle famiglie con la rilevazione CATI ha consentito sia di verificare l'attendibilità e l'accuratezza delle nuove stime su uno stesso periodo di riferimento, sia di effettuare la ricostruzione dei principali aggregati sulla domanda turistica, al fine di ovviare ai problemi di continuità e dare la possibilità agli utenti di fare analisi di breve, medio e lungo periodo.

La nuova indagine è stata condotta parallelamente all'indagine CATI durante il terzo e il quarto trimestre del 2012 e per tutto il 2013.

In generale, rispetto all'indagine CATI, la rilevazione CAPI ha stimato un maggior numero di eventi di viaggio (+11,4% nel 2013) soprattutto tra le tipologie più "critiche" dal punto di vista dell'effetto ricordo, cioè le vacanze brevi e i viaggi di lavoro (nel 2013, rispettivamente +34,8% e +38,1%) (Tavola 4.1). La riduzione del periodo di riferimento (da 3 mesi a 1 mese) ha indubbiamente favorito, per questi aggregati, il ricordo di eventi meno importanti o di durata inferiore (si pensi a quanti fanno dei week-end di vacanza e a quanti, come i liberi professionisti, sono soliti viaggiare molto per lavoro).

**Tavola 4.1 - Viaggi e notti per tipologia. Secondo semestre 2012 e anno 2013. Valori assoluti in migliaia e variazioni % CAPI sul corrispondente periodo CATI (a)**

PERIODO	VACANZA BREVE (1-3 notti)			VACANZA LUNGA (4 o più notti)			VIAGGI DI LAVORO			TOTALE VIAGGI		
	CATI	CAPI	Var %	CATI	CAPI	Var %	CATI	CAPI	Var %	CATI	CAPI	Var %
Viaggi												
2° semestre 2012	14.174	17.595	+24,1*	27.309	25.640	-6,1	4.429	5.277	19,1	45.912	48.512	5,7
2013	24.782	33.400	+34,8*	32.668	29.073	-11	5.703	7.876	+38,1*	63.154	70.350	+11,4*
Notti												
2° semestre 2012	28.287	33.593	18,8	325.921	289.488	-11,2	14.249	19.783	<b>+38,8*</b>	368.457	342.863	-6,9
2013	48.418	61.659	27,3	353.237	284.797	<b>-19,4*</b>	15.471	24.154	<b>+56,1*</b>	417.126	370.611	-11,2

(a) Nei confronti spazio-temporali è necessario tenere conto dell'errore campionario poiché limitate differenze tra i valori osservati nel campione possono non essere statisticamente significative in ragione di tale errore. Le differenze statisticamente significative tra le due indagini (cioè statisticamente diverse da zero) sono indicate con \*

Inoltre, va sottolineato che, sebbene in ambedue le indagini l'intervista sia proxy, nella nuova indagine essa si svolge presso l'abitazione della famiglia, in un momento concordato in cui vi sia la presenza di tutti i componenti. Ciò può facilitare il ricordo dei singoli eventi di viaggio di ognuno.

<sup>1</sup> I paragrafi 4.1, 4.2 e 4.5 sono stati curati e redatti da M. Di Torrice. Il paragrafo 4.3 è stato curato e redatto da A. Spizzichino. Il paragrafo 4.4 è stato curato e redatto da C. Graziani.

L'effetto positivo della riduzione del periodo di riferimento e della nuova organizzazione dell'indagine è di rilievo anche nel caso delle stime delle visite in giornata che, nel 2013, hanno fatto registrare un incremento del 61,6% rispetto all'indagine CATI: il 59% in più tra quelle dirette entro i confini nazionali e addirittura il quadruplo di visite all'estero.

Diversamente, le vacanze lunghe non mostrano significative differenze nell'ammontare complessivo rilevato durante il periodo di sovrapposizione, sebbene nei trimestri estivi di entrambi gli anni di sovrapposizione le stime fornite dalla nuova indagine CAPI siano inferiori rispetto a quelle prodotte dalla rilevazione CATI. In particolare, nel trimestre estivo del 2013, la stima CAPI delle vacanze lunghe estive è significativamente minore (-23,7% rispetto all'indagine CATI).

Nel corso del periodo di sovrapposizione tra le due indagini, diverse analisi di approfondimento sono state condotte per verificare se le tempistiche dell'indagine CAPI potessero avere un impatto negativo per il reperimento della famiglia nella propria abitazione durante i mesi estivi e quindi conseguenze sulla qualità delle stime sui viaggi. In realtà, questo aspetto era stato considerato sin dalla progettazione dell'indagine, considerando tuttavia la necessità di limitare i casi di sovrapposizione tra fasi di rilevazione, per semplificare il lavoro del rilevatore. Quindi, solo per le famiglie che compilavano il diario a luglio e agosto, era stato previsto un numero di giorni superiore, pari a 25, per effettuare l'intervista iniziale rispetto a quello stabilito negli altri mesi dell'anno, pari a 15 giorni (Tavola 4.2). Inoltre, soltanto per le interviste finali delle famiglie che compilavano il diario a luglio, il periodo era stato esteso a 25 giorni invece di 15, poiché l'intervista ricadeva interamente nel mese di maggior irreperibilità delle famiglie (agosto).

**Tavola 4.2 - Numero di giorni per l'intervista iniziale e finale nelle edizioni di indagini CAPI 2012, 2013 e 2017**

EDIZIONE DI INDAGINE	NUMERO DI GIORNI PER IL CONTATTO E L'INTERVISTA INIZIALE				NUMERO DI GIORNI PER L'INTERVISTA FINALE			
	Mese di compilazione del diario				Mese di compilazione del diario			
	Luglio	Agosto	Settembre	Altri mesi	Luglio	Agosto	Settembre	Altri mesi
Indagine CAPI 2012	25	25	15	15	25	10	10	10
Indagine CAPI 2013	25	25	<b>20</b>	<b>20</b>	25	<b>15</b>	<b>15</b>	<b>15</b>
Indagine CAPI 2017	25	25	<b>25</b>	20	25	15	15	15

Tuttavia, nel 2012, la difficoltà maggiore si è verificata nelle interviste iniziali delle famiglie che compilano il diario a settembre, le quali erano contattate dai rilevatori per l'intervista iniziale tra la seconda metà di agosto e la prima metà di settembre. Infatti, il numero di famiglie intervistate è stato inferiore a quello osservato negli altri periodi del 2012, per cui è stato necessario allargare il periodo di intervista iniziale, sebbene tale scelta fosse stata inizialmente scartata perché il periodo sarebbe comunque ricaduto nel mese di agosto. Nel 2013, a seguito delle riflessioni scaturite dall'analisi delle criticità riscontrate nella gestione dei contatti delle famiglie, si è deciso di ampliare il periodo di intervista iniziale per tutti i mesi dell'anno, portandolo da 15 a 20 giorni, fino ad arrivare, per il mese di settembre, ai 25 giorni attualmente previsti. Infine, il numero di giorni disponibili per effettuare l'intervista finale è stato aumentato da 10 a 15 per tutti i mesi dell'anno, senza ulteriori necessità di adeguamento negli anni successivi.

A fronte di questi aggiustamenti, nel 2013, si è comunque riscontrata la diminuzione significativa sopra citata per le vacanze lunghe nel trimestre estivo, rispetto alla stima della vecchia indagine. Nel caso delle vacanze lunghe, si può supporre che il passaggio a un pe-

riodo di riferimento mensile non abbia apportato benefici, probabilmente perché la natura del fenomeno ne favorisce comunque un buon ricordo anche con quesito avente un riferimento temporale a 3 mesi (si tratta di viaggi lunghi che rappresentano, nella maggior parte dei casi, l'unica vacanza nel corso dell'anno). E' ipotizzabile, invece, che la vecchia indagine CATI, basata su un campione di liste di abbonati alla telefonia fissa, fornisca una sovrastima degli eventi di vacanza lunga. Le analisi condotte in passato (cfr. Istat, 2009a e 2009b), infatti, hanno evidenziato che il numero di turisti per vacanza lunga tra le famiglie che possiedono un telefono fisso non riservato, che sono le famiglie raggiungibili dall'indagine telefonica, risulta essere superiore rispetto al totale delle famiglie. Quindi, i risultati della nuova indagine CAPI, che stimano un numero inferiore di vacanze lunghe rispetto all'indagine CATI, sembrano maggiormente rappresentativi, realizzando l'atteso superamento dei problemi di copertura.

### 4.2 Le serie da ricostruire

Il periodo di sovrapposizione ha permesso di applicare tecniche di ricostruzione di serie storiche consolidate in Istituto, per mezzo delle quali i principali indicatori prodotti dalla rilevazione effettuata con tecnica CATI nel periodo 1997-2013 sono stati resi coerenti con quelli prodotti dalla nuova indagine CAPI.

La disponibilità di sei trimestri di sovrapposizione, durante i quali le due rilevazioni sono state condotte in parallelo, ha rappresentato uno strumento chiave per osservare e valutare le differenze tra la nuova e la vecchia indagine.

