

INDAGINE SUI CONSUMI DELLE FAMIGLIE

**DOCUMENTAZIONE TECNICA E DESCRIZIONE
DEL FILE STANDARD NON GERARCHICO**
Anni 1987-1996

PREMESSA

Il Decreto Legislativo n.322 del 6/9/1989 regola la diffusione delle informazioni statistiche prodotte nell'ambito del Sistema Statistico Nazionale al fine di garantire la riservatezza dei rispondenti.

In particolare, per la diffusione di dati elementari, l'articolo 10, comma 2, dispone quanto segue: "Sono distribuite altresì, ove disponibili, su richiesta motivata e previa autorizzazione del Presidente dell'Istat, collezioni campionarie di dati elementari, resi anonimi e privi di ogni riferimento che ne permetta il collegamento con singole persone fisiche e giuridiche".

Nell'osservanza di tale legge l'Istat ha adottato misure e tecniche che rendono impossibile, o altamente improbabile, il collegamento dei dati rilasciati con l'unità statistica a cui si riferiscono. Per tale motivo sono state apportate alcune modifiche sui files originali delle indagini, nell'intento di garantire la massima protezione ai dati contenendo al minimo l'eventuale perdita di informazione.

Le metodologie applicate si concretizzano nell'accorpamento e/o riclassificazione di modalità di variabili e nell'oscuramento di variabili. In quest'ultimo caso, nei campi del tracciato record è riportata la dicitura "RISERVATO ISTAT".

Va considerato inoltre che la stessa dicitura è stata utilizzata anche per quelle variabili non attendibili dal punto di vista campionario e quindi non analizzabili statisticamente.

Il file non gerarchico permette di effettuare analisi sulle spese familiari annuali relative a 61 categorie di consumi.

Non contiene le caratteristiche strutturali dei singoli componenti della famiglia, ma una sola variabile relativa alla tipologia della famiglia (in termini di componenti e composizione per età).

Il tasso di campionamento è quello previsto per l'indagine e pari allo 0,16%.

La soglia territoriale minima è costituita dalla regione Molise (328.000 abitanti). La variabile chiave è costituita dalla tipologia della famiglia (14 modalità)

ASPETTI GENERALI DELLA RILEVAZIONE SUI CONSUMI DELLE FAMIGLIE

Generalità

La rilevazione campionaria sui Consumi delle famiglie italiane, comunemente nota anche come indagine sui Bilanci di Famiglia (BF), è fondamentalmente mirata ad accertare i consumi familiari e, quindi, ha lo scopo di osservare particolari aspetti sociali ed economici delle condizioni di vita delle famiglie italiane, d'interesse sia per le autorità pubbliche locali e nazionali, sia per gli operatori privati.

Inizialmente concepita ed utilizzata dai Contabili nazionali per controllare e migliorare la congruità e l'attendibilità delle spese familiari e per fornire un valido ausilio alla costruzione del "paniere" degli indici dei prezzi al consumo, ha via via ampliato le sue finalità allargando il campo degli utilizzatori fino a divenire una fonte statistica autonoma, ora rappresenta una fonte d'informazioni sempre più utilizzata in campo statistico.

La rilevazione, di cui vengono divulgati i risultati in questo manuale, riguarda i consumi delle famiglie residenti nel territorio nazionale con esclusione, quindi, dei consumi dei membri delle convivenze (caseme, ospedali, ecc.) e di quelli effettuati in Italia da persone abitualmente residenti all'estero.

Principali finalità della rilevazione

L'indagine sui consumi delle famiglie ha lo scopo, come già accennato in precedenza, di rilevare la struttura e il livello dei consumi secondo modalità di carattere economico-sociale e territoriale delle famiglie. Essendo continuativa, l'indagine consente di conoscere e seguire l'evoluzione in senso qualitativo e quantitativo degli standard di vita familiari e di operare confronti territoriali e sociali, costituendo una base conoscitiva utilizzata da ricercatori ed operatori economici pubblici e privati. Peraltro, la simultanea conoscenza di molteplici aspetti socio-economici delle famiglie permette studi trasversali tra gruppi sociali diversi moltiplicando, in tal modo, l'utilizzazione delle informazioni acquisite.

Oggetto essenziale della rilevazione sono i *consumi privati*, costituiti da *tutti i beni e servizi acquistati o autoconsumati dalle famiglie per il soddisfacimento dei propri bisogni*. Nelle definizioni rientrano anche i beni che provengono dal proprio orto o azienda agricola, i beni e servizi forniti dal datore di lavoro ai dipendenti a titolo di salario o per prestazioni di servizio, i fitti stimati delle abitazioni godute in proprietà. Ogni altra spesa effettuata dalle famiglie per scopo diverso dal consumo è esclusa dalla rilevazione (come per esempio l'acquisto di case e di terreni, il pagamento delle imposte, le spese connesse con l'attività professionale come spese telefoniche di un negozio e simili).

