

**La procedura di controllo e correzione dei dati come insieme integrato
di diverse metodologie: sperimentazione sull'indagine pilota 1997 sul
Conto Economico delle Aziende Agricole**

Marco Di Zio (*), Orietta Luzi (*)

(*) ISTAT - Servizio Studi Metodologici

1. Introduzione

Le operazioni di controllo e correzione (*editing*) di dati statistici rilevati su soggetti economici quali imprese e istituzioni rappresenta una fase particolarmente critica del processo di indagine, a causa sia della complessità dei fenomeni indagati e delle relazioni generalmente esistenti fra essi, sia delle diverse e molteplici tipologie di errore non campionario riscontrabili in questo tipo di dati. Generalmente, una parte considerevole delle risorse impiegate in un'indagine statistica viene utilizzata nella fase di revisione e correzione (automatica e/o interattiva) dei dati (Granquist, 1996,1997). Nel caso delle indagini di tipo economico e particolarmente di quelle di carattere congiunturale è importante tenere sotto controllo non solo la qualità dei risultati finali, ma anche i costi e la tempestività con cui essi vengono prodotti. Pertanto, al fine di ottimizzare l'intero processo di revisione dei dati, sono state proposte una serie di tecniche e metodologie che possono essere impiegate o nella fase di editing vera e propria, oppure in una delle fasi precedenti del processo di indagine. E' questo il caso dell'uso delle tecniche di intervista assistita da computer (CAI), in cui una serie di controlli di coerenza sui dati raccolti sono effettuati direttamente all'atto dell'intervista dell'unità statistica, con la possibilità di rettifica immediata da parte del rispondente degli eventuali errori riscontrati. Anche le tecniche di registrazione controllata hanno lo scopo di "anticipare" la fase di verifica della coerenza dei dati, ed in particolare di minimizzare l'ammontare degli errori non campionari dovuti alla fase di registrazione stessa. Per il trattamento degli errori che comunque permangono nei dati è necessario definire, in fase di revisione, una strategia di controllo e correzione che consenta di individuare il massimo numero possibile di errori "veri" e di ripristinare per essi una situazione il più vicina possibile al "dato vero", ottimizzando al contempo l'uso delle risorse. A questo scopo sono state proposte una serie di metodologie, di tipo sia interattivo sia automatico, per il miglioramento dell'efficienza e la razionalizzazione dei diversi passi della procedura di editing, metodologie con cui sostituire o integrare l'approccio esclusivamente deterministico. Tale approccio, infatti, se da un lato consente il trattamento di alcune tipologie di errore (in particolare gli errori sistematici), dall'altro soffre di limiti quali la non garanzia di correttezza dei risultati finali e il non rispetto sia del criterio del minimo cambiamento (numero di variabili da modificare per liberare il singolo record dagli errori), sia della distribuzione multivariata iniziale. Tali limiti sono superabili solo mediante l'approccio probabilistico.

Tra le metodologie esistenti per la fase di individuazione dei valori anomali, le tecniche di *macroediting* (Granquist, 1992) e di tipo *selettivo* (Latouche e Berthelot, 1992) consentono di ottimizzare il ricorso ai controlli di tipo interattivo, concentrandoli sulle sole unità rilevanti o *influenti* sulle stime dei fenomeni di interesse; in tal modo si ottiene anche un contenimento del fenomeno dell'*over editing*, cioè di quelle operazioni di revisione che, riguardando errori poco influenti, assorbono risorse e tempo senza produrre un apprezzabile miglioramento della qualità dei risultati. Metodologie per l'analisi delle regole di controllo del questionario sono state proposte per l'ottimizzazione e la razionalizzazione del piano di compatibilità di un'indagine, cioè per l'eliminazione dal piano stesso di ridondanze, incongruenze, contraddittorietà fra le regole da utilizzare. Algoritmi di tipo probabilistico per la localizzazione degli errori, in particolare la *metodologia Fellegi-Holt* (Fellegi e Holt, 1976), consentono di ottimizzare la fase di individuazione degli errori stocastici sulla base dell'analisi *simultanea* dell'insieme delle regole del piano di compatibilità, nel rispetto dei requisiti di qualità precedentemente citati (correttezza dei dati finali, minimo cambiamento e mantenimento delle distribuzioni originali).

In termini di qualità e costi, il miglioramento della qualità dell'intero processo di editing può essere quindi ottenuto adottando una strategia di controllo e correzione *integrata*, in cui cioè relativamente ad ogni passo della procedura e ad ogni tipologia di errore si adottino metodologie ottimali, e tutte le metodologie utilizzate risultino fra loro complementari ed interagiscano in modo efficiente. Per *procedura integrata* si intende quindi un processo unico composto da fasi che devono essere disegnate in maniera che il loro output sia l'input di una fase successiva: ciò implica che i sottoprocessi (*fasi*) non possano essere disegnati indipendentemente l'uno dall'altro.

In questo documento è descritta una proposta operativa per la realizzazione di una procedura integrata per il controllo e la correzione dei dati quantitativi rilevati dall'indagine pilota sul Conto Economico delle Aziende Agricole relativa al 1997. In questa proposta tecniche di tipo deterministico per il controllo di errori non casuali risultano integrate con metodologie e tecniche innovative quali l'*editing selettivo* per l'individuazione degli errori influenti e gli algoritmi probabilistici di localizzazione degli errori implementati nel sistema automatico generalizzato per il trattamento di variabili quantitative GEIS (Generalised Editing and Imputation System). In particolare, nel paragrafo 2 viene descritta la struttura generale della procedura, mentre il paragrafo 3 contiene la descrizione dell'approccio adottato per la selezione dei casi da sottoporre a

controllo interattivo. Nel paragrafo 4 è descritta la procedura automatica progettata per il trattamento degli errori stocastici e che utilizza il software GEIS. Il paragrafo 5 è dedicato ad analisi complessive di tipo quantitativo e qualitativo dei risultati ottenuti dall'applicazione della procedura ai dati dell'indagine, infine il paragrafo 6 è dedicato alle conclusioni.

2. Struttura della procedura *integrata* di controllo e correzione

L'indagine campionaria sul Conto Economico delle Aziende Agricole fa parte della più complessa indagine sulla Struttura delle Aziende Agricole condotta annualmente dall'ISTAT. I dati sottoposti ad elaborazione sono relativi all'indagine pilota effettuata nel 1997, cui hanno risposto 2067 aziende distribuite sul territorio nazionale.

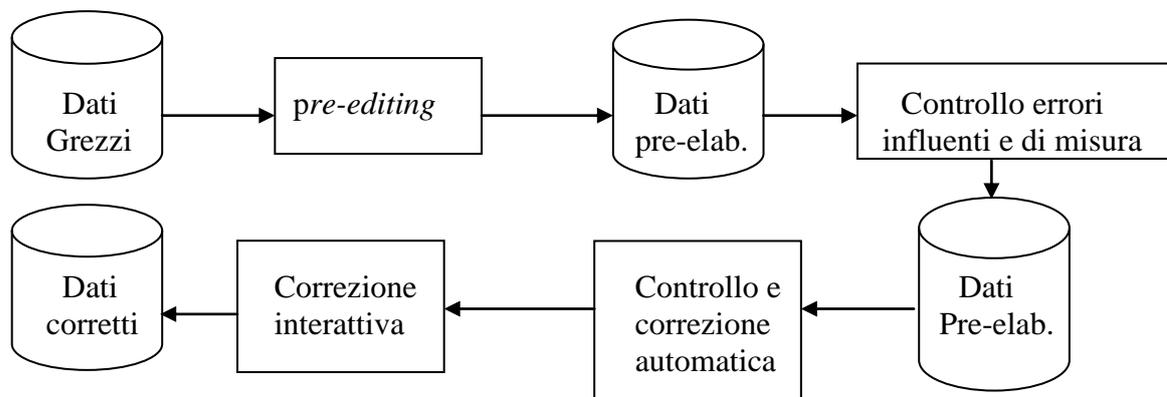
Il questionario dell'indagine è costituito da 9 sezioni (vedi allegato 1), di cui le prime due e l'ultima sono composte principalmente da variabili di tipo qualitativo il cui trattamento esula dagli scopi dell'applicazione, e che pertanto non sono state prese in considerazione. Fra le variabili di tipo numerico del questionario sono state esaminate quelle relative alle sezioni "*Struttura dei costi*" (sezione 3 del questionario), "*Giacenze e scorte*" (sezione 4), "*Acquisti e vendite di capitali fissi nell'esercizio*" (sezione 5), "*Reimpieghi*" (sezione 6) ed infine "*Ricavi*" (sezione 7). Per la sezione 8 ("*Contributi*") si ritiene più appropriato un controllo di tipo non probabilistico, per i motivi illustrati nel seguito. Sulle 2067 aziende rispondenti sono state esaminate complessivamente 72 variabili numeriche.

Propedeutica all'applicazione della procedura di controllo e correzione descritta nel lavoro è l'effettuazione sui dati di una serie di controlli preliminari (*pre-editing*) riguardanti principalmente la verifica di errori sui codici identificativi delle aziende agricole.

La procedura oggetto del presente lavoro è composta dalle seguenti tre fasi principali (vedi figura 1):

1. individuazione interattiva di errori di misura o di registrazione gravi oppure influenti (Fase 1)
2. controllo e correzione automatica dei dati (Fase 2)
3. revisione ed eventuale correzione interattiva degli errori di cui al punto 1 e dei casi residui al punto 2 (Fase 3)

Figura 1 – Schema generale della procedura di controllo e correzione



Nella prima fase vengono individuati in modo deterministico alcuni errori che non possono essere trattati efficientemente da GEIS (perché di tipo sistematico o per i quali non è comunque plausibile l'ipotesi di casualità) e vengono segnalati per il controllo interattivo questionari con potenziali errori di misura gravi su una o più variabili del questionario. Sempre in questa fase, mediante l'uso di una tecnica multivariata di tipo *selettivo*, viene selezionato un sottoinsieme di aziende che, violando almeno un vincolo del piano di controllo ed avendo potenzialmente un elevato impatto sulle stime finali di alcuni fenomeni di primario interesse, devono essere sottoposte a controllo interattivo. La fase 2 invece prevede il controllo e la correzione automatica dei dati utilizzando il software generalizzato di tipo probabilistico GEIS. Nella strategia proposta è previsto inoltre che l'output di GEIS costituisca a sua volta un input importante per la terza fase (la fase di controllo interattivo). Infatti, anche per i record da controllare interattivamente (record selezionati nella fase 1) è prevista l'applicazione del passo probabilistico di localizzazione degli errori e di eventuale imputazione. Ciò al fine di poter utilizzare le azioni di GEIS come *suggerimento* sui valori da modificare e sui possibili valori da assegnare per risolvere le incoerenze rilevate. E' chiaro che il principio di integrazione delle suddette tecniche non solo è possibile, ma bensì necessario, perché utilizzate separatamente esse sarebbero o non sufficienti a garantire i livelli di qualità richiesti (ad esempio procedendo con il controllo interattivo solo di alcuni dati lasciando inalterati gli altri) o errate (come l'utilizzo di GEIS per il trattamento di errori sistematici). Nell'ultima fase della procedura gli errori gravi o influenti ed i casi non risolvibili mediante l'approccio automatico sono sottoposti ad un processo di verifica che può essere di tipo interattivo (*follow-up*, revisione manuale dei

questionari ecc.), oppure di tipo deterministico. Quest'ultima fase si ricongiunge circolarmente all'inizio dell'intera procedura di controllo: infatti, ad esempio, attraverso l'analisi delle frequenze di violazione degli edit e delle frequenze di errore delle singole variabili, si possono avere delle indicazioni riguardo la validità delle regole di incompatibilità oppure riguardo la permanenza di errori sistematici nei dati. Una volta validato il piano di trattamento di questa tipologia di errori, si può ripartire ed eseguire l'intero piano di controllo e correzione dei dati.

3. Controllo degli errori influenti e degli errori di misura

Questa fase è volta all'individuazione nei dati grezzi (pre-elaborati) di alcune tipologie di errore non campionario che non possono essere adeguatamente trattate dagli algoritmi probabilistici implementati in GEIS. Gli errori oggetto del controllo in questo passo rientrano sostanzialmente nelle tre categorie seguenti:

1. errore nell'unità di misura generalizzato su almeno una delle sezioni del questionario;
2. errore di compilazione o registrazione su un addendo che però non si ripercuote sul corrispondente totale di sezione;
3. errore di compilazione o di registrazione su un addendo che si ripercuote sul totale di sezione.

In particolare nella tabella 1 vengono riportate le diverse tipologie di errore, le tecniche utilizzabili per individuarle e le conseguenti azioni possibili.

In termini generali, l'individuazione delle tipologie di errore di cui ai punti 1 e 3 è basata sull'analisi congiunta di alcuni rapporti caratteristici fra i totali di sezione e delle distribuzioni dei totali di sezione stessi all'interno di domini omogenei opportunamente definiti, nel paragrafo 3.1 è data una descrizione dettagliata della metodologia.

Il controllo degli errori di cui al punto 2, ed in particolare dei casi che si configurano come potenzialmente *influenti* sulle stime dei parametri di interesse (in questo caso i totali di sezione), è stato invece effettuato sulla base di una procedura selettiva multivariata in cui come misura dell'errore potenziale commesso sulle variabili di interesse è stata utilizzata la differenza fra il valore registrato di ciascuna variabile ed il valore ottenuto come somma dei corrispondenti addendi. Questa tecnica verrà descritta nel paragrafo 3.2.

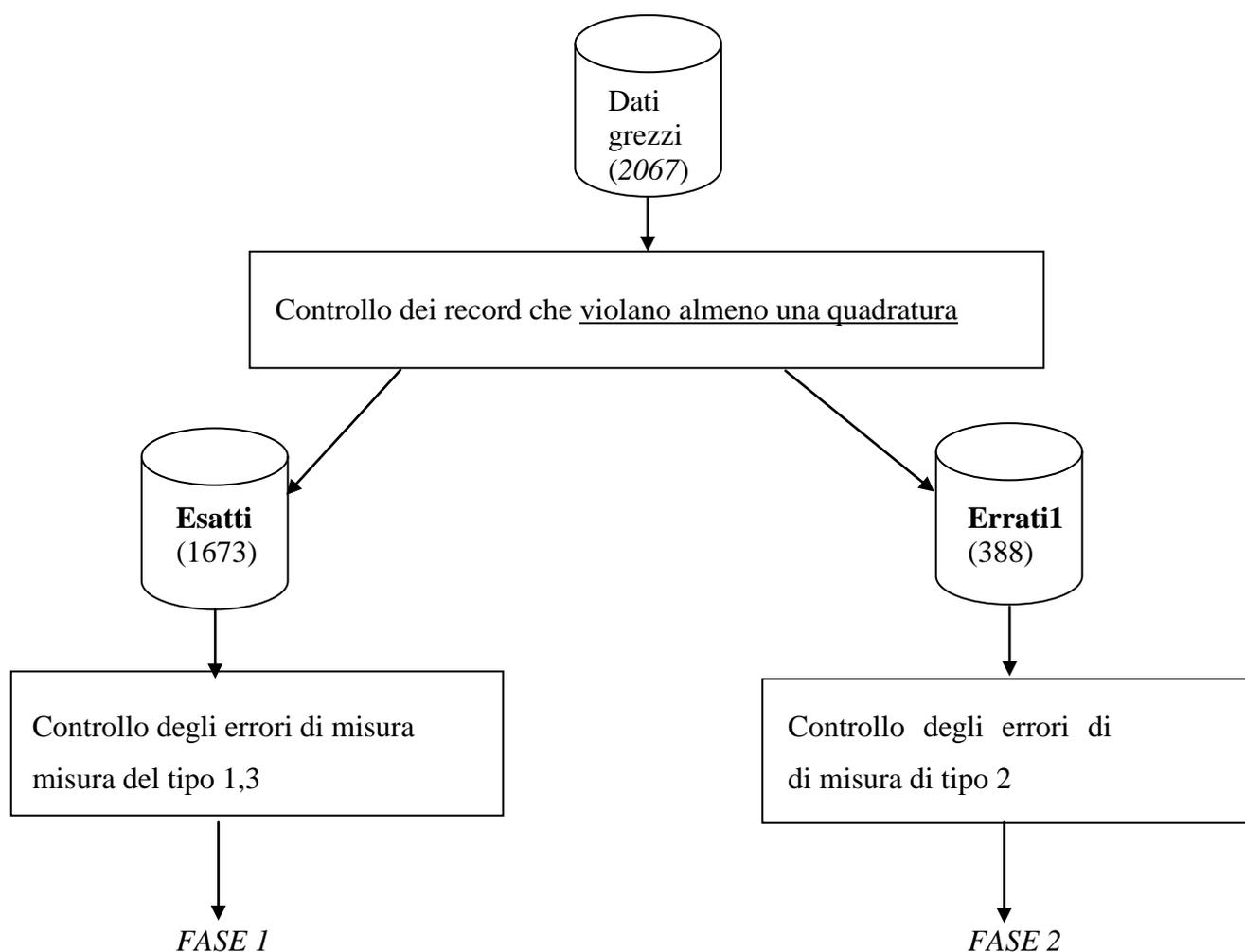
Tabella 1 - Tipologie di errore di misura

| <i>Tipologia di errore</i> | <i>Strumenti per la localizzazione</i> | <i>Strumenti per la correzione</i> |
|--|---|--|
| 1. Errore di unità di misura generalizzato su una delle sezioni del questionario | <ul style="list-style-type: none"> • Analisi delle distribuzioni dei totali di sezione nel campione • Analisi dei rapporti caratteristici fra totali di sezione | <ul style="list-style-type: none"> • Divisione per 1000 di tutte le variabili della sezione • Controllo interattivo |
| 2. Errore di compilazione o registrazione su un addendo che non si ripercuote sul corrispondente totale di sezione | <ul style="list-style-type: none"> • Analisi degli addendi > del totale corrispondente • Analisi dei rapporti fra addendi e totale corrispondente • Analisi della variabile 'Differenza fra totale registrato e totale calcolato' (editing selettivo) | <ul style="list-style-type: none"> • Imputazione del valore dell'addendo con la differenza fra il totale e la somma degli altri addendi • Divisione per 1000 • Introduzione di missing • Controllo interattivo |
| 3. Errore di registrazione o di compilazione che si ripercuote sul totale di sezione | <ul style="list-style-type: none"> • Analisi della distribuzione dei totali di sezione nel campione • Analisi dei rapporti fra addendi e totale corrispondente • Analisi dei rapporti caratteristici totali di sezione | <ul style="list-style-type: none"> • Divisione per 1000 degli addendi e dei totali • Introduzione di missing • Controllo interattivo |

Dal momento che una caratteristica delle aziende potenzialmente affette dalle tipologie di errore 1-3 rispetto a quelle con eventuali errori del tipo 2 è il fatto di violare o meno almeno uno dei vincoli di quadratura relativi ad ognuna delle 5 sezioni del questionario in esame, il trattamento di tali errori è stato preceduto da una fase preliminare in cui le aziende sono state classificate in due gruppi: un primo gruppo (costituito da 1673 aziende) in cui tutte le quadrature risultano soddisfatte, e per le quali si controllerà la presenza di errori del tipo 1 o 3, ed un secondo gruppo (costituito dalle rimanenti 388 unità) su cui sarà effettuata una ricerca dei valori errati influenti.

Il flusso dei dati relativo a questa fase è rappresentato nella figura 2.

