

# File Standard

## Percorsi di studio e di lavoro dei diplomati Indagine 2001

*Manuale utente e tracciato record*

Il manuale, frutto della collaborazione fra il Servizio Popolazione, Istruzione e Cultura e il Servizio Metodologie per la produzione statistica, è stato curato da Simona Pace e Andrea Cutillo.

L'Appendice D è stata redatta da Claudia De Vitiis (D.1, D.2 ed D.3.1) e Adriano Pareto (D.3.2)

L'Appendice E è stata redatta da Alessandra Capobianchi

La progettazione dell'indagine e la rilevazione dei dati sono state curate da Paola Ungaro.

## *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> .....	6
2. Le variabili contenute nel file .....	10
 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 e il disegno dell'indagine**

Nel 2001 l'Istituto Nazionale di Statistica ha realizzato la seconda indagine sui percorsi di studio e di lavoro dei diplomati.

L'obiettivo dell'indagine è l'analisi delle condizioni di studio o di lavoro dei giovani a poco più di tre anni dal conseguimento del diploma di scuola secondaria superiore.

In considerazione delle particolarità che in Italia caratterizzano i percorsi all'uscita della scuola secondaria superiore, l'indagine propone, per quei giovani che decidono di affrontare gli studi accademici, un adeguato approfondimento del percorso universitario e offre, per quanti si confrontano con il mercato del lavoro, un'ampia analisi degli esiti occupazionali.

La scelta di condurre le interviste a distanza di tre anni dal conseguimento del diploma permette di indagare sia sul primo inserimento dei giovani nel mondo del lavoro sia sul loro impegno negli studi accademici (numero di esami sostenuti, frequenza delle lezioni, etc.) o sull'eventuale interruzione della frequenza. Il fenomeno dell'abbandono degli studi, infatti, è particolarmente rilevante proprio nei primi anni di università.

Nel tentativo di cogliere le determinanti dei diversi percorsi, l'analisi tiene in considerazione anche il contesto economico e sociale in cui si è formato lo studente consentendo, in questo modo, di valutare sia l'interazione esistente tra estrazione sociale e selezione/espulsione operata dall'università, sia la capacità delle politiche del diritto allo studio di costituire un eventuale correttivo.

L'indagine è campionaria, a due stadi di selezione con stratificazione delle unità primarie: le unità di primo stadio sono le scuole secondarie superiori (stratificate per regione e tipo di scuola), quelle di secondo stadio i diplomati dell'anno 1998 (vedi Appendice D). L'indagine è svolta con cadenza triennale ed è riferita a una singola leva di studenti intervistati a circa tre anni dal conseguimento del titolo. Nel corso del 2001 sono stati quindi intervistati i giovani che nel 1998 hanno conseguito un diploma di scuola secondaria superiore.

La tecnica di rilevazione utilizzata, coerentemente con le altre indagini sulla transizione scuola-lavoro (sull'inserimento professionale dei laureati e dei diplomati universitari), è quella telefonica attraverso sistema CATI (*Computer Assisted Telephone Interview*).

## **1.2 Il processo di produzione dei dati**

L'indagine si è svolta in due diverse fasi: nella prima, sono state contattate per posta le scuole selezionate al fine di ottenere la lista dei diplomati, con i loro recapiti telefonici, su cui condurre l'indagine. Nella seconda fase i diplomati selezionati sono stati contattati telefonicamente da una ditta specializzata che ha raccolto i dati mediante intervista telefonica assistita da computer (CATI).

Una volta raccolti i dati, questi sono stati sottoposti a controllo e successiva correzione. Le incoerenze e le incompatibilità sono state eliminate mediante specifici *software* basati su metodologie deterministiche e probabilistiche. In particolare i metodi probabilistici utilizzati si basano sulla ricerca di un diplomato-donatore che possa donare l'informazione e che abbia caratteristiche simili a quelle del diplomato che presenta il dato mancante o errato.

Si è proceduto, quindi, alla costruzione delle stime utilizzando degli stimatori post-stratificati per sesso, regione e tipo di scuola, secondo la metodologia di stima standard utilizzata dall'Istat (vedi appendice D).

I dati sono stati elaborati e sottoposti a validazione mediante l'utilizzo di serie storiche e il confronto con altre indagini, per esempio l'Indagine sulle forze di lavoro o altre indagini sulla transizione scuola-lavoro.

Il processo di produzione si conclude con la diffusione dei dati. I canali di diffusione utilizzati sono molteplici: il tradizionale volume, il sito internet dell'Istat, il file standard ma anche il Laboratorio di Analisi dei Dati Elementari - ADELE -.

## **1.3 I contenuti informativi**

Il questionario è costituito da tre 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.

