

Percorsi di studio e di lavoro dei
diplomati
Indagine 2004

Manuale utente e tracciato record

Il manuale, frutto della collaborazione fra il Servizio Popolazione, Istruzione e Cultura e il Servizio progettazione e supporto metodologico nei processi di produzione statistica, è stato curato da Valentino Parisi e Fabio Massimo Rottino.

L'Appendice D è stata redatta da Claudia De Vitiis ed Emanuela Scavalli
L'Appendice E è stata redatta da Alessandra Capobianchi

Per informazioni rivolgersi all'U.O. DEM/D tel. 06/46737592

Indice

PREMESSA	4
1. L'indagine sui percorsi di studio e di lavoro dei diplomati	5
<i>1.1 Le principali caratteristiche dell'indagine</i>	5
<i>1.2 Il processo di produzione dei dati</i>	6
<i>1.3 I contenuti informativi</i>	8
2. Le variabili contenute nel file	12
 APPENDICE A: TRACCIATO RECORD	 A-1
 APPENDICE B: ALLEGATI	 B-1
 APPENDICE C: DEFINIZIONI	 C-1
 APPENDICE D: LA STRATEGIA DI CAMPIONAMENTO E LA VALUTAZIONE DEL LIVELLO DI PRECISIONE DELLE STIME	 D-1
 APPENDICE E: LA METODOLOGIA PER LA STIMA DEL RISCHIO DI VIOLAZIONE	 E-1

Premessa

Il Decreto Legislativo n° 322 del 6/9/1989 regola la diffusione delle informazioni statistiche prodotte nell'ambito del Sistema Statistico Nazionale al fine di garantire la riservatezza dei rispondenti. In particolare, per la diffusione di dati elementari, l'articolo 10, comma 2, dispone quanto segue: *"Sono distribuite altresì, ove disponibili, su richiesta motivata e previa autorizzazione del Presidente dell'ISTAT, collezioni campionarie di dati elementari, resi anonimi e privi di ogni riferimento che ne permetta il collegamento con singole persone fisiche e giuridiche"*.

Nell'osservanza di tale legge l'ISTAT ha adottato misure e tecniche che rendono impossibile, o altamente improbabile, il collegamento dei dati rilasciati con l'unità statistica a cui si riferiscono. Per tale motivo sono state apportate alcune modifiche sui files originali delle indagini, nell'intento di garantire la massima protezione ai dati contenendo al minimo l'eventuale perdita di informazione.

Le metodologie applicate si concretizzano nell'accorpamento e/o riclassificazione di modalità di variabili e nell'oscuramento di variabili per alcuni record.

Dopo una breve descrizione delle fasi principali dell'indagine sui percorsi di studio e di lavoro dei diplomati, viene riportato nell'appendice A il tracciato record che descrive le variabili contenute nel file standard.

Per agevolare l'elaborazione dei dati e l'interpretazione dei risultati, nelle appendici sono riportate le classificazioni adottate e vengono illustrate le caratteristiche del disegno di campionamento e la metodologia adottata per la protezione dei dati.

Con l'occasione si ricorda al richiedente che si impegna a:

- utilizzare i dati soltanto per gli scopi dichiarati;
- non fornire a terzi i dati elementari, consentendone l'accesso, sotto la propria responsabilità, soltanto alle persone direttamente coinvolte nel lavoro per il quale essi sono stati richiesti;
- citare la fonte Istat nell'eventuale divulgazione di elaborazioni dei dati;
- inviare alla Biblioteca dell'Istat due copie delle pubblicazioni eventualmente prodotte con l'utilizzo dei dati ottenuti.

1. L'indagine sui percorsi di studio e di lavoro dei diplomati

1.1 Le principali caratteristiche dell'indagine e il disegno dell'indagine

Nel 2004 l'Istituto Nazionale di Statistica ha realizzato la terza edizione¹ dell'Indagine sui percorsi di studio e di lavoro dei diplomati, che ha avuto come oggetto la coorte dei diplomati del 2001 intervistati a tre anni dal conseguimento del titolo.

L'indagine fa parte del sistema integrato di indagini sulla transizione scuola-lavoro che l'Istat ha realizzato con l'obiettivo di analizzare le attività di formazione e di lavoro svolte da coloro che hanno conseguito un titolo di studio post-obbligo; tale sistema comprende, oltre alla rilevazione sui percorsi di studio e di lavoro dei diplomati, anche l'indagine sull'inserimento professionale dei laureati² e l'indagine sull'inserimento professionale dei diplomati universitari³.

Tutte le indagini del sistema sulla transizione scuola-lavoro hanno caratteristiche metodologiche comuni: le rilevazioni hanno periodicità triennale e sono condotte su singole coorti di diplomati/laureati⁴ intervistati a circa tre anni dal conseguimento del titolo.

In tutte e tre le rilevazioni, a partire dal 2001⁵, per effettuare le interviste viene utilizzata la tecnica C.A.T.I. (Computer Assisted Telephone Interview). Le indagini risultano integrate anche per ciò che concerne i principali contenuti informativi: i modelli di rilevazione utilizzati nelle indagini hanno una struttura affine e molti quesiti in comune che presentano, per quanto possibile, una formulazione omogenea.

La rilevazione si svolge in due fasi: la prima, coinvolge le scuole secondarie superiori, la seconda è rivolta direttamente ai diplomati.

¹ La prima edizione si è svolta nel 1998 e ha intervistato i diplomati che hanno conseguito il titolo di scuola secondaria superiore nell'anno 1995.

² Cfr. Istat (2003): I laureati e il mercato del lavoro – Inserimento professionale dei laureati - Indagine 2001, Collana Informazioni.

Istat (2004): I laureati e lo studio – Inserimento professionale dei laureati - Indagine 2001, Collana Informazioni.

³ Cfr. Istat (2005): Inserimento professionale dei diplomati universitari – Indagine 2002, Collana Informazioni.

In considerazione dell'esaurimento dei corsi di diploma universitario, dovuto all'applicazione della riforma dei cicli universitari, l'indagine condotta nel 2002, seconda edizione, è stata l'ultima.

⁴ Le coorti di diplomati/laureati sono definite in base all'anno solare di conseguimento del titolo.

⁵ L'indagine sull'inserimento professionale dei laureati prevedeva, fino al 1998, l'autocompilazione del questionario da parte dei soggetti intervistati ed il successivo invio del modello per posta ordinaria. Le altre due rilevazioni hanno utilizzato fin dal loro inizio la tecnica C.A.T.I..

In occasione di questa edizione d'indagine, nella prima fase della rilevazione sono state introdotte principalmente due innovazioni: la prima ha interessato le modalità di selezione dei diplomati (le scuole hanno fornito l'elenco completo di tutti i diplomati), la seconda ha riguardato le modalità di compilazione e di trasmissione dati da parte delle scuole che per l'inserimento, la validazione e la trasmissione delle informazioni relative ai diplomati hanno potuto utilizzare anche un apposito sito web. Allo scopo di migliorare la qualità delle informazioni di base è stato predisposto un sistema di acquisizione controllata dei dati che prevede la segnalazione immediata della presenza di errori nella compilazione di alcuni campi (o variabili) e l'assenza di valori nei campi obbligatori (quali ad esempio il nominativo, il sesso o il recapito del diplomato).

Altre innovazioni hanno riguardato la seconda fase della rilevazione.

La struttura e i contenuti di alcune sezioni del modello sono state rivisitate e aggiornate: in particolare, in collaborazione con l'Isfol, è stata approfondita la sezione riguardante la formazione professionale post-diploma per indagare le caratteristiche dei percorsi formativi extra-universitari.

La seconda ha riguardato lo sviluppo in-house del software per la gestione, tramite il sistema BLAISE, del questionario elettronico CATI, del piano di schedulazione delle chiamate e di monitoraggio di tutti i tentativi di contatto, tramite produzione di reportistica. All'interno del questionario sono stati inoltre introdotti dei controlli di coerenza⁶ tra le informazioni fornite dagli intervistati ai diversi quesiti e tra queste e le notizie raccolte dalle scuole ove è stato conseguito il titolo.

La terza innovazione ha portato all'adozione di due nuovi sistemi di codifica assistita nel corso dell'intervista telefonica che, grazie alla possibilità di interagire direttamente con l'intervistato, hanno prodotto un consistente miglioramento nella codifica dei quesiti aperti relativi al corso universitario frequentato o interrotto e dei quesiti relativi alle professioni svolte dal diplomato e dal padre⁷.

1.2 Il processo di produzione dei dati

L'insieme degli istituti di scuola secondaria superiore ove nel 2001 si sono tenuti esami di stato per il conseguimento del diploma di scuola secondaria superiore è stato suddiviso in strati sulla base della regione di ubicazione della scuola, del tipo di indirizzo di studi e della dimensione in termini di

⁶ I controlli di coerenza introdotti sono di due tipi: hard e soft. In entrambi i casi al verificarsi dell'incompatibilità, il sistema visualizza un messaggio che segnala l'incoerenza; l'intervistatore ha il compito di formulare nuovamente uno o più quesiti coinvolti e di digitare la nuova risposta. Per i controlli di tipo hard l'incompatibilità deve comunque essere sanata pena l'impossibilità di proseguire l'intervista.

⁷ Per la rilevazione della variabile Professione è stata utilizzata la Classificazione delle professioni 2001 (Istat, Collana Metodi e norme, n.12,2001)

numero di diplomati. Si è proceduto all'estrazione di 1.868 scuole (unità di primo stadio) appartenenti ai diversi strati: le informazioni richieste e fornite dalle scuole estratte per ogni diplomato hanno riguardato le notizie anagrafiche e il percorso scolastico dei singoli diplomati, informazioni utilizzate anche in corso d'intervista per effettuare controlli di coerenza con le risposte fornite dagli intervistati.

Obiettivo principale della prima fase dell'indagine è stato quello di ottenere la lista e i recapiti telefonici dei diplomati sui quali condurre l'indagine vera e propria. La lista comprendeva tutte le informazioni necessarie per reperire i diplomati, con particolare interesse per il recapito telefonico e postale del diplomato e i nominativi dei genitori.

E' stato così ottenuto l'elenco da cui estrarre i nominativi da contattare (unità di secondo stadio).

Nella seconda fase della rilevazione, al fine di favorire la disponibilità a rilasciare l'intervista, a tutti i diplomati appartenenti al campione base e al campione suppletivo, è stata inviata una lettera di avviso, a firma del presidente dell'Istituto, al fine di presentare le finalità dell'indagine e rendere note le modalità di selezione e di partecipazione all'indagine.

I diplomati selezionati sono stati poi contattati telefonicamente ed intervistati da una ditta specializzata.

Le interviste sono state effettuate nell'arco di 12 settimane, dal 28 settembre al 18 dicembre 2004, a poco più di 3 anni dal conseguimento del diploma di scuola secondaria superiore. Tutti i nominativi forniti alla società (poco più di 29.000, di cui 20.430 unità base) sono stati oggetto di tentativi di contatto. Per i recapiti telefonici risultati errati la società esterna e il personale interno dell'Istituto hanno provveduto alla ricerca dei nuovi recapiti sulla base dei nominativi dei genitori, della provincia di ubicazione della scuola e del comune di residenza del diplomato durante gli studi superiori.