Figura 2 – Flusso iniziale del controllo degli errori di misura e degli errori influenti



3.1. Controllo degli errori di registrazione o di misura gravi

In questa fase del controllo si è proceduto all'individuazione dei potenziali errori gravi nell'ambito delle aziende agricole "esatte" rispetto ai vincoli di quadratura. A questo scopo, ci si è basati sull'uso di alcuni rapporti caratteristici fra variabili significativamente correlate del questionario. In particolare, la ricerca di questi errori avviene sulla base di tre diversi indicatori di potenziale errore di misura, calcolati analizzando:

- rapporti caratteristici (ratio edits) fra totali di sezione significativamente correlati all'interno di domini corrispondenti alla classificazione per OTE (*Orientamento Tecnico-Economico* dell'azienda agricola);
- la presenza di un potenziale errore di misura generalizzato su una o più sezioni del questionario (basato sulla funzione $\text{mod} \lfloor x / 1000 \rfloor = 0$);

- la distribuzione per UDE (*Unità di Dimensione Economica* dell'azienda agricola) dei totali di sezione coinvolti nei rapporti caratteristici violati.

E' stato deciso che un'azienda agricola è da sottoporre a revisione interattiva nel caso in cui si verifichi il primo tipo di errore potenziale e congiuntamente ad esso almeno uno degli altri due.

Una volta definiti domini omogenei di aziende agricole sulla base della variabile OTE, sono state prese in considerazione le seguenti variabili:

- i tre totali di sezione del questionario *Costi, Acquisto di capitali fissi, Ricavi*
- la variabile $Giacenze = |Giacenze a fine mese - Giacenze a inizio mese|$

Sulla base dell'analisi delle correlazioni fra tali variabili sono stati selezionati i seguenti rapporti caratteristici:

- $R1 = Costi / Ricavi$
- $R2 = Costi / Giacenze$
- $R3 = Giacenze / Ricavi$
- $R4 = Acquisto di Capitali fissi / Ricavi$

Per ogni dominio, è stato quindi selezionato il sottoinsieme di rapporti caratteristici da utilizzare per il controllo sulla base del livello di correlazione esistente in quel dominio fra le variabili in essi coinvolte (vedi tabella 2).

Tabella 2 - Rapporti caratteristici utilizzati per dominio (dominio=OTE)

| <i>OTE</i> | <i>Rapporti considerati</i> | | | |
|------------|-----------------------------|-----------|-----------|-----------|
| 1 | <i>R1</i> | - | - | - |
| 2 | <i>R1</i> | <i>R2</i> | <i>R3</i> | <i>R4</i> |
| 3 | <i>R1</i> | - | <i>R3</i> | <i>R4</i> |
| 4 | <i>R1</i> | <i>R2</i> | - | - |
| 5 | <i>R1</i> | <i>R2</i> | <i>R3</i> | <i>R4</i> |
| 6 | <i>R1</i> | <i>R2</i> | - | - |
| 7 | <i>R1</i> | <i>R2</i> | <i>R3</i> | <i>R4</i> |
| 8 | <i>R1</i> | <i>R2</i> | <i>R3</i> | - |
| 9 | <i>R1</i> | <i>R2</i> | <i>R3</i> | - |

I limiti di accettazione dei rapporti caratteristici e delle distribuzioni univariate dei totali di sezione sono stati individuati utilizzando la metodologia D-MASO (*Distance Measurement Algorithm for the Selection of Outliers*) (Paletz, 1993; Winkler, 1994; Della Rocca e Luzi, 1997) e le potenzialità grafiche di SAS INSIGHT.

In generale, dato un edit-rapporto fra due variabili x_i e x_j del tipo:

$$L_{ij} \leq x_i/x_j \leq U_{ij}$$

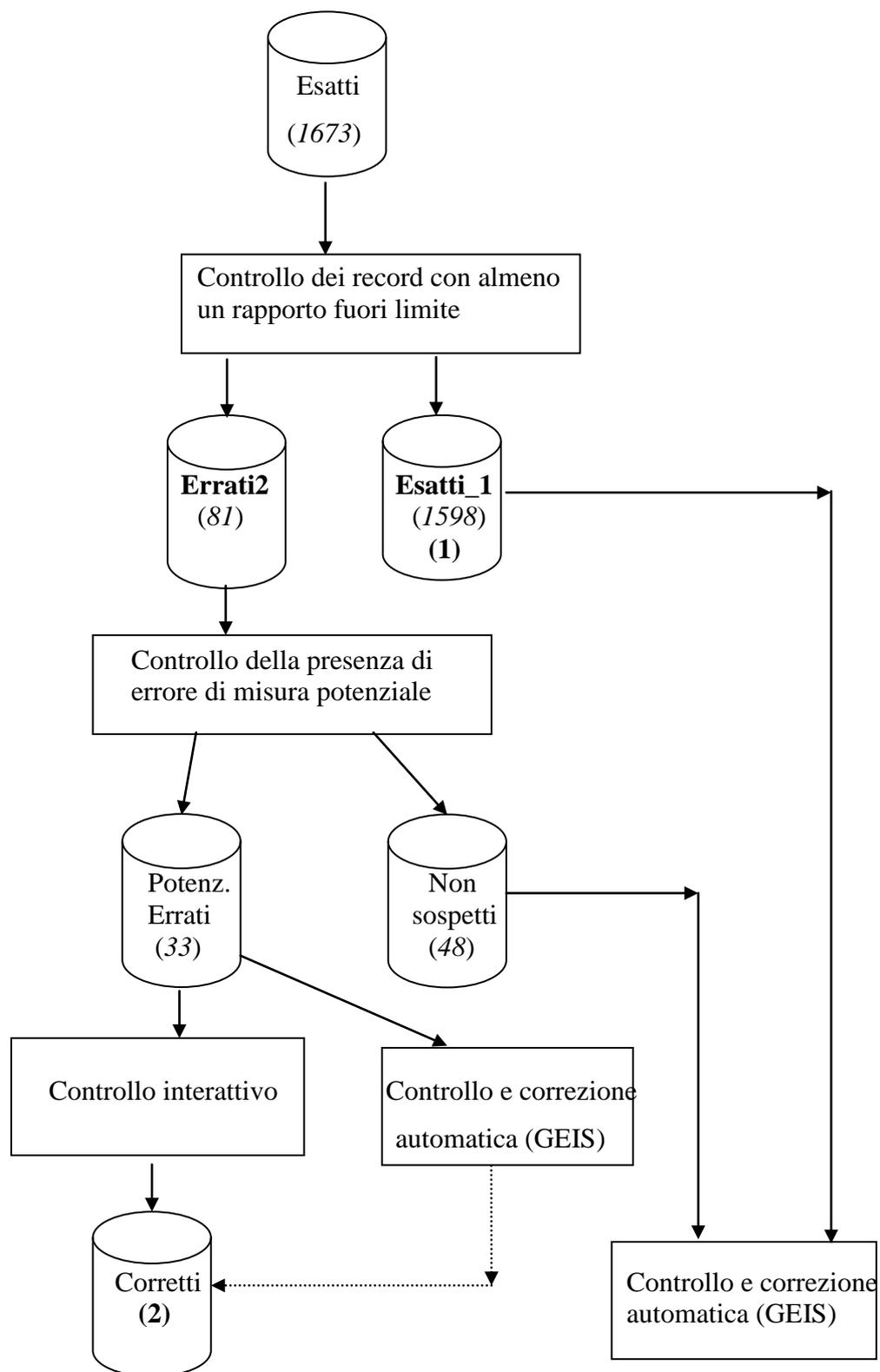
il *Distance Measurement Algorithm for Selection of Outliers* per la determinazione dei limiti L_{ij} ed U_{ij} è basato sul seguente assunto: anziché scegliere un intervallo di ampiezza tale da escludere una percentuale fissa di casi nelle code di sinistra e di destra della distribuzione il che può portare o a includere valori che sono outlier o ad escluderne di ammissibili, è preferibile procedere all'individuazione di quelli che sono i primi probabili outlier procedendo dal centro della distribuzione verso sinistra e destra. L'individuazione si basa sulla considerazione degli *intervalli* tra i valori dei quozienti: l'ipotesi è che *se c'è un intervallo tra due valori prossimi ad una estremità della distribuzione che è significativamente più ampio della maggior parte degli altri intervalli, il più estremo dei due valori è probabilmente un outlier*. Il meno estremo dei due valori è allora assunto come limite (inferiore o superiore, a seconda che ci si trovi nella coda sinistra o in quella destra) dell'intervallo di accettazione.

Nell'ambito della procedura di controllo descritta in questo paragrafo, mentre i limiti di accettazione dei rapporti caratteristici sono stati individuati per OTE, nel caso dell'analisi delle distribuzioni univariate di ciascun totale di sezione, la ricerca dei valori anomali è stata effettuata all'interno di domini individuati dalla variabile UDE, essendo tale variabile maggiormente esplicativa della variabilità dei fenomeni esprimenti la *dimensione* delle aziende agricole (in termini di costi, ricavi, ecc).

L'applicazione del sistema di indicatori così costruito ha consentito di individuare i seguenti insiemi di aziende:

- su 1673 aziende esatte rispetto ai vincoli di quadratura, 81 casi risultano violare almeno uno dei rapporti caratteristici applicati per strato; gli altri 1598 casi sono destinati al controllo automatico;
- degli 81 casi *sospetti*, 33 violano almeno uno degli altri due indicatori di errore potenziale: essi costituiscono un primo gruppo di aziende da sottoporre a controllo interattivo, con l'ausilio delle informazioni eventualmente fornite da GEIS sulla base dell'algoritmo probabilistico per la localizzazione degli errori, e da altre informazioni opportunamente ottenute. Il flusso dei dati e dei processi è illustrato nella figura 3.

Figura 3 - Flusso del controllo degli errori di misura del tipo 1,3



3.2. Controllo dei potenziali errori influenti

In questa fase vengono analizzati gli errori nelle aziende per le quali è attivata almeno una quadratura, al fine di individuare i casi influenti sulle stime dei totali delle variabili principali del questionario. Come variabili principali sono state considerate i totali delle sezioni *Costi*, *Giacenze*, *Acquisti di capitale fisso*, *Reimpieghi* e *Ricavi*.

Questo controllo viene effettuato attraverso una procedura multivariata di *editing selettivo* (Latouche, Berthelot 1992; Luzi, 1998). Tale procedura prevede la stratificazione preliminare delle aziende agricole per UDE ed il calcolo, per ogni azienda in ogni strato, di una funzione (detta *funzione punteggio*) che tiene conto:

- del numero di violazioni dei vincoli di quadratura;
- della dimensione dell'errore potenziale misurato come differenza fra valore registrato del totale e valore calcolato sulla base degli addendi del totale stesso;
- del peso campionario;
- del contributo alla stima preliminare dei totali di sezione coinvolti nelle quadrature violate dall'azienda nel dominio di appartenenza.

La funzione punteggio utilizzata per ogni unità *i-esima* del dominio *k-esimo* del campione è quindi del tipo (per non appesantire la notazione non verrà riportato l'indice relativo al dominio *k*)

$$f_{pi} = \sum_{j=1,m} [e_{ij} \pi_{ij} z_{ij}] \quad i=1, \dots, n$$

dove:

- *m* è il numero di vincoli di quadratura utilizzati;
- *e_{ij}* è la misura dell'errore presente nell'unità *i-esima* misurato dalla differenza fra valore registrato *Y_{ij}* e valore calcolato *Y*_{ij}* del totale di sezione *Y_j*;
- la quantità

$$\pi_{ij} = \frac{Y_{ij} \omega_i}{\sum_{i=1}^{n_k} Y_{ij}} = \frac{Y_{ij} \omega_i}{\hat{Y}_j}$$

dove ω_i è il peso diretto dell'unità *i-esima* del campione, esprime l'influenza dell'unità *i* appartenente al dominio *k* sulla stima (preliminare) del totale \hat{Y}_j della variabile *Y_j* nel dominio *k*.

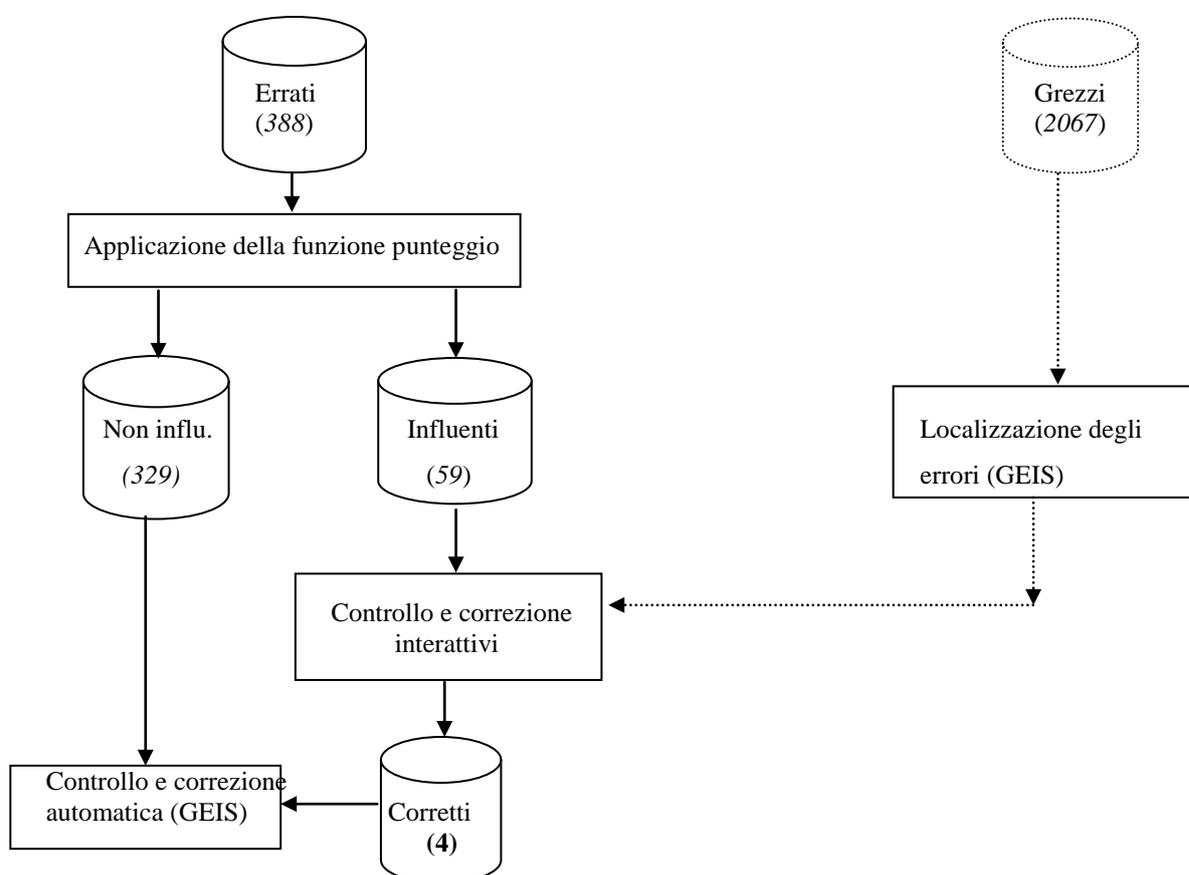
- *z_{ij}* è una variabile indicatrice che vale 1 se il vincolo di quadratura relativo al totale *Y_j* è violato dall'unità *i-esima* e 0 altrimenti.

La procedura di selezione delle unità influenti nel campione è pertanto formata dai passi seguenti:

- stratificazione del campione in 9 domini corrispondenti ciascuno ad una classe UDE;
- per ogni dominio, selezione delle unità statistiche che violano almeno un vincolo di quadratura del questionario e calcolo della f_{pi} ;
- in ogni dominio, ordinamento dei valori della f_{pi} in senso decrescente e visualizzazione grafica della corrispondente distribuzione;
- a partire dalle aziende con punteggio più alto, sulla base dell'analisi interattiva delle caratteristiche di ogni azienda e della distribuzione della f_{pi} , selezione delle unità statistiche ritenute critiche e pertanto da sottoporre a revisione accurata.

Sulla base di questa procedura, dei 388 casi errati rispetto ai vincoli di quadratura, 59 sono stati classificati come influenti e destinati al controllo interattivo, avvalendosi anche in questo caso dell'ausilio delle informazioni fornite dal processo di localizzazione degli errori effettuato con GEIS. Il flusso delle operazioni effettuate è descritto nella figura 4.

Figura 4 – Flusso del controllo degli errori di misura del tipo 2



4. Controllo e correzione automatica mediante GEIS

Il controllo e la correzione automatica dei dati sono stati effettuati utilizzando le tecniche implementate nel software automatico GEIS (Cotton, 1991; Kovar et al. 1988; Barcaroli e Luzi, 1995). Realizzato in linguaggio C, il sistema è disponibile in ISTAT nella sua versione UNIX. In particolare GEIS richiede la presenza di un DBMS ORACLE per la costruzione del data base dei dati da sottoporre a controllo e correzione.

Per il controllo dei dati GEIS implementa un approccio basato sulla metodologia probabilistica di *Fellegi-Holt*. Il sistema è costituito da un insieme di *moduli*, ciascuno dei quali implementa una particolare sottofunzione di una delle funzioni principali (definizione del piano dei controlli, individuazione degli errori, imputazione, individuazione degli outlier) in esso previste.

Per il controllo e la correzione automatica dei dati è stato predisposto dagli esperti dell'indagine un piano di incompatibilità composto da vincoli di quadratura ed esistenza, dove per vincoli di quadratura intendiamo che la variabile "totale" (ove presente nel questionario) deve essere uguale alla somma degli addendi corrispondenti, mentre un vincolo di esistenza esprime che se esiste una variabile, necessariamente deve essere presente una variabile collegata ad essa, e.g se la voce dei costi per alimenti per animali è maggiore di zero allora necessariamente l'azienda deve avere degli animali, e viceversa. L'elenco delle regole utilizzate è riportato nell'allegato 2.

Al fine di validare il piano dei controlli, di effettuare una prima valutazione dell'ammontare dei casi contenenti errori e di fare una prima analisi delle tipologie di errori presenti nei dati, i 2067 dati del campione sono stati preliminarmente analizzati con GEIS sfruttando il modulo che fornisce statistiche generali sugli effetti dell'applicazione ai dati stessi del piano di compatibilità (*Summary Statistic Tables*). In generale, le aziende per le quali risulta violato almeno un vincolo sono 531 (25,7%). Di queste, 457 ne violano solo uno (22,1%), 56 violano 2 soli edit, mentre le rimanenti 18 aziende violano non più di 4 vincoli. In particolare, dall'analisi delle frequenze di violazione per edit fornita da GEIS, si ha che:

- tra i vincoli di quadratura, quello relativo alla sezione dei costi è violato da 286 aziende, mentre le quadrature delle altre sezioni sono violate da un minimo di 5 volte (sezione *Acquisto di capitali fissi*) ad un massimo di 67 volte (sezione *Ricavi*).

Ciò porta ad una prima considerazione relativa al fatto che essendo la sezione dei costi formata da un numero di variabili elevato, per migliorare la qualità dei dati e facilitarne il controllo sarebbe opportuno apportare una modifica sul questionario inserendo dei sub-totali in modo opportuno

- dei 48 vincoli di esistenza, solo 28 sono violati almeno una volta. Di questi, due vincoli hanno frequenze di attivazione molto elevate (49 volte l'edit ESIS3_1B, 28 volte ESIS3_41). Questa circostanza rende necessaria una riflessione su tali edit, in quanto va verificato che essi debbano essere necessariamente soddisfatti affinché un questionario possa essere considerato accettabile, e che tali vincoli siano ben specificati, nel senso che mettano in relazione in modo corretto tutte e sole le variabili necessarie al controllo voluto.