La rilevazione viene svolta secondo il metodo della "spesa" che consiste nel rilevare tutti i beni e servizi acquistati o comunque consumati dalle famiglie per le proprie necessità: spese e quantità per generi alimentari e spese per generi non alimentari e servizi. Viene indicato l'intero ammontare delle spese con riferimento al valore al momento dell'acquisto, prescindendo dalle modalità di pagamento anche per gli acquisti a rate o a credito.

Unità di rilevazione

L'unità di rilevazione è la famiglia di fatto, intesa come *un insieme di persone coabitanti e legati da vincoli di matrimonio, parentela, affinità, adozione, tutela o affettivi*. Sono considerate facenti parte della famiglia, come membri aggregati, tutte le persone che, a qualsiasi titolo, convivono abitualmente con la famiglia stessa.

Poichè molte informazioni nelle precedenti pubblicazioni sono riferite al "capo famiglia", va detto che tale termine a partire dall'edizione del 1992 è stato sostituito da "persona di riferimento" che nel regolamento anagrafico vigente corrisponde all'intestatario della scheda familiare".

Sono escluse dalla rilevazione le convivenze (caseme, ospedali, befofrof, istituti religiosi, convitti, ecc.).

Disegno di campionamento

La rilevazione si basa su un disegno di campionamento di tipo complesso: in ciascuna regione geografica vengono definiti due domini territoriali, nei quali si applicano due diverse strategie campionarie:

a) Per il dominio territoriale costituito da Comuni capoluogo di provincia o con più di 50.000 abitanti, definiti Comuni di gruppo 1 o autorappresentativi (AR), viene adottato un campionamento ad uno stadio stratificato, in cui ogni comune costituisce strato a sé. Le unità di campionamento sono costituite dalle famiglie anagrafiche, selezionate senza reimmissione e probabilità uguali.

b) Per il dominio territoriale rappresentato dai rimanenti comuni, definiti Comuni di gruppo 2 o non autorappresentativi (NAR), è utilizzato un campionamento a due stadi stratificato. Le unità di primo stadio sono costituite dai Comuni, stratificati in base all'ampiezza demografica, alla zona altimetrica ed all'attività economica prevalente, per un totale di 135 strati. In ogni strato vengono selezionati tre Comuni con probabilità proporzionale alla dimensione demografica. Le unità di secondo stadio sono costituite dalle famiglie anagrafiche.

I Comuni AR, (circa 150) partecipano alla rilevazione tutti i mesi dell'anno. I Comuni NAR, (estratti, circa 400), invece vi partecipano suddivisi in tre sottogruppi:

- i Comuni del primo sottogruppo effettuano l'indagine nel primo mese di ciascun trimestre (gennaio, aprile, luglio, ottobre);
- quelli del secondo sottogruppo nel secondo mese (febbraio, maggio, agosto, novembre);
- i Comuni del terzo sottogruppo nell'ultimo mese (marzo, giugno, settembre, dicembre).

Riepilogando, ogni mese partecipano, all'indagine 285 Comuni di cui 150 AR (sempre gli stessi) e 135 NAR (che variano da un mese all'altro nell'ambito di un trimestre).

La determinazione del numero delle famiglie campione, pari a circa 3.384 famiglie in ciascun mese per un totale di 39.168 unità l'anno, è stata effettuata in modo da garantire l'attendibilità delle principali stime pubblicate, il cui livello territoriale minimo è rappresentato dalla regione.

L'indagine è condotta su campioni di famiglie indipendenti in quanto per ogni Comune AR si seleziona un campione di famiglie che viene successivamente suddiviso in 12 gruppi, ciascuno dei quali effettua l'indagine in un mese dell'anno. Per ogni Comune NAR si procede all'estrazione di un campione di famiglie che, invece, viene suddiviso in quattro gruppi che partecipano all'indagine nei 4 mesi nei quali il Comune stesso deve svolgere l'indagine.

Ciascuno di tali gruppi di famiglie viene a sua volta suddiviso in tre sub-gruppi: le famiglie del primo sub-gruppo eseguono l'indagine nella prima decade del mese; quelle del secondo sub-gruppo nella seconda decade e quelle del terzo nell'ultima decade del mese.