Il 70,2% dei nominativi forniti sono stati oggetto di una intervista completa (per un totale di 20.408), il 3,5% delle interviste sono state rifiutate, lo 0,4% interrotte, lo 0,9% hanno riguardato soggetti fuori target⁽⁸⁾, mentre con il restante 25% non è stato stabilito alcun contatto.

Gli intervistatori formati per effettuare l'indagine sono stati 131; ognuno ha effettuato mediamente 6,8 interviste al giorno. L'introduzione delle codifiche on line non ha comportato un significativo allungamento dei tempi di intervista rispetto alle precedenti edizioni d'indagine: si è passati da una durata media pari a 13 minuti e 20 secondi nell'indagine realizzata nel 2001 a 13 minuti e 43 secondi della rilevazione condotta nel 2004.

Al termine della fase di raccolta dei dati, sebbene fossero presenti regole di coerenza all'interno del questionario C.A.T.I., ha avuto inizio il processo di correzione dei dati.

(8) Si definisce "fuori target" un soggetto che non ha conseguito il diploma di scuola secondaria superiore nel 2001.

Attraverso l'utilizzo di regole di localizzazione e di correzione di tipo deterministico e di procedure, costruite con l'ausilio di software per la correzione di dati errati e l'imputazione dei dati mancanti, che utilizzano metodologie di tipo probabilistico, si è arrivati alla definizione di dati corretti.

Ai fini della ricostruzione dell'universo di riferimento, ai dati relativi alle singole interviste sono stati applicati stimatori di ponderazione vincolati per regione, tipo di scuola, sesso e strato di appartenenza, secondo la metodologia di stima utilizzata dall'Istat.

Le elaborazioni dei dati ha permesso la successiva fase di validazione con i dati provenienti dalle serie storiche, da altre rilevazioni condotte all'interno dell'Istituto (in particolare le altre indagini appartenenti al Sistema sulla transizione scuola-lavoro) e da rilevazioni della stessa area condotte da altri enti (soprattutto con i dati amministrativi sugli iscritti e i diplomati delle scuole secondarie superiori raccolti, con cadenza annuale, dal Ministero dell'Istruzione).

Il processo di produzione termina con la diffusione dei dati attraverso diverse modalità: la pubblicazione del volume cartaceo (all'interno della Collana Informazioni), il sito internet dell'Istat, il file standard, il Laboratorio di Analisi dei Dati Elementari ADELE.

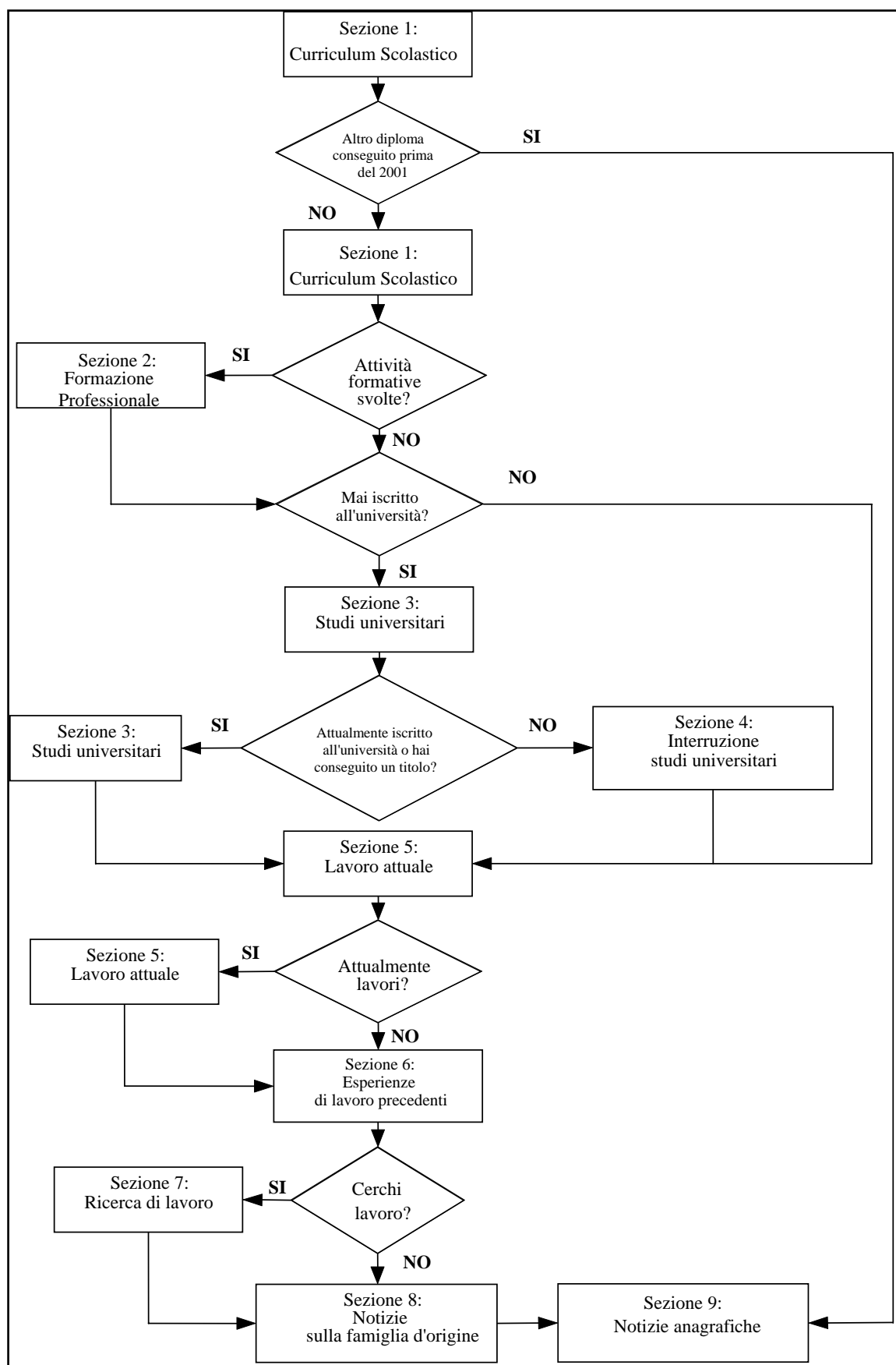
1.3 I contenuti informativi

Il questionario è costituito da nove sezioni tematiche, ulteriormente articolate all'interno, ed è strutturato in modo tale da assicurare la confrontabilità con le informazioni raccolte nelle altre due indagini sulla transizione, nonché l'integrazione con altre fonti, in particolare con l'indagine sulle forze di lavoro.

Le prime quattro sezioni del questionario sono dedicate al curriculum degli studi e alle attività di qualificazione dopo il diploma, le successive tre sezioni al lavoro e alla ricerca del lavoro, mentre nelle ultime due si chiedono notizie relative alla famiglia d'origine e informazioni anagrafiche.

Le diverse sezioni del questionario e i principali quesiti di snodo sono presentati nello schema 1; di seguito vengono illustrati, invece, gli obiettivi e i contenuti delle diverse sezioni.

Schema 1 - Sezioni e principali snodi del questionario sull'Inserimento professionale dei diplomati



SEZIONE 1 – CURRICULUM SCOLASTICO

La sezione riguarda tutti gli intervistati con eccezione di coloro che hanno conseguito nel 2001 un ulteriore titolo di studi superiori che esulano l'interesse specifico dell'indagine. Le domande presenti in questa sezione sono mirate a definire la carriera scolastica, il possesso di eventuali ulteriori titoli di scuola secondaria superiore, gli eventuali cambiamenti di corso di studi, l'eventuale frequenza di scuole per il recupero di anni scolastici, la conoscenza delle lingue straniere, le conoscenze informatiche e il giudizio sulla qualità dell'istruzione secondaria superiore ricevuta.

SEZIONE 2 - CORSI DI FORMAZIONE POST-SECONDARIA

Questa sezione interessa gli intervistati che hanno frequentato o stanno frequentando attività formative post-diploma (extrauniversitarie). I principali contenuti informativi di questa sezione riguardano la partecipazione ad attività formative post-diploma, la conclusione, l'interruzione e l'iscrizione a corsi di formazione, le caratteristiche del corso (durata, argomento, tipologia, certificazione rilasciata), le attività concomitanti e successive al corso, il grado di soddisfazione e le aspettative sugli esiti del corso.

SEZIONE 3 - STUDI UNIVERSITARI

Sezione destinata a coloro che al momento dell'intervista sono iscritti all'università o che hanno già conseguito un titolo universitario. Le domande relative a questa sezione raccolgono informazioni sull'iscrizione o meno all'università, sui motivi d'iscrizione o di non iscrizione, sul possesso di eventuali titoli accademici, sulla conoscenza e sul giudizio in merito alla riforma del sistema universitario, sulle modalità di orientamento per la scelta dell'indirizzo universitario, sul corso di studi universitario frequentato, sull'anno di prima immatricolazione, sulla frequenza alle lezioni, sulla valutazione della didattica dei docenti e sull'eventuale cambio di corso di studi.

SEZIONE 4 - INTERRUZIONE DEGLI STUDI

Rivolta a coloro che hanno abbandonato gli studi universitari. Le domande di questa sezione riguardano le caratteristiche del corso universitario interrotto, la motivazione dell'interruzione, l'anno d'immatricolazione e la valutazione della didattica dei docenti.

SEZIONE 5 – LAVORO ATTUALE

Indirizzata a coloro che al momento dell'intervista svolgono un'attività lavorativa retribuita. Le domande sono mirate ad individuare le attività d'inserimento lavorativo post-diploma, il motivo di eventuale inattività, il tipo di lavoro svolto, la posizione nella professione, la regolarizzazione del rapporto di lavoro, il tipo di contratto, la professione svolta, il settore di attività economica in cui opera, l'orario di lavoro, il guadagno mensile netto, la data d'inizio dell'attuale lavoro, i requisiti

richiesti per iniziare l'attività (titolo di studio e votazione minima), il grado di soddisfazione nei confronti del lavoro e la modalità attraverso cui hanno trovato lavoro.

SEZIONE 6 – ESPERIENZE DI LAVORO PRECEDENTI

La sezione ha l'obiettivo di rilevare informazioni sulla prima esperienza lavorativa iniziata e terminata dopo il conseguimento del diploma. I quesiti portano ad esaminare i seguenti aspetti: le opportunità di lavoro post-diploma, il motivo di un eventuale rifiuto, il tipo di lavoro interrotto o concluso, la data d'inizio della prima attività lavorativa conclusa, il motivo d'interruzione e/o di conclusione e l'eventuale ricerca di lavoro.

SEZIONE 7 - RICERCA DI LAVORO

La sezione è dedicata esclusivamente a quanti, tra coloro che lavorano o non lavorano, dichiarano di cercare lavoro. I quesiti indagano sul tempo trascorso dall'ultima iniziativa di ricerca di lavoro, sul tipo di lavoro preferito, sull'orario di lavoro preferito, sulla propensione a cambiare città, sul guadagno mensile minimo richiesto e, infine, sulla disponibilità ad iniziare un nuovo lavoro entro due settimane.

