

# istat working papers

N. 9  
2011

## **Procedura di limitazione del rischio di violazione della riservatezza per tabelle collegate gerarchiche**

*Alessandra Capobianchi*



# istat working papers

N. 9  
2011

## **Procedura di limitazione del rischio di violazione della riservatezza per tabelle collegate gerarchiche**

*Alessandra Capobianchi*

## **Comitato di redazione**

*Coordinatore:* Giulio Barcaroli

*Componenti:*

Rossana Balestrino	Francesca Di Palma	Luisa Picozzi
Marco Ballin	Alessandra Ferrara	Mauro Politi
Riccardo Carbini	Angela Ferruzza	Alessandra Righi
Claudio Ceccarelli	Danila Filipponi	Luca Salvati
Giuliana Coccia	Cristina Freguja	Giovanni Seri
Fabio Crescenzi	Aurea Micali	Leonello Tronti
Carla De Angelis	Nadia Mignolli	Sonia Vittozzi

*Segreteria:*

Lorella Appolloni, Maria Silvia Cardacino, Laura Peci, Gilda Sonetti, Antonio Trobia

## **Istat Working Papers**

**Procedura di limitazione del rischio di violazione della riservatezza per tabelle collegate gerarchiche**

N. 9/2011

ISBN 88-458-1692-3

Istituto nazionale di statistica  
Servizio Editoria  
Via Cesare Balbo, 16 – Roma

# Procedura di limitazione del rischio di violazione della riservatezza per tabelle collegate gerarchiche

Alessandra Capobianchi

## Sommario

*Nelle indagini in ambito economico per limitare il rischio di violazione della riservatezza dei rispondenti nella diffusione di tabelle spesse si procede all'applicazione di tecniche di protezione che permettono di mantenere il rischio di violazione della riservatezza all'interno di limiti ragionevoli. Una delle tecniche di protezione più diffusa è quella della "Soppressione" ovvero l'eliminazione di informazioni relative a celle opportunamente selezionate. Poiché la maggior parte delle informazioni rilasciate dagli Istituti Nazionali di Statistica è costituita da insiemi di tabelle che presentano alcune celle comuni (tabelle collegate) la loro protezione rappresenta un problema complesso e di non facile soluzione. L'obiettivo di questo lavoro è duplice. Da un lato si intende presentare l'elaborazione di uno schema concettuale per la protezione di un insieme di tabelle collegate attraverso l'analisi delle relazioni che intercorrono tra queste. Dall'altro il lavoro intende essere una guida all'effettivo utilizzo del software Tau-Argus (<http://neon.vb.cbs.nl/casc/..%5Ccasc%5Ctau.htm>.) attraverso la descrizione delle diverse fasi necessarie per ottenere la protezione delle tabelle collegate. Per descrivere operativamente l'intero schema procedurale in questo lavoro viene esposta un'applicazione al piano di pubblicazione delle principali tavole relative alle Statistiche Strutturali sulle Imprese.*

**Parole chiave:** tabelle collegate, Tau Argus, Statistiche Strutturali sulle Imprese.

## Abstract

*In order to limit the risk of breaching the confidentiality of respondents from released aggregates, statistical agencies applying protection methods that allow to keep the risk of a breach of confidentiality within reasonable limits. One of the most widely used is the "Cell Suppression" method that appropriately hide some values in published tables. Since from a single survey sets of tables that have common cells and totals (linked tables) have to be published, their protection is a complex problem with a not easy solution. The aim of this paper is twofold. On the one hand we intend to present the development of a conceptual framework to protect a set of linked tables through the analysis of the relationships between them. On the other hand the work is meant to be as a guide to the use of software Tau-Argus in complex situation (<http://neon.vb.cbs.nl/casc/..%5Ccasc%5Ctau.htm>.) through the description of various steps necessary to protect linked tables. In this paper, to describe the entire procedural scheme, an application to the main table of Structural Business Statistics is presented.*