4.1. Disegno della procedura di controllo e correzione automatica

Il processo di controllo e correzione automatica avviene all'interno di gruppi di dati (*domini*) che corrispondono alla variabile *UDE*. Questa scelta è stata fatta per tenere conto dell'omogeneità delle aziende agricole rispetto a tale variabile nella fase di imputazione con donatore: in questo modo infatti i potenziali donatori vengono scelti fra le unità con stessa dimensione economica.

Come già detto, la procedura automatica di controllo e correzione non riguarda la sezione dei *Contributi*, effettivamente priva di legami logico-matematici con le altre sezioni e per la quale esiste un unico criterio di correttezza rispetto al quale i dati devono risultare esatti (vincolo di quadratura della sezione). Per questo motivo si ritiene che un trattamento di tipo automatico con GEIS sia non raccomandabile, mentre un approccio di tipo deterministico (in parte interattivo ed in massima parte automatico del tipo *prorating*¹) possa essere più adatto. Questo è dovuto fondamentalmente al fatto che l'algoritmo Fellegi-Holt utilizzato da GEIS per individuare gli errori si basa sul criterio del minimo cambiamento, e quindi associa ad ogni variabile una maggiore o minore probabilità di essere considerata errata dipendentemente dal numero di vincoli violati in cui essa è coinvolta. Ovviamente nel caso in cui i vincoli formino una partizione dello spazio delle variabili (ovvero ogni variabile appare in un uno ed un solo edit) la probabilità che una variabile sia considerata errata risulta essere uguale a tutte le altre,

¹ Per *prorating* si intende comunemente la redistribuzione proporzionale fra gli addendi ed il totale della differenza fra totale osservato e totale calcolato

ciò si traduce nella pratica in una selezione assolutamente casuale della variabile da correggere. In questo caso potrebbe essere rischioso quindi affidare interamente al sistema automatico il controllo, e quindi la correzione, del dato.

La procedura automatica messa a punto è costituita dalle due fasi principali descritte nel seguito.

PASSO 1 - Controllo e correzione automatica dell'insieme di aziende agricole esatte rispetto ai vincoli di quadratura ed ai rapporti caratteristici (vedi paragrafo 3.1)

In questo caso, il controllo automatico ha il solo obiettivo di verificare l'esattezza dei dati rispetto ai vincoli di esistenza del piano di compatibilità. Ovviamente il piano di controllo e correzione contiene anche i vincoli di quadratura, necessari per garantire la correttezza finale dei dati in seguito al processo di imputazione con donatore.

Dal momento che per le aziende di questo gruppo (che nel seguito chiameremo *esatti_1*) è possibile avere una elevata fiducia sull'affidabilità dei totali delle sezioni del questionario - sia perché i vincoli di quadratura non sono violati, sia perché i rapporti caratteristici calcolati sui totali di sezione sono entro i limiti di accettazione - in fase di localizzazione dell'errore si è fatto in modo che la procedura automatica non potesse modificarli, ciò è stato ottenuto tecnicamente assegnando ad essi il massimo peso. Successivamente alla fase di localizzazione dell'errore è stata fatta una prima imputazione applicando il modulo di correzione denominato in GEIS "*correzione deterministica*" che consente di risolvere i casi in cui l'errore individuato ha un'unica soluzione analitica, per esempio il caso in cui una quadratura è violata ed un unico addendo della quadratura è considerato errato: in questo caso la soluzione è ottenibile analiticamente. Successivamente si è proceduti all'imputazione con donatore di minima distanza. Il processo di imputazione con donatore è stato effettuato separatamente per i 9 domini di aziende agricole definiti sulla base dell'UDE, soprattutto allo scopo di garantire in fase di imputazione con donatore la similitudine fra record da imputare e record donatori rispetto a tale criterio. In questa ultima fase, al fine di aumentare la probabilità di individuare donatori adatti per i vari record errati, sono stati utilizzati insiemi di regole di controllo rilassate (*edit di post-imputazione*) in cui i vincoli di quadratura sono stati sostituiti da intervalli centrati sul valore del totale e con una ampiezza pari al 50% di quest'ultimo, per esempio se abbiamo che un vincolo è $X+Y=Z$ esso viene sostituito da $0.5 * Z \leq X+Y \leq 1.5 * Z$.

Ovviamente, per garantire comunque la coerenza finale dei dati rispetto ai vincoli originali, si è effettuato un ulteriore passo di localizzazione dell'errore (solo sulle variabili imputate) e ad una ulteriore operazione di correzione utilizzando i vincoli nella loro forma originale. I risultati ottenuti in questa fase del processo sono descritti nel paragrafo 4.2.1.

PASSO 2 - Controllo e correzione automatica dei questionari che violano almeno un vincolo di quadratura (ad esclusione dei casi riconosciuti influenti sulla base della procedura automatica descritta al paragrafo 3.2) e dei questionari che, pur violando almeno un rapporto caratteristico, sono risultati non sospetti di contenere errori di misura o di registrazione gravi sulla base dell'analisi descritta nel paragrafo 3.1.²

Anche per questo gruppo di aziende (a cui ci si riferirà nel seguito come **Errati**) il piano di controllo è costituito dai vincoli di esistenza e di quadratura utilizzati al passo precedente. Non sono stati utilizzati invece vincoli di tipo rapporto in quanto il loro uso presuppone, in generale, la non accettabilità di tutti i casi in cui le variabili coinvolte assumono valore nullo. Questo non è il caso dei totali delle sezioni del questionario in esame, essendo possibile e in molti casi verosimile la completa mancanza di importi all'interno di una o più sezioni. La conseguenza di questa scelta è stata la permanenza nei dati trattati automaticamente di casi in cui alcuni rapporti caratteristici risultano fuori limite (cfr. paragrafo 5): ciò renderà necessario tenere sotto controllo la gravità e/o l'influenza sulle stime finali di tali casi in fase di post-editing.

Dal momento che per le aziende di questo secondo gruppo non è possibile avere fiducia sull'affidabilità dei totali di sezione, in fase di localizzazione dell'errore non si è assegnato ad essi alcun peso, rendendo equiprobabile la loro selezione per l'imputazione rispetto a quella delle altre variabili. Anche in questo secondo insieme di dati, successivamente alla localizzazione dell'errore, è stato applicato il modulo GEIS di "*correzione deterministica*" descritto precedentemente.

² In realtà, in questo passo vengono sottoposte a controllo e correzione automatica anche le aziende ritenute influenti ed i casi sospetti di contenere errori di misura o di registrazione gravi, ma in questi casi l'obiettivo è solo quello di ottenere ulteriori informazioni su tali aziende dal trattamento di tipo probabilistico (variabili ritenute da imputare, valori imputati). Questi casi saranno comunque trattati in modo interattivo.

I dati sono stati poi sottoposti a correzione col metodo del donatore utilizzando come possibili donatori anche le aziende controllate e corrette nel passo 1 precedente, in modo da aumentare la probabilità di individuare donatori accettabili per i record errati.

In fase di imputazione con donatore sono stati utilizzati gli stessi insiemi di regole di post-imputazione del primo gruppo di aziende, ed anche in questo caso si è effettuato un ulteriore passo di localizzazione dell'errore (solo sulle variabili imputate) e ad una operazione di correzione utilizzando i vincoli nella loro forma originale. I risultati ottenuti in questa fase del processo sono descritti nel paragrafo 4.2.2.

Il flusso dei dati e delle elaborazioni che, a partire dalla selezione delle aziende del primo gruppo, ha portato agli insiemi finali dei dati (corretti interattivamente, corretti automaticamente, non corretti per eccessiva complessità o per mancanza di donatori) è illustrato nel diagramma riportato nella figura 5. Come si può vedere, il controllo e la correzione automatica ha coinvolto tutte le 2067 aziende iniziali, comprese quindi le aziende che in realtà dovranno subire un trattamento interattivo perché contenenti valori influenti o potenziali errori di misura gravi.

Una volta effettuate le operazioni di selezione dei potenziali errori sistematici ed influenti (vedi paragrafo 3) ed ottenuti i due insiemi di dati da trattare automaticamente *Esatti_1* ed *Errati*, si procede al primo passo della procedura di controllo automatico. I dati vengono caricati in una tabella ORACLE tramite un'apposita procedura SAS (PROC DBLOAD), specificando i formati delle variabili da trattare onde evitare problemi di arrotondamento e/o troncamento dei valori.

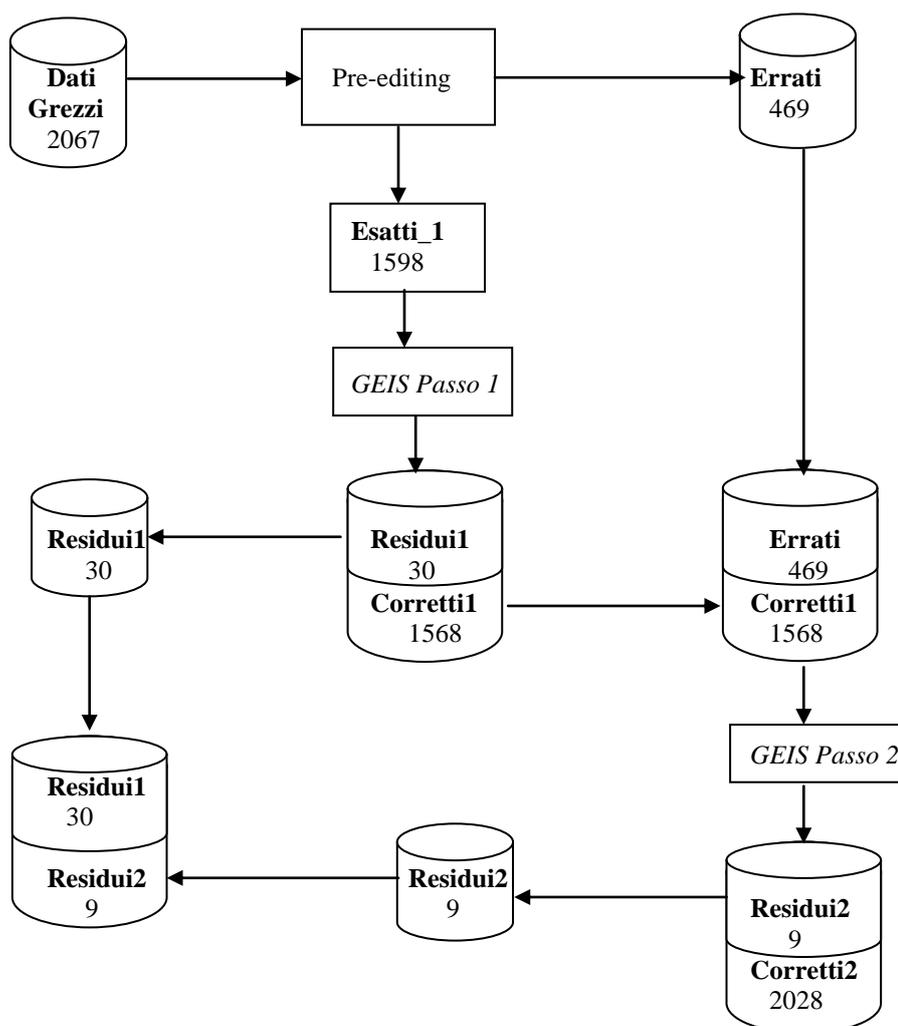
Al termine delle elaborazioni relative al primo passo della procedura automatica, dell'insieme dei 1598 casi trattati, 1568 risultano soddisfare il piano di controllo mentre per 30 record il sistema non è stato in grado di trovare una soluzione (in fase di localizzazione o di imputazione degli errori). Di queste, 15 aziende non sono state risolte o perché non è stato trovato un donatore plausibile, o perché dopo l'imputazione con donatore non è stato possibile fare in modo che i vincoli di compatibilità iniziali siano soddisfatti.

Queste 30 aziende (*Residui1*) vengono escluse dalle elaborazioni automatiche successive, per essere destinate ad un controllo di tipo interattivo. In alternativa si poteva utilizzare l'imputazione con stimatore.

Le rimanenti 1568 aziende vengono accorpate con le 469 aziende del file *Errati* perché accrescano il serbatoio dei potenziali donatori nel secondo passo della procedura automatica. I dati così ottenuti vengono nuovamente caricati in ORACLE per essere

elaborati successivamente elaborati. Delle 2037 aziende trattate in questo secondo passo, 9 (**Residui2**) risultano non corrette a causa dell'impossibilità di trovare una soluzione di minimo cambiamento nei tempi concessi (TLE), mentre i rimanenti 2028 record risultano soddisfare l'intero piano dei controlli.

Figura 5 - Flusso generale delle elaborazioni automatiche dei dati



Si fa notare che l'insieme delle 92 aziende (59 casi influenti, 33 potenziali errori di misura gravi) selezionate per il controllo interattivo, si trovano: all'inizio della procedura automatica nel file **Errati**, mentre al termine di essa 89 sono nell'insieme **Corretti2** mentre 3 sono in **Residui2**. L'insieme dei dati selezionati per il trattamento automatico è stato pertanto ottenuto estraendo da **Corretti2** le 89 aziende di cui sopra, e le 6 aziende contenenti valori nulli per tutte le variabili del questionario, ottenendo il un insieme di 1933 aziende (93,5 % del totale) composte da aziende corrette sin dall'inizio ed aziende controllate e corrette in modalità completamente automatica. Allo stesso

modo, eliminando da **Residui2** le 3 aziende da verificare interattivamente ed unendo le rimanenti 6 a **Residui1**, si ottiene il file dei 36 casi residui per i quali GEIS non ha trovato soluzione (1,74% del totale).

4.2.1 Risultati del primo gruppo di aziende (quadrature e rapporti caratteristici esatti)

Questo gruppo di dati è composto da 1598 unità di cui 1465 risultano essere corrette rispetto alle regole di compatibilità composte da vincoli di quadratura ed esistenza e le rimanenti 133 violano almeno una delle regole di esistenza. Di questi 133 record la maggior parte viola una sola regola di esistenza mentre le altre ne violano da un minimo di due ad un massimo di quattro contemporaneamente. Tutte queste informazioni, unitamente ad altre, sono riassunte nella tavola (output di GEIS) delle "summary statistics" mostrata in appendice (allegato 3). Prima di passare alla fase effettiva di controllo e correzione, è stata effettuata un'analisi (tramite un modulo di GEIS) della coerenza e ridondanza del piano di vincoli utilizzato. Questo controllo non ha messo in evidenza ridondanze nè tantomeno incoerenze fra gli edit introdotti.

Successivamente a questa fase di descrizione del gruppo di dati ed analisi degli edit si è passati alla fase di localizzazione di errore. Questa procedura, come tutte le altre utilizzate in seguito, agirà per singoli sottogruppi di dati caratterizzati dall'appartenenza alle singole UDE. Al fine di utilizzare al meglio le risorse della CPU, in un primo momento in questa analisi è stato dato un limite di tempo massimo per trovare l'errore per l'algoritmo utilizzato da GEIS di 5 secondi, in seguito i record per cui non è stato possibile localizzare un possibile errore in questo lasso di tempo sono stati di nuovo sottoposti a localizzazione di errore con un tempo limite massimo di 240 secondi. Alla fine di questa procedura si ottengono record con indicazione delle variabili da correggere, record per cui non è possibile trovare una soluzione (MIR) e record per cui non è stato possibile trovare una soluzione per limiti di tempo (TLE). I record che abbiamo indicato con TLE e MIR dovranno essere corretti interattivamente o deterministicamente. Sottolineiamo che anche aumentando considerevolmente i 240 secondi di tempo limite il numero di record TLE rimane sostanzialmente invariato. Diamo di seguito una tabella che illustra il numero di record totali, errati, corretti, TLE e MIR, tutto stratificato per UDE.

| | <i>N° record</i> | <i>N° record errati</i> | <i>TLE</i> | <i>MIR</i> |
|-------------------|------------------|-------------------------|------------|------------|
| <i>UDE1</i> | 129 | 29 | 3 | 1 |
| <i>UDE2</i> | 99 | 9 | 0 | 0 |
| <i>UDE3</i> | 63 | 7 | 1 | 0 |
| <i>UDE4</i> | 60 | 5 | 1 | 0 |
| <i>UDE5</i> | 66 | 6 | 1 | 0 |
| <i>UDE6</i> | 42 | 4 | 1 | 0 |
| <i>UDE7</i> | 390 | 22 | 0 | 0 |
| <i>UDE8</i> | 520 | 34 | 4 | 0 |
| <i>UDE9</i> | 229 | 17 | 3 | 0 |
| <i>TOT</i> | 1598 | 133 | 14 | 1 |

In questa prima fase notiamo che l'11.3% dei dati errati non può essere risolto automaticamente. Il passo successivo prevede la prima operazione di imputazione, denominata in GEIS "deterministica", in cui vengono corretti quei campi per cui la soluzione è univocamente ed analiticamente determinata (vedi paragrafo precedente). Per le relative statistiche rimandiamo alle tabelle in appendice (allegato 4) dove sono inoltre riportati per i campi cambiati sia il valore originario che quello modificato. Conclusa questa prima fase di imputazione, per i record ancora da correggere si passa all'imputazione con donatore. Qui vengono scelti i donatori con minima distanza rispetto al record da imputare. Tra i donatori più vicini vengono accettati quelli i cui valori, una volta sostituiti ai campi da imputare, soddisfano gli edit di post-imputazione (edit rilassati). Anche con questo accorgimento può però accadere che per alcuni record non sia possibile trovare un donatore. Nella tabella successiva mostriamo il numero di record per cui si è cercato un donatore ed il numero di record per cui non è stato possibile trovare un donatore.

| | <i>Record da imputare</i> | <i>Record senza donatore</i> |
|-------------------|---------------------------|------------------------------|
| <i>UDE1</i> | 20 | 5 |
| <i>UDE2</i> | 5 | 0 |
| <i>UDE3</i> | 5 | 0 |
| <i>UDE4</i> | 3 | 0 |
| <i>UDE5</i> | 5 | 2 |
| <i>UDE6</i> | 3 | 1 |
| <i>UDE7</i> | 19 | 2 |
| <i>UDE8</i> | 21 | 1 |
| <i>UDE9</i> | 10 | 1 |
| <i>TOT</i> | 91 | 12 |

Queste frequenze sono state tratte dal rapporto di GEIS riportato in appendice (allegato 5) e da cui è inoltre possibile ottenere un'altra importante informazione relativa al numero di volte che un record è stato usato come donatore, infatti più è alto il numero

record usati una sola volta come donatori, più è preservata la variabilità della popolazione. E' evidente come effettivamente questa richiesta sia soddisfatta nell'applicazione in esame. Per i 12 casi che dopo questa fase devono essere ancora da imputare, sono disponibili più soluzioni: essi potranno andare a far parte dell'insieme di record che devono essere imputati interattivamente, oppure potrebbe essere utilizzato il modulo di GEIS per l'imputazione dei dati con stimatore, oppure si potrebbe non delimitare il bacino di donatori alle sole aziende appartenenti alla stessa UDE ma ampliarlo a quelle vicine, aumentando così il numero di potenziali donatori.

