



**File ad uso pubblico
micro.STAT**

Indagine Viaggi e vacanze

Periodo di riferimento: anno 2019

Descrizione del file

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>), per le quali, a partire dal 2013, sono stati sviluppati anche i corrispondenti file per la ricerca (MFR¹). Il contenuto informativo di questi ultimi è superiore rispetto a quello del file ad uso pubblico².

Per la predisposizione dei file mlcro.STAT dell'indagine *Viaggi e vacanze*, periodo di riferimento 2019, è stata adottata una opportuna metodologia, allo scopo di limitare 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 statistiche 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 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.

¹ 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 (<https://www.istat.it/it/archivio/178695>).

² I ricercatori, che necessitano di informazioni maggiormente dettagliate, possono, su richiesta motivata e previa autorizzazione del Presidente dell'Istituto, richiedere il file di microdati per la ricerca.

Le misure di protezione adottate

Per la predisposizione dei tre file (individui, viaggi ed escursioni) ad uso pubblico dell'indagine *Viaggi e vacanze*, periodo di riferimento 2019, sono state adottate le misure di protezione appresso specificate.

Soppressione di variabili comuni ai tre file:

1. Codice famiglia
2. Relazione di parentela
3. Anno di matrimonio
4. Tipologia comunale

Ricodifiche di variabili categoriali o quantitative discrete comuni ai tre file:

1. Codice individuo: numerazione progressiva univoca sull'intero file individui.
2. Residenza: il comune e la provincia di residenza sono stati sostituiti dalla regione con accorpamento delle province di Trento e Bolzano.
3. Classi di età: [0,15), [15,25), [25,35), [35,45), [45,55), [55,65), [65,75), [75 e oltre).
4. Classificazione dell'attività economica in 3 modalità:
1='Settore primario'
2='Settore secondario'
3='Settore terziario'
5. Classificazione dello stato civile in 4 modalità:
1='Celibe/nubile'
2='Coniugato'
3='Separato/divorziato'
4='Vedovo';
6. Classificazione del titolo di studio in 4 modalità:
1='Nessun titolo/licenza di scuola elementare'
2='Licenza di scuola media (o avviamento professionale)'
3='Diploma di scuola secondaria superiore o qualifica professionale'
4='Diploma universitario, ex scuole parauniversitarie, laurea, post laurea'
7. Classificazione della condizione lavorativa autopercipita in 4 modalità:
1='Occupato'
2='In cerca di occupazione'
3='Casalinga/studente/altra condizione'
4='Persona ritirata dal lavoro'
8. Classificazione della posizione nella professione in 4 modalità:
1='Dirigente, direttivo, quadro, impiegato'
2='Operaio apprendista lavorante a domicilio per conto d'impresa'

3='Imprenditore, libero professionista'

4='Lavoratore in proprio, socio cooperativa, coadiuvante, co.co.pro, prestazione opera occasionale'

9. Classificazione della codifica della professione in 7 modalità:

1='Imprenditori e alta dirigenza'

2='Professioni della conoscenza'

3='Professioni esecutive d'ufficio'

4='Professioni qualificate nel commercio e nei servizi'

5='Professioni manuali qualificate'

6='Professioni manuali non qualificate'

7='Forze armate'

10. Numero di componenti della famiglia: top-coding dal sesto componente.

Ricodifiche di variabili categoriali o quantitative discrete limitatamente al file *viaggi*:

1. Classificazione del motivo principale dei viaggi per motivi personali

1='Piacere, svago, vacanza'

2='Visita a parenti e/o amici'

3='Motivo religioso, pellegrinaggio'

4='Trattamento di salute, cura termale (inclusi i viaggi prescritti dal medico)'

2. Classificazione del motivo per il quale non è stato effettuato alcun viaggio per motivi personali nel mese di riferimento dei dati:

1='Motivi economici'

2='Motivi di lavoro/studio/familiari'

3='Mancanza di abitudine'

4='Altri motivi (inclusi motivi di salute)'

99='Non sa, non risponde'

3. Numero di componenti della famiglia che hanno viaggiato insieme: top-coding dal quinto componente.

Ricodifiche di variabili categoriali o quantitative discrete limitatamente al file *escursioni*:

1. Classificazione del motivo per il quale è stata effettuata l'escursione nel mese di riferimento dei dati:

1='Piacere, svago, vacanza'

2='Visita a parenti e/o amici'

3='Motivi religiosi o di pellegrinaggio'

4='Cultura'

5='Cure termali, trattamenti di salute/ Visite e cure mediche'

6='Shopping'

7='Motivi di lavoro'

8='Accompagnare un familiare, parente, amico'

9='Bellezze naturali'

10='Gastronomia, folklore o spettacoli'

11='Altro, Non sa/Non risponde'

2. Numero di componenti della famiglia che hanno effettuato escursioni insieme: top-coding dal quinto componente.

Ricodifiche di variabili quantitative continue limitatamente al file *viaggi*:

1. Raggruppamento e top-coding delle variabili 'spesa media del viaggio' e 'spesa media giornaliera del viaggio'.

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

In alcuni record sono stati modificati i valori puntuali in corrispondenza delle variabili *motvac* (classificazione delle vacanze) e *tipo* (classificazione dell'attività principale svolta durante le vacanze di piacere/svago/relax) al fine di proteggere l'informazione sensibile di quanti hanno effettuato viaggi per motivi di salute (con prescrizione medica) nei casi in cui non è stata raggiunta la 2-diversità: essi sono stati sostituiti rispettivamente con il codice 1 per *motvac* (viaggi di piacere/svago) e codice 3 per *tipo* (trattamenti di salute e benessere senza prescrizione medica).

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 documento è stato redatto da Barbara Dattilo.