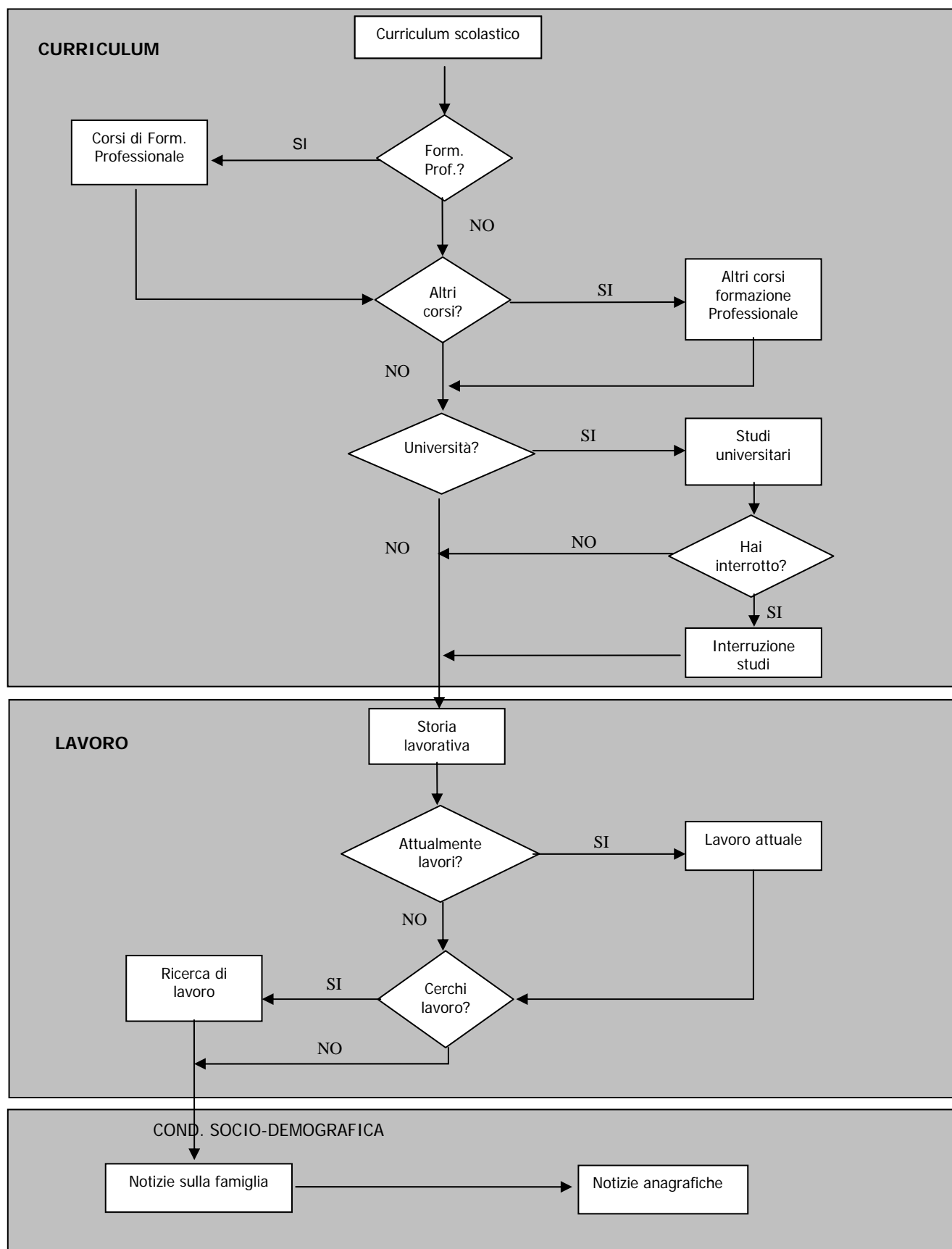
La prima sezione del questionario è dedicata al curriculum degli studi e alle attività di qualificazione dopo il diploma, la seconda al lavoro e alla ricerca del lavoro, mentre nella terza si chiedono notizie relative alla famiglia d'origine e informazioni anagrafiche.

La sezione relativa agli studi riguarda tanto il percorso scolastico già concluso, quanto eventuali studi post-secondari; quella relativa al lavoro, anch'essa articolata in due aree, riguarda le esperienze lavorative e la ricerca di lavoro.

La sezione relativa al curriculum scolastico e alla famiglia di origine, situate rispettivamente all'inizio e alla fine del questionario, sono rivolte a tutti i soggetti intervistati, mentre quella relativa al lavoro e agli studi post-secondari sono riferite a quanti, al momento dell'intervista, risultano "attivi" o studenti.

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 – GLI STUDI**

### **IL CURRICULUM SCOLASTICO**

L'obiettivo della sezione è di verificare quale sia il rapporto tra il percorso scolastico e le scelte successive al diploma, siano esse di studio o di lavoro, nonché la capacità predittiva della valutazione scolastica rispetto al "successo" futuro dei giovani.

I quesiti riguardano il tipo di scuola frequentata, il voto di diploma, i passaggi da scuola pubblica a privata, la frequenza di scuole per il recupero di anni scolastici, le classi eventualmente ripetute, etc.

### **CORSI DI FORMAZIONE POST-SECONDARIA**

Si chiedono notizie riguardo alla frequenza di eventuali corsi di formazione professionale sia privati che regionali.

### **GLI STUDI UNIVERSITARI**

La sezione intende analizzare le modalità di frequenza dei corsi universitari, siano essi di diploma o di laurea. Per questo si chiedono una serie di informazioni tra cui l'ateneo e il corso frequentato, eventuali cambi di corso o di sede, il numero di esami superati, la regolarità nella frequenza delle lezioni.

A tutti quelli che hanno frequentato le lezioni di almeno un insegnamento fondamentale si chiede inoltre una valutazione del docente titolare del corso.

### **L'INTERRUZIONE DEGLI STUDI UNIVERSITARI**

Questa sezione punta a fornire una misura dell'impegno degli studenti che poi abbandonano gli studi (numero di esami superati, frequenza delle lezioni) e ottenere anche da questi, così come per i frequentanti, una valutazione della didattica per verificare se e quanto questa abbia inciso sulla decisione di interrompere gli studi.