Tecnica di rilevazione

L'indagine è basata su tre diversi questionari :

- il libretto degli acquisti, sul quale la famiglia campione annota giornalmente, per un periodo di dieci giorni, le spese e le quantità acquistate di generi alimentari (pane, pasta, carne, ecc.) e le spese sostenute per beni e servizi correnti (giornali, tabacco, biglietti autobus, ecc.), ,
- il taccuino degli autoconsumi, sul quale la famiglia riporta giornalmente, sempre per una decade, le quantità consumate e la stima dei rispettivi costi di generi alimentari autoprodotti;
- il riepilogo delle spese, compilato dal rilevatore mediante intervista diretta alla famiglia alla fine del mese di rilevazione, contenente le informazioni socio-demografiche della famiglia e le principali spese sostenute dalla medesima per beni e servizi e per l'acquisto di beni durevoli.

Le operazioni di rilevazione sono affidate agli uffici di statistica dei Comuni che provvedono alla estrazione delle famiglie campione, al reclutamento degli intervistatori ed alla spedizione di lettere di preavviso, a firma del Presidente dell'Istat, per agevolare la collaborazione delle famiglie stesse.

In pratica l'intervistatore contatta la famiglia, illustra gli scopi dell'indagine e consegna il libretto degli acquisti, fornendo le opportune spiegazioni per la compilazione; inoltre, accertata la possibilità di consumo di generi alimentari autoprodotti, consegna ed illustra anche il taccuino degli autoconsumi. Nei primi giorni del mese successivo a quello di rilevazione (ad esempio se la famiglia ha registrato le spese giornaliere dal 11 al 20 aprile, nei primi giorni di maggio) l'intervistatore torna dalla famiglia, controlla e ritira il libretto degli acquisti, con l'eventuale taccuino degli autoconsumi compilato, e conduce l'intervista finale relativa alle restanti spese sostenute nel mese (nell'esempio ad aprile) e per alcune spese rilevanti sostenute negli ultimi tre mesi (nell'esempio febbraio, marzo ed aprile).

Ciascuna famiglia registra le spese per una sola decade nel mese e sostiene una sola intervista finale, senza essere più contattata nel tempo.

In caso di impossibilità della famiglia a partecipare all'indagine (rifiuto, irreperibilità, momentanea assenza, ecc.) si procede alla sostituzione con una famiglia che abbia caratteristiche socio-demografiche simili.

Revisione e spoglio dei dati

I questionari compilati vengono trasmessi mensilmente all'ISTAT, presso il quale sono sottoposti ad una revisione manuale al fine di eliminare errori ed omissioni; dopo la registrazione su nastro magnetico vengono effettuati i controlli e le correzioni automatiche che si avvalgono di "griglie di compatibilità" predisposte sulla base di schemi logici precostituiti. Tali procedure informatiche permettono di correggere gli errori di registrazione e controllano le singole variabili e le loro interrelazioni. La presenza di casi anomali viene segnalata affinché sia verificata e modificata in maniera interattiva.

Costruzione delle stime

La costruzione delle stime è basata su un processo di postratificazione delle famiglie per numero di componenti, variabile strettamente correlata con il livello delle spese sostenute dalla famiglia per consumi. Quindi per il riporto all'universo (che viene eseguito trimestralmente in ciascun strato nell'ambito regionale) si utilizzano dei coefficienti ottenuti in due fasi successive. Nella prima fase, per lo strato "i-esimo" della regione "j-esima", viene determinato il coefficiente "c_{ij}" derivato dal rapporto:

$$c_{ij} = \frac{P_{ij}}{p_{ij}} \quad (1)$$

dove P_{ij} è la popolazione universo dello strato "i-esimo" della regione "j-esima" e p_{ij} è il numero degli individui appartenenti alle famiglie intervistate in tale strato. Successivamente si procede ad una postratificazione delle famiglie per ampiezza, quindi il coefficiente (1) viene trasformato nei coefficienti "C_{ijs}" che si ottengono in base alla relazione

$$C_{ijs} = c_{ij} \frac{F_{js}}{f_{js}} \quad (s = 1, 2, \dots, 5 \text{ e più}) \quad (2)$$

in cui F_{js} sono le famiglie universo di ampiezza s della regione j-esima e f_{js} la stima delle stesse ottenuta tramite le famiglie campione. In pratica si costruisce un record per ciascuna famiglia intervistata, nel quale oltre alle notizie rilevate viene riportato il coefficiente di espansione.

Va tenuto presente che l'universo di riferimento è costituito dalla popolazione residente delle famiglie al netto delle convivenze; tale valore viene stimato sottraendo dalla popolazione anagrafica la percentuale ottenuta al Censimento della Popolazione 1991 di individui appartenenti alle convivenze. Il numero totale delle famiglie è stato stimato in ogni Regione, distintamente per i due domini territoriali AR e NAR, dividendo la popolazione sopra citata per il numero medio di componenti per famiglia risultante al Censimento della Popolazione 1991. Quindi a tale totale è stata applicata la distribuzione delle famiglie per ampiezza risultante dai dati censuari.