SEZIONE 8 - NOTIZIE SULLA FAMIGLIA D'ORIGINE

Le domande di questa sezione, rivolte a tutti gli intervistati, concernono il contesto della famiglia d'origine al momento in cui l'intervistato ha iniziato gli studi superiori (mediamente a 14 anni). Gli aspetti rilevati riguardano: l'aiuto familiare nello studio, l'influenza familiare nella scelta scolastica, il livello di istruzione, l'attività, la posizione nella professione e il settore di attività economica in cui i genitori sono occupati e, per il padre, la professione svolta.

SEZIONE 9 - NOTIZIE ANAGRAFICHE

Rivolta a tutti gli intervistati, con quesiti che mirano a raccogliere le principali notizie anagrafiche quali la cittadinanza, la provincia di residenza, l'eventuale diverso domicilio, lo svolgimento del servizio militare o civile, l'anno di nascita e lo stato civile.

1.4 Le variabili contenute nel file

Per la protezione della riservatezza dei rispondenti all'indagine sui percorsi di studio e di lavoro dei diplomati del 2001, l'Istat ha adottato una metodologia di protezione dei dati basata su una definizione di rischio di violazione della riservatezza a livello individuale. Tale tecnica ha permesso un ulteriore miglioramento della tutela della riservatezza delle unità effettivamente a rischio e il rilascio di un maggiore contenuto informativo. Conseguenza dell'utilizzo di tale metodologia di soppressione (vedi Appendice D) è la presenza di alcuni *blank* nelle variabili regione della scuola ed anno di nascita del diplomato. Inoltre per alcune variabili sono state costruite delle aggregazioni.

Nel tracciato record (Appendice A) vengono riportati il nome, l'ampiezza del campo in *bytes*, la posizione ed il contenuto informativo di ciascuna variabile.

Nella colonna VARIABILE sono riportati i nomi delle variabili che, per facilitare il collegamento con il questionario, sono definiti dal numero del quesito corrispondente. Nella colonna DESCRIZIONE vengono invece indicati i contenuti sintetici delle variabili, descrivendo le codifiche adottate solo nei casi in cui si sia proceduto a codifiche diverse da quelle presenti nel questionario.

La variabile PESOFIN individua il coefficiente di riporto all'universo, cioè il peso che ogni unità campionaria assume nella popolazione di riferimento. Ad esempio, se in un record il valore di PESOFIN è pari a 30, vuol dire che questa unità rappresenta se stessa ed altre 29 unità della popolazione che non sono state incluse nel campione.

APPENDICE A: TRACCIATO RECORD

Variabili	Lunghezza	Colonna iniziale	Colonna Finale	Descrizione quesiti	
tipo_scuola	2	1	2	Tipo di scuola	vedi Allegato TIPO DI SCUOLA, appendice B
gestione	1	3	3	Gestione della scuola	1=pubblica, 2=privata
regscu	2	4	5	Regione della scuola	vedi Allegato REGIONI, appendice B
ripscu	1	6	6	Ripartizione della scuola	vedi Allegato RIPARTIZIONI, appendice B
q_sesso	1	7	7	Sesso	
q1_1	1	8	8	Altro diploma, oltre quello del 2001?	
q1_2	1	9	9	Tipo di scuola altro diploma	
q1_3	1	10	10	Conseguito altro diploma prima/dopo 2001	
q1_4	1	11	11	Motivo conseguimento altro diploma	
q1_5	1	12	12	Cambiato tipo di scuola durante studi superiori?	
q1_6	1	13	13	Tipo scuola (gruppo) di iscrizione dopo medie	
q1_7	1	14	14	Scuola di iscrizione dopo medie pubblica o privata	
q1_8	1	15	15	Frequentato scuole per recupero anni?	
q1_9	1	16	16	Sei stato respinto?	
q1_10	1	17	17	Numero di ripetenze	1=1; 2=2 o più
q1_11	1	18	18	Esame diploma come privatista?	
q1_12	1	19	19	Voto diploma	1=60-69; 2=70-79; 3=80-89; 4=90-100
q1_13	1	20	20	Voto licenza media inferiore	
q1_14	1	21	21	Studiato lingue straniere?	
q1_15	1	22	22	Soddisfazione lingue straniere	
q1_16	1	23	23	Studiato informatica?	
q1_17	1	24	24	Soddisfazione studio informatica	
q1_18a	1	25	25	Soddisfazione su istruzione: rapporto insegnanti	
q1_18b	1	26	26	Soddisfazione su istruzione: capacità professionali insegnanti	
q1_18c	1	27	27	Soddisfazione su istruzione: contenuti studio	
q1_18d	1	28	28	Soddisfazione su istruzione: strutture scolastiche	
q1_19	1	29	29	Svolto stage o tirocinio organizzato dalla scuola?	
q2_1	1	30	30	Iscrizione a corso di formazione post diploma	
q2_2	2	31	32	Perché non iscritto?	
q2_3	1	33	33	Iscritto a quanti corsi?	
q2_4	1	34	34	Concluso corso?	

Variabili	Lunghezza	Colonna iniziale	Colonna Finale	Descrizione quesiti	
q2_5	1	35	35	Frequenza corso	
q2_6	1	36	36	Ottenuta certificazione?	
q2_7	1	37	37	Motivo interruzione corso	
q2_8a	1	38	38	Quando interrotto/concluso (anno)?	
q2_8b	1	39	39	Quando interrotto/concluso (mese)?	1=da gennaio a giugno; 2=da luglio a dicembre; 9=non indicato
q2_9	1	40	40	Tipo di attività formativa	
q2_10	1	41	41	Tipo di corso pubblico	
q2_11	1	42	42	Argomento del corso	
q2_12	1	43	43	Durata del corso in ore	1=meno di 200 ore; 2=da 200 a 599 ore; 3=da 600 a 1199 ore; 4=da 1200 ore in poi; 9=non indicato
q2_13	1	44	44	Durata del corso in mesi	1=fino a 3 mesi; 2=da 4 a 5 mesi; 3=da 6 a 11 mesi; 4=da 12 mesi in poi; 9=non indicato;
q2_14	1	45	45	Rilasciata certificazione?	
q2_14a	1	46	46	Tipo di certificazione - (1^ indicazione)	
q2_14b	1	47	47	Tipo di certificazione - (2^ indicazione)	
q2_14c	1	48	48	Tipo di certificazione - (3^ indicazione)	
q2_15a	1	49	49	Nel corso: attività di gruppo?	
q2_15b	1	50	50	Nel corso: laboratori?	
q2_15c	1	51	51	Nel corso: formazione a distanza?	
q2_15d	1	52	52	Nel corso: stage/tirocinio?	
q2_15e	1	53	53	Nel corso: visita studio in azienda?	
q2_15f	1	54	54	Nel corso: assistenza ricerca lavoro?	
q2_16	2	55	56	Motivo iscrizione/frequenza corso	
q2_17	2	57	58	Modi di informazione sul corso	
q2_18a	1	59	59	Altri impegni oltre al corso - 1^ indicazione	
q2_18b	1	60	60	Altri impegni oltre al corso - 2^ indicazione	
q2_18c	1	61	61	Altri impegni oltre al corso - 3^ indicazione	
q2_19	1	62	62	Cosa hai fatto nei 6 mesi post-corso?	
q2_20a	1	63	63	Soddisfazione su corso: strutture e strumenti didattici	
q2_20b	1	64	64	Soddisfazione su corso: competenza docenti	
q2_20c	1	65	65	Soddisfazione su corso: organizzazione	

Variabili	Lunghezza	Colonna iniziale	Colonna Finale	Descrizione quesiti	
q2_20d	1	66	66	Soddisfazione su corso: attualità contenuti	
q2_20e	1	67	67	Soddisfazione su corso: opportunità lavoro	
q2_21	1	68	68	Utili corsi per futuro professionale?	
q2_22a	1	69	69	Utilità corso per: cultura generale	
q2_22b	1	70	70	Utilità corso per: competenze specialistiche	
q2_22c	1	71	71	Utilità corso per: capacità lavorare in gruppo	
q2_22d	1	72	72	Utilità corso per: uso tecnologie informatiche	
q2_22e	1	73	73	Utilità corso per: uso lingue straniere	
q3_1	1	74	74	Mai iscritto all'università?	
q3_2	2	75	76	Perché non iscritto?	
q3_3	1	77	77	Motivo iscrizione	
q3_4	1	78	78	Conseguito titolo universitario?	
q3_6	1	79	79	Attualmente iscritto?	
q3_7	1	80	80	Motivo inizio nuovo corso	
q3_8	1	81	81	Tipo di corso	
q3_9	1	82	82	Ti iscriverai a corso 2° livello?	
q3_10	1	83	83	Puoi dare parere su sistema 3+2?	
q3_11a	1	84	84	Parere su sistema 3+2: qualità offerta formativa	
q3_11b	1	85	85	Parere su sistema 3+2: preparazione culturale laureati	
q3_11c	1	86	86	Parere su sistema 3+2: formare profili adeguati al mercato lavoro	
q3_11d	1	87	87	Parere su sistema 3+2: fenomeno abbandoni studi	
q3_11e	1	88	88	Parere su sistema 3+2: fenomeno fuori corso	
q3_12	1	89	89	Giudizio complessivo su riforma	
q3_13	1	90	90	Ti sei informato sui corsi?	
q3_14	1	91	91	Modi di informazione sui corsi	
reg_ateneo	1	92	92	Regione dell'ateneo	1=regione diversa rispetto a quella della scuola; 2=stessa regione
rip_ateneo	1	93	93	Ripartizione dell'ateneo	vedi Allegato RIPARTIZIONI, appendice B
q3_17	2	94	95	Gruppo di corsi universitari	vedi Allegato GRUPPI DI CORSI UNIVERSITARI, appendice B
q3_18	1	96	96	Utilità preparazione scuola superiore	
q3_19	1	97	97	Anno di immatricolazione all'università	
q3_19bis	1	98	98	Rivolto istituti privati per preparare esami?	
q3_20	1	99	99	Svolto periodo formazione all'estero?	

