

PRESIDENZA DEL CONSIGLIO DEI MINISTRI

COMMISSIONE PER LA GARANZIA DELL'INFORMAZIONE STATISTICA

**ANALISI DELLE PROCEDURE DI CORREZIONE/IMPUTAZIONE
UTILIZZATE DALL'ISTAT
NELLE PRINCIPALI INDAGINI SULLE FAMIGLIE**

VOLUME II

Rapporto di Ricerca

00.03

Gennaio 2000

a cura di:

Daniela Cotzia

Maria Elena Graziani

Cristina Panattoni

La Commissione per la garanzia dell'informazione statistica (CGIS), istituita presso la Presidenza del Consiglio dei Ministri con il decreto legislativo n. 322 del 1989, art.12, è un organo collegiale indipendente chiamato a garantire il principio della imparzialità e della completezza dell'informazione statistica. A tal fine la CGIS vigila: (a) sulla imparzialità e completezza dell'informazione statistica; (b) sulla tutela della riservatezza delle informazioni fornite agli enti del Sistema Statistico Nazionale; (c) sulla qualità delle metodologie statistiche e delle tecniche informatiche impiegate nella raccolta, nella conservazione e nella diffusione dei dati; (d) sulla conformità delle rilevazioni alle direttive degli organismi internazionali e comunitari.

La serie "Rapporti di ricerca" raccoglie i risultati di attività di appositi gruppi di lavoro, promossi e coordinati dalla CGIS in relazione all'adempimento dei propri compiti.

La responsabilità del contenuto del rapporto è degli autori, e non coinvolge la Commissione.

PRESIDENZA DEL CONSIGLIO DEI MINISTRI

COMMISSIONE PER LA GARANZIA DELL'INFORMAZIONE STATISTICA

Ugo Trivellato, *Presidente*

Graziella Caselli

Pierluigi Ciocca

Bruno De Leo

Antonio Golini

Vittorio Grilli

Renato Guarini

Cesare Imbriani

Luisa Torchia

Commissione per la Garanzia dell'Informazione Statistica

Via Po, n.16/A

00198 Roma

tel. ++39-6-8598.2010/8598.2132

fax ++39-6-8598.2012

INDICE

Introduzione	5
Sezione I: Indagine Multiscopo	7
a cura di Maria Elena Graziani	
1 Il trattamento delle mancate risposte nell'Indagine Multiscopo (Istat, 1993: 13-158)	7
1.1 <i>Disegno dell'indagine</i>	7
1.2 <i>Controllo e correzione dei dati: aspetti generali</i>	8
1.3 <i>Controllo e correzione dei dati: progettazione logico fisica dell'indagine</i>	9
2 L'Indagine sulla vittimizzazione: un esperimento di intervista computer assisted. Il controllo e la correzione dei dati	14
2.1 <i>L'Indagine sulla vittimizzazione</i>	14
2.2 <i>Aspetti metodologici generali</i>	15
2.3 <i>Il computer nella raccolta dei dati (W.E. Saris, 1991)</i>	18
3 Applicazione del questionario on line all'Indagine Multiscopo sulle famiglie: una simulazione e una semplificazione per algoritmi	18
3.1 <i>Premessa</i>	18
3.2 <i>La simulazione</i>	19
3.3 <i>Gli algoritmi</i>	31
3.4 <i>Conclusioni</i>	34
Sezione II: Indagine sulle Forze di Lavoro	35
a cura di Cristina Panattoni	
4 Cenni preliminari sull'indagine	35
5 Come si struttura il controllo e la correzione dei dati sulle Forze di Lavoro	35
6 Il nuovo piano di incompatibilità per la RTFL	40
6.1 <i>Correzione deterministica</i>	40
6.2 <i>Correzione probabilistica</i>	42
6.3 <i>Principali risultati</i>	46

7	Piano di controllo e correzione dell'Indagine sulle Forze di Lavoro: quali prospettive?	47
8	Alcuni spunti propositivi sul controllo on line dei dati della RTFL per grandi aree tematiche	48
	Sezione III:Indagine sui Consumi delle Famiglie	56
	a cura di Daniela Cotzia	
9	Caratteristiche dell'indagine	56
9.1	<i>Il disegno di campionamento e di rilevazione dei dati</i>	56
9.2	<i>Tecnica di rilevazione</i>	58
9.3	<i>Indagine sperimentale con sistema CATI</i>	59
10	Il fenomeno delle mancate risposte	60
10.1	<i>Le mancate risposte totali di primo stadio (UPS)</i>	61
10.2	<i>Le mancate risposte totali di secondo stadio (USS)</i>	61
11	La revisione e correzione dei dati nell'Indagine sui Consumi delle Famiglie	72
12	Regole di revisione manuale utilizzate dai revisori dell'Istat	73
13	Acquisizione controllata dei dati - CADI	74
13.1	<i>Il sistema BLAISE</i>	75
13.2	<i>RULES impostate in BLAISE nell'indagine sui Consumi delle Famiglie</i>	78
13.2.1	<i>Riepilogo delle spese (Mod. BF/2)</i>	78
13.2.2	<i>Libretto degli acquisti (Mod. BF/1)</i>	78
13.2.3	<i>Taccuino degli autoconsumi (Mod. BF/1A)</i>	80
14	Controllo e correzione dei dati	80
14.1	<i>Correzione delle variabili qualitative</i>	81
14.1.1	<i>Descrizione del Sistema di Correzione ed Imputazione Automatica (SCIA)</i>	81
14.1.2	<i>La revisione delle variabili qualitative nell'Indagine sui Consumi delle Famiglie</i>	84
14.2	<i>Correzione delle variabili quantitative</i>	88
14.2.1	<i>Descrizione del Sistema di Ricostruzione delle Informazioni con Donazione Automatica (RIDA)</i>	88
14.2.2	<i>La revisione delle variabili quantitative nell'Indagine sui Consumi delle Famiglie</i>	92
	Allegato	96
	Bibliografia	112

Introduzione

Il progetto di ricerca relativo alla “*Analisi delle procedure di Correzione/Imputazione utilizzate dall’ISTAT nelle principali Indagini sulle Famiglie*” è indirizzato all’esame delle metodologie adottate dall’ISTAT per il trattamento dei dati mancanti o inammissibili nelle principali indagini sulle famiglie.

In linea con gli obiettivi prefissati, il presente rapporto (volume II), articolato in tre sezioni, documenta il lavoro di indagine empirica condotto dal gruppo di ricerca direttamente presso i Servizi dell’ISTAT.

Ogni sezione è dedicata all’esame di una delle tre indagini campionarie condotte dall’ISTAT sulle famiglie (*Indagine Multiscopo, Indagine sulle Forze di Lavoro, Indagine sui Consumi delle Famiglie*), specificando la logica dei controlli e delle correzioni effettuate sui dati errati o mancanti.

Il trattamento delle mancate risposte nell’Indagine Multiscopo viene articolato in tre momenti successivi.

Dapprima, si analizzano, all’interno di un sintetico esame dell’impianto dell’indagine, le modalità di trattamento delle mancate risposte; successivamente, ci si sofferma sull’*Indagine sulla sicurezza dei cittadini* condotta nell’ambito delle Multiscopo. In questa indagine l’ausilio del computer nell’inserimento dei dati in tempo reale ha rappresentato una vera novità che ha enormi potenzialità nel trattamento dei dati errati o mancanti.

Infine, si conduce una simulazione di intervista *computer assisted* applicata alla sezione generale dell’Indagine Multiscopo, approfondendo, in particolare, le modalità di correzione. Si propongono delle strutture logiche che permettano controlli incrociati tra le risposte date a differenti domande e evidenzino le incoerenze o le combinazioni di risposte troppo contraddittorie.

Per l’Indagine sulle Forze di Lavoro, vengono esaminate le attuali strategie di controllo e correzione dei dati utilizzate dal Servizio FIL\C, con specifica attenzione al piano di incompatibilità, alle tecniche di correzione e alle procedure informatiche utilizzate a partire dalla revisione di aprile 1999.

Si descrivono i miglioramenti ottenuti a seguito di tali modifiche, in termini di qualità dei dati e di affidabilità dei risultati.

Dall'indagine effettuata è emerso che i dati delle rilevazioni a cadenza trimestrale sulle Forze di Lavoro sono utilizzati da molti anni esclusivamente in un'ottica "trasversale". L'informazione aggiuntiva risultante dall'intervistare gli stessi individui in differenti occasioni è quindi in parte "sprecata".

In particolare, nell'ambito degli obiettivi di lavoro attribuiti a questo gruppo di ricerca non emerge da parte dell'ISTAT alcun impiego dell'informazione longitudinale ai fini del controllo e della correzione/imputazione dei dati sulle Forze di Lavoro.

Infine, si offrono spunti di riflessione propositivi sulla possibilità di porre in essere controlli *on line* in un'ottica longitudinale.

Per quanto attiene all'indagine sui Consumi delle Famiglie, si approfondisce il fenomeno delle mancate risposte analizzando sia le cadute di primo stadio, relative al rifiuto dei Comuni campione di partecipare all'indagine, sia le cadute di secondo stadio, dovute all'impossibilità di intervistare le famiglie selezionate e alla scarsa qualità delle loro risposte.

Vengono, inoltre, approfondite le tecniche relative all'immissione controllata dei dati per mezzo del programma BLAISE e alla verifica e correzione delle variabili qualitative e quantitative analizzando le modalità di utilizzo rispettivamente dei programmi SCIA e RIDA e la metodologia dei Massimi e minimi randomizzati.

E' doveroso ringraziare i responsabili delle tre indagini e i loro collaboratori – la dott.ssa Linda Laura Sabbadini e la dott.ssa Luciana Quattrococchi per la Multiscopo, il dott. Pietro Gennari e il dott. Carlo Lucarelli per l'indagine Forze di Lavoro, la dott.ssa Giuliana Coccia e le dott.sse Nicoletta Pannuzzi e Donatella Grasso per l'indagine sui Consumi - per la disponibilità accordata e per aver fornito tutta la documentazione e il materiale possibili ai fini del presente lavoro.

Sezione I: Indagine Multiscopo

A cura di Maria Elena Graziani

1 Il trattamento delle mancate risposte nell'Indagine Multiscopo (Istat, 1993: 13-158)

1.1 Disegno dell'indagine

L'Indagine Multiscopo sulle famiglie è basata su un campione mensile di circa 4.000 famiglie in 207 comuni campione e la raccolta delle informazioni avviene mediante intervista “faccia a faccia” presso l’abitazione della famiglia. Il campione è costruito mediante estrazione casuale dall’anagrafe della popolazione residente di ciascun comune campione; tale estrazione individua una lista di indirizzi e di corrispondenti nomi di “intestatari” delle schede di famiglia: la rilevazione riguarda tutti i componenti familiari delle famiglie campione, considerate interamente nella loro composizione “di fatto”.

L’indagine è ciclica, con un progetto che vede, complessivamente, sei cicli di rilevazione, ciascuno con durata di sei mesi. Il campione relativo a ciascun ciclo dell’indagine è costituito da circa 23.000 famiglie in 1.132 comuni campione.

I questionari utilizzati per le interviste sono articolati in due tipologie di quesiti:

⇒ i quesiti comuni, che sono ogni mese oggetto di intervista anche se su rispondenti diversi, riguardano i dati individuali e familiari di carattere strutturale; queste informazioni sono necessarie per analizzare tutti i fenomeni socio-demografici considerati nell’indagine;

⇒ i quesiti specifici sono relativi ad aspetti socio-demografici oggetto di rilevazione a cadenza periodica. In particolare:

- il primo ciclo si caratterizza per la rilevazione di eventi ‘negativi’ (incidenti in ambienti domestici e vittimizzazione da atti criminosi) e di aspetti connessi all’impiego ‘culturale’ del tempo libero (letture e ascolto di radio e televisione), nonché l’uso di lingua italiana, dialetti e lingue straniere;

- nel secondo ciclo si indaga l'attività sportiva, l'impiego del tempo, la storia coniugale-riproduttiva e i cambiamenti di abitazione;
- nel terzo, oltre che all'uso del tempo (già rilevato nel precedente), vengono analizzati i viaggi, la partecipazione all'attività scolastica e la condizione dell'infanzia;
- il quarto ciclo si caratterizza come un ciclo dedicato alla costruzione di informazioni sulla vita degli anziani, sui disabili, sullo stato di salute e sull'uso di servizi socio-sanitari;
- questi ultimi aspetti sono oggetto di rilevazione anche nel quinto ciclo che è, infatti, dedicato al tema della salute;
- il sesto ed ultimo riguarda essenzialmente la famiglia come struttura in cui si integrano le storie di vita e le attività individuali.

1.2 Controllo e correzione dei dati: aspetti generali

Ciascun ciclo rappresenta una indagine, con una propria fase preparatoria, una fase di lavoro sul campo e una fase di elaborazione. E' nella fase di elaborazione che si effettua il controllo e la correzione dei dati.

Il primo controllo avviene già durante il trasferimento delle informazioni dal supporto cartaceo a quello informatico e consiste nell'applicazione di una procedura di registrazione controllata, in cui i controlli sono estremamente ridotti e concentrati su dati 'indicativi' dei questionari stessi (mese di rilevazione, codici di provincia e comune, numero d'ordine della famiglia).

Gli interventi di correzione vera e propria vengono, però, rinviati ad una fase successiva e sono di due tipi: interattivi e automatici.

⇒ Le correzioni interattive vengono fatte manualmente a terminale, dietro segnalazione di errore effettuata tramite procedura informatica e sono concentrati sulle notizie di tipo anagrafico (sesso, anno di nascita e stato civile) e su quelle riguardanti la struttura familiare: per molti di tali aspetti risulta di notevole aiuto la disponibilità degli stati di famiglia anagrafici delle famiglie campione.

⇒ La procedura automatica, che segue la fase della correzione interattiva, è diversa a seconda del tipo di variabile:

- per il grado di istruzione, la condizione, la posizione nella professione degli occupati e il titolo di godimento dell'abitazione, il valore errato o mancante viene sostituito da un valore compatibile calcolato utilizzando le informazioni corrette desunte dalla stessa rilevazione;
- per le altre variabili, il dato errato viene trasformato in risposta mancante e resta tale anche sul record finale.

1.3 Controllo e correzione dei dati: progettazione logico fisica dell'indagine

Ognuno dei sei cicli di rilevazione in cui è suddivisa l'Indagine Multiscopo è divisibile in 15 passi. Nel seguito del paragrafo verranno elencati e analizzati dando particolar rilievo a quelli relativi, o in qualche modo connessi, alla correzione dei dati. Si farà riferimento ad un generico ciclo di rilevazione perché lo schema della procedura è prevalentemente lo stesso, cambiano di volta in volta i programmi che vengono attivati..

- 1) *Predisposizione del piano di spedizione e produzione automatica di lettere circolari, etichette e prospetti di spunta.*
- 2) *Predisposizione dell'archivio di controllo e documentazione e di elaborati sugli errori rilevati.* Ai fini del controllo e della gestione dell'indagine, è stato predisposto un file che contiene, oltre che alle notizie sulla struttura del campione, anche una serie di informazioni che vengono inserite dalla sezione responsabile dell'indagine durante lo svolgimento della rilevazione e che riguardano gli arrivi, la revisione i pagamenti, l'invio in magazzino e l'invio in registrazione. Questo file si chiama "FAMB DATA".
- 3) *Registrazione dei dati statistici elementari.* I dati vengono raccolti e registrati mensilmente a cura di una ditta esterna su supporto magnetico.
- 4) *Produzione del piano di spoglio e produzione delle relative posizioni e tavole di verifica.* I nastri provenienti dalla registrazione esterna sono sottoposti ad un primo

controllo sia della qualità della registrazione che del contenuto del nastro e ad un primo trattamento.

- 5) *Controllo della scheda sulla struttura familiare e delle interviste non effettuate e predisposizione dei pannelli per le correzioni interattive.* Questa fase è piuttosto complessa e comprende anche il controllo e la correzione delle schede relative alle mancate interviste (mancate risposte totali).

Alla base di questo controllo vi è:

- ⇒ il confronto tra la scheda generale (scheda 1) e le mancate interviste (scheda 6) con riferimento al numero teorico di famiglie campione;
- ⇒ il confronto tra la famiglia di fatto e quella anagrafica;
- ⇒ la definizione precisa della famiglia di fatto, degli individui che la compongono e della loro composizione familiare secondo una classificazione per nuclei.

Le variabili che vengono controllate sono tutte quelle presenti nella scheda generale del questionario.

Il primo input è costituito dal file di dati grezzi comprendenti tutti i tipi di scheda; da questo vengono estratte le schede 1 (SCHEDE1 DATA) e le schede 6 (SCHEDE6 DATA) e ordinate per dati identificativi della famiglia. Il secondo input è costituito dal file "FAMB DATA" da cui viene prelevato il numero teorico di famiglie campione.

Il check è strutturato in 3 fasi:

- ⇒ nella prima viene effettuato il controllo tra il numero teorico delle famiglie campione e il numero delle schede 6;
- ⇒ nella seconda fase vengono analizzate contemporaneamente le schede 1 e le schede 6 e si segnalano gli eventuali doppi e le schede mancanti e viene costituito l'archivio "ERRORI DATA";
- ⇒ nella terza fase viene realizzato il controllo capillare della scheda 1 (oltre 500 controlli di compatibilità tra le variabili).

Il risultato del programma di check è costituito da un file di record esatti, un file di record errati, un listing delle schede errate con indicazione dei codici di errore, il file "ERRORI DATA" aggiornato.

Il processo è iterativo e si ripete fino alla rimozione di tutti gli errori.

Terminata la fase di correzione, il file “SCHEDE6 DATA” viene definitivamente archiviato mentre il file “SCHEDE1 DATA” viene integrato con una serie di variabili costruite sulla base di quelle presenti.

- 6) *Determinazione dei coefficienti di riporto all’universo e creazione di variabili relative alla scheda sulla struttura familiare.* Terminata la correzione della scheda generale, è possibile attivare la procedura di calcolo dei coefficienti di riporto all’universo. Il coefficiente viene riportato sul file “SCHEDE1 DATA”.
- 7) *Agganci tra la scheda sulla struttura familiare e le schede specifiche generate dal questionario per la ricostruzione della famiglia di fatto e predisposizione dei pannelli per le correzioni interattive.* Viene a questo punto ricomposto il file complessivo dei dati rilevati (SCHEDEALL DATA) comprendente il tipo-record 1 pulito e gli altri tipo-record ancora grezzi.

A questo punto è necessario ricostruire esattamente la famiglia di fatto: per ogni famiglia deve essere presente un tipo-record 1, tanti tipi-record 2 quanti sono i componenti della famiglia di fatto, un tipo-record 3 e tipi-record 4 e 5 se il ciclo di rilevazione lo richiede. Tale ricostruzione è strutturata in modo tale da prevedere una serie di correzioni automatiche ed una serie di correzioni interattive, è interamente realizzata in COBOL ed è basata sul confronto tra le variabili principali.

Un primo programma esegue la fase di aggancio e produce in output il file “record esatti”, il file “record errati”, il file “ERRORI DATA” con i codici di errore riscontrati, un tabulato con l’illustrazione delle forzature automatiche e un tabulato con la descrizione dei record errati e le segnalazioni di errore.

Un secondo programma predispone dei pannelli ad hoc per consentire l’effettuazione delle correzioni interattive da parte dell’utente. Il processo si ripete fino alla completa eliminazione degli errori.

- 8) *Definizione del file “ERRORI DATA”.* Terminate le correzioni interattive, prima di passare alle fasi successive, completamente automatiche, è necessario effettuare la definizione finale del file “ERRORI DATA” che comprende, per ciascun tipo-scheda

che ha dato luogo ad almeno un errore, uno o più record contenenti le segnalazioni di errore per codice di errore.

9) *Trasferimento delle variabili create sulla scheda generale e predisposizione del file "DATI COMUNI"*. In questa fase viene effettuato il trasferimento delle variabili contenute sulla scheda generale (SCHEDE1 DATA) sugli altri tipi di record e la costruzione del file dei dati comuni (SCHEDE0 DATA). In particolare, il file dei dati comuni è costituito da tanti records per quanti sono gli individui rilevati e riporta inizialmente tutte le notizie relative alla parte comune dell'Indagine Multiscopo. Nel file dei dati comuni vengono anche riportate le variabili create nella scheda generale, i coefficienti di riporto all'universo e vengono calcolati i coefficienti annuali in modo da permettere delle elaborazioni su due cicli di indagine.

10) *Controllo e correzione automatica del file "DATI COMUNI"*. La procedura di controllo e correzione dei dati comuni si basa sulla individuazione di tre categorie di errore:

- ⇒ errori di *range* riguardanti il dominio delle possibili risposte (ad esempio per il sesso M=1 e F=2, il dominio è 1,2);
- ⇒ errori di rotta riguardanti la corretta somministrazione delle domande presenti nel questionario (ad esempio il gruppo di domande relative alla condizione lavorativa deve essere significativo solo per le persone in età di 14 anni e oltre);
- ⇒ errori di consistenza riguardanti le compatibilità tra le varie risposte (ad esempio una persona di sesso femminile non può dichiarare di svolgere la leva).

Gli errori generano correzioni esclusivamente automatiche di due tipi: deterministiche e probabilistiche (cioè basate sulle distribuzioni di frequenza delle variabili stesse nei record corretti) e le variabili che vengono corrette casualmente per le quali non sono ammesse le mancate risposte sono: il grado di istruzione, la condizione professionale, la posizione nella professione e il titolo di godimento dell'abitazione.

In questo caso è stato necessario creare delle tabelle temporanee riportanti la distribuzione di frequenza di ciascuna variabile determinata sull'insieme "record esatti"; in secondo luogo i record da correggere sono stati ordinati secondo la data di

nascita (giorno, mese, anno). L'ordinamento così predisposto ha permesso l'assegnazione casuale dei codici sulla base delle tabelle precedentemente calcolate.

La fase di check dei dati comuni è piuttosto complessa ed ha richiesto lo sviluppo di 11 programmi che individuano e correggono circa 100 tipi di errori.

L'input è costituito da due file: quello dei dati comuni "SCHEDE0 DATA" e quello degli errori "ERRORI DATA", definito nella fase 8.

- ⇒ Il primo programma effettua il calcolo dell'età in anni compiuti, la ricodifica della "cittadinanza" ed il check della data di nascita;
- ⇒ il secondo programma effettua la determinazione della variabile "componente più anziano della famiglia";
- ⇒ nel terzo viene predisposta la tabella per la correzione della variabile "grado di istruzione" che viene poi corretta integralmente nei due programmi successivi (il quarto e il quinto);
- ⇒ il sesto predispone la tabella per il controllo della variabile "condizione professionale" che viene corretta nei due programmi che seguono (il settimo e l'ottavo) in base anche ai valori del sesso, dell'età e del grado di istruzione;
- ⇒ inoltre nell'ottavo programma si costruisce anche la tabella per la correzione della variabile "posizione nella professione" che viene corretta nel nono programma;
- ⇒ il decimo programma effettua il check e la correzione del ramo di attività economica e della ricerca di lavoro in rapporto all'età e alla condizione professionale, delle fonti di reddito e della condizione di salute in relazione al sesso e all'età. Inoltre viene predisposta la tabella per il controllo della variabile "titolo di godimento dell'abitazione";
- ⇒ l'ultimo programma effettua la correzione dei dati relativi all'abitazione.

In ogni passo viene aggiornato il file "ERRORI DATA".

- 11) *Creazione di variabili relative al file "DATI COMUNI"*. Ultimata la fase di check dei dati comuni, il file "SCHEDE0 DATA" viene ulteriormente integrato: un programma crea una serie di variabili nuove sulla base di quelle esistenti che verranno poi utilizzate in fase di tabulazione.

Con questa fase termina la procedura di trattamento dei dati comuni. Il file "SCHEDE0 DATA" così strutturato viene archiviato e costituisce l'input definitivo per tutte le elaborazioni sui dati comuni.

- 12) *Controllo e correzioni automatica dei files specifici di ciascun ciclo.* Prima di procedere all'effettuazione dei controlli sui dati specifici di ciascun ciclo, è necessario effettuare il trasferimento dei dati comuni sottoposti a check e registrati nel file "SCHEDE0 DATA" sugli altri tipi-records al fine di omogeneizzare gli archivi.

Inizia così la fase di controllo e correzione delle notizie specifiche rilevate in ciascun ciclo.

- 13) *Creazione di variabili relative ai files specifici.*

- 14) *Definizione degli archivi definitivi di ciascun ciclo.*

- 15) *Tabulazione comprendente la predisposizione delle tavole relative al Notiziario sui dati comuni, al Volume sui dati comuni, ai Notiziari tematici per ciascun ciclo, ai Volumi tematici per ciascun ciclo.*

2 *L'Indagine sulla vittimizzazione: un esperimento di intervista computer assisted. Il controllo e la correzione dei dati*

2.1 *L'Indagine sulla vittimizzazione*

Tra settembre '97 e gennaio '98 l'Istat ha condotto, all'interno di un ciclo di indagini multiscopo presso le famiglie, un'indagine telefonica: la "Indagine sulla sicurezza dei cittadini". L'impostazione data a questa ricerca sulla 'vittimizzazione' rientra in un nuovo campo della criminologia, chiamato, appunto, 'vittimologia'. Questa disciplina, ha assunto una ragguardevole importanza negli Stati Uniti negli anni '60 ma altrove si è diffusa solo più recentemente (G. Muratore, *D.I.*).

Con l'introduzione del concetto di vittima e, soprattutto, dell'analisi dei suoi racconti si apre la possibilità di leggere il crimine dal punto di vista nuovo di chi è stato vittimizzato/a, cioè di chi ha subito un crimine, in relazione alla natura dello stesso, alla sua gravità, alla relazione con l'offensore e alle sue caratteristiche.

Un'indagine che assume l'ottica del vittimizzato/a rende possibile condurre un'analisi quantitativa del complesso dei fatti delittuosi anche non denunciati e, nello stesso tempo, permette di cogliere anche il vissuto sociale di una realtà, la sua incidenza nel quotidiano, i rapporti di aiuto che l'individuo è capace di attivare con altri individui e con le istituzioni, il senso civico, il costo fisico, sociale, psicologico ed economico che il crimine comporta (L. Quattrocioni, *D.I.*).

2.2 Aspetti metodologici generali

Dall'esame dell'impianto metodologico (S. Falorsi, 1997) emerge che:

⇒ La popolazione oggetto d'indagine è costituita da 'famiglie di fatto', intese come insieme di persone legate da vincoli di parentela, affinità, ecc. e dagli 'individui' residenti in famiglia aventi almeno 14 anni, al netto dei membri permanenti in convivenze. Un solo individuo per famiglia rappresenta l'unità finale di campionamento che, oltre a fornire le informazioni richieste per se stesso/a, fornisce alcune informazioni riguardanti la famiglia in cui vive.

⇒ Le principali variabili d'interesse sono:

- sugli individui: informazioni generali, dati anagrafici, tempo fuori casa, percezione della sicurezza; alcuni tipi di fatti delittuosi sia contro il patrimonio (scippi, borseggi, furto di oggetti personali, rapine) sia contro l'individuo (minacce, aggressioni, molestie e violenze sessuali); comportamento e caratteristiche delle persone che hanno subito eventi di vittimizzazione; comportamenti che tendono a prevenire il crimine;
- sulle famiglie: informazioni generali, caratteristiche strutturali della famiglia (ampiezza, composizione per sesso, età e stato civile dei componenti), abitazione, reddito; alcuni tipi di fatti delittuosi subiti dalla famiglia (furto di oggetti esterni all'abitazione, furto in abitazione, tentato furto in abitazione, ingresso abusivo, furto di veicoli, atti di vandalismo contro l'abitazione o i veicoli).

⇒ Il disegno di campionamento adottato è a due stadi con stratificazione delle unità primarie.

- Le unità primarie sono costituite dagli indirizzi telefonici presenti sull'Archivio Telefonico Informatizzato delle Famiglie della Telecom e sono state stratificate per regione geografica e tipologia di comune (caratteristica fornita dall'Istat sulla base delle caratteristiche socio-demografiche dei comuni italiani rilevate al Censimento della Popolazione e in base ai dati anagrafici). Per ciascuna unità primaria appartenente all'elenco base (50.000 numeri di telefono) sono selezionate tre unità primarie aggiuntive che entrano a far parte dell'elenco sostitutivo (15.000).
- Le unità secondarie sono costituite dagli individui: per ciascuna famiglia selezionata al primo stadio si estrae un individuo campione tra i componenti della famiglia aventi almeno 14 anni.