I 79 record per cui è stato possibile trovare un donatore potenzialmente soddisfano ora solo gli edit rilassati, mentre invece un record per essere considerato corretto deve soddisfare gli edit iniziali di quadratura ed esistenza. Questi 79 record vengono quindi sottoposti nuovamente al passo di localizzazione dell'errore, sottolineando però che l'algoritmo esaminerà solo i campi che sono stati imputati con donatore. Anche in questo caso non è assicurato che sia trovata una soluzione. Riportiamo nella tabella il numero dei record tra quelli imputati che non soddisfano le regole di compatibilità iniziali, fra questi il numero di record per cui è stato possibile trovare una soluzione:

| | <i>N° record da imputare</i> | <i>N° record con soluzione</i> |
|-------------------|------------------------------|--------------------------------|
| <i>UDE1</i> | 12 | 12 |
| <i>UDE2</i> | 5 | 5 |
| <i>UDE3</i> | 4 | 4 |
| <i>UDE4</i> | 3 | 2 |
| <i>UDE5</i> | 3 | 3 |
| <i>UDE6</i> | 2 | 2 |
| <i>UDE7</i> | 15 | 15 |
| <i>UDE8</i> | 19 | 17 |
| <i>UDE9</i> | 8 | 8 |
| <i>TOT</i> | 71 | 68 |

Per 3 dei 71 record non stato è possibile trovare soluzione, per cui anche questi si aggiungeranno a quelli che devono essere corretti interattivamente, o con stimatore, o deterministicamente. I rimanenti 68 record vengono sottoposti ad una ultima correzione "deterministica" (secondo il significato di GEIS). Anche da questa ultima fase si avranno dei campi per cui non è possibile trovare una soluzione unica (caratteristica della imputazione "deterministica" di GEIS), e quindi anche questi verranno risolti interattivamente.

Ricapitolando, su 133 record errati GEIS non ne corregge automaticamente 15, ovvero l'11.3%. Per altre 15 unità è possibile procedere, come già detto, a correzione

interattiva o con stimatore. Nel caso di correzione interattiva o comunque esterna a GEIS, queste unite alle precedenti formeranno un insieme di 30 aziende. Giova comunque osservare che il 40% di questi 30 record, e precisamente 12, non sono stati risolti perché non è stato possibile trovare un donatore per essi. Questo fattore dipende chiaramente dalla dimensione del campione disponibile, e di conseguenza dal basso numero di possibili donatori, ci si aspetta quindi che, nel caso di applicazione all'indagine a regime, questo problema subisca un naturale un ridimensionamento.

4.2.2 *Secondo gruppo di aziende (quadrature o rapporti caratteristici violati)*

Questo gruppo di dati è composto da 2037 unità di cui 1639 (a cui si aggiungerebbero quelle corrette con stimatore) risultano essere corrette rispetto alle regole di compatibilità del piano di controllo e le rimanenti 398 violano almeno una di tali regole. Di questi 398 record la maggior parte viola vincoli di quadratura, in particolare la quadratura relativa alla sezione dei costi è violata ben 286 volte (vedi par. 4). Anche in questo caso la maggior parte delle volte viene violato un solo edit (340 volte). Tutte queste informazioni, unitamente ad altre, sono riassunte nella tavola (output di GEIS) delle "*summary statistics*" mostrata in appendice (allegato 6). In questo caso non è stato necessario effettuare il controllo della coerenza e ridondanza del piano di vincoli utilizzato poiché il piano di vincoli è lo stesso del piano precedente.

Successivamente a questa fase di descrizione del gruppo di dati ed analisi degli edit si è passati alla fase di localizzazione di errore. Analogamente al passo precedente si agirà per singoli sottogruppi di dati caratterizzati dall'appartenenza alle singole UDE. Anche in questo passo, al fine di utilizzare al meglio le risorse della CPU, in un primo momento in questa analisi è stato dato un limite di tempo massimo per trovare l'errore per l'algoritmo utilizzato da GEIS di 5 secondi, in seguito i record per cui non è stato possibile localizzare un possibile errore in questo lasso di tempo sono stati di nuovo sottoposti a localizzazione di errore con un tempo limite massimo di 240 secondi. Alla fine di questa procedura si ottengono record con indicazione delle variabili da correggere, record per cui non è possibile trovare una soluzione (MIR) e record per cui non è stato possibile trovare una soluzione per limiti di tempo (TLE). I record che abbiamo indicato con TLE e MIR dovranno essere corretti interattivamente. Diamo di seguito una tabella che illustra il numero di record totali, errati, corretti, TLE e MIR, tutto stratificato per UDE:

| | <i>N° record</i> | <i>N° record errati</i> | <i>TLE</i> | <i>MIR</i> |
|-------------------|------------------|-------------------------|------------|------------|
| <i>UDE1</i> | 143 | 14 | 1 | 0 |
| <i>UDE2</i> | 112 | 8 | 0 | 0 |
| <i>UDE3</i> | 79 | 14 | 1 | 0 |
| <i>UDE4</i> | 63 | 5 | 0 | 0 |
| <i>UDE5</i> | 73 | 7 | 0 | 0 |
| <i>UDE6</i> | 52 | 9 | 1 | 0 |
| <i>UDE7</i> | 523 | 116 | 0 | 0 |
| <i>UDE8</i> | 680 | 149 | 3 | 0 |
| <i>UDE9</i> | 312 | 76 | 3 | 0 |
| <i>TOT</i> | 2037 | 398 | 9 | 0 |

In questa prima fase notiamo che il 2.3% dei dati errati non può essere risolto automaticamente. Per le statistiche relative alla successiva fase di imputazione "deterministica" rimandiamo alle tabelle in appendice (allegato 7) dove sono inoltre riportati per i campi cambiati sia il valore originario che quello modificato. Conclusa questa prima fase di imputazione, per i record ancora da correggere si passa all'imputazione con donatore. Tra i donatori più vicini vengono accettati quelli i cui valori, una volta sostituiti ai campi da imputare, soddisfano gli edit di post-imputazione (edit rilassati), gli edit di post-imputazione utilizzati sono gli stessi introdotti per il primo gruppo di dati. Nella tabella successiva mostriamo il numero di record per cui è stato cercato un donatore, ed il numero di record per cui non è stato possibile trovarne uno accettabile:

| | <i>Record da imputare</i> | <i>Record senza donatore</i> |
|-------------------|----------------------------------|-------------------------------------|
| <i>UDE1</i> | 3 | 0 |
| <i>UDE2</i> | 1 | 0 |
| <i>UDE3</i> | 0 | 0 |
| <i>UDE4</i> | 0 | 0 |
| <i>UDE5</i> | 1 | 0 |
| <i>UDE6</i> | 0 | 0 |
| <i>UDE7</i> | 2 | 0 |
| <i>UDE8</i> | 3 | 0 |
| <i>UDE9</i> | 1 | 0 |
| <i>TOT</i> | 11 | 0 |

Notiamo che per tutti i record è stato possibile trovare un donatore. Come nel paragrafo precedente rimandiamo in appendice (allegato 8) per ottenere l'informazione relativa al numero di volte che ogni record è stato usato come donatore. Gli 11 record per cui è stato possibile trovare un donatore potenzialmente soddisfano ora solo gli edit rilassati, per cui essi vengono sottoposti nuovamente al passo di localizzazione

dell'errore sulla base dei vincoli in forma originale, sottolineando però che l'algoritmo esaminerà solo i campi che sono stati imputati. Anche in questo caso non è assicurato che sia trovata una soluzione. Riportiamo nella tabella il numero dei record tra quelli imputati che non soddisfano le quadrature finali, fra questi il numero di record per cui è stato possibile trovare una soluzione:

| | <i>N° record da imputare</i> | <i>N° record con soluzione</i> |
|-------------------|------------------------------|--------------------------------|
| <i>UDE1</i> | 3 | 3 |
| <i>UDE2</i> | 1 | 1 |
| <i>UDE3</i> | 0 | 0 |
| <i>UDE4</i> | 0 | 0 |
| <i>UDE5</i> | 1 | 1 |
| <i>UDE6</i> | 0 | 0 |
| <i>UDE7</i> | 2 | 2 |
| <i>UDE8</i> | 3 | 3 |
| <i>UDE9</i> | 1 | 1 |
| <i>TOT</i> | 11 | 11 |

Notiamo che per tutti i record è stato trovato un campo da imputare e di conseguenza tutti sono stati sottoposti ad un ultimo passo di correzione "*deterministica*". In quest'ultima fase tutti i record vengono corretti con successo.

Ricapitolando, su 398 record errati GEIS ne corregge automaticamente 387, ovvero il 97.2% mentre i rimanenti 9 verranno trattati esternamente al sistema.

5. Analisi dei risultati

5.1. Analisi dei dati corretti da GEIS

In questo paragrafo si analizzano i risultati prodotti da GEIS sul sottoinsieme di aziende agricole indirizzate a controllo automatico e per le quali GEIS ha trovato soluzione nella fase di localizzazione dell'errore e di imputazione. Dei 531 record errati rispetto a vincoli di quadratura e/o di esistenza 434 (82%) sono stati corretti automaticamente. Esaminando media e deviazione standard per UDE e i livelli di correlazione delle variabili per OTE (vedi allegato 9) risulta che complessivamente la struttura univariata e di relazione bivariata delle variabili principali del questionario (totali di sezione) nel file dei dati grezzi è preservata, come ci auspicavamo e come l'algoritmo probabilistico basato sul minimo cambiamento e l'imputazione con donatore di minima distanza tendono a garantire.

Se da un punto di vista di macroanalisi possiamo ritenere soddisfacente il livello di qualità dei risultati ottenuti, da una microanalisi risulta che alcune tipologie di record necessitano ancora di una ulteriore validazione. Questa esigenza nasce dalla particolare struttura dei dati e degli edit dell'indagine in esame, infatti come accennato nel paragrafo 3, la maggior parte delle unità statistiche violano quasi esclusivamente vincoli di quadratura. Nel caso in cui una azienda viola solo edit di quadratura, in mancanza di altre informazioni (date dalla violazione di altre regole), l'algoritmo di localizzazione dell'errore basato sul minimo cambiamento sceglierà casualmente come variabile da imputare uno degli addendi della quadratura in questione. Questo evento è trascurabile nel caso in cui la differenza fra il totale di sezione e quello osservato nel questionario sia piccola, ma potrebbe essere non trascurabile nel caso opposto e quindi per maggior sicurezza si ritiene opportuna una analisi più approfondita dei risultati prodotti da GEIS per questo sottoinsieme di unità e di variabili. Per questo motivo è raccomandabile in fase di post-editing effettuare un ulteriore controllo dei casi influenti. Dei 434 record da correggere automaticamente:

- 313 violano solo quadrature (72%)
- 111 violano solo vincoli di esistenza (26%)
- 10 violano sia quadrature che esistenze (2%)

Dei 313 record che violano solo quadrature, solo 86 presentano una variazione fra dato grezzo e dato imputato di almeno una delle variabili maggiore del 70% esclusi i casi in cui si è passato da un valore nullo ad uno diverso da zero e viceversa.

Un ulteriore controllo è stato fatto selezionando i record che hanno subito una modifica per una data variabile superiore al 90° percentile della distribuzione delle differenze assolute fra dati grezzi e dati corretti; dei 22 casi selezionati, 16 sono già inclusi negli 86 precedenti.

Pertanto, dei 434 casi iniziali, 342 possono essere considerati corretti in modo soddisfacente dalla procedura automatica, quindi possiamo affermare il 78.8% dei dati errati che dovevano essere corretti automaticamente (corrispondente ad un ammontare pari al 64.4% di tutti i record errati di questa indagine), viene corretto automaticamente con successo.

Una ulteriore verifica fatta sulle aziende corrette da GEIS è stata basata sull'analisi per OTE dei rapporti caratteristici descritti nel paragrafo 3.1. Fra le aziende corrette automaticamente, vengono selezionate quelle per le quali almeno un rapporto è fuori limite (22 aziende), in particolare siamo interessati, al fine di valutare errori indotti da

GEIS, ai casi in cui è stato modificato uno dei totali coinvolti nel rapporto violato (2 aziende). Di questi due casi solo uno risulta essere stato modificato in modo non verosimile dalla procedura automatica: si tratta di un'azienda per la quale solo il vincolo di quadratura sulla sezione dei ricavi era violato e per la quale GEIS ha selezionato casualmente per l'imputazione il totale di sezione, rientrando così nella tipologia di record, discussa precedentemente. Notiamo che questo record era stato selezionato nella lista di quelli che violando solo quadrature avevano subito una modifica maggiore del 90° percentile della distribuzione delle differenze assolute tra valore grezzo e corretto.

5.2. Analisi dei dati per i quali è richiesto il controllo interattivo

Questi dati sono relativi alle aziende che:

1. contengono possibili errori di misura gravi (33 aziende, primo gruppo)
2. contengono valori influenti (59 aziende, secondo gruppo)
3. non è stato possibile correggere automaticamente mediante GEIS (36 aziende, terzo gruppo).

I record di cui ai punti 1 e 2 sono stati trattati automaticamente da GEIS in modo da avere informazioni aggiuntive sulla possibile localizzazione degli errori e sui possibili valori da attribuire in fase di controllo e correzione interattiva dei dati.

Per l'analisi di queste tipologie di dati sono state ricavate le seguenti informazioni:

- informazioni dalla tabella ORACLE contenente l'indicazione delle variabili da modificare o modificate da GEIS.
- elenco delle aziende che hanno subito la modifica di almeno una variabile del questionario in seguito al trattamento automatico, con l'indicazione, per ogni azienda dell'elenco, del valore originale e corretto per ogni variabile modificata.
- elenco dei vincoli del piano di controllo violati da ogni azienda di cui ai punti 1,2 e3.

Sulle 92 aziende di cui ai punti 1 e 2, GEIS ha operato come segue: 1) per 2 aziende del primo gruppo ed 1 azienda del secondo gruppo non è stata individuata una soluzione ammissibile; 2) 58 aziende sono state modificate; 3) le rimanenti 31 aziende (tutte appartenenti al primo gruppo) non sono state corrette in quanto esatte rispetto a tutte le regole del piano di controllo, giova ricordare che queste aziende sono state localizzate per il controllo interattivo in quanto sospette rispetto ai rapporti caratteristici fra totali di

sezione. Per queste aziende il controllo interattivo della presenza di un possibile errore di misura non può avvalersi di informazioni aggiuntive da GEIS.

Le 58 aziende modificate da GEIS corrispondono tutte ad aziende individuate per il controllo interattivo perché contenenti errori influenti.

6. Conclusioni

L'obiettivo naturale di una procedura di controllo e correzione dei dati deve essere quello di trovare gli errori non campionari e ripristinare per essi il dato "vero", relativamente a questa affermazione del tutto intuitiva il miglior piano di controllo e correzione risulta essere ovviamente il follow-up completo, è chiaro altresì che questa situazione appare del tutto utopica in quanto nella statistica ufficiale tempi e costi sono un altro importante elemento da considerare per la valutazione delle prestazioni di una procedura di editing. Gli elementi fondamentali da tenere congiuntamente sotto controllo in un tale processo diventano quindi: qualità, tempestività e costi. L'idea è quindi puntare ad un compromesso fra qualità dei risultati e tempestività, ovvero attuare una vera e propria analisi di costi/benefici la cui risposta però ovviamente non può essere univoca ma basata sull'effettiva disponibilità di risorse per l'indagine. E' proprio a questo fine che si ritiene necessaria l'adozione di una procedura integrata, le cui sottofasi fondamentali consistono proprio nell'individuazione dei record che necessariamente devono essere corretti con più attenzione (elevata qualità) e dalla correzione automatica dei rimanenti record (tempestività, coerenza). In questo lavoro l'accento non è stato posto quindi sui risultati veri e propri dell'applicazione ma sull'attuabilità di una tale procedura, questo perché sarà poi compito del responsabile dell'indagine decidere di volta in volta a quale caratteristica dare peso maggiore.

Congiuntamente ai vantaggi precedentemente introdotti, questa procedura presenta un'altra caratteristica, l'elevata produzione di report che possono essere utilizzati sia come documentazione che come fonte di informazioni da utilizzare per analizzare il processo attuato. Anche in questo caso nei paragrafi precedenti non ci siamo tanto soffermati sulla valutazione della bontà dei risultati dell'applicazione all'indagine in questione, quanto piuttosto ad illustrare tutte le possibili applicazioni e interpretazioni dei vari report a disposizione.

Un'altra importante caratteristica dell'approccio utilizzato in questa indagine è che esso è generalizzabile ad altre indagini di tipo economico, in termini sia di architettura

generale che di tecniche utilizzate. In particolare, lo strumento utilizzato per il controllo e la correzione automatica (GEIS) risulta già di sua natura generalizzato, mentre nel caso delle tecniche di tipo selettivo solo la filosofia è generalizzabile, mentre le caratteristiche delle funzioni utilizzate dipendono strettamente dalle caratteristiche dell'indagine e delle variabili oggetto del controllo. Questa caratteristica è completamente assente nei controlli di tipo deterministico, che evidentemente sono strettamente legati alle caratteristiche della singola indagine e vanno quindi approntati caso per caso, con un notevole ed inevitabile dispendio di risorse.

L'applicazione effettuata ha evidenziato alcune problematiche di tipo generale relative al trattamento di alcune tipologie di errore, ovvero dei casi in cui esiste un solo vincolo per variabile, oppure non sono violati più vincoli simultaneamente. In queste situazioni l'approccio di Fellegi-Holt non sembra quello migliore, per cui andrebbero effettuati ulteriori studi e sperimentazioni di approcci alternativi, quali ad esempio la tecnica della proration (prevista nella nuova versione di GEIS), oppure diversi algoritmi di individuazione dell'errore, come ad esempio quelli utilizzati nel software automatico generalizzato NIM (*New Imputation Methodology*, Bankier et al, 1994) in cui, sempre nel rispetto dei criteri del minimo cambiamento e di imputazione con donatori di minima distanza, le fasi di localizzazione e di correzione degli errori sono completamente "guidate" dai dati disponibili.

Bibliografia

- BANKIER M., FILION J.-M., LUC M., NADEAU C. (1994), "Imputing Numeric and Qualitative Variables Simultaneously", *Proceedings of the Section on Survey Research Methods*, American Statistical Association, pp. 242-247.
- BARCAROLI G., LUZI O. (1995), "GEIS - Sistema Generalizzato per l'Editing e l'Imputazione di Variabili Quantitative: una Sperimentazione", *Quaderni di Ricerca ISTAT*, n.1.
- COTTON C. (1991), "Functional description of the generalized edit and imputation system", Statistics Canada, Business Survey Methods Division, July 25.
- DELLA ROCCA G., LUZI O. (1997), "La procedura D-MASO per la localizzazione dei valori outlier", *Atti del XIII Convegno Utenti Italiani di SAS System (SUGITALIA97)*, Sorrento, 29-31.

- FELLEGI I.P., HOLT D. (1976), "A systematic approach to edit and imputation", *Journal of the American Statistical Association*, vol.71, pp.17-35.
- GRANQUIST L. (1992), "A review of methods for rationalizing the editing of survey data", *Statistical Data Editing Methods and Techniques*, vol. 1.
- GRANQUIST L. (1996), "The new view on editing", *International Statistical Review*, Vol. 65, No. 3, pp. 381-387.
- GRANQUIST L. (1997), "An overview of methods of evaluating data editing procedures", *Statistical Data Editing*, Vol. 2, Methods and Techniques, *Statistical Standards and Studies*, n. 48.,UN/ECE, pp.112-122.
- KOVAR J.G., MAC MILLIAN J.H., WHITRIDGE P. (1988), "Overview and strategy for the generalized edit and imputation system", *Statistics Canada, Methodology Branch*.
- LATOUCHE M., BERTHELOT J.M. (1992), "Use of Score Function to Prioritize and Limit Recontacts in Editing Business Surveys", *Journal of Official Statistics*, Vol.8, n. 3, Part II.
- LUZI O. (1998), "L'editing selettivo come strumento per la razionalizzazione del processo di editing: il caso", *Quaderni di Ricerca ISTAT*, n. 3.
- PALETZ D. (1993), "Documentation of Distance Measurement Algorithm for Selection of Outliers (D-MASO)", documento interno U.S. Bureau of the Census.
- WINKLER W.E. (1994), "SPEER Edit System", computer system and unpublished documentation, *Statistical Research Division, U.S. Bureau of the Census, Washington D.C., USA*.

