



File di microdati per la ricerca

Consumi energetici delle famiglie

Periodo di riferimento: anno 2020-2021

Descrizione del file

INDICE

Introduzione	3
Le misure di protezione adottate	4
Riferimenti bibliografici	7

Introduzione ¹

I file di microdati per scopi di ricerca scientifica (MFR²) possono essere rilasciati su richiesta motivata e previa autorizzazione del Presidente dell'Istituto, purché siano resi anonimi e privi di ogni riferimento che ne permetta il collegamento con le unità intervistate.

Per la predisposizione del file MFR dell'indagine *Consumi Energetici delle famiglie*, periodo di riferimento 2020-2021, è stata adottata una opportuna metodologia, allo scopo di limitare il rischio di violazione della riservatezza. Lo scenario di intrusione considerato è quello relativo all'identificazione spontanea, ossia conseguente a conoscenze a priori dell'utente che potrebbero permettere di attribuire correttamente i dati rilasciati alle unità della popolazione rilevata.

Le variabili coinvolte nel processo di protezione sono quelle che possono consentire l'associazione tra le informazioni e i rispondenti, ossia:

- gli *identificativi diretti*, che individuano in maniera univoca le unità statistiche di rilevazione/analisi (come ad esempio l'indirizzo);
- gli *identificativi indiretti* (come ad esempio il numero di componenti della famiglia), che permettono di circoscrivere la popolazione alla quale appartengono i rispondenti e, da sole o in combinazione con altre, possono portare alla re-identificazione di uno o più record.

Mentre le prime sono cancellate dal file, le seconde vengono trattate statisticamente riducendone il contenuto informativo, allo scopo di limitare il rischio di violazione della riservatezza. Oltre agli identificativi ("diretti" ed "indiretti") sono state coinvolte nella procedura di protezione anche altre variabili non annoverate tra gli identificativi ma che sono state opportunamente trattate a causa del potenziale danno che potrebbero causare agli interessati in caso di violazione. Per il significato delle variabili e delle rispettive modalità si rinvia al questionario di indagine contenuto nella cartella dei metadati.

Si ricorda che gli utenti del file di microdati per scopi di ricerca scientifica sono obbligati ad adottare tutte le misure per la tutela della riservatezza alle quali si sono già impegnati all'atto della sottoscrizione del modulo di richiesta³

¹ Il presente documento è a cura di Emilia Arcaleni, Marco Caputi e Luigi Virgili, con la collaborazione di Giovanna Tagliacozzo e Antonio Damiani.

² MFR è l'acronimo per il file per la ricerca (Microdata File for Research).

³ Per le modalità di acquisizione dei dati si veda: <https://www.istat.it/it/dati-analisi-e-prodotti/microdati>

Le misure di protezione adottate

Per la predisposizione del file per la ricerca scientifica *Consumi energetici delle famiglie*, periodo di riferimento 2020-2021, sono state adottate le misure di protezione di seguito specificate.

Soppressione di variabili

1. Codice famiglia: il progressivo famiglia univoco a livello indagine è stato sostituito con un numero fittizio,
2. Comune e provincia di residenza,
3. Tutte le variabili individuali dal sesto componente in poi sono state soppresse (vedi più avanti Numero di componenti),
4. Tipologia comunale.

Ricodifica di variabili categoriali o quantitative discrete

Per alcune variabili sono state effettuate aggregazioni in classi e top-coding, che consiste nel raggruppare le unità statistiche collocate lungo la coda dei valori più elevati della distribuzione.

1. Numero di componenti della famiglia (*ncomp_mfr*):
top-coding dal sesto componente (6 e più componenti). Tutte le variabili individuali dal sesto componente in poi sono state soppresse (*Sesso6-Sesso12*, *eta6-eta12*, *relpar6-relpar12*, *giorni6-giorni12*, *istr6-istr12* e *condiz6-condiz12*). La famiglia con la maggiore numerosità rilevata ha 12 componenti.
2. Età dei componenti (*eta_mfr1 - eta_mfr5*) - raggruppamento e top-coding:

Etichetta	Descrizione
1	0-5 anni
2	6-10 anni
3	11-14 anni
4	15-19 anni
5	20-29 anni
6	30-39 anni
7	40-49 anni
8	50-59 anni
9	60-69 anni
10	70-79 anni
11	80 anni e più

3. Relazione di parentela con la Persona di Riferimento (*relpar_mfr2* - *relpar_mfr5*) - raggruppamento delle modalità originarie:

Etichetta	Descrizione	Etichetta originaria
1	persona di riferimento (PR)	1
2	coniuge/convivente (coniugalmente) di PR	2
4	figlia/o di PR (o del coniuge/convivente di PR)	4
5	altro parente di PR (o del coniuge/convivente di PR): sorella/fratello, genitore, suocera/o, nonna/o, zia/o, nipote ecc.	3, 5
6	persona coabitante non legata da vincoli di parentela (persona legata da amicizia, colf e altro personale coabitante addetto ai servizi della famiglia, persona in affitto, pensionante ecc.)	6, 7, 8