**Keywords:** Linked table, Tau-Argus, Structural business statistics.

## 1. Introduzione

Gli Istituti Nazionali di Statistica, nel predisporre piani di pubblicazione di dati statistici, devono provvedere ad una sempre maggiore richiesta di informazione da parte di utenti esterni. La necessità di studiare fenomeni complessi comporta una crescente richiesta di informazione, basata sulla stessa rilevazione statistica, non solo sempre più dettagliata ma anche molto articolata. Ciò comporta la pubblicazione di un crescente numero di tabelle tra loro collegate ovvero tabelle

che presentano dati relativi alla stessa variabile risposta e condividono almeno alcune categorie della stessa variabile classificatrice,<sup>1</sup> che presentano strutture gerarchiche complesse e in alcuni casi sovrapponibili.

Alcune celle di tali tabelle possono contenere informazioni relative a pochi se non a singoli rispondenti che possono così risultare identificabili. In questi casi gli Istituti di Statistica sopprimono tali informazioni in modo tale da garantire la protezione della riservatezza dei rispondenti (Hundepool *et al.*, 2009). La scelta dell'insieme delle celle da sopprimere deve risultare coordinata in maniera tale da tener conto di tutti i legami che intercorrono tra le tabelle (tabelle collegate) (de Wolf, Giessing, 2008) (Jansson, Bernstrom, Calson 2010) (Giessing, Schmidt 2010) (de Wolf 2007) (de Wolf, Hundepool 2010) (Capobianchi, Franconi 2009) definite nel piano di diffusione e tale da minimizzare la perdita di informazione.

L'obiettivo del presente lavoro è quello di definire uno schema procedurale per la protezione di un insieme di tabelle collegate attraverso l'utilizzo del software Argus (Hundepool *ed al.*). Tale schema si basa sull'analisi dei legami che intercorrono tra gli aggregati specificati in un determinato piano di diffusione per poi passare alla definizione di un insieme di tabelle analizzabili con il software Argus. Poiché tale software non tratta in maniera automatica tabelle collegate, è stato necessario definire una procedura semi-automatica che permetta di tener conto dei legami tra le tabelle.

Per descrivere operativamente l'intero schema procedurale in questo lavoro viene esposta un'applicazione al piano di pubblicazione delle principali tavole relative alle Statistiche Strutturali sulle Imprese. Come noto tali statistiche sono armonizzate a livello europeo tramite il Regolamento (CE) n.295/2008 che prevede una struttura comune per la collezione, compilazione, trasmissione e valutazione dei dati relativi alle "statistiche strutturali sulle imprese industriali e dei servizi" (nel seguito SBS) che ne garantisce la qualità e la comparabilità internazionale. Le informazioni richieste ai singoli paesi sono regolamentate sia nella sostanza, elenco delle variabili e dettaglio della relativa classificazione, che nella forma, ovvero nella struttura del file da inviare ad Eurostat, dal regolamento stesso. Con il precedente regolamento (CE n.58/97) l'Istat provvedeva alla trasmissione delle suddette tabelle per le quali veniva segnalata la "soppressione primaria" (definita in base alla regola della minima frequenza o regola della soglia)<sup>2</sup> e solo successivamente Eurostat, attraverso un esame complessivo delle tabelle stesse e utilizzando il software CIF (Confidentiality InterFace), suggeriva le "soppressioni secondarie". L'attuale regolamento, richiede ai singoli stati membri di provvedere in maniera autonoma, prima della trasmissione dei dati, alla determinazione delle soppressioni secondarie ovvero alla protezione dal rischio di violazione della riservatezza dell'intero piano di pubblicazione.

In virtù di questa norma si è effettuata l'analisi di una possibile procedura di protezione delle tabelle SBS richieste da Eurostat in cui per l'individuazione delle soppressioni secondarie viene utilizzato il software Tau-Argus. Tale software è stato sviluppato all'interno del progetto europeo CASC (Computational Aspects of Statistical Confidentiality) e in successivi progetti in ambito europeo che promuovono lo sviluppo di strumenti e metodologie nell'ambito della teoria della tutela della riservatezza <http://neon.vb.cbs.nl/casc/..%5Ccase%5Ctau.htm>.