⇒ Per la selezione del campione si ordinano tutti gli indirizzi dell'Archivio Telefonico, per regione, secondo l'ampiezza demografica del comune, la strada, il numero civico e il piano; si definisce il passo di estrazione regionale e seleziona 6 punti di partenza, uno per ciascuna tipologia comunale. Si procede, poi, alla selezione degli indirizzi campione sulla base del passo regionale; per ogni indirizzo selezionato, sono considerati anche i tre indirizzi successivi o precedenti che entrano a far parte dell'elenco sostitutivo, garantendo, così la selezione di unità sostitutive simili in termini di prossimità territoriale. Quando all'indirizzo selezionato risulta risiedere una famiglia, si procede alla selezione casuale del componente da intervistare: si chiede alla persona che ha risposto al telefono l'elenco dei componenti della famiglia, il sesso, l'età e la relazione con il capo famiglia; si forma la graduatoria per età dei componenti assegnando un numero progressivo ai componenti intervistabili (età maggiore o uguale a 14 anni); si estrae un numero casuale compreso tra 1 e il numero totale dei componenti intervistabili; si intervista il componente estratto.

⇒ La lista telefonica, ovviamente, non copre tutta la popolazione oggetto di indagine ma restano escluse alcune sottopopolazioni senza telefono. Queste possono avere caratteristiche socio-demografiche differenti dalle famiglie con telefono; ciò comporta una distorsione delle stime dovuta sia all'incidenza della sottocopertura della lista, sia alla differenza tra le caratteristiche delle famiglie con telefono e di quelle senza (o con telefono ma non riportate sugli elenchi).

⇒ Per ridurre gli effetti distorsivi dovuti alla mancanza risposta totale si è cercato di intervistare le unità appartenenti all'elenco base e a tal fine sono stati previsti, per ciascun indirizzo, cinque tentativi, in giorni e orari differenti, prima di passare alla sostituzione con un indirizzo dell'elenco sostitutivo. Poiché la regola generale è quella di non far cadere una famiglia contattata, se l'individuo selezionato non si dichiara disposto a rispondere, si sostituisce all'interno della stessa famiglia, con una regola di tipo non casuale che tende a recuperare le sottoclassi di individui che maggiormente si sottraggono all'intervista (ad esempio i maschi giovani).

Lo strumento utilizzato per l'acquisizione dei dati è l'intervista telefonica che rappresenta una novità rispetto alle altre indagini dell'Istat negli studi sulla popolazione (L. Quattrocchi, *D.I.*). La scelta dell'intervista telefonica è stata fatta considerati i notevoli vantaggi che tale strumento consente, specie se utilizzato congiuntamente ad un sistema interattivo computerizzato per assistere le rilevatrici nella raccolta delle informazioni.

L'utilizzo del telefono presenta, però, uno svantaggio: nonostante la rete telefonica sia sufficientemente diffusa in maniera omogenea sul territorio nazionale, rimane il problema della distorsione dovuta proprio alla copertura inferiore al 100%.

Se si indica con P_t e P_{nt} rispettivamente la parte di popolazione con il telefono e senza il telefono, supponendo di voler effettuare inferenze sul carattere quantitativo X e, in particolare sul valor medio $E(X) = \mu$, si ottiene:

$$\begin{aligned} \mu &= 1/P (\mu_t P_t + \mu_{nt} P_{nt}) = \frac{P_t}{P} \mu_t + \frac{P_{nt}}{P} \mu_{nt} = 1/P [\mu_t (P - P_{nt}) + (\mu_{nt} P_{nt})] = \\ &= 1/P (\mu_t P - \mu_t P_{nt} + \mu_{nt} P_{nt}) = \mu_t P/P + (\mu_{nt} - \mu_t) (P_{nt}/P) = \mu_t + (\mu_{nt} - \mu_t) (P_{nt}/P) \end{aligned}$$

La stima di μ , quindi, è dovuta a due fattori:

$(\mu_{nt} - \mu_t) \rightarrow$ la differenza tra i valori medi della variabile d'interesse nelle due sottopopolazioni;

$(P_{nt}/P) \rightarrow$ il tasso di mancata copertura telefonica;

se una delle due componenti è prossima ad essere nulla, sarà trascurabile anche l'errore di copertura.

2.3 *Il computer nella raccolta dei dati (W.E. Saris, 1991)*

La vera novità dell'impianto metodologico di questa indagine è l'introduzione dei computer nella raccolta dei dati, cioè l'uso del CATI *computer-assisted telephone interviewing*¹. L'assistenza del computer permette di elevare la qualità dei dati in fase di acquisizione grazie all'uso di un software di gestione delle interviste telefoniche che garantisce l'acquisizione controllata dei dati secondo schemi e regole precedentemente predisposti volte ad assicurare la coerenza delle risposte ottenute.

Ciò rappresenta, senza dubbio, un motore importantissimo della rivoluzione che sta, in questi anni, attraversando il terreno della misura dei fenomeni perché riduce enormemente i tempi e i costi delle indagini: il programma informatico, infatti, mostrando sullo schermo del computer le domande e permettendo di digitare direttamente le risposte degli intervistati, riunifica in un solo passo la stampa dei questionari, la spedizione, la codifica e l'immissione dei dati.

3 *Applicazione del questionario on line all'Indagine Multiscopo sulle famiglie: una simulazione e una semplificazione per algoritmi*

3.1 *Premessa*

In questo capitolo si presenta una simulazione di piano di editing facendo riferimento all'Indagine Multiscopo e si analizzano i passaggi necessari alla trasformazione del questionario cartaceo in questionario *on line*.

Le domande a cui si fa riferimento sono quelle iniziali sulla famiglia e, cioè, quelle raccolte nella 'scheda generale'. Si tratta di quesiti che consentono l'analisi delle cosiddette 'informazioni di base' e possono essere considerate tutte non gerarchiche; lo studio delle relazioni tra le variabili è, quindi, piuttosto complesso, poiché non è possibile sfruttare la relazione causa-effetto propria dei rapporti gerarchici o di dipendenza.

¹ Per approfondimenti si veda il primo volume del rapporto.

Nel seguito si procede analizzando un esempio di applicazione riferiti all'Indagine e poi gli algoritmi, di volta in volta più articolati; la norma che regola tutta la simulazione è la seguente: si pongono tutte le domande una ad una, dopo ogni risposta fornita il computer esegue il controllo di coerenza con le risposte precedenti, se l'*edit* non si attiva sullo schermo compare la domanda successiva, altrimenti compaiono le domande in contraddizione da correggere.

E' stata prevista la possibilità di aree di 'allarme', cioè *range* di risposte che sono al margine tra l'incompatibilità e la compatibilità: in questi casi il computer dà l'allarme ma non obbliga a correggere i le risposte immesse.

3.2 *La simulazione*

Tra tutte le domande poste nel questionario, alcune sono vincolate tra loro da legami logici di coerenza; si può imporre, quindi, che queste domande attivino dei controlli su altre variabili a cui si è risposto in precedenza. In base alla risposta data (cioè in base alla modalità scelta), ogni controllo può portare a delle correzioni.

Supponiamo di dividere il questionario in 'batterie' di domande: una batteria finisce con la prima domanda che attiva un controllo:

DOMANDA 1	DOMANDA 2	DOMANDA i	DOMANDA che attiva i controlli
------------------	------------------	------	------------------	------	---------------------------------------

Sviluppando il questionario, cioè ponendo le domande nell'ordine, se una domanda attiva un controllo, ci si chiede:

- ⇒ se la risposta data entra in contraddizione con altre e quindi va effettuata una correzione;
- ⇒ in tal caso, in quale batteria sta la variabile da modificare.

Nell'Indagine che stiamo considerando le batterie di domande sono le seguenti 9:

1) N° d'ordine del componente Posizione con riferimento all'indirizzo della famiglia anagrafica
--

Relazione di parentela o di convivenza con la persona di riferimento del questionario

Sesso

Anno di nascita → attiva 2 controlli (con 'anno attuale' e 'relazione di parentela')

2) Stato civile alla data della rilevazione → attiva 2 controlli (con 'relazione di parentela' e 'anno di nascita')

3) Titolo di studio più alto conseguito alla data della rilevazione → attiva 1 controllo (con 'anno di nascita')

4) Condizione unica o prevalente alla data della rilevazione → attiva 2 controlli (con 'sesso' e 'anno di nascita')

5) Posizione nella professione alla data della rilevazione → attiva 2 controlli (con 'condizione economica' e 'anno di nascita')

6) Attività economica alla data della rilevazione → attiva 1 controllo (con 'condizione economica')

7) Fonte principale di reddito alla data della rilevazione → attiva 3 controlli (con 'anno di nascita', 'posizione nella professione' e 'condizione')

Stato civile alla stessa data dell'anno precedente

Titolo di studio più alto conseguito alla stessa data dell'anno precedente

Condizione unica o prevalente alla stessa data dell'anno precedente

Posizione nella professione alla stessa data dell'anno precedente

Attività economica alla stessa data dell'anno precedente

queste variabili nell'esempio non vengono considerate

8) Anno del matrimonio attuale → attiva 3 controlli (con 'stato civile', 'anno di nascita' e 'relazione di parentela')

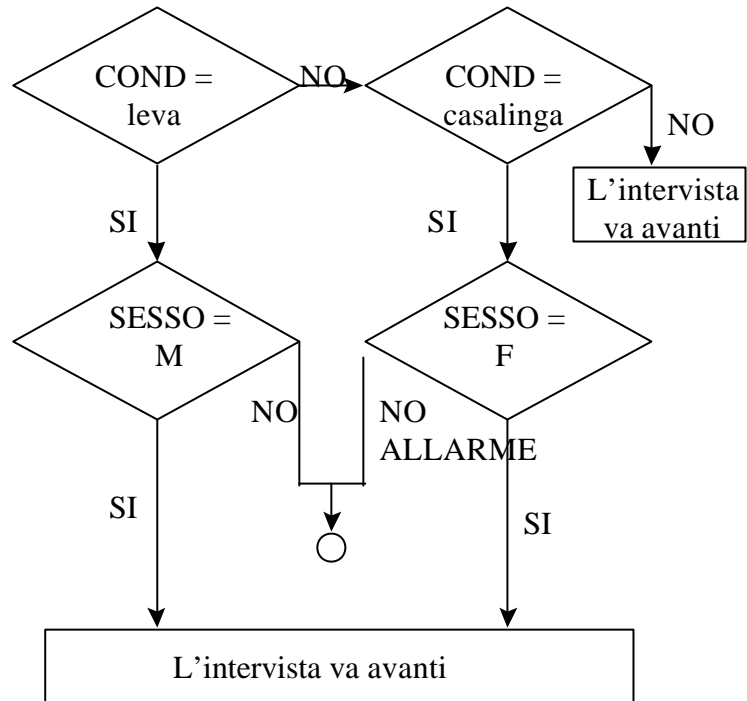
9) Stato civile prima del matrimonio attuale → attiva 2 controlli (con 'stato civile attuale' e 'anno di nascita')

In ogni batteria, l'ultima domanda attiva dei controlli che possono modificare domande precedenti che appartengono a diverse batterie: se prendiamo ad esempio la batteria n° 4, la variabile 'condizione unica o prevalente alla data della rilevazione' attiva due controlli, con 'sesso' e 'anno di nascita', che si trovano entrambe nella prima batteria; se si considera, invece, come esempio la batteria n° 9, la variabile 'stato civile prima del matrimonio attuale' attiva due controlli, con 'anno di nascita' (primo blocco) e 'stato civile attuale' (secondo blocco).

Batteria 4	Condizione economica	controlli con	sesso anno di nascita	batteria 1 batteria 1
Batteria 9	Stato civile prima	controlli con	anno di nascita stato civile	batteria 1 batteria 2

Ogni controllo verifica se le risposte date alle domande che sono legate tra loro dal controllo stesso sono in contraddizione e, in tal caso, il sistema permette di immettere le correzioni. Riprendendo il due esempi della batteria 4 e 9, si attivano i seguenti controlli:

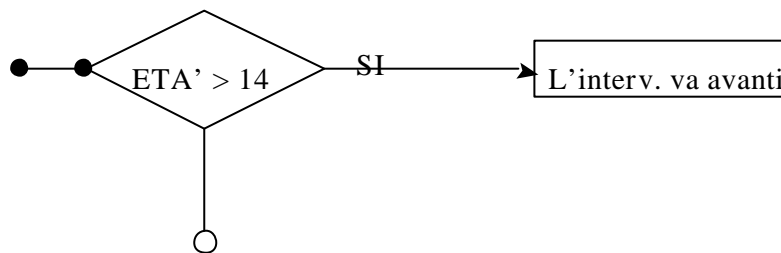
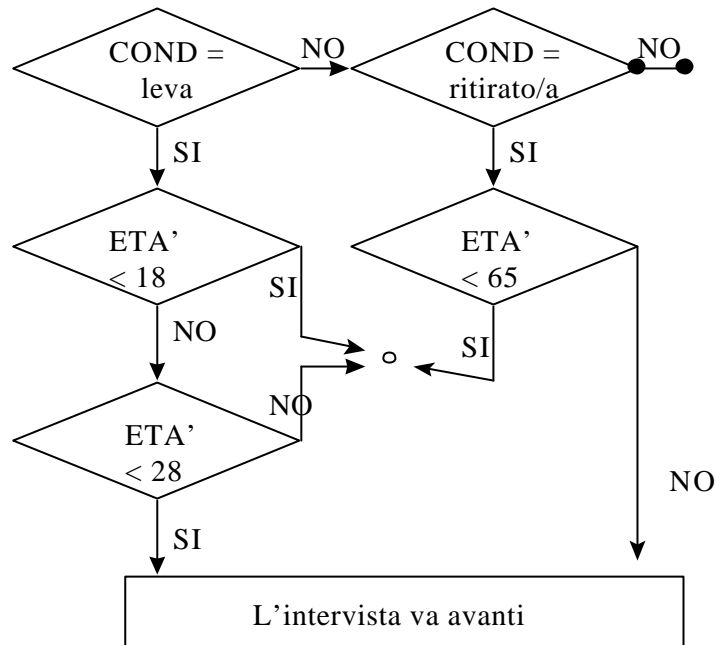
CONTROLLO: CONDIZIONE ECONOMICA - SESSO





SESSO	M	F	
CONDIZIONE UNICA E PREVALENTE (se ha ≥ 14 anni)			
Occupato	ǒ		
In cerca di nuova occupazione		ǒ	
In cerca di prima occupazione		ǒ	
In servizio di leva o servizio sostitutivo		ǒ	
Casalinga	ǒ	Studente	ǒ
Inabile al lavoro	ǒ	Persona ritirata dal lavoro	ǒ
In altra condizione		ǒ	

**CONTROLLO:
CONDIZIONE ECONOMICA - ETA'**



ANNO DI NASCITA _____

CONDIZIONE UNICA E PREVALENTE (se ha ≥ 14 anni)

Occupato	ǒ	In cerca di nuova occupazione	ǒ
In cerca di prima occupazione	ǒ	In servizio di leva o servizio sostitutivo	ǒ
Casalinga	ǒ	Studente	ǒ
Persona ritirata dal lavoro	ǒ	Inabile al lavoro	ǒ
		In altra condizione	ǒ



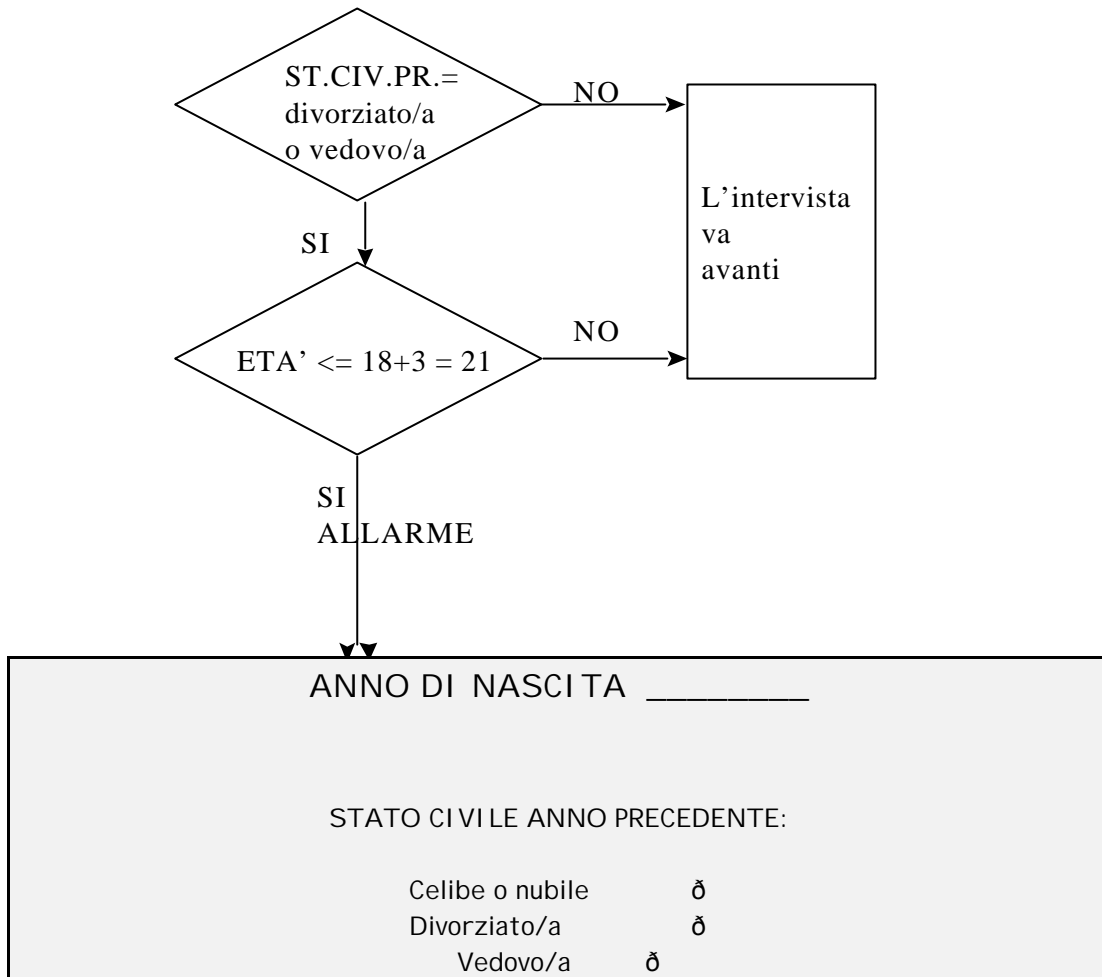
STATO CIVILE ATTUALE

Celibe o nubile ð
Coniugato/a coabitante col coniuge ð
Coniugato/a non coabitante col coniuge (separato/a di fatto) ð
Separato/a legalmente ð
Divorziato/a ð
Vedovo/a ð

STATO CIVILE ANNO PRECEDENTE

Celibe o nubile ð
Divorziato/a ð
Vedovo/a ð

**CONTROLLO: STATO CIVILE PRIMA
DEL MATRIMONIO ATTUALE - ETA'**



Il discorso fatto per le due variabili appena viste e i quattro controlli attivati può essere riprodotto per tutte le batterie di domande e tutti i controlli che vengono attivati nello svolgere delle domande dell'Indagine Multiscopo. Le schede che seguono mostrano i blocchi e i controlli:

ISTAT CADAC INDAGINE MULTISCOPO 2 0 0 _			
<u>COMPONENTE N°</u> ___ → 01			
<u>Residenza all'indirizzo (la sua posizione con riferimento all'indirizzo della famiglia anagrafica è)</u>			
PRESENTE	ð	TEMPORANEAMENTE ASSENTE	ð
<u>Può indicare la sua relazione di parentela con la Persona di Riferimento?</u>			
Persona di Riferimento ð			
Coniuge di PR ð			
Convivente coniugalmente di PR ð			
Genitore (coniuge del genitore, convivente del genitore) di PR ð			
Genitore (coniuge o convivente del genitore) del coniuge o del convivente di PR ð			
Figlio di PR nato dall'ultimo matrimonio o convivenza ð			
Figlio di PR o del coniuge o del convivente nato da precedente matrimonio o convivenza ð			
Coniuge del figlio di PR (o del coniuge o convivente di PR) ð			
Convivente del figlio di PR (o del coniuge o convivente di PR) ð			
Nipote (figlio del figlio) di PR (o del coniuge o convivente di PR) ð			
Nipote (figlio del fratello/sorella) di PR (o del coniuge o convivente di PR) ð			
Fratello/sorella di PR ð			
Fratello/sorella di PR del coniuge o convivente di PR ð			
Coniuge del fratello/sorella di PR (o del coniuge o convivente di PR) ð			
Convivente del fratello/sorella di PR (o del coniuge o convivente di PR) ð			
Altro parente di PR (o del coniuge o convivente di PR) ð			
Persona legata da amicizia ð			
<u>Sesso (lei è maschio o femmina)</u>			
M	ð	F	ð
<u>Anno di nascita (può dire il suo anno di nascita)</u>			
1_ _ _			

CONTROLLO N.1: ANNO NASCITA - ANNO ATTUALE

(CALCOLO AUTOMATICO DELL'ETA' = ANNO ATTUALE - ANNO NASCITA)

CONTROLLO N.2: ETA' - RELAZIONE DI PARENTELA CON PR

STATO CIVILE (se ha ≥ 6 anni)

Celibe o nubile

Coniugato/a coabitante col coniuge

Coniugato/a non coabitante col coniuge (separato/a di fatto)

Separato/a legalmente

Divorziato/a

Vedovo/a

CONTROLLO N.3: STATO CIVILE - RELAZIONE DI PARENTELA

CONTROLLO N.4: STATO CIVILE - ANNO DI NASCITA (ETA')

TITOLO DI STUDIO PIU' ALTO CONSEGUITO (se ha ≥ 6 anni)

Dottorato di ricerca o specializzazione post laurea

Laurea

Diploma universitario

Diploma o qualifica di scuola media sup. (4-5) anni

Diploma o qualifica di scuola media sup. (2-3) anni

Licenza di scuola media inferiore

Licenza elementare

Nessun titolo - sa leggere e scrivere

Nessun titolo - non sa leggere e/o scrivere

CONTROLLO N.5: TITOLO DI STUDIO - ANNO DI NASCITA (ETA')

CONDIZIONE UNICA E PREVALENTE (se ha ≥ 14 anni)

Occupato

In cerca di nuova occupazione

In cerca di prima occupazione

In servizio di leva o servizio sostitutivo

Casalinga Studente

Inabile al lavoro Persona ritirata dal lavoro

In altra condizione

CONTROLLO N.6: CONDIZIONE ECONOMICA - SESSO

CONTROLLO N.7: CONDIZIONE ECONOMICA - ANNO DI NASCITA

POSIZIONE NELLA PROFESSIONE (se è occupato/a)		
<u>Alle dipendenze come:</u>		
Dirigente	ð	
Direttivo quadro	ð	
Impiegato, intermedio	ð	
Capo operaio, operaio subalterno ed assimilati	ð	
Apprendista	ð	
Lavorante a domicilio per conto d'impresa		ð
<u>Autonomo come:</u>		
Imprenditore	ð	
Libero professionista	ð	
Lavoratore in proprio	ð	
Socio di cooperativa di produzione di beni e/o prestazioni di servizio		ð
Coadiuvante	ð	

CONTROLLO N.8: POSIZIONE NELLA PROFESSIONE - COND. ECONOMICA

CONTROLLO N.9: POSIZIONE NELLA PROFESSIONE - ANNO DI NASCITA

ATTIVITA' ECONOMICA (se è occupato/a)		
Agricoltura, caccia, pesca	ð	
Industria, estrazione, manifatturiere, energia		ð
Costruzioni	ð	
Commercio all'ingrosso e al dettaglio, alberghi e ristoranti		ð
Trasporti, magazzinaggio e comunicazioni		ð
Intermediazioni, noleggio, altre attività professionali		ð
Pubblica amministrazione e difesa	ð	
Istruzione, sanità e altri servizi sociali		ð
Altri servizi	ð	

CONTROLLO N.10: ATTIVITA' ECONOMICA - CONDIZIONE ECONOMICA

FONTE PRINCIPALE DI REDDITO (se 14 anni e più)

Reddito da lavoro dipendente	ð
Reddito da lavoro autonomo	ð
Pensione	ð
Indennità e provvidenze varie	ð
Redditi patrimoniali	ð
Mantenimento da parte dei familiari	ð

CONTROLLO N.11: FONTE PRINCIPALE DI REDDITO - ANNO DI NASCITA

CONTROLLO N.12: FONTE PRINCIP. DI REDDITO - POSIZ. NELLA PROFESS.

CONTROLLO N.13: FONTE PRINCIPALE DI REDDITO - CONDIZIONE

SITUAZIONE ALLA STESSA DATA DELL'ANNO PRECEDENTE

ANNO DEL MATRIMONIO ATTUALE

19 _ _

(Solo per le persone coniugate e coabitanti)

CONTROLLO N.14: ANNO DEL MATRIMONIO ATTUALE - STATO CIVILE

CONTROLLO N.15: ANNO DEL MATRI. ATTUALE - ANNO DI NASCITA

CONTROLLO N.16: ANNO DEL MATRI. ATTUALE - RELAZ. CON PR

Questo controllo è già automatico e deriva dai controlli

ANNO DI NASCITA- ANNO DI MATRIMONIO

ANNO DI NASCITA RELAZIONE DI PARENTELA

STATO CIVILE PRIMA DEL MATRIMONIO ATTUALE

Celibe o nubile	ð
Coniugato/a coabitante col coniuge	ð
Coniugato/a non coabitante col coniuge (separato/a di fatto)	ð
Separato/a legalmente	ð
Divorziato/a	ð
Vedovo/a	ð

CONTROLLO N.17: STATO CIVILE PRIMA DEL MATRI. - STATO CIVILE ATT.

CONTROLLO N.18: STATO CIVILE PRIMA DEL MATRI. - ANNO DI NASCITA

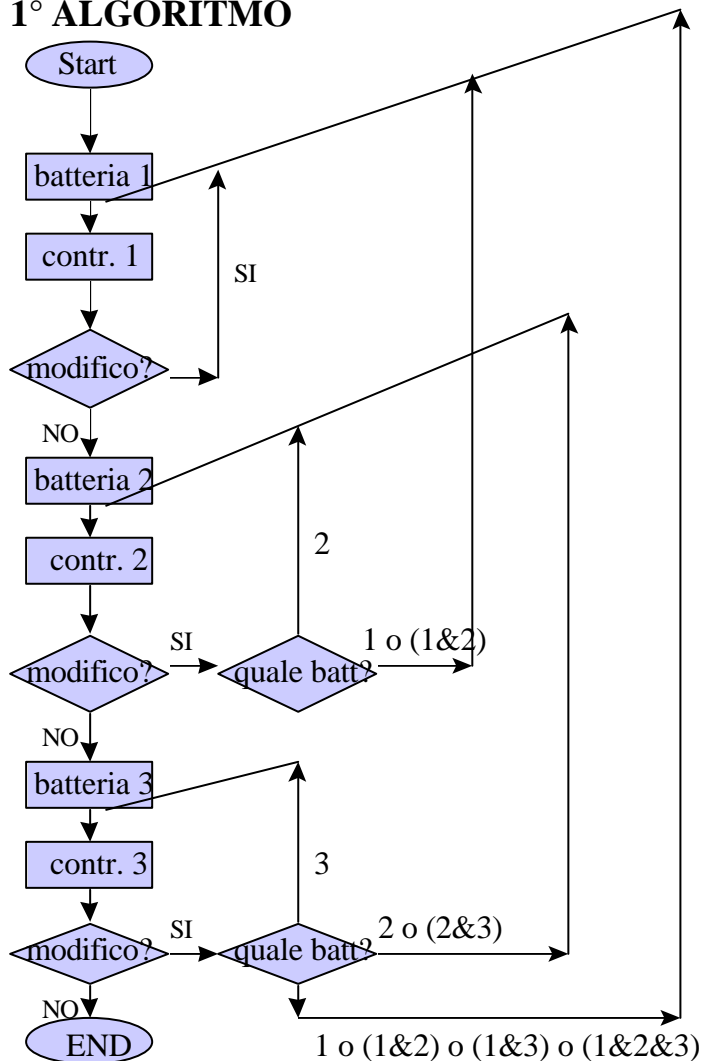
3.3 Gli algoritmi

Per esemplificare quanto appena analizzato, si osservi il seguente algoritmo, supponendo di avere solo tre batterie di domande.

Al termine della prima batteria, l'ultima domanda attiva il controllo: se la risposta data entra in contraddizione con un'altra precedente, si deve modificare una delle risposte; in tal caso il sistema riesegue i controlli alla luce delle correzioni.