La ricostruzione ha previsto la produzione delle principali serie sui viaggi e le notti per le tre macro ripartizioni territoriali (Nord, Centro e Mezzogiorno) e per ogni trimestre, a partire dal 1° trimestre 1997 fino al 2° trimestre 2012. Le 18 serie storiche ricostruite sono quindi:

- il numero di notti e il numero di viaggi per lavoro;
- il numero di notti e il numero di viaggi per vacanze brevi (fino a tre notti);
- il numero di notti e il numero di viaggi per vacanze lunghe.

### 4.3 La metodologia

La metodologia adottata per la ricostruzione delle serie si è basata su quanto già fatto in passato per ricostruire le stime dell'indagine sulle Forze di Lavoro, in occasione del passaggio dalla rilevazione trimestrale alla continua (nel 2004).

Come già avvenuto nel caso della suddetta indagine, occorre considerare che si hanno pochi dati a disposizione per l'elaborazione (appena sei trimestri di sovrapposizione), un gran numero di serie da ricostruire e un insieme complesso di vincoli di aggregazione che lega le serie tra loro. Per questi motivi e per esigenze di trasparenza, l'Istituto propende per una procedura semplice ed efficace, facilmente meccanizzabile, escludendo approcci teoricamente più complessi alla ricostruzione, per esempio quelli basati su tecniche come filtro di Kalman, o simili. Gli istituti nazionali di statistica, in generale, preferiscono infatti procedure automatizzate per ridurre al minimo l'intervento a posteriori dei ricercatori, preservando così la trasparenza del processo e l'oggettività del dato.

In particolare, è stato adottato un approccio macro-fondato, model based e per componenti.

L'approccio è denominato macro-fondato poiché i numerosi cambiamenti occorsi nell'indagine hanno riguardato diverse fasi del processo produttivo; l'adozione di un approccio micro-fondato è apparsa quindi inadatta, in quanto non avrebbe consentito di ottenere risultati omogenei. Sfruttando invece il periodo di sovrapposizione, è stato possibile ricostruire direttamente i dati aggregati per ognuna delle 18 serie trovando la relazione funzionale (1) che lega ogni dato stimato dalla nuova indagine al corrispettivo della vecchia.

$$Y_{nt} = f(Y_{vt}) \quad (1)$$

Si è scelto un approccio model based in quanto in grado di esplicitare in modo chiaro ed esaustivo le ipotesi formulate e di elaborare in maniera rapida e generalizzabile un gran numero di serie. Questo approccio recepisce anche le *best practices* in uso nelle banche centrali e nel mondo accademico. In generale, l'aggiustamento stagionale è condotto in maniera semi-automatica, e quindi con minore arbitrarietà da parte dell'analista. Nello specifico, per l'analisi delle serie storiche e l'esplicitazione delle relazioni funzionali tra le componenti della nuova e quelle della vecchia indagine, sono stati utilizzati congiuntamente pacchetti per l'analisi delle serie storiche (JDemetra+)<sup>2</sup> e funzioni avanzate disponibili nei fogli elettronici.

Visto il tipo di fenomeno oggetto dell'indagine, è apparso del tutto fondato ritenere che con il passaggio alla nuova indagine le serie siano interessate non solo da level-shift, ma anche da cambiamenti nel pattern stagionale. In generale, si è valutato che le numerose cause di differenza tra le serie abbiano prodotto effetti che si sono manifestati in modo diverso sulle componenti.

Con l'approccio per componenti, per ciascuna serie sono ricostruite separatamente le tre componenti temporali: il ciclo-trend, che ne rappresenta la dinamica di medio-lungo periodo, la stagionalità, che ne costituisce il pattern infra-annuale, e la componente erratica, di brevissimo periodo.

Analizzando le serie a disposizione, si è osservato che quelle prodotte dall'indagine CATI Viaggi e vacanze sono lunghe abbastanza per poter identificare e stimare modelli ARIMA che le rappresentino e che permettano di scomporle in ciclo-trend, stagionalità e componente erratica.

Per il processo di scomposizione di queste serie, è stata impiegata la procedura TRAMO-SEATS, implementata nel software JDemetra+; si è optato per un processo quasi completamente automatico, imponendo solo di non eseguire la trasformazione logaritmica sulle serie trattate.

La procedura TRAMO-SEATS è la procedura adottata dall'Istat (Istat, 1999) per destagionalizzare dati di carattere congiunturale. Essa nasce tipicamente per l'analisi di dati di natura economica e si compone di due parti: la prima, chiamata TRAMO (Time series Regression with Arima noise, Missing observations and Outliers) è diretta a eliminare dalla serie storica d'interesse effetti deterministici, quali quelli dovuti al diverso numero di giorni

<sup>2</sup> Il programma JDemetra+ è un software sviluppato da EUROSTAT, user friendly in ambiente MS Windows, che implementa entrambe le procedure di destagionalizzazione più usate dagli istituti nazionali di statistica europei: TRAMO-SEATS, appunto, e X12ARIMA. È disegnato principalmente per la gestione di data set di numerose serie, come è il caso delle serie soggette al raccordo nell'indagine Viaggi e Vacanze dell'Istat. Maggiori informazioni sul software sono disponibili al seguente link [http://ec.europa.eu/eurostat/cros/content/software-jdemetra\\_en](http://ec.europa.eu/eurostat/cros/content/software-jdemetra_en).

lavorativi, alla presenza di festività “mobili” (ad esempio, la Pasqua) e a valori anomali. Inoltre, TRAMO identifica e stima i modelli ARIMA che descrivono efficacemente l’evoluzione nel tempo della serie storica.

Una volta che la serie originaria è stata depurata dagli effetti deterministici ottenendo la serie “linearizzata”, quest’ultima viene scomposta nelle sue componenti non osservabili dalla procedura SEATS (Signal Extraction in Arima Time Series). La scomposizione della serie si basa sull’ipotesi che ogni componente segua un particolare modello ARIMA coerente con quello individuato per la serie originaria. Oltre alla scomposizione della parte stocastica, SEATS assegna i vari effetti deterministici individuati da TRAMO alle diverse componenti, ottenendo così le componenti finali. Inoltre, SEATS offre l’opportunità di convalidare i risultati raggiunti sia attraverso una diagnostica molto accurata, sia tramite un’analisi grafica delle componenti non osservabili.

La scomposizione ottenuta per le vecchie serie ( $v$ ) è del tipo:

$$Y_{vt} = CT_{vt} + S_{vt} + E_{vt} \quad (2)$$

dove CT è la componente di ciclo-trend, S la componente stagionale ed E la componente erratica.

Per scomporre invece le serie della nuova indagine CAPI sulle Spese delle famiglie ( $n$ ), lunghe appena sei trimestri, le informazioni provenienti dalle vecchie serie (sei valori di ogni serie) sono state utilizzate per stimare i parametri del modello di relazione tra ramo vecchio e nuovo di ogni serie.

Per quanto detto, le serie possono essere scomposte come segue:

$$y_{nt} = ct_{nt} + s_{nt} + e_{nt} \quad (3)$$

dove  $ct$  è la componente di ciclo-trend,  $s$  la componente stagionale ed  $e$  la componente erratica. L’approccio per componenti prevede che ogni componente del ramo nuovo sia funzione della stessa componente del ramo vecchio:

$$y_{nt} = f_1(CT_{vt}) + f_2(S_{vt}) + f_3(E_{vt}) \quad (4)$$

ponendo il ciclo-trend del ramo nuovo della serie proporzionale a quello del ramo vecchio, una relazione moltiplicativa per la componente stagionale, con un parametro diverso per ogni trimestre  $e$ , infine, di nuovo una relazione proporzionale che lega la componente erratica del ramo nuovo a quella del ramo vecchio, ne consegue che

$$y_{nt} = \alpha * CT_{vt} + \delta_i * \beta_i * S_{vt} + \gamma * E_{vt} \quad (5)$$

dove  $i$  indica il trimestre e  $\delta_i$  è una variabile indicatrice pari a 1 quando il tempo  $t$  corrisponde al trimestre  $i$  e uguale a zero altrimenti.

Risolvendo il sistema di regressione lineare a cinque<sup>3</sup> parametri con il metodo dei minimi quadrati, è possibile stimare i parametri  $\alpha$ ,  $\beta_i$  e  $\gamma$  e ricostruire all’indietro il ramo nuovo,

3 Per le proprietà dell’analisi per componenti la stagionalità si risolve nell’anno, quindi la somma delle quattro stagionalità di ogni singolo anno risulta pari a zero; da ciò è possibile definire un  $\beta_4$  come funzione lineare degli altri tre per cui i parametri della regressione lineare sono  $\alpha$ ,  $\beta_1$ ,  $\beta_2$ ,  $\beta_3$  e  $\gamma$ .

