

File Standard

Inserimento professionale dei
laureati dell'anno 1995
Indagine 1998

Manuale utente e tracciato record

Il manuale, frutto della collaborazione fra il Progetto istruzione e lavoro minorile e il Servizio Metodologie per la produzione statistica, è stato curato da Simona Pace.

In particolare: la premessa, i capitoli 1 e 2 e l'Appendice C sono stati realizzati da S. Pace, l'Appendice B da R. De Nisco e V. Parisi, l'Appendice D da S. Falorsi e L. Di Consiglio, l'Appendice E da G. Seri e A. Capobianchi.

Indice

PREMESSA	4
1. L'indagine sull'inserimento professionale dei laureati	5
1.1 <i>Le principali caratteristiche dell'indagine</i>	5
1.2 <i>Il processo di produzione dei dati</i>	6
1.3 <i>I contenuti informativi</i>	6
2. Le variabili contenute nel file	10
 APPENDICE A: TRACCIATO RECORD	 A-1
 APPENDICE B: CARTELLINI	 B-1
 APPENDICE C: DEFINIZIONI	 C-1
 APPENDICE D: LA STRATEGIA DI CAMPIONAMENTO E GLI ERRORI CAMPIONARI	 D-1
1. Il disegno di campionamento	D-1
2. Il calcolo delle stime	D-3
3. Valutazione del livello di precisione delle stime, gli errori campionari	D-5
 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.

Dopo una breve descrizione delle fasi principali dell'indagine sull'inserimento professionale dei laureati, 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, il questionario 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 sull'inserimento professionale dei laureati

1.1 Le principali caratteristiche e il disegno dell'indagine

Nel 1998 l'Istituto Nazionale di Statistica ha realizzato la quarta edizione dell'indagine sull'inserimento professionale dei laureati.

Obiettivo di quest'indagine è analizzare la condizione occupazionale dei laureati a circa tre anni dal conseguimento del titolo per tenere conto da una parte dell'eventuale espletamento del servizio militare o la conclusione di ulteriori attività di qualificazione, universitarie e non, dall'altra del fatto che i tempi di inserimento nel mercato del lavoro non sono brevi, soprattutto se si vuole anche indagare la qualità del lavoro trovato.

La popolazione di riferimento è costituita dai **laureati dell'anno solare 1995 in tutte le sedi universitarie italiane** per un totale di **105.097 individui** (49.393 maschi e 55.704 femmine).

L'indagine è campionaria e prevede la formazione di un **campione a uno stadio stratificato per sesso, sede universitaria di appartenenza e corso di laurea**. Il campione teorico è risultato pari a 25.716 unità; dei modelli di rilevazione inviati ne sono tornati all'ISTAT 17.326, con un tasso di risposta pari a circa il 67%.

L'indagine si è svolta in due fasi, entrambe postali: la prima, che ha coinvolto gli atenei, ha consentito di costruire l'elenco da cui estrarre i nominativi da contattare; la seconda, rivolta ai laureati facenti parte del campione, ha permesso la raccolta dei dati relativi ai singoli percorsi universitari e post-laurea.

Nella prima fase è stato fatto un censimento dei laureati dell'anno 1995 presso tutti gli atenei italiani; nella seconda è stato inviato a ciascun laureato facente parte del campione un questionario da compilare e restituire all'Istat.

A partire dall'elenco dei laureati di tutte le università italiane è stato costruito un campione di 25.716 laureati da intervistare (con un tasso di campionamento del 24,5%).

Il campione di laureati è stato intervistato mediante **questionario postale autocompilato**. I rispondenti all'indagine sono stati **17.326** con un tasso di risposta pari al **67,4%**.

1.2 Il processo di produzione dei dati

L'indagine prevede una serie di fasi, di seguito sintetizzate.

I questionari, compilati e restituiti per posta, vengono controllati dal personale che ne registra l'arrivo, l'avvenuta compilazione o, eventualmente, i motivi della non compilazione. I singoli questionari vengono quindi revisionati e, in particolare, si procede alla codifica manuale di alcuni quesiti tra i quali la professione svolta dall'intervistato e dai genitori, la provincia e il comune. I questionari revisionati e codificati vengono quindi mandati in registrazione. La fase successiva è relativa al controllo e alla correzione dei dati. Le incoerenze e le incompatibilità vengono eliminate mediante specifici software basati su metodologie deterministiche e probabilistiche. In particolare i metodi probabilistici utilizzati si basano sulla ricerca di un laureato-donatore con caratteristiche simili a quelle del laureato che presenta il dato mancante o errato, che possa donare l'informazione.

Si procede, quindi, alla costruzione delle stime utilizzando degli stimatori di ponderazione vincolati a sesso, corso e sede universitaria.

I dati vengono 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.

A questo punto i dati possono essere diffusi. 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 utilizzato per l'indagine presenta diverse novità rispetto ai precedenti. La sua ristrutturazione è stata motivata dall'esigenza di assicurare sia la confrontabilità con le informazioni raccolte nelle altre due indagini sulla transizione dei diplomati¹ e dei diplomati universitari², sia della necessaria integrazione con altre fonti, in particolare con l'indagine sulle forze di lavoro.

¹ Istat (2000): Percorsi di studio e di lavoro dei diplomati – Indagine 1998, Collana Informazioni

² Istat (2001): Inserimento professionale dei diplomati universitari – Indagine 1999, Collana Informazioni (in corso di pubblicazione)

Rispetto alla precedente edizione la sequenza dei quesiti è stata modificata cercando di seguire maggiormente l'ordine cronologico degli eventi che caratterizzano il percorso formativo e professionale dei laureati (così come nelle altre due indagini sulla transizione): la prima sezione del questionario è stata dedicata al curriculum degli studi e alle attività di qualificazione, la seconda al lavoro, la terza alla ricerca del lavoro, mentre nella quarta e nella quinta si chiedono notizie relative alla famiglia d'origine e informazioni anagrafiche. Le diverse sezioni sono state ampliate, tenendo conto della necessità di sviluppare maggiormente alcuni temi presenti anche nelle altre indagini sulla transizione.

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, evidenziando i principali cambiamenti apportati.

SEZIONE 1 – CURRICULUM

Il percorso formativo intrapreso rappresenta certamente una determinante degli esiti occupazionali: per tale motivo l'indagine propone un'ampia sezione nella quale viene analizzato il curriculum degli studi chiedendo informazioni sulla carriera sia scolastica che universitaria.

L'analisi del curriculum si amplia di informazioni relative alla frequenza di corsi e alla valutazione della didattica.

SEZIONE 2 – LAVORO

La parte relativa al lavoro tiene conto del fatto che spesso la mobilità occupazionale nei primi anni di "vita lavorativa" è assai elevata: molti giovani, infatti, passano per fasi di precariato sperimentando condizioni occupazionali diverse. Dopo una prima parte dedicata a ricostruire la storia lavorativa dell'intervistato dopo il conseguimento della laurea, la sezione prosegue con l'analisi del lavoro svolto.

La parte relativa al lavoro, e in particolare quella relativa al lavoro di tipo "stabile", è più ampia rispetto al passato e descrive più compiutamente le caratteristiche del lavoro svolto: non solo, quindi, il tipo di lavoro (la posizione nella professione, la branca di attività economica, la mansione specifica, l'area professionale ecc...), ma anche - ed è questa la novità rispetto alle precedenti edizioni - la qualità del lavoro svolto (il numero di ore lavorate, la retribuzione, la regolarizzazione o meno del rapporto di lavoro, la formazione professionale ecc...).