I dati sulla condizione sociale di ciascuno, raccolti alla fine del questionario, consentono inoltre di tracciare il profilo dello studente "scoraggiato" e di confrontarlo con quello di quanti proseguono gli studi o di quanti, al contrario, non li hanno mai intrapresi.

## **SEZIONE 2 – IL LAVORO**

### **IL LAVORO ATTUALE**

La sezione analizza il primo inserimento lavorativo dei diplomati, in particolare per quanti svolgono un lavoro continuativo. I quesiti mirano a valutare sia il tipo di lavoro svolto (ad esempio, posizione nella professione, settore di attività economica, numero di ore lavorate, specifica mansione), sia la sua qualità (grado di regolarizzazione del rapporto di lavoro, orario, retribuzione, necessità o meno del diploma).

## LA RICERCA DI LAVORO

La sezione indaga, da una parte, sull'intensità con cui viene condotta la ricerca di lavoro (tempo intercorso dall'ultima iniziativa, disponibilità ad iniziare il lavoro entro le due settimane successive all'intervista); dall'altra, sulle attese retributive, professionali e di carriera, con cui i giovani si presentano sul mercato del lavoro a conclusione di un ciclo di scuola secondaria superiore. Si chiedono quindi le aspettative riguardo al tipo di orario preferito, alla stabilità dell'impiego, al trattamento economico desiderato, alla disponibilità o meno a lavorare fuori dal proprio comune.

## SEZIONE 3 – LA FAMIGLIA D'ORIGINE

In questa parte del questionario si raccolgono informazioni sul titolo di studio e la condizione occupazionale dei genitori. L'analisi dell'inserimento professionale dei diplomati viene svolta, infatti, tenendo conto anche del contesto sociale ed economico in cui si è formato lo studente. In questo modo è possibile analizzare se e fino a che punto l'estrazione sociale condizioni la decisione di proseguire gli studi o l'inserimento professionale.

### 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 1998, 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 E) è la presenza di alcuni *blank* nelle variabili scuola, sesso, regione di residenza ed età. Inoltre per alcune variabili sono state costruite delle aggregazioni.

Nel tracciato record (Appendice A) vengono riportate la posizione, l'ampiezza del campo in *bytes*, il nome e 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 COEF 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 COEF è pari a 5, significa che le informazioni contenute in quel record sono riferite a 5 diplomati dell'anno 1998.

## APPENDICE A: TRACCIATO RECORD

VARIABILE	LUNGHEZZA	COLONNA INIZIALE	COLONNA FINALE	DESCRIZIONE	
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
regsc	2	4	5	regione della scuola	vedi allegato REGIONI, appendice B
rip5sc	1	6	6	ripartizione della scuola	vedi allegato RIPARTIZIONI, appendice B
votomed	1	7	7	voto delle medie	1=sufficiente; 2=buono; 3=distinto; 4=ottimo; 9=non indicato.
votomat	1	8	8	voto di maturità	1=36-41; 2=42-47; 3=48-53; 4=54-60;
q1	1	9	9	Scuola diploma è la stessa di iscrizione dopo medie?	
q2	2	10	11	Tipo scuola di iscrizione dopo medie	vedi allegato TIPO DI SCUOLA, appendice B
q3	1	12	12	Scuola di iscrizione dopo medie pubblica o privata	1=pubblica; 2=privata
q4	1	13	13	Frequentato scuole per recupero anni?	
q5	1	14	14	Sei stato respinto?	
q6_1	1	15	15	Respinto in prima	
q6_2	1	16	16	Respinto in seconda	
q6_3	1	17	17	Respinto in terza	
q6_4	1	18	18	Respinto in quarta	
q6_5	1	19	19	Respinto in quinta	
q7	1	20	20	Frequentato anno integrativo?	
q8	1	21	21	Maturità come privatista?	
q9_1	1	22	22	Soddisfazione lingue straniere	
q9_2	1	23	23	Soddisfazione conoscenze informatiche	
q10	1	24	24	Iscrizione a corso della Regione	
q11	1	25	25	Frequenza del corso	
q12	1	26	26	Durata del corso	
q13	1	27	27	Lavoravi quando ti sei iscritto?	

VARIABILE	LUNGHEZZA	COLONNA INIZIALE	COLONNA FINALE	DESCRIZIONE	
q14	1	28	28	Meglio il corso o un lavoro?	
q15	1	29	29	Corso post-maturità organizzato da scuola statale?	
q16	1	30	30	Corso di formazione privata a pagamento?	
q17	2	31	32	Su che argomento?	
q18	1	33	33	Conseguito diploma universitario?	
q19	1	34	34	Iscritto a corso universitario?	
q20	1	35	35	Ti sei informato sui corsi?	
q21	2	36	37	Modi di informazione sui corsi	
q22	2	38	39	Regione dell'Ateneo universitario	Vedi allegato REGIONI, appendice B
q24	2	40	41	Gruppo corso di laurea/diploma/SDFS	Vedi allegato GRUPPI, appendice B
q25	1	42	42	Anno di immatricolazione all'università	
q26	1	43	43	Hai sostenuto esami?	
q27	2	44	45	Numero esami sostenuti	
q28	1	46	46	Frequenza alle lezioni	
q29	1	47	47	Frequenza a lezioni esame fondamentale	
q30_1	1	48	48	Soddisfazione su docente: competenza	
q30_2	1	49	49	Soddisfazione su docente: chiarezza	
q30_3	1	50	50	Soddisfazione su docente: presenza lezioni	
q30_4	1	51	51	Soddisfazione su docente: presenza ricevimento	
q30_5	1	52	52	Soddisfazione su docente: disponibilità	
q31	1	53	53	Cambio di corso	
q32	1	54	54	Lavoravi quando ti sei iscritto a università?	
q33	1	55	55	Meglio università o lavoro?	
q34_1	1	56	56	Interruzione corso di laurea	
q34_2	1	57	57	Interruzione diploma universitario	
q34_3	1	58	58	Interruzione SDFS	
q34_4	1	59	59	Interruzione: nessun corso	
q35	2	60	61	Motivo interruzione	
q36	1	62	62	Anno immatricolazione università	
q37	1	63	63	Avevi frequentato lezioni di fondamentale?	