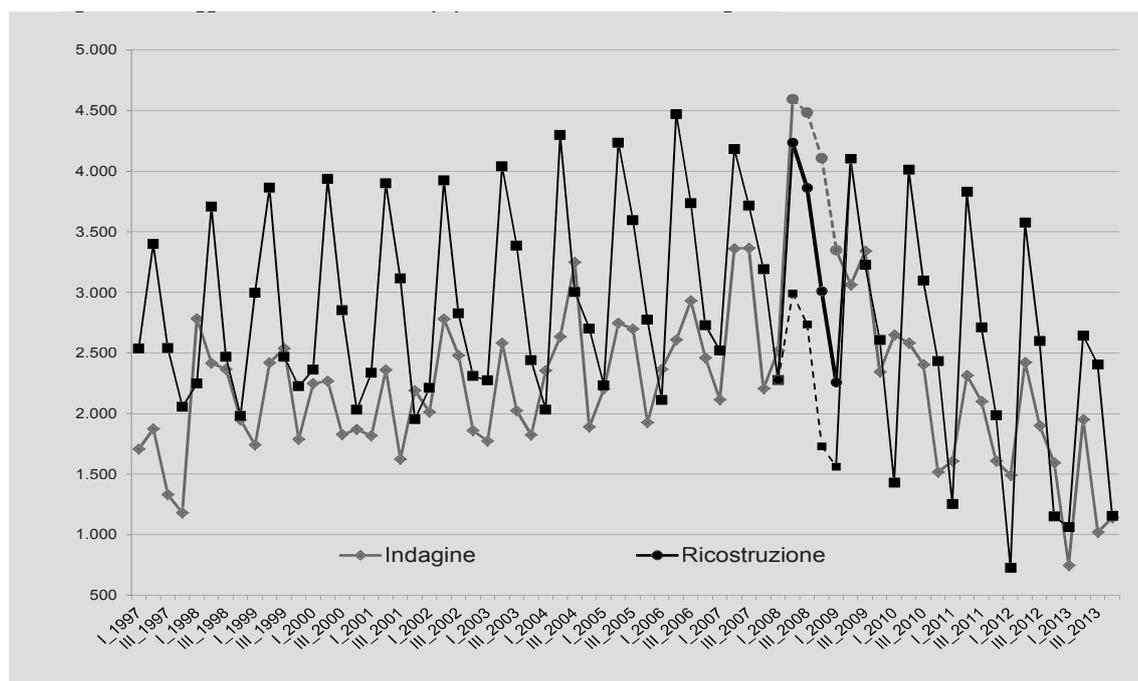
fino al I trimestre 1997.

Nelle varie fasi della ricostruzione vera e propria, è stato necessario affrontare alcune problematiche relative ai dati di input, per le quali sono state adottate delle soluzioni ad hoc prendendo spunto dalla bibliografia disponibile sul tema di ricostruzioni di serie storiche:

- è stata di fondamentale importanza la definizione della stagionalità da associare alle diverse serie da ricostruire; nelle precedenti esperienze, era stata considerata quella derivante dalla vecchia serie e relativa al periodo più recente, che coincideva con quello di sovrapposizione. Per questa ricostruzione si è scelto invece di considerare, come stagionalità di riferimento dei quattro trimestri dell'anno, la media delle stagionalità degli anni 2006 e 2007<sup>4</sup>, in quanto non caratterizzati da un periodo di forte crisi economica, come è stato invece il 2013, e pertanto ritenuti più rappresentativi dell'intero periodo da ricostruire. Tale media è stata ulteriormente pesata, al fine di rispettare il vincolo di risoluzione nell'anno della componente stagionale;
- nel caso di serie in cui, dopo la scomposizione, la componente erratica avesse un peso troppo elevato rispetto alle altre componenti, si è scelto di aggregarla al ciclo-trend e effettuare la ricostruzione seguendo la stessa procedura, ma senza stimare  $\gamma$ . La scelta di unire la componente erratica a quella di ciclo trend, piuttosto che a quella stagionale, è conforme a quanto viene fatto correntemente nella diffusione di dati destagionalizzati, che uniscono appunto il ciclo trend con la componente irregolare;
- in presenza di outliers nella serie, non giustificabili da effettivi eventi straordinari, per eliminarli è stato adottato un correttore della componente erratica che sostituisce, all'errore calcolato per il singolo trimestre, la media degli errori di tutti i corrispondenti trimestri dell'intera serie.

Il grafico 4.1 mostra come è cambiato il risultato della ricostruzione a seguito di questo

Figura 4.1 - Viaggi di vacanza breve della popolazione residente nel Mezzogiorno



4 Attraverso riproporzionamenti è stato rispettato anche il vincolo per cui la somma delle stagionalità calcolate per i quattro trimestri è pari a 0.

tipo di correzione; in particolare, nella serie relativa ai viaggi di vacanza breve nel Mezzogiorno, nella fase di destagionalizzazione il ramo di serie compreso tra il II trimestre del 2008 e il I trimestre del 2009 (parte tratteggiata in grigio) veniva colto come un *transitory change* che determinava una forte distorsione nella serie ricostruita (ramo tratteggiato in nero); applicando la media degli errori di tutti i corrispondenti trimestri dell'intera serie a questi quattro trimestri, le corrispondenti stime ricostruite sono risultate coerenti con il resto della serie.

L'ultima problematica affrontata ha riguardato la scelta tra la definizione diretta o indiretta delle serie aggregate dei viaggi e delle notti per lavoro e per vacanze brevi o lunghe.

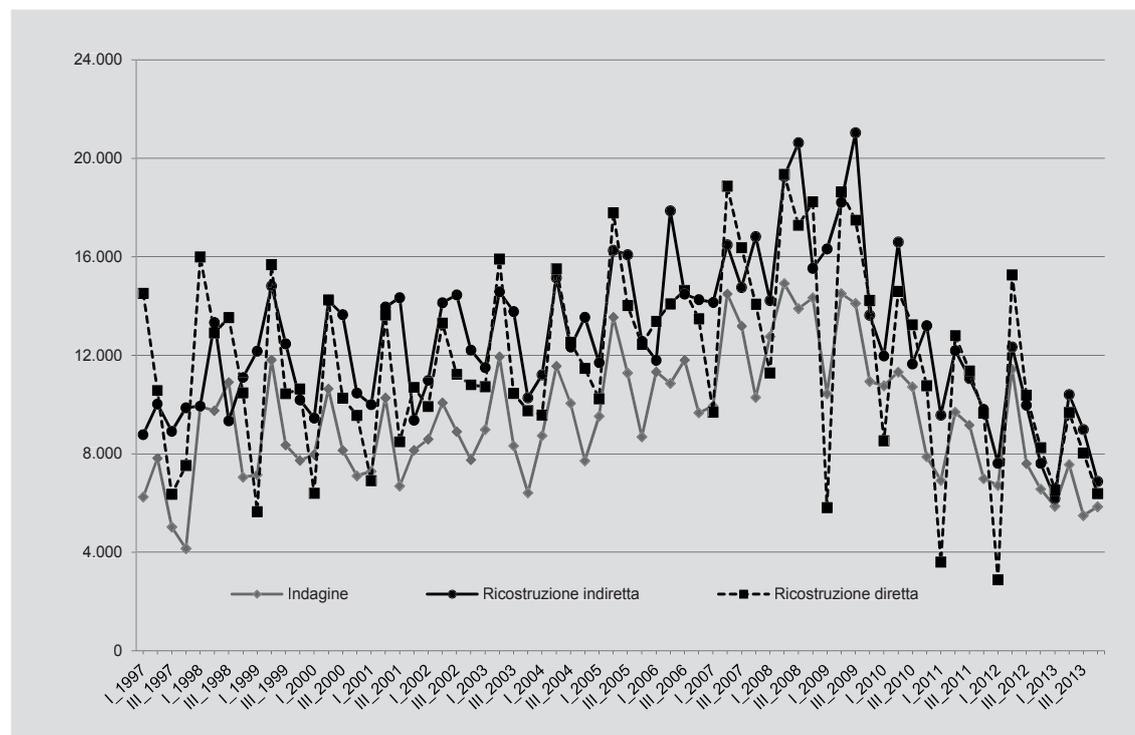
Utilizzando il metodo diretto, vengono prima ricostruite le sei serie a livello nazionale e poi definite quelle con la distinzione territoriale attraverso le tecniche di disaggregazione già richiamate (cfr. nota 2).

Con il metodo indiretto, invece, si ricostruiscono prima le serie con il massimo dettaglio (18 serie), poi si aggregano per ripartizione geografica per ottenere le serie a livello nazionale.

Il secondo metodo è risultato più efficiente, in quanto tiene conto di un maggior numero di informazioni e, in particolare, di tutti gli accorgimenti inseriti nelle fasi della ricostruzione delle singole serie, che altrimenti non potrebbero essere considerati.

Anche dal confronto grafico tra i due metodi di ricostruzione (Figura 4.2) emerge quanto il metodo indiretto sia più efficiente di quello diretto, poiché non produce *outliers* o discontinuità non presenti né nella vecchia serie d'indagine, né nella nuova.

Figura 4.2 - Ricostruzione diretta e indiretta dei viaggi per vacanze brevi, I trimestre 1997-IV trimestre 2013 (dati in migliaia)



#### 4.4 I risultati della ricostruzione

La tecnica di ricostruzione scelta si è dimostrata particolarmente performante rispetto alla tipologia di dati a disposizione, soprattutto in virtù della notevole stagionalità che caratterizza le serie oggetto d'analisi.