Il patrimonio relazionale dei giovani si rivela di estrema importanza per trovare lavoro. Per questo motivo nell'ultima parte della sezione dedicata alle modalità attraverso cui si è trovato

l'attuale lavoro è stata introdotta una batteria di quesiti relativi alla persona che eventualmente si fosse rivelata utile per iniziare la propria attività.

SEZIONE 3 – RICERCA DI LAVORO

La sezione riguarda le modalità di ricerca del lavoro nonché le attese retributive, professionali e di carriera dei giovani intervistati. In questa edizione la sezione è stata ampliata con i quesiti relativi al tipo di lavoro cercato, all'orario preferito e alla retribuzione minima richiesta.

SEZIONE 4 – NOTIZIE SULLA FAMIGLIA D'ORIGINE

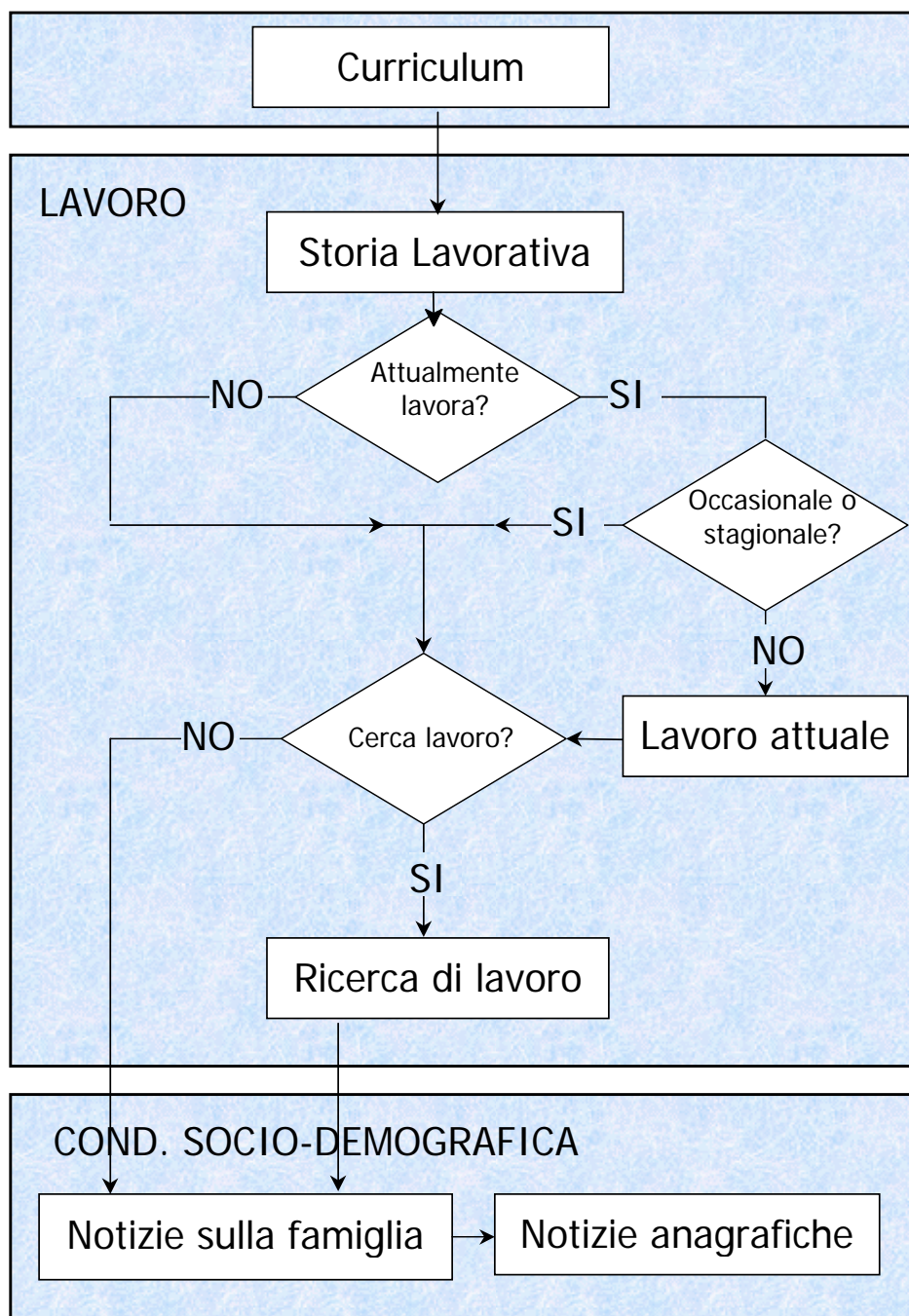
L'analisi dell'inserimento professionale dei laureati viene svolta tenendo conto anche del contesto sociale ed economico in cui si è formato lo studente. In questo modo è possibile analizzare sia l'effetto delle condizioni sociali di origine sul successo negli studi, sia confrontare la posizione sociale dei giovani rispetto a quelle della famiglie d'origine.

Rispetto al passato, la sezione si arricchisce dei quesiti riguardanti la professione dei genitori.

SEZIONE 5 – NOTIZIE ANAGRAFICHE

Questa parte è stata uniformata alle altre indagini dell'Istat, in base a definizioni internazionali.

Schema 1 - Sezioni e principali quesiti di snodo del questionario sull'Inserimento professionale dei laureati



1.4 Le variabili contenute nel file

Per la nuova edizione dell'indagine sull'Inserimento professionale dei laureati l'Istat ha deciso di adottare una nuova metodologia di protezione dei dati. Ciò implica l'individuazione e la soppressione di alcune informazioni sui record che presentano un rischio di identificazione elevato e, di conseguenza, un maggior numero di informazioni disponibili rispetto al passato. Conseguenza dell'utilizzo di tale metodologia di soppressione (vedi Appendice E) è la presenza di alcuni blank nelle variabili corso di laurea, sede universitaria, sesso, residenza ed età.

Inoltre per queste e per alcune altre 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 descrizione vengono invece indicati i contenuti delle variabili, descrivendo le codifiche adottate o rimandando ai cartellini (Appendice B) 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 laureati dell'anno 1995.