Variabili	Lunghezza	Colonna iniziale	Colonna Finale	Descrizione quesiti	
q3_21	1	100	100	Frequentato lezioni esame fondamentale?	
q3_22a	1	101	101	Soddisfazione su docente: competenza	
q3_22b	1	102	102	Soddisfazione su docente: chiarezza espositiva	
q3_22c	1	103	103	Soddisfazione su docente: presenza lezioni	
q3_22d	1	104	104	Soddisfazione su docente: presenza ricevimento	
q3_22e	1	105	105	Soddisfazione su docente: disponibilità	
q3_23	1	106	106	Frequentato lezioni nell'ultimo a.a.?	
q3_24	1	107	107	Frequenza lezioni ultimo a.a.	
q3_25	1	108	108	Cambiato corso?	
q3_26	1	109	109	Stavi lavorando quando iscritto università?	
q3_27	1	110	110	Meglio università o lavoro?	
q4_1a	1	111	111	Interruzione corso di laurea - 1 ^ indicazione	
q4_1b	1	112	112	Interruzione corso di laurea - 2 ^ indicazione	
q4_2	1	113	113	Tipo di corso interrotto	1=Gruppo ingegneria; 2=Gruppo economico-statistico; 3=Gruppo scienze politiche; 4=Gruppo giurisprudenza; 5=Altri Gruppi
q4_3	1	114	114	Motivo interruzione	
q4_4	1	115	115	Anno immatricolazione università	
q4_5	1	116	116	Numero esami sostenuti	1=nessuno; 2=da 1 a 5; 3=più di 5
q4_6	1	117	117	Frequentato lezioni insegnamento fondamentale?	
q4_7	1	118	118	Quante lezioni?	
q4_8a	1	119	119	Soddisfazione su docente: competenza	
q4_8b	1	120	120	Soddisfazione su docente: chiarezza espositiva	
q4_8c	1	121	121	Soddisfazione su docente: presenza lezioni	
q4_8d	1	122	122	Soddisfazione su docente: presenza ricevimento	
q4_8e	1	123	123	Soddisfazione su docente: disponibilità	
q5_1a	1	124	124	Attività post diploma: tirocinio/stage	
q5_1b	1	125	125	Attività post diploma: praticantato	
q5_1c	1	126	126	Attività post diploma: borsa di studio/lavoro	
q5_1d	1	127	127	Attività post diploma: PIP	

Variabili	Lunghezza	Colonna iniziale	Colonna Finale	Descrizione quesiti	
q5_2	1	128	128	Attualmente lavori?	
q5_3	2	129	130	Perché non lavori?	
q5_4	1	131	131	Posizione professione	
q5_5	1	132	132	Posizione professione lavoratori autonomi	
q5_6	1	133	133	Retribuito con partita IVA?	
q5_7	1	134	134	Retribuito con ritenuta d'acconto?	
q5_8	1	135	135	Posizione professione lavoratori dipendenti	
q5_9	1	136	136	Datore di lavoro versa contributi?	
q5_10	1	137	137	Tipo di lavoro (occasionale/continuativo)	
q5_11	1	138	138	Perché lavoro occasionale?	
q5_12	1	139	139	Lavoro a tempo determinato/indeterminato	
q5_13	1	140	140	Perché a tempo determinato?	
q5_14	2	141	142	Tipo di contratto	
q5_15	1	143	143	Codifica Istat professione diplomato	vedi Allegato PROFESSIONI, appendice B
q5_16	1	144	144	Settore lavorativo	
q5_17	1	145	145	Settore industriale	
q5_18	2	146	147	Settore servizi	
q5_19	1	148	148	Tempo pieno o part time?	
q5_20	1	149	149	Perché part-time?	
q5_21	1	150	150	Ore di lavoro settimanali	1=fino a 20 ore; 2=21-35 ore; 3=36-40 ore; 4=oltre 40 ore
q5_22	4	151	154	Guadagno mensile netto	fino a 1500 € arrotondato ai 25 €; da 1501 a 2000 arrotondato ai 50€; oltre 2000 € arrotondato ai 100 €
q5_23	1	155	155	Hai dovuto trasferirti?	
q5_24	1	156	156	Lavoro cominciato prima o dopo diploma?	
q5_25_a	1	157	157	Quando è cominciato il lavoro (anno)	
q5_25_b	1	158	158	Quando è cominciato il lavoro (mese)	1=da gennaio a marzo; 2=da aprile a giugno 3=da luglio a settembre; 4=da ottobre a dicembre;
q5_26	1	159	159	Diploma era requisito necessario per lavoro?	
q5_27	1	160	160	Diploma qualsiasi o specifico?	
q5_28	1	161	161	Votazione minima richiesta?	
q5_29	1	162	162	Reputi necessario diploma per lavoro?	
q5_30a	1	163	163	Soddisfazione lavoro: trattamento economico	
q5_30b	1	164	164	Soddisfazione lavoro: possibilità carriera	

Variabili	Lunghezza	Colonna iniziale	Colonna Finale	Descrizione quesiti	
q5_30c	1	165	165	Soddisfazione lavoro: mansioni svolte	
q5_30d	1	166	166	Soddisfazione lavoro: stabilità posto	
q5_30e	1	167	167	Soddisfazione lavoro: grado di autonomia	
q5_30f	1	168	168	Soddisfazione lavoro: conoscenze scuola superiore	
q5_31	2	169	170	Come hai trovato lavoro?	
q5_32	1	171	171	C'è stata persona che ti ha aiutato a trovare lavoro?	
q5_33	1	172	172	In che modo ti ha aiutato?	
q5_34	1	173	173	Chi ti ha aiutato?	
q6_1	1	174	174	Opportunità di lavoro?	
q6_2	1	175	175	Hai accettato?	
q6_3	1	176	176	Motivo rifiuto	
q6_4	1	177	177	Primo lavoro (occasionale/continuativo)	
q6_5	1	178	178	Lavoro a tempo determinato/indeterminato	
q6_6_a	1	179	179	In che anno era iniziato?	
q6_6_b	1	180	180	In che mese?	1=da gennaio a marzo; 2=da aprile a giugno 3=da luglio a settembre; 4=da ottobre a dicembre; 9=non indicato
q6_7	2	181	182	Perché concluso/interrotto?	
q6_8	1	183	183	Cerchi lavoro?	
q6_9	1	184	184	Perché non cerchi lavoro?	
q6_10	1	185	185	Cerchi nuovo lavoro?	
q7_1	1	186	186	Quanti mesi fa hai preso ultima iniziativa di ricerca lavoro?	
q7_2	1	187	187	Preferisci tempo pieno o part-time?	
q7_3	1	188	188	Preferisci dipendente o autonomo?	
q7_4	1	189	189	Disponibilità a lavorare all'estero	
q7_5	1	190	190	Disponibilità a cambiare città	
q7_6	4	191	194	Cifra minima mensile accettabile per lavoro	fino a 1500 € arrotondato ai 25 €; da 1501 a 2000 arrotondato ai 50€; oltre 2000 € arrotondato ai 100 €
q7_7	1	195	195	Lavoreresti subito?	
q8_1a	1	196	196	Aiuto della famiglia nello studio - 1^ indicazione	
q8_1b	1	197	197	Aiuto della famiglia nello studio - 2^ indicazione	
q8_1c	1	198	198	Aiuto della famiglia nello studio - 3^ indicazione	
q8_2	1	199	199	Peso della famiglia nello scegliere scuola superiore	

Variabili	Lunghezza	Colonna iniziale	Colonna Finale	Descrizione quesiti	
q8_3	1	200	200	Titolo studio padre quando intervistato 14-enne	1=licenza elementare/analfabeta; 2=licenza media; 3=diploma di scuola secondaria superiore; 4=laurea/diploma universitario; 5=non indicato
q8_4	1	201	201	Titolo studio madre quando intervistato 14-enne	1=licenza elementare/analfabeta; 2=licenza media; 3=diploma di scuola secondaria superiore; 4=laurea/diploma universitario; 5=non indicato
q8_5	1	202	202	Occupazione padre quando intervistato 14-enne	
q8_6	1	203	203	Posizione nella professione padre	
q8_7	1	204	204	Posizione nella professione padre lav. autonomo	
q8_8	1	205	205	Posizione professione padre lav. dipendente	
q8_9	1	206	206	Codifica Istat professione padre	vedi Allegato PROFESSIONI, appendice B
q8_10	1	207	207	Settore lavoro padre	
q8_11	1	208	208	Occupazione madre quando intervistato 14-enne	
q8_12	1	209	209	Posizione nella professione madre	
q8_13	1	210	210	Posizione professione madre lav. autonoma	
q8_14	1	211	211	Posizione professione madre lav. dipendente	
q8_15	1	212	212	Settore lavoro madre	
q9_1	1	213	213	Cittadinanza italiana?	
resid	1	214	214	Regione di residenza	1=regione diversa rispetto a quella della scuola; 2=stessa regione
ripres	1	215	215	Ripartizione di residenza	vedi Allegato RIPARTIZIONI, appendice B
q9_4	1	216	216	Provincia residenza è la stessa del domicilio?	
domic	1	217	217	Regione del domicilio	1=regione diversa rispetto a quella della scuola; 2=stessa regione
ripdom	1	218	218	Ripartizione del domicilio	vedi Allegato RIPARTIZIONI, appendice B
q9_6	1	219	219	Hai fatto servizio militare/civile?	
q9_7	1	220	220	Anno di nascita	1=prima del 1982; 2=1982; 3=1983 e successivi
q9_8	1	221	221	Stato civile	1=celibe/nubile; 2=altro

Variabili	Lunghezza	Colonna iniziale	Colonna Finale	Descrizione quesiti	
primodip	1	222	222	Flag per i diplomati che hanno conseguito il primo diploma prima del 2001	0=primo diploma nel 2001; 1=primo diploma conseguito prima del 2001
pesofin	12	223	234	Coefficiente di espansione all'universo	

APPENDICE B: ALLEGATI

Allegato TIPO DI SCUOLA

Codice	Tipo di scuola
---------------	-----------------------

01	ISTITUTO PROF INDUSTRIA E ARTIGIANATO
02	ISTITUTO PROF PER I SERVIZI COMMERCIALI E TURISTICI
03	ALTRO ISTITUTO PROFESSIONALE
04	ISTITUTO TECNICO INDUSTRIALE
05	ISTITUTO TECNICO COMMERCIALE
06	ISTITUTO TECNICO PER GEOMETRI
07	ALTRO ISTITUTO TECNICO
08	LICEO SCIENTIFICO
09	LICEO CLASSICO
10	LICEO LINGUISTICO
11	ISTRUZIONE MAGISTRALE
12	ISTRUZIONE ARTISTICA

Allegato REGIONI

Codice	Regione
01	Piemonte
02	Valle d'Aosta
03	Lombardia
04	Trentino-Alto Adige
05	Veneto
06	Friuli-Venezia Giulia
07	Liguria
08	Emilia-Romagna
09	Toscana
10	Umbria
11	Marche
12	Lazio
13	Abruzzo
14	Molise
15	Campania
16	Puglia
17	Basilicata
18	Calabria
19	Sicilia
20	Sardegna

Allegato RIPARTIZIONI

Codice	Ripartizione
1	Nord-Ovest
2	Nord-Est
3	Centro
4	Sud
5	Isole
9	Estero

Allegato GRUPPI

Codice Gruppo di corsi di laurea e di diploma

- | | |
|----|-----------------------------|
| 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 |

Allegato PROFESSIONI

Codice Grandi Gruppi

- | | |
|---|---|
| 1 | Legislatori, dirigenti e imprenditori |
| 2 | Professioni intellettuali, scientifiche e di elevata specializzazione |
| 3 | Professioni tecniche |
| 4 | Impiegati |
| 5 | Professioni qualificate nelle attività commerciali e nei servizi |
| 6 | Artigiani, operai specializzati e agricoltori |
| 7 | Conduttori di impianti e operai semiqualeficati addetti a macchinari fissi e mobili |
| 8 | Professioni non qualificate |
| 9 | Forze Armate |
| 0 | Non indicato |

Appendice C: Definizioni

Ripartizioni geografiche

Costituiscono una suddivisione geografica del territorio e sono così articolate:

Nord

Nord-Ovest comprende: Piemonte, Valle d'Aosta, Lombardia, Liguria

Nord-Est comprende: Trentino-Alto Adige, Veneto, Friuli-Venezia Giulia, Emilia-Romagna

Centro comprende: Toscana, Umbria, Marche, Lazio

Mezzogiorno

Sud comprende: Abruzzo, Molise, Campania, Puglia, Basilicata, Calabria

Isole comprende: Sicilia, Sardegna

Voto di diploma: è espresso in centesimi.