Allegato 1

Questionario dell'indagine

3. STRUTTURA DEI COSTI (IVA esclusa)

| | Cod. | Importi in lire |
|--|------|-----------------|
| 3.1 Spese per acquisto di beni e servizi per le coltivazioni | | |
| a) Sementi e piantine | 01 |000 |
| b) Concimi | 02 |000 |
| c) Prodotti fitosanitari | 03 |000 |
| d) Altre spese | 04 |000 |
| <i>(specificare)</i> | | |
| 3.2 Spese per acquisto di beni e servizi per gli allevamenti | | |
| a) Alimenti per animali | 05 |000 |
| b) Uova da cova e pulcini | 06 |000 |
| c) Animali, esclusi quelli da riproduzione, latte, lana e lavoro (b) | 07 |000 |
| d) Spese veterinarie e prodotti farmaceutici | 08 |000 |
| e) Altre spese | 09 |000 |
| <i>(specificare)</i> | | |
| 3.3 Spese per la meccanizzazione | | |
| a) Carburanti | 10 |000 |
| b) Lubrificanti | 11 |000 |
| c) Manutenzione ordinaria, riparazioni e pezzi di ricambio | 12 |000 |
| d) Altre spese | 13 |000 |
| <i>(specificare)</i> | | |
| 3.4 Spese generali e varie | | |
| a) Acqua | 14 |000 |
| b) Elettricità | 15 |000 |
| c) Combustibili | 16 |000 |
| d) Assicurazioni | 17 |000 |
| e) Consulenze tecniche e contabili | 18 |000 |
| f) Spese per servizi bancari | 19 |000 |
| g) Quote associative | 20 |000 |
| h) Affitti e noli | 21 |000 |
| i) Trasformazione dei prodotti agricoli | 22 |000 |
| l) Trasporto, commerciali e immagazzinaggio | 23 |000 |
| m) Lavori agricoli eseguiti da terzi | 24 |000 |
| n) Materiali e piccoli attrezzi | 25 |000 |
| o) Manutenzioni e riparazioni ordinarie dei fabbricati agricoli non residenziali | 26 |000 |
| p) Altre spese | 27 |000 |
| <i>(specificare)</i> | | |
| 3.5 Interessi passivi | 28 |000 |
| 3.6 Imposte indirette | 29 |000 |
| TOTALE | 30 |000 |

4. GIACENZE E SCORTE (IVA esclusa)

| | Cod. | Inizio esercizio Importi in lire | Cod. | Fine esercizio Importi in lire |
|---|------|-------------------------------------|------|-----------------------------------|
| 4.1 Prodotti vegetali, compresi i trasformati e in corso di trasformazione (a) | 31 |000 | 36 |000 |
| 4.2 Prodotti zootecnici, compresi i trasformati e in corso di trasformazione (b) | 32 |000 | 37 |000 |
| 4.3 Animali, esclusi quelli da riproduzione, latte, lana e lavoro (b) | 33 |000 | 38 |000 |
| 4.4 Mezzi tecnici extraziendali | 34 |000 | 39 |000 |
| TOTALE | 35 |000 | 40 |000 |

(a) Dati compresi nella scheda ausiliaria n. 1

(b) Dati compresi nella scheda ausiliaria n. 2

9. OCCUPAZIONE E REDDITI

| 9.1 Famiglia del conduttore manodopera dei familiari e dei parenti | C o d i c e | Sesso | Totale giornate di lavoro prestate in azienda | Ore medie lavorate in azienda nella giornata | | | Classi di redditi extraziendali dei componenti la famiglia agricola (*) | | | | Contributi sociali per la sola attività agricola aziendale (**) |
|--|----------------------------|---|---|---|----------------------------|----------------------------|---|---------------------------|--------------------------|--------------------------------|--|
| | | | | < 4 | 4-8 | > 8 | lavoro indipen- dente | lavoro dipen- dente | pensione | capitale e altre entrate | |
| | | | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | |
| a) Conduttore | 01 | 1 <input type="checkbox"/> M 2 <input type="checkbox"/> F | □ □ □ □ | 1 <input type="checkbox"/> | 2 <input type="checkbox"/> | 3 <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |000 |
| b) Coniuge | 02 | 1 <input type="checkbox"/> M 2 <input type="checkbox"/> F | □ □ □ □ | 1 <input type="checkbox"/> | 2 <input type="checkbox"/> | 3 <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |000 |
| c) Altri familiari - che lavorano in azienda | 03 | 1 <input type="checkbox"/> M 2 <input type="checkbox"/> F | □ □ □ □ | 1 <input type="checkbox"/> | 2 <input type="checkbox"/> | 3 <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |000 |
| | 03 | 1 <input type="checkbox"/> M 2 <input type="checkbox"/> F | □ □ □ □ | 1 <input type="checkbox"/> | 2 <input type="checkbox"/> | 3 <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |000 |
| | 03 | 1 <input type="checkbox"/> M 2 <input type="checkbox"/> F | □ □ □ □ | 1 <input type="checkbox"/> | 2 <input type="checkbox"/> | 3 <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |000 |
| - che non lavorano in azienda | 04 | 1 <input type="checkbox"/> M 2 <input type="checkbox"/> F | □ □ □ □ | 1 <input type="checkbox"/> | 2 <input type="checkbox"/> | 3 <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |000 |
| | 04 | 1 <input type="checkbox"/> M 2 <input type="checkbox"/> F | □ □ □ □ | 1 <input type="checkbox"/> | 2 <input type="checkbox"/> | 3 <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |000 |
| | 04 | 1 <input type="checkbox"/> M 2 <input type="checkbox"/> F | □ □ □ □ | 1 <input type="checkbox"/> | 2 <input type="checkbox"/> | 3 <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |000 |
| d) Parenti del conduttore che lavorano in azienda | 05 | 1 <input type="checkbox"/> M 2 <input type="checkbox"/> F | □ □ □ □ | 1 <input type="checkbox"/> | 2 <input type="checkbox"/> | 3 <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |000 |
| | 05 | 1 <input type="checkbox"/> M 2 <input type="checkbox"/> F | □ □ □ □ | 1 <input type="checkbox"/> | 2 <input type="checkbox"/> | 3 <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |000 |
| | 05 | 1 <input type="checkbox"/> M 2 <input type="checkbox"/> F | □ □ □ □ | 1 <input type="checkbox"/> | 2 <input type="checkbox"/> | 3 <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |000 |

(*) Indicare le classi di reddito mediante la seguente codifica:
 Redditi fino a 5 milioni..... 1
 Redditi da 5,1 a 10 milioni 2
 Redditi da 10,1 a 20 milioni 3
 Redditi da 20,1 a 30 milioni 4
 Redditi superiori a 30 milioni..... 5

| 9.2 Altra manodopera aziendale | C o d i c e | Totale giornate di lavoro prestate in azienda | Ore medie lavorate in azienda nella giornata | | | Retribuzioni lorde (**) | Contributi sociali (**) | Accantonamenti per TFR (**) |
|--|----------------------------|--|---|----------------------------|----------------------------|-------------------------------|-------------------------------|-----------------------------------|
| | | | < 4 | 4-8 | > 8 | | | |
| | | | 1 | 2 | 3 | | | |
| a) Operai a tempo indeterminato, categorie speciali, impiegati, dirigenti di cui impiegati | 06 | □ □ □ □ □ | 1 <input type="checkbox"/> | 2 <input type="checkbox"/> | 3 <input type="checkbox"/> |000 |000 |000 |
| | 07 | □ □ □ □ □ | 1 <input type="checkbox"/> | 2 <input type="checkbox"/> | 3 <input type="checkbox"/> |000 |000 |000 |
| b) Operai a tempo determinato di cui extracomunitari | 08 | □ □ □ □ □ | 1 <input type="checkbox"/> | 2 <input type="checkbox"/> | 3 <input type="checkbox"/> |000 |000 |000 |
| | 09 | □ □ □ □ □ | 1 <input type="checkbox"/> | 2 <input type="checkbox"/> | 3 <input type="checkbox"/> |000 |000 |000 |
| c) Coloni impropri ed assimilati | 10 | □ □ □ □ □ | 1 <input type="checkbox"/> | 2 <input type="checkbox"/> | 3 <input type="checkbox"/> |000 |000 |000 |

(**) Importi in Lire

| 9.3 Manodopera extraziendale | C o d i c e | Totale giornate di lavoro prestate in azienda | Ore medie lavorate in azienda nella giornata | | |
|--|----------------------------|--|---|----------------------------|----------------------------|
| | | | < 4 | 4-8 | > 8 |
| | | | 1 | 2 | 3 |
| a) Manodopera fornita da altre imprese e/o aziende agricole | 11 | □ □ □ □ □ | 1 <input type="checkbox"/> | 2 <input type="checkbox"/> | 3 <input type="checkbox"/> |

| | |
|--------------------------------------|--|
| ANNOTAZIONI: | IL RILEVATORE CODICE □ □ □ □ FIRMA DATA □ □ □ □ 1998 <small style="display: block; text-align: center;">GIORNO MESE</small> |
|--------------------------------------|--|

Allegato 2

Vincoli del piano di compatibilità

Vincoli di quadratura

Sez. 3 *Struttura dei costi*

(COS_QUAD) cod. 01+cod. 02+...+cod. 29=cod. 30

Sez. 4 *Giacenze e scorte*

(GIA_QUAD1) cod. 31+cod. 32+ cod. 33+cod. 34=cod. 35

(GIA_QUAD2) cod. 36+cod. 37+ cod. 38+cod. 39=cod. 40

Sez. 5 *Acquisti e vendite di capitali fissi nell'esercizio*

(FIS_QUAD1) cod. 41+cod. 42+...+cod. 49=cod. 50

(FIS_QUAD2) cod. 51+cod. 52+...+cod. 56=cod. 57

Sez. 6 *Reimpieghi*

(REIM_QUAD) cod. 58+cod. 59=cod. 60

Sez. 7 *Ricavi*

(RIC_QUAD) cod. 61+cod. 62+...+cod. 71=cod. 72

Altri vincoli

Sez. 3.1 *Spese per acquisto di beni e servizio per le coltivazioni*

(ESIS3_1B) Se (cod. 1+cod. 2+cod. 3+cod. 4)>0 deve esistere:

e/o alla Sez. 4.1 il cod. 36;

e/o alla Sez. 4.4 cod. 39;

e/o alla Sez. 6 il cod. 58;

e/o alla Sez. 7.1 cod. 61 e/o 62 e/o 66 e/o 67.

Sez. 3.2 *Spese per acquisto di beni e servizio per gli allevamenti:*

(ESIS3_2B) Se esiste cod. 5 deve esistere:

alla Sez. 3.2 il cod. 6, e/o 7 e/o 8 e/o 9;

e/o alla Sez. 4 cod. 37 e/o cod. 39;

e/o alla Sez. 5 cod. 42 e/o cod. 52;

e/o alla Sez. 6 il cod. 59;

e/o alla Sez. 7.1 cod. 63 e/o 64 e/o 65 e/o 66 e/o 67 e/o 69.

(ESIS3_21B) Se esiste cod. 6 deve esistere:

alla Sez. 3.2 il cod. 5 e/o 8 e/o 9;

e/o alla Sez. 4.3 cod. 38;

e/o alla Sez. 6 il cod. 59;

e/o alla Sez. 7.1 cod. 63 e/o 64 e/o 65 e/o 66 e/o 67.

(ESIS3_22B) Se esiste cod. 7 deve esistere:

alla Sez. 3.2 il cod. 5 e/o 8 e/o 9;

e/o alla Sez. 3.4 cod. 22;

e/o alla Sez. 4 cod. 33 e/o 37 e/o 38 e/o 39;

e/o alla Sez. 6 il cod. 59;

e/o alla Sez. 7.1 cod. 63 e/o 64 e/o 65 e/o 66 e/o 67 e/o 69.

(ESIS3_23B) Se (cod. 7 + cod. 9)>0 deve esistere:

alla Sez. 3.2 il cod. 5 e/o 6 e/o 7;
e/o alla Sez. 4 cod. 33 e/o 37 e/o 38 e/o 39;
e/o alla Sez. 5 cod. 42 e/o 52;
e/o alla Sez. 6 il cod. 59;
e/o alla Sez. 7.1 cod. 63 e/o 64 e/o 65 e/o 66 e/o 67 e/o 69.

Sez. 3.3 Spese per la meccanizzazione

(ESIS3_3B) Se (cod. 10 + cod. 11 + cod. 12 + cod. 13)>0 deve esistere:
alla Sez. 3.1 il cod. 1 e/o 2 e/o 3 e/o 4;
e/o alla Sez. 3.2 il cod. 5 e/o 6 e/o 7 e/o 8 e/o 9;
e/o alla Sez. 3.3 il cod. 11 e/o 12 e/o 13;
e/o alla Sez. 3.4 il cod. 17 e/o 22;
e/o alla Sez. 3.6 il cod. 29;
e/o alla Sez. 4.4 cod. 39;
e/o alla Sez. 5 cod. 44 e/o 45 e/o 49 e/o 54 e/o 55 e/o 56;

Sez. 3.4 Spese generali e varie

(ESIS3_4) Se (cod. 14 + cod. 15)>0 deve esistere:
alla Sez. 3.1 il cod. 1 e/o 2 e/o 3 e/o 4;
e/o alla Sez. 3.2 il cod. 5 e/o 6 e/o 7 e/o 8 e/o 9;
e/o alla Sez. 3.4 il cod. 22;
e/o alla Sez. 5 cod. 42 e/o 43 e/o 52 e/o 53;
e/o alla Sez. 6 cod. 58 e/o 59;
e/o alla Sez. 7.1 cod. 61 e/o 62 e/o 63 e/o 64 e/o 65 e/o 67.

(ESIS3_41) Se esiste cod. 16 deve esistere:
alla Sez. 3.4 il cod. 21 e/o 25;
e/o alla Sez. 4 il cod. 39;
e/o alla Sez. 5 cod. 42 e/o 43 e/o 44 e/o 49 e/o 52 e/o 53 e/o 54 e/o 56;
e/o alla Sez. 6 cod. 58 e/o 59;
e/o alla Sez. 7.1 cod. 61 e/o 62 e/o 63 e/o 64 e/o 65 e/o 67 e/o 68 e/o 69.

(ESIS3_42) Se esiste cod. 17 deve esistere:
alla Sez. 3.1 il cod. 1 e/o 2 e/o 3 e/o 4;
e/o alla Sez. 3.2 il cod. 5 e/o 6 e/o 7 e/o 8 e/o 9;
e/o alla Sez. 3.3 il cod. 10 e/o 11 e/o 12 e/o 13;
e/o alla Sez. 5 cod. 41 e/o 42 e/o 43 e/o 44 e/o 45 e/o 49 e/o 51 e/o 52 e/o 53 e/o 54 e/o 56.

(ESIS3_43) Se esiste cod. 18 deve esistere:
alla Sez. 3.1 il cod. 1 e/o 2 e/o 3 e/o 4;
e/o alla Sez. 3.2 il cod. 5 e/o 6 e/o 7 e/o 8 e/o 9;
e/o alla Sez. 3.5 il cod. 28;
e/o alla Sez. 5 cod. 41 e/o 42 e/o 46 e/o 47 e/o 49.

(ESIS3_44) Se esiste cod. 19 deve esistere:
alla Sez. 3.1 il cod. 1 e/o 2 e/o 3 e/o 4;
e/o alla Sez. 3.2 il cod. 5 e/o 6 e/o 7 e/o 8 e/o 9;
e/o alla Sez. 3.5 il cod. 28;

e/o alla Sez. 5 cod. 50 e/o 57.

(ESIS3_45) Se esiste cod. 20 deve esistere:
alla Sez. 3.1 il cod. 1 e/o 2 e/o 3 e/o 4;
e/o alla Sez. 3.2 il cod. 5 e/o 6 e/o 7 e/o 8 e/o 9;
e/o alla Sez. 3.3 il cod. 10 e/o 11 e/o 12 e/o 13;
e/o alla Sez. 7.1 cod. 61 e/o 62 e/o 63 e/o 64 e/o 65 e/o 67 e/o 68 e/o 69.

(ESIS3_46) Se esiste cod. 21 deve esistere:
alla Sez. 3.1 il cod. 1 e/o 2 e/o 3 e/o 4;
e/o alla Sez. 3.2 il cod. 5 e/o 6 e/o 7 e/o 8 e/o 9;
e/o alla Sez. 3.4 il cod. 23;
e/o alla Sez. 7.1 cod. 61 e/o 62 e/o 63 e/o 64 e/o 65 e/o 69.

(ESIS3_47) Se esiste cod. 22 deve esistere:
alla Sez. 3.1 il cod. 1 e/o 2 e/o 3 e/o 4;
e/o alla Sez. 3.2 il cod. 5 e/o 6 e/o 7 e/o 8 e/o 9;
e/o alla Sez. 3.4 il cod. 14 e/o 15 e/o 16 e/o 23;
e/o alla Sez. 3.6 il cod. 29;
e/o alla Sez. 4 il cod. 31 e/o 32 e/o 36 e/o 37;
e/o alla Sez. 5 cod. 43 e/o 44 e/o 49 e/o 53 e/o 54 e/o 56;
e/o alla Sez. 6 cod. 58 e/o 59;
e/o alla Sez. 7.1 cod. 61 e/o 62 e/o 65 e/o 66 e/o 67.

(ESIS3_48) Se esiste cod. 23 deve esistere:
alla Sez. 4 il cod. 31 e/o 32 e/o 36 e/o 37;
e/o alla Sez. 5 cod. 42 e/o 43 e/o 44 e/o 45 e/o 49;
e/o alla Sez. 6 cod. 58 e/o 59;
e/o alla Sez. 7.1 cod. 61 e/o 62 e/o 63 e/o 64 e/o 65 e/o 67 e/o 68.

(ESIS3_49) Se esiste cod. 24 deve esistere:
alla Sez. 3.1 il cod. 1 e/o 2 e/o 3 e/o 4;
e/o alla Sez. 3.2 il cod. 5 e/o 6 e/o 7 e/o 8 e/o 9;
e/o alla Sez. 3.4 il cod. 26;
e/o alla Sez. 5 cod. 41 e/o 42 e/o 45 e/o 46 e/o 49 e/o 51 e/o 52 e/o 53;

(ESIS3_50) Se esiste cod. 25 deve esistere:
alla Sez. 3.1 il cod. 1 e/o 2 e/o 3 e/o 4;
e/o alla Sez. 3.2 il cod. 5 e/o 6 e/o 7 e/o 8 e/o 9;
e/o alla Sez. 3.3 il cod. 10 e/o 11 e/o 12 e/o 13;
e/o alla Sez. 3.4 il cod. 26 e/o 27;
e/o alla Sez. 5 cod. 50 e/o 57;
e/o alla Sez. 6 cod. 60;
e/o alla Sez. 7.1 cod. 72.

(ESIS3_51) Se esiste cod. 26 deve esistere:
alla Sez. 3.4 il cod. 22 e/o 25;
e/o alla Sez. 3.6 il cod. 29;
e/o alla Sez. 4 il cod. 35 e/o 40;
e/o alla Sez. 5.1 cod. 43 e/o 53;

e/o alla Sez. 7.1 cod. 61 e/o 62 e/o 63 e/o 64 e/o 65.