ALLEGATO A:

TRACCIATO RECORD

Posizione		Ampiezza del campo	Variabile	Descrizione	
1	2	2	Gruppo	Gruppo di corsi di laurea	vedi cartellino GRUPPI
3	4	2	Corso	Corso di Laurea	vedi cartellino CORSI
5	5	1	Sede	Sede Universitaria	vedi cartellino SEDI
6	7	2	Q.1	Tipo di maturità	
8	9	2	Q.2	Voto maturità	
10	10	1	Q.3	Altra laurea	
11	12	2	Q.4	Gruppo prima laurea	vedi cartellino GRUPPI
13	13	1	Q.5	Motivo nuovo corso	
14	17	4	Q.8	Anno accademico di prima immatricolazione	
18	18	1	Q.9	Durata corso	
19	19	1	Q.10	Cambio corso	
20	21	2	Q.11	Gruppo corso precedente	vedi cartellino GRUPPI
22	22	1	Q.12	Trasferimento per studi	
23	23	1	Q.13	Motivo mancato trasferimento	
24	24	1	Q.14	Abitazione	
25	25	1	Q.15	Lavori durante gli studi	
26	27	2	Q.16	Durata studi universitari	
28	30	3	Q.17	Voto laurea	
31	31	1	Q.18	Lode	
32	32	1	Q.19_1	Votazione massima	
33	35	3	Q.19_2	Altra votazione massima	
36	36	1	Q.20	Nuova iscrizione	
37	37	1	Q.21_1	Tipo corso nuova iscrizione	
38	39	2	Q.21_2	Gruppo del corso di nuova iscrizione	vedi cartellino GRUPPI
40	40	1	Q.22	Motivo non iscrizione allo stesso corso	
41	41	1	Q.23	Frequenza lezioni insegnamento fondamentale	
42	42	1	Q.24_1	Competenza docente sulla materia	1 = molto 2 = abbastanza 3 = poco 4 = per niente 5 = non indicato
43	43	1	Q.24_2	Chiarezza espositiva del docente	1 = molto 2 = abbastanza 3 = poco 4 = per niente 5 = non indicato
44	44	1	Q.24_3	Presenza/puntualità alle lezioni	1 = molto 2 = abbastanza 3 = poco 4 = per niente 5 = non indicato
45	45	1	Q.24_4	Presenza/puntualità orari di ricevimento	1 = molto 2 = abbastanza 3 = poco 4 = per niente 5 = non indicato
46	46	1	Q.24_5	Disponibilità rapporto con gli studenti	1 = molto 2 = abbastanza 3 = poco 4 = per niente 5 = non indicato

Posizione		Ampiezza del campo	Variabile	Descrizione	
47	47	1	Q.25	Corsi privati di preparazione agli esami	
48	48	1	Q.26	Esami di stato	
49	50	2	Q.27_1	Dottorato di ricerca	
51	52	2	Q.27_2	Specializzazione post-laurea	
53	54	2	Q.27_3	Ulteriori studi universitari	
55	56	2	Q.27_4	Borse di studio	
57	58	2	Q.27_5	Tirocinio, praticantato, stage	
59	60	2	Q.27_6	Volontariato	
61	62	2	Q.27_7	Altre attività di qualificazione	
63	64	2	Q.27_8	Nessuna attività in corso	
65	66	2	Q.27_9	Nessuna attività conclusa	
67	68	2	Q.27_10	Nessuna attività interrotta	
69	69	1	Q.28	Opportunità di lavoro post-laurea	
70	70	1	Q.29	Accettazione opportunità di lavoro post-laurea	
71	71	1	Q.30	Motivo principale rifiuto lavoro post-laurea	
72	72	1	Q.31	Esperienze di lavoro interrotte	
73	73	1	Q.32	Tipo di lavoro interrotto	
74	75	2	Q.33_1	Mese inizio attività interrotta	
76	77	2	Q.33_2	Anno inizio attività interrotta	
78	78	1	Q.34	Occupazione	
79	79	1	Q.35	Svolgimento ore di lavoro dei non occupati	
80	81	2	Q.36	Ore di lavoro svolte dai non occupati	
82	82	1	Q.37	Ricerca di lavoro dei non occupati	
83	83	1	Q.38	Tipo di lavoro attualmente svolto	
84	84	1	Q.39	Inizio attuale lavoro	
85	86	2	Q.40_1	Mese inizio attuale lavoro	
87	88	2	Q.40_2	Anno inizio attuale lavoro	
89	89	1	Q.41	Laurea requisito necessario	
90	90	1	Q.42	Tipo di laurea richiesta	
91	91	1	Q.43	Tipo di impiego	
92	92	1	Q.44	Ore di lavoro settimanali	1 = fino a 20 ore 2 = 21-35 3 = 36-40 4 = oltre 40
93	93	1	Q.45_2	Guadagno mensile netto	1 = fino a 1.500.000 2 = 1.501.000-2.000.000 3 = 2.001.000-3.000.000 4 = più di 3.000.000
94	94	1	Q.46	Professione laureato	vedi cartellino PROFESSIONI
95	96	2	Q.47	Ruolo professionale	
97	97	1	Q.48	Partita IVA	
98	98	1	Q.49	Contributi per la pensione	
99	99	1	Q.50	Tipo di settore di attività economica	
100	101	2	Q.51	Settore di attività economica	
102	102	1	Q.52	Numero di persone impegnate nell'impresa	
103	104	2	Q.53	Area di attività	
105	105	1	Q.54	Partecipazione a corsi di formazione professionale	
106	106	1	Q.55	Durata corsi di formazione professionale	
107	107	1	Q.56_1	Grado di soddisfazione relativo a trattamento economico	
108	108	1	Q.56_2	Grado di soddisfazione relativo a possibilità di carriera	
109	109	1	Q.56_3	Grado di soddisfazione relativo a stabilità/sicurezza	
110	110	1	Q.56_4	Grado di soddisfazione utilizzo conoscenze universitarie	
111	111	1	Q.57	Necessità della laurea per lo svolgimento dell'attuale lavoro	

Posizione		Ampiezza del campo	Variabile	Descrizione	
112	113	2	Q.58	Mezzo utilizzato per trovare l'attuale lavoro	
114	114	1	Q.59	Tipo di persona e aiuto utile per l'inizio dell'attuale lavoro	
115	115	1	Q.60	Persona che ha fornito l'aiuto	
116	117	2	Q.61	Ruolo professionale della persona che ha fornito l'aiuto	
118	118	1	Q.62	Ricerca di un nuovo lavoro	
119	120	2	Q.63_1	Iniziative prese - nessuna	
121	122	2	Q.63_2	Iniziative prese - richiesta a familiari	
123	124	2	Q.63_3	Iniziative prese - richiesta ad altri parenti	
125	126	2	Q.63_4	Iniziative prese - richiesta ad amici/conoscenti	
127	128	2	Q.63_5	Iniziative prese - segnalazione università	
129	130	2	Q.63_6	Iniziative prese - contatti diretti con datori di lavoro	
131	132	2	Q.63_7	Iniziative prese - inserzioni	
133	134	2	Q.63_8	Iniziative prese - concorsi	
135	136	2	Q.63_9	Iniziative prese - lavoro in proprio	
137	138	2	Q.63_10	Iniziative prese - domande a Provveditori o Presidi	
139	140	2	Q.63_11	Iniziative prese - iscrizione ufficio di collocamento	
141	142	2	Q.63_12	Iniziative prese - agenzie private di collocamento	
143	144	2	Q.63_13	Iniziative prese - altre iniziative	
145	145	1	Q.64	Tempo intercorso dall'ultima iniziativa	
146	146	1	Q.65	Tipo di lavoro cercato	
147	147	1	Q.66	Tipo di orario preferito	
148	148	1	Q.67	Possibilità di iniziare subito un'attività	
149	149	1	Q.68	Disponibilità a trasferirsi	
150	150	1	Q.69	Cifra minima desiderata	1 = fino a 1.500.000 2=1.501.000-2.000.000 3=2.001.000-3.000.000 4= più di 3.000.000
151	151	1	Q.70	Fratelli/sorelle	
152	153	2	Q.71	Numero fratelli/sorelle	
154	154	1	Q.72_1	Titolo di studio del padre	1-8, 8=non indicato
155	155	1	Q.72_2	Titolo di studio della madre	1-8, 8=non indicato
156	156	1	Q.73_1	Condizione professionale del padre	
157	157	1	Q.73_2	Condizione professionale della madre	
158	158	1	Q.74_1	Ramo di attività economica del padre	
159	159	1	Q.74_2	Ramo di attività economica della madre	
160	160	1	Q.75_1	Professione del padre	vedi cartellino PROFESSIONI
161	161	1	Q.75_2	Professione della madre	vedi cartellino PROFESSIONI
162	163	2	Q.76_1	Ruolo professionale del padre	
164	165	2	Q.76_2	Ruolo professionale della madre	
166	167	2	Q.77	Eta'	24=fino a 24 anni 25=25 anni 26=26 anni 27=27 anni 28=28-29 anni 30=dai 30 anni in su
168	168	1	Q.78	Sesso	
169	169	1	Q.79	Obblighi di leva	
170	170	1	Q.80_1	Convivenza - da solo o con amici	
171	171	1	Q.80_2	Convivenza - con suo padre	
172	172	1	Q.80_3	Convivenza - con sua madre	
173	173	1	Q.80_4	Convivenza - con fratelli/sorelle	
174	174	1	Q.80_5	Convivenza - con coniuge/convivente	
175	175	1	Q.80_6	Convivenza - con figli	
176	176	1	Q.80_7	Convivenza - con altri parenti o affini	