Il presente lavoro espone l'applicazione dello schema procedurale utilizzando quale riferimento il precedente Regolamento; tutte le indicazioni relative alla numerazione degli Annex e alla struttura delle tabelle fanno riferimento ad esso.

Nella seconda sezione vengono descritte le fasi concettuali del processo di protezione di tabelle collegate in Tau Argus.

Successivamente, nella terza sezione vengono analizzate le tabelle nei primi 4 Annex del Regolamento Eurostat 58/97 del 20 dicembre del 1996 (EC, EURATOM). Con riferimento a tali tabelle vengono quindi descritte le variabili classificatrici e le unità statistiche di riferimento. Nella quarta

<sup>1</sup> La variabile classificatrice è quella variabile le cui modalità vengono utilizzate per classificare il collettivo di riferimento.

<sup>2</sup> La "regola della minima frequenza" o "regola della soglia" definisce una cella "a rischio" se il numero dei rispondenti della cella è minore di una certa soglia  $\alpha$ . Nel caso delle tabelle SBS il valore soglia considerato è  $\alpha = 3$ .

sezione viene analizzata la prima fase del processo ovvero quella in cui a partire dalle indicazioni contenute nei primi 4 Annex, vengono definite nuove tabelle che, rispettando alcune caratteristiche, possano essere protette con il software Argus. Nella quinta sezione vengono analizzati i legami intercorrenti tra queste ultime tabelle in modo tale da individuare un ordine nello schema di protezione che permetta di definire un insieme coerente di soppressioni. Nel sesto sezione vengono brevemente descritti metodi e parametri successivamente adottati per la protezione delle tabelle in esame. Nella sezione successiva (sezione 7) viene descritta l'intera fase di applicazione del software Argus all'insieme delle tabelle definite nella quarta sezione. Tale fase viene suddivisa in diversi passi: predisposizione e lettura dei dati (sezione 7.1); definizione celle a rischio (sezione 7.2); determinazione soppressioni secondarie (sezione 7.3); salvataggio delle informazioni (sezione 7.4); costruzione e applicazione di History file (sezione 7.5). L'applicazione del software a tabelle successive alla prima viene analizzata nella sezione 8. Infine, nell'ultima sezione (sezione 9) viene sintetizzata l'applicazione del procedimento nel suo complesso.

## 2. Fasi del processo di protezione di tabelle collegate in Tau Argus

L'intero processo di protezione applicato alle tabelle SBS può essere schematizzato considerando quattro fasi fondamentali:

- FASE 0: analisi delle variabili classificatrici utilizzate per la determinazione delle tabelle definite nel piano di pubblicazione;
- FASE 1: analisi delle tabelle definite dal Piano di pubblicazione in modo da definire un insieme di nuove tabelle che, rispettando determinati requisiti, possano essere successivamente protette con l'utilizzo del software Argus (analizzato nella Sezione 4);
- FASE 2: individuazione dei legami esistenti tra le suddette tabelle così da determinare un ordine nel processo di protezione delle tabelle in questione (analizzato nella Sezione 5);
- FASE 3: applicazione del software Argus ad ogni tabella identificata nei passi precedenti in modo tale da ottenere uno schema complessivo di soppressioni secondarie coerente per l'intero insieme delle tabelle stesse (analizzato nelle Sezioni 6 e 7).

## 3. Analisi delle variabili classificatrici

Le tabelle che sono state considerate per la definizione del processo in esame sono quelle definite nei primi 4 Annex del precedente regolamento SBS. Le **variabili classificatrici** utilizzate nella determinazione delle tabelle SBS richieste da Eurostat, definite nei primi quattro Annex, sono:

- attività economica prevalente;
- variabile territoriale;
- variabile dimensionale numero di addetti;
- variabile dimensionale fatturato prodotto;

Le prime tre variabili sono di tipo gerarchico.