I risultati prodotti mostrano come le serie ricostruite mantengano le caratteristiche della vecchia indagine, pur valorizzando le informazioni della nuova con riferimento sia al livello, sia alla stagionalità.

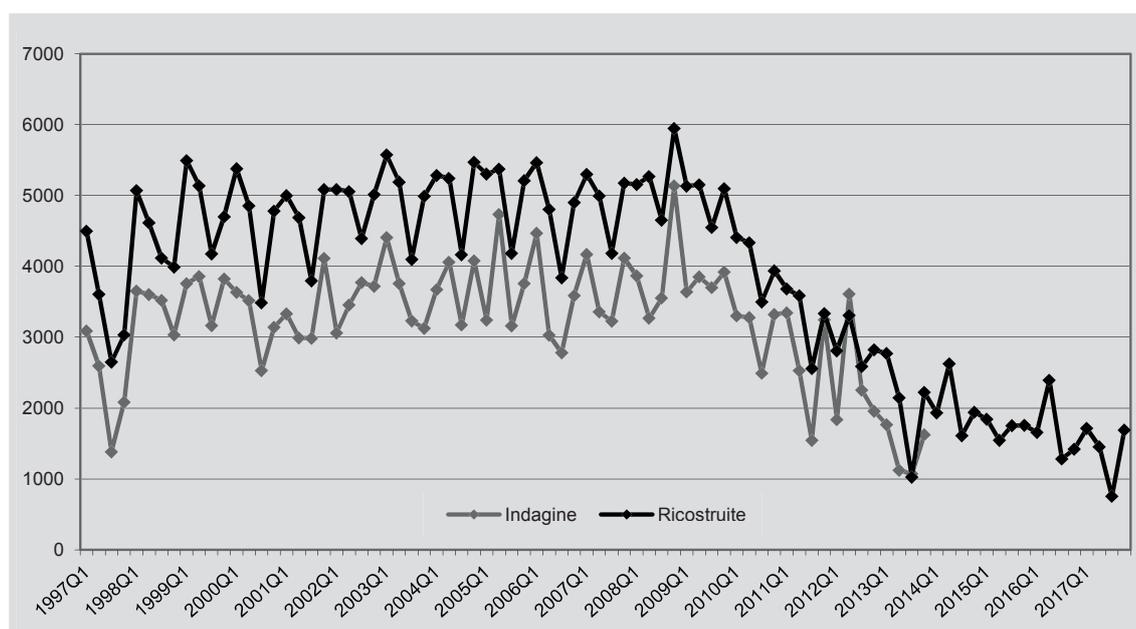
È importante notare come gli andamenti dei dati rilevati con la nuova indagine, emersi nei soli 6 trimestri di sovrapposizione, siano stati poi confermati nei successivi quattro anni (2014-2017); tutto ciò fa venir meno il rischio che i dati utilizzati in fase di ricostruzione, prodotti per la prima volta secondo le nuove regole d'indagine, potessero essere affetti da problematiche dovute all'innovazione.

Di seguito vengono riportati i principali risultati della ricostruzione, relativi ai viaggi a livello nazionale; tali aggregati sono tutti ottenuti come somma dei livelli ripartizionali. Non vengono proposte le corrispondenti analisi per il numero di notti, poiché le dinamiche risultano simili, sebbene su livelli diversi. In particolare, vengono analizzati i viaggi per lavoro, per vacanze brevi e lunghe.

Tutte le figure che seguono hanno la caratteristica di mostrare l'andamento dei dati rilevati con l'indagine CATI Viaggi e vacanze dal primo trimestre del 1997 al quarto trimestre del 2012; dei dati rilevati con l'indagine CAPI sulle Spese delle famiglie dal terzo trimestre del 2012 fino a oggi<sup>5</sup> e, infine, di riportare tutti i valori ricostruiti dall'inizio della serie all'inizio del periodo di sovrapposizione (terzo trimestre 2012-quarto trimestre 2013).

La serie dei viaggi per lavoro (figura 4.3), è quella che meno mostra un andamento stagionale all'interno del singolo anno; necessariamente il terzo trimestre è quello in cui questo tipo di viaggi sono meno frequenti e ciò non sempre veniva colto dalla vecchia inda-

Figura 4.3 - Viaggi per lavoro della popolazione residente in Italia, I trimestre 1997-IV trimestre 2017 (dati in migliaia)



<sup>5</sup> Coincidenti, nel periodo di sovrapposizione, con i valori ricostruiti.

#### 4. La ricostruzione delle sedi storiche

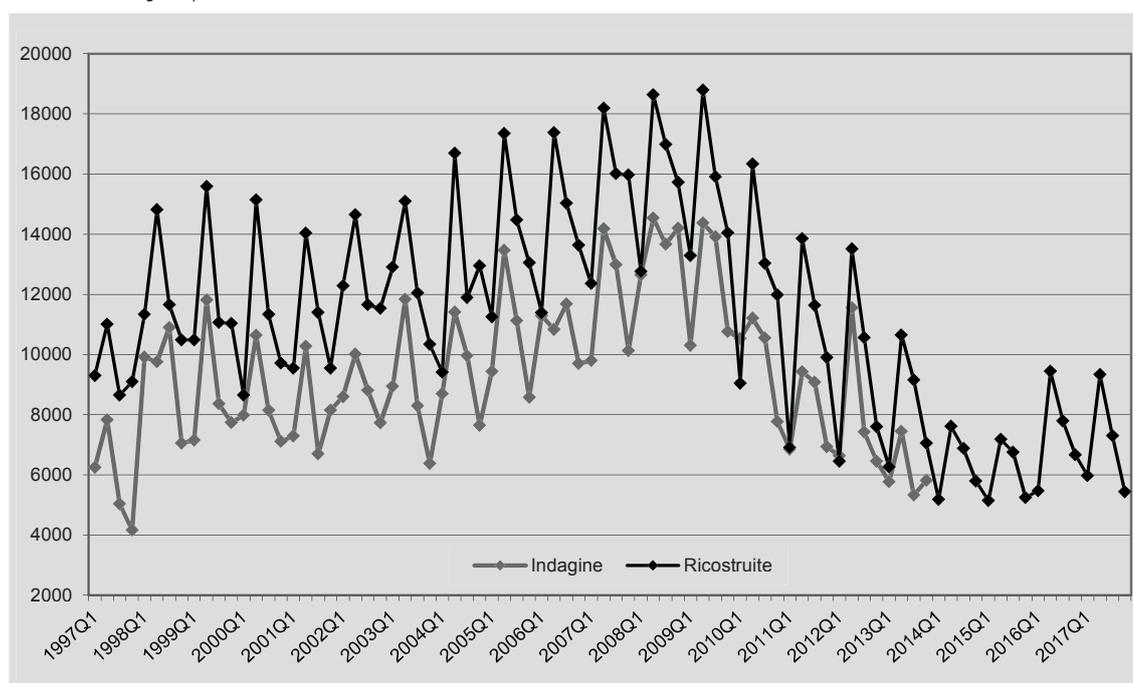
gine<sup>6</sup>. Il dato ricostruito mostra invece un livello sensibilmente inferiore nel terzo trimestre rispetto agli altri in tutta la serie.

Va sottolineato come la ricostruzione riesca anche a depurare la serie da eventuali outliers, cioè trimestri che evidenziavano andamenti non omogenei, riconducendoli ad evidenze più coerenti. Emerge chiaramente, infatti, come alcuni dati fortemente discontinui rispetto al resto della serie (ad esempio il quarto trimestre del 2008 o il primo trimestre 2012) dopo la ricostruzione presentano variazioni più contenute, mantenendo comunque la dinamica d'origine.

La figura 4.4 riporta la serie dei viaggi per vacanze brevi, in cui, come nel caso dei viaggi di lavoro, si osserva che la serie ricostruita si attesta su livelli sostanzialmente superiori a quelli rilevati dall'indagine CATI Viaggi e vacanze. Più nel dettaglio, la nuova indagine e, di conseguenza, la serie ricostruita, evidenzia una stagionalità particolarmente marcata.

Il secondo trimestre di ogni anno si caratterizza per un numero di vacanze di breve durata sensibilmente superiore a quello degli altri trimestri, ciò non era sempre verificato nella vecchia indagine quando, ad esempio nel 2006, nel primo e terzo trimestre i viaggi brevi erano stati più che nel secondo. Le variazioni nel livello e nella stagionalità della serie ricostruita rispetto alla serie in vecchia indagine danno nota delle caratteristiche della nuova indagine e di come la ricostruzione sia riuscita a riprodurle sull'intera serie.