VARIABILE	LUNGHEZZA	COLONNA INIZIALE	COLONNA FINALE	DESCRIZIONE	
q38_1	1	64	64	Soddisfazione su docente: competenza	
q38_2	1	65	65	Soddisfazione su docente: chiarezza	
q38_3	1	66	66	Soddisfazione su docente: presenza lezioni	
q38_4	1	67	67	Soddisfazione su docente: presenza ricevimento	
q38_5	1	68	68	Soddisfazione su docente: disponibilità	
q39	1	69	69	Opportunità di lavoro	
q40	1	70	70	Hai accettato?	
q41	2	71	72	Motivi rifiuto	
q42	1	73	73	Hai iniziato un lavoro poi interrotto?	
q43	1	74	74	Che tipo di lavoro?	
q44	1	75	75	Era autonomo?	
q45_1	1	76	76	In che anno era iniziato?	
q45_2	2	77	78	In che mese?	
q46	2	79	80	Motivo di interruzione	
q47_1	1	81	81	Attività post diploma: tirocinio	
q47_2	1	82	82	Attività post diploma: stage	
q47_3	1	83	83	Attività post diploma: praticantato	
q47_4	1	84	84	Attività post diploma: borsa di studio	
q47_5	1	85	85	Attività post diploma: PIP	
q48	1	86	86	Attualmente lavori? (oltre attività formativa)	
q49	1	87	87	Attualmente lavori?	
q50	1	88	88	Hai svolto ore di lavoro?	
q51	1	89	89	Quante?	1=fino a 20ore; 2=21-35 ore; 3=36-40 ore; 4=oltre 40 ore.
q52	1	90	90	Cerchi lavoro?	
q53	2	91	92	Perché non cerchi?	
q54	1	93	93	Tipo di lavoro (saltuario/continuativo)	
q55	1	94	94	Lavoro cominciato prima o dopo il diploma?	
q56_1	1	95	95	Quando è cominciato il lavoro (anno)	
q56_2	2	96	97	Quando è cominciato il lavoro (mese)	

VARIABILE	LUNGHEZZA	COLONNA INIZIALE	COLONNA FINALE	DESCRIZIONE	
q57	1	98	98	Necessario diploma per lavoro?	
q58	1	99	99	Diploma qualsiasi o specifico?	
q59_1	1	100	100	Maturità richiesta: professionale	
q59_2	1	101	101	Maturità richiesta: tecnica	
q59_3	1	102	102	Maturità richiesta: liceale	
q59_4	1	103	103	Maturità richiesta: magistrale	
q59_5	1	104	104	Maturità richiesta: arte applicata?	
q60	1	105	105	Votazione minima richiesta?	
q61	1	106	106	Hai dovuto trasferirti?	
q62	1	107	107	Tempo pieno o part time?	
q63	1	108	108	Motivo part-time	
q64	1	109	109	Ore di lavoro settimanali	1=fino a 20 ore; 2=21-35 ore; 3=36-40 ore; 4=oltre 40 ore.
q65	1	110	110	Guadagno netto	1=fino a 600,00 euro; 2=da 600,01 a 800,00 euro; 3=da 800,01 a 1000,00 euro; 4=oltre 1000,00 euro; 9=non indica;
q66	2	111	112	Professione	Vedi ISTAT, Classificazione delle professioni, metodi e norme - serie C - n.12, edizione 1991
q67	1	113	113	Settore lavorativo	
q68	2	114	115	Settore industriale	
q69	2	116	117	Settore servizi	
q70	1	118	118	Numero dipendenti dell'impresa	
q71	1	119	119	Posizione professione	
q72	2	120	121	Posizione professione indipendenti	
q73	1	122	122	Partita IVA	
q74	1	123	123	Vieni pagato con ritenuta d'acconto?	
q75	2	124	125	Tipo di contratto	
q76	2	126	127	Posizione professione dipendenti	
q77	1	128	128	Datore di lavoro versa contributi?	
q78	1	129	129	Lavori nel pubblico o nel privato?	
q79	1	130	130	Partecipato a corsi formazione organizzati dal datore di lavoro?	
q80	1	131	131	Quanti giorni sono durati?	
q81	2	132	133	Argomento del corso più lungo	

