



Percorsi di studio e di lavoro dei diplomati dell'anno 2004

Indagine 2007

Descrizione del file

INDICE

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

1. Introduzione

I file di microdati sono collezioni campionarie di dati elementari relative ad alcune indagini svolte dall'Istat. I file di microdati per scopi di ricerca 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 singole persone fisiche o giuridiche. Per la costruzione del file per la ricerca relativo all' Indagine 2007 sui percorsi di studio e di lavoro dei diplomati dell'anno 2004 la valutazione statistica del rischio di violazione della riservatezza avviene a fronte di ipotesi su possibili modalità di violazione chiamate scenari di identificazione.

L'approccio utilizzato prevede quindi l'individuazione di possibili scenari, l'adozione della definizione di rischio, la sua misurazione, e, per ciascun scenario individuato, l'applicazione delle metodologie più idonee per la riduzione del rischio di identificazione.

Nella sezione 2 si forniscono indicazioni su ricodifiche o variazioni delle variabili del MFR rispetto al questionario di indagine. La metodologia statistica per la tutela della riservatezza è brevemente illustrata nella sezione 3.

L'utente del file di microdati per la ricerca deve rispettare le condizioni stabilite nel contratto di rilascio.

2. Descrizione delle variabili

Variabile	Nome variabile	SDC status	Breve Dscrizione
qh	Tipo di diploma specifico conseguito	Variato	Vedi variabile tipsc
SEZIONE 1: CURRICULUM SCOLASTICO			
q1_12	Voto diploma in centesimi	Invariato	
SEZIONE 3: STUDI UNIVERSITARI			
q3_11	Crediti acquisiti	Variato	1=0; 2=1-59; 3=60-119; 4=120-179; 5=180 o più; 9=non indicato
q3_12	Esami superati	Variato	1=0; 2=1-14; 3=15-24; 4=25 o più; 9=non indicato
q3_18	Regione dell'ateneo	Invariato	
Prov_ateneo	Provincia dell'ateneo		
q3_19	Stato estero dell'ateneo	Variato	Non rilasciato
q3_20	Quale ateneo	Invariato	
q3_22	Corso di laurea/diploma specifico	Variato	01=Gruppo scientifico 02=Gruppo chimico-farmaceutico 03=Gruppo geo-biologico 04=Gruppo medico 05=Gruppo ingegneria 06=Gruppo architettura 07=Gruppo agrario 08=Gruppo economico-statistico 09=Gruppo politico-sociale 10=Gruppo giuridico 11=Gruppo letterario 12=Gruppo linguistico 13=Gruppo insegnamento 14=Gruppo psicologico 15=Gruppo educazione fisica
q3_34	Corso 1^ immatricolazione specifico	Variato	1=Gruppo ingegneria; 2=Gruppo economico-statistico; 3=Gruppo scienze politiche;

4=Gruppo
giurisprudenza;
5=Altri Gruppi

SEZIONE 4: INTERRUZIONE DEGLI STUDI UNIVERSITARI

q4_2	Tipo di corso interrotto	Variato	1=Gruppo ingegneria; 2=Gruppo economico- statistico; 3=Gruppo politico sociale; 4=Gruppo giuridico; 5=Altri gruppi
q4_6	Crediti acquisiti	Variato	1=0; 2=1-59; 3=60 o più; 9=non indicato
q4_7	Numero esami sostenuti	Variato	1=0; 2=1 o più; 9=non indicato

SEZIONE 5: LAVORO ATTUALE

q5_26	Guadagno mensile netto (specifico)	Variato	fino a 1500 € arrotondato ai 25 €; da 1501 a 2000 arrotondato ai 50€; oltre 2000 € arrotondato ai 100 €
-------	---------------------------------------	---------	--

SEZIONE 8: NOTIZIE SULLA FAMIGLIA D'ORIGINE

q8_3	Titolo studio padre quando intervistato 14-enne	Variato	1=licenza elementare/analfabeta; 2=licenza media; 3=diploma di scuola secondaria superiore; 4=laurea/diploma universitario
q8_4	Titolo studio madre quando intervistato 14-enne	Variato	1=licenza elementare/analfabeta; 2=licenza media; 3=diploma di scuola secondaria superiore; 4=laurea/diploma universitario
q8_5	Condizione occupazionale padre quando intervistato 14-enne	Variato	1=occupato; 2=in cerca di occupazione;

			3=pensionato; 4=in altra condizione
q8_8	Posizione professione padre lav. autonomo	Variato	1 - imprenditore 2- libero professionista 3- lavoratore in proprio 4- coadiuvante nell'azienda di un familiare; socio di una cooperativa 5- lavoratore autonomo senza specifica qualificazione
q8_9	Posizione professione padre lav. dipendente	Variato	1- dirigente 2- quadro, un funzionario 3- impiegato ad alta/media qualificazione 4- impiegato esecutivo; operaio o lavoratore senza specifica qualificazione
q8_11	Condizione occupazionale madre quando intervistato 14-enne		1=occupata; 2=in cerca di occupazione; 3=casalinga; 4=pensionata; 5=in altra condizione
q8_13	Posizione professione madre lav. autonoma	Variato	1 - imprenditore 2- libero professionista 3- lavoratore in proprio 4- coadiuvante nell'azienda di un familiare; socio di una cooperativa 5- lavoratore autonomo senza specifica

			qualificazione
			1- dirigente 2- quadro, un funzionario 3- impiegato ad alta/media qualificazione 4- impiegato esecutivo; operaio o lavoratore senza specifica qualificazione
q8_14	Posizione professione madre lav. dipendente	Variato	