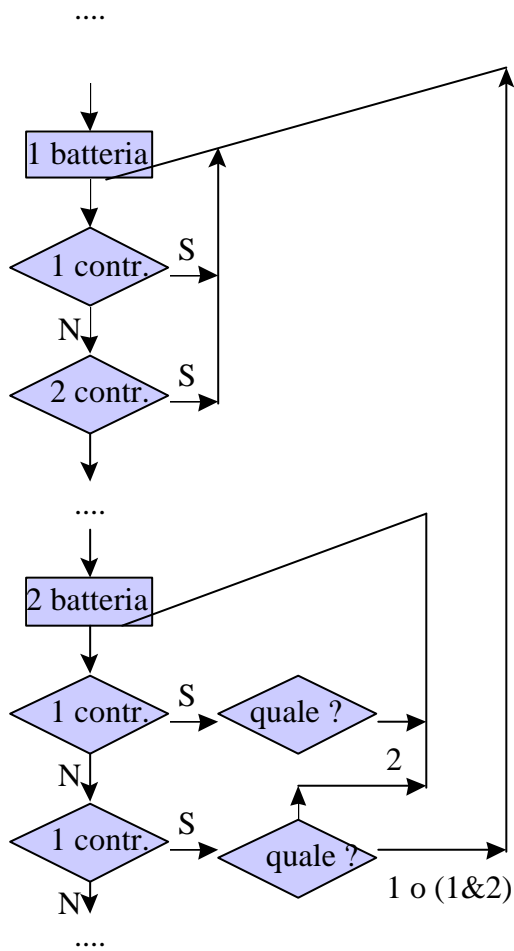
Alla fine dei controlli e delle correzioni, si passa alla batteria successiva: in questo caso, se si esegue una correzione, il sistema deve saper individuare a quale blocco appartenga la variabile su cui agisce, per ricominciare i controlli dal punto giusto.

1° ALGORITMO



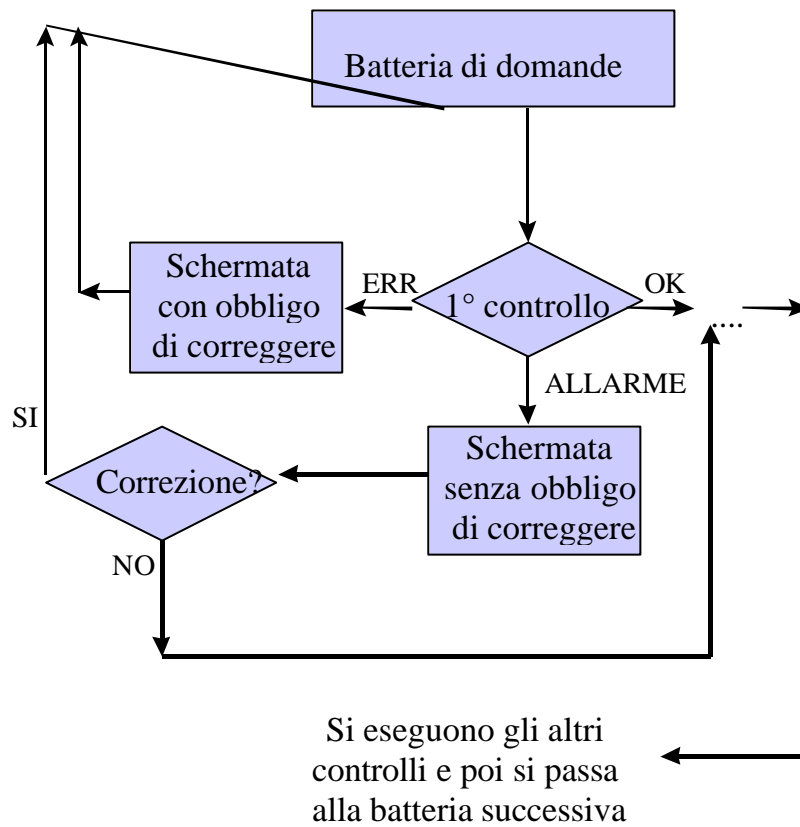
Però, l'algoritmo può essere ulteriormente articolato, perché ogni variabile, come ormai è noto, può attivare più di un controllo:

2° ALGORITMO



E ancora: se analizziamo in particolare la logica del passaggio tra una batteria di domande e l'altra, si ottiene il seguente algoritmo:

3° ALGORITMO



3.4 Conclusioni

E' stata presentata fin qui una simulazione di piano di editing sull'Indagine Multiscopo, analizzando i passaggi necessari alla trasformazione del questionario cartaceo in questionario *on line*, approfondendo, in particolare, le modalità di correzione. Le procedure di controllo e correzione possono in realtà essere varie, ma tutte partono dallo stesso principio: il programma confronta le risposte date a differenti domande e evidenzia le incoerenze o le combinazioni di risposte troppo contraddittorie. Chiaramente, ciò che non può fare il programma è segnalare quale tra quelle evidenziate sia la risposta sbagliata; ma l'intervista *on line* permette di consultare l'intervistato in tempo reale per verificare la correttezza delle risposte.

La letteratura sul tema illustra, e l'applicazione dovrebbe contribuire, che le tecniche che supportano le interviste *computer assisted* sono piuttosto complesse, soprattutto se il questionario *on line* deve, cosa auspicata, apportare miglioramenti al tradizionale supporto cartaceo.

Sezione II: Indagine sulle Forze di Lavoro

A cura di Cristina Panattoni

4 Cenni preliminari sull'indagine

L'indagine sulle Forze di Lavoro è una rilevazione a cadenza trimestrale condotta in base ad un disegno campionario a due stadi in cui le unità del primo stadio sono i Comuni e quelle del secondo sono le famiglie.

Ciascuna famiglia viene intervistata *face to face* per due trimestri consecutivi, per i successivi due viene esclusa dall'indagine per poi rientrare ed essere intervistata ancora per due occasioni consecutive (schema di rotazione 2-2-2).

5 Come si struttura il controllo e la correzione dei dati sulle Forze di Lavoro

Nella RTFL (Rilevazione Trimestrale sulle Forze di Lavoro) il controllo sui dati raccolti tramite questionario cartaceo avviene già a partire dalla fase di registrazione dei dati, compito questo affidato a una società esterna. Tale ditta utilizzando *softwares* sviluppati al proprio interno effettua un preliminare controllo sui domini di ciascuna variabile e segnala la presenza di eventuali errori.

Le mancate risposte totali vengono minimizzate sostituendo nei due stadi le unità non rispondenti: i Comuni del gruppo B (comuni non capoluogo di provincia con popolazione inferiore ai 20.000 abitanti) vengono sostituiti se le mancate risposte superano il 50% del previsto (se, cioè, la numerosità dello stadio si riduce in misura maggiore del 50% a causa delle mancate risposte).

La sostituzione non riguarda i Comuni capoluogo di provincia e tutti quelli con più di 20.000 abitanti che devono necessariamente essere inclusi nel campione (comuni autorappresentativi).

Dalla tabella seguente si desume che mediamente il tasso di non risposta dei Comuni, per le rilevazioni da aprile '96 a luglio '98, è stato dell'1,5% circa.

Nel grafico 1 viene presentata la distribuzione dei comuni non pervenuti, a partire dalla rilevazione dell'ottobre 1992 fino a quella dell'ottobre 1999, suddivisi in base alla rappresentatività.

Tab.1 Tassi di non risposta (Aprile '96- Luglio '98)

	Apr. 96	Lug. 96	Ott. 96	Gen. 97	Apr. 97	Lug. 97	Ott. 97	Gen. 98	Apr. 98	Lug. 98
Comuni	1.351	1.351	1.351	1.351	1.351	1.351	1.351	1.351	1.351	1.351
- non pervenuti	17	22	15	22	20	16	25	16	18	24
-% non risposta	1,26	1,63	1,11	1,63	1,48	1,18	1,85	1,18	1,33	1,78
Famiglie	75.394	75.463	75.491	75.463	75.332	75.426	75.405	75.331	75.362	75.258
- non pervenute	122	53	25	53	184	90	111	185	154	258
- % non risposta	0,16	0,07	0,03	0,07	0,24	0,12	0,15	0,25	0,20	0,34

Fonte: ISTAT- Forze di lavoro: documento interno.

Le singole famiglie non rispondenti vengono sostituite attingendo nominativi da apposite liste suppletive (per questo i tassi di non risposta delle famiglie sono irrilevanti).

In un'ottica longitudinale questa strategia crea non pochi problemi relativi all'abbinamento di record individuali e familiari a discapito della ricostruzione dell'informazione nel tempo².

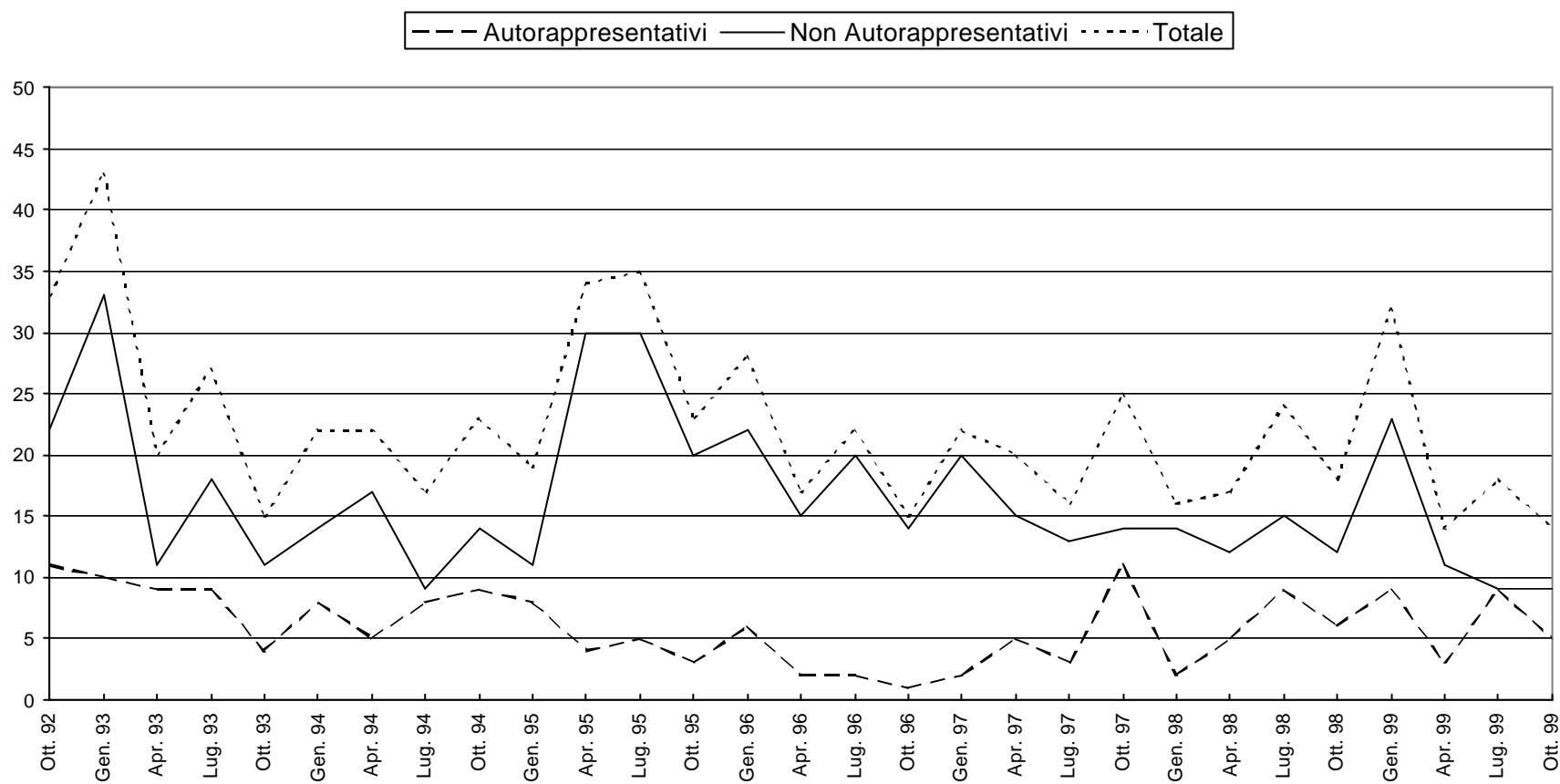
² Per approfondimenti e applicazioni delle procedure di abbinamento al caso dell'Indagine sulle Forze di lavoro si rimanda a:

Paggiaro A., Torelli N. (1999) Una procedura per l'abbinamento di record nella rilevazione trimestrale delle forze di lavoro. Progetto di ricerca Murst "Lavoro e disoccupazione: questioni di misura e analisi", working paper n.15, Dipartimento di Scienze Statistiche, Università di Padova.

Favro-Paris M., Gennari P., Oneto G.P. (1996) La durata della disoccupazione in Italia. In: *Quaderni di ricerca*, n.4, ISTAT, Roma.

Grafico 1 – Comuni non pervenuti

Rilevazioni da ottobre '92 a ottobre '99



Fonte : ISTAT – Forze di lavoro: documento interno.

I controlli relativi alla coerenza dei dati e all'individuazione/correzione delle mancate risposte parziali vengono effettuati attraverso l'interazione di diversi programmi che implementano le due differenti metodologie di correzione: quella deterministica e quella probabilistica.

Una sequenza di programmi SAS e COBOL "pulisce" il file proveniente dalla registrazione e lo organizza in modo tale da poter essere sottoposto ai successivi passi di correzione.

In SAS e COBOL vengono effettuati, anche, i controlli deterministici che riguardano sia le variabili anagrafiche che quelle occupazionali. Le poche variabili quantitative presenti sul questionario vengono verificate tutte con espressioni di tipo deterministico.

Per quanto attiene alle correzioni probabilistiche, prima di entrare nel dettaglio, è necessario ricordare che, a partire dalla rilevazione di aprile 1999, è stato adottato per la RTFL un *software* probabilistico diverso da quello utilizzato nelle rilevazioni precedenti, a seguito del progetto di ristrutturazione organizzativa che sta coinvolgendo tutti i servizi dell'ISTAT e che dovrà concludersi entro il 2000.

Tale ristrutturazione prevede il passaggio da un sistema informativo centralizzato ad uno distribuito, con *server* dedicati per singoli servizi e, di conseguenza, il cambiamento di ambiente operativo in grado di gestire la nuova situazione (da CMS a UNIX).

Pertanto, mentre fino al gennaio 1999 il *software* utilizzato per la correzione/imputazione probabilistica dei dati era DAISY (*Design Analysis and Imputation SYstem*) che operava in ambiente CMS, dalla rilevazione di aprile 1999, si è passati all'utilizzo di SCIA (Sistema Controllo e Imputazione Automatici) che opera in ambiente UNIX e che è in uso già presso la maggior parte dei servizi dell'Istituto.

Entrambi i *softwares* implementano l'algoritmo di Fellegi-Holt per variabili qualitative.

Conseguentemente, è stato revisionato il piano di incompatibilità, ma solo per esigenze di adeguamento rispetto al nuovo *software* utilizzato. Con l'occasione sono state apportate alcune modifiche volte a rendere più efficiente la correzione.

I programmi attualmente utilizzati per la pulizia dei dati sulle Forze di Lavoro, eseguiti a cascata, e le relative funzionalità sono sintetizzate nella tabella seguente³:

Tab.2 Sequenza dei programmi attualmente utilizzati per la “pulizia” dei dati nella RTFL

Denominazione	Tipo	Funzionalità
AWK250	UNIX	rende uniformi le lunghezze dei record
Nodup.sas	SAS	elimina i record vuoti
AWK250	UNIX	rende uniformi le lunghezze dei record
Contaco.sas	SAS	controlla il numero di comuni e di famiglie
Efldet01	COBOL	effettua controlli sul tracciato record (allineamenti, etc.) verifica la composizione familiare e l'età degli individui
Postdet1.sas	SAS	correzione deterministica su var. quantitative (es: "ore e giorni di lavoro nella settimana di riferimento"; "anno e mese di inizio e fine dell'ultima occupazione")
Preter.sas	SAS	correzione deterministica in relazione alle modifiche del questionario nel 1995
SCIA01	SCIA	correzione probabilistica - anagrafica
Separa	COBOL	crea due files: < 15 anni e > 15 anni
Det02	COBOL	correzione deterministica sulle modalità di lavoro (turni, etc., ore abituali/ore effettive, etc.)
Postdet02.sas	SAS	Postdet1.sas, di controllo dopo il 1° passo probabilistico + controlli professione/posizione nella professione
SCIA02	SCIA	correzione probabilistica - occupazione (unione del secondo e terzo passo di DAISY)
SCIA03	SCIA	correzione probabilistica - ricerca di lavoro
SCIA04	SCIA	correzione probabilistica -istruzione e formazione
Prg		riunificazione dei files < 15 anni e > 15 anni
Bisdet04	COBOL	correzione deterministica controlli familiari (relazione con il capo famiglia, etc.)
Postfin.sas	SAS	correzione deterministica controlli su inizio/fine attività/periodo rilevazione

Al fine di rendere più snella la procedura di correzione probabilistica SCIA viene applicato separatamente su tre sezioni distinte del questionario (quattro quando viene aggiunto il modulo di rilevazione su istruzione e formazione).

³ Le informazioni relative ai programmi utilizzati e ai cambiamenti apportati al piano di incompatibilità sono state desunte da: Cascioli R. (1999) “Il passaggio da DAISY a SCIA: migrazione dal sistema operativo CMS a UNIX del piano di controllo e correzione dei dati relativi all’indagine sulle Forze di lavoro”, *FIL/C: documento interno (resoconto del lavoro svolto da R. Cascioli, P. Gennari, V. Spinelli e C. Lucarelli)*, ISTAT, Roma.

Ciascuna applicazione prende in considerazione un diverso tipo di variabili: nel primo passo probabilistico vengono corrette le variabili anagrafiche e vengono controllati, in base a queste, i successivi salti del questionario in considerazione soprattutto dei filtri sull'età e sul motivo dell'assenza dal comune di residenza.

Il secondo passo di SCIA riguarda solo il file di coloro che hanno un'età maggiore di 15 anni ed è quello di maggiore rilevanza poiché in esso sono contenute le regole di incompatibilità relative all'occupazione; i controlli sono indirizzati alla correzione di errori formali, nel caso ci siano risposte non dovute, e di errori sostanziali, nel caso in cui si riscontrino delle illogicità tra le risposte fornite ai diversi quesiti (titolo di studio/professione, professione/branchia di attività economica, professione/posizione nella professione, titolo di studio/età, posizione nella professione/età, etc.).

Il terzo passo probabilistico prende in esame le variabili relative alla ricerca di lavoro, alla disponibilità al lavoro e alla condizione precedente alla ricerca di lavoro.

6 *Il nuovo piano di incompatibilità per la RTFL*

La sostanziale differenza rispetto al precedente piano di incompatibilità risiede, in generale, nella riduzione della numerosità delle regole, a favore di una maggiore integrità del dato originale, e nell'introduzione di alcune incompatibilità che non erano state valutate in precedenza.

In particolare, la maggior parte delle modifiche ha riguardato le associazioni tra la "professione" e la "posizione nella professione" e tra la "professione" e la "branchia di attività economica".

6.1 *Correzione deterministica*

Analizzando le relazioni tra "professione" e "posizione nella professione", nel nuovo piano sono state introdotte alcune regole deterministiche che cancellano la professione per quelle combinazioni di modalità che palesemente risultano essere incompatibili (cfr. **Tab.3**).

In questo modo, è stato attribuito un maggior grado di fiducia alle risposte fornite alla domanda sulla posizione nella professione che, del resto, è di più facile compilazione rispetto alla domanda sulla professione per cui l'intervistato deve dichiarare la propria professione e l'intervistatore deve codificarla individuando l'esatta corrispondenza nella classificazione ISTAT.

Tab.3 Incompatibilità tra la professione e la posizione nella professione.

PROFESSIONE	POSIZIONE NELLA PROFESSIONE
legislatori, dirigenti, imprenditori impresa	impiegato o intermedio operaio, subalterno e assimilati apprendista lavoratore a domicilio per conto di
professioni intellettuali, scientifiche e di elevata specializzazione impresa	operaio, subalterno e assimilati apprendista lavoratore a domicilio per conto di
professioni intermedie (tecnici) impresa	lavoratore a domicilio per conto di
professioni esecutive relative all'amministrazione e gestione	libero professionista lavoratore in proprio
professioni relative alla vendita e ai servizi alle famiglie	dirigente
artigiani, operai specializzati e agricoltori	dirigente direttivo-quadro libero professionista
conduttori di impianti, operatori di macchinari fissi e mobili (anche in agricoltura), operai di montaggio industriale	dirigente direttivo-quadro libero professionista
personale non qualificato	dirigente direttivo-quadro impiegato o intermedio libero professionista

Non sono state introdotte, altresì, correzioni deterministiche in combinazione con le posizioni “socio di cooperativa” e “coadiuvante” poiché si è ritenuto che non entrassero in contraddizione con nessuna categoria professionale. Allo stesso modo, si è proceduto per la posizione “imprenditore” ma sulla base di diverse considerazioni.

Infatti, a rigor di logica, sono imprenditori solo quei soggetti che dichiarano di essere tali sia nella domanda sulla professione, sia in quella sulla posizione nella professione: accade, però, che molti imprenditori tendano a dichiarare come professione il lavoro che svolgono o la professionalità acquisita. La scelta di non cancellare la professione dichiarata nasce, allora, dall’esigenza di non perdere troppa informazione sulla variabile stessa.

Di particolare rilievo risultano le modifiche apportate alle correzioni di carattere deterministico per le variabili:

anni di residenza: per valori > 11 (fuori dominio) è imputato il valore 11 (persone con più di 10 anni di residenza in Italia).

In precedenza, attraverso il primo passo probabilistico, venivano individuati i valori fuori dominio e si procedeva, quindi, a cancellarli lasciando vuoto il campo. L’inconveniente nasceva dal fatto che per i cittadini italiani la variabile doveva essere lasciata in bianco: una tale correzione aumentava del 14% il numero di cittadini italiani;

relazione con il capo famiglia: è stata eliminata la regola che, nel vecchio piano di incompatibilità, sostituiva la modalità “coniugato-convivente” con “altro parente o affine” se lo stato civile dichiarato era diverso da “coniugato”. La perdita delle convivenze (- 3,1%) era la conseguenza di tale regola.

6.2 *Correzione probabilistica*

Nell’ambito della correzione probabilistica sono state ridotte le incompatibilità tra la professione e la branca di attività economica in quanto il *software* SCIA (2° passo) non era in grado di generare l’insieme completo degli *edit* a partire dalla stessa quantità di regole esplicite introdotte, in merito, nella procedura implementata con il *software*

DAISY. Sono risultati, altresì, eccessivi alcuni *edit* che venivano attivati troppo spesso poiché introducevano delle condizioni troppo rigide, non sempre riscontrabili nella realtà.

Nel precedente piano di incompatibilità, i controlli sulla variabile “professione” avvenivano sulle codifiche a quattro cifre, ora su quelle a tre cifre.

Ciascuna sottocategoria professionale, specificata alla quarta cifra, veniva considerata incompatibile con tutte le branche di attività economica diverse da quella di appartenenza, producendo, in questo modo, un elevato numero di regole; ora, invece, tali incompatibilità vengono riassunte in una unica regola più generale riguardante la categoria professionale a livello superiore, ovvero specificata alla terza cifra, e le branche di attività economica con cui risulta essere, in generale, incompatibile.

Per esempio, nella classificazione ISTAT delle professioni, sotto la voce “legislatori, dirigenti e imprenditori” (prima cifra uguale a 1), la categoria professionale “imprenditori e amministratori di aziende private”, propone sottocategorie professionali analoghe, differenziate unicamente per settore di operatività dell’azienda. Nel vecchio piano di controllo ciascun sottogruppo veniva considerato incompatibile con tutti i settori di attività economica diversi da quello specifico dell’azienda. Nella nuova formulazione, viene proposta un’unica regola che riguarda l’incompatibilità tra “imprenditori e amministratori di aziende private” e i soli settori “pubblica amministrazione e difesa” e “organizzazione e organismi internazionali”.

Nei casi in cui l’accorpamento alla terza cifra non ha permesso la specificazione di regole generali, gli *edit* relativi a quelle categorie professionali sono stati eliminati (per esempio, nell’ambito della categoria “professioni relative alla vendita e ai servizi per le famiglie – prima cifra uguale a 5 – sono state eliminate due regole in riferimento alle “professioni relative alla sicurezza” poiché, accorpando alla terza cifra, non aveva senso riprodurle).

Nel seguito, riportiamo solo le categorie professionali che hanno subito maggiori variazioni nel nuovo piano di incompatibilità.

Le “professioni intermedie (tecnici)” (prima cifra uguale a 3) accorpate alla terza cifra hanno consentito di eliminare alcune incompatibilità che si manifestavano specificando nel

dettaglio tale categoria: sono stati accorpate i “tecnici paramedici” e tutte le “professioni intermedie delle attività turistico-alberghiere”; è stata eliminata l’incompatibilità tra le “professioni intermedie in imprese commerciali e assimilati” e il settore “industria” poiché ritenuta poco coerente; alle categorie “professioni intermedie finanziario assicurative” e “professioni intermedie dell’insegnamento” sono stati aggiunti rispettivamente i settori “pubblica amministrazione, difesa assicurazione sociale e obbligatoria” e “sanità ed altri servizi sociali”, così da renderli compatibili.

Nella categoria “professioni relative alla vendita e ai servizi per le famiglie” (prima cifra uguale a 5), il sottogruppo “professioni commerciali” è stato unificato in un’unica regola che esplicita l’incompatibilità con solo alcune branche di attività economica. Le “professioni nelle attività turistiche e alberghiere” sono state unificate in un solo *edit* ed è stata eliminata l’incompatibilità con i settori “commercio” e “altri servizi pubblici, sociali e personali”: nel precedente piano tale regola riguardante la sottocategoria (specificata alla quarta cifra) degli “esercenti e addetti alla ristorazione e ai pubblici servizi” veniva attivata un numero considerevole di volte (circa 650).

Le “professioni relative ai servizi personali” sono state considerate compatibili con i settori “sanità e altri servizi sociali e personali” e “pubblica amministrazione” mentre prima lo erano solo con i settori “alberghi e ristoranti” e “altri servizi pubblici, sociali e personali”.

Le professioni “artigiani, operai specializzati e agricoltori” (prima cifra uguale a 6) avevano, nel vecchio piano di controllo, numerose incompatibilità poiché queste coinvolgevano anche la “posizione nella professione”. Lo scopo era quello di discriminare tra un lavoro autonomo e uno dipendente; per la stessa professione si poteva scrivere una regola con ogni attività diversa dal commercio al dettaglio o dall’industria.

Per la stesura del nuovo piano, è stata eliminata la variabile “posizione nella professione”, al fine di ridurre la numerosità delle regole.

Inoltre, sono state accorpate le professioni “artigiani e operai del tessile e dell’abbigliamento” e “artigiani e operai della lavorazione del cuoio, delle pelli, delle calzature e assimilati”: non vengono più esplicitate le incompatibilità di ciascuna

sottocategoria di queste professioni con i vari settori industriali ma, nel nuovo piano, è presente un'unica regola riguardante l'incompatibilità delle professioni così accorpate con alcuni settori industriali diversi dai settori specifici della professione e dal commercio al dettaglio.

Anche altre professioni della stessa categoria (“artigiani e operai delle lavorazioni alimentari”, “ebanisti, attrezzisti, operai, e artigiani del trattamento del legno e assimilati”), al loro interno, sono state unificate in un'unica regola che esprime le incompatibilità esclusivamente con alcuni settori di attività economica producendo una riduzione e una minore rigidità delle regole (in precedenza, alcune di esse venivano attivate un eccessivo numero di volte).

All'interno della categoria “personale non qualificato” (prima cifra uguale a 8), la voce “personale non qualificato nel settore dell'istruzione”, è stata considerata compatibile con il settore della “pubblica amministrazione” oltre che con la branca di attività economica “istruzione”.

Per quanto riguarda, infine, le “forze armate” (prima cifra uguale a 9), nella revisione dei controlli, è stato rilevato e corretto un errore nella formulazione di una regola che, nel vecchio piano, esplicitava l'incompatibilità con un'attività a tempo pieno piuttosto che un'attività a tempo parziale. Questo errore, nel passato, ha comportato la sottostima di tale classe professionale nella RTFL.

Sempre in riferimento alla correzione probabilistica, altre relazioni tra variabili, hanno subito delle modifiche. Le più significative hanno riguardato:

posizione nella professione/titolo di studio: è stata eliminata l'incompatibilità tra “dirigente”, “direttivo-quadro” e titolo di studio inferiore al “diploma di qualifica professionale”, tra “libero professionista” e tutti i tipi di titolo di studio corrispondenti ai diplomi di qualifica professionale. E' rimasta l'incompatibilità tra “libero professionista” e alcuni titoli di studio ma, con una minore restrizione sul titolo di studio;

professione/titolo di studio: sono state rese compatibili le professioni di “imprenditore, amministratore, dirigente e direttore di aziende private” con i titoli di studio diversi da “dottorato” e “laurea”, causando un recupero della categoria stessa.. Alcune professioni

intellettuali e intermedie (tecnici) sono state rese incompatibili con titoli di studio inferiori al diploma di qualifica professionale (2-3 anni di corso) invece del diploma di maturità (4-5 anni di corso);

posizione nella professione/professione: è stata aggiunta l'incompatibilità tra la posizione di "imprenditore" e la risposta "no" alla variabile "ha dipendenti";

paese di nascita/cittadinanza/anni di residenza: è stata eliminata la regola che legava erroneamente il paese di nascita alla cittadinanza e agli anni di residenza in Italia.