<b>VARIABILE</b>	<b>LUNGHEZZA</b>	<b>COLONNA INIZIALE</b>	<b>COLONNA FINALE</b>	<b>DESCRIZIONE</b>	
q82	1	134	134	Quanto spesso usi il computer	
q83_1	2	135	136	Per cosa usi il computer: scrivere	
q83_2	2	137	138	Per cosa usi il computer: disegnare	
q83_3	2	139	140	Per cosa usi il computer: internet	
q83_4	2	141	142	Per cosa usi il computer: consultare database	
q83_5	2	143	144	Per cosa usi il computer: trattamento suono	
q83_6	2	145	146	Per cosa usi il computer: inserimento dati	
q83_7	2	147	148	Per cosa usi il computer: posta elettronica	
q83_8	2	149	150	Per cosa usi il computer: prodotti multimediali	
q83_9	2	151	152	Per cosa usi il computer: foglio elettronico	
q83_10	2	153	154	Per cosa usi il computer: programmare	
q83_11	2	155	156	Per cosa usi il computer: altro	
q84_1	1	157	157	Quanto leggi o scrivi una lingua straniera?	
q84_2	1	158	158	Quanto parli una lingua straniera?	
q85_1	1	159	159	Soddisfazione lavoro: trattamento economico	
q85_2	1	160	160	Soddisfazione lavoro: possibilità carriera	
q85_3	1	161	161	Soddisfazione lavoro: mansioni svolte	
q85_4	1	162	162	Soddisfazione lavoro: stabilità posto	
q85_5	1	163	163	Soddisfazione lavoro: autonomia	
q86	1	164	164	Giudizio sul proprio diploma rispetto al lavoro	
q87_1	1	165	165	Preparazione tecnico-professionale	
q87_2	1	166	166	Cultura generale fornita da scuola	
q88	2	167	168	Argomento di cui necessiti più conoscenza	
q89	2	169	170	Come hai trovato lavoro?	

VARIABILE	LUNGHEZZA	COLONNA INIZIALE	COLONNA FINALE	DESCRIZIONE	
q90	1	171	171	C'è stata persona che ti ha aiutato a cercare lavoro?	
q91	1	172	172	Chi ti ha aiutato?	
q92	1	173	173	In che modo ti ha aiutato?	
q93	1	174	174	Cerchi un nuovo lavoro?	
q94	1	175	175	Quanti mesi fa hai preso ultima iniziativa di ricerca lavoro?	
q95	1	176	176	Preferisci saltuario o continuativo?	
q96	1	177	177	Preferisci tempo pieno o ridotto?	
q97	1	178	178	Preferisci dipendente o autonomo?	
q98	1	179	179	Disponibilità a lavorare all'estero	
q99	1	180	180	Disponibilità a cambiare città	
q100	1	181	181	Cifra minima accettabile per lavoro	1=fino a 600,00 euro; 2=da 600,01 a 800,00 euro; 3=da 800,01 a 1000,00 euro; 4=oltre 1000,00 euro; 9=non indica.
q101	1	182	182	Lavoreresti subito?	
q102	1	183	183	Hai fratelli o sorelle?	
q103	1	184	184	Quanti?	
q104	2	185	186	Titolo studio padre a 14 anni intervistato	
q105	2	187	188	Titolo studio madre a 14 anni intervistato	
q106	1	189	189	Un nonno aveva istruzione secondaria?	
q107	1	190	190	Occupazione padre a 14 anni intervistato	
q108	2	191	192	Professione padre	Vedi ISTAT, Classificazione delle professioni, metodi e norme - serie C - n.12, edizione 1991
q109	1	193	193	Posizione nella professione padre	
q110	2	194	195	Posizione nella professione padre indipendente	
q111	2	196	197	Posizione nella professione padre dipendente	
q112	1	198	198	Settore lavoro padre	



VARIABILE	LUNGHEZZA	COLONNA INIZIALE	COLONNA FINALE	DESCRIZIONE	
q113	1	199	199	Occupazione madre a 14 anni intervistato	
q114	1	200	200	Posizione nella professione madre	
q115	2	201	202	Posizione nella professione madre indipendente	
q116	2	203	204	Posizione nella professione madre dipendente	
q117	1	205	205	Settore lavoro madre	
q118	1	206	206	Cittadinanza italiana?	
regres	2	207	208	Provincia di residenza	vedi allegato REGIONI, appendice B
rip5res	1	209	209	Ripartizione di residenza	vedi allegato RIPARTIZIONI, appendice B
q121	1	210	210	Provincia residenza è la stessa del domicilio?	
regdom	2	211	212	Regione di domicilio	vedi allegato REGIONI, appendice B
rip5dom	1	213	213	Ripartizione di domicilio	vedi allegato RIPARTIZIONI, appendice B
q123	1	214	214	Sesso	
q124	1	215	215	Hai fatto il militare?	
eta	1	216	216	età	1=fino a 21 anni; 2=22 anni; 3=23 anni; 4=24 anni e oltre
q126	1	217	217	Stato civile	1=celibe; 2=altro
coef	9	218	226	Coefficiente di espansione all'universo	Nove campi di cui cinque destinati ai decimali