SEZIONE 9: NOTIZIE ANAGRAFICHE

q9_1	Cittadinanza italiana?	Variato	cancellata
q9_2	Che cittadinanza hai?	Variato	cancellata
			1=1982 o precedente; 2=1983; 3=1984; 4=1985; 5=1986 o successivo
q9_8	Anno di nascita	Variato	1=celibe/nubile; 2=altro
q9_9	Stato civile	Variato	

DATI RELATIVI ALLE SCUOLE

regscu	Codice Regione/Provincia autonoma della scuola	Invariato	
provscu	Codice provincia della scuola		
tipsc19	Codice tipo scuola 19 modalità	Variato	cancellata
tipsc	Codice tipo scuola 5 modalità	Invariato	
gestione	Gestione della scuola (1=Statale, 2=Pubblica non statale, 3=Privata)	Variato	(1=Pubblica, 2=Privata)

3. Metodologia statistica per la tutela della riservatezza

La metodologia utilizzata per la costruzione del file per la ricerca relativo all' Indagine 2007 sui percorsi di studio e di lavoro dei diplomati dell'anno 2004 prevede l'individuazione di possibili scenari di intrusione. Il rischio di violazione è valutato in relazione alle frequenze di combinazioni di variabili (presenti nello scenario) nella popolazione e nel campione e in relazione alle frazioni campionarie. L'idea di base è che combinazioni rare (ovvero con frequenza bassa) nella popolazione e nel campione siano a rischio di identificazione così come una elevata frazione campionaria in tali strati della popolazione può portare a possibili identificazioni (Istat , 2004, Willenborg e De Waal, 2001).

3.1. Possibili scenari di violazione della riservatezza

In una prima fase vengono definiti gli scenari di intrusione; ovvero vengono valutate le possibili strategie e l'utilizzo di informazioni esterne disponibili ad un eventuale utente che tenti di individuare un'unità statistica per effettuare un'identificazione.

Definiamo come:

- variabili identificative indirette (o chiave) quelle variabili che da sole, o in combinazione con altre, possono portare all'identificazione di uno o più record del file rilasciato;
- l'identificazione si verifica quando, con un certo grado di sicurezza si riesce a stabilire una relazione biunivoca tra le combinazioni delle variabili identificative indirette di un'unità presente nel file e quelle dell'individuo presente nell'archivio in possesso dell'intruso.
- la definizione di violazione della riservatezza adottata è basata sul concetto di identificabilità del soggetto a cui si fa riferimento e con la possibilità per l'intruso di acquisire informazioni riservate relative all'unità statistica a partire dalle informazioni rilasciate.

Due sono gli scenari che prendiamo in considerazione: l'identificazione da archivi esterni e l'identificazione spontanea sulla base di conoscenze a priori (Circolo delle Conoscenze).

Identificazione da Archivi Esterni

L'ipotesi alla base di questo scenario è che esista un archivio esterno pubblico e nominativo contenente informazioni associabili a quelle identificative indirette rilasciate nel

file. In questo caso si assume che l'intruso avvalendosi di tale archivio nominativo riesca a identificare l'unità statistica ottenendo così informazioni su di essa.

Circolo delle conoscenze/ identificazione spontanea

Un altro tipo di identificazione è quella che può avvenire all'interno del cosiddetto "circolo delle conoscenze" ovvero l'unità statistica viene individuata da un utente sulla base delle conoscenze dirette dello stesso.

3.2. Ricodifica globale

Il metodo di protezione che si è scelto di adottare è il metodo della "ricodifica globale". Esso consiste nella ricodifica di una variabile che si ottiene aggregando più modalità in una sola per l'intero file.

I dettagli sulle variabili coinvolte nel processo di protezione e il tipo di ricodifica applicata sono desumibili dalla tabella riportata nel secondo paragrafo.

4. Riferimenti bibliografici

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/

Willenborg, L. e de Waal, T. (2001). Elements of statistical disclosure control. Lecture Notes in Statistics, 115, New York: Springer-Verlag.

Numerosi riferimenti bibliografici sul tema della tutela statistica della riservatezza sono reperibili sul sito <http://neon.vb.cbs.nl/casc/> nella documentazione prodotta nell'ambito dei diversi progetti europei che si sono occupati del tema: CASC project (2000-2003); CENEX project (2006); ESSnet project (2008-2009).

5. Contatti

Per ulteriori dettagli su metodi di tutela della riservatezza, contattare rilascio.microdati@istat.it.

Curatori dei capitoli

Il documento è stato redatto da Alessandra Capobianchi.