Figura 4.4 - Viaggi per vacanze brevi della popolazione residente in Italia, I trimestre 1997-IV trimestre 2017 (dati in migliaia)



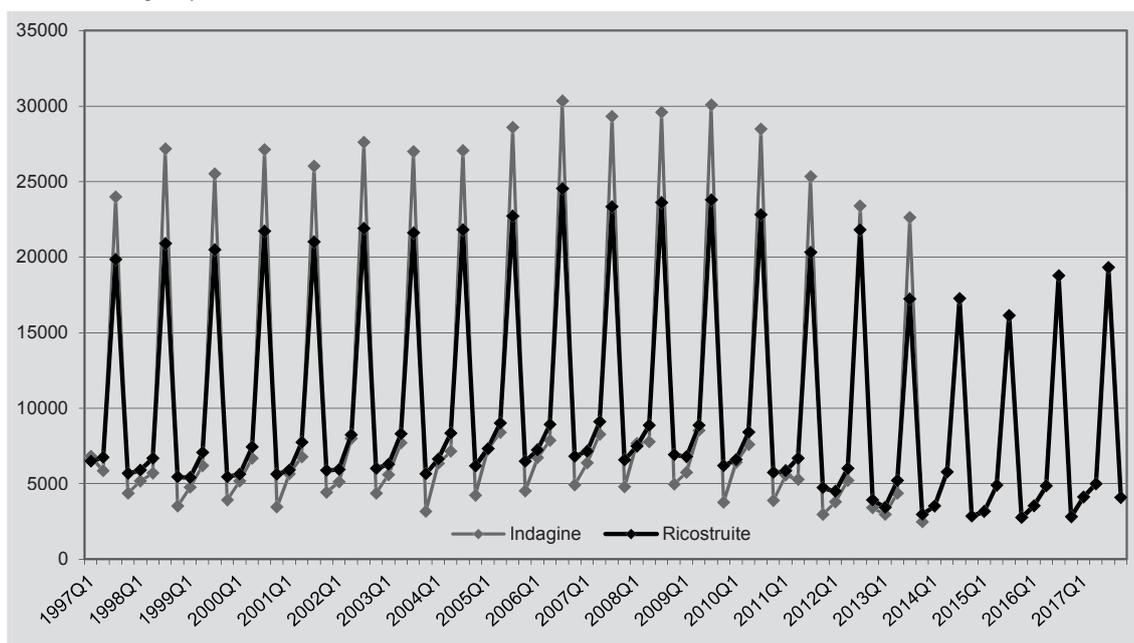
È interessante mostrare inoltre quanto la ricostruzione risulti efficace nel caso di serie caratterizzate da forte stagionalità, come quelle dei viaggi per vacanze lunghe. Come già

6 Ciò è spiegabile anche perché, fino al 2011, l'indagine CATI Viaggi e vacanze includeva, tra i viaggi di lavoro, gli impieghi presso il luogo di destinazione (lavori stagionali, supplenze, altri lavori a tempo determinato), che si concentravano nel trimestre estivo. Tali viaggi sono stati successivamente identificati tramite un apposito item di risposta e quindi decurtati dalle stime, recependo le raccomandazioni IRTS.

osservato nel paragrafo 4.1, il passaggio alla nuova indagine ha determinato forti cambiamenti rispetto al numero di viaggi e notti per vacanze lunghe nel terzo trimestre del 2013.

La figura 4.5 riporta il numero di viaggi per vacanze lunghe dei residenti; emerge che la stima prodotta dalla nuova indagine, nel terzo trimestre del 2013, è pari a circa l'80% della stima fornita dall'indagine CATI. Questa relazione riguardante il terzo trimestre, emersa durante la fase di sovrapposizione, si riproduce sull'intero periodo.

**Figura 4.5 - Viaggi per vacanze lunghe della popolazione residente in Italia, I trimestre 1997-IV trimestre 2017 (dati in migliaia)**



Al contrario, per i primi, secondi e quarti trimestri la stima della nuova indagine non è significativamente diversa da quella dell'indagine CATI. Quindi, è evidente che il problema della copertura precedentemente discusso fosse, nella vecchia indagine CATI, concentrato nel trimestre maggiormente turistico e che la nuova rilevazione sia riuscita a rappresentare meglio il fenomeno delle vacanze lunghe.

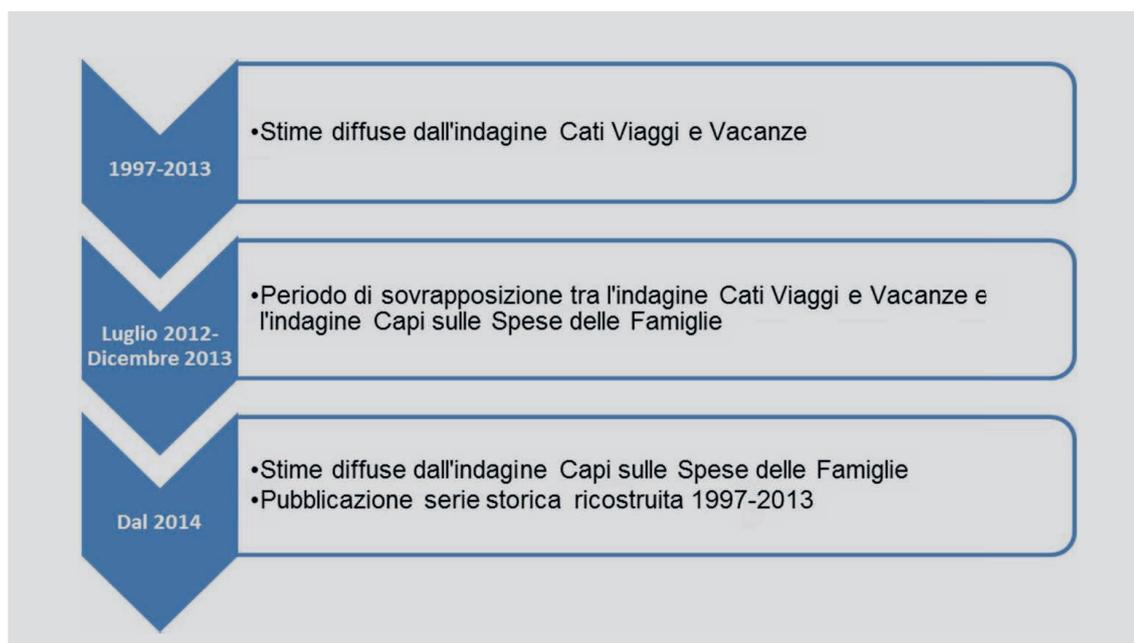
Tutto il lavoro di ricostruzione qui presentato rappresenta un patrimonio informativo fondamentale per lo studio dell'andamento dei viaggi e delle vacanze nel lungo periodo. La disponibilità delle nuove serie ricostruite, depurate dal break corrispondente al passaggio alla nuova rilevazione, ha consentito di continuare a produrre confronti tendenziali dei principali aggregati d'analisi e ha messo a disposizione degli utenti una serie storica che ha permesso di continuare a produrre analisi statistiche di lungo periodo.

Il grande vantaggio della tecnica di ricostruzione utilizzata è inoltre di non precludere, attraverso metodologie ormai standardizzate in Istituto, eventuali ulteriori disaggregazioni delle serie ricostruite mantenendone la coerenza.

### 4.5 La diffusione delle serie ricostruite

Nel 2014 sono state diffuse per la prima volta, nel datawarehouse dell'Istituto, le stime ufficiali sulla domanda turistica provenienti dall'indagine sulle Spese delle famiglie. Contestualmente, sono state diffuse le serie storiche dei principali aggregati, ricostruite per l'arco temporale di 17 anni, dal 1997 al 2013 (Figura 4.6).

Figura 4.6 - Schedulazione della diffusione delle stime della domanda turistica





## 5. PROSPETTIVE FUTURE<sup>1</sup>

### Quale futuro per l'indagine sulla domanda turistica?

La rilevazione sulla domanda turistica Viaggi e Vacanze ha indubbiamente accresciuto la sua portata informativa rispetto al passato, anche grazie al suo innesto nell'impianto dell'indagine CAPI sulle Spese delle famiglie. L'integrazione tra le due rilevazioni, infatti, apre nuove opportunità tematiche, sia per l'analisi della domanda turistica, sia per quella sulle spese per i consumi delle famiglie. I comportamenti di spesa delle famiglie, ad esempio, potranno essere analizzati contestualizzandoli nella loro vita quotidiana, ma anche in occasione di eventi specifici, quali viaggi e/o escursioni; ciò permetterà avanzamenti conoscitivi sul rapporto tra spese destinate al turismo e altre spese familiari, nonché sulla capacità di spesa del turista.

Inoltre, il questionario progettato per l'intervista iniziale sulle Spese delle famiglie consente di raccogliere un ventaglio di informazioni ampio e articolato sulle strutture familiari, sulle caratteristiche socio-demografiche di ciascun componente e sulle condizioni abitative delle famiglie. Da questo punto di vista, saranno possibili approfondimenti sui comportamenti turistici in una nuova prospettiva, quella familiare, che di fatto si affiancherebbe alle due tradizionali prospettive di analisi della domanda turistica, cioè quella individuale, che rileva i flussi dei turisti e le loro peculiarità e quella che considera gli eventi (viaggi, notti ed escursioni) in termini di numero, caratteristiche e spese destinate<sup>2</sup>.

L'utilizzo di procedure statistiche che ricostruiscono i nuclei all'interno delle famiglie sulla base delle relazioni di parentela tra i vari componenti, consentirà lo studio delle scelte e dei comportamenti di viaggio delle diverse tipologie familiari e dei membri che le compongono. In questo modo si potrà ricostruire il contesto socio-economico e familiare all'interno del quale si muove il turista.