## APPENDICE B: ALLEGATI

### Allegato TIPO DI SCUOLA

<b>Codice</b>	<b>Tipo di scuola</b>
11	Istituto professionale agrario
12	Istituto professionale industriale
13	Istituto professionale marinaro
14	Istituto professionale per i servizi commerciali, turistici e della pubblicità
15	Istituto professionale per i servizi alberghieri e della ristorazione
16	Istituto professionale per i servizi sociali
17	Altro istituto professionale
20	Istituto tecnico agrario
21	Istituto tecnico industriale
22	Istituto tecnico nautico
23	Istituto tecnico aeronautico
24	Istituto tecnico commerciale
25	Istituto tecnico per geometri
26	Istituto tecnico per il turismo
27	Istituto tecnico per periti aziendali
28	Istituto tecnico femminile
29	Altro istituto tecnico
31	Scuola magistrale
32	Istituto magistrale
41	Liceo scientifico
42	Liceo ginnasio
43	Liceo linguistico
51	Istituto d'arte
52	Liceo artistico

## **Allegato REGIONI**

<b>Codice</b>	<b>Regione</b>
00	Estero
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**

<b>Codice</b>	<b>Ripartizione</b>
0	Estero
1	Nord-Ovest
2	Nord-Est
3	Centro
4	Sud
5	Isole

## **Allegato GRUPPI**

<b>Codice</b>	<b>Gruppo di corsi di laurea e di diploma</b>
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
99	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 sessantesimi, anche quando la scala prevista è diversa.

**Tasso di occupazione:** il rapporto tra le persone che lavorano e il totale dei diplomati.

**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 e cercano lavoro).

**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:** i diplomati che dichiarano di essere in cerca di lavoro.

**Lavori continuativi:** sono 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.

**Lavoro part-time:** si ha il part-time quando il lavoratore, in base a un accordo con il datore di lavoro, è tenuto a prestare 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.

**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.

## APPENDICE D:

### La strategia di campionamento e la valutazione del livello di precisione delle stime

#### D.1. La strategia di campionamento

##### D.1.1. 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 1998.

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 1998. 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.

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 la maturità nell'anno 1998.

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.

##### D.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.

Nell'indagine in esame, 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<sup>1</sup>, 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 18 per gli istituti professionali e tecnici e pari a 12 per gli altri tipi di scuole;
- formazione di strati aventi ampiezza approssimativamente costante in termini di diplomati.

Il procedimento di stratificazione, attuato all'interno di ogni dominio, si articola 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 incluse con certezza nel campione; per quanto riguarda, invece, le scuole NAR, nell'ambito di ogni strato viene 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).

Ciascuna scuola estratta al primo stadio è stata contattata e ha proceduto alla selezione di un numero approssimativamente costante di diplomati campione, mediante scelta sistematica dalla lista dei diplomati dell'anno 1998.

Il campione di diplomati così determinato è stato intervistato mediante intervista telefonica. Tenendo conto che la rilevazione telefonica può dar luogo a un'alta percentuale di mancate risposte, si è

---

<sup>1</sup> In realtà, non essendo ancora disponibili da fonti ufficiali, i diplomati per tipo di scuola sono stati stimati a partire dagli iscritti all'ultimo anno tenendo conto dei tassi di promozione.

stabilito di estrarre un campione base e un campione suppletivo. Si è definito un campione suppletivo di dimensione pari approssimativamente a quella del campione base.

Al fine di illustrare la dimensione campionaria dell'indagine, viene riportata nel Prospetto D.1 la distribuzione per regione delle scuole e dei diplomati dell'anno 1998 nell'universo e nel campione dei rispondenti. A causa del procedimento di stratificazione utilizzato (vedi nota 1 nella presente appendice) il totale dei diplomati può differire da quello indicato su altre pubblicazioni dell'Istat.

**Prospetto D.1 – Distribuzione delle scuole e dei diplomati, nell'universo e nel campione, per regione - Anno 1998**

REGIONI	Scuole		Diplomati	
	Universo	Campione	Universo	Campione
Piemonte	492	90	27.353	1.247
Valle d'Aosta	19	17	741	280
Lombardia	967	198	61.878	2.742
Trentino-Alto Adige	129	35	6.097	275
Veneto	522	110	32.156	1.536
Friuli-Venezia Giulia	158	39	8.072	494
Liguria	211	57	9.983	662
Emilia-Romagna	410	92	27.039	1.294
Toscana	411	90	26.236	1.400
Umbria	115	45	7.138	602
Marche	196	53	13.145	783
Lazio	695	135	48.184	1.966
Abruzzo	162	70	12.080	930
Molise	53	44	2.991	802
Campania	702	129	59.891	1.672
Puglia	521	108	42.042	1.219
Basilicata	123	61	6.614	675
Calabria	334	92	22.936	967
Sicilia	809	198	49.005	2.494
Sardegna	257	89	15.323	1.222
<b>Italia</b>	<b>7.286</b>	<b>1.752</b>	<b>478.904</b>	<b>23.262</b>

## D.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, a una unità



campionaria viene attribuito un peso pari a 30, vuol dire che questa unità rappresenta se stessa e 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 maturo;
- $p$  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 maturo della scuola  $i$  e strato  $h$ .

Ipotizziamo di voler stimare con riferimento a 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 a 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 così articolata:

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

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

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:

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

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

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  ${}_s C_p$ :

$$W_{hij} = D_{hij} \cdot {}_s C_p .$$

Una volta assegnato a ogni individuo il coefficiente di riporto all'universo, è 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.

### D.3. Valutazione del livello di precisione delle stime

#### D.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:

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

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-maturo 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  $k=1.96$ .

#### *D.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 a 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;
- D2. ripartizioni geografiche (5 modalità) e totale Italia;
- D3. tipi di scuola 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 ed, inoltre, permette di costruire modelli regressivi del tipo (6) per la presentazione sintetica degli errori di campionamento.