6.3 *Principali risultati*

Poiché i nuovi controlli sono stati applicati all'intera serie storica revisionata, ottobre '92 - gennaio '99, è stata creata una versione *master* del piano di incompatibilità con periodo di riferimento aprile '95 – gennaio '98; le altre versioni sono state ottenute da quest'ultima per differenza, cioè sono stati inseriti e/o eliminati programmi deterministici e regole di incompatibilità.

In generale, l'adozione del nuovo piano di controllo e correzione ha avuto i seguenti risvolti⁴:

- una maggiore vicinanza tra dati grezzi e dati puliti (cfr. **Allegato** – tavole di transizione⁵);
- sono state recuperate le forze armate - del tutto assenti nelle rilevazioni pregresse - e la categoria dei dirigenti e imprenditori è raddoppiata;
- la distribuzione degli occupati per settori di attività economica è risultata più vicina a quella rilevata con il Censimento intermedio dell'Industria.

⁴ Gennari P., Spinelli V., "Stato di attuazione del lavoro di revisione delle serie statistiche dell'indagine forze di lavoro (ott. '92 – gen. '99) e proposta di modifica di alcune definizioni", *FIL/C: documento interno*, ISTAT, Roma, 18 maggio 1999.

⁵ L'analisi delle tavole di transizione dati grezzi/dati puliti permette di valutare il tipo di correzione effettuata dopo il passaggio del programma di "pulizia". La situazione ottimale si riscontra quando la distribuzione si concentra sulla diagonale principale e gli altri casi si equidistribuiscono nelle caselle rimanenti. Quando si verifica una concentrazione in celle diverse da quelle della diagonale principale e le distribuzioni marginali non hanno lo stesso andamento, la causa può essere individuata in una non corretta specificazione delle regole di incompatibilità e/o nell'esistenza di errori sistematici.

I risultati, esposti nel seguito, fanno riferimento all'applicazione relativa alla rilevazione di aprile 1997⁶.

Secondo gli indicatori di bontà delle correzioni calcolati (numero di correzioni, matrici di transizione dati grezzi/dati puliti, cfr. **Allegato**) a seguito dei cambiamenti intervenuti è risultato principalmente che:

- SCIA corregge più frequentemente utilizzando l'imputazione congiunta allargata rispetto a quella ristretta (cfr. **Allegato -Tab.1**);
- i record errati al 2° passo probabilistico (occupazione) si sono ridotti dal 17,2% al 14,9%: con l'adozione di SCIA ci sono 3.998 record in più su cui non interviene la procedura di correzione (cfr. **Allegato -Tab.1**);
- in totale la percentuale di record con nessuna variabile modificata dal programma di controllo (solo per il file degli individui con età al di sopra dei 14 anni) è passata dal 46,2% (DAISY) al 47,3% (SCIA) (cfr. **Allegato -Tab.2**);
- sulla variabile "branca di attività economica" si sono ridotte drasticamente le correzioni (-67,8 %) passando da 8.199 con DAISY a 2.632 con SCIA (cfr. **Allegato -Tab.3**).

Il confronto tra i due piani ha evidenziato (cfr. **Allegato** – Matrice di transizione per la variabile "attività economica") una riduzione degli addetti nel settore alberghi e ristoranti (22%), nel settore agricoltura (9,6%) e nei settori costruzione e commercio (3-4%). Al contrario, si sono riscontrati recuperi di unità nel settore "pubblica amministrazione, difesa e organismi internazionali" (12,7%), nelle "altre attività professionali e imprenditoriali" (7,5%) e nell'industria (5,9%).

7 Piano di controllo e correzione dell'Indagine sulle Forze di Lavoro: quali prospettive?

Le informazioni derivanti dall'indagine sulle Forze di lavoro possono essere sfruttate secondo due differenti ottiche: trasversale e longitudinale.

⁶ I risultati di aprile 1997 sono generalizzabili agli altri trimestri per i quali non si sono riscontrate differenze significative in relazione all'adozione del nuovo piano di controllo e correzione.

Nel primo caso, ogni occasione di indagine viene considerata indipendente dalle altre e le notizie acquisite precedentemente vengono utilizzate per effettuare confronti tra aggregati (la struttura panel ha la sola funzione di rendere più stabili le stime delle variazioni); il piano di incompatibilità e la successiva fase di correzione/imputazione sono progettati in funzione di ciò che è rilevato nella singola occasione sfruttando le conoscenze a priori sulle variabili e sulle relazioni esistenti tra le modalità di risposta (come avviene tuttora).

Nel secondo caso, si tiene conto della struttura campionaria secondo lo schema di rotazione 2-2-2 per cui la stessa unità viene intervistata a distanza di 3, 12, 15 mesi dalla prima occasione di contatto; le rilevazioni non sono indipendenti e si ottengono informazioni sui flussi (passaggi da una condizione all'altra; per es: da disoccupato ad occupato).

Per quanto attiene al controllo dei dati, si possono confrontare le risposte date in tempi diversi da uno stesso individuo e correggere le variabili in funzione dei valori che hanno assunto nelle precedenti occasioni.

Nel successivo paragrafo si considera la possibilità di sfruttare l'informazione longitudinale della RTFL in un piano di controllo inserito in un sistema di rilevazione *computer-assisted*.

Si propongono alcuni spunti riguardanti le aree tematiche relative all'indagine per la programmazione di controlli *on line* nella gestione informatizzata del questionario.

Si individuano alcune informazioni che possono essere tenute fisse durante le diverse occasioni di indagine, cioè quelle informazioni che acquisite al primo contatto non dovranno essere più richieste e modificate, poiché invarianti nel tempo.

8 *Alcuni spunti propositivi sul controllo on line dei dati della RTFL per grandi aree tematiche*

Composizione familiare

Se si vuole delineare nel tempo il percorso formativo e occupazionale di un individuo per rendere conto delle variazioni del fenomeno "occupazione", in un'indagine in cui l'unità

di rilevazione è la famiglia, un primo controllo da attivare è quello sulla composizione familiare nelle varie occasioni.

E' necessario verificare se un individuo sia entrato o uscito dalla famiglia nel periodo tra un'onda e l'altra.

Utilizzando le schermate dinamiche si può proporre all'intervistato la struttura familiare dichiarata nel precedente contatto: se non sono intervenuti cambiamenti si chiede, a video, di confermare i dati che compaiono sulla schermata; se, al contrario, si sono verificate nuove entrate o uscite si richiede di passare alla schermata successiva in cui si porranno quesiti relativi alle caratteristiche dei nuovi entrati e alle possibilità di contattare coloro che sono usciti dalla famiglia per poterli seguire durante l'arco dell'indagine.

La schermata sulla composizione familiare conterebbe anche la richiesta di informazioni circa la relazione di ciascun componente con il capo famiglia:

Composizione familiare al 31/12/99

NOME	RELAZIONE CON IL CAPOFAMIGLIA
Mario	Capofamiglia
Francesca	moglie
Giuseppe	figlio
Giorgio	figlio

Se la struttura familiare è rimasta invariata premere *return* per confermare.

Se la struttura familiare è cambiata premere F5*.

*accesso alla schermata in cui si chiede di indicare chi è uscito, perché e dove è andato, e chi è entrato.

Anagrafica

I dati che non richiedono la conferma nel tempo sono: il sesso, la data di nascita, il paese di nascita.

Ci sono poi quelle informazioni oggettive che delineano il profilo dell'individuo, hanno cioè un carattere strutturale, e hanno solo bisogno di eventuali aggiornamenti. Per questi dati basterebbe chiedere una conferma della situazione precedente, presentando a video l'ultimo *status* rilevato.

L'ipotesi potrebbe essere quella di verificare l'attuale parte strutturale del questionario della RTFL con una schermata in cui compaiono tutti i dati di carattere anagrafico tenendo inibita la possibilità di modificare l'anno di nascita, il sesso, il paese di nascita, se non su esplicito richiamo di una funzione (in caso di eventuali errori), e chiedendo conferma (modalità: modifica/conferma) di tutte le altre informazioni relative allo stato civile, al titolo di studio, alla cittadinanza, agli anni di residenza e all'assenza dal comune di residenza.

Per il titolo di studio e per gli anni di residenza si porrà il vincolo di una "non decrescita" nel tempo per verificare la coerenza delle risposte fornite da un'occasione all'altra.

Il vantaggio di una tale schermata risiede nell'evitare la richiesta di informazioni già acquisite in occasioni precedenti, diminuendo il numero dei quesiti a cui rispondere e, in generale, facilitando l'intervistato nella compilazione del questionario.

La controindicazione risiede nella possibilità che l'intervistato, per semplicità o stanchezza, confermi i dati anche laddove sono intervenuti dei cambiamenti introducendo una stabilità nelle stime dei flussi.

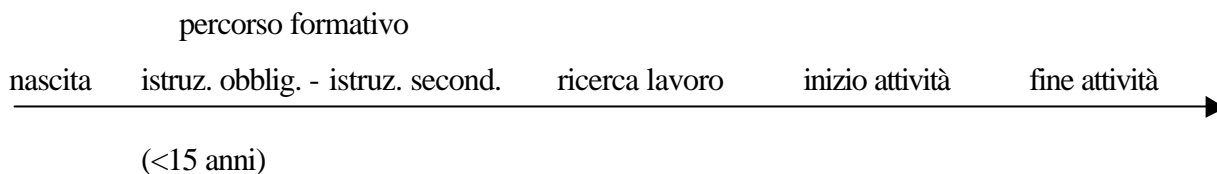
Tuttavia, per l'area anagrafica che precede la compilazione delle altre aree tematiche, questa soluzione potrebbe essere efficace in quanto l'intervistato è all'inizio del suo percorso e non risente della stanchezza relativa all'aver risposto già ad altri quesiti; nello stesso tempo, abbreviando l'onere della parte introduttiva si può prevedere un atteggiamento positivo dell'intervistato che avrà più cura nel rispondere alle domande specifiche dell'indagine.

Occupazione

Intervistando un individuo in più occasioni si deve tener conto che i percorsi che vengono delineati devono necessariamente seguire delle strutture logiche che permettono di identificare all'interno della vita lavorativa di ciascuno dei punti cardine la cui non verifica discrimina rispetto a una situazione di errore.

Dovendo rappresentare la vita di un individuo in termini lavorativi lungo un percorso rettilineo la si può supporre scandita dalla sequenza logica dei seguenti fatti:

Fig. 1 - Schema rettilineo



La mancata successione di questi semplici eventi costituisce un campanello di allarme per attivare controlli e verifiche.

Lo schema corredato da date di riferimento, permette all'intervistato di collocare temporalmente, nell'arco dell'indagine, i fatti di volta in volta rilevati.

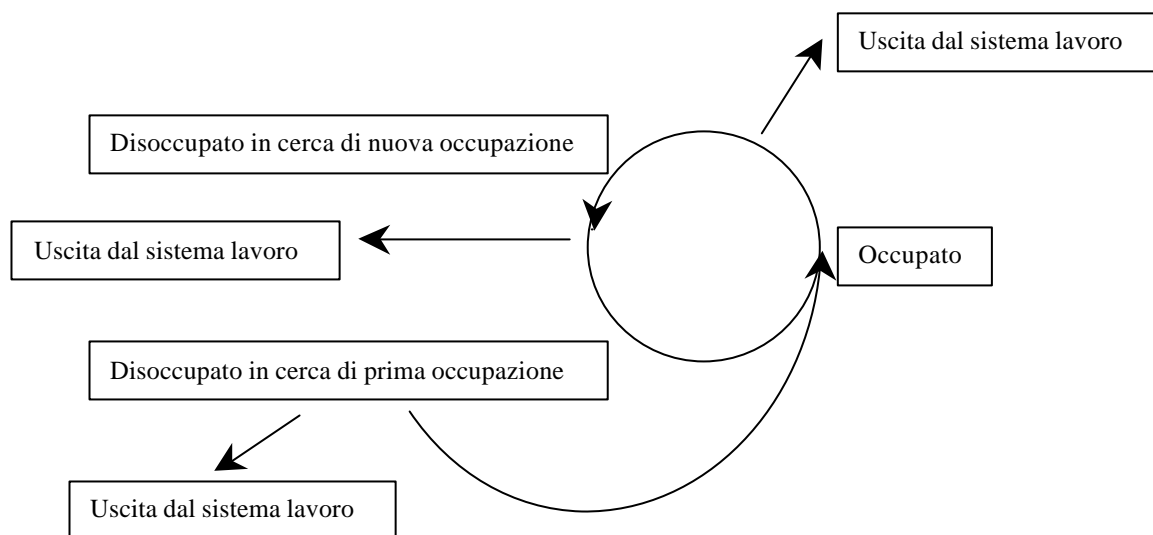
Le date che si possono ricavare dall'attuale questionario sulle forze di lavoro sono: la data di nascita, le date dei contatti (primo mese di contatto, tre mesi dopo, dodici mesi dopo, quindici mesi dopo), mese e anno d'inizio dell'attuale occupazione (dipendente o in proprio) – per gli occupati, data di fine dell'ultima occupazione – per i non occupati.

Inoltre, si possono ricavare riferimenti temporali incrociati attraverso i quesiti relativi alla durata in mesi dell'occupazione temporanea, durata in mesi della ricerca di lavoro, tempo (1 o più mesi) trascorsi dall'ultima azione di ricerca lavoro.

Peraltro, le condizioni inizio/fine attività possono verificarsi più di una volta nell'arco della vita eventualmente inframmezzate da periodi di ricerca di lavoro più o meno lunghi.

Nello stesso tempo alcuni "stati", risultanti dalle classificazioni ufficiali sull'occupazione, sono irripetibili e si verificano una sola volta durante l'*iter* lavorativo individuale, come l'essere "disoccupato in cerca di prima occupazione" da cui lo schema logico di riferimento risulta avere una struttura circolare il cui input è esterno e l'uscita può avvenire a partire da ciascuna condizione:

Fig. 2 - Schema circolare



Condizione professionale/non professionale

Da un controllo longitudinale sul quesito:

Qual è attualmente la sua condizione?

- Occupato
- Non occupato
 - Disoccupato in cerca di una nuova occupazione
 - In cerca di prima occupazione
 - Inizierà un'attività in futuro
 - Casalinga
 - Studente
 - Ritirato/a dal lavoro
 - Inabile al lavoro
 - In servizio di leva o in servizio civile sostitutivo
 - Altra condizione (benestante o simili)

deve risultare che la condizione “in cerca di prima occupazione” è incoerente rispetto alle modalità “occupato” e “disoccupato in cerca di una nuova occupazione” dichiarata nelle occasioni precedenti.

In riferimento alla stessa domanda, se il rispondente si dichiara “inabile al lavoro” non potrà rispondere di far parte di un'altra categoria lavorativa nelle successive *waves*, a meno che non si tratti persona appartenente a categorie protette. In questi casi possono essere approfonditi argomenti del tipo “*Ha partecipato a concorsi con posti riservati nel periodo intercorso dall'ultima intervista?*”, “*Ha vinto concorsi con posti riservati?*”, per capire se c'è stato realmente un cambiamento.

Posizione nella professione

Quale è la posizione nel lavoro?

Alle dipendenze come:

- Dirigente
- Direttivo –quadro
- Impiegato o intermedio
- Operaio subalterno ed assimilati
- Apprendista
- Lavorante a domicilio per conto di imprese

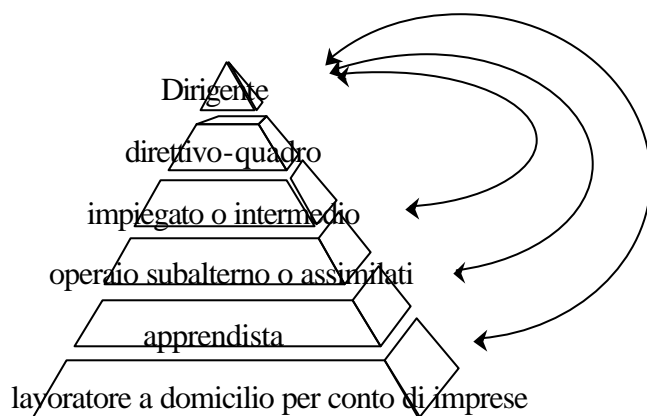
Autonomo come:

- Imprenditore
- Libero professionista
- Lavoratore in proprio
- Socio di cooperativa di produzione
- Coadiuvante

All'interno della categoria "lavoro alle dipendenze" si deve considerare che i cambiamenti rispetto a una situazione iniziale avvengono solitamente verso l'alto all'interno di una gerarchia di ruoli che difficilmente si muove in senso negativo per l'individuo.

Di conseguenza, l'aver dichiarato al primo contatto di essere un dirigente e nell'occasione successiva essersi definiti impiegati deve far scattare delle verifiche aggiuntive.

Fig. 3 Logica gerarchica nelle posizioni relative al lavoro dipendente



In questa situazione si potrebbe ricorrere al *probing*, avviando una serie di domande di approfondimento del tipo: "Ha cambiato società, ditta, ente negli ultimi mesi?", "Quale è il motivo di tale cambiamento?".

Infatti, si può supporre che la diversa condizione professionale dichiarata possa essere probabile se l'intervistato è alle dipendenze di un nuovo datore di lavoro e se il motivo del cambiamento è attribuibile a situazioni esterne non provocate dal dipendente: fallimento dell'azienda presso cui prestava precedentemente servizio, licenziamento, etc.; in caso contrario, se non c'è stato passaggio a nuovo lavoro o se la decisione del cambiamento è personale si può verosimilmente pensare che ci sia un errore nella risposta data, in quanto è facile intuire che cambi volontari di condizione lavorativa avvengano verso situazioni migliori e non peggiori.

Peraltro, alcune posizioni nella professione sono strettamente legate al titolo di studio per cui le verifiche devono includere il controllo su questa variabile anche per passaggi di condizione da livelli inferiori della scala gerarchica a livelli superiori: sono sospetti, comunque, salti di più livelli da una rilevazione all'altra.

Altrettanto improbabili risultano i passaggi da lavoratore dipendente a lavoratore autonomo.

Istruzione - Formazione

Per sviluppare controlli longitudinali, si può pensare al percorso formativo di un individuo in termini non decrescenti in funzione del tempo.

Si deve, inoltre, considerare che le interrelazioni tra l'istruzione e l'occupazione sono tali e si esplicano in modo tale da non permettere, innanzitutto, la corrispondenza dei concetti di istruzione e formazione e, in secondo luogo, la perfetta sequenza logico-temporale degli eventi relativi al percorso formativo.

E' necessario precisare che il concetto di istruzione è strettamente legato all' "istruzione scolastica" impartita nella scuola di qualsiasi grado (dalle elementari all'università).

Il concetto di formazione viene inteso, invece, in un'accezione più ampia, comprendendo il concetto di istruzione allargato attraverso una serie di eventi quali borse di studio, *stage* in azienda, praticantato, esperienza sul campo, ecc., venendo a coincidere con il concetto di "formazione professionale", cioè tutta quella "istruzione" che è orientata alla ricerca di lavoro, al mercato del lavoro.

A qualsiasi livello di istruzione, può inserirsi la formazione professionale, che non è necessariamente all'apice del percorso scolastico; nello stesso tempo, l'occupazione non conclude il percorso scolastico, né tantomeno, per quanto detto, quello formativo.

Sezione III: Indagine sui Consumi delle Famiglie

A cura di Daniela Cotzia

9 *Caratteristiche dell'indagine*

L'indagine sui consumi delle famiglie ha lo scopo di rilevare la struttura e il livello dei consumi secondo modalità di carattere economico-sociale e territoriale delle famiglie.

Oggetto delle rilevazioni sono i consumi privati, costituiti da tutti i beni e servizi acquistati o autoconsumati dalle famiglie per il soddisfacimento dei propri bisogni. Nella definizione rientrano anche i beni che provengono dal proprio orto o azienda agricola, i beni e servizi forniti dal datore di lavoro ai dipendenti a titolo di salario o per prestazioni di servizio, i fitti stimati delle abitazioni godute in proprietà.

Da gennaio 1997 l'Indagine sui Consumi delle Famiglie è stata completamente rinnovata dall'ISTAT. La ristrutturazione ha interessato tutte le fasi del processo di produzione dei dati, in particolare sono stati modificati:

- ✓ il disegno di campionamento;
- ✓ i questionari;
- ✓ le procedure di revisione, acquisizione e correzione dei dati.

Le modifiche apportate all'indagine hanno causato una rottura nella serie storica dei dati sulle spese per consumi pubblicati dall'ISTAT fino al 1996.

Attualmente, sono disponibili i primi risultati delle indagini condotte nel 1997 e nel 1998 su campioni annuali di circa 24.000 famiglie.

9.1 *Il disegno di campionamento e di rilevazione dei dati*

Nell'indagine sui consumi si utilizza un disegno a due stadi. Il primo stadio è costituito dai Comuni, il secondo dalle famiglie. Le unità del primo stadio sono stratificate, in ciascuna regione, in base alla tipologia del comune ed alla dimensione demografica.

I 103 comuni capoluogo di provincia, definiti autorappresentativi (AR), costituiscono uno strato a sé; gli altri comuni, indicati come non autorappresentativi (NAR), sono stratificati secondo la dimensione demografica. In particolare, gli strati vengono determinati quando la somma cumulata della popolazione risulta superiore ad una soglia che dipende:

- ✓ dalla frazione di campionamento trimestrale;
- ✓ dal numero medio di componenti della famiglia;
- ✓ dal numero minimo di interviste.

Sulla base di questa metodologia sono stati determinati 4 strati costituiti da un solo comune (Rovereto, Merano, Foligno e Termoli) e 127 strati da cui vengono estratti tre comuni campione che partecipano all'indagine rispettivamente il primo, il secondo ed il terzo mese di ogni trimestre.

Ogni mese partecipano all'indagine 2.000 famiglie estratte da 234 comuni. Nell'arco dell'anno i comuni interessati dall'indagine sono 488 e le famiglie intervistate 24.000.

Dalle liste anagrafiche dei comuni interessati vengono estratte in modo casuale le famiglie, suddivise in due liste:

- ✓ Elenco Base (Mod. ISTAT/BF/3);
- ✓ Elenco Suppletivo (Mod. ISTAT/BF/4).

La prima lista riporta i dati delle famiglie da intervistare, la seconda le famiglie che sostituiscono le prime in caso di rifiuto o impossibilità a collaborare all'indagine. La famiglia sostitutiva deve equivalere alla sostituita per numero di componenti della famiglia e per sezione di censimento di residenza.

Nelle indagini condotte negli anni precedenti le famiglie venivano chiamate a compilare un diario per 10 giorni consecutivi, nella nuova indagine tale periodo è stato ristretto a 7 giorni.

Nella nuova indagine, quindi, per ogni mese e per ogni regione vengono estratte casualmente due settimane di riferimento.

L'indagine è condotta su campioni di famiglie indipendenti in quanto per ogni Comune AR si seleziona un campione di famiglie che viene suddiviso in 12 gruppi, ciascuno dei

quali effettua l'indagine in un mese dell'anno. Per ogni comune NAR si procede all'estrazione di un campione di famiglie che viene suddiviso in 4 gruppi che partecipano all'indagine nei quattro mesi in cui il comune è coinvolto nell'indagine.

Ciascuno di tali gruppi viene a sua volta suddiviso in due sottogruppi. Le famiglie appartenenti al primo sottogruppo partecipano all'indagine nella prima settimana di riferimento, il secondo nella seconda settimana di riferimento del mese di rilevazione.

All'inizio del mese successivo, tutte le famiglie coinvolte nell'annotazione delle spese giornaliere sono sottoposte ad una intervista finale dove vengono chieste notizie riguardanti le spese della famiglia.

9.2 *Tecnica di rilevazione*

I dati sugli acquisti dei beni di consumo delle famiglie vengono acquisiti mediante tre questionari:

- ✓ il *Libretto degli Acquisti*, sul quale la famiglia campione annota giornalmente, per un periodo di sette giorni, le spese effettuate per l'acquisto di generi alimentari (pane, pasta, carne ecc.);
- ✓ il *Taccuino degli Autoconsumi*, sul quale la famiglia (di agricoltori, cacciatori, pescatori, ecc.) riporta giornalmente, sempre per sette giorni, le quantità consumate e la stima dei rispettivi costi di eventuali generi alimentari autoprodotti;
- ✓ il *Riepilogo delle Spese*, compilato dal rilevatore mediante intervista diretta alla famiglia nei primi sette giorni del mese successivo a quello di rilevazione; questo contiene le principali informazioni socio-demografiche della famiglia (relazione con l'intestatario della scheda anagrafica o persona di riferimento, età, titolo di studio, condizione professionale, attività economica, presenza in famiglia al momento dell'indagine) e le principali spese sostenute dalla stessa per beni e servizi e per l'acquisto di beni durevoli.

Le operazioni di rilevazione sono affidate agli uffici di statistica dei comuni che provvedono alle estrazioni delle famiglie campione, al reclutamento degli intervistatori ed alla spedizione di lettere di preavviso.

All'intervistatore è affidato il compito di contattare le famiglie, illustrare gli scopi dell'indagine e consegnare il libretto degli acquisti. Se l'intervistatore accerta la possibilità di consumo di generi alimentari autoprodotti, consegna il taccuino degli autoconsumi.

Nei primi giorni del mese successivo a quello di rilevazione l'intervistatore controlla e ritira il libretto degli acquisti ed, eventualmente, il taccuino degli autoconsumi e conduce l'intervista finale relativa alle altre spese sostenute nel mese e per alcune spese più rilevanti negli ultimi 3 mesi.

La compilazione dei questionari dell'indagine sui consumi si è rivelata spesso complessa. Le difficoltà riscontrate in prevalenza dai ricercatori sono:

- ✓ la reticenza a dichiarare alcune spese e soprattutto le informazioni relative al reddito;
- ✓ la difficoltà nel compilare il diario giornaliero;
- ✓ la difficoltà di mettersi in contatto con il rilevatore.

Per tale motivo con la nuova indagine sui consumi è stato istituito un "Numero verde" utilizzabile dalle famiglie come strumento di supporto sia nella fase iniziale, per ricevere informazioni sulle finalità e modalità dell'indagine, sia durante la settimana di rilevazione, per ricevere un'assistenza tecnica durante la compilazione del questionario.

9.3 Indagine sperimentale con sistema CATI

Nel periodo di giugno - dicembre 1997 e 1998, oltre all'indagine corrente condotta tramite la rete di rilevazione comunale, l'ISTAT ha effettuato un'indagine telefonica sperimentale sui consumi. Le famiglie interessate mensilmente dall'indagine sono circa 550 appartenenti a 105 differenti comuni.⁷

⁷ L'indagine è stata condotta con metodologia CATI dalle società private Metro R&C di Roma nel 1997 e dalla Pragma, sempre di Roma, nel 1998.

Le famiglie che entrano a far parte del campione sono state individuate nel seguente modo:

- ✓ si è proceduto all'estrazione casuale dei comuni facenti parte del campione ISTAT, con il vincolo che ciascun comune partecipasse all'indagine di controllo almeno due volte;
- ✓ sono state estratte con metodo casuale le famiglie tra quelle che risultavano aver partecipato all'indagine ISTAT.

Questa indagine ha avuto un duplice scopo. Da un lato, quello di verificare la qualità della rete di rilevazione comunale, dall'altro testare questionari aggiuntivi e verificare l'accuratezza delle interviste finali.

Nel periodo dal 15/6/99 al 30/10/99 sono state contattate circa 2.700 famiglie che avevano partecipato all'indagine sui consumi ed è emerso che l'8% delle interviste sono da considerarsi "mai svolte" ed il 20% non rispettano le procedure di indagine indicate dall'ISTAT.

10 Il fenomeno delle mancate risposte

Sulla base delle indagini condotte nei vari anni si è riscontrato che la numerosità effettiva dell'indagine, sulla quale si basano le stime, è sempre inferiore a quella teorica programmata. I motivi per cui si verifica possono essere ricondotti:

- ✓ al rifiuto di alcuni comuni campione di partecipare all'indagine (caduta di unità di 1° stadio);
- ✓ all'impossibilità di intervistare le famiglie selezionate dai comuni (caduta di unità di 2° stadio);
- ✓ alla scarsa qualità delle risposte fornite dalle famiglie coinvolte nell'indagine, che non permette l'utilizzo dei questionari ricevuti dall'ISTAT che vengono in tal caso annullati (caduta di 2° stadio).

Sia le cadute delle unità di 1° stadio (UPS) sia quelle delle unità di secondo stadio (USS) possono avere effetti distorcenti in fase di stima dei principali parametri di indagine.

Nei prossimi paragrafi verranno indicate le percentuali di caduta di unità di primo e secondo stadio, ricavate da documentazione interna fornita dal servizio CEF dell'ISTAT preposto alla conduzione dell'indagine (D. Grassi, 1998).

