



## **Inserimento professionale dei laureati 2011**

### **Descrizione del file**

## INDICE

1. Introduzione .....	3
2. Descrizione delle variabili .....	4
3. Metodologia statistica per la tutela della riservatezza.....	7
4. Riferimenti bibliografici.....	10
5. Contatti .....	10
5. Autori .....	10

## 1. Introduzione

I file di dati elementari per scopi di ricerca scientifica (MFR) possono essere rilasciati su richiesta motivata e previa autorizzazione del Presidente dell'Istituto, privi di elementi identificativi diretti e comunque secondo modalità che non ne permettano il collegamento con singole persone fisiche e/o giuridiche. Gli utenti dei file MFR sono obbligati a rispettare le condizioni specificate nel modulo di richiesta.

Per il rilascio di un file di dati elementari per la ricerca (MFR) per l'indagine *Inserimento professionale dei laureati*, periodo di riferimento 2011, è stata definita una metodologia statistica per la riduzione del rischio di violazione della riservatezza dei rispondenti.

Informazioni sullo *status* di ciascuna variabile nel MFR sono fornite nella sezione 2. La metodologia statistica per la tutela della riservatezza è brevemente illustrata nella sezione 3.

Questo documento illustra soltanto le modifiche dovute al metodo statistico di riduzione del rischio di violazione della riservatezza.

## 2. Descrizione delle variabili

In questo paragrafo s'illustrano in breve i cambiamenti indotti dal metodo di protezione dei dati adottato per l'indagine in questione *Inserimento professionale dei laureati*, periodo di riferimento 2011. Tali cambiamenti sono necessari per proteggere la riservatezza dei rispondenti. La protezione dei dati individuali è stata raggiunta modificando il contenuto informativo di alcune variabili.

Il file MFR può essere fornito in formato fisso oppure tab-delimited.

Le unità statistiche del MFR sono i laureati.

Il separatore decimale è “.”.

Il file MFR contiene 62000 records e 217 variabili.

La somma dei coefficienti di riporto all'universo (variabile COEFFIN) è pari a 300338.

La Tabella 1 sintetizza le variabili che sono state modificate o soppresse per motivi di tutela della riservatezza.

**Tabella 1. Lo status delle variabili nel file MFR modificate per motivi di tutela della riservatezza.**

Nome Campo	Descrizione Variabile	Modifica SDC
Progr_IL11	Progressivo del laureato	Soppressa
Classe_Corso_IL11	Aggregazione di classi/corsi	Ricodifica a 124 modalità
dI	Quale cittadinanza	Soppressa
dJ	Anno di nascita 19_ _	Sostituzione con l'età al momento del conseguimento della laurea
Q1_27	Mese di laurea	Soppressa
Q2_3_3pos	Codice della professione a 3 digit separati dal punto	Soppressa
Q2_14	Tipo di contratto	Modalità 07, 08 e 09 aggregate
Q2_22_AUGA	Reddito annuale netto (per i lavoratori autonomi)	Top e bottom coding
Q2_22_AUGC	Cifra a cui si avvicina di più il guadagno annuale netto (per i lavoratori autonomi che non l'hanno voluto indicare in modo puntuale)	Soppressa
Q2_22_DCGM	Reddito mensile netto (per i lavoratori dipendenti o a progetto)	Top e bottom coding

Q2_22_DCGC	Cifra a cui si avvicina di più il guadagno mensile netto (per i lavoratori dipendenti o a progetto che non l'hanno voluto indicare in modo puntuale)	Soppressa
Q2_25b	Mese inizio lavoro	Soppressa
Q2_42b	Mese inizio primo lavoro	Soppressa
Q4_3	Comune di residenza prima dell'iscrizione all'università	15 Comuni Grandi
Q4_4	Paese di residenza prima dell'iscrizione all'università	Soppressa
Q4_10	Attualmente vive: quale comune italiano	15 Comuni Grandi
Q4_11	Attualmente vive: quale altra nazione	Francia, Belgio, Germania, Regno Unito, Spagna, Svizzera, Stati Uniti d'America, Altro paese
Q4_20	Sede lavoro: quale altra nazione	Soppressa
Q5_4_3pos	Codice della professione del padre a 3 digit separati dal punto	Soppressa

La Tabella 2 sintetizza le variabili che sono state sopprese in quanto campi non significativi.