Posizione		Ampiezza del campo	Variabile	Descrizione	
177	177	1	Q.81	Stato civile	1=altro 2=coniugato-a/convivente
178	178	1	Q.82	Cittadinanza italiana	
179	179	1	Q.83	Cittadinanza estera	
180	180	1	Q.84	Provincia residenza precedente l'iscrizione	1=stessa provincia residenza attuale 2=altra provincia
181	181	1	Q.85	Provincia residenza attuale	Vedi cartellino MACRO-REGIONI
182	182	1	Q.86	Provincia attuale lavoro	1=stessa provincia residenza attuale 2=altra provincia
183	190	8	COEF	Coefficiente di riporto all'universo	numerico

ALLEGATO B:

CARTELLINI

Cartellino GRUPPI

Codice	Gruppo di corsi di laurea
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

Cartellino CORSI

Codice	Corso di laurea
01	Matematica
02	Altri corsi gruppo scientifico
03	Scienze dell'informazione
05	Altri corsi gruppo chimico-farmaceutico
06	Chimica industriale
09	Scienze biologiche
10	Scienze geologiche
11	Altri corsi gruppo geo-biologico
13	Medicina e chirurgia
14	Odontoiatria e protesi dentarie
15	Ingegneria meccanica
16	Altri corsi gruppo ingegneria
17	Ingegneria elettronica
18	Ingegneria chimica
20	Ingegneria civile
23	Corsi gruppo architettura
25	Scienze agrarie
26	Medicina veterinaria
27	Altri corsi gruppo agrario
28	Economia e commercio
29	Scienze economiche e bancarie
30	Scienze statistiche
31	Economia aziendale
32	Altri corsi gruppo economico-statistico
33	Scienze politiche
34	Altri corsi gruppo politico-sociale
36	Corsi gruppo giuridico
38	Lettere
39	Materie letterarie
40	Filosofia
41	Altri corsi gruppo letterario
44	Altri corsi gruppo linguistico
45	Lingue e letterature straniere
47	Corsi gruppo insegnamento
49	Psicologia

Cartellino SEDI

Codice	Regione della sede universitaria
1	Piemonte, Liguria
2	Lombardia
3	Trentino-Alto Adige, Veneto, Friuli-Venezia Giulia
4	Emilia-Romagna
5	Toscana, Umbria, Marche
6	Lazio
7	Abruzzo, Molise, Campania
8	Puglia, Basilicata, Calabria
9	Sicilia, Sardegna

Cartellino PROFESSIONI

Codice	Professione (primo digit)
1	Legislatori, dirigenti e imprenditori
2	Professioni intellettuali, scientifiche e di elevata specializzazione
3	Professioni intermedie (tecnici)
4	Professioni esecutive relative all'amministrazione e gestione
5	Professioni relative alle vendite ed ai servizi per le famiglie
6	Artigiani, operai specializzati e agricoltori
7	Conduttori di impianti, operatori di macchinari fissi e mobili e operai di montaggio industriale
8	Personale non qualificato
9	Forze armate

Cartellino MACRO-REGIONI

Codice	Regione di residenza
0	Esteri
1	Piemonte, Liguria
2	Lombardia
3	Trentino-Alto Adige, Veneto, Friuli-Venezia Giulia
4	Emilia-Romagna
5	Toscana, Umbria, Marche
6	Lazio
7	Abruzzo, Molise, Campania
8	Puglia, Basilicata, Calabria
9	Sicilia, Sardegna

ALLEGATO C:

DEFINIZIONI

Voto di laurea: è espresso in centodecimi, anche quando la scala prevista è diversa.

Corsi post-laurea: comprendono i corsi di perfezionamento (di durata non superiore a un anno), le scuole di specializzazione (con durata da 2 a 5 anni) e i corsi di dottorato di ricerca (3-4 anni).

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

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 (dottorato di ricerca, borsa di studio, tirocinio, stage eccetera) anche se retribuite.

Persona in cerca di occupazione: i laureati che dichiarano di non lavorare e 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.

ALLEGATO D:

STRATEGIA DI CAMPIONAMENTO E LIVELLO DI PRECISIONE DELLE STIME

Strategia di campionamento

Sono qui illustrati gli aspetti più significativi della *strategia di campionamento* dell'indagine sull'inserimento professionale dei laureati dell'anno 1995.

La *popolazione di interesse* – ossia l'insieme delle unità statistiche intorno alle quali si intende investigare – è costituita dagli oltre 100.000 studenti che hanno conseguito il diploma di laurea nel corso dell'anno 1995 - tre anni prima della data di osservazione dell'indagine - in tutte le sedi universitarie italiane, con esclusione di coloro che hanno conseguito diplomi universitari che costituiscono titoli non equiparabili alla laurea.

L'indagine è volta ad acquisire una vasta gamma di informazioni sulle condizioni dei laureati, a circa tre anni dal conseguimento della laurea, e si propone lo studio delle transizioni dall'università al mondo del lavoro. L'intervallo di tre anni dovrebbe consentire di analizzare l'inserimento nel mondo del lavoro in una fase di relativa stabilità nella condizione occupazionale, dopo che gli individui hanno concluso eventuali corsi di perfezionamento e specializzazione post-laurea ed espletato il loro servizio di leva.

Gli obiettivi conoscitivi più significativi riguardano lo stato occupazionale, la stabilità dell'occupazione, la posizione nella professione e l'area di attività economica distintamente per sesso, corso di laurea e ripartizione geografica di residenza.

I *domini di interesse* sono l'intero territorio nazionale, le ripartizioni geografiche e i gruppi di corsi di laurea incrociati con il sesso e le regioni per un numero limitato di stime.

Descrizione generale del disegno di campionamento

Per quanto riguarda la progettazione del disegno campionario occorre fare presente che per ciascuna sede universitaria si dispone della lista dei laureati di ciascun anno solare, inoltre la rilevazione è organizzata mediante autocompilazione dei questionari inviati per via postale; è possibile, quindi, definire un disegno di campionamento ad uno stadio, con stratificazione delle unità per sesso, sede universitaria e corso di laurea (gli strati sono risultati 878).