10.1 Le mancate risposte totali di primo stadio (UPS)

Nella tabella 1 vengono riportate le percentuali di caduta dei comuni e delle famiglie per gli anni 1996, 1997 e 1998.

Nelle indagini più recenti si riscontra una riduzione del numeri di comuni che rifiutano di partecipare all'indagine.

Tabella 1 Percentuale di cadute dei comuni e delle relative famiglie

Anno di riferimento	Comuni	Famiglie
1996	5,36	4,71
1997	4,56	2,93
1998	3,45	3,16

Nel momento dell'estrazione dei comuni campione il servizio CEF effettua una ricognizione con l'ausilio degli Uffici Regionali per verificare se nei comuni prescelti esistano situazioni che potrebbero causare un rifiuto di partecipare all'indagine.

Se si riscontrano delle difficoltà non sormontabili, si procede, nel caso di comuni non autorappresentativi, alla sostituzione con comuni appartenenti allo stesso strato.

Per i comuni autorappresentativi, non essendo possibile alcuna sostituzione, l'attività di organizzazione e di persuasione a collaborare all'indagine del servizio CEF e degli Uffici Regionali diventa fondamentale.

10.2 Le mancate risposte totali di secondo stadio (USS)

Nell'indagine sui consumi delle famiglie la selezione delle unità di secondo stadio viene effettuata in modo casuale sulla base delle liste anagrafiche comunali.

Le famiglie estratte entrano a far parte di due elenchi: l'elenco base, composto dalle famiglie da intervistare, e l'elenco suppletivo, da cui vengono estratte le famiglie sostitutive nell'ipotesi in cui alcune famiglie del primo elenco non possano far parte del campione.

Per tenere conto delle distorsioni prodotte dalla sostituzione, si richiede che le famiglie sostitutive abbiano le stesse caratteristiche delle famiglie sostituite. Le uniche due variabili prese in considerazione sono la numerosità della famiglia e l'appartenenza alla stessa sezione di censimento. Entrambe queste variabili sono correlate con la spesa media mensile.

Non sempre, però, la regola di sostituire le famiglie con famiglie di uguale ampiezza viene rispettata.

Le tabelle 2,3 e 4 riportano, rispettivamente, il numero di famiglie sostituite classificate in base all'ampiezza, la percentuale per riga e per colonna di sostituzione delle famiglie con famiglie di eguale ampiezza (A. Filippone e I. Foglia, 1997).

Tabella 2 Famiglie sostitutive e sostituite classificate in base all'ampiezza

Numero componenti Famiglia sostituita	Numero componenti famiglia sostitutiva						Totale
	1	2	3	4	5+	Non specificato	
1	635	234	143	106	29	84	1231
2	106	474	133	96	34	49	892
3	56	78	342	82	21	50	629
4	24	58	69	266	41	42	500
5+	11	23	29	66	77	8	214
Non specificato	47	44	29	28	13	2	163
Totale	879	911	745	644	215	235	3629

Tabella 3 Famiglie sostitutive e sostituite classificate in base all'ampiezza
(percentuali di riga)

Numero componenti Famiglia sostituita	Numero componenti famiglia sostitutiva						Totale
	1	2	3	4	5+	Non specificato	
1	51,58	19,01	11,62	8,61	2,36	6,82	100
2	11,88	53,14	14,91	10,76	3,81	5,49	100
3	8,90	12,40	54,37	13,04	3,34	7,95	100
4	4,80	11,60	13,80	53,20	8,20	8,40	100
5+	5,14	10,75	13,55	30,84	35,98	3,74	100
Non specificato	28,83	26,99	17,79	17,18	7,98	1,23	100
Totale	24,22	25,10	20,5	17,75	5,92	6,48	100

**Tabella 4 Famiglie sostitutive e sostituite classificate in base all'ampiezza
(percentuali di colonna)**

Numero componenti Famiglia sostituita	Numero componenti famiglia sostitutiva						Totale
	1	2	3	4	5+	Non specificato	
1	72,24	25,69	19,20	16,45	13,52	35,70	33,92
2	12,06	52,04	17,85	14,90	15,82	20,82	24,58
3	6,37	8,56	45,90	12,73	9,78	21,26	17,33
4	2,73	6,37	9,26	41,29	19,08	17,86	13,78
5+	1,25	2,53	3,89	10,25	35,84	3,40	5,90
Non specificato	5,35	4,83	3,89	4,35	6,05	0,85	4,49
Totale	100	100	100	100	100	100	100

La sostituzione di famiglie dell'elenco base con famiglie dell'elenco suppletivo viene documentata nel modello BF/5 compilato dal responsabile comunale, che deve annotare oltre alle caratteristiche delle famiglie il motivo di mancata risposta.

Il modello BF/5 così come predisposto induce ad una sottostima della mancata risposta in quanto non è stata prevista la possibilità di riportarvi le cadute attribuibili a famiglie dall'elenco suppletivo.

Per tale motivo, a partire dal 1998 è stato introdotto un ulteriore modello (BF/4/Bis) che viene impiegato nel caso di cadute di famiglie sostitutive.

Nella tabella 5 viene riportato il tasso complessivo di mancata risposta, calcolato come rapporto tra il numero di famiglie sostitutive ed il numero complessivo delle famiglie effettivamente rilevate (D. Grassi, 1998).

Tabella 5 Famiglie base sostituite sul totale delle intervistate

Regione	Vecchia indagine		Nuova indagine	
	1995	1996	1997	1998 ⁸
Piemonte	14,5	12,7	18,9	19,3
Valle d'Aosta	4,4	10,2	16,9	18,6
Lombardia	14,6	16,1	24,1	20,1
Trentino A.A.	10,7	12,3	17,5	18,0
Veneto	6,6	8,6	14,6	15,7
Friuli V. Giulia	19,9	16,0	26,8	25,0
Liguria	9,8	6,5	14,1	16,6
Emilia Romagna	12,7	12,5	19,1	20,9
Toscana	9,5	10,8	16,7	21,0
Umbria	5,6	6,0	15,6	9,8
Marche	2,9	3,0	11,1	6,3
Lazio	7,2	7,0	9,2	11,6
Abruzzo	4,0	7,0	7,6	7,0
Molise	7,5	6,5	13,0	10,7
Campania	4,5	5,2	15,0	11,9
Puglia	3,8	4,9	11,8	14,1
Basilicata	1,6	2,7	7,7	6,1
Calabria	9,4	8,5	16,1	11,2
Sicilia	6,3	8,3	14,3	12,8
Sardegna	7,9	10,0	12,1	15,9
Italia	8,7	9,2	15,7	15,3

Nel passaggio dalla vecchia alla nuova indagine si riscontra un incremento complessivo della mancata risposta. Il servizio CEF interpreta positivamente questa variazione, associandola ad una diminuzione delle interviste non utilizzabili (il 28% nella vecchia indagine il 16% nella nuova).

Come indicato in precedenza nel modello BF/5 viene registrato anche la motivazione che determina la mancata partecipazione di una famiglia all'indagine.

⁸ Stima per il 1998 (percentuale famiglie dell'elenco base sostituite)

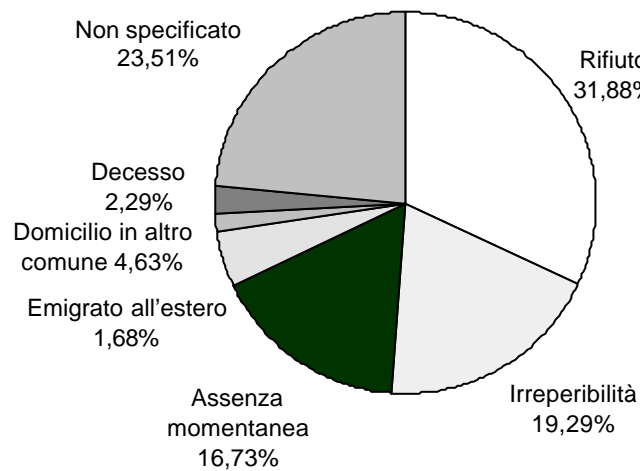
Le motivazioni contemplate sono le seguenti:

- ✓ rifiuto;
- ✓ irreperibilità;
- ✓ decesso;
- ✓ assenza momentanea;
- ✓ domicilio in altro comune;
- ✓ emigrato all'estero;
- ✓ non specificato.

Dall'analisi delle motivazioni registrate nel modello BF/5 è possibile verificare che il motivo principale per cui le famiglie decidono di non partecipare all'indagine è il rifiuto (31,88%).

Il seguente grafico indica le principali cause di caduta di secondo stadio.

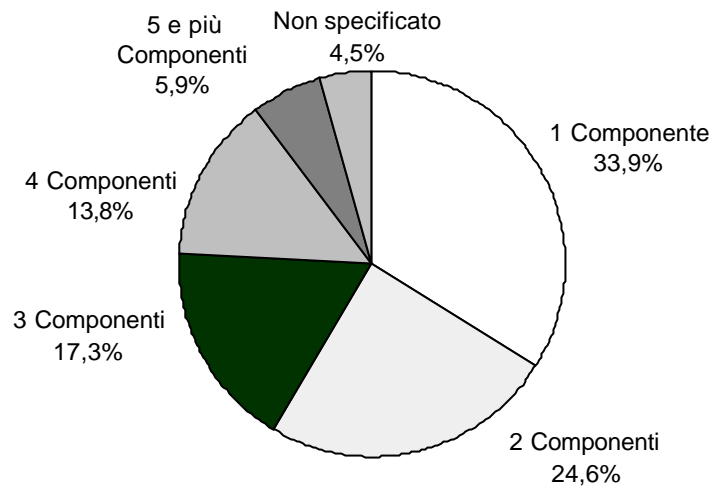
Grafico 1 Famiglie sostituite in base ai motivi della mancata risposta (1997)



Le ragioni di mancata risposta si differenziano anche in base al numero di componenti.

Il grafico 2 mostra come le sostituzioni decrescano all'aumentare dell'ampiezza familiare, passando dal 39,9% delle famiglie composte da un unico individuo, al 5,9% delle famiglie con almeno 5 componenti.

Grafico 2 Famiglie sostituite in base al numero dei componenti (1997)



Per maggior dettaglio i grafici 3 e 4 riportano, rispettivamente, le famiglie che sono state sostituite per assenza momentanea e per rifiuto in base al numero dei loro componenti.

Grafico 3 Famiglie sostituite per assenza momentanea in base al numero dei componenti (1997)

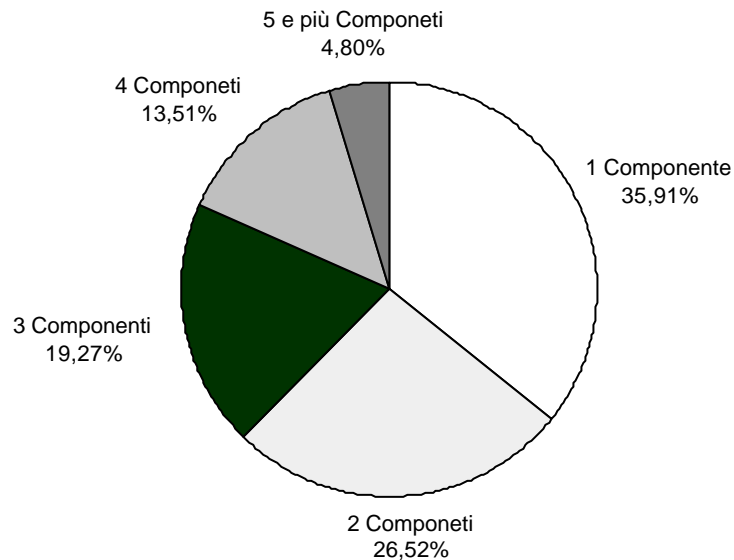
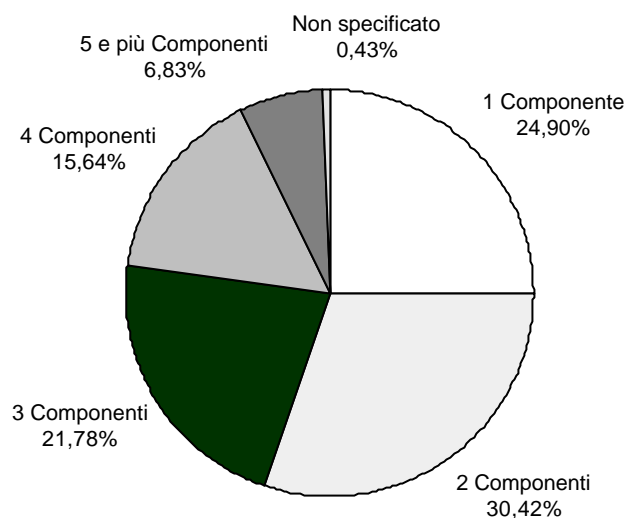


Grafico 4 Famiglie sostituite per rifiuto in base al numero dei componenti (1997)



Per l'indagine del 1997 sono stati studiati degli altri indicatori utili per la comprensione del fenomeno delle mancate risposte (A. Filippone e I. Foglia, 1997).

Per una chiara lettura degli indicatori si tenga presente che con:

- ✓ N_1 viene indicato il numero di famiglie che hanno *rifiutato* l'intervista;
- ✓ N_2 viene indicato il numero di famiglie che risultano *irreperibili*
- ✓ N_3 viene indicato il numero di famiglie che risultano *decedute*;
- ✓ N_4 viene indicato il numero di famiglie *momentaneamente assenti*;
- ✓ N_5 viene indicato il numero di famiglie *domiciliate in altro comune*;
- ✓ N_6 viene indicato il numero di famiglie *emigrate all'estero*;
- ✓ N_7 viene indicato il numero di famiglie per cui *non viene specificato* il motivo di mancata partecipazione all'indagine;
- ✓ N_L viene considerato il numero di famiglie *erroneamente immesse in lista* ed è pari a
$$N_L = N_2 + N_3 + N_5 + N_6$$
- ✓ N_{NR} viene indicato il numero delle famiglie *non rispondenti*;
- ✓ N_R viene indicato il numero di famiglie *rispondenti*;

✓ N viene indicata la *numerosità campionaria*.

Sulla base di questi parametri vengono definiti i seguenti tassi⁹:

↳ *Tasso di non risposta totale:* N_{NR}/N

che fornisce una sintesi dell'entità del fenomeno di mancata risposta preso nel suo complesso;

↳ *Tasso di errore di lista:* N_L/N

che individua i casi in cui la mancata risposta è riconducibile ad errori presenti nelle liste anagrafiche comunali;

↳ *Tasso di aggiornamento di lista:* $(N_3+N_5+N_6)/N_L$

che indica la porzione di errori di lista dovuti a decessi, cambi di domicilio ed emigrazione all'estero sul totale degli errori di lista;

↳ *Tasso di accuratezza di lista:* N_2/N_L

che è dato dal complemento ad uno del tasso di aggiornamento di lista e rappresenta la quota degli errori di lista dovuti all'irreperibilità della famiglia sul totale degli errori di lista

↳ *Tasso lordo di non risposta:* $(N_1+N_4)/N$

che individua il numero di famiglie che rifiutano di collaborare o che sono momentaneamente assenti sul totale delle famiglie appartenenti al campione;

↳ *Tasso netto di non risposta:* $(N_1+N_4)/(N_R+N_1+N_4)$

che rappresenta il numero di famiglie che rifiutano di collaborare o che sono momentaneamente assenti sul totale delle famiglie rintracciate dal rilevatore, indipendentemente dall'esito del tentato contatto;

↳ *Tasso netto di rifiuto:* $N_1/(N_R+N_1)$

che indica la proporzione delle famiglie contattate dal rilevatore che non hanno voluto partecipare all'indagine;

↳ *Tasso netto di non intervista:* $N_4/(N_R+N_1+N_4)$

⁹ I tassi riportati di seguito sono stati costruiti sulla base delle indicazioni di Platek R. e Gray G.B. (1986)

che rappresenta la quota delle famiglie momentaneamente assenti sul totale delle famiglie rintracciate dal rilevatore.

Nella tabella 6 è possibile leggere i tassi di non risposta totali per regione. Le tabelle 7 e 8 riportano, rispettivamente, i restanti tassi suddivisi per regione e per numero di componenti della famiglia.

Tabella 6 Percentuali di non risposta totali per regione

<i>Regione</i>	<i>Non risposta</i>	<i>Campione</i>	<i>Tasso di non risposta totale</i>
	(N_R)	(N)	(N_R/N)
Piemonte	274	1452	18,87
Valle d'Aosta	107	632	16,93
Lombardia	584	2427	24,06
Trentino A.A.	205	1173	17,48
Veneto	216	1474	14,65
Friuli V. Giulia	207	773	26,78
Liguria	122	865	14,10
Emilia Romagna	269	1408	19,11
Toscana	216	1292	16,72
Umbria	101	648	15,59
Marche	89	804	11,07
Lazio	152	1658	9,17
Abruzzo	62	817	7,59
Molise	83	641	12,95
Campania	260	1737	14,97
Puglia	148	1250	11,84
Basilicata	47	612	7,68
Calabria	149	923	16,14
Sicilia	225	1571	14,32
Sardegna	113	934	12,10
Italia	3629	23091	15,72

Tabella 7 Percentuali di non risposta per regione

Regione	<i>Tasso di errore di lista</i>	<i>Tasso di aggiornam. di lista</i>	<i>Tasso di accuratezza di lista</i>	<i>Tasso lordo di non risposta</i>	<i>Tasso netto di non risposta</i>	<i>Tasso netto di rifiuto</i>	<i>Tasso netto di non intervista</i>
	N_L/N	$(N_3+N_5+N_6)/N_L$	N_2/N_L	$(N_I+N_4)/N$	$(N_I+N_4)/(N_R+N_I+N_4)$	$N_I/(N_R+N_I)$	$N_4/(N_R+N_I+N_4)$
Piemonte	3,58	30,77	69,23	11,91	12,81	9,10	4,07
Valle d' Aosta	4,91	29,03	70,97	8,54	9,33	5,75	3,80
Lombardia	6,67	18,52	81,48	12,53	14,16	8,67	6,01
Trentino A.A.	5,29	29,03	70,97	9,04	9,87	7,46	2,61
Veneto	4,00	42,37	57,63	6,04	6,61	5,20	1,48
Friuli V. Giulia	9,31	22,22	77,78	13,20	15,27	10,02	5,84
Liguria	3,35	27,59	72,41	7,75	8,27	3,38	5,06
Emilia Romagna	3,84	20,37	79,63	9,80	10,81	7,92	3,13
Toscana	3,95	23,53	76,47	10,53	11,22	8,19	3,30
Umbria	3,09	40,00	60,00	8,95	9,59	6,66	3,14
Marche	2,24	44,44	55,56	5,22	5,55	4,16	1,45
Lazio	4,83	33,75	66,25	2,29	2,46	1,76	0,71
Abruzzo	1,47	58,33	41,67	4,04	4,19	2,71	1,52
Molise	3,74	54,17	45,83	7,18	7,62	5,26	2,48
Campania	3,86	50,75	49,25	4,15	4,65	2,19	2,52
Puglia	2,56	31,25	68,75	7,12	7,47	6,05	1,51
Basilicata	1,96	75,00	25,00	4,58	4,72	2,42	2,63
Calabria	3,79	28,57	71,43	7,69	8,40	5,72	2,84
Sicilia	6,30	30,30	69,70	4,96	5,48	3,93	1,62
Sardegna	4,39	26,83	73,17	4,28	4,65	3,86	0,81
Italia	4,38	30,83	69,17	7,64	8,31	5,61	2,86

Tabella 8 Percentuali di non risposta per numero di componenti

Numero componenti	<i>Tasso di errore di lista</i>	<i>Tasso di aggiornam. di lista</i>	<i>Tasso di accuratezza di lista</i>	<i>Tasso lordo di non risposta</i>	<i>Tasso netto di non risposta</i>	<i>Tasso netto di rifiuto</i>	<i>Tasso netto di non intervista</i>
	N_L/N	$(N_3+N_5+N_6)/N_L$	N_2/N_L	$(N_I+N_4)/N$	$(N_I+N_4)/(N_R+N_I+N_4)$	$N_I/(N_R+N_I)$	$N_4/(N_R+N_I+N_4)$
1	10,10	32,34	67,66	10,87	12,88	7,76	5,55
2	4,17	29,02	70,98	9,54	10,26	7,28	3,22
3	2,46	31,30	68,70	6,94	7,29	5,10	2,31
4	2,48	26,32	73,68	4,90	5,12	3,58	1,60
5+	2,35	33,33	66,67	4,98	5,24	3,89	1,41

11 La revisione e correzione dei dati nell'Indagine sui Consumi delle Famiglie

Le operazioni di controllo nell'indagine sui consumi delle famiglie si basano sia su metodi di controllo manuale, effettuati mediante revisori esperti, sia su procedure informatiche di tipo deterministico e probabilistico.

La revisione dei dati prevede la definizione del piano di compatibilità del questionario, ultimato il quale viene individuato il complesso di regole di correzione riferite sia alle variabili qualitative sia alle variabili quantitative.

Le procedure applicate nell'immissione controllata dei dati e la loro correzione sono le seguenti:

- ↳ **Revisione manuale**, che prevede il controllo da parte dei revisori dei questionari e, laddove si riscontri un'incompatibilità tra due o più variabili, il ricontattamento telefonico della famiglia intervistata (C. Ceccarelli, D. Grassi, N. Pannuzzi, 1997).
- ↳ **Acquisizione controllata**, che, essendo effettuata dagli stessi revisori esperti, è strettamente connessa con quella di revisione. In alcuni casi, infatti, le incoerenze di tipo formale vengono risolte attraverso i controlli inseriti nel programma di immissione (BLAISE), altre vengono segnalate e trattate opportunamente secondo criteri standard del revisore (S. Macchia, 1999).
- ↳ **Controllo e correzione**, che a sua volta viene suddivisa in (D. Grassi, 1999):
 - ✓ *controllo e correzione deterministica*, effettuata con l'ausilio di un programma SAS, verifica il rispetto delle regole di revisione per le eventuali incoerenze non eliminate in fase di acquisizione controllata;
 - ✓ *analisi delle variabili qualitative*, vengono effettuate le correzioni mediante due diversi programmi SCIA (Sistema Correzione Imputazione Automatica) ed alcuni programmi sviluppati in SAS;
 - ✓ *analisi delle variabili quantitative*, circa 2/3 delle 308 voci di spesa vengono controllate e corrette probabilisticamente attraverso il software RIDA (Ricostruzione delle Informazioni con Donazione Automatica). Per le variabili quantitative sulle quali

non è possibile applicare RIDA a causa delle basse frequenze, i valori anomali vengono riportati all'interno dell'intervallo di variazione mediante una procedura di attribuzione casuale di un valore compreso tra il massimo ed il minimo della variabile. Tale procedura prende il nome di “*massimi e minimi randomizzati*”.

Ne paragrafi successivi verranno descritte le fasi precedentemente elencate.

12 Regole di revisione manuale utilizzate dai revisori dell'Istat

Al revisore spetta il compito di effettuare le verifiche preliminari sui dati per non rischiare di effettuare errori di immissione non deducibili informaticamente.

In particolare all'apertura dei pacchi deve verificare la presenza dei modelli:

- ✓ due o più modelli BF/5 (Elenco delle famiglie da Intervistare);
- ✓ BF/2 (Riepilogo delle Spese Familiari);
- ✓ BF/1 (Libretto degli Acquisti);
- ✓ se necessario BF/1/A (Taccuino degli Autoconsumi).

Altre verifiche che richiedono necessariamente revisioni manuali sono:

- ✓ che il numero di modelli pervenuti coincida con il numero di modelli teorici riportato sulla posizione di verifica.
- ✓ che il numero di modelli BF/5 sia pari al numero di rilevatori utilizzati per ogni periodo di rilevazione;
- ✓ che, nel caso di utilizzo di un timbro nell'intestazione del modello BF/2 per apporre i dati del Comune, non siano invertiti i codici della Provincia e del Comune;
- ✓ che, nel caso sia dichiarata la compilazione del taccuino degli autoconsumi, il taccuino stesso sia stato effettivamente inserito nel pacco;
- ✓ che sia stata apposta la firma sia dell'intervistatore che del responsabile Comunale e che la data di consegna dei modelli corrisponda ai giorni del mese successivo a quello di rilevazione.

Una volta stabilito qual è il numero effettivo dei modelli per ogni Comune campione, viene controllata la corretta compilazione del Comune dei frontespizi dei modelli di rilevazione, ed in particolare:

- ✓ viene controllato se i codici dei frontespizi dei modelli relativi alla stessa famiglia coincidano;
- ✓ viene verificato sulla base del modello BF/5:
 - il codice intervistatore;
 - il numero d'ordine della famiglia;
 - se la famiglia appartiene all'elenco base o suppletivo
 - il periodo di rilevazione;
- ✓ viene verificato, inoltre, in presenza di una famiglia suppletiva o di una incompleta rilevazione che siano correttamente compilati i relativi campi del modello BF/5 e la coerenza della sostituzione (numero componenti e ubicazione dell'abitazione).

Nel caso in cui si verificassero delle incongruenze o dati mancanti, è compito del revisore ricontattare il Comune per verificarne i motivi.

Dopo aver revisionato i modelli di una famiglia campione, il revisore deve apporre la firma sulla copertina del modello BF/2 relativo alla famiglia stessa.

13 Acquisizione controllata dei dati - CADI

Come spesso avviene nelle indagini statistiche i dati rilevati sono soggetti ad errori, che possono presentarsi sia in fase di compilazione sia in fase di registrazione.

Nell'indagine sui consumi delle famiglie per minimizzare gli errori che si presentano in fase di registrazione si ricorre all'acquisizione controllata dei dati in modalità CADI (*Computer Assisted Data Input*).

Gli errori effettuati in fase di registrazione vengono classificati come:

- ✓ errori di *campo di variazione* ossia inserimento di un valore esterno al campo di definizione della variabile;

- ✓ errori di *rotta* riguardanti il cammino da seguire nella somministrazione delle domande di un questionario;
- ✓ errori di *coerenza* tra quanto viene risposto ad un quesito rispetto a quanto è stato dichiarato precedentemente in un altro quesito.

L'acquisizione controllata in modalità CADI consente di inserire controlli sui valori delle variabili sia qualitative sia quantitative. Il software utilizzato per l'acquisizione controllata dei dati nell'indagine sui consumi delle famiglie è il BLAISE.

BLAISE permette di impedire la registrazione di un dato errato (controllo *hard*) o di inserire il dato errato previa segnalazione di errore (controllo *soft*).

Qualora il software di controllo venga usato direttamente in fase di rilevazione dei dati l'impostazione è di tipo *hard* in quanto la correzione è immediata.

Applicando il CADI, invece, il trattamento degli errori non può essere così rigido, in quanto il dato errato è stato ormai acquisito in fase di rilevazione e non si può quindi inibire sempre in modo assoluto la registrazione.

13.1 Il sistema BLAISE

Il programma informatico BLAISE, sviluppato dall'Istituto nazionale di statistica olandese (Statistics Netherlands), è un sistema che permette l'applicazione delle tecniche CAPI e CATI, ma che può essere utilizzato per l'acquisizione controllata dei dati e la correzione interattiva di questi.

BLAISE opera in ambiente DOS dei personal computer, anche collegati in LAN, e gode di una larga diffusione, soprattutto all'interno degli istituti nazionali di statistica. Uno dei motivi di tale diffusione risiede nella grande facilità di sviluppo di programmi utilizzabili sia per applicazioni CAPI/CATI sia per il data entry.

La sintassi di BLAISE permette, una volta definita la struttura del questionario ed i quesiti che ne fanno parte, di associarvi delle regole di controllo, ognuna delle quali viene applicata ogni volta che il cursore giunge all'ultima delle variabili interessate. Ad esempio:

ANNO_RILEVAZIONE - ANNO_NASCITA≥23) And TITOLO_STUDIO = 1

“se la persona ha conseguito il dottorato di ricerca, deve avere almeno 23 anni”

In questa regola, che stabilisce un condizione di correttezza, vengono coinvolte tre variabili:

- ✓ l'anno cui si riferiscono i dati
- ✓ l'anno di nascita della persona
- ✓ il suo titolo di studio.

Se le variabili compaiono in questa sequenza nel questionario, ogni volta che si giunge ad inserire il valore TITOLO_STUDIO il sistema valuta se la regola è rispettata. In tal caso, viene proposto all'operatore il messaggio racchiuso tra virgolette, unitamente all'elenco delle tre variabili e dei valori che assumono nel questionario. L'operatore può scegliere quale/i variabile/i correggere e quale/i valore/i assegnare ad essa/e, oppure può decidere di ignorare la segnalazione di errore, proseguendo nell'attività di memorizzazione dati (controllo *soft*).