*Ricordiamo che una classificazione si dice **gerarchica** se si articola su più livelli di risoluzione ciascuno dei quali è incluso nel precedente (ovvero nel successivo).*

In particolare abbiamo che:

- la variabile *attività economica prevalente* viene codificata secondo la nomenclatura delle attività economiche della Comunità economica Europea o codice Nace (rev1.1). Nell'insieme delle tabelle il dettaglio massimo considerato per tale variabile è di 4 digit (livello classe); dal 2010 verrà utilizzata la classificazione NACE rev.2;

- la variabile *territoriale* viene codificata considerando la nomenclatura delle unità statistiche territoriali (NUTS); per tale variabile il livello NUTS2 è il massimo livello gerarchico considerato per la costruzione delle tabelle SBS (per l'Italia la NUTS2 corrisponde alla regione);
- la variabile *numero di addetti* presenta due classificazioni in parte coincidenti, sempre di tipo gerarchico, dipendenti dalle attività economiche considerate.

In particolare per i settori C-F  
si considera la classificazione

#### Classe Addetti Tipo A

1-19	
	1-9
	10-19
20 e oltre	
	20-49
	50-249
	250 +

mentre per i settori G, H, I, K

#### Classe Addetti Tipo B

1-19	
	1-9
	10-19
20 e oltre	
	20-49
	50-249
	250 +

Principalmente nelle tabelle SBS si considera come **unità statistica** di riferimento l'Impresa.

Con il termine Impresa<sup>3</sup> si indica l'unità giuridico-economica che produce beni e servizi destinati alla vendita e che, in base alle leggi vigenti o a proprie norme statutarie, ha facoltà di distribuire profitti realizzati ai soggetti proprietari siano essi privati o pubblici. Il responsabile è rappresentato da una o più persone fisiche, in forma individuale o associata, o da una o più persone giuridiche. Tra le imprese sono comprese: le imprese individuali, le società di persone, le società di capitali, le società cooperative, le aziende speciali di comuni o province o regioni. Sono considerate imprese anche i lavoratori autonomi e i liberi professionisti.

Alcune tabelle prendono in considerazione unità di riferimento differenti dall'impresa ovvero le unità locali e le unità funzionali (KAU - kind of activity units).

Per Unità locale<sup>4</sup> si intende un luogo fisico nel quale un'unità giuridico - economica (impresa, istituzione) esercita una o più attività economiche. In tale località, o a partire da tale località, si esercitano delle attività economiche per le quali una o più persone lavorano (eventualmente a tempo parziale) per conto della stessa unità giuridico - economica. Costituiscono esempi di unità locale le seguenti tipologie: agenzia, albergo, ambulatorio, bar, cava, deposito, domicilio, garage, laboratorio, magazzino, miniera, negozio, officina, ospedale, ristorante scuola, stabilimento studio professionale, ufficio, ecc.

Con il termine Unità funzionale (KAU)<sup>5</sup> si indica l'unità che all'interno di un'impresa raggruppa l'insieme delle parti che concorrono all'esercizio di un'attività economica a livello di classe (quattro cifre) della classificazione Nace Rev.1. Si tratta di un'entità che corrisponde a un sistema di informazioni che consentono di fornire o calcolare per ogni unità di attività economica almeno il valore della produzione, i consumi intermedi, le spese per il personale, il risultato di gestione, l'occupazione e gli investimenti fissi lordi.

Spesso queste tre tipologie di unità statistica coincidono tra loro, inoltre, per definizione, un'impresa contiene sempre almeno un'unità locali e almeno un'unità funzionale. Una qualche differenza si può verificare solo nel caso di imprese con più di 200 addetti, differenze che, come risulterà chiaro nel seguito dall'analisi dell'intero procedimento, non interferiscono sul processo di protezione applicato alle tabelle.

Nella tavola 1, per fornire un quadro d'insieme relativamente alle unità statistiche coinvolte negli Annex 1-4, sono riportate le variabili classificatrici e le unità statistiche di riferimento, con il relativo dettaglio.