**Tabella 2. Campi non significativi.**

Nome Campo	Descrizione Variabile
dA	Ha ricevuto lettera ISTAT
dB	Vuole rispondere
dC	Ha conseguito una laurea nel 2007
dD	In che anno l'ha conseguita
dE	Che tipo di laurea ha conseguito
dF	Filler
Q1_6	Nazione di laurea del titolo di accesso
Q1_14	Quale altro fattore?
Q1_20	In quale paese si è recato all'estero per studio
Q1_24	Voto massimo previsto alla laurea

Q1_25	Voto laurea
Q1_26	Lode
Q1_37_b	Mese di conseguimento laurea specialistica/magistrale
Q2_3_4pos	Codice della professione a 4 digit separati dal punto
Q2_3_5pos	Codice della professione a 5 digit separati dal punto
Q2_15_pr	Impresa attiva nel privato, no profit o pubblico
Q2_18	Quanti dipendenti ha l'impresa
Q2_21	Ore retribuite
Q4_5	Sede università: città di residenza o altra
Q4_6	Sede università: si è trasferito nella città della sede
Q5_4_4pos	Codice della professione del padre a 4 digit separati dal punto
Q5_4_5pos	Codice della professione del padre a 5 digit separati dal punto

### **3. Metodologia statistica per la tutela della riservatezza**

#### **3.1. Scenari di violazione della riservatezza**

La valutazione statistica del rischio di violazione della riservatezza avviene attraverso diverse ipotesi su quali siano le intenzioni di un *intruder* e su quali informazioni abbia a disposizione.

Le unità statistiche (laureati) a rischio di violazione della riservatezza sono state identificate ipotizzando i seguenti scenari di violazione della riservatezza:

- uno definito dalle informazioni registrate sia nell'archivio censuario dei laureati e sia nei Curriculum Vitae (CV) disponibili on-line;
- il secondo definito da una identificazione spontanea causata dalla natura o dal dettaglio di registrazione di alcune variabili (valori anomali) presenti nell'archivio indagine.

In riferimento ai dati dell'archivio censuario e dei CV disponibili on-line, si è valutato il rischio di violazione della riservatezza utilizzando criteri derivati da regole di frequenza.

#### **3.2. Modifiche del contenuto informativo delle variabili**

##### **3.2.1. Soppressione delle variabili**

Le variabili sopprese dal MFR per motivi di tutela della riservatezza sono elencate di seguito:

1. variabili identificative dirette: *Progr\_IL11*;
2. variabili considerate troppo dettagliate: *dl*, *Q2\_3\_3pos*, *Q4\_4*, *Q4\_20*, *Q5\_4\_3pos*;
3. variabili che hanno un ridotto valore aggiunto al potenziale di ricerca: *Q1\_27*, *Q2\_22\_AUGC*, *Q2\_22\_DCGC*, *Q2\_25b*, *Q2\_42b*.

##### **3.2.2. Ricodifica globale**

La ricodifica globale è uno dei più semplici e noti metodi di protezione dei dati individuali.

Si può applicare solamente alle variabili categoriali. Essa consiste nell'accorpamento di alcune modalità della variabile categoriale. Ovviamente, le nuove modalità non hanno più il significato precedente.

Per il rilascio del MFR *Inserimento professionale dei laureati* sono state ricodificate le seguenti variabili:

1. *Classe\_Corso\_IL11*: Aggregazione di classi/corsi di laurea, è stata ricodificata in 124 modalità (la variabile ricodificata è denominata *ClasseCorsoAGG*).
2. *dJ*: Anno di nascita 19\_ \_\_, è stata sostituita con la variabile derivata “età al momento del conseguimento della laurea”. Quest’ultima è stata ricodificata con le seguenti modalità (la variabile ricodificata è denominata *etaLaurea4classi*):
  - 01: 21 e 22 anni;
  - 02: 23 e 24 anni;
  - 03: da 25 a 29 anni;
  - 04: 30 anni e oltre.
3. *Q2\_14*: Tipo di contratto, è stata ricodificata con le seguenti modalità (la variabile ricodificata è denominata *Q2\_14AGG*):
  - 01: Con un contratto di Formazione e lavoro;
  - 02: Con un contratto di apprendistato;
  - 03: Con un contratto collettivo nazionale di lavoro;
  - 04: Lavora per una agenzia interinale;
  - 05: Con un contratto a progetto o di collaborazione coordinata e continuativa;
  - 06: Con un contratto di prestazione d’opera occasionale;
  - 07: Con un altro tipo di contratto a termine o lavora senza un contratto o con un contratto verbale;
  - 10: Con un contratto a tempo indeterminato.
4. *Q4\_3*: Comune di residenza prima dell’iscrizione all’università, è stata ricodificata a livello di 15 comuni più grandi in termini di popolazione residente (la variabile ricodificata è denominata *Q4\_3\_Comuni15Grandi*)<sup>1</sup>.

Lo stesso vale per la variabile *Q4\_10*: Attualmente vive: quale comune italiano, (la variabile ricodificata è denominata *Q4\_10\_Comuni15Grandi*);
5. *Q4\_11*: Attualmente vive: quale altra nazione, è ricodificata con le seguenti modalità (la variabile ricodificata è denominata *Q4\_11\_AGG*):

---

<sup>1</sup> Per i rimanenti comuni è possibile considerare l’aggregazione in provincia della variabile corrispondente con dettaglio provinciale (*Q4\_2*: Provincia di residenza prima dell’iscrizione all’università).



- 206: Belgio;
- 215: Francia;
- 216: Germania;
- 219: Regno Unito;
- 239: Spagna;
- 241: Svizzera;
- 536: Stati Uniti d'America;
- 201: Altri Paesi.

### **3.2.3. Top e bottom coding delle variabili guadagno**

#### ***Lavoratori autonomi***

In corrispondenza della variabile Q2\_22\_AUGA: Nel 2010, qual è stato il guadagno annuale netto per questo lavoro (per i lavoratori autonomi), sono stati applicati metodi di bottom-coding e top-coding (la variabile ricodificata è denominata Q2\_22\_AUGA\_AGG).

In particolare, i valori inferiori e pari a 10.000 euro annuali netti sono stati ricodificati in 10.000, con significato "10.000 euro o meno".

I valori uguali e superiori a 30.000 euro mensili sono stati ricodificati in 30.000, con significato "30.000 euro e oltre".

I valori 99999 hanno il significato "non risponde".

#### ***Lavoratori dipendenti***

In corrispondenza della variabile Q2\_22\_DCGM: Qual è il suo guadagno mensile netto per questo lavoro (per i lavoratori dipendenti) sono stati applicati metodi di bottom-coding e top-coding (la variabile ricodificata è denominata Q2\_22\_DCGM\_AGG).

In particolare, i valori inferiori e pari a 500 euro mensili netti sono stati ricodificati in 500, con significato "500 euro o meno".

I valori uguali e superiori a 2.500 euro mensili sono stati ricodificati in 2.500, con significato "2.500 euro e oltre".

I valori 9999 hanno il significato "non risponde".

I coefficienti di riporto all'universo (variabile COEFFIN) non sono stati modificati.

## **4. Riferimenti bibliografici**

1. Istat (2004), Metodologie e tecniche di tutela della riservatezza nel rilascio di informazione statistica (Metodi e norme, n. 20, 2004), [http://www.istat.it/dati/catalogo/20040706\\_00/](http://www.istat.it/dati/catalogo/20040706_00/).
2. AA.VV. (2010). *Handbook on Statistical Disclosure Control*. ESSnet on Statistical Disclosure Control. <http://neon.vb.cbs.nl/casc/handbook.htm>. 18/08/2011.

## 5. Contatti

Per eventuali informazioni sulla tutela della riservatezza, contattare [rilascio.microdati@istat.it](mailto:rilascio.microdati@istat.it).

## 6. Autore

Laura Corallo

DIQR/DCIQ/PSS/1

Piazza Indipendenza 4, 00185 Roma, [corallo@istat.it](mailto:corallo@istat.it).