In pratica il programma BLAISE è composto da due sezioni principali:

- ✓ FIELDS, con cui si definiscono le variabili e le loro caratteristiche;
- ✓ RULES, sono le regole che rilevano gli errori presenti nei valori delle variabili e attivano i messaggi che compaiono sullo schermo (es.: “*se la persona ha conseguito il dottorato di ricerca, deve avere almeno 23 anni*”)

Le regole di tipo *hard* sono indicate con il termine CHECK, mentre per una gestione *soft* dell'errore si utilizzano regole di tipo SIGNAL.

Una data regola viene valutata dal sistema non appena terminato l'inserimento del valore dell'ultima variabile coinvolta. Se la regola non è rispettata, il sistema propone all'operatore un definito messaggio di errore unitamente all'elenco delle variabili coinvolte e dei valori registrati in esse. Se la regola è di tipo CHECK l'operatore è obbligato a rimuovere l'errore per continuare l'inserimento dei dati. Se la regola è di tipo SIGNAL l'operatore può scegliere tra rimuovere l'errore e ignorare l'inserimento dei dati.

Nell'ambito dell'indagine sui consumi delle famiglie del 1997 la gestione degli errori ha seguito le seguenti regole:

- ✓ le *rotte* già previste sul modello cartaceo sono state trattate in modalità *hard*;
- ✓ i *quesiti di tipo tabellare*, che prevedono il riempimento di una colonna subordinatamente al fatto che alla precedente sia stato risposto “SI”, sono stati trattati in modalità *hard*;
- ✓ le regole di *coerenza tra quesiti di modelli diversi* sono state trattate in modalità *hard*;
- ✓ le regole di *coerenza tra quesiti nell’ambito dello stesso modello* sono state trattate in modalità *soft*;
- ✓ i controlli di quadratura tra gli importi parziali rispetto al totale sono stati tratti in modalità *soft*;
- ✓ i controlli sulle variabili quantitative di spesa rispetto ai limiti massimi e minimi precedentemente stimati sono trattati in modalità *soft*.

Si sviluppano con BLAISE tre moduli, uno per ciascun modello dell’indagine (Riepilogo delle spese, Libretto degli acquisti e Taccuino degli autoconsumi).

I dati di ciascuna famiglia campione sono memorizzati su un unico record. I dati vengono in seguito esportati dall’archivio di BLAISE in un file in formato ASCII.

Lo sviluppo dell’applicazione per l’indagine sui consumi delle famiglie è stato curato dal Servizio Studi Metodologici e successivamente modificato dal Servizio CEF dell’Istat.

Le operazioni di registrazione sono attualmente effettuate per la gran parte presso il servizio CEF/A e per la restante parte presso il servizio adibito alla registrazione dati, DINF/C. Il servizio CEF/A utilizza il pacchetto di acquisizione sia per la registrazione che come supporto alla revisione, mentre il servizio DINF/C procede alla sola registrazione di modelli preventivamente revisionati dal servizio CEF/A. I dati presso le diverse postazioni di lavoro sono successivamente accorpati e trasferiti in ambiente mainframe VM-CMS (è già prevista la conversione in ambiente UNIX). In ambiente VM-CMS un predisposto programma SAS provvede ad eseguire il compattamento delle spese del libretto dal dettaglio quotidiano al riepilogo settimanale.

13.2 RULES impostate in BLAISE nell'indagine sui Consumi delle Famiglie

Si riportano di seguito le principali regole che vengono impostate nel programma BLAISE per i tre questionari utilizzati nell'indagine.

L'insieme dei vincoli posti in fase di immissione di dati sono riportati in allegato. (Allegato C).

13.2.1 Riepilogo delle spese (Mod. BF/2)

Per il modello BF/2 facciamo riferimento alle regole della sezione dedicata ai componenti delle famiglie.

Le regole di tale sezione sono da considerarsi in prevalenza di tipo soft. In particolare, per quanto si riferisce alla relazione con la persona di riferimento, al sesso, all'anno di nascita, allo stato civile, al titolo di studio, alla condizione professionale, alla branca di attività economica, alla posizione nella professione si riporta fedelmente quanto dichiarato nel questionario.

Si apportano cambiamenti (regole hard) solo nei casi in cui:

- ✓ il sesso della persona di riferimento e del coniuge risulta uguale (viene cambiato solo quello del coniuge);
- ✓ se un figlio risultante militare non è maschio (viene imposto maschio);
- ✓ se l'anno di nascita del figlio è inferiore a quello di entrambi i genitori (viene lasciato in bianco);
- ✓ se una condizione professionale non implica la presenza della branca e della posizione professionale (viene lasciata in bianco la branca e la posizione professionale);
- ✓ se la presenza non risulta dichiarata (viene imposta SI come risposta).

13.2.2 Libretto degli acquisti (Mod. BF/1)

La maggioranza delle anomalie riscontrate nella registrazione dei dati relativi al libretto degli acquisti sono di difficile interpretazione e, prima di effettuare una correzione, i revisori

sono invitati a consultarsi con i ricercatori del servizio CEF. Le regole di revisione sono quindi di tipo soft.

I problemi riscontrati in fase di revisione si possono ricondurre a:

- ✓ errori di collocazione della spesa;
- ✓ valori di spesa annotati non riconducibili ad una normale conduzione familiare.

Poiché il libretto degli acquisti viene compilato direttamente dai componenti della famiglia, è più facile che le spese siano collocate in posizione non corretta a differenza di quanto avviene nel modello BF/2 compilato a cura del rilevatore.

Le ragioni che, in prevalenza, determinano un errore di collocazione dei consumi sono:

- ✓ che la spesa sia relativa ad altra spesa nella stessa sezione;
- ✓ che la spesa sia relativa ad altra spesa in altra sezione del Mod. BF/1;
- ✓ che la spesa sia relativa ad una spesa da riportare sul Mod. BF/2

Questi errori possono risultare non evidenti in fase di revisione e, quindi, non soggetti a correzione.

Quando la spesa viene spostata in altra voce della stessa sezione o in altra sezione del Mod. BF/1 o sul Mod. BF/2, può accadere che vi si trovi già segnata una spesa. In questo caso procedono sommando le due spese.

Può, inoltre, accadere che una voce di spesa sia registrata due volte. Ad esempio, si può trovare una voce di spesa non corrente riportata erroneamente sul Mod. BF/1 e correttamente sul Mod. BF/2 in fase di intervista finale. In questo caso il revisore non apporta nessuna modifica senza consultarsi con i ricercatori preposti all'indagine.

Ancora più difficili da interpretare sono gli errori di annotazione della spesa, che possono essere ricondotti ad:

- ✓ errori di registrazione (si può inserire uno 0 in più o in meno);
- ✓ errori concettuali (in alcuni casi le famiglie non riportano nel modello le spese giornaliere, ma il consumo giornaliero di un prodotto).

Spesso non è possibile correggere questi errori in fase di immissione di dati e si procede alla loro correzione con le procedure automatiche (RIDA).

13.2.3 Taccuino degli autoconsumi (Mod. BF/1A)

Al termine della settimana di rilevazione è compito del rilevatore sommare le quantità e le spese giornaliere, che devono essere riportate nella colonna predisposta.

I controlli che si devono fare in fase di immissione di dati sono:

- ✓ verifica della presenza delle somme e/o della correzione del calcolo;
- ✓ verifica della corretta collocazione dei generi alimentari.

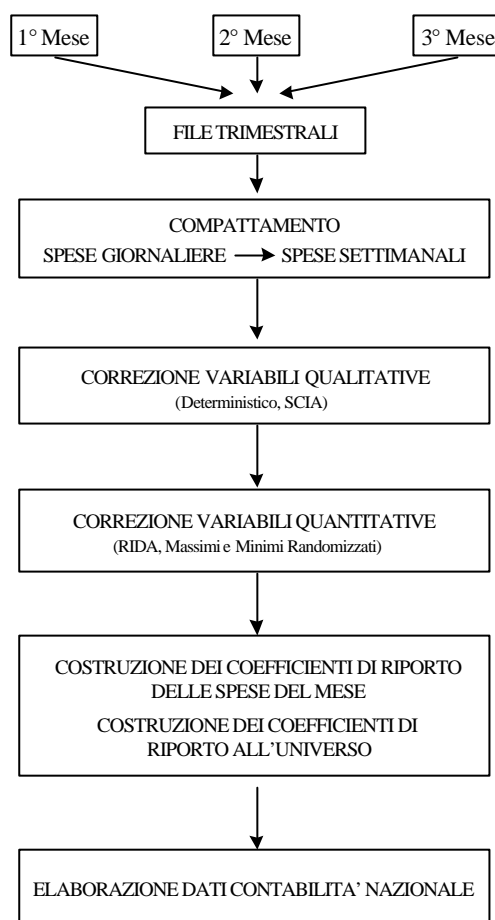
Se nel primo caso la correzione è immediata, nel secondo può risultare complessa ed, in molti casi, prede che alcune spese riportate nella voce “Altri Generi (da specificare)” siano copiate nelle voci prestampate del modello e cancellate della voce di spesa “Altri generi”.

14 *Controllo e correzione dei dati*

La correzione dei dati anomali immessi richiede l'utilizzo di una collezione di procedure, sia per l'individuazione dei dati che devono esservi sottoposti, verifica effettuata mediante procedure deterministiche che si avvalgono di programmi SAS, sia per la correzione delle variabili qualitative, realizzata mediante il software SCIA, sia per la correzione delle variabili quantitative, che viene attuata mediante due distinte procedure: RIDA e la metodologia denominata dei “Massimi e Minimi Randomizzati”.

Il seguente schema mette in luce in modo sintetico le varie fasi cui i dati dell'indagine sui consumi delle famiglie sono sottoposti.

Grafico 5 *Fasi di correzione dei dati per la determinazione della spesa media mensile per i consumi delle famiglie*



Nei paragrafi successivi verranno riprese le procedure di correzione delle variabili qualitative e quantitative, descritte nel grafico precedente, per chiarirne il funzionamento.

14.1 *Correzione delle variabili qualitative*

14.1.1 *Descrizione del Sistema di Correzione ed Imputazione Automatica (SCIA)*

SCIA è stato sviluppato in Istat a partire dal 1990 e attualmente è disponibile nelle versioni MVS/CMS su main frame IBM, AIX/RISC600 e UNIX/HP.

Il maggior utilizzo di questo algoritmo per la correzione delle variabili qualitative è dovuto ai seguenti vantaggi:

- ✓ eliminazione totale di scrittura di codice;
- ✓ facilità di definizione del piano di controllo;
- ✓ totale automatismo dell'applicazione.

L'algoritmo SCIA è un sistema per la correzione probabilistica delle variabili qualitative che fa riferimento alla metodologia di Fellegi-Holt (I. P. Fellegi, D. Holt, 1976).

Tale metodologia si basa su regole di incompatibilità che vengono chiamate *edit*. Un edit è costituito da un'insieme di condizioni poste sui valori delle variabili del record. L'edit è attivato da un dato record quando sono verificate simultaneamente tutte le condizioni in esso definite.

Gli edit devono essere in forma normale, cioè devono esprimere solo condizioni specifiche come intersezione (AND) tra due o più asserzioni. Ad esempio:

Sesso = femmina AND Professione = carabiniere - Errore

Età <15 anni AND Stato Civile = no celibe - Errore

Un edit è sempre riconducibile in forma normale, ad esempio:

Sesso = femmina AND Età <18 anni AND Professione = carabiniere

può essere convertito in:

Sesso = femmina AND Professione = carabiniere - Errore

Età <18 anni AND Professione = carabiniere - Errore

Sulla base del piano di incompatibilità deve essere definito l'insieme degli edit tenendo sempre presente che:

- ✓ un record per essere esatto non deve attivare una regola;
- ✓ le regole o edit descrivono incompatibilità tra le variabili;
- ✓ le regole vanno espresse in forma normale;
- ✓ è indifferente l'ordine delle regole, il sistema le verifica tutte senza alcuna gerarchia.

Essendo SCIA un sistema probabilistico viene definita esclusivamente l'incompatibilità tra le variabili, senza specificare cosa fare se si verifica l'incompatibilità. Tutte le regole sono, quindi, del tipo:

SE.....E.....E.....

e non

SE.....E.....ALLORA.....

Sulla base dell'insieme delle regole definito in fase di programmazione del sistema, l'algoritmo controlla i dati grezzi e separa i record esatti da quelli errati.

SCIA, una volta individuati i record errati, determina il numero minimo di variabili da modificare (insieme minimale) e permette di correggerli sulla base di tre strategie diverse:

↳ *imputazione congiunta*;

↳ *imputazione sequenziale*;

↳ *imputazione basata sulle distribuzioni marginali o imputazione forzata*.

Le prime due strategie di correzione sono del tipo “*da donatore*”. La terza strategia di correzione dei dati, o imputazione forzata, è basata sull'analisi e sull'utilizzo delle distribuzioni marginali semplici rilevate nell'indagine per le variabili dell'insieme minimale.

La tecnica di *imputazione congiunta* prevede a sua volta due diverse possibilità:

- ✓ *imputazione congiunta ristretta*: correzione con un donatore che abbia in tutte le variabili da non imputare gli stessi valori del record errato;
- ✓ *imputazione congiunta allargata*: correzione con un donatore che abbia le variabili da non imputare ammissibili, ovvero che non facciano fallire alcuna regola.

Nel caso di *imputazione sequenziale* per ciascuna variabile appartenente all'insieme minimo viene calcolato il campo di variazione dei valori ammissibili; per ciascuna di esse viene quindi cercato nel serbatoio e, se esiste, selezionato un record con valore nel corrispondente campo di variazione.

Se non viene trovato nel serbatoio un record con i valori ammissibili, la correzione viene effettuata utilizzando la tecnica di *imputazione forzata*. Tale tecnica è basata su un algoritmo random di estrazione del valore da assegnare alla variabile errata (selezionato tra

i valori ammissibili), che viene guidata da una funzione di probabilità definita sulla base della distribuzione di frequenze che la variabile stessa assume nel file dei dati originari.

14.1.2 La revisione delle variabili qualitative nell'Indagine sui Consumi delle Famiglie

Il controllo e la correzione delle variabili qualitative nell'Indagine sui Consumi delle Famiglie viene effettuata mediante l'utilizzo del sistema SCIA. In particolare, vengono controllate e corrette prima le variabili relative ai componenti della famiglia e successivamente quelle relative all'abitazione e alle abitudini di spesa.

A seconda delle variabili soggette a correzione, è possibile distinguere tre fasi di applicazione di SCIA:

- ✓ *vincoli inter - componenti* (1^a fase), si analizzano le relazioni di incompatibilità tra valori di variabili relative a differenti componenti di una stessa famiglia;
- ✓ *vincoli intra - componenti* (2^a fase), si analizzano le relazioni di incompatibilità tra valori di variabili relative a un singolo componente;
- ✓ *vincoli sull'abitazione, beni durevoli e abitudini di spesa* (3^a fase), si verificano le relazioni tra le seguenti variabili:
 - tipologia abitazione;
 - anno costruzione - occupazione;
 - superficie - numeri di stanze;
 - servizio di energia elettrica - elettrodomestici.

La fase di correzione consiste nell'attribuzione alle variabili con valori considerati "errati", sia perché incoerenti rispetto agli edit definiti sia perché mancanti laddove richiesti, i valori prelevati dalle corrispondenti variabili di record "esatti" applicando la metodologia di *imputazione da donatore*. Nel caso in cui non si riesca a definire un bacino di donatori, si fa ricorso alle distribuzioni marginali delle variabili stesse utilizzando la metodologia di *imputazione forzata*.

Nell'imputazione da donatore viene fatto riferimento alla metodologia "*hot deck*". All'interno di una delle tre fasi di correzioni descritte si impone il vincolo aggiuntivo che un

record che è stato corretto non possa diventare a sua volta donatore. Quindi, ad esempio, se un record è stato corretto nella prima fase può far parte dell'insieme dei donatori della 2^a e 3^a fase ma non dei restanti record della 1^a.

La scelta di suddividere in due fasi la correzione delle notizie riguardanti i componenti della famiglia è dovuta all'utilizzo nella prima fase di record familiari e nella seconda di record individuali per ogni singolo componente della famiglia.

La considerazione simultanea di entrambe le tipologie di vincoli (inter - componenti e intra - componenti) comporterebbe l'organizzazione dei dati esclusivamente in record familiari contenenti le variabili di tutti i componenti relativi ad una stessa famiglia e quindi nella denominazione delle variabili in modo che siano identificabili i numeri d'ordine dei componenti cui si riferiscono. L'utilizzo del record familiare comporta, di conseguenza, la gestione posizionale delle variabili relative ai diversi componenti, ovvero le caratteristiche individuali del generico *i - esimo* componente vengono corrette prelevando i valori delle corrispondenti variabili del componente *i - esimo* nel record donatore (imputazione da donatore) o in base alle distribuzioni marginali definite dalla posizione (imputazione forzata).

Tale procedura potrebbe non essere appropriata laddove si verificano l'uno o l'altra delle seguenti circostanze:

- ↳ diminuzione del serbatoio dei donatori (ultimi componenti di famiglie numerose);
- ↳ mancata corrispondenza tra la posizione nel record familiare del componente con valori errati e la posizione del componente con valori corretti. Ad esempio, il terzo componente del record errato potrebbe essere il figlio della persona di riferimento, mentre nel record esatto potrebbe essere la moglie.

Per evitare questi inconvenienti sono state divise le procedure di verifica dei dati in due distinte applicazioni sequenziali del sistema SCIA (CN971A e CN971B) che effettuano, rispettivamente, il controllo e la correzione:

- ✓ delle *variabili demografiche* - controllo di tutte le relazioni di incompatibilità tra le variabili, localizzazione degli errori e correzione delle variabili relative alle caratteristiche

demografiche individuali (*relazione con la persona di riferimento, sesso, anno di nascita, stato civile*);

- ✓ delle variabili *socio – economiche* - correzione delle variabili relative alle altre caratteristiche individuali (*titolo di studio, condizione professionale, branca di attività economica prevalente, posizione nella professione*).

Nell'applicazione CN971A si opera con i record familiari. Il sistema procede ad individuare i record errati, a localizzare gli errori e ad attribuire alle variabili selezionate i nuovi valori. I valori imputati per le variabili demografiche sono considerati validi mentre i valori attribuiti alle variabili socio - economiche vengono sostituiti con quelli originali.

Al termine della prima fase i record familiari vengono frammentati in tanti record individuali quanti sono i componenti della famiglia e si procede ad eseguire l'applicazione CN971B.

Nei record individuali sono presenti i valori originali delle variabili socio - economiche da correggere e i valori delle variabili demografiche *età* e *sesso* risultanti dalla procedura di controllo CN971A.

Sono, inoltre, presenti i valori di due nuove variabili:

- ✓ la *ripartizione geografica* ottenuta accorpare le provincie di rilevazione nelle cinque macro ripartizioni (nord - ovest; nord - est, centro, sud e isole) utilizzata per la stratificazione del serbatoio dei donatori;
- ✓ una variabile composta ottenuta dalla combinazione delle modalità della variabile *sesso* con le classi di età <15, 15 - 29, 30 - 49, 50 - 64, ≥65.

I record sono, quindi, ordinati secondo la ripartizione geografica e la combinazione delle due variabili sesso ed età.

L'intera procedura di controllo e correzione per le variabili relative alle notizie sui componenti della famiglia è articolata nelle seguenti fasi:

- ✓ copia delle modalità delle variabili socio - economiche, calcolo della variabile età in anni per ciascun componente la famiglia e predisposizione del file da sottoporre al sistema SCIA per la correzione delle variabili demografiche;

- ✓ esecuzione dell'applicazione SCIA per la correzione delle variabili demografiche;
- ✓ ripristino delle modalità originarie delle variabili socio - economiche, calcolo di alcune variabili aggiuntive quali la variabile di stratificazione ripartizione geografica e la variabile combinata sesso - età e predisposizione del file dei record individuali da sottoporre al sistema di SCIA per la correzione delle variabili socio - economiche.
- ✓ esecuzione dell'applicazione di SCIA per la correzione delle variabili socio - economiche.

Per quanto riguarda le caratteristiche dell'abitazione, il possesso di beni durevoli e le abitudini di spesa una prima correzione viene effettuata in modo deterministico con il programma SASDET sulla base delle incoerenze rilevate applicando il programma BLAISE.

In particolare si controllano e correggono in modo deterministico:

- ✓ la tipologia dell'abitazione fissandola come impropria quando mancano contemporaneamente tutte le informazioni successive, negli altri casi si interviene in modo probabilistico;
- ✓ l'anno di costruzione e l'anno di occupazione dell'abitazione se almeno uno dei due è presente;
- ✓ la superficie dell'abitazione con il numero di stanze se almeno una delle due è presente (si fissa un rapporto standard tra le due variabili);
- ✓ la presenza del servizio di acqua calda se risulta indicata una modalità riguardante il tipo di impianto di riscaldamento dell'acqua;
- ✓ la presenza di elettrodomestici e di energia elettrica e quindi della bolletta;
- ✓ la compatibilità tra servizio telefonico, apparecchio telefonico e bolletta telefonica;
- ✓ alcune informazioni riguardanti la seconda abitazione.

Con SCIA vengono controllate tutte le altre incompatibilità per cui, ad esempio, nel caso di tipologia dell'abitazione mancante ed in presenza delle altre informazioni riguardanti l'abitazione, SCIA imputerà con metodo probabilistico una tipologia diversa da quella impropria. Nel caso, invece, sia stata indicata la presenza di acqua calda ma non il tipo di

impianto di riscaldamento acqua, SCIA imputerà il tipo di impianto. SCIA, inoltre, controllerà ed eventualmente correggerà tutte le incompatibilità relative al titolo di godimento dell'abitazione, al proprietario dell'abitazione e al tipo di contratto. Si è inoltre stabilito che l'eventuale imputazione degli elettrodomestici debba avvenire solo in presenza di energia elettrica. Infine, SCIA controlla la dichiarazione delle modalità di pagamento.

L'imputazione avviene, oltre che nel rispetto delle regole di compatibilità, utilizzando come variabile di stratificazione una variabile ottenuta come combinazione della *tipologia familiare* e della *zona geografica di residenza*. In particolare, per le due variabili la disaggregazione presa in considerazione è la seguente:

↳ *zona geografica di residenza*:

- ✓ Nord - Centro;
- ✓ Sud

↳ *tipologia familiare*:

- ✓ 2 componenti entrambi anziani;
- ✓ 2 componenti adulti;
- ✓ 2 - 4 componenti con almeno un giovanissimo;
- ✓ 3 - 4 componenti adulti o anziani;
- ✓ 5 e più componenti.

Come sintesi delle due si ottengono 10 modalità di stratificazione.

14.2 *Correzione delle variabili quantitative*

14.2.1 Descrizione del Sistema di Ricostruzione delle Informazioni con Donazione Automatica (RIDA)

Il sistema RIDA (Ricostruzioni delle Informazioni con Donazione Automatica) è stato prodotto internamente all'ISTAT ed è disponibile in versione *Assembler*, realizzata da Giorgio Massimini, e in *C*, realizzata da Carlo Runci. RIDA permette la correzione di un file di dati di qualsiasi tipo tramite la tecnica del donatore.

Per poter applicare RIDA l'insieme dei record deve essere contenuto in un file dove viene riportata la matrice dei dati X di n unità e k variabili. Le unità rappresentano i vettori - riga o record, le variabili i vettori - colonna o campi. Ogni record viene identificato da uno o più campi che lo individuano in modo univoco, detti campi *chiave*.

Le variabili vengono distinte in:

- ✓ variabili affette da errore (in numero $h < k$);
- ✓ variabili esatte (in numero $k - h$).

All'inizio del processo RIDA distingue l'insieme dei record in due gruppi:

- ✓ insieme di record totalmente esatti;
- ✓ insieme di record che contengono almeno un errore.

Fondamentale nell'utilizzo di questa metodologia è la definizione delle distanze tra due unità in modo da individuare la distanza minima tra il donatore ed il ricevente. L'individuazione del donatore avviene, infatti, dopo aver calcolato tutte le distanze dell'unità ricevente da tutte le unità considerate esatte.

Per definire la distanza d tra due unità rispetto alle variabili esatte è necessario introdurre una metrica per tipologia di variabile:

↳ *Variabili qualitative sconnesse*. Si pone $d = 0$ se le unità presentano la stessa modalità, $d = 1$ altrimenti. $X_1 = X_2 \Rightarrow d = 0$, $X_1 \neq X_2 \Rightarrow d = 1$;

↳ *Variabili ordinate con m modalità.* Si pone $d = 0$ se sulle due unità è stata rilevata la stessa modalità, $d = 1$ se le modalità sono adiacenti, $d = 2$ se tra esse ce n'è una sola, e così via fino a $d = m - 1$, se le due modalità sono agli estremi opposti. Per rendere d variabile tra 0 ed 1, essa viene divisa per il suo massimo $m - 1$.
 $X_1 = X_2 \Rightarrow d = 0, X_1 = r, X_2 = s (r \neq s) \Rightarrow d = |r - s| / m - 1$.

↳ *Variabili qualitative telescopiche.* Tali variabili sono rappresentabili gerarchicamente tramite un insieme di gruppi primari di livello 1, contenenti ognuno più sottogruppi di livello 2. Ogni sottogruppo di livello 2 contiene più sottogruppi di livello 3 e così via fino ad un sottogruppi di livello j , contenente modalità non ulteriormente scomponibili in sottogruppi, che sono al livello più basso $j+1$. Una modalità siffatta può essere codificata con g gruppi di bit, ognuno dei quali è dimensionato in modo da poter rappresentare tutti i sottogruppi relativi a qual livello. Poniamo $d = 0$ se le due unità presentano la stessa modalità, $d = 1$ se le due modalità diverse sono nello stesso sottogruppo di livello j , $d = 2$ se esse sono in gruppi differenti di livello j , ma nello stesso sottogruppo di livello $j - 1$, $d = 3$ se sono in gruppi differenti di livello $j - 1$, ma nello stesso sottogruppo di livello $j - 2$ e così via fino ad un massimo di $d = j + 1$ se le due modalità sono in gruppi primari diversi di livello 1. Rendiamo la distanza variabile tra 0 ed 1 dividendola per il suo massimo pari a $j + 1$. Sia r il livello più alto a partire dal quale si riscontra una differenza tra X_1 ed X_2 .
 $X_1 = X_2 \Rightarrow d = 0, X_1 \neq X_2 \Rightarrow d = |j + 2 - r| / j + 1$.

↳ *Variabili quantitative.* Sia X_1 il valore assunto dalla variabile X nella prima unità, X_2 nella seconda. Poniamo $d = |X_1 - X_2|$. La distanza può essere resa variabile tra 0 e 1 dividendola per il suo massimo, pari alla differenza tra i valori massimo (X_{max}) e minimo (X_{min}) della variabile X presenti nel file.
 $X_1 = X_2 \Rightarrow d = 0, X_1 \neq X_2 \Rightarrow d = |X_1 - X_2| / X_{max} + X_{min}$.

Data una matrice di dati che presenta $k - h$ variabili non affette da errore, si definisce distanza minima ponderata D tra due generiche unità una espressione del tipo:

$$D = \sum_{i=1}^r W_i D_i,$$

dove D_i è la distanza tra due unità rispetto alla variabile i , misurata con una delle espressioni indicate nel precedente paragrafo, e W_i è un numero reale positivo che rappresenta l'importanza assegnata alla variabile i nel calcolo della distanza. Le r variabili sono scelte tra le $k-h$ non affette da errore e vengono chiamate *variabili di accoppiamento o di matching*.

Una volta individuata un'unità affetta da errore nella variabile k si determina l'unità esatta posta alla distanza minima o *unità donatrice*. L'unità donatrice dona il valore della variabile k all'unità affetta da errore. L'insieme delle unità tra le quali è scelta l'unità donatrice è detto il *serbatoio dei donatori*. Il serbatoio dei donatori può essere costruito in due modi:

- ✓ selezionando le unità esatte rispetto alla sola variabile k ;
- ✓ selezionando le unità esatte rispetto a tutte le variabili.

Nel primo caso si usa un serbatoio unico per ogni variabile errata, nel secondo caso si utilizza un serbatoio unico per tutte le variabili affette da errore. La scelta dell'unità donatrice è ulteriormente affinabile scegliendo nell'insieme delle variabili non affette da errore e non usate come variabili matching delle variabili dette di *strato*. Dopo aver formato il serbatoio dei donatori in uno dei due modi indicati, si seleziona l'unità donatrice tra quelle che, rispetto alle variabili di strato, presentano le stesse modalità dell'unità affetta da errore.

In sintesi per effettuare delle correzioni tramite RIDA l'utente deve definire 4 file:

- ✓ Il file contenente i record errati.
- ✓ Il file contenente i record esatti, costituenti il serbatoio dei potenziali donatori.
- ✓ Il file dei parametri. Esso contiene le variabili di stratificazione e di matching. Per ogni variabile è necessario specificare la posizione iniziale, la lunghezza, il tipo e il peso. Si possono poi inserire i parametri U , R , L . Essi sono, rispettivamente, il numero massimo di volte che la stessa unità può essere utilizzata come donatrice, il fattore moltiplicativo che penalizza l'uso ripetuto dello stesso donatore e la massima distanza a cui può essere considerato un donatore. L'inserimento dei parametri U e L implica la possibilità di non riuscire a trovare il donatore.