COSTRUZIONE DELLE STIME ED ERRORI DI CAMPIONAMENTO

Le informazioni riportate nel presente file sono di carattere campionario. Per ottenere stime relative all'intera popolazione oggetto d'indagine è necessario moltiplicare ciascuna informazione per il coefficiente di riporto all'universo.

E' tuttavia da evidenziare che il file contiene le informazioni delle quattro indagini trimestrali effettuate in ciascun anno; tale circostanza rende necessaria una particolare cautela nella costruzione di stime annuali. Nel diffondere i risultati di un'indagine campionaria occorre fornire agli utilizzatori le informazioni necessarie per valutare l'attendibilità delle stime ottenibili.

Limitatamente agli aspetti campionari, sarebbe opportuno calcolare per ogni stima prodotta il corrispondente errore di campionamento. E' ovvio che questa pratica comporterebbe notevoli difficoltà per l'utilizzatore, dovute al fatto che per problemi di tutela della riservatezza non vengono forniti i codici identificativi territoriali sui quali è basato il disegno dell'indagine. Tale problema viene risolto con l'introduzione di opportuni modelli regressivi che consentono di mettere in relazione l'errore relativo di campionamento con l'ampiezza della stima; infatti è noto che l'errore relativo decresce all'aumentare della stima.

Per l'indagine sui consumi delle famiglie è stato utilizzato il modello:

$$\ln(E.R.) = \ln \frac{\sigma(\hat{Y})}{\hat{Y}} = a + b \cdot \ln \hat{Y}$$

in cui E.R. è l'errore relativo, è la stima prodotta e $\sigma(\hat{Y}) = \sqrt{V(\hat{Y})}$ il corrispondente scarto quadratico medio.

Il calcolo dei parametri a e b è stato condotto, in ciascun dominio geografico, adattando il modello, mediante il metodo dei minimi quadrati ponderati, ad una nuvola costituita da un consistente numero di punti \hat{Y} e $V(\hat{Y})$.

Nel prospetto 1 sono riportati i valori dei coefficienti a e b e dell'indice di determinazione R^2 delle stime di frequenze assolute per regione relativi alle famiglie.

Ad esempio per determinare l'errore relativo percentuale, l'errore assoluto e l'intervallo di confidenza del numero di famiglie con tre componenti nella regione Lazio che hanno una spesa media mensile inferiore a 3.000.000 di lire, pari a 210.000 unità, dal prospetto 1 si ottiene:

$$\ln \frac{\sigma(\hat{Y})}{\hat{Y}} = 3,7798 - 0,50991 \cdot \ln 210.000 = -2,46908$$

da cui segue:

$$E.R. = \frac{\sigma(\hat{Y})}{\hat{Y}} = 0,085 = 8,5\%$$

Il corrispondente errore assoluto è uguale a:

$$0,085 \times 210.000 = 17.850$$

l'intervallo di confidenza avrà quindi come estremi, al livello di probabilità $P=95\%$:

$$210.000 - 2 \times 17.850 = 174.300$$

$$210.000 + 2 \times 17.850 = 245.700$$

Prospetto 1 - Coefficienti a e b e indice $R^2(\%)$, delle funzioni interpolatrici per gli errori campionari delle stime di frequenza assolute riferite ai bilanci di famiglia.

Zone Territoriali	a	b	$R^2(\%)$
ITALIA	3,9762	-0,51959	94,6
Regione			
Piemonte e Valle d'Aosta	3,0274	-0,46150	94,6
Lombardia	3,3435	-0,46475	90,6
Bolzano	2,8237	-0,50940	88,5
Trento	3,0711	-0,53508	87,4
Veneto	2,7294	-0,42167	91,8
Friuli-Venezia Giulia	2,7515	-0,44566	86,6
Liguria	3,5986	-0,53131	92,7
Emilia Romagna	4,0990	-0,54726	93,4
Toscana	3,8501	-0,55027	92,2
Umbria	3,4646	-0,54773	89,2
Marche	2,9511	-0,48470	85,5
Lazio	3,7798	-0,50991	89,7
Abruzzi	3,4780	-0,51113	88,9
Molise	2,3516	-0,47576	89,8
Campania	3,6090	-0,49450	93,4
Puglia	2,9777	-0,44236	86,9
Basilicata	2,4288	-0,43991	89,3
Calabria	2,9500	-0,44012	89,9
Sicilia	3,5510	-0,48188	88,6
Sardegna	3,3105	-0,48903	90,9