Il prospetto D.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 D.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 ( $a=2,20285$ ,  $b=-0,97756$ ) riportati nella prima riga del prospetto D.2 e ricavare il corrispondente errore relativo.

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

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

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 D.3 sono riportati i valori interpolati degli errori di campionamento relativi di alcune stime di frequenze assolute riferite ai diplomati.

Il prospetto D.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 D.4.

Le informazioni contenute nei prospetti D.3 e D.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 D.2 - Valori dei coefficienti  $a$  e  $b$  e dell'indice di determinazione  $R^2$  (%) del modello per l'interpolazione degli errori campionari delle stime riferite ai diplomati per tipo di scuola e per ripartizione geografica**

DOMINIO DI STIMA	a	b	R <sup>2</sup>
<b>TIPI DI SCUOLA</b>			
<b>Istituti Professionali</b>	<b>2,20285</b>	<b>-0,97756</b>	<b>97,01</b>
di cui: Industriali	2,27754	-0,97785	95,55
Serv comm., turist., pubbl.	2,28654	-0,94960	95,04
<b>Istituti Tecnici</b>	<b>3,19313</b>	<b>-0,88629</b>	<b>91,27</b>
di cui: Industriali	2,96829	-0,87050	88,15
Commerciali	3,91530	-0,92426	91,15
Per geometri	2,50021	-0,91078	86,94
<b>Licei</b>	<b>3,52898</b>	<b>-0,94599</b>	<b>94,00</b>
di cui: Scientifici	4,27030	-1,01154	96,16
Ginnasi	3,57028	-1,00332	95,67
<b>Istituti Magistrali</b>	<b>3,96235</b>	<b>-0,95958</b>	<b>88,46</b>
<b>Altri tipi di scuole</b>	<b>2,51164</b>	<b>-0,98252</b>	<b>90,77</b>
<b>RIPARTIZIONI</b>			
Nord-Ovest	2,76653	-0,86196	88,49
Nord-Est	2,61801	-0,85405	86,73
Centro	2,46062	-0,81745	84,43
Sud	3,20077	-0,90742	90,44
Isole	2,29310	-0,81390	86,10
<b>Italia</b>	<b>3,84437</b>	<b>-0,98536</b>	<b>95,64</b>

**Prospetto D.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	
1.000	20,3	19,4	20,3	21,6	18,9	22,7
2.000	15,1	14,4	15,3	15,8	14,3	16,2
3.000	12,7	12,1	13,0	13,1	12,1	13,2
4.000	11,2	10,7	11,5	11,5	10,8	11,5
5.000	10,2	9,7	10,5	10,4	9,8	10,3
6.000	9,4	9,0	9,8	9,6	9,1	9,4
7.000	8,8	8,4	9,2	8,9	8,6	8,7
8.000	8,3	8,0	8,7	8,4	8,1	8,2
9.000	7,9	7,6	8,3	8,0	7,7	7,7
10.000	7,5	7,3	7,9	7,6	7,4	7,3
15.000	6,3	6,1	6,7	6,3	6,3	6,0
20.000	5,6	5,4	6,0	5,5	5,6	5,2
25.000	5,1	4,9	5,5	5,0	5,1	4,7
30.000	4,7	4,5	5,1	4,6	4,7	4,3
35.000	4,4	4,2	4,8	4,3	4,5	3,9
40.000	4,1	4,0	4,5	4,0	4,2	3,7
45.000	3,9	3,8	4,3	3,8	4,0	3,5
50.000	3,8	3,6	4,1	3,7	3,9	3,3
75.000	3,2		3,5	3,0		2,7
100.000				2,7		2,4
150.000						1,9
200.000						1,7
250.000						1,5
300.000						1,4

**Prospetto D.4 - Valori dei coefficienti  $a$  e  $b$  e dell'indice di determinazione  $R^2$  (%) 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		a	b	$R^2$
Istituti Professionali	Nord	2,14182	-0,97596	95,36
	Centro	1,48678	-0,87743	89,51
	Mezzogiorno	1,87706	-0,92093	93,88
	<i>Italia</i>	<i>2,20285</i>	<i>-0,97756</i>	<i>97,01</i>
Istituti Tecnici	Nord	2,51697	-0,79706	85,07
	Centro	1,90832	-0,69753	70,10
	Mezzogiorno	2,48893	-0,80584	85,87
	<i>Italia</i>	<i>3,19313</i>	<i>-0,88629</i>	<i>91,27</i>
Licei	Nord	2,81625	-0,85094	88,10
	Centro	2,82322	-0,84356	87,78
	Mezzogiorno	3,36241	-0,92420	90,69
	<i>Italia</i>	<i>3,52898</i>	<i>-0,94599</i>	<i>94,00</i>
Istituti Magistrali	Nord	3,79063	-0,99767	70,78
	Centro	3,61662	-0,93280	71,33
	Mezzogiorno	3,83374	-0,93134	85,63
	<i>Italia</i>	<i>3,96235</i>	<i>-0,95958</i>	<i>88,46</i>
Altri tipi di scuole	Nord	2,71415	-1,02049	88,55
	Centro	2,56677	-0,98371	80,38
	Mezzogiorno	2,16293	-0,95826	83,51
	<i>Italia</i>	<i>2,51164</i>	<i>-0,98252</i>	<i>90,77</i>