(ESIS3_52) Se esiste cod. 28 deve esistere:
alla Sez. 3.1 il cod. 1 e/o 2 e/o 3 e/o 4;
e/o alla Sez. 3.2 il cod. 5 e/o 6 e/o 7 e/o 8 e/o 9;
e/o alla Sez. 3.3 il cod. 10 e/o 11 e/o 12 e/o 13;
e/o alla Sez. 3.4 il cod. 22 e/o 23 e/o 24 e/o 26;
e/o alla Sez. 5 cod. 50.

(ESIS3_53) Se esiste cod. 29 deve esistere:
alla Sez. 3.3 il cod. 10 e/o 11 e/o 12 e/o 13;
e/o alla Sez. 3.4 il cod. 21 e/o 22 e/o 26;
e/o alla Sez. 5 cod. 42 e/o 43 e/o 44 e/o 45 e/o 47 e/o 52 e/o 53 e/o 54 e/o 55;
e/o alla Sez. 7.1 cod. 61 e/o 62 e/o 63 e/o 64 e/o 65 e/o 67.

Sez. 4 Giacenze e scorte

(ESIS4) Se esiste cod. 31 deve esistere:
alla Sez. 4.1 il cod. 36;
e/o alla Sez. 6.1 il cod. 58;
e/o alla Sez. 7.1 cod. 61 e/o 62 e/o 66 e/o 67.

(ESIS4_1) Se esiste cod. 32 deve esistere:
alla Sez. 4.2 il cod. 37;
e/o alla Sez. 6.1 il cod. 59;
e/o alla Sez. 7.1 cod. 64 e/o 65 e/o 66 e/o 67.

(ESIS4_2) Se esiste cod. 33 deve esistere:
alla Sez. 4.3 il cod. 38;
e/o alla Sez. 7.1 cod. 63.

(ESIS4_3) Se esiste cod. 34 deve esistere:
alla Sez. 3.1 il cod. 1 e/o 2 e/o 3 e/o 4;
e/o alla Sez. 3.2 il cod. 5 e/o 6 e/o 7 e/o 8 e/o 9;
e/o alla Sez. 3.3 il cod. 10 e/o 11 e/o 12 e/o 13;
e/o alla Sez. 4.4 il cod. 39.

(ESIS4_4) Se esiste cod. 36 deve esistere:
alla Sez. 3.1 il cod. 1 e/o 2 e/o 3 e/o 4;
e/o alla Sez. 3.3 il cod. 10 e/o 11 e/o 12 e/o 13;
e/o alla Sez. 3.4 il cod. 14 e/o 15 e/o 16 e/o 22 e/o 23;
alla Sez. 4.1 il cod. 31.

(ESIS4_5) Se esiste cod. 37 deve esistere:
alla Sez. 3.2 il cod. 5 e/o 6 e/o 7 e/o 8 e/o 9;
e/o alla Sez. 3.3 il cod. 10 e/o 11 e/o 12 e/o 13;
e/o alla Sez. 3.4 il cod. 14 e/o 15 e/o 16 e/o 22 e/o 23;
alla Sez. 4.2 il cod. 32.

(ESIS4_6) Se esiste cod. 38 deve esistere:

alla Sez. 3.2 il cod. 5 e/o 6 e/o 7 e/o 8 e/o 9;
e/o alla Sez. 3.4 il cod. 14 e/o 15 e/o 16 e/o 17 e/o 23;
alla Sez. 4.3 il cod. 33.

(ESIS4_7) Se esiste cod. 39 deve esistere:
alla Sez. 3.1 il cod. 1 e/o 2 e/o 3 e/o 4;
e/o alla Sez. 3.2 il cod. 5 e/o 6 e/o 7 e/o 8 e/o 9;
e/o alla Sez. 3.3 il cod. 10 e/o 11 e/o 12 e/o 13;
alla Sez. 4.4 il cod. 34.

Sez. 5 Acquisti e vendite di capitali fissi nell'esercizio

(ESIS5_1) Se esiste cod. 41 deve esistere:
alla Sez. 3.1 il cod. 1 e/o 2 e/o 3 e/o 4;
e/o alla Sez. 3.3 il cod. 10 e/o 11 e/o 12 e/o 13;
e/o alla Sez. 3.4 il cod. 14 e/o 17 e/o 18 e/o 19 e/o 24;
e/o alla Sez. 3.5 il cod. 28;
e/o alla Sez. 5 cod. 48;
e/o alla Sez. 6.1 cod. 58;
e/o alla Sez. 7.1 cod. 61 e/o 62.

(ESIS5_2) Se esiste cod. 42 deve esistere:
alla Sez. 3.2 il cod. 5 e/o 6 e/o 7 e/o 8 e/o 9;
e/o alla Sez. 3.3 il cod. 10 e/o 11 e/o 12 e/o 13;
e/o alla Sez. 3.4 il cod. 14 e/o 15 e/o 16 e/o 17 e/o 18 e/o 19 e/o 23 e/o 24;
e/o alla Sez. 3.5 il cod. 28;
e/o alla Sez. 5 cod. 43 e/o 48;
e/o alla Sez. 6 cod. 59;
e/o alla Sez. 7.1 cod. 64 e/o 65.

(ESIS5_3) Se (cod. 43+ cod. 44+ cod. 45)>0 deve esistere:
alla Sez. 3.1 il cod. 1 e/o 2 e/o 3 e/o 4;
e/o alla Sez. 3.2 il cod. 5 e/o 6 e/o 7 e/o 8 e/o 9;
e/o alla Sez. 3.3 il cod. 10 e/o 11 e/o 12;
e/o alla Sez. 3.4 il cod. 15 e/o 16 e/o 17 e/o 19 e/o 23;
e/o alla Sez. 3.5 il cod. 28;
e/o alla Sez. 7.1 cod. 72.

(ESIS5_4) Se esiste cod. 46 deve esistere:
alla Sez. 3.3 il cod. 10 e/o 11 e/o 12;
e/o alla Sez. 3.4 il cod. 18 e/o 19 e/o 24;
e/o alla Sez. 3.5 il cod. 28.

(ESIS5_5) Se esiste cod. 47 deve esistere:
alla Sez. 3.3 il cod. 10 e/o 11 e/o 12;
e/o alla Sez. 3.5 il cod. 28;
e/o alla Sez. 5 cod. 41 e/o 42 e/o 43 e/o 44 e/o 45 e/o 46;
e/o alla Sez. 6 cod. 59;
e/o alla Sez. 7.1 cod. 64 e/o 65.

(ESIS5_6) Se esiste cod. 48 deve esistere:

alla Sez. 3.4 il cod. 19;
e/o alla Sez. 3.5 il cod. 28;
e/o alla Sez. 5 cod. 41 e/o 42 e/o 43 e/o 44 e/o 45 e/o 49.

(ESIS5_7) Se esiste cod. 49 deve esistere:
alla Sez. 3.4 il cod. 19;
e/o alla Sez. 3.5 il cod. 28.

(ESIS5_8) Se esiste cod. 51 deve esistere:
alla Sez. 3.1 il cod. 1 e/o 2 e/o 3 e/o 4;
e/o alla Sez. 3.4 il cod. 18;
e/o alla Sez. 3.6 cod. 29.

(ESIS5_9) Se esiste cod. 52 deve esistere:
alla Sez. 3.2 il cod. 5 e/o 6 e/o 7 e/o 8 e/o 9;
e/o alla Sez. 3.4 il cod. 14 e/o 15 e/o 16;
e/o alla Sez. 3.6 cod. 29.

(ESIS5_10) Se esiste cod. 53 deve esistere:
alla Sez. 3.4 il cod. 14 e/o 15 e/o 16 e/o 22;
e/o alla Sez. 3.6 cod. 29;
e/o alla Sez. 5 il cod. 47;
e/o alla Sez. 7.1 cod. 67.

(ESIS5_11) Se (cod. 54+cod. 55)>0 deve esistere:
alla Sez. 3.3 il cod. 10 e/o 11 e/o 12 e/o 13;
e/o alla Sez. 3.4 cod. 17 e/o 23;
e/o alla Sez. 5 il cod. 47.

Sez. 6 Reimpieghi

(ESIS6_1) Se esiste cod. 58 deve esistere:
alla Sez. 3.1 il cod. 1 e/o 2 e/o 3 e/o 4;
e/o alla Sez. 3.3 il cod. 10 e/o 11 e/o 12 e/o 13;
e/o alla Sez. 3.4 il cod. 14 e/o 15 e/o 16 e/o 22 e/o 23 e/o 24 e/o 25 e/o 26 e/o 27.

(ESIS6_2) Se esiste cod. 59 deve esistere:
alla Sez. 3.2 il cod. 5 e/o 6 e/o 7 e/o 8 e/o 9;
e/o alla Sez. 3.3 il cod. 10 e/o 11 e/o 12 e/o 13;
e/o alla Sez. 3.4 il cod. 14 e/o 15 e/o 16 e/o 22 e/o 23 e/o 24 e/o 25 e/o 26 e/o 27.

Sez. 7 Ricavi

(ESIS7_1) Se (cod. 61+ cod. 62)>0 deve esistere:
alla Sez. 3.1 il cod. 1 e/o 2 e/o 3 e/o 4;
e/o alla Sez. 3.3 il cod. 10 e/o 11 e/o 12 e/o 13;
e/o alla Sez. 3.4 il cod. 14 e/o 15 e/o 16 e/o 17 e/o 18 e/o 19 e/o 20 e/o 21 e/o 22 e/o 23
e/o 24 e/o 25 e/o 26;
e/o alla Sez. 3.6 il cod. 29;
e/o alla Sez. 4.1 cod. 31;
e/o alla Sez. 4.4 cod. 34.

(ESIS7_2) Se (cod. 63+ cod. 64+cod 65)>0 deve esistere:
alla Sez. 3.2 il cod. 5 e/o 6 e/o 7 e/o 8 e/o 9;
e/o alla Sez. 3.3 il cod. 10 e/o 11 e/o 12 e/o 13;
e/o alla Sez. 3.4 il cod. 14 e/o 15 e/o 16 e/o 17 e/o 18 e/o 20 e/o 21 e/o 22 e/o 23 e/o 25
e/o 26;
e/o alla Sez. 3.6 il cod. 29;
e/o alla Sez. 4 (cod. 32+cod. 33+cod. 34)>0 e/o (cod. 37+cod. 38+cod. 39)>0.

(ESIS7_3) Se esiste cod. 66 deve esistere:
alla Sez. 3.1 il cod. 1 e/o 2 e/o 3 e/o 4;
e/o alla Sez. 3.2 il cod. 5 e/o 6 e/o 7 e/o 8 e/o 9;
e/o alla Sez. 3.3 il cod. 10 e/o 11 e/o 12 e/o 13;
e/o alla Sez. 3.4 il cod. 14 e/o 15 e/o 16 e/o 17 e/o 18 e/o 20 e/o 21 e/o 22 e/o 23 e/o 25
e/o 26.

(ESIS7_4) Se esiste cod. 67 deve esistere:
alla Sez. 3 il cod. 30;
e/o alla Sez. 4 il cod. 35.

(ESIS7_5) Se esiste cod. 68 deve esistere:
alla Sez. 3.3 il cod. 10 e/o 11 e/o 12 e/o 13;
e/o alla Sez. 3.4 il cod. 17 e/o 25.

(ESIS7_6) Se esiste cod. 69 deve esistere:
alla Sez. 3.1 il cod. 1 e/o 2 e/o 3 e/o 4;
e/o alla Sez. 3.2 il cod. 5 e/o 6 e/o 7 e/o 8 e/o 9;
e/o alla Sez. 3.3 il cod. 10 e/o 11 e/o 12 e/o 13;
e/o alla Sez. 3.4 il cod. 14 e/o 15.

(ESIS7_7) Se esiste cod. 70 deve esistere:
alla Sez. 3.1 il cod. 1 e/o 2 e/o 3 e/o 4;
e/o alla Sez. 3.2 il cod. 5 e/o 6 e/o 7 e/o 8 e/o 9;
e/o alla Sez. 3.3 il cod. 10 e/o 11 e/o 12 e/o 13.

Allegato 3

Summary statistics (Passo 1)

EDIT SUMMARY STATISTICS

TABLE 1-1

COUNTS OF RECORDS THAT PASSED, MISSED AND FAILED FOR EACH EDIT

| EDIT | RECORDS PASSED | RECORDS MISSED | RECORDS FAILED |
|----------|----------------|----------------|----------------|
| ESIS7_7 | 1598 | 0 | 0 |
| ESIS7_6 | 1598 | 0 | 0 |
| ESIS7_5 | 1598 | 0 | 0 |
| ESIS7_4 | 1598 | 0 | 0 |
| ESIS7_3 | 1596 | 0 | 2 |
| ESIS7_2 | 1598 | 0 | 0 |
| ESIS7_1 | 1594 | 0 | 4 |
| ESIS6_2 | 1597 | 0 | 1 |
| ESIS6_1 | 1597 | 0 | 1 |
| ESIS5_9 | 1598 | 0 | 0 |
| ESIS5_8 | 1598 | 0 | 0 |
| ESIS5_7 | 1591 | 0 | 7 |
| ESIS5_6 | 1593 | 0 | 5 |
| ESIS5_5 | 1596 | 0 | 2 |
| ESIS5_4 | 1597 | 0 | 1 |
| ESIS5_3 | 1597 | 0 | 1 |
| ESIS5_2 | 1598 | 0 | 0 |
| ESIS5_11 | 1598 | 0 | 0 |
| ESIS5_10 | 1598 | 0 | 0 |
| ESIS5_1 | 1598 | 0 | 0 |
| ESIS4_7 | 1598 | 0 | 0 |
| ESIS4_6 | 1598 | 0 | 0 |
| ESIS4_5 | 1598 | 0 | 0 |
| ESIS4_4 | 1598 | 0 | 0 |
| ESIS4_3 | 1598 | 0 | 0 |

| | | | |
|-----------|------|---|----|
| ESIS4_2 | 1591 | 0 | 7 |
| ESIS4_1 | 1598 | 0 | 0 |
| ESIS4 | 1597 | 0 | 1 |
| ESIS3_53 | 1596 | 0 | 2 |
| ESIS3_52 | 1597 | 0 | 1 |
| ESIS3_51 | 1594 | 0 | 4 |
| ESIS3_50 | 1598 | 0 | 0 |
| ESIS3_49 | 1585 | 0 | 13 |
| ESIS3_48 | 1595 | 0 | 3 |
| ESIS3_47 | 1598 | 0 | 0 |
| ESIS3_46 | 1594 | 0 | 4 |
| ESIS3_45 | 1595 | 0 | 3 |
| ESIS3_44 | 1598 | 0 | 0 |
| ESIS3_43 | 1593 | 0 | 5 |
| ESIS3_42 | 1596 | 0 | 2 |
| ESIS3_41 | 1572 | 0 | 26 |
| ESIS3_4 | 1595 | 0 | 3 |
| ESIS3_3B | 1590 | 0 | 8 |
| ESIS3_2B | 1591 | 0 | 7 |
| ESIS3_23B | 1596 | 0 | 2 |
| ESIS3_22B | 1598 | 0 | 0 |
| ESIS3_21B | 1598 | 0 | 0 |
| ESIS3_1B | 1556 | 0 | 42 |
| COS_QUAD | 1598 | 0 | 0 |
| FIS_QUAD1 | 1598 | 0 | 0 |
| FIS_QUAD2 | 1598 | 0 | 0 |
| GIA_QUAD1 | 1598 | 0 | 0 |
| GIA_QUAD2 | 1598 | 0 | 0 |
| REIM_QUAD | 1598 | 0 | 0 |
| RIC_QUAD | 1598 | 0 | 0 |

TABLE 1-2

DISTRIBUTION OF RECORDS THAT PASSED, MISSED AND FAILED K EDITS

| NUMBER OF EDITS (K) | RECORDS PASSED | RECORDS MISSED | RECORDS FAILED |
|------------------------|-------------------|-------------------|-------------------|
| 0 | 0 | 1598 | 1465 |
| 1 | 0 | 0 | 117 |
| 2 | 0 | 0 | 10 |
| 3 | 0 | 0 | 4 |
| 4 | 0 | 0 | 2 |
| 5 | 0 | 0 | 0 |
| TOTAL RECORDS | 1598 | 1598 | 1598 |

TABLE 1-3

OVERALL COUNTS OF RECORDS THAT PASSED, MISSED AND FAILED

| RECORDS PASSED | RECORDS MISSED | RECORDS FAILED | TOTAL |
|-------------------|-------------------|-------------------|-------|
| 1465 | 0 | 133 | 1598 |

Allegato 4

Imputazione con unica soluzione (deterministica) (Passo 1)

DETERMINISTIC IMPUTATION SOLUTIONS REPORT(S)

UDE 1

Record Key is: CODICE

| CODICE | FieldId | VALUE |
|--------------|----------|----------------------|
| 046003010281 | COS10 | Old :200 New : -0 |
| | COS23 | Old : 0 New :200 |
| 052032003054 | FISSI45 | Old : 0 New :150 |
| | FISSI49 | Old :150 New : -0 |
| 083006001231 | COS1 | Old : 0 New : 30 |
| | COS23 | Old : 30 New : -0 |
| 088006011255 | COS16 | Old :400 New : -0 |
| | COS22 | Old : 0 New :400 |
| 091062002214 | RICAVI61 | Old :450 New : -0 |
| | RICAVI71 | Old : 0 New :450 |

DETERMINISTIC IMPUTATION STATISTICS

Number of fields to impute.....: 50
 Number of fields imputed.....: 10
 Number of records to impute.....: 25
 Number of records imputed (*).....: 5

(*) Number of records for which at least one field was imputed with deterministic solutions.

UDE 2

DETERMINISTIC IMPUTATION STATISTICS

Number of fields to impute.....: 18
Number of fields imputed.....: 8
Number of records to impute.....: 9
Number of records imputed (*).....: 4

(*) Number of records for which at least one field was imputed with deterministic solutions.

UDE 3

DETERMINISTIC IMPUTATION STATISTICS

Number of fields to impute.....: 12
Number of fields imputed.....: 2
Number of records to impute.....: 6
Number of records imputed (*).....: 1

(*) Number of records for which at least one field was imputed with deterministic solutions.

UDE 4

DETERMINISTIC IMPUTATION STATISTICS

Number of fields to impute.....: 8
Number of fields imputed.....: 2
Number of records to impute.....: 4
Number of records imputed (*).....: 1

(*) Number of records for which at least one field was imputed with deterministic solutions.

UDE 5

DETERMINISTIC IMPUTATION STATISTICS

| | |
|--------------------------------------|----|
| Number of fields to impute.....: | 10 |
| Number of fields imputed.....: | 0 |
| Number of records to impute.....: | 5 |
| Number of records imputed (*):.....: | 0 |

(*) Number of records for which at least one field was imputed with deterministic solutions.

UDE 6

DETERMINISTIC IMPUTATION STATISTICS

| | |
|--------------------------------------|---|
| Number of fields to impute.....: | 6 |
| Number of fields imputed.....: | 0 |
| Number of records to impute.....: | 3 |
| Number of records imputed (*):.....: | 0 |

(*) Number of records for which at least one field was imputed with deterministic solutions.

UDE 7

DETERMINISTIC IMPUTATION STATISTICS

| | |
|--------------------------------------|----|
| Number of fields to impute.....: | 44 |
| Number of fields imputed.....: | 6 |
| Number of records to impute.....: | 22 |
| Number of records imputed (*):.....: | 3 |

(*) Number of records for which at least one field was imputed with deterministic solutions.