<sup>3</sup> "Conti Economici delle Imprese" Anno 2003 – collana Informazioni n.8 2007 -Istat

<sup>4</sup> "Conti Economici delle Imprese" Anno 2003 – collana Informazioni n.8 2007 -Istat

<sup>5</sup> "Conti Economici delle Imprese" Anno 2003 – collana Informazioni n.8 2007 -Istat

Tutte le tabelle richieste negli Annex 1-4 presentano come variabile classificatrice la variabile Attività economica prevalente.

**Tavola 1 - Variabili classificatrici coinvolte nel piano di pubblicazione per Eurostat con relativo dettaglio della classificazione per ciascun tipo di unità statistica di riferimento**

Variabili Classificatrici	Unità statistica di riferimento
Attività economica prevalente (NACE 4)	Impresa e Unità funzionale
Attività economica prevalente (NACE 3) Numero di addetti (due classificazioni)	Impresa
Attività economica prevalente (NACE 3 o NACE 2) Variabile territoriale (NUTS 2)	Unità locali
Attività economica prevalente (NACE 2) Fatturato (classi)	Imprese

#### 4. Fase 1: costruzione di Tabelle di input per il software Argus

Per poter utilizzare il software Argus e proteggere le tabelle richieste dal Regolamento Eurostat è necessario riorganizzare le informazioni in modo tale da ottenere nuove tabelle che presentino la stessa unità statistica di riferimento e risultino omogenee in termini di definizione e livello delle variabili classificatrici utilizzate nella loro determinazione.

Nella tavola 2, per ogni tabella che verrà analizzata con il software Argus, vengono riportate le informazioni inerenti: alle variabili esplicative con relativa codifica; all'unità statistica di riferimento; agli Annex che contengono le informazioni incluse nella tabella predisposta per Argus.

**Tavola 2- Schema delle tabelle di Input per il software Argus**

TABELLE ANALIZZATE CON ARGUS	Variabili esplicative	Dettaglio Variabili esplicative	Unità statistica di riferimento	Annex
Tab1.1	Attività economica prevalente	NACE 4	Imprese	Annex 1A, Annex 2A, Annex 4A Annex 3B
Tab1.2	Attività economica prevalente	NACE 4	Unità funzionale	Annex 2E Annex 4E
Tab2.1	Attività economica prevalente per numero di addetti	NACE 3 per classi di addetti Tipo A	Imprese	Annex 2D Annex 4D
Tab2.2	Attività economica prevalente per numero di addetti	NACE 3 per classi di addetti Tipo B	Imprese	Annex 3C Annex 1B
Tab2.3	Attività economica prevalente per variabile territoriale	NACE 3 per NUTS-2	Unità locali	Annex 3E
Tab2.4	Attività economica prevalente per fatturato	NACE 3 per Classi di fatturato	Imprese	Annex 3C

## 5. Fase 2: analisi dei legami tra tabelle e definizione della sequenza di protezione

L'analisi dei legami tra tabelle svolta in questa fase definisce un ordinamento procedurale per la successiva applicazione del software Argus, ossia stabilisce una sequenza logica di passi successivi da intraprendere ai fini di protezione.

Le particolari caratteristiche, sia delle tabelle che delle variabili esplicative, che principalmente influenzano la procedura sono:

- la natura **gerarchica** di alcune classificazioni adottate per le variabili esplicative;
- il legame che intercorre tra le tabelle (*tabelle collegate*).

Prima di procedere al vero e proprio processo di protezione è necessario uno studio dei legami che intercorrono tra le tabelle che si vogliono proteggere. Durante tale analisi è necessario individuare la variabile di legame, che nel nostro caso è la variabile “attività economica prevalente”, e, relativamente a tale variabile, il livello gerarchico utilizzato nella costruzione delle tabelle coinvolte nel processo di protezione.

Una volta individuati gli eventuali collegamenti tra le tabelle, è importante definire uno schema di relazioni sulla base del quale si poggia l'intera procedura. In generale, nella costruzione di tale schema, si parte considerando la tabella che presenta un maggiore dettaglio relativamente alla variabile di collegamento. Tale scelta è determinata dal presupposto che un maggior dettaglio della variabile classificatrice implichi che tutte le celle marginali delle restanti tabelle rappresentino sempre delle aggregazioni di celle di tale tabella, come riportato nella figura 1.