La numerosità campionaria complessiva viene determinata in modo da garantire che gli errori di campionamento attesi delle principali stime riferite ai diversi domini di interesse non superino prefissati livelli, tenendo conto anche dei tassi di risposta attesi. La dimensione campionaria teorica complessiva è risultata pari a 25.716 unità; dei modelli di rilevazione inviati ne sono tornati all'ISTAT 17.326, con un tasso di risposta pari a circa il 67%.

Una volta definite le numerosità campionarie teoriche negli strati la selezione delle unità campionarie avviene senza reimmissione con probabilità uguali.

Allo scopo di illustrare l'allocazione della dimensione campionaria tra gli strati, viene riportata nel Prospetto 1 la distribuzione dei laureati dell'anno 1995 distinta per sesso e per gruppo di corsi di laurea, nell'universo dei laureati e nel campione estratto.

Prospetto 1 - Distribuzione dei laureati dell'anno 1995, nell'universo e nel campione per gruppo di corsi di laurea e sesso.

GRUPPI DI CORSI	MASCHI		FEMMINE		TOTALE	
	Laureati Universo	Laureati Campione	Laureati Universo	Laureati Campione	Laureati Universo	Laureati Campione
Gruppo scientifico	2380	720	2004	719	4384	1439
Gruppo chimico-farmaceutico	1705	644	2500	853	4205	1497
Gruppo geo-biologico	1824	675	3028	940	4852	1615
Gruppo medico	4061	991	3580	937	7641	1928
Gruppo ingegneria	9004	2307	1159	687	10163	2994
Gruppo architettura	2996	511	2643	493	5639	1004
Gruppo agrario	1619	461	925	371	2544	832
Gruppo economico-statistico	11065	2154	8655	1907	19720	4061
Gruppo politico-sociale	3381	704	3841	848	7222	1552
Gruppo giuridico	7455	1438	8906	1839	16361	3277
Gruppo letterario	2370	796	7454	1809	9824	2605
Gruppo linguistico	736	338	6649	1405	7385	1743
Gruppo insegnamento	417	195	2838	600	3255	795
Gruppo psicologico	367	80	1535	294	1902	374
Totale	49380	12014	55717	13702	105097	25716

Procedimento per il calcolo delle stime

Le stime prodotte dall'indagine sono principalmente stime di frequenze assolute e relative riferite ai laureati dell'anno 1995; una stima di interesse è data, ad esempio, dal numero totale di laureati che lavorano a un anno dal conseguimento del 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 a 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, ciò significa 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 del livello di riferimento delle stime (domini di interesse);
- i , indice di laureato;
- h , indice dello strato (sede \times corso \times sesso);
- N_h , totale dei laureati dello strato h ;
- n_h^* , campione teorico dei laureati dello strato h ;
- n_h , campione effettivo dei laureati dello strato h ;
- H , numero totale di strati;
- $D_{hi} = \frac{N_h}{n_h^*}$, peso diretto da attribuire all' i -mo laureato dello strato h ;
- W_{hi} , peso finale da attribuire all' i -mo laureato dello strato h ;
- y , generica variabile oggetto di indagine;
- Y_{hi} , valore osservato della variabile y sull' i -mo laureato dello strato h ;
- δ_{hid} , variabile indicatrice pari a 1 se l'unità i dello strato h appartiene al dominio di interesse.

Se, ad esempio, y rappresenta la condizione lavorativa (espressa dalle due modalità lavora, non lavora), si avrà $Y_{hi} = 1$ se il laureato i -mo lavora e $Y_{hi} = 0$ altrimenti.

Si supponga di voler stimare con riferimento a un generico dominio d , il totale della variabile in esame, espresso dalla relazione:

$$Y_d = \sum_{h=1}^H \sum_{i=1}^{N_h} Y_{hi} \delta_{hid} \quad (1)$$

La stima del totale (1), viene determinata mediante uno stimatore del rapporto separato espresso mediante la seguente formula:

$$\hat{\hat{Y}}_d = \sum_{h=1}^H \frac{\hat{Y}_h}{\hat{M}_h} M_h \quad (2)$$

dove

$$\hat{M}_h = \sum_{i=1}^{n_h} D_{hi} \quad \text{e} \quad \hat{Y}_h = \sum_{i=1}^{n_h} Y_{hi} \delta_{hid} D_{hi} \quad (3)$$

rappresentano rispettivamente, con riferimento allo strato h , la stima diretta del totale dei laureati, M_h (valore noto dai database delle Università) e la stima diretta del totale dei laureati che possiedono una determinata caratteristica, Y_h . La quantità M_h differisce da N_h in quanto l'estrazione del campione è stata effettuata da un archivio non definitivo dei laureati del 1995 che non risulta esattamente coincidente con l'archivio successivamente pubblicato dal quale sono stati desunti i valori di M_h . Lo stimatore appena illustrato rientra nella classe degli *stimatori di ponderazione vincolata*³, 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. Una delle caratteristiche dei suddetti stimatori è di garantire l'uguaglianza tra i totali noti delle variabili ausiliarie e le corrispondenti stime ottenute con i pesi finali. Nell'indagine in oggetto i totali noti sono costituiti, come detto sopra, dal numero di laureati classificati per sesso, corso e sede universitaria desunte dai database delle Università.

Per la valutazione dei pesi da attribuire a ciascuna unità è stato utilizzato un software generalizzato che costruisce il peso finale di ciascuna unità hi per mezzo delle seguenti fasi:

- 1) determinazione del peso diretto, D_{hi} , ottenuto come inverso della probabilità di inclusione della unità; poiché in ciascuno strato si adotta il campionamento semplice senza ripetizione, il peso diretto è dato da N_h / n_h^*
- 2) calcolo di un fattore correttivo, γ_{hi} , per mancata risposta totale; ottenuto come l'inverso della probabilità di risposta dell'unità hi ; poiché nella presente indagine non si dispone di altre informazioni legate alla probabilità di risposta tale probabilità viene stimata mediante il tasso di risposta dello strato h cui l'unità i appartiene, $\gamma_{hi} = n_h^* / n_h$;
- 3) calcolo del fattore correttivo, c_{hi} , che consente di soddisfare, in ogni strato, la condizione di uguaglianza tra i totali noti della popolazione e le corrispondenti stime campionarie;
- 4) calcolo del peso finale, W_{hi} , ottenuto come prodotto del peso diretto per i fattori correttivi determinati nei passi due e tre.