Nuove opportunità e sfide per le statistiche sul turismo vengono poi dai *Big Data*, cioè dai dati originati dall'utilizzo degli strumenti digitali, che registrando eventi, spesso comportamenti spontanei, ampliano le possibilità di analisi e la disponibilità di dati tempestivi generati a un costo estremamente contenuto. In particolare, l'utilizzo quotidiano dei telefoni cellulari da parte degli individui crea un flusso enorme di dati di "posizionamento"; i cosiddetti *mobile positioning data*, infatti, contengono informazioni sugli eventi di chiamata e/o connessione a internet da cellulare, che consentono di individuare la *real time location* di un individuo. Tali dati sono registrati dai gestori di telefonia mobile (MNOs - *Mobile Network Operators*) come *Call Detail Records* e *Data Detail Records* (CDRs/DDRs); la loro acquisizione per finalità statistiche consentirebbe di disporre di un potenziale enorme di informazioni sugli spostamenti territoriali degli individui, in generale, e sui movimenti turistici, in particolare.

<sup>1</sup> Il capitolo è stato curato e redatto da M. Sabato.

<sup>2</sup> Un'analisi familiare dei comportamenti turistici è stata condotta in via sperimentale nel 2008 utilizzando i dati degli anni 2006 e 2007 dell'indagine CATI "Viaggi e Vacanze". In tale occasione grazie all'applicazione della Procedura Famiglia, che ha consentito di ricostruire i nuclei all'interno delle famiglie, è stato possibile studiare le scelte e i comportamenti di viaggio delle diverse tipologie familiari e dei membri che le compongono (Baiocchi, Dattilo, 2008).

L'accesso e l'utilizzo dei *mobile positioning data* a fini statistici richiede la risoluzione di diverse questioni riguardanti la dotazione tecnologica e le competenze per l'acquisizione e l'elaborazione di tali dati, i modelli di business e di accordo con i MNOs per la fornitura, in un quadro di rigorosa tutela della privacy e di riservatezza del dato. Un ulteriore tema da affrontare riguarda la comunicazione al cittadino in merito all'uso dei dati telefonici come ulteriore fonte, affinché tale uso non venga percepito, in maniera errata, come forma di controllo o monitoraggio delle abitudini personali degli individui<sup>3</sup>.

Oltre a tali questioni di carattere generale, l'elaborazione dei dati di telefonia mobile per produrre indicatori sul turismo richiede il superamento di problematiche sia concettuali, per rispettare le definizioni statistiche internazionali sul turismo, sia metodologiche di rappresentatività. Un aspetto cruciale rimane la necessità di identificare correttamente la popolazione di interesse, distinguendo le attività turistiche da quelle non turistiche. Ciò passa principalmente attraverso la corretta identificazione del luogo di residenza e dunque dell'ambiente abituale degli utenti di telefonia mobile, al fine di caratterizzare i movimenti sul territorio degli individui, differenziando tra viaggi abituali e spostamenti turistici, tra pendolari e turisti, ecc. Un ulteriore aspetto problematico è quello della copertura: i *mobile positioning data* sono affetti da problemi di sottocopertura dovuti, ad esempio, al fatto che non tutti i turisti utilizzano i cellulari tout court o in viaggio, ma soprattutto al fatto che alcune categorie di turisti (ragazzi o bambini, per esempio) non dispongono di telefono cellulare. Problemi di sovracopertura, invece, sono causati dall'eventuale utilizzo, in viaggio, di più di un mobile *device* e/o scheda SIM di cui il turista non sempre è l'intestatario.

Va sottolineato, comunque, che, allo stato attuale, le informazioni desumibili dalla fonte *mobile phone data* non possono certamente rispondere alle esigenze espresse dal Regolamento per le Statistiche del Turismo; tuttavia, i vantaggi che un utilizzo sistematico di questi dati potrebbe garantire in futuro (economicità, tempestività, accuratezza, riduzione del fastidio statistico) impongono di esplorarne le potenzialità attraverso sperimentazioni ad hoc.

In questa prospettiva si colloca lo studio progettuale "*Sperimentazione per l'utilizzo di big data su telefonia mobile nell'ambito delle statistiche sulla domanda turistica*"<sup>4</sup>. La prima fase di questa sperimentazione è stata dedicata alla raccolta di informazioni ausiliarie sull'utilizzo del cellulare in viaggio. Infatti, se la qualità dei *mobile phone data* per le statistiche sul turismo è influenzata dal tipo e dall'intensità d'uso del cellulare in viaggio, è necessario disporre di informazioni ausiliarie sui comportamenti degli utenti di telefonia mobile durante uno spostamento turistico per trattare e correggere il *selectivity bias* eventualmente insito in questo tipo di dati. Al tal fine, dal 2015, è stato inserito nel questionario del Focus Viaggi un modulo ad hoc permanente finalizzato a investigare il rapporto tra viaggi e uso delle SIM in viaggio. Nello specifico, i quesiti mirano, per ogni viaggio, a raccogliere informazioni su: 1) l'utilizzo del cellulare in viaggio (quanto il cellulare è utilizzato durante lo spostamento turistico); 2) le modalità di utilizzo del cellulare in viaggio (per effettuare chiamate, inviare

3 La validità e l'accuratezza dei *mobile positioning data* sono state ampiamente dibattute nello studio di fattibilità che Eurostat ha condotto tra il 2012 e il 2014 per promuovere la conoscenza e l'utilizzo dei dati di *mobile phone* nell'ambito delle statistiche sul turismo. Per approfondimenti si veda European Commission, 2014, *Feasibility Study on the Use of Mobile Positioning Data for Tourism Statistics*, Eurostat Contract No. 30501.2012.001-2012.452, voll 1-4, <http://ec.europa.eu/eurostat/web/tourism/methodology/projects-and-studies>.

4 Tale sperimentazione promossa dall'ISTAT è stata inserita nel Piano Statistico Nazionale (STU IST-02669) ed è stata avviata nel 2015 all'interno dell'indagine CAPI "Viaggi e Vacanze". Essa rientra nell'ambito della *Roadmap per l'uso dei Big Data a supporto della produzione di Statistica Ufficiale*, risultato dei lavori della Commissione di Studio sui Big Data istituita dall'Istituto nel febbraio del 2013. Per approfondimenti si veda Baldacci, E., M. Scannapieco, edit by, 2015. *ISTAT's Roadmap for the Adoption of Big Data for Official Statistics*, Roma: ISTAT.

messaggi, connettersi a internet); 3) il numero delle SIM viaggianti; 4) le tipologie di intestatari delle SIM viaggianti (partecipanti/non partecipanti al viaggio, aziende, ecc.).

Le informazioni raccolte possono essere così utilizzate, in primo luogo, per delineare il contesto nel quale ci si muove, dal momento che in Italia non ci sono indagini che forniscano stime sull'uso del cellulare durante un viaggio. In secondo luogo, il patrimonio informativo raccolto può consentire di valutare l'affidabilità e il grado di copertura dei *mobile phone data* rispetto al fenomeno del turismo. Infine, questi dati forniscono la base informativa per calcolare, eventualmente, coefficienti correttivi da applicare in futuro ai CDRs/DDRs per stimare correttamente i flussi turistici<sup>5</sup>.

La sfida principale per l'uso dei dati di telefonia mobile per le statistiche sul turismo è quella di passare dalla sperimentazione alla produzione e tale passaggio coinvolge diversi aspetti, tra cui la necessità di dotarsi di nuove infrastrutture tecnologiche e organizzative, nonché di nuove competenze metodologiche. Nondimeno, uno step fondamentale è la conclusione degli accordi con i gestori di telefonia mobile per la fornitura dei dati. E' indispensabile che gli MNOs siano coinvolti in una partnership collaborativa che rappresenti una "Win-win opportunity" sia per la statistica ufficiale che per i providers.

Nel futuro delle statistiche sul turismo, si può ipotizzare uno scenario in cui i *mobile phone data* e, più in generale, i *Big Data*, si affianchino alle fonti tradizionali, al fine di ampliare le opportunità di analisi, aumentare la tempestività e il dettaglio spazio-temporale delle informazioni, contribuendo così a migliorare la qualità delle stime.

La chiave che garantisce guadagni in termini di accuratezza, coerenza e completezza delle informazioni statistiche prodotte è nella capacità di integrare dati che provengono da fonti differenti: "We can and must move from a paradigm of producing the best estimates possible from a survey to that of producing the best possible estimates to meet user needs from multiple data sources" (C. Citro, 2014, p.1). Per raggiungere questo obiettivo, è necessario sviluppare nuove capacità, soprattutto nella direzione che porta a una completa integrazione tra fonti tradizionali e innovative sulla domanda e sull'offerta turistica.

