



**File ad uso pubblico  
micro.STAT**

## **Indagine sulle spese delle famiglie**

**Periodo di riferimento: anno 2018**

### **Descrizione del file**

Anno 2019

## INDICE

Introduzione .....	3
Le misure di protezione adottate.....	4
Riferimenti bibliografici.....	6

## Introduzione

I file di microdati ad uso pubblico (mlcro.STAT) sono collezioni di dati elementari, liberamente scaricabili via web (<http://www.istat.it/it/archivio/microdati+ad+uso+pubblico>), resi anonimi e privi di ogni riferimento che ne permetta il collegamento con le unità intervistate. Si segnala che relativamente a questa stessa indagine è disponibile anche il file per la ricerca (MFR<sup>1</sup>), il cui contenuto informativo è superiore rispetto a quello del file ad uso pubblico<sup>2</sup>.

Per la predisposizione del file mlcro.STAT relativo all'Indagine sulle spese delle famiglie, periodo di riferimento 2018, è stata adottata una metodologia che limitasse il rischio di violazione della riservatezza. Gli scenari di intrusione considerati sono stati:

- identificazione attraverso archivi esterni, ovvero attraverso il collegamento con i dati rilasciati da altre fonti pubbliche;
- identificazione spontanea, ovvero 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 di *rilevazione/analisi* (come ad esempio indirizzo e codice fiscale);
- gli *identificativi indiretti*, o *variabili chiave* (come ad esempio il numero di componenti della famiglia), che permettono di circoscrivere la popolazione cui appartengono i rispondenti e che, da sole o in combinazione con altre, possono portare alla re-identificazione di uno o più record.

Mentre i primi vengono cancellati dal file, i secondi vengono trattati secondo le modalità descritte di seguito allo scopo di limitare il rischio di violazione della riservatezza. Per il significato delle variabili e delle rispettive modalità si rinvia al questionario di indagine contenuto nella cartella metadati.

---

<sup>1</sup> MFR è l'acronimo per il file per la ricerca (Microdata File for Research). La documentazione metodologica di tali file è consultabile sul sito Istat (<http://www.istat.it/it/archivio/file+per+la+ricerca>).

<sup>2</sup> I ricercatori, che necessitano di informazioni maggiormente dettagliate, possono, previa autorizzazione del Presidente dell'ISTAT, richiedere il file di microdati per la ricerca.

## Le misure di protezione adottate

Per la predisposizione del file ad uso pubblico per l'Indagine sulle spese delle famiglie, periodo di riferimento 2018, sono state adottate le seguenti misure di protezione.

### *Soppressione di variabili*

1. Codice famiglia
2. Codice fiscale dei componenti la famiglia
3. Percentuale di reddito percepito dai componenti la famiglia sul totale dei redditi familiari: sostituita dal principale percettore
4. Comune e provincia di residenza
5. Tipologia di comune

### *Ricodifiche di variabili categoriali o quantitative discrete*

1. Età dei componenti la famiglia: [0,18), [18,35),[35,65),[65 e oltre)
2. Numero di componenti: top-coding dal sesto componente (6 e più componenti).  
Tutte le variabili individuali dal settimo componente in poi sono state soppresse.
3. Relazione di parentela dei componenti la famiglia  
1=persona di riferimento (PR)  
2=coniuge/convivente/unito civilmente con PR  
3=figlio di PR  
4=altra relazione
4. Titolo di studio dei componenti la famiglia  
1=nessun titolo  
2=scuola primaria (elementari)  
3=scuola secondaria di primo grado  
4=scuola secondaria di secondo grado  
5=laurea e post laurea
5. Attività economica dei componenti la famiglia  
1=settore primario  
2=settore secondario  
3=settore terziario
6. Posizione nella professione dei componenti la famiglia  
1=dirigenti e impiegati  
2=operai e assimilati

3=imprenditori e liberi professionisti

4=lavoratori in proprio

7. Fonte principale di reddito dei componenti la famiglia

1=reddito da lavoro

2=reddito da pensione

3=altri redditi

8. Condizione autodichiarata dei componenti la famiglia

1=occupato/a

2=in cerca di occupazione

3=casalinga/studente/in altra condizione

4=persona ritirata dal lavoro

9. Professione dei componenti la famiglia

1=imprenditori e alta dirigenza

2=professioni della conoscenza

3=professioni esecutive di ufficio

4=professioni qualificate nel commercio e nei servizi

5=professioni manuali qualificate

6=professioni manuali non qualificate

7=forze armate

*Ricodifiche di variabili quantitative continue*

1. Raggruppamento e top-coding delle variabili sulle spese delle famiglie relative a variabili sensibili:

- spesa per occhiali e lenti da vista, spesa per apparecchi acustici, spesa per protesi dentarie e apparecchi ortodontici, spesa per sedie e carrozzine e spesa per riparazione di attrezzature sanitarie sono state raggruppate insieme;
- spesa per medico generico (certificati, vaccinazioni o altre prestazioni a pagamento), spesa per visite specialistiche, spesa per ricoveri in strutture pubbliche; e spesa per ricoveri in strutture private sono state raggruppate insieme;
- spesa per analisi cliniche e esami diagnostici (analisi del sangue, radiografie, ecografie, elettrocardiogrammi, ecc.), spesa per infermieri e altri servizi paramedici (fisioterapisti, podologi, logopedisti, agopunturisti), spesa per cure termali, ginnastica correttiva e spesa per noleggio di attrezzature sanitarie o servizi di ambulanza sono state raggruppate insieme.

2. Per tutte le variabili relative alle spese sanitarie è stato effettuato un top-coding che seguisse la metodologia di Hubert e Vandervieren (2008).

### *Soppressioni locali*

Sono stati inseriti valori mancanti in corrispondenza di tutte le variabili individuali dal settimo componente in poi. A causa delle soppressioni locali si possono verificare scostamenti rispetto ai dati pubblicati dall'Istat.

## **Riferimenti bibliografici**

- Hubert, M., Vandervieren, E. (2008). An adjusted boxplot for skewed distributions. *Computational Statistics and Data Analysis* 52, 5186-5201.
- 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.

## **Curatori**

Il presente documento, relativo all'indagine del 2018, è stato curato da Andrea Cutillo e Letizia Valdoni.

La procedura di di anonimizzazione è stata ideata da Flavio Foschi e Luigi Virgili.

Il file di microdati ad uso pubblico è stato creato da Letizia Valdoni.