UDE 8

DETERMINISTIC IMPUTATION STATISTICS

Number of fields to impute.....: 60
Number of fields imputed.....: 18
Number of records to impute.....: 30
Number of records imputed (*).....: 9

(*) Number of records for which at least one field was imputed with deterministic solutions.

UDE 9

DETERMINISTIC IMPUTATION STATISTICS

Number of fields to impute.....: 29
Number of fields imputed.....: 8
Number of records to impute.....: 14
Number of records imputed (*).....: 4

(*) Number of records for which at least one field was imputed with deterministic solutions.

Allegato 5

Imputazione con donatore (Passo 1)

DONOR IMPUTATION SUMMARY STATISTICS

UDE 1

Donor Count: 105
 Donor Percent: 81.40
 Recipients with Matching Fields: 20
 with Donor: 15
 with no Donor: 5
 Recipients with no Matching Fields: 0
 with Donor: 0
 with no Donor: 0
 Imputation Criteria: 30
 Imputation Criteria Percent: 30.00
 Bucket Size: 16

DONOR IMPUTATION SUMMARY STATISTICS

TABLE 4-3

FREQUENCY OF
 NUMBER OF TIMES RECORDS WERE USED AS DONORS

| NUMBER OF TIMES USED AS A DONOR | ALL POTENTIAL DONORS | | | DONORS ACTUALLY USED | | |
|--|----------------------|---------|-----------------------|----------------------|---------|-----------------------|
| | NUMBER OF RECORDS | PERCENT | CUMULATIVE PERCENT | NUMBER OF RECORDS | PERCENT | CUMULATIVE PERCENT |
| NEVER USED | 93 | 88.57% | 88.57% | - | - | - |
| USED 1 TIME | 10 | 9.52% | 98.10% | 10 | 83.33% | 83.33% |
| USED 2 TIMES | 1 | 0.95% | 99.05% | 1 | 8.33% | 91.67% |
| USED 3 TIMES | 1 | 0.95% | 100.00% | 1 | 8.33% | 100.00% |
| TOTAL | 105 | | | 12 | | |

UDE 2

Donor Count: 94

Donor Percent: 94.95

Recipients with Matching Fields: 5
 with Donor: 5
 with no Donor: 0

Recipients with no Matching Fields: 0
 with Donor: 0
 with no Donor: 0

Imputation Criteria: 30
 Imputation Criteria Percent: 30.00

Bucket Size: 16

FREQUENCY OF NUMBER OF TIMES RECORDS WERE USED AS DONORS

| | ALL POTENTIAL DONORS | DONORS ACTUALLY USED |
|---------------------------------|----------------------|----------------------|
| NUMBER OF TIMES USED AS A DONOR | NUMBER OF RECORDS | PERCENT |
| | PERCENT | CUMULATIVE PERCENT |
| NEVER USED | 90 | 95.74% |
| USED 1 TIME | 3 | 3.19% |
| USED 2 TIMES | 1 | 1.06% |
| TOTAL | 94 | |

UDE 3

Donor Count: 57

Donor Percent: 90.48

Recipients with Matching Fields: 5
 with Donor: 5
 with no Donor: 0

Recipients with no Matching Fields: 0
 with Donor: 0
 with no Donor: 0

Imputation Criteria: 30
 Imputation Criteria Percent: 30.00

Bucket Size: 16

FREQUENCY OF
NUMBER OF TIMES RECORDS WERE USED AS DONORS

| NUMBER OF TIMES USED AS A DONOR | ALL POTENTIAL DONORS | | | DONORS ACTUALLY USED | | |
|---------------------------------|----------------------|---------|--------------------|----------------------|---------|--------------------|
| | NUMBER OF RECORDS | PERCENT | CUMULATIVE PERCENT | NUMBER OF RECORDS | PERCENT | CUMULATIVE PERCENT |
| NEVER USED | 52 | 91.23% | 91.23% | - | - | - |
| USED 1 TIME | 5 | 8.77% | 100.00% | 5 | 100.00% | 100.00% |
| TOTAL | 57 | | | 5 | | |

UDE 4

Donor Count: 56

Donor Percent: 93.33

Recipients with Matching Fields: 3
with Donor: 3
with no Donor: 0

Recipients with no Matching Fields: 0
with Donor: 0
with no Donor: 0

Imputation Criteria: 30
Imputation Criteria Percent: 30.00

Bucket Size: 16

FREQUENCY OF
NUMBER OF TIMES RECORDS WERE USED AS DONORS

| NUMBER OF TIMES USED AS A DONOR | ALL POTENTIAL DONORS | | | DONORS ACTUALLY USED | | |
|---------------------------------|----------------------|---------|--------------------|----------------------|---------|--------------------|
| | NUMBER OF RECORDS | PERCENT | CUMULATIVE PERCENT | NUMBER OF RECORDS | PERCENT | CUMULATIVE PERCENT |
| NEVER USED | 53 | 94.64% | 94.64% | - | - | - |
| USED 1 TIME | 3 | 5.36% | 100.00% | 3 | 100.00% | 100.00% |
| TOTAL | 56 | | | 3 | | |

UDE 5

Donor Count: 60

Donor Percent: 90.91

Recipients with Matching Fields: 5
 with Donor: 3
 with no Donor: 2

Recipients with no Matching Fields: 0
 with Donor: 0
 with no Donor: 0

Imputation Criteria: 30
 Imputation Criteria Percent: 30.00

Bucket Size: 16

FREQUENCY OF NUMBER OF TIMES RECORDS WERE USED AS DONORS

| NUMBER OF TIMES USED AS A DONOR | ALL POTENTIAL DONORS NUMBER OF RECORDS | PERCENT | CUMULATIVE PERCENT | DONORS ACTUALLY USED NUMBER OF RECORDS | PERCENT | CUMULATIVE PERCENT |
|--|--|---------|-----------------------|--|---------|-----------------------|
| NEVER USED | 57 | 95.00% | 95.00% | - | - | - |
| USED 1 TIME | 3 | 5.00% | 100.00% | 3 | 100.00% | 100.00% |
| TOTAL | 60 | | | 3 | | |

UDE 6

Donor Count: 38

Donor Percent: 90.48

Recipients with Matching Fields: 3
 with Donor: 2
 with no Donor: 1

Recipients with no Matching Fields: 0
 with Donor: 0
 with no Donor: 0

Imputation Criteria: 30
 Imputation Criteria Percent: 30.00

Bucket Size: 16

FREQUENCY OF
NUMBER OF TIMES RECORDS WERE USED AS DONORS

| NUMBER OF TIMES USED AS A DONOR | ALL POTENTIAL DONORS | | | DONORS ACTUALLY USED | | |
|--|----------------------|---------|-----------------------|----------------------|---------|-----------------------|
| | NUMBER OF RECORDS | PERCENT | CUMULATIVE PERCENT | NUMBER OF RECORDS | PERCENT | CUMULATIVE PERCENT |
| NEVER USED | 36 | 94.74% | 94.74% | - | - | - |
| USED 1 TIME | 2 | 5.26% | 100.00% | 2 | 100.00% | 100.00% |
| TOTAL | 38 | | | 2 | | |

UDE 7

Donor Count: 371

Donor Percent: 95.13

Recipients with Matching Fields: 19
with Donor: 17
with no Donor: 2

Recipients with no Matching Fields: 0
with Donor: 0
with no Donor: 0

Imputation Criteria: 30
Imputation Criteria Percent: 30.00

Bucket Size: 16

FREQUENCY OF
NUMBER OF TIMES RECORDS WERE USED AS DONORS

| NUMBER OF TIMES USED AS A DONOR | ALL POTENTIAL DONORS | | | DONORS ACTUALLY USED | | |
|--|----------------------|---------|-----------------------|----------------------|---------|-----------------------|
| | NUMBER OF RECORDS | PERCENT | CUMULATIVE PERCENT | NUMBER OF RECORDS | PERCENT | CUMULATIVE PERCENT |
| NEVER USED | 356 | 95.96% | 95.96% | - | - | - |
| USED 1 TIME | 13 | 3.50% | 99.46% | 13 | 86.67% | 86.67% |
| USED 2 TIMES | 2 | 0.54% | 100.00% | 2 | 13.33% | 100.00% |
| TOTAL | 371 | | | 15 | | |

UDE 8

Donor Count: 495
Donor Percent: 95.19
Recipients with Matching Fields: 21
 with Donor: 20
 with no Donor: 1
Recipients with no Matching Fields: 0
 with Donor: 0
 with no Donor: 0
Imputation Criteria: 30
Imputation Criteria Percent: 30.00
Bucket Size: 16

FREQUENCY OF NUMBER OF TIMES RECORDS WERE USED AS DONORS

| NUMBER OF TIMES USED AS A DONOR | ALL POTENTIAL DONORS | DONORS ACTUALLY USED | | | | |
|---------------------------------|----------------------|----------------------|--------------------|-------------------|---------|--------------------|
| | NUMBER OF RECORDS | PERCENT | CUMULATIVE PERCENT | NUMBER OF RECORDS | PERCENT | CUMULATIVE PERCENT |
| NEVER USED | 477 | 96.36% | 96.36% | - | - | - |
| USED 1 TIME | 16 | 3.23% | 99.60% | 16 | 88.89% | 88.89% |
| USED 2 TIMES | 2 | 0.40% | 100.00% | 2 | 11.11% | 100.00% |
| TOTAL | 495 | | | 18 | | |

UDE 9

Donor Count: 216
Donor Percent: 94.32
Recipients with Matching Fields: 10
 with Donor: 9
 with no Donor: 1
Recipients with no Matching Fields: 0
 with Donor: 0
 with no Donor: 0
Imputation Criteria: 30
Imputation Criteria Percent: 30.00
Bucket Size: 16

FREQUENCY OF

NUMBER OF TIMES RECORDS WERE USED AS DONORS

| NUMBER OF TIMES USED AS A DONOR | ALL POTENTIAL DONORS | | | DONORS ACTUALLY USED | | |
|--|----------------------|---------|-----------------------|----------------------|---------|-----------------------|
| | NUMBER OF RECORDS | PERCENT | CUMULATIVE PERCENT | NUMBER OF RECORDS | PERCENT | CUMULATIVE PERCENT |
| NEVER USED | 208 | 96.30% | 96.30% | - | - | - |
| USED 1 TIME | 7 | 3.24% | 99.54% | 7 | 87.50% | 87.50% |
| USED 2 TIMES | 1 | 0.46% | 100.00% | 1 | 12.50% | 100.00% |
| TOTAL | 216 | | | 8 | | |

Allegato 6

Summary statistics (Passo 2)

EDIT SUMMARY STATISTICS

TABLE 1-1

COUNTS OF RECORDS THAT PASSED, MISSED AND FAILED FOR EACH EDIT

| EDIT | RECORDS PASSED | RECORDS MISSED | RECORDS FAILED |
|----------|----------------|----------------|----------------|
| ESIS7_7 | 2037 | 0 | 0 |
| ESIS7_6 | 2037 | 0 | 0 |
| ESIS7_5 | 2037 | 0 | 0 |
| ESIS7_4 | 2037 | 0 | 0 |
| ESIS7_3 | 2036 | 0 | 1 |
| ESIS7_2 | 2037 | 0 | 0 |
| ESIS7_1 | 2035 | 0 | 2 |
| ESIS6_2 | 2036 | 0 | 1 |
| ESIS6_1 | 2036 | 0 | 1 |
| ESIS5_9 | 2037 | 0 | 0 |
| ESIS5_8 | 2037 | 0 | 0 |
| ESIS5_7 | 2035 | 0 | 2 |
| ESIS5_6 | 2035 | 0 | 2 |
| ESIS5_5 | 2037 | 0 | 0 |
| ESIS5_4 | 2037 | 0 | 0 |
| ESIS5_3 | 2037 | 0 | 0 |
| ESIS5_2 | 2037 | 0 | 0 |
| ESIS5_11 | 2037 | 0 | 0 |
| ESIS5_10 | 2037 | 0 | 0 |
| ESIS5_1 | 2037 | 0 | 0 |
| ESIS4_7 | 2037 | 0 | 0 |
| ESIS4_6 | 2037 | 0 | 0 |
| ESIS4_5 | 2037 | 0 | 0 |
| ESIS4_4 | 2037 | 0 | 0 |
| ESIS4_3 | 2036 | 0 | 1 |

| | | | |
|-----------|------|---|-----|
| ESIS4_2 | 2035 | 0 | 2 |
| ESIS4_1 | 2036 | 0 | 1 |
| ESIS4 | 2036 | 0 | 1 |
| ESIS3_53 | 2035 | 0 | 2 |
| ESIS3_52 | 2037 | 0 | 0 |
| ESIS3_51 | 2036 | 0 | 1 |
| ESIS3_50 | 2037 | 0 | 0 |
| ESIS3_49 | 2037 | 0 | 0 |
| ESIS3_48 | 2036 | 0 | 1 |
| ESIS3_47 | 2037 | 0 | 0 |
| ESIS3_46 | 2037 | 0 | 0 |
| ESIS3_45 | 2037 | 0 | 0 |
| ESIS3_44 | 2037 | 0 | 0 |
| ESIS3_43 | 2034 | 0 | 3 |
| ESIS3_42 | 2036 | 0 | 1 |
| ESIS3_41 | 2035 | 0 | 2 |
| ESIS3_4 | 2037 | 0 | 0 |
| ESIS3_3B | 2036 | 0 | 1 |
| ESIS3_2B | 2034 | 0 | 3 |
| ESIS3_23B | 2035 | 0 | 2 |
| ESIS3_22B | 2037 | 0 | 0 |
| ESIS3_21B | 2037 | 0 | 0 |
| ESIS3_1B | 2030 | 0 | 7 |
| COS_QUAD | 1751 | 0 | 286 |
| FIS_QUAD1 | 2028 | 0 | 9 |
| FIS_QUAD2 | 2032 | 0 | 5 |
| GIA_QUAD1 | 2014 | 0 | 23 |
| GIA_QUAD2 | 2012 | 0 | 25 |
| REIM_QUAD | 2020 | 0 | 17 |
| RIC_QUAD | 1970 | 0 | 67 |

TABLE 1-2

DISTRIBUTION OF RECORDS THAT PASSED, MISSED AND FAILED K EDITS

| NUMBER OF EDITS (K) | RECORDS PASSED | RECORDS MISSED | RECORDS FAILED |
|------------------------|-------------------|-------------------|-------------------|
| 0 | 0 | 2037 | 1639 |
| 1 | 0 | 0 | 340 |
| 2 | 0 | 0 | 46 |
| 3 | 0 | 0 | 11 |
| 4 | 0 | 0 | 1 |
| 5 | 0 | 0 | 0 |
| TOTAL RECORDS | 2037 | 2037 | 2037 |

TABLE 1-3

OVERALL COUNTS OF RECORDS THAT PASSED, MISSED AND FAILED

| RECORDS PASSED | RECORDS MISSED | RECORDS FAILED | TOTAL |
|-------------------|-------------------|-------------------|-------|
| 1639 | 0 | 398 | 2037 |

Allegato 7

Imputazione con unica soluzione (deterministica) (Passo 2)

DETERMINISTIC IMPUTATION

UDE 1

| CODICE | FieldId | VALUE |
|--------------|----------|----------------------------|
| 014019001097 | REIMP60 | Old : 860 New : 960 |
| 023045001125 | RICAVI61 | Old : 3735 New : 3585 |
| 028013002202 | COS5 | Old : 1201080 New : 200 |
| 063078001006 | COS11 | Old : 20 New : -0 |
| 065148001037 | REIMP60 | Old : 100 New : -0 |
| 069076002194 | COS14 | Old : 0 New : 350 |
| | COS5 | Old : 350 New : -0 |
| 072020036011 | COS22 | Old : 680 New : -0 |
| 073006999143 | COS30 | Old : 13650 New : 19165 |
| 073006999228 | COS10 | Old : 1000 New : 300 |
| 075090666001 | COS30 | Old : 310 New : 470 |
| 082069999148 | COS2 | Old : 1500 New : 1470 |
| 087013001195 | COS3 | Old : 1065 New : -0 |

DETERMINISTIC IMPUTATION STATISTICS

Number of fields to impute.....: 19
Number of fields imputed.....: 13
Number of records to impute.....: 13
Number of records imputed (*).....: 12

(*) Number of records for which at least one field was imputed with deterministic solutions.

UDE 2

DETERMINISTIC IMPUTATION STATISTICS

Number of fields to impute.....: 9
Number of fields imputed.....: 7
Number of records to impute.....: 8
Number of records imputed (*).....: 7

(*) Number of records for which at least one field was imputed with deterministic solutions.

UDE 3

DETERMINISTIC IMPUTATION STATISTICS

Number of fields to impute.....: 13
Number of fields imputed.....: 13
Number of records to impute.....: 13
Number of records imputed (*).....: 13

(*) Number of records for which at least one field was imputed with deterministic solutions.

UDE 4

DETERMINISTIC IMPUTATION STATISTICS

| | |
|-----------------------------------|---|
| Number of fields to impute.....: | 5 |
| Number of fields imputed.....: | 5 |
| Number of records to impute.....: | 5 |
| Number of records imputed (*): | 5 |

(*) Number of records for which at least one field was imputed with deterministic solutions.

UDE 5

DETERMINISTIC IMPUTATION STATISTICS

| | |
|-----------------------------------|---|
| Number of fields to impute.....: | 9 |
| Number of fields imputed.....: | 7 |
| Number of records to impute.....: | 7 |
| Number of records imputed (*): | 6 |

(*) Number of records for which at least one field was imputed with deterministic solutions.

UDE 6

DETERMINISTIC IMPUTATION STATISTICS

| | |
|-----------------------------------|---|
| Number of fields to impute.....: | 8 |
| Number of fields imputed.....: | 8 |
| Number of records to impute.....: | 8 |
| Number of records imputed (*): | 8 |

(*) Number of records for which at least one field was imputed with deterministic solutions.

UDE 7

DETERMINISTIC IMPUTATION STATISTICS

Number of fields to impute.....: 134
Number of fields imputed.....: 130
Number of records to impute.....: 116
Number of records imputed (*).....: 114

(*) Number of records for which at least one field was imputed with deterministic solutions.

UDE 8

DETERMINISTIC IMPUTATION STATISTICS

Number of fields to impute.....: 172
Number of fields imputed.....: 166
Number of records to impute.....: 146
Number of records imputed (*).....: 143

(*) Number of records for which at least one field was imputed with deterministic solutions.

UDE 9

DETERMINISTIC IMPUTATION STATISTICS

Number of fields to impute.....: 88
Number of fields imputed.....: 86
Number of records to impute.....: 73
Number of records imputed (*).....: 72

(*) Number of records for which at least one field was imputed with deterministic solutions.