Occupato: chi svolge un'attività lavorativa retribuita, anche se non regolarizzata da contratto; sono invece escluse tutte le attività di qualificazione (borsa di studio, tirocinio, stage eccetera) anche se retribuite.

Persona in cerca di occupazione: diplomato che, oltre a dichiarare di essere in cerca di lavoro, ha intrapreso un'azione di ricerca di lavoro negli ultimi trenta giorni ed è disposto ad iniziare un lavoro entro due settimane.

Tasso di occupazione: il rapporto tra gli occupati e il totale della popolazione.

Tasso di disoccupazione: il rapporto tra le persone che non lavorano e sono in cerca di occupazione e gli attivi (persone che lavorano e persone che non lavorano in cerca di occupazione).

Lavori continuativi: tutti i lavori che vengono svolti con continuità, cioè con una cadenza regolare, con o senza regolare contratto.

Lavori stagionali: sono lavori che vengono svolti in modo continuativo, ma solo in alcuni periodi dell'anno.

Lavori occasionali: sono quelli che vengono svolti di tanto in tanto (per brevi periodi), senza una cadenza regolare o prevedibile.

Lavoratore dipendente: chi presta il suo lavoro alle dipendenze di un datore di lavoro pubblico o privato, è pagato con un salario o stipendio e non potrebbe svolgere il proprio lavoro se altri non gli mettessero a disposizione i mezzi per farlo.

Lavoratore indipendente: chi svolge un lavoro autonomo da solo o con altri soci, essendo proprietario di tutti gli strumenti tecnici e organizzativi necessari.

Lavoratore part-time: lavoratore che, in base a un accordo con il datore di lavoro, presta un orario inferiore a quello normalmente in vigore per gli altri occupati dell'ambiente di lavoro in cui opera. Anche un lavoratore indipendente può lavorare part-time, ad esempio un negoziante che svolge la propria attività nel suo negozio solo la mattina o il pomeriggio.

Guadagno mensile netto: guadagno percepito mensilmente da cui sono stati detratti i contributi, le imposte, eccetera.

APPENDICE D:

La strategia di campionamento

1. Descrizione del disegno di campionamento

Nelle pagine che seguono si illustrano gli obiettivi conoscitivi e gli aspetti più significativi della strategia di campionamento dell'indagine sui percorsi di studio e di lavoro dei diplomati dell'anno 2001.

La *popolazione di interesse* dell'indagine in oggetto, ossia l'insieme delle unità statistiche intorno alle quali si intende investigare, è costituita dagli studenti che hanno conseguito il diploma nelle scuole secondarie superiori nell'anno 2001. Gli obiettivi conoscitivi più rilevanti riguardano le condizioni di lavoro o di studio, mentre i domini di riferimento delle stime sono:

- l'intero territorio nazionale;
- le cinque ripartizioni geografiche (Italia Nord-Occidentale, Italia Nord-Orientale, Italia Centrale, Italia Meridionale, Italia Insulare);
- le 19 tipologie di scuola;
- le quindici modalità ottenute dall'incrocio tra l'aggregazione delle tipologie di scuola in cinque gruppi e la ripartizione geografica a tre modalità (Nord, Centro e Mezzogiorno).

Il disegno di campionamento è a due stadi di selezione con stratificazione delle unità di primo stadio. Le unità di primo stadio sono le scuole secondarie superiori, stratificate per regione geografica, tipo di scuola e dimensione in termini di diplomati. Le unità di secondo stadio sono gli alunni che hanno conseguito il diploma nell'anno 2001.

La numerosità campionaria di primo e di secondo stadio è stata definita tenendo conto sia esigenze organizzative e di costo, sia gli errori di campionamento attesi delle principali stime di interesse a livello dei domini di stima sopra menzionati. La dimensione complessiva del campione di diplomati da intervistare è stata fissata intorno a 20.000.

1.2 Stratificazione e selezione delle scuole

L'obiettivo della stratificazione è quello di formare gruppi (o strati) di unità caratterizzate, relativamente alle variabili oggetto d'indagine, da massima omogeneità interna agli strati e massima eterogeneità fra gli strati. Il raggiungimento di tale obiettivo si traduce in termini statistici in un guadagno nella precisione delle stime, ossia in una riduzione dell'errore campionario a parità di numerosità campionaria.

Le scuole sono state stratificate nei domini ottenuti come incrocio della regione e del tipo di scuola (355 domini) e, all'interno di tali domini, in base alla loro dimensione in termini di diplomati, nel rispetto delle seguenti condizioni:

- autoponderazione del campione a livello dei domini;
- selezione di una scuola campione nell'ambito di ciascuno strato;
- definizione del numero di diplomati da intervistare in ciascuna scuola campione: tale numero è stato posto pari a 12;
- formazione di strati aventi ampiezza approssimativamente costante in termini di diplomati.

Poiché in un notevole numero di scuole il numero atteso di diplomati da intervistare non veniva garantito, prima di procedere alla stratificazione e alla selezione, le scuole dell'universo (6.685) sono state aggregate creando dei *cluster* di scuole che contenessero un numero minimo di diplomati.

Tale aggregazione è stata fatta all'interno dei domini (definiti dall'incrocio di tipo di scuola e regione): dopo aver ordinato in modo casuale le scuole al di sotto di una certa dimensione (fissata pari a 33), sono stati formati i *cluster* di scuole. Ciononostante, in alcuni casi non è stato possibile garantire la dimensione campionaria minima in quanto per alcuni domini era presente solo una scuola piccola e quindi non è stato possibile creare il relativo *cluster*.

Lo schema campionario è stato quindi applicato all'universo di *cluster* di scuole così definito (5.170), fissando il numero minimo di interviste per scuola pari a 12. Il numero di scuole rispondenti è risultato alla fine pari a 1.802 (corrispondente a 1.434 *cluster*).

La dimensione complessiva del campione è stata distribuita tra i domini ottenuti dall'incrocio delle variabili tipo di scuola e regione in modo da garantire che gli errori di campionamento attesi delle principali stime riferite ai diversi domini di interesse non superassero prefissati livelli. A questo scopo è stata utilizzata una metodologia basata su una generalizzazione del metodo di allocazione multivariata di Bethel⁹ al caso di più tipologie di domini di stima¹⁰.

Le variabili di interesse utilizzate per l'allocazione del campione riguardano la condizione di studio o di lavoro dei giovani diplomati e le stime considerate sono le frequenze relative delle principali modalità assunte da tali variabili (lavora, cerca lavoro, studia).

Il procedimento di stratificazione delle scuole¹¹, attuato all'interno di ogni dominio, è stato articolato nelle seguenti fasi:

- ordinamento delle scuole del dominio in ordine decrescente secondo la loro dimensione in termini di diplomati;
- determinazione di una soglia di popolazione per la definizione delle scuole auto rappresentative (AR) e non auto rappresentative (NAR), mediante la relazione:

$${}_d\lambda = \frac{{}_d\bar{m}}{{}_df}$$

in cui ${}_d\bar{m}$ indica il numero di individui da intervistare in ciascuna scuola campione del dominio d e f denota la frazione di campionamento del dominio d ;

- suddivisione di tutte le scuole nei due sottoinsiemi AR e NAR: le scuole di dimensione superiore o uguale a ${}_d\lambda$ sono state definite come scuole AR e le rimanenti come NAR;
- suddivisione delle scuole dell'insieme NAR in strati aventi dimensione, in termini di diplomati, approssimativamente costante e all'incirca pari alla soglia ${}_d\lambda$.

Una volta effettuata la stratificazione, le scuole AR, che fanno strato a sé stante, sono state incluse con certezza nel campione; per quanto riguarda, invece, le scuole NAR, nell'ambito di ogni strato è stata estratta una scuola campione con probabilità proporzionale alla dimensione in termini di diplomati, mediante la procedura di selezione sistematica proposta da Madow (1949, *On the theory of systematic sampling*, II. Ann Math. Stat., 20, 333-354).

A ciascuna scuola estratta nel primo stadio campionario è stato chiesto di inviare un elenco completo degli studenti che hanno conseguito il diploma nel 2001.

A partire da questa lista si è proceduto all'estrazione di un campione di maturi che successivamente è stato contattato mediante intervista telefonica. Tenuto conto che la rilevazione telefonica dà usualmente luogo ad un'alta percentuale di mancate risposte, al fine di garantire la

⁹ Bethel J. (1989), *Sample Allocation in Multivariate Surveys*, Survey Methodology, Vol. 15.

¹⁰ Si veda: Falorsi P.D., Ballin M., De Vitiis C., Scepi G. (1998) "Principi e metodi del software generalizzato per la definizione del disegno di campionamento nelle indagini sulle imprese condotte dall'ISTAT", *Statistica Applicata* Vol. 10, n.2

¹¹ D'ora in avanti con il termine "scuola" si intenderà "*cluster*" di scuole.

numerosità campionaria prefissata di interviste si è deciso di estrarre un campione sovradimensionato rispetto al campione iniziale previsto di 20.000 diplomati. A tal fine sono stati utilizzati, come tassi di sovracampionamento, i tassi di non risposta osservati nell'indagine precedente.

Al fine di illustrare la dimensione campionaria dell'indagine, viene riportata nel Prospetto 1 la distribuzione per regione delle scuole e dei diplomati dell'anno 2001 nell'universo e nel campione dei rispondenti.

REGIONI	Scuole		Diplomati	
	Universo	Campione	Universo	Campione
Piemonte	419	94	26.989	1.074
Valle d'Aosta	10	8	610	104
Lombardia	876	225	58.483	2.486
Trentino-Alto Adige	124	38	6.311	348
Veneto	482	142	31.026	1.725
Friuli-Venezia Giulia	142	36	7.772	368
Liguria	172	51	8.963	472
Emilia Romagna	376	109	25.006	1.258
Toscana	368	125	23.686	1.404
Umbria	111	34	6.756	388
Marche	178	62	11.900	757
Lazio	634	220	46.700	2.716
Abruzzo	157	40	11.970	351
Molise	46	19	3.165	206
Campania	660	135	56.686	1.690
Puglia	495	106	38.512	1.388
Basilicata	108	26	6.525	272
Calabria	311	74	22.112	738
Sicilia	772	190	46.878	1.966
Sardegna	244	68	14.853	697
Italia	6.685	1.802	454.903	20.408

2. Procedimento per il calcolo delle stime

Le stime prodotte dall'indagine sono principalmente stime di frequenze assolute e relative, come, ad esempio, il numero totale (e la frequenza relativa percentuale) dei diplomati che lavorano tre anni dopo il diploma.

Il principio su cui è basato ogni metodo di stima campionaria è che le unità appartenenti al campione rappresentino anche le unità della popolazione che non sono incluse nel campione.

Questo principio viene realizzato attribuendo ad ogni unità campionaria un peso che denota il numero di unità della popolazione rappresentate dalla unità medesima. Se, ad esempio, ad una unità campionaria viene attribuito un peso pari a 30, vuol dire che questa unità rappresenta se stessa ed altre 29 unità della popolazione che non sono state incluse nel campione.