A partire dalla formula (2), mediante semplici passaggi, è possibile esprimere lo stimatore del rapporto come funzione del peso finale, si ottiene infatti:

$$\hat{Y}_d = \sum_{h=1}^H \sum_{i=1}^{n_h} Y_{hi} \delta_{hid} W_{hi} , \quad (4)$$

dove

$$W_{hi} = D_{hi} \frac{M_h}{\sum_{i=1}^{n_h} D_{hi}} \gamma_{hi} = D_{hi} c_{hi} \gamma_{hi} = \frac{M_h}{n_h} . \quad (5)$$

Per cui la stima di Y_d è semplicemente

$$\hat{Y}_d = \sum_{h=1}^H \bar{y}_{hd} M_h ,$$

dove \bar{y}_{hd} è data da

$$\bar{y}_{hd} = \frac{1}{n_h} \sum_{i=1}^{n_h} Y_{hi} \delta_{hid} .$$

Valutazione del livello di precisione delle stime

Calcolo della varianza campionaria

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

La stima dell'errore di campionamento assoluto e dell'errore di campionamento relativo di una generica stima \hat{Y}_d sono definite rispettivamente dalle seguenti espressioni:

$$\hat{\sigma}(\hat{Y}_d) = \sqrt{\hat{V}ar(\hat{Y}_d)} \quad (6)$$

$$\hat{\varepsilon}(\hat{Y}_d) = \frac{\hat{\sigma}(\hat{Y}_d)}{\hat{Y}_d} \quad (7)$$

La stima della varianza di \hat{Y}_d , indicata nella (6) come $\hat{V}ar(\hat{Y}_d)$, è stata calcolata utilizzando il metodo di linearizzazione di Woodruff, che consente di ottenere un'espressione approssimata della varianza campionaria nel caso generale di stimatori come quello descritto nel paragrafo 1.2, che non sono funzione lineare dei dati campionari. In simboli si ha:

Noti nella letteratura anglosassone sull'argomento come calibration estimators, si veda Deville, J.C. and Sarndäl, C.E. (1992) Calibration Estimators in Survey Sampling. Journal of the American Statistical Association, 87, 376-382

$$\hat{\text{Var}}(\hat{\hat{Y}}_d) \cong \sum_{h=1}^H \frac{N_h^2}{n_h} \left(1 - \frac{n_h}{N_h}\right) \sum_{i=1}^{n_h} \frac{(\hat{Z}_{hid} - \hat{\bar{Z}}_{hd})^2}{n_h - 1} \quad (8)$$

in cui

$$\hat{Z}_{hid} = \left(Y_{hi} \delta_{hid} - \frac{\hat{Y}_h}{\hat{M}_h} \right) \quad (9)$$

è l'espressione della variabile linearizzata relativa allo stimatore del rapporto utilizzato per l'indagine in questione.

Gli errori campionari espressi dalla (6) e dalla (7) consentono di valutare il grado di precisione delle

stime; inoltre, la conoscenza dell'errore assoluto permette anche di costruire un intervallo di confidenza α , sulla base della relazione

$$\Pr\left\{Y_d - k\hat{\sigma}(\hat{\hat{Y}}_d) \leq \hat{\hat{Y}}_d \leq Y_d + k\hat{\sigma}(\hat{\hat{Y}}_d)\right\} = \alpha \quad (10)$$

Avendo, ad esempio, prefissato $\alpha = 0,95$ l'intervallo di confidenza è dato da:

$$\left\{ \hat{\hat{Y}}_d - 1,96 \cdot \hat{\sigma}(\hat{\hat{Y}}_d); \hat{\hat{Y}}_d + 1,96 \cdot \hat{\sigma}(\hat{\hat{Y}}_d) \right\}$$

Presentazione sintetica degli errori campionari

Ad ogni stima $\hat{\hat{Y}}_d$ è associato un errore campionario relativo $\hat{\varepsilon}(\hat{\hat{Y}}_d)$ e pertanto, per consentire un uso corretto delle stime fornite dall'indagine, sarebbe necessario pubblicare anche il corrispondente errore di campionamento relativo. Non è, tuttavia, possibile soddisfare questa esigenza di informazione, sia per motivi di tempi e costi eccessivi di elaborazione, sia perché le tavole di pubblicazione risulterebbero eccessivamente appesantite e di non agevole consultazione per l'utente. Inoltre, non sarebbero in ogni caso disponibili gli errori delle stime non pubblicate che l'utente può ricavare in modo autonomo.

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

Il modello utilizzato, con riferimento alle stime di frequenze assolute, è del tipo seguente:

$$\log \hat{\varepsilon}^2(\hat{\hat{Y}}_d) = a + b \log(\hat{\hat{Y}}_d) \quad (11)$$

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

Nella presente indagine è stato stimato un modello di tipo (11) per ciascuno dei seguenti domini di interesse:

- D1. il totale Italia e il totale Italia distintamente per sesso;
- D2. le 5 ripartizioni geografiche in totale e distintamente per sesso;
- D3. i gruppi di corsi di laurea in totale e distintamente per sesso.

Per calcolare gli errori di campionamento è stato utilizzato un software generalizzato messo a punto dall'ISTAT che consente di calcolare per le variabili prescelte tra quelle rilevate, oltre agli errori di campionamento assoluti e relativi, anche gli intervalli di confidenza; inoltre permette di costruire modelli regressivi del tipo (11) 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 6 per i domini definiti dai gruppi di corsi di laurea distintamente per sesso.

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

Dalla (11) mediante semplici passaggi si ricava:

$$\hat{\varepsilon}(\hat{Y}_d^*) = \sqrt{\exp(a + b \log(\hat{Y}_d^*))}. \quad (12)$$

Prospetto 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 laureati dell'anno 1995 distinti per sesso e in totale per l'Italia e le 5 ripartizioni geografiche.

DOMINIO DI STIMA:		a	b	R^2
Ripartizione per sesso				
NORD-OVEST	M	2,503	-1,113	98,98
	F	2,309	-1,102	98,96
	<i>Totale</i>	2,483	-1,108	98,99
NORD-EST	M	2,058	-1,071	98,79
	F	1,989	-1,074	99,20
	<i>Totale</i>	2,128	-1,080	99,22
CENTRO	M	2,222	-1,081	99,26
	F	2,169	-1,084	99,27
	<i>Totale</i>	2,282	-1,087	99,38
SUD	M	2,074	-1,049	99,39
	F	2,258	-1,085	99,15
	<i>Totale</i>	2,267	-1,074	99,37
ISOLE	M	2,367	-1,095	99,10
	F	1,920	-1,067	96,94
	<i>Totale</i>	2,126	-1,072	97,72
TOTALE	M	2,435	-1,089	99,35
	F	2,329	-1,089	99,27
	<i>Totale</i>	2,802	-1,137	98,75

Se ad esempio la generica stima \hat{Y}_d^* si riferisce alla ripartizione Italia Centro per gli individui di sesso maschile, è possibile introdurre nella (12) i valori dei parametri a e b (a= 2,222, b= -1,081) riportati nella settima riga del prospetto 1 e ricavare il corrispondente errore relativo.

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

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

Allo scopo di facilitare il calcolo degli errori campionari per le stime che si riferiscono alle 5 ripartizioni geografiche e al totale Italia, nei prospetti 3, 4 e 5 sono riportati per alcuni valori crescenti di stime di frequenze percentuali (colonna 1) i corrispondenti valori assoluti delle stime (colonne 2, 4, 6, 8 10 e 12) e i valori interpolati degli errori relativi percentuali (colonne 3, 5, 7, 9, 11 e 13), calcolati introducendo nella (12) i valori di a e b riportati nel prospetto 2.

Per ciascun dominio il suddetto prospetto ha la seguente struttura:

Stime di frequenze percentuali	Corrispondenti stime di frequenze assolute	Errori relativi percentuali interpolati
1%	$\hat{Y}_d(1)$	$\hat{\varepsilon} [\hat{Y}_d(1)]$
2%	$\hat{Y}_d(2)$	$\hat{\varepsilon} [\hat{Y}_d(2)]$
.....
50%	$\hat{Y}_d(50)$	$\hat{\varepsilon} [\hat{Y}_d(50)]$

Le informazioni contenute nei tre prospetti, permettono di calcolare l'errore relativo di una generica stima di frequenza assoluta o relativa mediante due procedimenti che risultano di facile applicazione, anche se conducono a risultati meno precisi di quelli che possono essere ottenuti applicando direttamente l'espressione (12).