Allegato 8

Imputazione con donatore (Passo 2)

DONOR IMPUTATION SUMMARY STATISTICS

UDE 1

Donor Count: 139
 Donor Percent: 97.20
 Recipients with Matching Fields: 3
 with Donor: 3
 with no Donor: 0
 Recipients with no Matching Fields: 0
 with Donor: 0
 with no Donor: 0
 Imputation Criteria: 5
 Imputation Criteria Percent: 30.00
 Bucket Size: 16

FREQUENCY OF
 NUMBER OF TIMES RECORDS WERE USED AS DONORS

| NUMBER OF TIMES USED AS A DONOR | ALL POTENTIAL DONORS | | | DONORS ACTUALLY USED | | |
|--|----------------------|---------|-----------------------|----------------------|---------|-----------------------|
| | NUMBER OF RECORDS | PERCENT | CUMULATIVE PERCENT | NUMBER OF RECORDS | PERCENT | CUMULATIVE PERCENT |
| NEVER USED | 136 | 97.84% | 97.84% | - | - | - |
| USED 1 TIME | 3 | 2.16% | 100.00% | 3 | 100.00% | 100.00% |
| TOTAL | 139 | | | 3 | | |

UDE 2

Donor Count: 111
 Donor Percent: 99.11
 Recipients with Matching Fields: 1
 with Donor: 1
 with no Donor: 0
 Recipients with no Matching Fields: 0
 with Donor: 0
 with no Donor: 0
 Imputation Criteria: 5
 Imputation Criteria Percent: 30.00
 Bucket Size: 16

FREQUENCY OF

NUMBER OF TIMES RECORDS WERE USED AS DONORS

| NUMBER OF TIMES USED AS A DONOR | ALL POTENTIAL DONORS | | | DONORS ACTUALLY USED | | |
|--|----------------------|---------|-----------------------|----------------------|---------|-----------------------|
| | NUMBER OF RECORDS | PERCENT | CUMULATIVE PERCENT | NUMBER OF RECORDS | PERCENT | CUMULATIVE PERCENT |
| NEVER USED | 110 | 99.10% | 99.10% | - | - | - |
| USED 1 TIME | 1 | 0.90% | 100.00% | 1 | 100.00% | 100.00% |
| TOTAL | 111 | | | 1 | | |

UDE 3

Donor Count: 72

Donor Percent: 98.63

Recipients with Matching Fields: 1
with Donor: 1
with no Donor: 0

Recipients with no Matching Fields: 0
with Donor: 0
with no Donor: 0

Imputation Criteria: 5
Imputation Criteria Percent: 30.00

Bucket Size: 16

FREQUENCY OF
NUMBER OF TIMES RECORDS WERE USED AS DONORS

| NUMBER OF TIMES USED AS A DONOR | ALL POTENTIAL DONORS | | | DONORS ACTUALLY USED | | |
|--|----------------------|---------|-----------------------|----------------------|---------|-----------------------|
| | NUMBER OF RECORDS | PERCENT | CUMULATIVE PERCENT | NUMBER OF RECORDS | PERCENT | CUMULATIVE PERCENT |
| NEVER USED | 71 | 98.61% | 98.61% | - | - | - |
| USED 1 TIME | 1 | 1.39% | 100.00% | 1 | 100.00% | 100.00% |
| TOTAL | 72 | | | 1 | | |

UDE 4

Donor Count: 521
Donor Percent: 99.62
Recipients with Matching Fields: 2
 with Donor: 2
 with no Donor: 0
Recipients with no Matching Fields: 0
 with Donor: 0
 with no Donor: 0
Imputation Criteria: 5
Imputation Criteria Percent: 30.00
Bucket Size: 16

FREQUENCY OF NUMBER OF TIMES RECORDS WERE USED AS DONORS

| NUMBER OF TIMES USED AS A DONOR | ALL POTENTIAL DONORS | PERCENT | CUMULATIVE PERCENT | DONORS ACTUALLY USED | PERCENT | CUMULATIVE PERCENT |
|---------------------------------|----------------------|---------|--------------------|----------------------|---------|--------------------|
| NEVER USED | 519 | 99.62% | 99.62% | - | - | - |
| USED 1 TIME | 2 | 0.38% | 100.00% | 2 | 100.00% | 100.00% |
| TOTAL | 521 | | | 2 | | |

UDE 5

FREQUENCY OF NUMBER OF TIMES RECORDS WERE USED AS DONORS

| NUMBER OF TIMES USED AS A DONOR | ALL POTENTIAL DONORS | PERCENT | CUMULATIVE PERCENT | DONORS ACTUALLY USED | PERCENT | CUMULATIVE PERCENT |
|---------------------------------|----------------------|---------|--------------------|----------------------|---------|--------------------|
| NEVER USED | 671 | 99.55% | 99.55% | - | - | - |
| USED 1 TIME | 3 | 0.45% | 100.00% | 3 | 100.00% | 100.00% |
| TOTAL | 674 | | | 3 | | |

UDE 6

Donor Count: 308

Donor Percent: 98.72

Recipients with Matching Fields: 1
 with Donor: 1
 with no Donor: 0

Recipients with no Matching Fields: 0
 with Donor: 0
 with no Donor: 0

Imputation Criteria: 5
 Imputation Criteria Percent: 30.00

Bucket Size: 16

FREQUENCY OF
 NUMBER OF TIMES RECORDS WERE USED AS DONORS

| NUMBER OF TIMES USED AS A DONOR | | ALL POTENTIAL DONORS | | | DONORS ACTUALLY USED | | |
|---------------------------------|--|----------------------|---------|--------------------|----------------------|---------|--------------------|
| | | NUMBER OF RECORDS | PERCENT | CUMULATIVE PERCENT | NUMBER OF RECORDS | PERCENT | CUMULATIVE PERCENT |
| NEVER USED | | 307 | 99.68% | 99.68% | - | - | - |
| USED 1 TIME | | 1 | 0.32% | 100.00% | 1 | 100.00% | 100.00% |
| TOTAL | | 308 | | | 1 | | |

Allegato 9

**Statistiche univariate sulle distribuzioni dei dati grezzi e
corretti da GEIS**

Dati corretti da GEIS - Statistiche univariate

| OBS | UDE | MCOS | MGIA1 | MGIA2 | MFIS1 | MFIS2 | MREIM |
|-----|-----|-----------|-----------|-----------|----------|----------|----------|
| OBS | UDE | MCOS | MGIA1 | MGIA2 | MFIS1 | MFIS2 | MREIM |
| 1 | 1 | 2340.55 | 360.62 | 437.31 | 1072.17 | 0.00 | 186.47 |
| 2 | 2 | 4543.84 | 1190.88 | 1276.08 | 1313.11 | 311.32 | 673.31 |
| 3 | 3 | 5736.40 | 2813.41 | 2736.15 | 1814.11 | 180.15 | 1079.79 |
| 4 | 4 | 7919.38 | 1545.05 | 1742.42 | 1587.48 | 150.00 | 1324.58 |
| 5 | 5 | 28108.01 | 31117.91 | 40163.64 | 22732.68 | 9271.96 | 1813.97 |
| 6 | 6 | 19566.98 | 7779.96 | 6509.51 | 1611.44 | 851.11 | 3982.73 |
| 7 | 7 | 38347.71 | 17067.89 | 16725.00 | 8112.93 | 1526.65 | 5818.54 |
| 8 | 8 | 95423.77 | 47138.19 | 35396.74 | 20323.65 | 3175.15 | 15764.51 |
| 9 | 9 | 360192.19 | 171477.08 | 183593.86 | 64938.16 | 11918.79 | 23592.11 |

| OBS | MRIC | STDCOS | STDGIA1 | STDGIA2 | STDFIS1 | STDFIS2 | STDREIM | STDRIC | MAXCOS |
|-----|-----------|-----------|-----------|-----------|-----------|----------|----------|------------|---------|
| 1 | 4805.89 | 3417.02 | 1094.41 | 1564.30 | 7462.62 | 0.00 | 727.95 | 10006.00 | 21600 |
| 2 | 11205.43 | 8077.41 | 4597.57 | 4679.32 | 6967.69 | 2153.02 | 2899.41 | 33052.22 | 73200 |
| 3 | 10359.47 | 5377.06 | 7061.15 | 6085.44 | 7291.46 | 857.12 | 4312.26 | 11014.88 | 33980 |
| 4 | 13853.60 | 8821.29 | 3750.49 | 3636.31 | 5093.82 | 1038.66 | 3892.31 | 18754.91 | 58468 |
| 5 | 61691.39 | 129976.98 | 202274.13 | 268735.50 | 151857.32 | 68158.15 | 4342.06 | 316410.02 | 1084965 |
| 6 | 30991.16 | 25052.64 | 22530.57 | 19613.82 | 3077.93 | 4094.04 | 8401.69 | 43020.73 | 149593 |
| 7 | 72932.27 | 49232.48 | 53875.32 | 53456.72 | 20224.48 | 7003.70 | 14590.57 | 84242.39 | 660257 |
| 8 | 180809.70 | 143393.87 | 371880.32 | 115219.93 | 55812.53 | 14924.96 | 73256.26 | 358839.47 | 1721517 |
| 9 | 647209.45 | 614178.17 | 464480.04 | 520625.67 | 167203.62 | 59104.98 | 68124.97 | 1035233.89 | 4798400 |

| OBS | MAXGIA1 | MAXGIA2 | MAXFIS1 | MAXFIS2 | MAXREIM | MAXRIC | MEDCOS | MEDGIA1 | MEDGIA2 | MEDFIS1 |
|-----|---------|---------|---------|---------|---------|---------|----------|---------|---------|---------|
| 1 | 10000 | 14000 | 70000 | 0 | 6000 | 73271 | 1000.0 | 0 | 0 | 0 |
| 2 | 41000 | 39500 | 53000 | 18000 | 27000 | 330000 | 2645.0 | 0 | 0 | 0 |
| 3 | 39265 | 30400 | 50000 | 5651 | 35000 | 55200 | 3900.0 | 0 | 0 | 0 |
| 4 | 19000 | 16200 | 30000 | 8000 | 20100 | 133000 | 6157.5 | 0 | 0 | 0 |
| 5 | 1677443 | 2231199 | 1258266 | 565627 | 18575 | 2642997 | 9405.0 | 0 | 0 | 0 |
| 6 | 132000 | 124000 | 12000 | 25000 | 37000 | 270000 | 11010.0 | 0 | 0 | 0 |
| 7 | 706534 | 740388 | 150200 | 84000 | 177315 | 851000 | 26050.0 | 0 | 0 | 0 |
| 8 | 9194843 | 1890000 | 685650 | 198363 | 1700000 | 7692050 | 52100.0 | 280 | 270 | 750 |
| 9 | 4355671 | 5573909 | 1203300 | 844000 | 510769 | 7600000 | 148200.0 | 0 | 0 | 5000 |

| OBS | MEDFIS2 | MEDREIM | MEDRIC | MINCOS | MINGIA1 | MINGIA2 | MINFIS1 | MINFIS2 | MINREIM | MINRIC |
|-----|---------|---------|--------|--------|---------|---------|---------|---------|---------|--------|
| 1 | 0 | 0 | 2170 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 2 | 0 | 0 | 4755 | 10 | 0 | 0 | 0 | 0 | 0 | 0 |
| 3 | 0 | 0 | 7000 | 350 | 0 | 0 | 0 | 0 | 0 | 0 |
| 4 | 0 | 0 | 9750 | 220 | 0 | 0 | 0 | 0 | 0 | 600 |
| 5 | 0 | 0 | 17385 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 6 | 0 | 0 | 20137 | 690 | 0 | 0 | 0 | 0 | 0 | 0 |
| 7 | 0 | 0 | 53800 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 8 | 0 | 0 | 109000 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 9 | 0 | 0 | 291000 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

Dove MCOS = Media variabile Costi
STDCOS = Deviazione Standard della variabile Costi
MAXCOS = Valore massimo assunto dalla variabile Costi
MEDCOS = Mediana della variabile Costi
MINCOS = Valore minimo assunto dalla variabile Costi
Ecc...

Correlation Analysis (*)

----- OTE=3 -----

N = 645

Pearson Correlation Coefficients

| | COS30 | GIA35 | GIA40 | FISSI50 | FISSI57 | REIMP60 | RICAVI72 |
|----------|---------|---------|---------|---------|----------|----------|----------|
| COS30 | 1.00000 | 0.51574 | 0.64154 | 0.71034 | 0.35956 | 0.11730 | 0.72992 |
| GIA35 | 0.51574 | 1.00000 | 0.73821 | 0.55566 | 0.20977 | 0.18535 | 0.78988 |
| GIA40 | 0.64154 | 0.73821 | 1.00000 | 0.62344 | 0.36535 | 0.26560 | 0.63905 |
| FISSI50 | 0.71034 | 0.55566 | 0.62344 | 1.00000 | 0.61392 | 0.07111 | 0.63271 |
| FISSI57 | 0.35956 | 0.20977 | 0.36535 | 0.61392 | 1.00000 | -0.00404 | 0.29623 |
| REIMP60 | 0.11730 | 0.18535 | 0.26560 | 0.07111 | -0.00404 | 1.00000 | 0.12883 |
| RICAVI72 | 0.72992 | 0.78988 | 0.63905 | 0.63271 | 0.29623 | 0.12883 | 1.00000 |

(*) Per la stratificazione rispetto ad OTE viene presa la sola OTE = 3 a scopo esemplificativo, i risultati per le altre OTE sono simili.

Dati originali - Statistiche univariate

| OBS | UDE | MCOS | MGIA1 | MGIA2 | MFIS1 | MFIS2 | MREIM | MRIC | STDCOS |
|-----|-----|-----------|-----------|-----------|----------|----------|----------|-----------|-----------|
| 1 | 1 | 2337.84 | 360.62 | 437.31 | 1072.17 | 0.00 | 178.42 | 4805.89 | 3418.70 |
| 2 | 2 | 4538.14 | 1190.88 | 1276.08 | 1313.11 | 311.32 | 673.31 | 11205.43 | 8080.36 |
| 3 | 3 | 5719.14 | 2813.41 | 2736.15 | 1814.11 | 180.15 | 1067.47 | 10357.41 | 5378.46 |
| 4 | 4 | 7919.38 | 1545.05 | 1742.42 | 1587.48 | 150.00 | 1324.58 | 13853.60 | 8821.29 |
| 5 | 5 | 28108.01 | 31117.91 | 40163.64 | 22680.51 | 9271.96 | 1813.97 | 61694.29 | 129976.98 |
| 6 | 6 | 19566.98 | 7779.96 | 6509.51 | 1611.44 | 851.11 | 3982.73 | 30991.16 | 25052.64 |
| 7 | 7 | 38230.00 | 16949.65 | 16710.57 | 8092.89 | 1525.75 | 5806.17 | 72934.27 | 49268.53 |
| 8 | 8 | 95137.93 | 47130.53 | 35395.94 | 20323.65 | 2984.65 | 15755.49 | 180618.46 | 143491.98 |
| 9 | 9 | 359323.85 | 167409.97 | 179597.73 | 64904.26 | 12000.15 | 23596.38 | 645132.01 | 614452.15 |

| OBS | STDGIA1 | STDGIA2 | STDFIS1 | STDFIS2 | STDREIM | STDRIC | MAXCOS | MAXGIA1 | MAXGIA2 |
|-----|-----------|-----------|-----------|----------|----------|------------|---------|---------|---------|
| 1 | 1094.41 | 1564.30 | 7462.62 | 0.00 | 725.38 | 10006.00 | 21600 | 10000 | 14000 |
| 2 | 4597.57 | 4679.32 | 6967.69 | 2153.02 | 2899.41 | 33052.22 | 73200 | 41000 | 39500 |
| 3 | 7061.15 | 6085.44 | 7291.46 | 857.12 | 4314.07 | 11014.84 | 33980 | 39265 | 30400 |
| 4 | 3750.49 | 3636.31 | 5093.82 | 1038.66 | 3892.31 | 18754.91 | 58468 | 19000 | 16200 |
| 5 | 202274.13 | 268735.50 | 151864.47 | 68158.15 | 4342.06 | 316409.46 | 1084965 | 1677443 | 2231199 |
| 6 | 22530.57 | 19613.82 | 3077.93 | 4094.04 | 8401.69 | 43020.73 | 149593 | 132000 | 124000 |
| 7 | 53845.64 | 53459.42 | 20227.56 | 7003.21 | 14584.43 | 84243.75 | 660257 | 706534 | 740388 |
| 8 | 371880.18 | 115219.63 | 55812.53 | 14205.85 | 73280.62 | 358895.34 | 1721517 | 9194843 | 1890000 |
| 9 | 460671.86 | 517641.34 | 167195.46 | 59137.44 | 68131.13 | 1036186.41 | 4798400 | 4355671 | 5573909 |

| OBS | MAXFIS1 | MAXFIS2 | MAXREIM | MAXRIC | MEDCOS | MEDGIA1 | MEDGIA2 | MEDFIS1 | MEDFIS2 | MEDREIM |
|-----|---------|---------|---------|---------|----------|---------|---------|---------|---------|---------|
| 1 | 70000 | 0 | 6000 | 73271 | 1000.0 | 0 | 0 | 0 | 0 | 0 |
| 2 | 53000 | 18000 | 27000 | 330000 | 2645.0 | 0 | 0 | 0 | 0 | 0 |
| 3 | 50000 | 5651 | 35000 | 55200 | 3900.0 | 0 | 0 | 0 | 0 | 0 |
| 4 | 30000 | 8000 | 20100 | 133000 | 6157.5 | 0 | 0 | 0 | 0 | 0 |
| 5 | 1258266 | 565627 | 18575 | 2642997 | 9405.0 | 0 | 0 | 0 | 0 | 0 |
| 6 | 12000 | 25000 | 37000 | 270000 | 11010.0 | 0 | 0 | 0 | 0 | 0 |
| 7 | 150200 | 84000 | 177315 | 851000 | 26042.0 | 0 | 0 | 0 | 0 | 0 |
| 8 | 685650 | 198363 | 1700000 | 7692050 | 51940.0 | 280 | 270 | 750 | 0 | 0 |
| 9 | 1203300 | 844000 | 510769 | 7600000 | 148200.0 | 0 | 0 | 5000 | 0 | 0 |

| OBS | MEDRIC | MINCOS | MINGIA1 | MINGIA2 | MINFIS1 | MINFIS2 | MINREIM | MINRIC |
|-----|--------|--------|---------|---------|---------|---------|---------|--------|
| 1 | 2170 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 2 | 4755 | 10 | 0 | 0 | 0 | 0 | 0 | 0 |
| 3 | 7000 | 350 | 0 | 0 | 0 | 0 | 0 | 0 |
| 4 | 9750 | 220 | 0 | 0 | 0 | 0 | 0 | 600 |
| 5 | 17385 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 6 | 20137 | 690 | 0 | 0 | 0 | 0 | 0 | 0 |
| 7 | 53800 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 8 | 108671 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 9 | 289936 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

Correlation Analysis (*)

Correlation Analysis (*)
Correlation Analysis (*)

----- OTE=3 -----

N = 645

Pearson Correlation Coefficients

| | COS30 | GIA35 | GIA40 | FISSI50 | FISSI57 | REIMP60 | RICAVI72 |
|----------|---------|---------|---------|---------|----------|----------|----------|
| COS30 | 1.00000 | 0.51580 | 0.64162 | 0.71080 | 0.35972 | 0.11746 | 0.72967 |
| GIA35 | 0.51580 | 1.00000 | 0.73821 | 0.55567 | 0.20976 | 0.18536 | 0.78935 |
| GIA40 | 0.64162 | 0.73821 | 1.00000 | 0.62344 | 0.36533 | 0.26562 | 0.63824 |
| FISSI50 | 0.71080 | 0.55567 | 0.62344 | 1.00000 | 0.61394 | 0.07114 | 0.63292 |
| FISSI57 | 0.35972 | 0.20976 | 0.36533 | 0.61394 | 1.00000 | -0.00399 | 0.29619 |
| REIMP60 | 0.11746 | 0.18536 | 0.26562 | 0.07114 | -0.00399 | 1.00000 | 0.12895 |
| RICAVI72 | 0.72967 | 0.78935 | 0.63824 | 0.63292 | 0.29619 | 0.12895 | 1.00000 |

(*) Per la stratificazione rispetto ad OTE viene presa la sola OTE = 3 a scopo esemplificativo, i risultati per le altre OTE sono simili.