Al fine di rendere più chiara la successiva esposizione, introduciamo la seguente simbologia:

- d indice generico del livello di riferimento delle stime;
- i indice della scuola;
- j indice del diplomato;
- p indice di dominio ottenuto dall'incrocio della regione per il tipo di scuola;

- h indice dello strato di scuole;
- M_h totale dei diplomati dello strato h ;
- M_{hi} totale dei diplomati della scuola i dello strato h ;
- m_{hi} campione dei diplomati della scuola i dello strato h ;
- N_h totale di scuole nello strato h ;
- n_h scuole campione nello strato h (pari sempre a 1);
- H_d numero totale di strati nel dominio d ;
- H_p numero totale di strati nel dominio p ;
- sM_p totale dei diplomati nel dominio p per sesso ($s=1$, maschi; $s=2$, femmine);
- x generica variabile oggetto di indagine;
- X_{hij} valore osservato della variabile x sul j -mo diplomato della scuola i e strato h .

Ipotizziamo di voler stimare con riferimento ad un generico dominio d , il generico totale di popolazione espresso dalla seguente relazione:

$$X_d = \sum_{h=1}^{H_d} \sum_{i=1}^{N_h} \sum_{j=1}^{M_{hi}} X_{hij} \quad (1)$$

La stima del totale (1), si ottiene in generale mediante la seguente formula:

$$\hat{X}_d = \sum_{h=1}^{H_d} \sum_{i=1}^{n_h} \sum_{j=1}^{m_{hi}} X_{hij} W_{hij} \quad (2)$$

dove W_{hij} è il *peso finale* dell'individuo (hij).

I pesi finali da attribuire agli individui campione sono stati calcolati in base ad uno stimatore del rapporto combinato post-stratificato per sesso, che utilizza la conoscenza di totali noti di popolazione, disponibili da fonti esterne all'indagine. Tali totali sono il totale dei maschi e delle femmine a livello dei domini di stratificazione ottenuti dall'incrocio della regione con il tipo scuola, desunti dall'archivio delle scuole secondarie superiori. La post-stratificazione per sesso garantisce che sussista l'uguaglianza tra tali totali noti e le corrispondenti stime campionarie.

La procedura di costruzione dei pesi è stata così articolata:

- 1) si è determinato dapprima un *peso base*, o *peso diretto*, D_{hij} , uguale per tutti gli individui appartenenti alla medesima scuola (o cluster) i dello strato h , ottenuto come l'inverso della probabilità di inclusione nel campione:

$$D_{hij} = \frac{1}{n_h} \frac{M_{hi}}{m_{hi}} ;$$

- 2) si è calcolato il fattore correttivo del peso base che ha la finalità di far coincidere le stime campionarie dei totali di dominio per sesso con i corrispettivi totali noti:

$${}_s C_p = \frac{{}_s M_p}{{}_s \hat{M}_p} = \frac{{}_s M_p}{\sum_{h=1}^{H_p} \sum_{i=1}^{n_h} \sum_{j=1}^{m_{hi}} D_{hij} \cdot {}_s \delta_{hij}} ,$$

dove ${}_s \delta_{hij} = 1$ se l'individuo (hij) presenta il sesso s e ${}_s \delta_{hij} = 0$ altrimenti;

- 3) il peso finale dell'individuo (hij) di sesso s (con h appartenente al dominio p) è stato ottenuto moltiplicando il peso diretto D_{hij} per il correttore ${}_sC_p$:

$$W_{hij} = D_{hij} \cdot {}_sC_p.$$

Una volta assegnato a ogni individuo il coefficiente di riporto all'universo, è stato possibile ottenere le stime di interesse dei parametri di popolazione del tipo (1) come indicato nella (2).

E' utile sottolineare che lo stimatore appena illustrato rientra nella classe degli stimatori di ponderazione vincolata, che è il metodo di stima standard per la maggior parte delle indagini ISTAT sulle imprese e sulle famiglie. Tale classe di stimatori viene utilizzata quando si dispone di informazioni espresse in forma di totali noti di variabili ausiliarie legate alle variabili di interesse.

3. Valutazione del livello di precisione delle stime

3.1 Calcolo della varianza campionaria

Le principali statistiche di interesse per valutare la variabilità campionaria delle stime prodotte dall'indagine, sono l'errore di campionamento assoluto e l'errore di campionamento relativo.

La stima dell'errore di campionamento assoluto di \hat{X}_d è definita dalla seguente espressione:

$$\hat{\sigma}(\hat{X}_d) = \sqrt{\hat{\text{Var}}(\hat{X}_d)} \quad (3)$$

La stima dell'errore di campionamento relativo di \hat{X}_d è definita dall'espressione:

$$\hat{\varepsilon}(\hat{X}_d) = \frac{\hat{\sigma}(\hat{X}_d)}{\hat{X}_d} \quad (4)$$

La stima della varianza di \hat{X}_d , indicata nella (3) come $\hat{\text{Var}}(\hat{X}_d)$, viene calcolata utilizzando il metodo di linearizzazione di Woodruff, che consente di ottenere un'espressione approssimata della varianza campionaria nel caso di stimatori, come quello qui utilizzato, che non sono funzione lineare dei dati campionari. In simboli si ha:

$$\hat{\text{Var}}(\hat{X}_d) \cong \sum_{h=1}^{H_d} \frac{n_h}{n_h - 1} (\hat{Z}_{hi} - \hat{\bar{Z}}_h)^2$$

in cui

$$\hat{Z}_{hi} = \sum_{j=1}^{m_{hij}} \sum_{s=1}^2 \left(X_{hij} - \frac{s \hat{X}_h}{s \hat{M}_h} \right) {}_s\delta_{hij} W_{hij}$$

è l'espressione della variabile linearizzata relativa allo stimatore del rapporto e ${}_s\delta_{hij} = 1$ se il j -mo diplomato è del sesso s e ${}_s\delta_{hij} = 0$ altrimenti.

Gli errori campionari espressi dalla (3) e dalla (4) consentono di valutare il grado di precisione delle stime; inoltre, l'errore assoluto permette di costruire un intervallo di confidenza, nel quale con una certa probabilità si trova il parametro oggetto di stima :

$$\Pr\{\hat{X}_d - k\hat{\sigma}(\hat{X}_d) \leq X_d \leq \hat{X}_d + k\hat{\sigma}(\hat{X}_d)\} = P \quad (5)$$

Nella (5) il valore di k dipende dal valore fissato per la probabilità P ; ad es., per $P=0,95$ si ha $k=1.96$.

3.2 Presentazione sintetica degli errori campionari

Ad ogni stima \hat{X}_d è associato un errore campionario relativo $\hat{\varepsilon}(\hat{X}_d)$ e quindi, per consentire un uso corretto delle stime fornite dall'indagine, sarebbe necessario presentare, per ogni stima pubblicata, anche il corrispondente errore di campionamento relativo.

Tuttavia, non è possibile soddisfare questa esigenza di informazione, sia per motivi di tempi e di costi di elaborazione, sia perché le tavole della pubblicazione risulterebbero eccessivamente appesantite e di non agevole consultazione per l'utente finale. Inoltre, non sarebbero in ogni caso disponibili gli errori delle stime non pubblicate, che l'utente può ricavare in modo autonomo.

Per questi motivi, generalmente, si ricorre ad una presentazione sintetica degli errori relativi, basata sul *metodo dei modelli regressivi*. Tale metodo si basa sulla determinazione di una funzione matematica che mette in relazione ciascuna stima con il proprio errore relativo.

Il modello utilizzato per le stime di frequenze assolute è il seguente:

$$\log \hat{\varepsilon}^2(\hat{X}_d) = a + b \log(\hat{X}_d) \quad (6)$$

dove i parametri a e b vengono stimati mediante il metodo dei minimi quadrati.

Nella presente indagine è stato stimato un modello di tipo (6) per ciascuno dei seguenti domini di interesse:

- D1. tipi di scuola (19 modalità);
- D2. ripartizioni geografiche (5 modalità) e totale Italia;
- D3. tipi di scuola (5 modalità) per ripartizioni geografiche (3 modalità).

Per calcolare gli errori di campionamento è stato utilizzato un software generalizzato, messo a punto dall'Istat, che consente di calcolare gli errori campionari e gli intervalli di confidenza e, inoltre, permette di costruire modelli regressivi del tipo (6) per la presentazione sintetica degli errori di campionamento.

Il prospetto 2 riporta i valori dei coefficienti a e b e dell'indice di determinazione R^2 del modello utilizzato per l'interpolazione degli errori campionari delle stime riferite ai domini D1 e D2. Informazioni analoghe sono riportate nel prospetto 4 per il dominio D3.

Sulla base delle informazioni contenute nei suddetti prospetti è possibile calcolare l'errore relativo di una determinata stima di frequenza assoluta \hat{X}_d^* , riferita ai domini D1 e D2, e al dominio D3, nel modo di seguito descritto.

Dalla (6) mediante semplici passaggi si ricava:

$$\hat{\varepsilon}(\hat{X}_d^*) = \sqrt{\exp(a + b \log(\hat{X}_d^*))} \quad (7)$$

Se, per esempio, la generica stima \hat{X}_d^* si riferisce agli Istituti Professionali, è possibile introdurre nella (7) i valori dei parametri a e b ($a=2,19097$, $b=-0,92652$) riportati nella prima riga del prospetto 2 e ricavare il corrispondente errore relativo.

Una volta calcolato l'errore relativo è possibile costruire l'intervallo di confidenza al 95% come:

$$\{\hat{X}_d^* - 1,96 \cdot \hat{\varepsilon}(\hat{X}_d^*) \cdot \hat{X}_d^*; \hat{X}_d^* + 1,96 \cdot \hat{\varepsilon}(\hat{X}_d^*) \cdot \hat{X}_d^*\}.$$

Allo scopo di facilitare il calcolo degli errori campionari per le stime che si riferiscono alle cinque ripartizioni geografiche e al totale Italia, nel prospetto 3 sono riportati i valori interpolati degli errori di campionamento relativi di alcune stime di frequenze assolute riferite ai diplomati.

Il prospetto 5, invece, contiene, per ciascun di tipo di scuola e ripartizione geografica, gli errori relativi percentuali corrispondenti a valori crescenti di stime di frequenze assolute, calcolati introducendo nella (7) i valori di a e b riportati nel prospetto 4.

Le informazioni contenute nei prospetti 3 e 5 consentono di calcolare l'errore relativo di una generica stima di frequenza assoluta mediante due procedimenti che risultano di facile applicazione, anche se conducono a risultati meno precisi di quelli ottenibili applicando direttamente la formula (7).

Il primo metodo consiste nell'approssimare l'errore relativo della stima di interesse con quello, riportato nei prospetti, corrispondente al livello di stima che più vi si avvicina.