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.

Con il secondo metodo, più preciso del primo, si ricorre alla formula di interpolazione lineare per calcolare gli errori di stime non comprese tra i valori dati nei prospetti; pertanto, l'errore campionario della stima \hat{Y}_d^* , si ricava mediante l'espressione:

$$\hat{\varepsilon}(\hat{Y}_d^*) = \hat{\varepsilon}(\hat{Y}_d^{k-1}) + \frac{\hat{\varepsilon}(\hat{Y}_d^{k-1}) - \hat{\varepsilon}(\hat{Y}_d^k)}{\hat{Y}_d^k - \hat{Y}_d^{k-1}} (\hat{Y}_d^* - \hat{Y}_d^{k-1}) \quad (13)$$

dove \hat{Y}_d^{k-1} e \hat{Y}_d^k sono i valori in cui è compresa la stima \hat{Y}_d^* , letti nelle colonne pari dei prospetti, e $\hat{\varepsilon}(\hat{X}_d^{k-1})$ e $\hat{\varepsilon}(\hat{X}_d^k)$ sono i corrispondenti errori relativi.

Prospetto 3 - Valori interpolati degli errori campionari delle stime riferite ai laureati dell'anno 1995 per l'Italia e le ripartizioni geografiche.

Perc.	RIPARTIZIONI GEOGRAFICHE											
	Nord-Ovest		Nord-Est		Centro		Sud		Isole		Italia	
	Stima valore assol.	Errore relat. perc.	Stima valore assol.	Errore relat. perc.	Stima valore assol.	Errore relat. perc.	Stima valore assol.	Errore relat. perc.	Stima valore assol.	Errore relat. perc.	Stima valore assol.	Errore relat. perc.
1	143,28	22,06	93,32	24,68	92,18	26,35	115,89	23,32	45,04	40,57	489,72	11,59
2	286,56	15,00	186,65	17,03	184,36	18,11	231,79	16,21	90,08	27,76	979,44	7,95
3	429,85	11,97	279,97	13,70	276,54	14,55	347,68	13,11	135,12	22,23	1469,17	6,38
4	573,13	10,20	373,29	11,75	368,73	12,45	463,58	11,27	180,16	18,99	1958,89	5,45
5	716,41	9,01	466,61	10,43	460,91	11,04	579,47	10,03	225,20	16,80	2448,61	4,83
6	859,69	8,14	559,94	9,46	553,09	10,00	695,37	9,11	270,24	15,21	2938,33	4,37
7	1002,98	7,47	653,26	8,71	645,27	9,20	811,26	8,40	315,29	13,98	3428,06	4,02
8	1146,26	6,93	746,58	8,11	737,45	8,56	927,16	7,83	360,33	12,99	3917,78	3,74
9	1289,54	6,49	839,91	7,61	829,63	8,04	1043,05	7,37	405,37	12,18	4407,50	3,51
10	1432,82	6,12	933,23	7,19	921,81	7,59	1158,95	6,97	450,41	11,50	4897,22	3,31
15	2149,23	4,89	1399,84	5,79	1382,72	6,10	1738,42	5,63	675,61	9,21	7345,83	2,65
20	2865,65	4,16	1866,46	4,96	1843,63	5,22	2317,90	4,84	900,81	7,86	9794,44	2,27
25	3582,06	3,68	2333,07	4,40	2304,53	4,63	2897,37	4,31	1126,02	6,96	12243,05	2,01
30	4298,47	3,32	2799,68	4,00	2765,44	4,19	3476,85	3,92	1351,22	6,30	14691,67	1,82
35	5014,88	3,05	3266,30	3,68	3226,35	3,86	4056,32	3,61	1576,43	5,79	17140,28	1,67
40	5731,29	2,83	3732,91	3,42	3687,25	3,59	4635,80	3,37	1801,63	5,38	19588,89	1,56
45	6447,70	2,65	4199,53	3,22	4148,16	3,37	5215,27	3,17	2026,83	5,04	22037,50	1,46
50	7164,12	2,50	4666,14	3,04	4609,07	3,18	5794,75	3,00	2252,04	4,76	24486,11	1,38

Prospetto 4- Valori interpolati degli errori campionari delle stime riferite alle laureate dell'anno 1995 per l'Italia e le ripartizioni geografiche.

Perc.	RIPARTIZIONI GEOGRAFICHE											
	Nord-Ovest		Nord-Est		Centro		Sud		Isole		Italia	
	Stima valore assol.	Errore relat. perc.	Stima valore assol.	Errore relat. perc.	Stima valore assol.	Errore relat. perc.	Stima valore assol.	Errore relat. perc.	Stima valore assol.	Errore relat. perc.	Stima valore assol.	Errore relat. perc.
1	149,53	20,11	105,47	22,16	109,50	23,19	131,20	21,96	56,04	30,48	551,74	10,29
2	299,07	13,72	210,94	15,27	219,00	15,93	262,40	15,07	112,07	21,06	1103,48	7,05
3	448,60	10,98	316,41	12,28	328,50	12,78	393,60	12,10	168,11	16,96	1655,22	5,65
4	598,13	9,37	421,88	10,52	438,00	10,94	524,80	10,35	224,14	14,55	2206,97	4,83
5	747,67	8,28	527,35	9,34	547,51	9,69	656,01	9,17	280,18	12,91	2758,71	4,28
6	897,20	7,49	632,82	8,47	657,01	8,78	787,21	8,31	336,21	11,72	3310,45	3,88
7	1046,73	6,88	738,30	7,79	766,51	8,07	918,41	7,64	392,25	10,79	3862,19	3,56
8	1196,26	6,39	843,77	7,25	876,01	7,51	1049,61	7,11	448,28	10,05	4413,93	3,31
9	1345,80	5,99	949,24	6,81	985,51	7,05	1180,81	6,67	504,32	9,44	4965,67	3,11
10	1495,33	5,66	1054,71	6,43	1095,01	6,66	1312,01	6,30	560,35	8,92	5517,41	2,93
15	2243,00	4,52	1582,06	5,18	1642,52	5,34	1968,02	5,05	840,53	7,19	8276,12	2,35
20	2990,66	3,86	2109,42	4,43	2190,02	4,57	2624,02	4,32	1120,71	6,16	11034,83	2,01
25	3738,33	3,41	2636,77	3,93	2737,53	4,05	3280,03	3,83	1400,88	5,47	13793,53	1,78
30	4485,99	3,09	3164,12	3,57	3285,03	3,67	3936,03	3,47	1681,06	4,96	16552,24	1,61
35	5233,66	2,84	3691,48	3,28	3832,54	3,37	4592,04	3,19	1961,23	4,57	19310,95	1,48
40	5981,32	2,63	4218,83	3,06	4380,04	3,14	5248,04	2,97	2241,41	4,26	22069,65	1,38
45	6728,99	2,47	4746,19	2,87	4927,55	2,94	5904,05	2,78	2521,59	4,00	24828,36	1,29
50	7476,65	2,33	5273,54	2,71	5475,06	2,78	6560,06	2,63	2801,76	3,78	27587,07	1,22

Prospetto 5- Valori interpolati degli errori campionari delle stime riferite al totale dei laureati dell'anno 1995 per l'Italia e le ripartizioni geografiche.