4. Titolo di studio (*istr_mfr1* - *istr_mfr5*) - raggruppamento delle modalità originarie:

Etichetta	Descrizione	Etichetta originaria
1	nessun titolo o licenza di scuola elementare	1, 2
3	licenza di scuola media (secondaria di primo grado)	3
4	diploma di scuola secondaria di secondo grado (2-3 anni o 4-5 anni)	4, 5
6	laurea, post laurea, Accademia (diploma di Belle Arti, Conservatorio ecc.)	6, 7, 8

5. Condizione occupazionale (*condiz_mfr1* - *condiz_mfr5*) - raggruppamento delle modalità originarie:

Etichetta	Descrizione	Etichetta originaria
1	occupato/a	1
2	in cerca di occupazione	2, 3
4	casalingo/a	4
5	studente/essa	5
7	persona ritirata dal lavoro	7
8	in altra condizione	6, 8

6. Superficie dell'abitazione (*v1_7_mfr*) - raggruppamento e top-coding:

Etichetta	Descrizione
1	meno di 40 metri quadri
2	da 40 a 49 metri quadri
3	da 50 a 59 metri quadri
4	da 60 a 69 metri quadri
5	da 70 a 79 metri quadri
6	da 80 a 89 metri quadri
7	da 90 a 99 metri quadri
8	da 100 a 109 metri quadri

9	da 110 a 119 metri quadri
10	da 120 a 129 metri quadri
11	da 130 a 149 metri quadri
12	da 150 a 169 metri quadri
13	da 170 a 199 metri quadri
14	da 200 a 249 metri quadri
15	da 250 a 299 mq
16	300 mq e oltre

7. Numero di stanze dell'abitazione (*v1_8_mfr*) - raggruppamento e top-coding:

Etichetta	Descrizione
puntuali	per valori minori o uguali a 15
16	da 16 a 20 stanze
17	da 21 a 25 stanze
18	più di 25 stanze

8. Numero di pannelli dell'impianto solare (*v4_3a_mfr*) - raggruppamento e top-coding.

Il numero più alto di pannelli rilevato è stato 45.

Etichetta	Descrizione
1	1 pannello
2	2 pannelli
3	3-5 pannelli
4	6-10 pannelli
5	più di 10 pannelli

9. Numero di infissi.

È stato applicato un top-coding dal valore 16 in poi al numero totale di infissi in legno, in metallo o in pvc (rispettivamente *v1_11e_mfr*, *v1_14e_mfr*, *v1_17e_mfr*). Qualora una delle suddette variabili avesse superato il valore 15, è stata inoltre effettuata una **soppressione locale** della ripartizione dettagliata di tali infissi per numero di ante (*v1_11a_mfr* - *v1_11d_mfr*, *v1_14a_mfr* - *v1_14d_mfr*, *v1_17a_mfr* - *v1_17d_mfr*), salvaguardando l'informazione solo per la tipologia dominante (cioè quella col numero maggiore). Le variabili oscurate sono state poste missing. Il top-coding è stato applicato anche al numero di infissi con doppi/tripli vetri per i tre diversi materiali (rispettivamente *v1_12_mfr*, *v1_15_mfr*, *v1_18_mfr*).

Perturbazione di variabili quantitative

Per le variabili quantitative dei consumi energetici (spese e quantità), è stata applicata la tecnica statistica della **microaggregazione**, per preservare la riservatezza dei dati. La

tecnica prevede che il valore rilevato sia sostituito dal valore medio calcolato in gruppi omogenei di record di numerosità costante⁴. Le variabili trattate con questa tecnica sono:

- legna: quantità consumata (*v6_6_mfr*) e spesa (*v6_9_mfr*)
- pellet: quantità consumata (*v6_12_mfr*) e spesa (*v6_14_mfr*)
- spesa per energia elettrica (*elettricità_eu_mfr*)
- spesa per metano (*metano_eu_mfr*)
- spesa per gasolio (*gasolio_eu_mfr*)
- spesa per GPL (*gpl_eu_mfr*)
- spesa residua (per impianto centralizzato alimentato a biomasse, energia solare o fonte non identificata) (*altro_eu_mfr*).

La spesa totale (*totale_eu_mfr*) è stata quindi ottenuta come somma delle componenti microaggregate (elettricità, metano, gasolio, gpl, legna, pellet, altro).

Le variabili inglobano anche le risposte alle corrispondenti domande di recupero espresse in classi (rispettivamente quesiti 6.7, 6.13, 8.1c, 8.3, 8.4c, 8.5c, 8.6c).

A causa delle misure di protezione adottate si possono verificare scostamenti rispetto ai dati pubblicati dall'Istat.

Riferimenti bibliografici

Hundepool, A., Domingo-Ferrer, J., Franconi, L., Giessing, S., Nordholt, E. S., Spicer, K. E de Wolf, P.-P. (2012). Statistical Disclosure Control. Wiley.

Willenborg, L. e de Waal, T. (1996). Statistical Disclosure Control in Practice. Lecture Notes in Statistics, 111, New York: Springer-Verlag.

Willenborg, L. e de Waal, T. (2000). Elements of statistical disclosure control. Lecture Notes in Statistics, 115, New York: Springer-Verlag.

⁴ I record sono stati ordinati per la variabile quantitativa d'interesse (salvaguardando la stratificazione per regione) e sono stati definiti gruppi di record vicini di numerosità costante (triplette): la media ponderata calcolata in ciascun gruppo è stata poi assegnata ai record del gruppo, in sostituzione del valore rilevato.