Il secondo metodo, più preciso del primo, si basa sull'uso di una formula di interpolazione lineare per il calcolo degli errori di stime non comprese tra i valori forniti nei prospetti. In tal caso, l'errore campionario della stima \hat{X}_d^* , si ricava mediante l'espressione:

$$\hat{\varepsilon}(\hat{X}_d^*) = \hat{\varepsilon}(\hat{X}_d^{k-1}) + \frac{\hat{\varepsilon}(\hat{X}_d^{k-1}) - \hat{\varepsilon}(\hat{X}_d^k)}{\hat{X}_d^k - \hat{X}_d^{k-1}} (\hat{X}_d^* - \hat{X}_d^{k-1})$$

dove \hat{X}_d^{k-1} e \hat{X}_d^k sono i valori delle stime entro i quali è compresa la stima \hat{X}_d^* , mentre $\hat{\varepsilon}(\hat{X}_d^{k-1})$ e $\hat{\varepsilon}(\hat{X}_d^k)$ sono i corrispondenti errori relativi.

Prospetto 2 - Valori dei coefficienti a e b e dell'indice di determinazione R² (%) del modello per l'interpolazione degli errori campionari delle stime riferite ai diplomati per tipo di scuola e per ripartizione geografica

	a	b	R ²
TIPO DI SCUOLA			
Istituti professionali	2,19097	-0,92652	95,42
<i>di cui:</i>			
agricoltura e ambiente	1,30597	-1,00650	96,37
industria e artigianato	2,81153	-0,99097	96,22
servizi commerciali e turistici	2,88454	-0,99597	95,52
servizi alberghieri e ristorazione	2,12584	-1,02410	96,63
servizi sociali	1,16851	-0,98328	95,71
Istituti tecnici	2,50603	-0,85624	93,80
<i>di cui:</i>			
agrario	1,52757	-0,97335	94,52
industriale	3,10118	-0,92800	93,41
aeronautico/nautico	0,91503	-0,97032	90,82
commerciale	3,21524	-0,91892	95,45
geometri	2,62542	-0,96560	91,79
turismo	0,69163	-0,97689	95,43
attività sociali	1,44420	-0,91725	81,09
Licei	2,12540	-0,76633	84,28
<i>di cui:</i>			
scientifico	3,77104	-0,95852	94,92
classico	3,50989	-0,99624	97,15
linguistico	0,98723	-0,95941	88,89
Istruzione magistrale	2,05152	-0,75019	76,40
<i>di cui:</i>			
istituto magistrale	2,69711	-0,83724	79,89
scuola magistrale	1,11163	-1,08968	85,03
Istruzione artistica	2,31560	-0,98195	86,76
<i>di cui:</i>			
istituto d'arte	2,36839	-0,96683	80,50
liceo artistico	2,60423	-1,08865	86,48
RIPARTIZIONE GEOGRAFICA			
Nord-Ovest	2,43385	-0,86050	90,58
Nord-Est	2,44374	-0,89923	92,22
Centro	2,04648	-0,83899	90,87
Sud	2,06751	-0,73658	80,57
Isole	2,26090	-0,81028	87,82
Italia	2,65077	-0,88487	95,32

Prospetto 3 - Valori interpolati degli errori campionari relativi percentuali delle stime riferite ai diplomati per ripartizione geografica

STIMA	RIPARTIZIONE GEOGRAFICA					Italia
	Nord-Ovest	Nord-Est	Centro	Sud	Isole	
200	34,6	31,3	30,1	39,9	36,2	36,1
500	23,3	20,8	20,5	28,5	25,0	24,1
1.000	17,3	15,2	15,3	22,1	18,9	17,7
2.000	12,8	11,1	11,5	17,1	14,2	13,0
3.000	10,8	9,3	9,7	14,7	12,1	10,9
4.000	9,5	8,1	8,6	13,3	10,8	9,6
5.000	8,7	7,4	7,8	12,2	9,8	8,7
6.000	8,0	6,8	7,2	11,4	9,1	8,0
7.000	7,5	6,3	6,8	10,8	8,6	7,5
8.000	7,1	6,0	6,4	10,3	8,1	7,1
9.000	6,7	5,7	6,1	9,8	7,7	6,7
10.000	6,4	5,4	5,8	9,5	7,4	6,4
15.000	5,4	4,5	4,9	8,1	6,3	5,3
20.000	4,8	4,0	4,4	7,3	5,6	4,7
25.000	4,3	3,6	4,0	6,7	5,1	4,3
30.000	4,0	3,3	3,7	6,3	4,8	3,9
35.000	3,7	3,1	3,5	6,0	4,5	3,7
40.000	3,5	2,9	3,3	5,7	4,2	3,5
45.000	3,4	2,7	3,1	5,4	4,0	3,3
50.000	3,2	2,6	3,0	5,2	3,9	3,1
75.000	2,7		2,5	4,5		2,6
100.000				4,1		2,3
150.000						1,9
200.000						1,7
250.000						1,5
300.000						1,4

Prospetto 4 - Valori dei coefficienti a e b e dell'indice di determinazione R² (%) del modello per l'interpolazione degli errori campionari delle stime riferite ai diplomati per tipo di scuola e ripartizione geografica

DOMINIO DI STIMA TIPO DI SCUOLA PER RIPARTIZIONE GEOGRAFICA		a	b	R ²
Istituti professionali	Nord	2,20859	-0,88830	91,91
	Centro	2,25638	-0,97072	94,96
	Mezzogiorno	1,77259	-0,88167	92,08
	<i>Italia</i>	<i>2,19097</i>	<i>-0,92652</i>	<i>95,42</i>
Istituti tecnici	Nord	2,50977	-0,88435	92,92
	Centro	2,14778	-0,88865	91,80
	Mezzogiorno	2,23654	-0,74915	82,99
	<i>Italia</i>	<i>2,50603</i>	<i>-0,85624</i>	<i>93,80</i>
Licei	Nord	1,92122	-0,75879	84,45
	Centro	1,58980	-0,65482	65,94
	Mezzogiorno	2,83673	-0,82472	80,07
	<i>Italia</i>	<i>2,12540</i>	<i>-0,76633</i>	<i>84,28</i>
Istruzione magistrale	Nord	2,67111	-0,88237	73,00
	Centro	1,73151	-0,85420	83,55
	Mezzogiorno	3,24455	-0,86849	68,48
	<i>Italia</i>	<i>2,05152</i>	<i>-0,75019</i>	<i>76,40</i>
Istruzione artistica	Nord	2,59692	-1,02121	87,30
	Centro	2,17374	-0,96443	93,13
	Mezzogiorno	5,42000	-1,53231	73,12
	<i>Italia</i>	<i>2,31560</i>	<i>-0,98195</i>	<i>86,76</i>

Prospetto 5 - Valori interpolati degli errori campionari relativi percentuali delle stime riferite ai diplomati per tipo di scuola e ripartizione geografica

STIMA	Istituti professionali	Istituti tecnici	Licei	Istruzione magistrale	Istruzione artistica
NORD					
100	39,0	*	*	*	34,9
200	28,7	33,7	35,0	36,7	24,5
400	21,1	24,8	26,9	27,0	17,2
600	17,6	20,7	23,1	22,6	14,0
800	15,5	18,3	20,7	19,9	12,1
1.000	14,0	16,5	19,0	18,0	10,8
2.000	10,3	12,2	14,6	13,3	7,6
3.000	8,6	10,2	12,5	11,1	6,1
4.000	7,6	9,0	11,2	9,8	5,3
5.000	6,9	8,1	10,3	8,9	4,7
6.000	6,3	7,5	9,6	8,2	4,3
7.000	5,9	7,0	9,1	7,6	
8.000	5,6	6,6	8,6	7,2	
9.000	5,3	6,3	8,3	6,8	
10.000	5,0	6,0	7,9	6,5	
15.000	4,2	5,0	6,8		
20.000	3,7	4,4	6,1		
25.000	3,4	4,0	5,6		
30.000		3,7	5,2		
35.000		3,4	4,9		
40.000		3,2	4,7		
45.000		3,1	4,5		
50.000		2,9	4,3		
75.000					

* Errori campionari superiori al 40%

Prospetto 5 segue - Valori interpolati degli errori campionari relativi percentuali delle stime riferite ai diplomati per tipo di scuola e ripartizione geografica

STIMA	Istituti professionali	Istituti tecnici	Licei	Istruzione magistrale	Istruzione artistica
CENTRO					
100	33,1	37,8	*	33,3	32,2
200	23,6	27,8	39,1	24,7	23,0
400	16,9	20,4	31,1	18,4	16,5
600	13,9	17,1	27,3	15,5	13,6
800	12,0	15,0	24,8	13,7	11,8
1.000	10,8	13,6	23,1	12,4	10,6
2.000	7,7	10,0	18,4	9,2	7,6
3.000	6,3	8,3	16,1	7,8	6,2
4.000	5,5	7,3	14,7	6,9	
5.000	5,0	6,7	13,6	6,3	
6.000	4,5	6,1	12,8		
7.000	4,2	5,7	12,2		
8.000	3,9	5,4	11,7		
9.000	3,7	5,1	11,2		
10.000	3,5	4,9	10,9		
15.000		4,1	9,5		
20.000		3,6	8,7		
25.000		3,3	8,0		
30.000		3,0	7,6		
35.000					
40.000					
45.000					
50.000					
75.000					

* Errori campionari superiori al 40%

Prospetto 5 segue - Valori interpolati degli errori campionari relativi percentuali delle stime riferite ai diplomati per tipo di scuola e ripartizione geografica

STIMA	Istituti professionali	Istituti tecnici	Licei	Istruzione magistrale	Istruzione artistica
MEZZOGIORNO					
100	31,9	*	*	*	*
200	23,5	*	*	*	25,9
400	17,3	32,4	34,9	37,6	15,3
600	14,5	27,9	29,5	31,5	11,2
800	12,7	25,0	26,2	27,8	9,0
1.000	11,5	23,0	23,9	25,2	7,6
2.000	8,5	17,7	18,0	18,7	4,4
3.000	7,1	15,2	15,2	15,7	3,3
4.000	6,3	13,7	13,5	13,8	2,6
5.000	5,7	12,6	12,3	12,5	2,2
6.000	5,2	11,8	11,4	11,6	1,9
7.000	4,9	11,1	10,7	10,8	
8.000	4,6	10,6	10,2	10,2	
9.000	4,4	10,1	9,7	9,7	
10.000	4,2	9,7	9,3	9,3	
15.000	3,5	8,3	7,8	7,8	
20.000	3,1	7,5	7,0	6,9	
25.000	2,8	6,9	6,3		
30.000		6,4	5,9		
35.000		6,1	5,5		
40.000		5,8	5,2		
45.000		5,5	5,0		
50.000		5,3	4,8		
75.000		4,6			

* Errori campionari superiori al 40%

Prospetto 5 segue - Valori interpolati degli errori campionari relativi percentuali delle stime riferite ai diplomati per tipo di scuola e ripartizione geografica