5 I primi risultati della sperimentazione sono stati presentati al 14th Global Forum on Tourism Statistics e all'NTTS 2017. Per approfondimenti si veda Dattilo B., Radini R., Sabato M., 2016, *How many SIM cards in your luggage? A strategy to make mobile phone data usable in tourism statistics*, in Proceedings of the 14th Global Forum on Tourism Statistics, Venice, Italy, November 23-25, 2016 disponibile all'indirizzo [http://tsf2016venice.enit.it/images/articles/Papers\\_Forum/1.2\\_How%20many%20SIM%20in%20your%20luggage%20A%20strategy%20to%20make%20mobile%20phone%20data%20usable%20in%20tourism%20statistics.pdf](http://tsf2016venice.enit.it/images/articles/Papers_Forum/1.2_How%20many%20SIM%20in%20your%20luggage%20A%20strategy%20to%20make%20mobile%20phone%20data%20usable%20in%20tourism%20statistics.pdf) e Dattilo B., Sabato M., 2017, *Travelling SIM and Trips: an approach to make mobile phone data usable in tourism statistics*, in Proceedings of the NTTS - New Techniques and Technologies for Statistics, 13-17 March 2017, Brussels, Belgium, disponibile all'indirizzo [https://www.conference-service.com/NTTS2017/documents/agenda/data/abstracts/abstract\\_42.html](https://www.conference-service.com/NTTS2017/documents/agenda/data/abstracts/abstract_42.html).



## GLOSSARIO

### **Alloggio prevalente del viaggio**

Alloggio dove si è trascorso il maggior numero di notti.

### **Destinazione prevalente del viaggio**

Località italiana o estera dove si è trascorso il maggior numero di notti.

### **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 gli spostamenti di andata e ritorno. Sono escluse le eventuali escursioni che si effettuano durante i soggiorni di vacanza/lavoro, poiché il luogo di partenza e 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 esempio, 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, figurano: 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.

### **Mezzo di trasporto prevalente del viaggio**

Mezzo con cui è stata coperta la maggiore distanza durante il viaggio.

### **Notti**

Numero dei pernottamenti che compongono la durata di un viaggio, trascorsi dai turisti residenti presso qualunque tipologia di alloggio (strutture ricettive collettive e alloggi privati).

### **Prenotazione del viaggio**

Si intende la prenotazione dell'alloggio turistico e/o del trasporto che può essere fatta direttamente presso la struttura ricettiva o la compagnia di viaggio (prenotazione diretta), oppure presso un'agenzia o tour operator o tramite internet.

**Tipo di attività**

Per le vacanze di piacere/svago, questa informazione le connota in base all'attività prevalente svolta.

**Tipo di luogo visitato**

Per un viaggio di vacanza, questa informazione connota le caratteristiche del luogo visitato (città, mare, campagna, montagna, altro tipo), oppure se si è trattato di una crociera.

**Turismo**

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. Sono, pertanto, esclusi: il pendolarismo per lavoro, studio o per obblighi di famiglia, i movimenti migratori, anche stagionali, i trasferimenti di residenza (anche se temporanei), inclusi i trasferimenti dei diplomatici e dei militari (e delle loro famiglie), nonché i movimenti dei rifugiati, dei nomadi, ecc.

**Turista**

Persona che ha effettuato uno o più viaggi nel periodo di riferimento dell'indagine.

**Vacanza breve**

Viaggio di vacanza con durata del soggiorno inferiore a quattro pernottamenti.

**Vacanza lunga**

viaggio di vacanza con durata del soggiorno di quattro o più pernottamenti.

**Viaggio**

Spostamento realizzato, per turismo di vacanza o per ragioni di lavoro, fuori dal luogo dove si vive abitualmente e che comporta almeno un pernottamento nel luogo visitato. Vengono esclusi i viaggi e gli spostamenti effettuati nelle località frequentate tutte le settimane con soste di uno o più pernottamenti, nonché i viaggi di durata superiore ad un anno. In questi casi, infatti, il viaggio non costituisce flusso turistico poiché la località visitata viene associata al luogo in cui si vive.

**Viaggio abituale**

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

**Viaggio di vacanza**

Viaggio svolto per uno dei seguenti motivi prevalenti: relax, piacere, svago o ri-

poso (inclusi viaggi fatti per fare cure termali o trattamenti di salute con o senza una prescrizione o indicazione medica); visite a parenti e amici; motivi religiosi e/o di pellegrinaggio.

**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. Sono esclusi gli impieghi presso il luogo di destinazione (lavoro stagionale, supplenze, altri lavori temporanei).



## RIFERIMENTI BIBLIOGRAFICI

- AA.VV., 2000. Guida all'utilizzo di TRAMO-SEATS per la destagionalizzazione delle serie storiche. Documenti ISTAT, n.4/2000, Roma. <https://www.istat.it/it/files/2013/12/Standard-destagionalizzazione.pdf> (28 marzo 2018).
- Baiocchi, B., B. Dattilo, 2008, *Italian households on travel: who travels with who, where, when and how families travel?*, in Proceedings of the 9th Global Forum on Tourism Statistics, Paris, France, November 19-21, 2008.
- Baldacci, E., M. Scannapieco, a cura di, 2015. *ISTAT's Roadmap for the Adoption of Big Data for Official Statistics*, ISTAT, Roma.
- Bethel J. (1989). Sample Allocation in Multivariate Surveys, *Survey Methodology*, 15, 47-57
- Ceccarelli, C., A. Pezone, S. Rosati, 2013. L'utilizzo delle liste anagrafiche comunali nella statistica ufficiale. *Rivista italiana di Economia Demografia Statistica*, Vol. LXVII n.3/4 2013.
- Citro, C., 2014. From multiple modes for surveys to multiple data sources for estimates. *Survey Methodology*, 40(2):137–161. <http://www.statcan.gc.ca/pub/12-001-x/2014002/article/14128-eng.htm> (28 marzo 2018).
- Crossley, T., J. Winter, 2013. *Asking Households About Expenditures: What Have We Learned?* Draft paper. <http://www.nber.org/chapters/c12666.pdf> (28 marzo 2018).
- Dattilo, B., R. Radini, M. Sabato, 2016. *How many SIM cards in your luggage? A strategy to make mobile phone data usable in tourism statistics*. Paper for the 14th Global Forum on Tourism Statistics, November 23-25 2016, Venice, Italy. [http://tsf2016venice.enit.it/images/articles/Papers\\_Forum/1.2\\_How%20many%20SIM%20in%20your%20luggage%20A%20strategy%20to%20make%20mobile%20phone%20data%20usable%20in%20tourism%20statistics.pdf](http://tsf2016venice.enit.it/images/articles/Papers_Forum/1.2_How%20many%20SIM%20in%20your%20luggage%20A%20strategy%20to%20make%20mobile%20phone%20data%20usable%20in%20tourism%20statistics.pdf) (28 marzo 2018).
- Dattilo, B., M. Sabato, 2017. *Travelling SIM and Trips: an approach to make mobile phone data usable in tourism statistics*, paper for the NTTS - New Techniques and Technologies for Statistics, 13-17 March 2017, Brussels, Belgium. [https://www.conference-service.com/NTTS2017/documents/agenda/data/abstracts/abstract\\_42.html](https://www.conference-service.com/NTTS2017/documents/agenda/data/abstracts/abstract_42.html) (28 marzo 2018).
- Deville, J.C., C.E. Särndal, 1992. Calibration Estimators in Survey Sampling. *Journal of the American Statistical Association*, Volume 87, n.418.
- De Vitiis, C., P. Righi, 2012. Evaluations on list undercoverage bias and possible solutions: the case of Istat CATI survey Trips, holidays and daily life. *Rivista di statistica ufficiale*, n. 2-3/2011. <https://www.istat.it/it/files/2012/04/primocapitolo.pdf> (28 marzo 2018).
- Di Fonzo, T, M. Marini. 2011. Simultaneous and two-step reconciliation of systems of time series: methodological and practical issues. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 60 (2): 143–164. <http://www.jstor.org/stable/41057568> (28 marzo 2018).
- Direttiva 97/57/EC del Consiglio del 23 Novembre 1995 relativa alla raccolta di dati statistici nel settore del turismo. <http://eur-lex.europa.eu/eli/dir/1995/57/2004-05-01/ita/pdf> (28 marzo 2018).
- Di Torrice, M., M. Morricone, M. Sabato, 2012. *Beyond the CATI technique: a benchmarking analysis on demand-side surveys in EU and the Italian choice*. Paper for the 11th Global Forum on Tourism Statistics, November 14-16 2012, Reykjavík, Iceland. [http://www.congress.is/11thtourismstatisticsforum/papers/Beyond\\_the\\_CATI\\_technique.pdf](http://www.congress.is/11thtourismstatisticsforum/papers/Beyond_the_CATI_technique.pdf) (28 marzo 2018).
- Eurostat, 2013. JDemetra+ User Manual. [https://ec.europa.eu/eurostat/cros/system/files/jdemetra\\_user\\_guide.pdf](https://ec.europa.eu/eurostat/cros/system/files/jdemetra_user_guide.pdf) (28 marzo 2018).
- Eurostat, 2014. *Methodological manual for tourism statistics – Version 3.1*. <http://ec.europa.eu/eurostat/web/products-manuals-and-guidelines/-/KS-GQ-14-013> (28 marzo 2018).