**Prospetto D.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	Istituti Magistrali	Altri tipi di scuole
NORD					
200	22,0	42,6	42,9	47,3	26,0
400	15,7	32,3	31,9	33,5	18,3
600	12,9	27,5	26,9	27,4	14,9
800	11,2	24,5	23,8	23,7	12,8
1.000	10,0	22,4	21,6	21,2	11,4
2.000	7,1	17,0	16,1	15,0	8,0
3.000	5,9	14,5	13,6	12,3	6,5
4.000	5,1	12,9	12,0	10,6	5,6
5.000	4,6	11,8	10,9	9,5	5,0
6.000	4,2	11,0	10,1	8,7	
7.000	3,9	10,3	9,5	8,0	
8.000	3,6	9,8	8,9	7,5	
9.000	3,4	9,3	8,5	7,1	
10.000	3,3	9,0	8,1	6,7	
15.000	2,7	7,6	6,8		
20.000	2,3	6,8	6,0		
25.000	2,1	6,2	5,5		
30.000	1,9	5,8	5,1		
35.000		5,4	4,8		
40.000		5,2	4,5		
45.000		4,9	4,3		
50.000		4,7	4,1		
75.000		4,0			



**Prospetto D.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	Istituti Magistrali	Altri tipi di scuole
CENTRO					
200	20,6	40,9	43,9	51,5	26,6
400	15,2	32,1	32,8	37,3	18,9
600	12,7	27,9	27,6	30,9	15,5
800	11,2	25,2	24,5	27,0	13,5
1.000	10,2	23,3	22,3	24,3	12,1
2.000	7,5	18,3	16,6	17,6	8,6
3.000	6,3	15,9	14,0	14,6	7,0
4.000	5,5	14,4	12,4	12,7	
5.000	5,0	13,3	11,3	11,5	
6.000	4,6	12,5	10,5		
7.000	4,3	11,8	9,8		
8.000	4,1	11,3	9,3		
9.000	3,9	10,8	8,8		
10.000	3,7	10,5	8,4		
15.000	3,1	9,1	7,1		
20.000		8,2	6,3		
25.000		7,6	5,7		
30.000		7,1	5,3		
35.000		6,8			

**Prospetto D.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	Istituti Magistrali	Altri tipi di scuole
MEZZOGIORNO					
200	22,3	41,1	46,4	57,7	23,3
400	16,2	31,0	33,7	41,8	16,7
600	13,4	26,4	27,9	34,6	13,8
800	11,8	23,5	24,5	30,2	12,0
1.000	10,6	21,5	22,1	27,3	10,8
2.000	7,7	16,2	16,0	19,7	7,7
3.000	6,4	13,8	13,3	16,3	6,4
4.000	5,6	12,3	11,6	14,3	5,5
5.000	5,1	11,2	10,5	12,9	5,0
6.000	4,7	10,4	9,6	11,8	4,6
7.000	4,3	9,8	9,0	11,0	
8.000	4,1	9,3	8,4	10,3	
9.000	3,9	8,9	8,0	9,8	
10.000	3,7	8,5	7,6	9,3	
15.000	3,1	7,2	6,3	7,7	
20.000	2,7	6,4	5,5	6,8	
25.000	2,4	5,9	5,0	6,1	
30.000	2,2	5,5	4,6		
35.000		5,1	4,3		
40.000		4,9	4,0		
45.000		4,6	3,8		
50.000		4,4	3,6		
75.000		3,8			

**Prospetto D.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	Istituti Magistrali	Altri tipi di scuole
ITALIA					
200	22,6	47,2	47,6	57,1	26,0
400	16,1	34,7	34,3	40,9	18,5
600	13,2	29,0	28,3	33,7	15,2
800	11,5	25,5	24,7	29,3	13,2
1.000	10,3	23,1	22,2	26,4	11,8
2.000	7,3	17,0	16,0	18,9	8,4
3.000	6,0	14,2	13,2	15,6	6,9
4.000	5,2	12,5	11,5	13,6	6,0
5.000	4,7	11,3	10,4	12,2	5,3
6.000	4,3	10,4	9,5	11,2	4,9
7.000	4,0	9,8	8,9	10,4	4,5
8.000	3,7	9,2	8,3	9,7	4,2
9.000	3,5	8,7	7,9	9,2	4,0
10.000	3,3	8,3	7,5	8,7	3,8
15.000	2,7	7,0	6,2	7,2	3,1
20.000	2,4	6,1	5,4	6,3	
25.000	2,1	5,6	4,9	5,6	
30.000	1,9	5,1	4,5	5,2	
35.000	1,8	4,8	4,1	4,8	
40.000	1,7	4,5	3,9	4,5	
45.000	1,6	4,3	3,7		
50.000	1,5	4,1	3,5		
75.000	1,2	3,4	2,9		
100.000		3,0	2,5		
150.000		2,5			

## **Appendice E:**

### **La metodologia per la stima del rischio di violazione della riservatezza**

Per la protezione della riservatezza dei rispondenti all'indagine sui percorsi di studio e di lavoro dei diplomati dell'anno 1998, 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 un 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=23.262)	
s=0	s=1
20.458 (87,95%)	2.804 (12,05%)

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

Variabili	Sesso	Regione di residenza	Età	Tipo scuola superiore
<b>Numero di soppressioni (percentuale sul totale=23.262)</b>	10 (0,04%)	627 (2,69%)	2.142 (9,21%)	25 (0,11%)