



Rilevazione continua sulle Forze di Lavoro Dati longitudinali a 3 mesi di distanza

Descrizione del file

Anno 2016

INDICE

1. Introduzione.....	3
2. Descrizione delle variabili	3
3. Metodologia statistica per la tutela della riservatezza.....	3
4. Riferimenti bibliografici.....	5
5. Contatti	6

1. Introduzione

La necessità di tutela della riservatezza discende da vincoli di legge e da obblighi assuntivo i rispondenti. Seguendo l'Handbook on Statistical Disclosure Control (AA.VV., 2009), l'intrusione in informazioni non pubbliche si concretizza quando tramite i dati rilasciati sia possibile l'identificazione del rispondente. Limitando l'attenzione al rilascio di microdati, l'intrusione avviene quando il singolo record è correttamente assegnato ad una unità nella popolazione. Le unità statistiche oggetto dell'analisi sono gli individui e, condizionatamente alla stratificazione di interesse, sono considerate a rischio quelle che, uniche nel campione, siano tali anche nella popolazione. Le variabili specificate nel questionario di indagine hanno suggerito due possibili ambiti di riferimento; il primo attiene al contributo di singole variabili ad identificazioni spontanee o accidentali; nel secondo, si assume venga perseguito intenzionalmente l'abbinamento tra i dati di indagine e una lista di microdati disponibile secondo un prefissato livello di dettaglio.

2. Descrizione delle variabili

In riferimento all'insieme di informazioni ottenuto secondo il questionario di indagine, la riduzione del rischio di violazione della riservatezza è stata perseguita combinando soppressioni e modifiche nel livello di aggregazione delle variabili. Tali interventi sono esposti nel paragrafo 3.2. Per significati e articolazioni delle restanti variabili, rimaste immutate, si rinvia al questionario di indagine e al relativo tracciato record.

3. Metodologia statistica per la tutela della riservatezza

3.1 Apprezzamento del Rischio

In ordine al primo ambito di riferimento, riferito alle cosiddette intrusioni accidentali, l'identificazione origina "spontaneamente" dall'osservazione di un ristretto numero di variabili, alcune delle quali aventi un livello di dettaglio sufficiente ad isolare poche unità statistiche in base ad una distribuzione marginale di frequenze. Per ottenere una plausibile approssimazione degli ordini di grandezza coinvolti, in riferimento alle mutabili maggiormente suscettibili di criticità sono stati conteggiati il numero di modalità rappresentate da una sola unità statistica nonché i casi nei quali si siano manifestati cambiamenti a distanza di tre mesi dalla prima rilevazione. Con riferimento al secondo ambito, poiché il passaggio dai dati sezionali (disponibili dall'indagine sulle forze di lavoro) al panel su due unità di tempo è reso possibile dal link tra rispondenti a 3 mesi dalla prima intervista, il contenuto informativo del file longitudinale non ha natura autonoma rispetto ai

dati trasversali. In considerazione delle modalità realizzative del file longitudinale, la valutazione del rischio è stata mutuata dalla graduatoria sezionale.

3.2. Modifiche e Integrazioni delle variabili

3.2.1. Soppressione delle variabili

Diverse variabili, rilevate secondo il questionario di indagine, sono state rese non disponibili nel file. Tra di esse è significativo ricordare:

Tabella 1: variabili rese non disponibili

NOME VARIABILE	TIPO	PROVENIENZA (*)	DESCRIZIONE VARIABILE	SDC STATUS
CODFAM	CHAR	r	codice famiglia	soppressa
CODQUA	CHAR	r	codice famiglia	soppressa

(*) Le variabili rilevate nel questionario e ricodificate nella colonna sono indicate con r.

3.2.2. Ricodifica globale

La ricodifica globale è tra i più semplici metodi di riduzione del rischio di violazione della riservatezza nei dati individuali. Per caratteri raggruppati in classi, essa è definita dall'accorpamento di alcune modalità con ovvi cambiamenti per significati e frequenze. Le scelte effettuate a questo proposito sono state evidenziate nella tabella 2.

Tabella 2: variabili recanti modificazioni delle modalità

NOME VARIABILE(*)	TIPO	PROVENIENZA (**)	DESCRIZIONE VARIABILE	SDC STATUS	VARIAZIONE
MFRFAM_LONG	CHAR	r	codice famiglia	variato	Ricodifica con metodo casuale
MFRIND_LONG	CHAR	r	Identificativo dell'individuo all'interno della famiglia	variato	Ricodifica con metodo casuale

(*) Le variabili rilevate il t_{mo} trimestre del primo anno sono individuate dal suffisso _0Xt ove 0X è il primo anno; analogamente, per quelle attinenti il t_{mo} trimestre del secondo anno il suffisso è del tipo _0Yt.

(**) Le variabili rilevate nel questionario e ricodificate, nella colonna sono indicate con r.

4. Riferimenti bibliografici

AA.VV. *EU Labour Force Survey database User Guide*. European Commission, Eurostat, Directorate F, Unit F-2.

http://circa.europa.eu/irc/dsis/employment/info/data/eu_lfs/lfs_main/LFSuserguide/EULFS_Database_UserGuide_2008.pdf . 04/09/2009.

AA.VV. *Handbook on Statistical Disclosure Control*. CENTre of EXcellence for Statistical Disclosure Control, 2007. <http://neon.vb.cbs.nl/casc/handbook.htm>. 04/09/2009.

AA.VV. *La rilevazione sulle forze di lavoro: contenuti, metodologie, organizzazione*. Roma: Istat, 2006 (Metodi e Norme n. 32).

Skinner, Chris, e Natalie Shlomo. "Assessing Identification Risk in Survey Microdata Using Log-Linear Models". *Journal of the American Statistical Association* 103, 483 (2008): 989-1001

5. Contatti

ISTAT - Servizio Sistema integrato lavoro, istruzione e formazione

Via Oceano Pacifico, 171 - 00144 – Roma

tel. +39 06 46734565 / 4555 / 4714

e-mail: infolav@istat.it

ISTAT - Servizio Metodi, qualità e metadati

Via Cesare Balbo, 16 - 00198 – Roma

Curatori dei paragrafi

Paragrafi 1, 2, 3.1: Flavio Foschi

Paragrafo 3.2 : Luisa Franconi