- ✓ Il file eseguibile. Esso contiene il nome del file dei parametri, nonché dei file dei record esatti ed errati. Inoltre, l'utente deve specificare il nome dei due file che saranno creati nel corso dell'esecuzione del programma, uno contenente i record modificati, l'altro quelli che la procedura non è riuscita a correggere.

Dopo la corretta esecuzione di RIDA, l'utente dovrà provvedere a fondere in un file unico il file dei record esatti, quello dei corretti e quello degli eventuali record che non è stato possibile correggere.

I record non corretti dovranno essere corretti con una tecnica alternativa.

14.2.2 La revisione delle variabili quantitative nell'Indagine sui Consumi delle Famiglie

Nell'Indagine sui Consumi delle Famiglie viene utilizzato il software RIDA per l'imputazione, relativamente alle variabili quantitative, dei valori sia mancanti sia individuati come errati, che vengono sostituiti con altri provenienti dalle corrispondenti variabili relative ad osservazioni (unità donatrici) con valori completi e corretti.

Nell'applicazione di RIDA si è posto il limite che un'unità non possa diventare donatrice più di tre volte.

Per valutare l'erroneità o la mancata indicazione sul questionario di alcuni dati ci si basa su variabili qualitative presenti nel questionario che siano strettamente correlate alle variabili quantitative sottoposte a controllo. Ad esempio, se la famiglia dichiara che *“Ha effettuato spese per la manutenzione ordinaria negli ultimi 3 mesi”* ma non indica l'ammontare della spesa si ha un dato mancante. Le variabili qualitative di controllo associate alle variabili quantitative prendono il nome di variabili indicatrici.

Sulla base della verifica effettuata mediante le variabili indicatrici viene deciso di:

- ✓ mantenere inalterato il valore della variabile;
- ✓ essere dichiarato mancante il valore e, quindi, essere sottoposto ad imputazione;
- ✓ essere eliminato il valore, perché si è desunto dalle variabili indicatrici che l'immissione del valore è dovuta ad errore.

Viene, invece, definito come valore anomalo il valore esterno ad un intervallo di accettazione definito tenendo conto:

- ✓ dell'informazione relativa ai prezzi minimi e massimi sul territorio nazionale forniti dal Servizio Prezzi per vari beni e servizi oggetto di indagine;
- ✓ della distribuzione campionaria delle spese relative alle singole voci;
- ✓ dell'individuazione dei limiti inferiori e superiori della regione di accettazione, massimizzando la probabilità di escludere i soli outlier tramite la tecnica D - Maso (Distance Measurement Algorithm for Selection Outliers).¹⁰

Per l'individuazione dei limiti di accettazione dei valori sono utilizzate tutte le informazioni disponibili, in quanto nessuna delle tecniche sopra esposte, singolarmente utilizzate, può fornire un'informazione soddisfacente.

Prima di passare al processo di correzione vero e proprio, per ogni variabile viene analizzata la frequenza con cui i dati grezzi si discostano in misura eccessiva rispetto ai limiti di accettazione. Successivamente si decide quali variabili correggere con RIDA e quali con la tecnica denominata dei *Massimi e Minimi Randomizzati*.

Alcuni dei beni considerati, infatti, non possono essere imputati tramite RIDA a causa della esigua numerosità dei possibili donatori; in tali casi si è adottata la tecnica dei massimi e minimi randomizzati, ovvero:

- ✓ se il valore dichiarato supera il valore massimo dell'intervallo, esso viene sostituito con il valore massimo dell'intervallo meno una componente casualmente compresa nel 10 % del valore del limite;

¹⁰ L'algoritmo D-Maso (Della Rocca, Luzi, 1997) è composto dalle seguenti fasi:

1. data la distribuzione della variabile di interesse si calcolano i rapporti tra il valore corrente e quello successivo;
2. si determina la soglia, detta *cutoff*, tale che tutte le distanze maggiori di essa possono essere analizzate come possibili identificatrici di *outlier*. La soglia *cutoff* è data dal prodotto tra la distanza mediana ed un opportuno fattore moltiplicativo che permetta di restringere l'arco delle possibilità;
3. viene definito un valore *a* che è la massima percentuale di casi esclusi, in ciascuna delle due code, dai limiti dell'intervallo di accettazione da scegliersi;
4. viene assunto come limite inferiore (superiore) dell'intervallo di accettazione quel valore che risponde ai seguenti requisiti:
 - ha una distanza dal precedente (successivo) maggiore del *cutoff*;
 - taglia non più dell'*a* % dei casi nella coda sinistra (destra).

- ✓ se il valore dichiarato è inferiore al valore minimo dell'intervallo, esso viene sostituito con il valore minimo dell'intervallo più una componente casualmente generata compresa nel 10 % del valore del limite.

Questa tecnica viene adottata per tutte le voci di spesa che presentano una frequenza di acquisto molto ridotta.

Per le principali voci di spesa il sistema RIDA sostituisce ciascun valore considerato errato con il valore per la stessa variabile presente nel record del donatore prescelto.

Per definire il donatore si definiscono:

- ✓ la funzione di distanza tra unità donatrice ed unità da imputare;
- ✓ l'insieme di variabili da inserire nella funzione di distanza (variabili di matching);
- ✓ il criterio con cui pesare il contributo relativo di ciascuna variabile alla funzione di distanza;
- ✓ l'insieme delle variabili rispetto alle quali i possibili donatori devono avere le stesse caratteristiche dell'unità affetta da errore (variabili di stratificazione).

Come funzione di distanza è stata adottata il valore assoluto della differenza tra i valori assunti dalle due unità. Ovvero, sia X_1 il valore assunto dalla variabile X nella prima unità, X_2 nella seconda si utilizza come *funzione di distanza* $d = |X_1 - X_2|$

Le variabili quantitative da sottoporre a controllo nella nuova indagine sui consumi delle famiglie sono circa 280, di queste circa 180 sono state corrette tramite RIDA.

L'unica variabile di stratificazione considerata è la ripartizione geografica:

- ✓ nord - ovest;
- ✓ nord - est;
- ✓ centro;
- ✓ sud;
- ✓ isole.

La scelta della ripartizione geografica come variabile di strato è motivata sia dall'ampiezza del campione che impedisce un impoverimento del serbatoio dei donatori, sia perché tiene conto della differente struttura dei prezzi a livello nazionale.

Le variabili di matching sono state invece definite rispetto a ciascuna variabile oggetto di correzione, sulla base che queste presentano con la variabile di interesse. Per la maggior parte delle variabili si è, comunque, scelta la classe di reddito come variabile di matching.

ALLEGATO*

* Le seguenti tavole sono tratte da: Cascioli R. (1999) “Il passaggio da DAISY a SCIA: migrazione dal sistema operativo CMS a UNIX del piano di controllo e correzione dei dati relativi all’indagine sulle Forze di lavoro”, *FIL/C: documento interno (resoconto del lavoro svolto da R. Cascioli, P. Gennari, V. Spinelli e C. Lucarelli)*, ISTAT, Roma.

Tabella 1. Numero e tipo di correzioni nei passi probabilistici con i programmi DAISY e SCIA

		DAISY Aprile 1997		SCIA Aprile 1997	
I passo					
(Anagrafico)	Totale record	202017		201988	
	Record esatti	178942	88,6%	179105	88,7%
	Record errati	23075	11,4%	22883	11,3%
	Correzione				
	Ristretta	10917	47,3%	5290	23,1%
	Allargata	11604	50,3%	17075	74,6%
	Sequenziale	554	2,4%	518	2,3%
	Forzature	120	0,5%	14	0,1%
II passo					
(Occupazione)	Totale record	172311		172285	
	Record esatti	142663	82,8%	146635	85,1%
	Record errati	29648	17,2%	25650	14,9%
	Correzione				
	Ristretta	8253	27,8%	1324	5,2%
	Allargata	16199	54,6%	21806	85,0%
	Sequenziale	5196	17,5%	2520	9,8%
	Forzature	562	1,9%	44	0,2%
III passo					
(Ricerca di lavoro)	Totale record	172311		172285	
	Record esatti	158563	92,0%	158338	91,9%
	Record errati	13748	8,0%	13947	8,1%
	Correzione				
	Ristretta	8869	64,5%	6830	49,0%
	Allargata	4092	29,8%	6274	45,0%
	Sequenziale	787	5,7%	843	6,0%
	Forzature	0	0,0%	0	0,0%
IV passo					
(Formazione e corsi)	Totale record	172311		172285	
	Record esatti	151220	87,8%	151107	87,7%
	Record errati	21091	12,2%	21178	12,3%
	Correzione				
	Ristretta	19065	90,4%	16978	80,2%
	Allargata	1390	6,6%	3462	16,3%
	Sequenziale	636	3,0%	738	3,5%
	Forzature	408	1,9%	1	0,0%

**Tabella 2. Distribuzione dei record per numero di correzioni (aprile 1997)
(solo per gli individui di età maggiore di 14 anni)**

DAISY		
numero di variabili modificate	frequenza	percentuale
0	78967	46,2
1	48139	28,2
2	20213	11,8
3	8724	5,1
4	5676	3,3
5	3037	1,8
>5	6242	3,6
	170998	100,0

SCIA		
numero di variabili modificate	frequenza	percentuale
0	80866	47,3
1	47262	27,6
2	19682	11,5
3	8361	4,9
4	5549	3,3
5	2933	1,7
>5	6312	3,7
	170965	100,0

Tabella 3. CORREZIONI APPORTATE ALLE VARIABILI DAI DUE PIANI DI CONTROLLO E CORREZIONE ORDINATE SECONDO L'AMMONTARE DELLA DIFFERENZA (solo per gli individui di età maggiore di 14 anni)

VARIABILI	DAISY	SCIA	differenza	variazione percentuale
	numero	numero		
iscrizione al collocamento	2993	4015	1022	34,1
ha frequentato corsi*	7689	8506	817	10,6
anno di inizio del lavoro	82	870	788	961,0
mese di inizio del lavoro	3961	4648	687	17,3
titolo di studio	2029	2679	650	32,0
professione	4473	4859	386	8,6
durata dell'occupazione	331	699	368	111,2
durata del corso*	2298	2568	270	11,7
condizione precedente la ricerca	1676	1898	222	13,2
anni di residenza	147	362	215	146,3
luogo di lavoro desiderato	1365	1570	205	15,0
orario di lavoro desiderato	1183	1386	203	17,2
immediatamente disponibile a lavorare	1292	1493	201	15,6
cerca lavoro	2741	2909	168	6,1
motivo della non ricerca	5702	5861	159	2,8
ore di lavoro effettuate	3346	3473	127	3,8
ha dipendenti	5770	5894	124	2,1
tipo di occupazione cercato	1215	1338	123	10,1
mesi di ricerca del lavoro	2468	2578	110	4,5
cittadinanza	238	340	102	42,9
mese di termine dell'ultima occupazione	3671	3769	98	2,7
ha mai avuto un'occupazione	2139	2218	79	3,7
ore settimanali corso*	3130	3206	76	2,4
azioni di ricerca del lavoro	1608	1655	47	2,9
anno di termine dell'ultima occupazione	3406	3444	38	1,1
motivo di abbandono dell'ultima occupazione	1432	1456	24	1,7
professione nell'ultima occupazione	611	634	23	3,8
mesi trascorsi dall'ultima azione	2391	2409	18	0,8
stato civile	3018	3034	16	0,5
attività economica anno precedente*	3010	3024	14	0,5
professione seconda attività	58	70	12	20,7
ore di lavoro svolte (si/no)	5078	5087	9	0,2
tempo pieno o parziale	1563	1571	8	0,5
branca economica dell'ultima occupazione	532	540	8	1,5
posizione nella professione	1477	1484	7	0,5
giorni di lavoro abituali	1862	1868	6	0,3
giorni di lavoro effettuati	1832	1837	5	0,3
ore di lavoro abituali	2585	2590	5	0,2
ore di lavoro seconda attività	271	276	5	1,8
cerca altro lavoro	1603	1607	4	0,2
ultima azione effettuata	2313	2316	3	0,1
posizione nella professione dell'ultima occupazione	1192	1194	2	0,2
non compiute azioni di ricerca	2255	2257	2	0,1
condizione anno precedente*	6605	6607	2	0,0
numero occupati	750	751	1	0,1
ha una seconda attività	2387	2388	1	0,0
tipo di occupazione secondaria	55	55	0	0,0
dove lavora	225	224	-1	-0,4
livello di formazione raggiunto*	7290	7289	-1	0,0
dove risiedeva anno precedente*	12929	12928	-1	0,0
condizione professionale	943	940	-3	-0,3

posizione nella professione seconda attività	143	139	-4	-2,8
motivo dell'assenza	55	50	-5	-9,1
durata dell'assenza	261	255	-6	-2,3
branca economica seconda attività	83	77	-6	-7,2
percepisce indennità o sussidi	33059	33052	-7	0,0
comunque disposto a lavorare	9858	9849	-9	-0,1
posizione nella professione anno precedente*	2314	2300	-14	-0,6
giorno di nascita	152	134	-18	-11,8
mese di nascita	175	133	-42	-24,0
località dell'assenza	384	342	-42	-10,9
sexso	560	490	-70	-12,5
lavoro a turni	1120	1046	-74	-6,6
lavoro notturno	1355	1273	-82	-6,1
lavoro serale	1270	1186	-84	-6,6
lavoro di sabato	1250	1166	-84	-6,7
lavoro di domenica	1256	1172	-84	-6,7
lavoro a domicilio	1563	1479	-84	-5,4
motivo della differenza	362	225	-137	-37,8
anno di nascita	767	562	-205	-26,7
motivo di ricerca altro lavoro	685	475	-210	-30,7
carattere dell'occupazione	7102	6754	-348	-4,9
scopo del corso*	4284	3344	-940	-21,9
mesi trascorsi dall'ultimo contatto	2934	1871	-1063	-36,2
tipo di titolo di studio	4582	3414	-1168	-25,5
relazione con il capofamiglia	3125	1642	-1483	-47,5
branca di attività economica	8199	2636	-5563	-67,8
TOTALE	216118	211740	-4378	-2,0

*variabili del modulo aggiuntivo di Aprile sull'istruzione e formazione professionale

**Tabella 4. PERCENTUALI DI CORREZIONI APPORTATE ALLE
VARIABILI DAI DUE PIANI DI CONTROLLO E CORREZIONE
(solo per gli individui di età maggiore di 14 anni)**

VARIABILI	DAISY	SCIA
percepisce indennità o sussidi dove risiedeva anno precedente*	19,33	19,33
comunque disposto a lavorare	7,56	7,56
branca di attività economica	5,76	5,76
ha frequentato corsi*	4,79	1,54
livello di formazione raggiunto*	4,5	4,98
carattere dell'occupazione	4,26	4,26
condizione anno precedente*	4,15	3,95
ha dipendenti	3,86	3,86
motivo della non ricerca	3,37	3,45
ore di lavoro svolte (si/no)	3,33	3,43
tipo di titolo di studio	2,97	2,98
professione	2,68	2
scopo del corso*	2,62	2,84
mese di inizio del lavoro	2,51	1,96
mese di termine dell'ultima occupazione	2,32	2,72
anno di termine dell'ultima occupazione	2,15	2,2
ore di lavoro effettuate	1,99	2,01
relazione con il capofamiglia	1,96	2,03
ore settimanali corso*	1,83	0,96
stato civile	1,83	1,88
attività economica anno precedente*	1,76	1,77
iscrizione al collocamento	1,76	1,77
mesi trascorsi dall'ultimo contatto	1,75	2,35
cerca lavoro	1,72	1,09
ore di lavoro abituali	1,6	1,7
mesi di ricerca del lavoro	1,51	1,51
ha una seconda attività	1,44	1,51
mesi trascorsi dall'ultima azione	1,4	1,4
ultima azione effettuata	1,4	1,41
posizione nella professione anno precedente*	1,35	1,35
durata del corso*	1,35	1,35
non compiute azioni di ricerca	1,34	1,5
ha mai avuto un'occupazione	1,32	1,32
titolo di studio	1,25	1,3
giorni di lavoro abituali	1,19	1,57
giorni di lavoro effettuati	1,09	1,09
condizione precedente la ricerca	1,07	1,07
cerca altro lavoro	0,98	1,11
azioni di ricerca del lavoro	0,94	0,94
lavoro a domicilio	0,94	0,97
tempo pieno o parziale	0,91	0,87
posizione nella professione	0,91	0,92
motivo di abbandono dell'ultima occupazione	0,86	0,87
luogo di lavoro desiderato	0,84	0,85
lavoro notturno	0,8	0,92
immediatamente disponibile a lavorare	0,79	0,74
lavoro serale	0,76	0,87
lavoro di sabato	0,74	0,69
lavoro di domenica	0,73	0,68
tipo di occupazione cercato	0,73	0,69
posizione nella professione dell'ultima occupazione	0,71	0,78
orario di lavoro desiderato	0,7	0,7
lavoro a turni	0,69	0,81
condizione professionale	0,65	0,61
	0,55	0,55

anno di nascita	0,45	0,33
numero occupati	0,44	0,44
motivo di ricerca altro lavoro	0,4	0,28
professione nell'ultima occupazione	0,36	0,37
sexso	0,33	0,29
branca economica dell'ultima occupazione	0,31	0,32
località dell'assenza	0,22	0,2
motivo della differenza	0,21	0,13
durata dell'occupazione	0,19	0,41
ore di lavoro seconda attività	0,16	0,16
durata dell'assenza	0,15	0,15
cittadinanza	0,14	0,2
dove lavora	0,13	0,13
mese di nascita	0,1	0,08
giorno di nascita	0,09	0,08
anni di residenza	0,09	0,21
posizione nella professione seconda attività	0,08	0,08
paese di nascita	0,06	0,83
branca economica seconda attività	0,05	0,05
anno di inizio del lavoro	0,05	0,51
motivo dell'assenza	0,03	0,03
tipo di occupazione secondaria	0,03	0,03
professione seconda attività	0,03	0,04

***variabili del modulo aggiuntivo di Aprile sull'istruzione e formazione professionale**

Tabella 5. Distribuzioni di frequenza delle principali variabili dopo il passaggio del piano di controllo e correzione DAISY e SCIA (dati di APRILE 1997)

	DAISY	SCIA	variazioni percentuali
ETA'			
<15	29289	29296	0,02
15-29	41171	41139	-0,08
30-44	44217	44221	0,01
45-59	39330	39318	-0,03
60-65	14323	14334	0,08
>65	33687	33677	-0,03
TOTALE	202017	201985	-0,02
RELAZIONE CON IL CAPOFAMIGLIA			
capofamiglia	75422	75415	-0,01
coniuge/convivente	48205	49714	3,13
figlio	70022	70056	0,05
ascendente	1599	1558	-2,56
altro parente	6193	4669	-24,61
altra persona convivente	576	573	-0,52
TOTALE	202017	201985	-0,02
TITOLO DI STUDIO			
dottorato	430	315	-26,74
laurea	8976	8786	-2,12
diploma universitario	784	781	-0,38
diploma	36204	36428	0,62
qualifica non universitaria	8294	8145	-1,80
licenza media	58530	58663	0,23
licenza elementare	54477	54588	0,20
nessun titolo	34322	34279	-0,13
TOTALE	202017	201985	-0,02
CONDIZIONE PROFESSIONALE			
blank	29706	29700	-0,02
occupato	69149	69153	0,01
disoccupato	5915	5910	-0,08
in cerca di prima occupazione	6064	6073	0,15
inizierà un'attività in futuro	442	444	0,45
casalinga	32209	32203	-0,02
studente	15763	15780	0,11
ritirato dal lavoro	36852	36820	-0,09
inabile al lavoro	2749	2744	-0,18
in servizio di leva	657	648	-1,37
altra condizione	2511	2510	-0,04
TOTALE	202017	201985	-0,02
ORE DI LAVORO SVOLTE (SI/NO)			
blank	33112	33092	-0,06
si	67083	67067	-0,02
no	101822	101826	0,00
TOTALE	202017	201985	-0,02

POSIZIONE NELLA PROFESSIONE

blank	131984	131951	-0,03
dirigente	1161	1118	-3,70
direttivo, quadro	3092	3082	-0,32
impiegato o intermedio	21064	21068	0,02
operaio subalterno ed assimilati	23330	23360	0,13
apprendista	505	506	0,20
lavorante a domicilio per conto di imprese	169	169	0,00
imprenditore	1459	1449	-0,69
libero professionista	2632	2627	-0,19
lavoratore in proprio	12676	12705	0,23
socio di cooperativa di produzione	1065	862	-19,06
coadiuvante	2880	3088	7,22
TOTALE	202017	201985	-0,02

HA DIPENDENTI

blank	185250	185204	-0,02
si	8124	8344	2,71
no	8643	8437	-2,38
TOTALE	202017	201985	-0,02

PROFESSIONE

blank	131984	131951	-0,03
dirigenti, imprenditori, amministratori	814	2020	148,16
professioni intellettuali	5336	5209	-2,38
professioni intermedie (tecnici)	12985	13039	0,42
professioni esecutive amministrative	8914	8045	-9,75
professioni relative alle vendite ed ai servizi per le famiglie	12115	11866	-2,06
artigiani, operai specializzati e agricoltori	16475	16564	0,54
conduttori di impianti e macchinari	6297	6190	-1,70
personale non qualificato	7088	6456	-8,92
forze armate	9	645	7066,67
TOTALE	202017	201985	-0,02

ATTIVITA' ECONOMICA

blank	131984	131954	-0,02
agricoltura	5214	4711	-9,65
estrazioni di minerali e produzione-distribuzione energia	783	771	-1,53
industria	14941	15819	5,88
costruzioni	5737	5517	-3,83
commercio	11571	11159	-3,56
alberghi e ristoranti	3274	2551	-22,08
trasporti	3779	3752	-0,71
intermediaz. monet. finanz., attività immobiliari	2459	2398	-2,48
altre attività professionali ed imprenditoriali	3316	3566	7,54
pubblica amministrazione, difesa, organismi internaz.	5653	6371	12,70
istruzione e sanità	9671	9814	1,48
altri servizi pubblici, sociali e personali	3635	3605	-0,83
TOTALE	202017	201988	-0,01

NON HA COMPIUTO AZIONI DI RICERCA

blank	196254	196233	-0,01
nonblank	5763	5752	-0,19
TOTALE	202017	201985	-0,02

IMMEDIATAMENTE DISPONIBILE A LAVORARE

blank	182948	182828	-0,07
si	17674	17773	0,56
no	1395	1384	-0,79
TOTALE	202017	201985	-0,02

TIPO DI OCCUPAZIONE

blank	182948	182828	-0,07
alle dipendenze	18712	18709	-0,02
autonomo	357	448	25,49
TOTALE	202017	201985	-0,02

ISCRIZIONE AL COLLOCAMENTO

blank	29706	29700	-0,02
si	20021	19239	-3,91
no	152290	153046	0,50
TOTALE	202017	201985	-0,02

MESI TRASCORSI DALL'ULTIMO CONTATTO

blank	181996	182746	0,41
non blank	20021	19239	-3,91
TOTALE	202017	201985	-0,02

CONDIZIONE PROFESSIONALE

occupati	70033	70034	0,00
in cerca di lavoro	9544	9549	0,05
forze di lavoro potenziali	3601	3600	-0,03
non forze di lavoro	118381	118358	-0,02
emigrati	458	447	-2,40
TOTALE	202017	201988	-0,01

Matrice di transizione grezzi/puliti per la variabile RELAZIONE CON IL CAPOFAMIGLIA
(dati Aprile 1997)

DAISY	DATI PULITI							Totale
	capofamiglia	coniuge-convivente	figlio	ascendente	parente	altra persona convivente		
blank	68	15	80	0	0	50	0	213
capofamiglia	75143	0	63	0	0	166	2	75374
coniuge-convivente	160	48187	339	0	0	1800	0	50486
figlio	30	0	68962	0	0	295	0	69287
ascendente	17	3	549	1599	0	211	3	2382
parente	2	0	0	0	0	3671	0	3673
altra persona convivente	2	0	0	0	0	0	571	573
non ammissibili	0	0	29	0	0	0	0	29
Totale	75422	48205	70022	1599	6193	576	202017	

SCIA	DATI PULITI							Totale
	capofamiglia	coniuge-convivente	figlio	ascendente	parente	altra persona convivente		
blank	57	15	73	0	0	34	0	179
capofamiglia	75145	0	70	0	0	161	0	75376
coniuge-convivente	164	49693	349	0	0	282	0	50488
figlio	32	0	68970	0	0	283	0	69285
ascendente	13	6	566	1558	0	237	2	2382
parente	2	0	0	0	0	3671	0	3673
altra persona convivente	2	0	0	0	0	0	571	573
non ammissibili	0	0	28	0	0	1	0	29
Totale	75415	49714	70056	1558	4669	573	201985	

Matrice di transizione grezzi/puliti per la variabile TITOLO DI STUDIO
(dati Aprile 1997)

DAISY		DATI PULITI										
DATI GREZZI		dottorato	laurea	diploma universitario	diploma	qualifica non universitaria	licenza media	licenza elementare	nessun titolo	Totale		
blank		6	81	2	221	27	560	644	2971	4512		
dottorato		310	0	0	1	0	0	0	10	321		
laurea		0	8470	0	3	2	16	4	13	8508		
diploma universitario		3	26	760	1	0	2	0	2	794		
diploma		5	58	0	35740	8	5	5	8	35829		
qualifica non universitaria		15	51	4	67	8253	6	6	7	8409		
licenza media		73	252	17	46	4	57941	67	86	58486		
licenza elementare		16	35	1	22	0	0	53749	184	54007		
nessun titolo		2	3	0	2	0	0	0	31039	31046		
non ammissibili		0	0	0	1	0	0	2	2	5		
Totale		430	8976	784	36104	8294	58530	54477	34322	201917		

SCIA		DATI PULITI										
DATI GREZZI		dottorato	laurea	diploma universitario	diploma	qualifica non universitaria	licenza media	licenza elementare	nessun titolo	Totale		
blank		2	65	5	221	44	420	673	3047	4477		
dottorato		296	0	0	12	3	0	0	10	321		
laurea		0	8418	17	53	8	0	2	10	8508		
diploma universitario		4	45	688	44	10	1	0	2	794		
diploma		3	75	52	35403	280	3	5	10	35831		
qualifica non universitaria		2	48	17	544	7780	1	10	8	8410		
licenza media		8	116	2	128	16	58238	34	45	58587		
licenza elementare		0	18	0	21	4	0	53862	101	54006		
nessun titolo		0	1	0	1	0	0	0	31044	31046		
non ammissibili		0	0	0	1	0	0	2	2	5		
Totale		315	8786	781	36428	8145	58663	54588	34279	201985		

Matrice di transizione grezzi/puliti per la variabile PROFESSIONE
(dati Aprile 1997)

DAISY	DATI PULITI									Totale	
	blank	1	2	3	4	5	6	7	8		9
blank	131884	38	37	232	454	303	269	56	166	0	133439
1	6	749	14	265	544	115	155	23	100	0	1971
2	10	11	5263	298	242	37	32	19	54	0	5966
3	8	2	2	12034	83	38	15	14	48	0	12244
4	3	2	0	5	7316	5	0	2	1	0	7334
5	22	0	4	9	21	11460	9	3	7	0	11535
6	31	0	0	7	12	11	15947	2	12	0	16022
7	5	1	0	2	12	3	9	6156	1	0	6189
8	14	1	2	6	14	8	4	5	6656	0	6710
9	1	10	14	127	216	135	35	17	43	9	607
Totale	131984	814	5336	12985	8914	12115	16475	6297	7088	9	202017

SCIA	DATI PULITI									Totale	
	blank	1	2	3	4	5	6	7	8		9
blank	131852	51	18	264	187	384	394	110	144	6	133410
1	6	1885	0	32	28	5	6	5	3	1	1971
2	10	25	5161	349	184	57	82	36	57	5	5966
3	8	4	7	11967	78	53	34	25	56	12	12244
4	3	6	5	35	7194	26	39	14	13	0	7335
5	22	11	2	107	81	11151	77	20	56	4	11531
6	30	14	2	66	57	74	15688	46	37	8	16022
7	5	9	5	56	57	30	113	5887	25	2	6189
8	14	15	9	161	170	86	130	46	6059	20	6710
9	1	0	0	2	9	0	1	1	6	587	607
Totale	131951	2020	5209	13039	8045	11866	16564	6190	6456	645	201985

- 1 - dirigenti, imprenditori, amministratori
- 2 - professioni intellettuali
- 3 - professioni intermedie (tecnici)
- 4 - professioni esecutive amministrative
- 5 - professioni relative alle vendite ed ai servizi per le famiglie
- 6 - artigiani, operai specializzati e agricoltori
- 7 - conduttori di impianti e macchinari
- 8 - personale non qualificato
- 9 - forze armate