Figura 1 - Esempio di legami tra tabelle gerarchiche

Nace	
C	50
CA	50
CA10	X
CA101	33
CA1010	
CA102	
CA1020	
CA103	
CA1030	
CA11	X
CA111	
CA1110	

	Classe di addetti					
Nace	1_9	10_19	20_49	50-249	250	
C						50
CA						50
CA10						X
CA101						33
CA102						...
CA103						
CA11						X
CA111						

Considerando inizialmente il caso delle imprese, dall'analisi della Tavola 2 si osserva che la tabella che presenta il maggior dettaglio della variabile classificatrice è la **Tab.1.1** che in quanto rappresenta le imprese a livello nazionale classificate per attività economica prevalente considerando la classificazione Nace con un dettaglio di 4 digit. Tale tabella verrà definita come “**Tabella di partenza**”. Lo schema che sintetizza i legami tra le tabelle riportate nella tavola 2 è presentato nella figura 2.

Analizziamo ora nel dettaglio ciascun legame esistente tra le tabelle. Per la costruzione dello schema, seguendo l'argomentazione sopra esposta, si procede considerando le tabelle che presentano un livello gerarchico immediatamente successivo a quello della Tabella di partenza (Nace 3-digit). Per quanto riguarda l'insieme delle tabelle coinvolte nel processo di protezione (Vedi tavola 2) due sono quelle che vengono prese in considerazione (**Tab2.1 e Tab2.2**) per definire un collegamento diretto con la tabella di partenza Tab1.1. Entrambe le tabelle presentano:

- la variabile “Attività economica prevalente” con un dettaglio di 3 digit.;
- la caratteristica di non essere tra loro collegate in quanto si riferiscono a settori diversi della variabile “Attività economica”.

due differenti classificazioni relativamente alla variabile “Classi di addetti” (seconda variabile esplicativa) dipendenti dal settore di attività economica a cui fanno riferimento.

Dall’analisi della tavola 2 si evince che il successivo livello gerarchico della variabile di legame (NACE 2-digit) coinvolge esclusivamente la tabella **Tab3.1 solo per le unità locali**. Tale tabella presenta dei legami sia con la Tab2.1 sia con la Tab2.2 in base ai settori in esse considerati.

Da notare che la Tab3.1 esclude le informazioni relative al settore **G** della classificazione Nace. Tali informazioni sono veicolate nella Tab2.3.

In particolare per tale settore:

- l’informazione relativa all’incrocio delle variabili “Attività economica” con “Localizzazione territoriale” viene rilasciata considerando la classificazione NACE a 3-digit (Tab 2.3) a differenza di quanto rilasciato per gli altri settori (NACE a 2-digit, Tab 3.1);
- Inoltre, viene diffusa una tabella aggiuntiva (Tab2.4), con la quale viene rilasciata l’informazione relativa all’incrocio tra le variabili “Fatturato” con “Attività economica prevalente” considerando per quest’ultima una classificazione NACE a 3-digit.