STIMA	Istituti professionali	Istituti tecnici	Licei	Istruzione magistrale	Istruzione artistica
ITALIA					
100	35,4	*	*	*	33,2
200	25,7	36,2	38,0	38,2	23,6
400	18,6	26,9	29,1	29,5	16,8
600	15,4	22,6	24,9	25,3	13,8
800	13,5	20,0	22,3	22,7	12,0
1.000	12,2	18,2	20,5	20,9	10,7
2.000	8,8	13,5	15,7	16,1	7,6
3.000	7,3	11,4	13,5	13,8	6,2
4.000	6,4	10,0	12,1	12,4	5,4
5.000	5,8	9,1	11,1	11,4	4,9
6.000	5,3	8,4	10,3	10,7	4,4
7.000	4,9	7,9	9,7	10,1	4,1
8.000	4,7	7,5	9,2	9,6	3,9
9.000	4,4	7,1	8,8	9,2	3,6
10.000	4,2	6,8	8,5	8,8	3,5
15.000	3,5	5,7	7,3	7,6	2,8
20.000	3,0	5,0	6,5	6,8	
25.000	2,7	4,6	6,0	6,2	
30.000	2,5	4,2	5,6	5,8	
35.000	2,3	4,0	5,3	5,5	
40.000	2,2	3,7	5,0	5,2	
45.000	2,1	3,6	4,8		
50.000	2,0	3,4	4,6		
75.000		2,9	3,9		
100.000		2,5	3,5		
150.000		2,1			

* Errori campionari superiori al 40%

Prospetto 6 - Valori dei coefficienti a e b e dell'indice di determinazione R² (%) del modello per l'interpolazione degli errori campionari delle stime riferite ai diplomati per regione

DOMINIO DI STIMA REGIONE	a	b	R ²
Piemonte	3,97515	-1,09132	95,08
Valle d'Aosta	0,86432	-1,43330	50,63
Lombardia	4,24555	-1,10794	97,66
Trentino-Alto Adige	4,31310	-1,25583	93,57
Veneto	4,34763	-1,17731	95,27
Friuli-Venezia Giulia	3,48468	-1,12649	87,31
Liguria	4,53019	-1,30796	91,46
Emilia-Romagna	4,29806	-1,19148	94,92
Toscana	2,99972	-1,02186	93,05
Umbria	3,68598	-1,23390	91,83
Marche	2,25196	-0,92989	88,08
Lazio	3,33326	-1,04388	93,31
Abruzzo	2,24153	-0,90048	63,48
Molise	2,03651	-1,61344	97,90
Campania	3,99817	-0,98364	88,42
Puglia	3,41785	-0,97008	85,34
Basilicata	0,56805	-0,71825	30,74
Calabria	3,04989	-0,98635	82,31
Sicilia	4,44823	-1,12570	96,38
Sardegna	2,02056	-0,84608	78,87

Prospetto 7 - Valori interpolati degli errori campionari relativi percentuali delle stime riferite ai diplomati per regione

STIMA	Piemonte	Valle d'Aosta	Lombardia	Trentino-Alto Adige	Veneto
50	*	9,3	*	*	*
100	*	5,7	*	*	*
200	*	3,5	*	31,0	38,9
500	24,6	1,8	26,7	17,5	22,7
1.000	16,8		18,2	11,3	15,1
2.000	11,5		12,4	7,3	10,0
3.000	9,2		9,9	5,7	7,9
4.000	7,9		8,4	4,7	6,7
5.000	7,0		7,5	4,1	5,8
6.000	6,3		6,7	3,7	5,2
7.000	5,8		6,2		4,8
8.000	5,4		5,8		4,4
9.000	5,1		5,4		4,1
10.000	4,8		5,1		3,9
15.000	3,8		4,1		3,1
20.000	3,3		3,5		2,6
25.000	2,9		3,1		2,3
30.000			2,8		2,0

Prospetto 7 segue - Valori interpolati degli errori campionari relativi percentuali delle stime riferite ai diplomati per regione

STIMA	Friuli-Venezia Giulia	Liguria	Emilia-Romagna	Toscana	Umbria
50	*	*	*	*	*
100	*	*	*	*	36,9
200	28,9	30,1	36,5	29,9	24,0
500	17,2	16,5	21,2	18,7	13,7
1.000	11,7	10,5	14,0	13,1	8,9
2.000	7,9	6,7	9,3	9,2	5,8
3.000	6,3	5,1	7,3	7,5	4,5
4.000	5,3	4,2	6,1	6,5	3,8
5.000	4,7	3,7	5,4	5,8	3,3
6.000	4,3	3,3	4,8	5,3	2,9
7.000	3,9	2,9	4,4	4,9	
8.000		2,7	4,1	4,5	
9.000			3,8	4,3	
10.000			3,6	4,1	
15.000			2,8	3,3	
20.000			2,3	2,8	
25.000			2,1		
30.000					

* Errori campionari superiori al 40%

Prospetto 7 segue - Valori interpolati degli errori campionari relativi percentuali delle stime riferite ai diplomati per regione

STIMA	Marche	Lazio	Abruzzo	Molise	Campania
50	*	*	*	11,8	*
100	36,2	*	38,6	6,7	*
200	26,3	33,3	28,2	3,9	*
500	17,1	20,7	18,7	1,8	34,7
1.000	12,4	14,4	13,7	1,1	24,7
2.000	9,0	10,0	10,0	0,6	17,6
3.000	7,5	8,1	8,3	0,4	14,4
4.000	6,5	7,0	7,3		12,5
5.000	5,9	6,2	6,6		11,2
6.000	5,4	5,6	6,1		10,2
7.000	5,0	5,2	5,7		9,5
8.000	4,7	4,9	5,4		8,9
9.000	4,5	4,6	5,1		8,4
10.000	4,3	4,3	4,9		8,0
15.000		3,5			6,5
20.000		3,0			5,7
25.000		2,7			5,1
30.000		2,4			4,6

Prospetto 7 segue - Valori interpolati degli errori campionari relativi percentuali delle stime riferite ai diplomati per regione

STIMA	Puglia	Basilicata	Calabria	Sicilia	Sardegna
50	*	32,6	*	*	*
100	*	25,4	*	*	39,1
200	*	19,8	33,7	*	29,2
500	27,1	14,3	21,4	28,0	19,8
1.000	19,4	11,1	15,2	18,9	14,8
2.000	13,8	8,7	10,8	12,8	11,0
3.000	11,4	7,5	8,9	10,2	9,3
4.000	9,9	6,8	7,7	8,7	8,2
5.000	8,9	6,2	6,9	7,7	7,5
6.000	8,1	5,8	6,3	6,9	6,9
7.000	7,5		5,8	6,3	6,5
8.000	7,1		5,5	5,9	6,1
9.000	6,7		5,2	5,5	5,8
10.000	6,3		4,9	5,2	5,6
15.000	5,2		4,0	4,1	
20.000	4,5		3,5	3,5	
25.000	4,1			3,1	
30.000	3,7			2,8	

* Errori campionari superiori al 40%

3.3 Esempi di calcolo degli errori campionari

Esempio 1

La stima del numero di diplomati degli Istituti Professionali del Nord che lavorano è pari a 21.879 unità.

Nella colonna relativa agli Istituti Professionali del Nord, del prospetto C.5, si individua il valore corrispondente al livello di stima che più si avvicina a 21.879, ossia 20.000. A questo valore è associato un errore relativo percentuale del 3,7%.

Pertanto, l'errore assoluto della stima sarà uguale a:

$$\sigma(21.879) = 0,037 \times 21.879 = 810$$

e l'intervallo di confidenza avrà come estremi :

$$21.879 - (1,96 \times 810) = 20.291$$

$$21.879 + (1,96 \times 810) = 23.467$$

Esempio 2

Considerando la stima precedente, si possono ottenere valori più precisi dell'errore di campionamento operando mediante interpolazione lineare dei due livelli di stima consecutivi tra i quali è compreso il valore della stessa. Tali livelli sono 20.000 e 25.000 ai quali corrispondono, rispettivamente, i valori percentuali 3,7 e 3,4.

L'errore relativo corrispondente a 21.879 è pari a:

$$\hat{\varepsilon}(21.879) = 3,7 - [(3,7 - 3,4) / (25.000 - 20.000)] \times (21.879 - 20.000) = 3,59.$$

L'errore assoluto sarà il seguente:

$$\sigma(21.879) = 0,0359 \times 21.879 = 785$$

e l'intervallo di confidenza avrà come estremi:

$$21.879 - (1,96 \times 785) = 20.340$$

$$21.879 + (1,96 \times 785) = 23.418$$

Esempio 3

Il calcolo dell'errore può essere effettuato, direttamente, tramite la funzione interpolante:

$$\hat{\varepsilon}(\hat{Y}) = \sqrt{\exp(a + b \log(\hat{Y}))}$$

i cui parametri, riportati nel prospetto C.4 alla riga degli Istituti Professionali del Nord, sono i seguenti:

$$a = 2,20859 \quad b = -0,88830.$$

Per $\hat{Y} = 21.879$ si ha:

$$\hat{\varepsilon}(\hat{Y}) = \sqrt{\exp(2,20859 - 0,88830 \times \log(21.879))} = 0,0356.$$

L'errore relativo percentuale è quindi pari al 3,56% e il calcolo dell'errore assoluto e dell'intervallo di confidenza è del tutto analogo a quello degli esempi 1 e 2.

Appendice E:

Metodologia per la stima del rischio di violazione della riservatezza

Per la protezione della riservatezza dei rispondenti all'indagine sui percorsi post-diploma dei diplomati dell'anno 2001, l'Istat ha adottato una metodologia basata su una definizione di rischio di violazione della riservatezza a livello individuale. Tale approccio prevede la stima del rischio per ogni singolo individuo coinvolto nell'indagine che si basa sulla combinazione di modalità che esso assume nell'insieme delle variabili identificative indirette (ovvero che potenzialmente permettono l'identificazione) e sulle informazioni del disegno campionario.

La definizione di un rischio a livello individuale permette di adottare una tecnica di protezione maggiormente mirata al controllo di quegli individui che presentano un più elevato rischio di identificazione. A tale scopo alla consueta tecnica di "ricodifica globale" è stata affiancata quella di "soppressione locale".

Ricordiamo che la tecnica di ricodifica globale consiste nella ricodifica di una variabile unendo più modalità in una sola mentre, con la soppressione locale, i valori di una variabile vengono sostituiti, in una o più unità, con un valore mancante. Una differenza sostanziale tra dette tecniche di protezione sta nel fatto che mentre la prima (ricodifica globale) viene applicata sull'intero insieme di dati la seconda (soppressione locale) viene applicata ad una parte dei dati.

In particolare, stimato il rischio di identificazione per tutti gli individui coinvolti nell'indagine e fissata una soglia (al di sotto della quale il rischio è considerato accettabile) la tecnica di "soppressione locale" è stata applicata a tutti quei record il cui rischio supera detta soglia.

Tale tecnica è stata applicata utilizzando una procedura di ottimizzazione che permette di minimizzare il numero di soppressioni necessarie affinché il rischio di violazione di ciascun individuo non superi la soglia prefissata. Ciò ha permesso un ulteriore miglioramento della tutela della riservatezza delle unità effettivamente a rischio e il rilascio di un maggiore contenuto informativo sia per le altre unità che per le stesse unità a rischio.

Le tabelle allegate presentano in maniera sintetica l'impatto delle soppressioni sul contenuto informativo del file.

Tabella 1: *Distribuzione degli individui per numero di soppressioni*

Individui che presentano s soppressioni (percentuale sul totale=20.408)	
s=0	s=1
19.698 (96,52%)	710 (3,48%)

Tabella 2: *Distribuzione del numero di soppressione per variabile coinvolta.*

Variabili	Regione scuola	Anno di nascita
Numero di soppressioni (percentuale sul totale=20.408)	207 (1,02%)	503 (2,46%)