- European Commission, 2014. *Feasibility Study on the Use of Mobile Positioning Data for Tourism Statistics*, Eurostat Contract No. 30501.2012.001-2012.452, voll 1-4. <http://ec.europa.eu/eurostat/web/tourism/methodology/projects-and-studies> (28 marzo 2018).
- European Commission, 2017, *Tourism statistics: Early adopters of big data? 2017 edition, Statistical working papers*. <http://ec.europa.eu/eurostat/documents/3888793/8234206/KS-TC-17-004-EN-N.pdf/a691f7db-d0c8-4832-ae01-4c3e38067c54> (28 marzo 2018).
- 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.
- Fellegi, P. I., D. Holt. 1976. A systematic approach to automatic edit and imputation. *Journal of the American Statistical Association*, Applications Section, 71:17-35.
- Gatto, R., P. Gennari, N. Massarelli. 2001. La ricostruzione e il riallineamento delle serie storiche delle forze di lavoro 1984 - 1992. Paper for the workshop *Occupazione e disoccupazione in Italia: misura e analisi dei comportamenti*, 15-16 Gennaio 2001, Bressanone, Italia.
- Gatto, R. 2007. *The SUH Approach in Time Series Back-Recalculation*. Paper for the 18th Annual EC2 Conference, *Recent Advances in Econometric Time Series Analysis*, 14-15 dicembre 2007, Faro, Portogallo.
- ILO, 2003. *Household Income and expenditure statistics. Report II*, Seventeenth International Conference of Labour statisticians, 24 November – 3 December 2003, Geneva, Switzerland. <http://www.ilo.org/public/english/bureau/stat/download/17thicls/r2hies.pdf> (28 marzo 2018).
- Istat. 1999a. *La nuova strategia di destagionalizzazione degli indicatori congiunturali*. Nota Rapida, n.3. Roma.
- Istat. 1999b. *La revisione delle serie storiche delle forze di lavoro ottobre 1992 - aprile 1999*. *Statistiche in breve*. Roma.
- Istat, 2003. *Metodologia e organizzazione dell'indagine multiscopo sulla domanda turistica "Viaggi e vacanze"*. Collana Metodi e norme, n.17. Roma. [https://ebiblio.istat.it/digibib/metodi\\_e\\_norme/IST0043540Ed2003N17.pdf](https://ebiblio.istat.it/digibib/metodi_e_norme/IST0043540Ed2003N17.pdf) (28 marzo 2018).
- Istat, 2004, Concord V. 1.0. *Controllo e correzione dei dati. Manuale utente e aspetti metodologici*, collana Tecniche e Strumenti, n.1, Roma. <http://www3.istat.it/strumenti/metodi/software/MTSFload/TRATTAMENTOERRORload/CONCORDload/manualeconcord.pdf> (28 marzo 2018).
- Istat. 2006a. *Il sistema di indagini sociali multiscopo*. Collana Metodi e norme, n. 31. Roma. [http://www3.istat.it/dati/catalogo/20060615\\_00/met\\_%20norme\\_06\\_31\\_il\\_sistema\\_di\\_indagini\\_multiscopo.pdf](http://www3.istat.it/dati/catalogo/20060615_00/met_%20norme_06_31_il_sistema_di_indagini_multiscopo.pdf) (28 marzo 2018).
- Istat, 2006b. *La rilevazione sulle forze di lavoro: contenuti, metodologie, organizzazione*. Collana Metodi e norme, n. 32. Roma. [http://www3.istat.it/dati/catalogo/20060830\\_00/met\\_norme\\_06\\_32\\_%20rilevazione\\_forze\\_lavoro.pdf](http://www3.istat.it/dati/catalogo/20060830_00/met_norme_06_32_%20rilevazione_forze_lavoro.pdf) (28 marzo 2018).
- Istat, 2007. *Come si progetta il monitoraggio del lavoro sul campo di un'indagine sulle famiglie*. Collana Metodi e norme, n. 34. Roma. [http://www3.istat.it/dati/catalogo/20080201\\_00/men\\_0734\\_come\\_si\\_%20progetta\\_il\\_monitoraggio-lavoro\\_campo\\_indagine\\_famiglie.pdf](http://www3.istat.it/dati/catalogo/20080201_00/men_0734_come_si_%20progetta_il_monitoraggio-lavoro_campo_indagine_famiglie.pdf) (28 marzo 2018).
- Istat. 2009a. *Documento di sintesi sullo stato delle indagini CATI presso le famiglie e sull'analisi delle criticità legate alla copertura delle indagini basate su liste di telefoni fissi*. Technical Report. Roma.
- Istat, 2009b. *Documento di sintesi sulla prospettiva di ampliamento delle indagini CATI sulle famiglie e sulle possibili soluzioni da adottare per risolvere i problemi di copertura e qualità delle indagini telefoniche*. Technical Report. Roma.
- Istat, 2014. *La modernizzazione delle tecniche di rilevazione nelle indagini socioeconomiche sulle famiglie*. Collana Letture statistiche – Metodi. Roma. <https://www.istat.it/it/files/2015/01/Modernizzazione-indagini-famiglie.pdf> (28 marzo 2018).
- Istat, 2015. *La nuova indagine sulle spese per consumi in Italia*. Collana Letture statistiche – Metodi. Roma. <http://www.istat.it/it/files/2016/03/Indagine-spesa-per-consumi.pdf> (28 marzo 2018).

- Maravall, A., 1995. Unobserved Components in Economic Time Series, in M.H. Pesaran, M. Wickens, edit by, *Handbook of Applied of Econometrics. Macroeconomics*. Oxford: Basil Blackwell.
- Maravall, A., V. Gomez., 1992. *Signal Extraction in ARIMA Time Series. Program SEATS*. European University Working Paper, n. 92/65. Firenze.
- Polivka, A. E., 1994. *Comparisons of Labor Force Estimates from the Parallel Survey and the CPS during 1993*. CPS Overlap Analysis Team Technical Report, 1. Washington.
- Pollock, D., G. Stephen., 1993. *Lectures in Time-Series Analysis and Forecasting*. University Queen Mary and Westfield College, Department of Economics. London.
- Raghunathan, E., J. M. Lepkowski, J. Van Hoewyk, P. Solenberger, 2001. A multivariate technique for multiply imputing missing values using a sequence of regression models, *Survey Methodology*, 27 (1): 85-95. <http://www.statcan.gc.ca/pub/12-001-x/2001001/article/5857-eng.pdf> (28 marzo 2018).
- Regolamento (CE) n. 1893/2006 del Parlamento Europeo e del Consiglio del 20 dicembre 2006 *che definisce la classificazione statistica delle attività economiche NACE Revisione 2 e modifica il regolamento (CEE) n. 3037/90 del Consiglio nonché alcuni regolamenti (CE) relativi a settori statistici specifici* <https://eur-lex.europa.eu/legal-content/IT/TXT/PDF/?uri=CELEX:32006R1893&from=IT>. (28 marzo 2018).
- Regolamento (UE) n. 692/2011 del Parlamento Europeo e del Consiglio del 6 luglio 2011 *relativo alle statistiche europee sul turismo e che abroga la Direttiva 95/57/CE del Consiglio*. <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2011:192:0017:0032:IT:PDF> (28 marzo 2018).
- Regolamento di esecuzione (UE) n. 1051/2011 della Commissione del 20 ottobre 2011 *recante disposizioni di attuazione del regolamento (UE) n. 692/2011 del Parlamento europeo e del Consiglio relativo alle statistiche europee sul turismo per quanto riguarda la struttura delle relazioni sulla qualità e la trasmissione dei dati*. <http://eur-lex.europa.eu/legal-content/IT/TXT/PDF/?uri=CELEX:32011R1051&from=IT> (28 marzo 2018).
- Regolamento (UE) n. 557/2013 della Commissione del 17 giugno 2013 che applica il Regolamento (CE) n. 223/2009 del Parlamento europeo e del Consiglio relativo alle statistiche europee per quanto riguarda l'accesso ai dati riservati destinati a fini scientifici che abroga il regolamento (CE) n. 831/2002 della Commissione. [https://www.istat.it/it/files/2011/05/Reg.-557\\_2013.pdf](https://www.istat.it/it/files/2011/05/Reg.-557_2013.pdf) (28 marzo 2018).
- Spizzichino, A., S. Loriga, R. Gatto. 2007. *Quality Evaluation of the Back-recalculated Italian Labour Force Survey Time Series*. Paper for the Q2008 - European Conference on Quality in Official Statistics, 9-11 Luglio 2007, Roma.
- UNWTO, 2003. Tourism satellite account (TSA) implementation project. In E. Paci *Papers on measuring the economic significance of tourism, Volume 3*. <https://www.e-unwto.org/doi/abs/10.18111/9789284406425> (28 marzo 2018).
- UNWTO, 2005. *Measuring domestic tourism and the use of household surveys*, Working Paper. <http://statistics.unwto.org/sites/all/files/pdf/doc7.pdf> (28 marzo 2018).
- UNWTO, 2010. *International Recommendations on tourism statistics 2008* (IRTS 2008), ST/ESA/STAT/SER.M/83/Rev.1, Madrid and New York. <https://unstats.un.org/unsd/tradeserv/irts%202008%20edited%20whitecover.pdf> (28 marzo 2018).