Perc.	RIPARTIZIONI GEOGRAFICHE											
	Nord-Ovest		Nord-Est		Centro		Sud		Isole		Italia	
	Stima valore assol.	Errore relat. perc.	Stima valore assol.	Errore relat. perc.	Stima valore assol.	Errore relat. perc.	Stima valore assol.	Errore relat. perc.	Stima valore assol.	Errore relat. perc.	Stima valore assol.	Errore relat. perc.
1	292,82	14,88	198,79	16,66	201,68	17,51	247,10	16,11	101,08	24,38	1050,97	7,77
2	585,63	10,14	397,59	11,46	403,36	12,01	494,19	11,10	202,15	16,81	2101,94	5,24
3	878,45	8,10	596,38	9,21	605,05	9,64	741,29	8,93	303,23	13,53	3152,91	4,16
4	1171,26	6,91	795,17	7,88	806,73	8,24	988,38	7,65	404,30	11,60	4203,88	3,53
5	1464,08	6,10	993,97	6,99	1008,41	7,30	1235,48	6,79	505,38	10,29	5254,85	3,11
6	1756,89	5,52	1192,76	6,33	1210,09	6,61	1482,58	6,15	606,46	9,33	6305,82	2,80
7	2049,71	5,07	1391,56	5,83	1411,78	6,08	1729,67	5,66	707,53	8,59	7356,79	2,57
8	2342,52	4,70	1590,35	5,42	1613,46	5,65	1976,77	5,27	808,61	8,00	8407,76	2,38
9	2635,34	4,41	1789,14	5,09	1815,14	5,30	2223,86	4,95	909,68	7,51	9458,73	2,23
10	2928,15	4,16	1987,94	4,81	2016,82	5,01	2470,96	4,68	1010,76	7,10	10509,70	2,10
15	4392,23	3,32	2981,90	3,86	3025,24	4,02	3706,44	3,76	1516,14	5,71	15764,55	1,67
20	5856,31	2,83	3975,87	3,31	4033,65	3,44	4941,92	3,22	2021,52	4,89	21019,40	1,41
25	7320,38	2,50	4969,84	2,93	5042,06	3,04	6177,40	2,86	2526,90	4,34	26274,25	1,25
30	8784,46	2,26	5963,81	2,66	6050,47	2,76	7412,88	2,59	3032,28	3,94	31529,10	1,12
35	10248,54	2,08	6957,78	2,44	7058,89	2,54	8648,36	2,39	3537,66	3,63	36783,95	1,03
40	11712,62	1,93	7951,74	2,27	8067,30	2,36	9883,84	2,22	4043,04	3,38	42038,80	0,95
45	13176,69	1,81	8945,71	2,13	9075,71	2,21	11119,32	2,09	4548,42	3,17	47293,65	0,89
50	14640,77	1,70	9939,68	2,02	10084,12	2,09	12354,80	1,97	5053,80	2,99	52548,50	0,84

Prospetto 6 - 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 laureati dell'anno 1995 distinti per sesso e in totale per gruppi di corsi

DOMINIO DI STIMA: Gruppi di corsi per sesso		a	b	R^2
Gruppo scientifico	M	1,694	-1,082	95,20
	F	0,691	-0,968	96,08
	<i>Totale</i>	0,603	-0,925	93,63
Gruppo chimico-farmaceutico	M	1,736	-1,120	95,28
	F	1,982	-1,138	98,00
	<i>Totale</i>	2,071	-1,145	98,14
Gruppo geo-biologico	M	1,431	-1,069	98,13
	F	1,644	-1,087	98,19
	<i>Totale</i>	1,447	-1,057	97,52
Gruppo medico	M	2,103	-1,068	97,79
	F	2,269	-1,131	98,99
	<i>Totale</i>	2,098	-1,075	98,28
Gruppo ingegneria	M	2,220	-1,122	97,42
	F	0,755	-1,089	95,79
	<i>Totale</i>	1,949	-1,093	97,11
Gruppo architettura	M	2,610	-1,133	98,60
	F	2,610	-1,144	98,92
	<i>Totale</i>	2,596	-1,122	98,91
Gruppo agrario	M	1,672	-1,072	96,02
	F	1,313	-1,099	97,41
	<i>Totale</i>	1,572	-1,070	96,97
Gruppo economico-statistico	M	2,725	-1,140	98,40
	F	2,240	-1,108	98,69
	<i>Totale</i>	2,670	-1,135	98,65
Gruppo politico-sociale	M	2,529	-1,134	98,23
	F	2,608	-1,180	98,85
	<i>Totale</i>	2,705	-1,161	98,56
Gruppo giuridico	M	2,404	-1,107	99,31
	F	2,517	-1,096	99,04
	<i>Totale</i>	2,520	-1,098	99,25
Gruppo letterario	M	1,958	-1,109	98,42
	F	1,789	-1,073	97,77
	<i>Totale</i>	2,016	-1,100	98,61
Gruppo linguistico	M	2,574	-1,174	97,98
	F	2,249	-1,131	98,81
	<i>Totale</i>	2,446	-1,150	98,93
Gruppo insegnamento	M	2,289	-1,159	96,25
	F	2,062	-1,117	98,39
	<i>Totale</i>	2,183	-1,127	98,10
Gruppo psicologico	M	2,435	-1,190	97,16
	F	2,218	-1,096	97,79
	<i>Totale</i>	2,351	-1,117	98,31

Appendice E:

Metodologia per la stima del Rischio di Violazione

Per la protezione dei dati relativi all'indagine sull'*Inserimento professionale dei laureati del 1995*, l'Istat ha adottato una nuova metodologia basata su una differente definizione di rischio di identificazione. Tale approccio prevede la stima di un rischio per ogni singolo individuo coinvolto nell'indagine che si basa sulla combinazione di modalità che esso assume nell'insieme delle variabili pubbliche (o chiave). In questo modo è possibile tenere conto delle caratteristiche specifiche di ogni singolo individuo, oltre che della natura delle variabili rilevate, per la produzione di *file standard*.

Avendo definito un rischio a livello individuale è possibile adottare una tecnica di protezione maggiormente mirata al controllo di quei record (o 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". Con la soppressione locale i valori di una variabile vengono sostituiti, in una o più unità, con un valore mancante. In particolare, stimato il rischio di identificazione per tutti gli individui coinvolti nell'indagine e fissata una soglia di accettazione del rischio, la tecnica di "soppressione locale" è stata applicata a tutti quei record il cui rischio supera detta soglia.

Ciò ha permesso da un lato una maggiore tutela della riservatezza delle unità effettivamente più a rischio e dall'altro il rilascio di maggiore contenuto informativo per le altre unità.

Rispetto al precedente file standard tutte le variabili, ad eccezione del "sesso" sono rilasciate con un dettaglio notevolmente maggiore.

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

Tabella 1: *Distribuzione dei record per numero di soppressioni*

Record che presentano s soppressioni (percentuale sul totale=17326)		
$s=0$	$s=1$	$s=2$
9.518 (55%)	7557 (44%)	251 (1%)

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

Variabili	Corso di laurea	Sede Università	Sesso	Residenza	Età
Numero di soppressioni (percentuale sul totale=17.326)	533 (3%)	1.156 (7%)	2.990 (17%)	597 (3%)	2783 (16%)