**Matrice di transizione grezzi/puliti per le variabili ISCRIZIONE AL COLLOCAMENTO e
MESI TRASCORSI DALL'ULTIMO CONTATTO
(dati Aprile 1997)**

DAISY

ISCRIZIONE AL COLLOCAMENTO	DATI PULITI			Totale
	blank	si	no	
blank	29637	147	3268	33052
si	6	19866	0	19872
no	63	8	149022	149093
Totale	29706	20021	152290	202017

MESI DALL'ULTIMO CONTATTO	DATI PULITI			Totale
	blank	nonblank	non blank	
blank	181640	2638	17383	184278
non blank	356	17383	17739	198578
Totale	181996	20021	17739	202017

SCIA

ISCRIZIONE AL COLLOCAMENTO	DATI PULITI			Totale
	blank	si	no	
blank	29631	106	3307	33044
si	6	18976	887	19869
no	63	157	148852	149072
Totale	29700	19239	153046	201985

MESI DALL'ULTIMO CONTATTO	DATI PULITI			Totale
	blank	nonblank	non blank	
blank	182547	1701	17737	184248
non blank	199	17538	17737	198578
Totale	182746	19239	17737	201985

Matrice di transizione grezzi/puliti per la variabile CONDIZIONE PROFESSIONALE RICOSTRUITA
(dati Aprile 1997)

DAISY	DATI PULITI					Totale
	occupati	in cerca di lavoro	forze di lavoro potenziali	non forze di lavoro	emigrati	
blank	514	96	232	2187	456	3485
occupati	69390	0	0	0	0	69390
in cerca di lavoro	0	9287	1030	77	0	10394
forze di lavoro potenziali	0	0	2089	4	0	2093
non forze di lavoro	129	161	250	116113	2	116655
Totale	70033	9544	3601	118381	458	202017

SCIA	DATI PULITI					Totale
	occupati	in cerca di lavoro	forze di lavoro potenziali	non forze di lavoro	emigrati	
blank	524	95	247	2174	446	3086
occupati	69395	0	0	0	0	69395
in cerca di lavoro	0	9274	1027	93	0	10394
forze di lavoro potenziali	0	0	2086	7	0	2093
non forze di lavoro	115	180	240	116084	1	116620
Totale	70034	9549	3600	118358	447	201988

Matrice di transizione grezzi/puliti per la variabile ATTIVITA' ECONOMICA
(dati Aprile 1997)

DAISY		DATI PULITI													Totale
		blank	agricoltura	energia	estrazioni- industria	costruzioni	commercio	alberghi e ristoranti	trasporti	attività monetarie- finanz.- immobiliari	altre attività profes. ed imprendit.	pubb.amm, difesa, org. internaz.	sanità	altri servizi pubblici, sociali, personali	
blank	131888	131888	492	11	150	71	256	83	36	27	4	31	31	87	52
agricoltura	7	4392	2	32	6	22	4	4	9	4	1	9	9	8	10
estrazioni-energia	0	9	669	2	32	20	10	4	6	3	1	2	2	5	6
industria	17	57	61	14225	121	797	56	7	38	41	13	15	1	48	35
costruzioni	24	11	7	55	5232	41	41	3	19	3	0	1	1	5	3
commercio	10	40	6	138	11	10075	377	7	15	2	2	4	4	14	49
alberghi e ristoranti	18	12	1	9	3	27	2307	25	5	4	1	4	4	4	19
trasporti	1	20	3	66	39	29	25	25	3489	7	1	21	21	10	19
attività monetarie-finanz.-immobiliari	3	3	0	14	0	0	25	11	6	7	1	5	5	9	9
altre attività profes. ed imprendit.	3	13	6	60	34	45	36	11	16	12	12	15	15	34	97
pubb.amm, difesa, org. internaz.	2	78	5	53	121	22	72	36	83	13	9	5491	9	328	3606
istruzione e sanità	7	29	6	26	23	41	142	72	10	17	1	30	1	9057	240
altri servizi pubblici,sociali,personali	4	58	6	81	56	181	150	47	4	44	20	25	25	62	2892
Totale	131984	5214	783	14941	5737	11571	3274	3779	2459	3316	5653	9671	3635	202017	

SCIA		DATI PULITI													Totale
		blank	agricoltura	energia	estrazioni- industria	costruzioni	commercio	alberghi e ristoranti	trasporti	attività monetarie- finanz.- immobiliari	altre attività profes. ed imprendit.	pubb.amm, difesa, org. internaz.	sanità	altri servizi pubblici, sociali, personali	
blank	131859	133	9	263	85	365	75	52	28	1	54	50	146	69	
agricoltura	7	4451	0	20	1	16	0	3	1	0	0	1	4	2	
estrazioni-energia	0	4	735	12	3	3	1	1	0	1	1	2	3	2	
industria	16	16	23	15335	44	15	12	11	7	7	7	9	15	14	
costruzioni	24	4	0	32	5322	16	3	3	1	1	1	0	2	0	
commercio	10	24	1	38	2	10592	9	4	14	6	6	5	15	22	
alberghi e ristoranti	18	4	0	4	2	25	2352	1	2	2	1	2	2	1	
trasporti	1	10	0	28	5	16	5	3651	3	3	2	1	3	5	
attività monetarie-finanz.-immobiliari	3	1	0	10	0	15	8	0	2325	0	0	5	2	0	
altre attività profes. ed imprendit.	3	6	0	20	1	25	13	3	1	1	3486	6	23	19	
pubb.amm, difesa, org. internaz.	2	36	3	15	49	22	25	11	1	1	6268	68	16	3606	
istruzione e sanità	7	5	0	10	0	23	37	2	5	1	9491	6	88	16	
altri servizi pubblici,sociali,personali	4	17	0	32	3	26	11	10	10	6	16	16	40	5	
Totale	131954	4711	771	15819	5517	11159	2551	3752	2398	3566	6371	9814	3605	201988	

BIBLIOGRAFIA

- Abbate C. (1998), La completezza delle informazioni e l'imputazione da donatore con distanza mista minima, *documento interno ISTAT*.
- Balestrino R., Capelli D., De Angelis R., Macchia S., Mazza L. (1998) L'introduzione delle tecnologie CASIC in un Istituto produttore di statistiche ufficiali. *Quaderni di ricerca*, n.2, ISTAT, Roma.
- Barbagli M., Muratore M.G., Quattrociochi L., Sabbadini L.L., L'indagine pilota sulle vittime dei fatti delittuosi - Valutazione delle possibilità di adottare il sistema CATI, *FAM/C INDAGINI MULTISCOPO: documento interno ISTAT*.
- Barcaroli G., D'Aurizio L., Luzi O., Manzari A., Pallara A. (1999), Metodi e software per la correzione dei dati, *Documenti ISTAT*, n.1, ISTAT, Roma.
- Barcaroli G., Venturi M. (1994), The probabilistic approach to automatic edit and imputation: improvements of the Fellegi-Holt methodology, *Documenti ISTAT*.
- Cascioli R. (1999) Il passaggio da DAISY a SCIA: migrazione dal sistema operativo CMS a UNIX del piano di controllo e correzione dei dati relativi all'indagine trimestrale sulle forze di lavoro, *FIL/C: documento interno ISTAT - resoconto del lavoro svolto da Cascioli R., Gennari P., Spinelli V. e Lucarelli C.*
- Ceccarelli C. (1998), Indagine sui Consumi delle Famiglie – Anno 1997: Alcune considerazioni riguardanti il disegno di campionamento, *documento interno ISTAT*.
- Ceccarelli C., Grassi D., Pannuzzi, N. (1997), Regole di revisione, *documento interno ISTAT*.
- Cocchi D. (1999) La qualità delle indagini longitudinali. *Gruppo di lavoro sullo studio di aspetti generali delle indagini longitudinali (Giornata di studio: Roma, 25 febbraio 1999)*, ISTAT.
- Cocchi D., Montanari G.E. (1999) Considerazioni sulla ristrutturazione in senso longitudinale di indagini ISTAT su individui e famiglie. *Gruppo di lavoro sullo studio di indagini longitudinali sugli individui e sulle famiglie (Giornata di studio: Roma, 25 febbraio 1999)*, ISTAT.
- D'Aurizio L., Pannuzzi N., La revisione dei dati nella nuova indagine sui consumi delle famiglie: il trattamento delle variabili quantitative, *documento interno ISTAT*.
- Fabbris L. (1994) *L'error profile* di un'indagine statistica, In: Colombo B., Cortese A., Fabbris L. (a cura di) *La produzione di statistiche ufficiali*, CLEUP, Padova: 287-290.

- Fabbris L. (1999) Rilevazione di dati assistita da computer e per via telematica nelle indagini prospettiche. *Gruppo di lavoro sullo studio di aspetti generali delle indagini longitudinali (Giornata di studio: Roma, 25 febbraio 1999)*, ISTAT.
- Fabbris L., Bassi F. (1997) On-line likelihood controls in computer assisted interviewing. In: *Proceedings Invited papers (Istanbul, August 18-26 1997)*, International Statistical Institute: 33-47.
- Falorsi S. (1997) Descrizione dei principali aspetti metodologici delle indagini telefoniche Istat sulle famiglie, *Intervento al seminario ISTAT 'Esperienze di data capturing', Roma 18 giugno 1997*.
- Falorsi S., Rinaldelli C., Russo A. (1999) Obiettivi conoscitivi e disegno di rilevazione delle principali indagini ripetute nel tempo dall'ISTAT sugli individui e sulle famiglie. *Gruppo di lavoro sullo studio di indagini longitudinali sugli individui e sulle famiglie (Giornata di studio: Roma, 25 febbraio 1999)*, ISTAT.
- Falorsi S., Russo A. (1999) Definizione di famiglia e di universo nelle indagini longitudinali. *Gruppo di lavoro sullo studio di indagini longitudinali sugli individui e sulle famiglie (Giornata di studio: Roma, 25 febbraio 1999)*, ISTAT.
- Favro-Paris M., Gennari P., Oneto G. P. (1996) La durata della disoccupazione in Italia, *Quaderni di ricerca*, n.4, ISTAT, Roma.
- Fellegi I.P., Holt D. (1976) A systematic approach to automatic editing & imputation, *Journal of the American Statistical Association*, **71**: 17-35.
- Filippone A., Foglia I. (1998) Il fenomeno delle mancate risposte: rapporto relativo all'indagine ISTAT sui consumi delle Famiglie nel 1997, *documento interno ISTAT*.
- Gennari P., Sestito P. (1996) Durata dell'attività lavorativa e turn-over in Italia. XI *Convegno Nazionale dell'Associazione degli Economisti del Lavoro (AIEL)*: 3-5 ottobre 1996, Napoli.
- Gennari P. (1999) Progetti di aggiornamento e integrazione delle statistiche derivate dalla rilevazione trimestrale sulle forze lavoro, *FIL/C (ISTAT): relazione per l'audizione su "Lo stato dell'informazione statistica sul lavoro, con particolare riguardo alla partecipazione al lavoro e a retribuzioni e costo del lavoro" presso la Commissione per la Garanzia dell'Informazione Statistica*, Presidenza del Consiglio dei Ministri, Roma.
- Gennari P. (1999) Proposta di adeguamento di alcune definizioni della RTFL agli standard stabiliti in sede EUROSTAT, *FIL/C: documento interno ISTAT*.
- Gennari P., Spinelli V. (1999) Stato di attuazione del lavoro di revisione delle serie statistiche dell'indagine forze di lavoro (ott. '92 – gen. '99) e proposta di modifica di alcune definizioni, *FIL/C: documento interno ISTAT*.

- Grassi D. (1998) La rete di rilevazione, *documento interno ISTAT*.
- Grassi D. (1999) La revisione dei dati nella nuova indagine sui consumi delle famiglie: il trattamento delle variabili qualitative, *documento interno ISTAT*.
- ISTAT (1993) Indagine multiscopo sulle famiglie. Obiettivi, disegno e metodologia dell'indagine.
- ISTAT (1996) Rilevazione Trimestrale delle Forze di Lavoro, Istruzioni per gli intervistatori. Ciclo aprile 1996 – Gennaio 1997.
- ISTAT (1999) La revisione delle serie storiche delle Forze di lavoro. Ottobre 1992-Aprile 1999. *Statistiche in breve (16 luglio 1999)*.
- Leti G., Falorsi P.D., Falorsi S., Russo A., Tranquilli G. (1999) Indagini ripetute nel tempo: obiettivi e disegni di rilevazione. *Gruppo di lavoro sullo studio di aspetti generali delle indagini longitudinali (Giornata di studio: Roma, 25 febbraio 1999)*, ISTAT.
- Macchia S. (1999) L'acquisizione dei dati, *documento interno ISTAT*.
- Masselli M. (1994) L'*error profile* dell'indagine sulle forze di lavoro. In: Colombo B., Cortese A., Fabbris L. (a cura di) *La produzione di statistiche ufficiali*, CLEUP, Padova: 291-312.
- Montanari G.E. (1999) La ponderazione dei dati nelle rilevazioni longitudinali mediante campioni panel. *Gruppo di lavoro sullo studio di indagini longitudinali sugli individui e sulle famiglie (Giornata di studio: Roma, 25 febbraio 1999)*, ISTAT.
- Muratore M.G., Commissione indagine multiscopo - Le indagini sulla vittimizzazione nelle più importanti esperienze internazionali, *FAM/C INDAGINI MULTISCOPO: documento interno ISTAT*.
- Muratore M.G., La dinamica di una indagine telefonica: ipotesi di monitoraggio, *SERVIZIO SDS - STRUTTURE E DINAMICHE SOCIALI, documento interno ISTAT*.
- Paggiaro A., Torelli N. (1999) Una procedura per l'abbinamento di record nella rilevazione trimestrale delle forze di lavoro. *Progetto di ricerca Murst "Lavoro e disoccupazione: questioni di misura e analisi"*, Working paper n.15, Dipartimento di Scienze Statistiche, Università di Padova.
- Pannuzzi N. (1999), I questionari utilizzati per la nuova indagine sui consumi delle famiglie, *documento interno ISTAT*.
- Quattrocioni L., L'indagine pilota sulla vittimizzazione: tecnica di indagine e riflessioni metodologiche, *FAM/C INDAGINI MULTISCOPO, documento interno ISTAT*.

- Quattrocioni L., Qualità nell'indagine telefonica: l'esperienza dell'indagine pilota sulla vittimizzazione, *SERVIZIO SDS - STRUTTURE E DINAMICHE SOCIALI: documento interno ISTAT.*
- Riccini Margarucci E., Silvestri F., Floris P. (1997), S.C.I.A. Sistema Correzione Imputazione Automatica – Manuale utente, *documento interno ISTAT.*
- Sabbadini L.L., Primi passi nelle indagini telefoniche: riflessioni a partire dalle esperienze delle indagini multiscopo, *SERVIZIO SDS - STRUTTURE E DINAMICHE SOCIALI, documento interno ISTAT.*
- Sabbadini L.L., (1998) La sicurezza dei cittadini - Molestie e violenze sessuali, *Presentazione dei risultati della ricerca, Roma 22 settembre 1998.*
- Saris W.E. (1991) Computer-Assisted Interviewing, Number 07-080, *Sage publications*, Newbury Park California.
- Servizio CEF (1997) Procedura informatica per il processo di acquisizione/imputazione/correzione dei dati, *documento interno ISTAT.*
- Servizio CEF (1999) Nuova indagine sui consumi delle famiglie, *documento interno ISTAT.*
- Trivellato U., Ghellini G., Martelli C., Regoli A. (1995) Prospettive per possibili analisi longitudinali nella statistica ufficiale italiana. *Rapporti di ricerca, 95.09*, Commissione per la Garanzia dell'Informazione Statistica, Presidenza del Consiglio dei Ministri, Roma.

SERIE “RAPPORTI DI RICERCA”

- 93.01 Valutazioni di procedure di oscuramento delle informazioni individuali e di canoni di pubblicazione di informazioni a minimo rischio di individuazione, (*M. Angrisani*)
- 93.02 Gli investimenti pubblici: problemi di contabilità pubblica e di contabilità nazionale, (*G. Trupiano*)
- 93.03 Investimenti pubblici lordi e netti: problemi analitici, (*V. Selan*)
- 93.04 L'indice dei prezzi al consumo in Italia, (*F. Franceschini, G. Marliani, M. Martini*)
- 94.01 Privatizzazione e sistema statistico nazionale, (*G. Di Gaspare*)
- 94.02 Stato delle statistiche sociali in Italia, (*G.B. Sgritta*)
- 94.03 Statistica sociale e Statistiche sociali, (*L. Bernardi*)
- 94.04 Prospettive preliminari per possibili analisi longitudinali nella statistica ufficiale italiana, (*U. Trivellato, G. Ghellini, C. Martelli, A. Regoli*)
- 94.05 Analisi di alcune caratteristiche del Programma Statistico Nazionale 1995-1997, (*D. Cotzia, S. D'Andrea, E. Mastantuoni*)
- 94.06 Verifica dei ritardi rispetto alle previsioni di stampa delle pubblicazioni ISTAT negli anni 1993 e 1994, (*D. Cotzia*)
- 94.07 Analisi sulla tempestività della Produzione di informazione statistica (Esame di alcune rilevazioni ed elaborazioni dell'Istat), (*D. Cotzia*)
- 94.08 La suddivisione territoriale della spesa pubblica per investimenti, (*G. Trupiano*)
- 94.09 Il consolidamento della spesa pubblica per investimenti, (*G. Trupiano*)
- 94.10 Investimenti netti, ammortamenti e spese di manutenzione. Stock di capitale: un'ipotesi censuaria, (*V. Selan*)
- 94.11 Le spese per investimenti nelle statistiche Eurostat sui conti delle amministrazioni pubbliche, (*M. Colazingari*)
- 94.12 Gli investimenti pubblici del Comune di Roma, (*P. Palmarini*)

- 94.13 La revisione del Sistema dei Conti Nazionali: problemi e prospettive per l'Italia, (*B. Bracalente, G. Carbonaro, M. Carlucci, M. Di Palma, L. Esposito, G. Ferrari, R. Zelli*)
- 94.14 La tutela della riservatezza e l'identificazione dei rispondenti alle rilevazioni statistiche svolte nell'ambito del Sistan: rapporto preliminare, (*M. Angrisani, L. Buzzigoli, A. Giusti, L. Grassini, G. Marliani*)
- 94.15 I dati statistici produttivi di effetti giuridici determinati e la loro sindacabilità, (*G. Manto*)
- 94.16 Ufficialità del dato e Programma Statistico Nazionale, (*G. D'Alessio*)
- 94.17 Valutazioni preliminari sulla qualità dei dati dell'ultimo censimento generale della popolazione e delle abitazioni, (*G. De Santis, A. Bonaguidi, A. Santini*)
- 94.18 La revisione del Sistema dei Conti Nazionali: problemi e prospettive per l'Italia - rapporto finale, (*B. Bracalente, G. Carbonaro, M. Carlucci, M. Di Palma, L. Esposito, G. Ferrari, R. Zelli*)
- 95.01 Classificazione delle province italiane in clusters e determinazione delle province outliers in riferimento alle correzioni degli errori di coerenza e di range del censimento dell'agricoltura 1991, (*S. D'Andrea*)
- 95.02 La qualità dei dati dell'ultimo censimento generale della popolazione e delle abitazioni, (*G. De Santis, S. Salvini, A. Santini*)
- 95.03 Stato delle Statistiche sociali in Italia - Sintesi del rapporto, (*G. B. Sgritta*)
- 95.04 Lo Stato dell'informazione statistica nei comuni e negli altri enti territoriali intermedi del Sistan: le province di Ferrara e Siena, (*A. Buzzi Donato, I. Drudi, M.R. Ferrante, C. Filippucci, G. Gesano, G. Ghellini, T. Giovani, A. Lemmi*)
- 95.05 Analisi delle funzioni del Sistema di Informazione Geografica-GISCO della Commissione delle Comunità Europee, (*E. Mastantuoni*)
- 95.06 Stato ed evoluzione delle statistiche ambientali in Italia, (*L. Fabbris, M. Lo Cascio*)
- 95.07 Rapporto sugli aspetti statistici nella Legislazione Ambientale - I. Aria, (*S. Bordignon, A. C.S. Capelo, G. Lovison, G. Masarotto*)
- 95.08 Il Sistema Statistico delle Imprese in Italia: rapporto preliminare, (*S. Biffignandi, M. Pratesi, T. Proietti, L. Schionato*)

- 95.09 Prospettive per possibili analisi longitudinali nella statistica ufficiale italiana, (*U. Trivellato, G. Ghellini, C. Martelli, A. Regoli*)
- 95.10 Per una estensione dei compiti della Commissione per la Garanzia dell'informazione statistica, (*G. Calvi, M.T. Crisci, S. Draghi, L. Ferrari, A. Rizzi*)
- 95.11 Rapporto sugli aspetti statistici nella legislazione ambientale - II. Rumore, (*S. Bordignon, A. C.S. Capelo, G. Lovison, G. Masarotto*)
- 95.12 Innovazioni integrazioni nel sistema dei conti nazionali: Problemi aperti e soluzioni possibili - Sintesi e suggerimenti -, (*B. Bracalente, G. Carbonaro, M. Carlucci, M. Di Palma, L. Esposito, G. Ferrari, R. Zelli*)
- 95.13 Disaggregazione spaziale e temporale delle statistiche ufficiali sulla qualità dell'aria, (*L. Fabbris*)
- 95.14 Disaggregazione spaziale e temporale delle statistiche ufficiali sulla qualità delle acque, (*L. Fabbris*)
- 95.15 L'esercizio della funzione statistica a livello locale: lo stato degli uffici di statistica comunali dopo il d.lgs. n.322/89, (*G. Manto*)
- 95.16 Gli uffici di statistica dei Ministeri, (*C. Gallucci*)
- 95.17 Le statistiche comunitarie e le statistiche nazionali: evoluzione, coordinamento, integrazione e processi di uniformazione, (*G. Di Gaspare*)
- 95.18 Organizzazione ed attività statistica delle regioni nel contesto del Sistan, (*G. D'Alessio*)
- 96.01 Rapporto sullo stato dell'informazione statistica nei comuni della provincia di Bari, (*C. Cecchi, V. Nicolardi, A. Pollice, N. Ribocco*)
- 96.02 Sistemi Nazionali di statistica: loro organizzazione e funzionamento in alcuni paesi dell'unione europea, (*B. Carelli*)
- 96.03 L'attività delle amministrazioni centrali dello Stato per il programma statistico nazionale del triennio 1996-98, (*G. Filacchione*)
- 96.04 Rapporto sugli aspetti statistici nella legislazione ambientale - III. Dati mancanti -, (*S. Bordignon, A.C.S. Capelo, G. Lovison, G. Masarotto*)
- 96.05 Osservatorio Statistico Locale: Studio di un modello per il Sistan, (*P. Bellini, S. Campostrini, T. Di Fonzo, M.P. Bellini*)

- 96.06 La tutela della riservatezza e l'identificazione dei rispondenti alle rilevazioni statistiche svolte nell'ambito del Sistan - rapporto finale, (*M. Angrisani, L. Buzzigoli, A. Giommi, A. Giusti, L. Grassini, G. Marliani*)
- 96.07 Analisi dell'organizzazione e delle iniziative del Sistan - Esame delle pubblicazioni presenti nel Catalogo Sistan 1994, (*A. De Nardo, S. Sagramora*)
- 96.08 Sistema Statistico delle Imprese, (*S. Biffignandi, M. Pratesi, T. Proietti, L. Schionato*)
- 96.09 Monitoraggio della diffusione dei dati riguardanti alcuni indicatori dell'Istat su prezzi, lavoro e commercio con l'estero, (*A. De Nardo, E. Mastantuoni, M. Notarnicola, S. Sagramora*)
- 96.10 Monitoraggio della qualità e tempestività dell'indice della produzione industriale, (*V. Napoli, F. Tagliaferro*)
- 96.11 La qualità dei dati del VII censimento dell'industria e dei servizi: alcune valutazioni dal punto di vista dell'utilizzatore, (*R. Guarini, R. Zelli*)
- 96.12 Analisi del processo di revisione corrente delle stime provvisorie dei dati del Commercio con l'Estero, (*E. Mastantuoni, S. Sagramora*)
- 96.13 Prime indagini sull'accesso ai dati statistici individuali nell'ambito del Sistan, (*L. Buzzigoli, C. Martelli, N. Torelli*)
- 97.01 Interconnessione di basi di dati: problemi di sfruttamento statistico, (*A. Cortese*)
- 97.02 La formazione statistica nelle amministrazioni dello Stato: profili comparativi ed elementi propositivi, (*F. Covino*)
- 97.03 Rapporto sull'autonomia degli uffici di statistica nelle amministrazioni centrali dello Stato, (*F. Covino*)
- 97.04 Rapporto sulle regioni e le province autonome nel sistema statistico nazionale, (*N. Belvedere*)
- 97.05 Il sistema statistico europeo. Stato attuale e possibile riforma, (*I. Savi*)
- 97.06 Rapporto preliminare sulla statistica in Francia e nel Regno Unito, (*E. Marotta*)
- 97.07 Verifica della programmazione nell'attività del Sistan e dell'attività di vigilanza, (*F. Bigazzi*)

- 97.08 Indagine sulle statistiche della Sanità, (*P. Golini*)
- 98.01 Evoluzione e prospettive della statistica comunitaria: un aggiornamento, (*I. Savi*)
- 98.02 L'incidenza sul SISTAN delle leggi di riforma amministrativa e della disciplina in materia di privacy, (*N. Belvedere, I Savi*)
- 98.03 Analisi sullo stato di attuazione degli uffici di statistica dei comuni. Analisi preliminari e progetto di rilevazione, (*A. De Nardo, M. Notarnicola*)
- 98.04 Documentazione statistica su fenomeni di emarginazione sociale: offerta e fabbisogni: Tossicodipendenze, (*B. Colombo, G. Filacchione*)
- 98.05 Analisi delle caratteristiche dei non rispondenti con riferimento alle principali indagini campionarie sulle famiglie condotte dall'ISTAT, (*E. Mastantuoni, S. Sagramora, F. Tagliafierro*)
- 98.06 La razionalizzazione della statistica giudiziaria, (*F. Giusti, S. Andreano, M. Fabri, V. Napoli, R. Santoro*)
- 99.01 Validità e qualità degli indici dei prezzi al consumo. *Atti del Seminario, Roma, 12 dicembre 1997*
- 99.02 Analisi della disponibilità delle statistiche di genere, (*M.E. Graziani*)
- 99.03 La razionalizzazione della statistica giudiziaria - Rapporto finale, (*F. Giusti, S. Andreano, M. Fabri, V. Napoli, R. Santoro*)
- 99.04 Le procedure di destagionalizzazione di serie storiche economiche: esperienze internazionali e pratica nell'ambito dell'Istat, (*T. Di Fonzo, B. Fischer, T. Proietti*)
- 99.05 Lo stato dell'informazione statistica sul lavoro, con particolare riguardo alla partecipazione al lavoro ed a retribuzioni e costo del lavoro, (*G. Faustini, E. Rettore, P. Sestito*)
- 99.06 Analisi delle caratteristiche dei non rispondenti con riferimento alle principali indagini campionarie sulle famiglie condotte dall'Istat, (*E. Mastantuoni, S. Sagramora*)
- 99.07 Statistiche dei rifiuti, (*L. Fabbris, G. Nebbia*)
- 99.08 Problemi di adeguamento della legislazione italiana alla normativa comunitaria e internazionale sulla tutela della riservatezza di dati personali utilizzati per finalità statistiche, (*N. Belvedere, I. Savi, F. Tufarelli*)

- 99.09 Stato di attuazione degli uffici di statistica dei comuni, (*A. De Nardo, M. Notarnicola*)
- 99.10 Il confronto tra censimento ed anagrafe: per un maggior grado di coerenza tra le due fonti, (*L. Ciucci, G. De Santis, M. Natale, M. Ventisette*)
- 99.11 Censimenti economici e schedari delle imprese, (*R. Castellano, C. Quintano, G. Screpis, F. Tassinari*)
- 99.12 Accesso ai dati statistici individuali: l'esperienza di altri paesi, (*L. Buzzigoli, C. Martelli, N. Torelli*)
- 00.01 Analisi della qualità delle operazioni sul campo con riferimento alle principali indagini campionarie dell'Istat sulle famiglie, (*C. Filippucci, B. Buldo, V. Napoli, R. Bernardini Papalia*)
- 00.02 Analisi delle procedure di correzione/imputazione utilizzate dall'Istat nelle principali indagini sulle famiglie: volume I, (*L. Fabbris, C. Panattoni, M. Graziani*)
- 00.03 Analisi delle procedure di correzione/imputazione utilizzate dall'Istat nelle principali indagini sulle famiglie: volume II, (*C. Panattoni, M. Graziani, D. Cotzia*)

Il presente rapporto di ricerca è stato riprodotto nel mese di luglio 2000