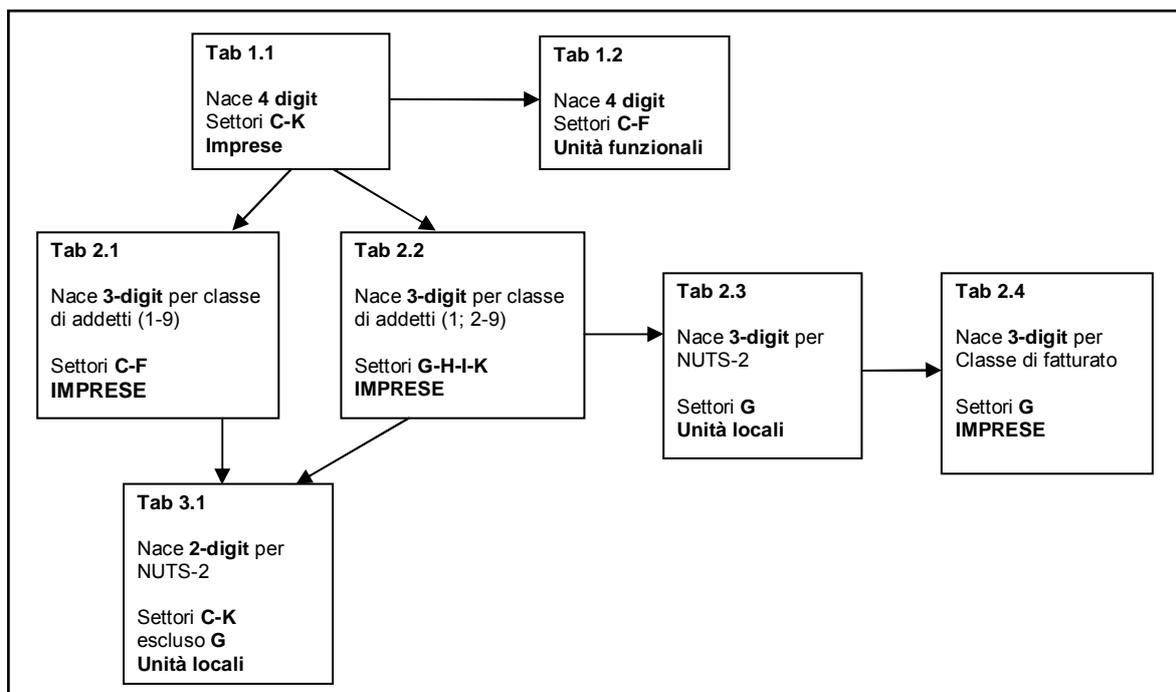
Le tabelle appena descritte (**Tab2.3 e Tab2.4**), nonostante presentino entrambi la variabile “Attività economica prevalente” con un dettaglio a 3-digit, non vengono collegate direttamente alla Tab1.1, ma vengono legate, in sequenza, alla tabella Tab2.2 (vedi Figura 2). Come risulterà più chiaro nei successivi paragrafi questo modo di collegare le tabelle è stato considerato in virtù di alcune potenzialità del software Argus. Infatti, tale software, permette di trasferire informazioni relative al processo di protezione da una tabella all’altra e di applicare un algoritmo di protezione condizionandolo alle informazioni trasferite.

Una differente considerazione deve invece essere fatta per la tabella **Tab1.2**.

Tale tabella presenta delle differenze con la tabella di partenza esclusivamente per le unità di riferimento considerate; in un caso l’impresa mentre nell’altro le unità funzionali. Le due tipologie di unità di riferimento sono tra loro strettamente collegate; per definizione una unità funzionale è sempre contenuta in un’impresa. Applicando alla Tab1.2 la protezione della tabella di partenza si otterrà la protezione desiderata. Infatti, un controllo sulle imprese è sempre più restrittivo rispetto a un controllo fatto sulle unità funzionali.

Tutti legami sopra descritti sono schematizzati nella figura 2.

**Figura 2 - “Schema dei legami tra tabelle”**



## 6. Fase 3: applicazione di Tau Argus alla sequenza di tabelle

Una volta definite sia le tabelle che lo schema che ne sintetizza i legami è possibile procedere alla fase di applicazione del software Argus. Tale applicazione deve essere fatta in maniera tale che sia coerente lo schema complessivo di soppressioni primarie e secondarie con l'intero insieme delle tabelle stesse. L'ordine generato dall'analisi dei legami tra le tabelle, come descritto nella figura 2, ha lo scopo di identificare celle comuni in tabelle collegate come nell'esempio riportato nella figura 1.

L'attuale versione del software Tau-argus (Versione 3.3.1 build 15), non riesce a gestire in maniera automatica tabelle collegate per le quali la variabile di legame presenti livelli gerarchici differenti. Di conseguenza, per poter proteggere tabelle così caratterizzate, si procede in maniera parzialmente manuale, usando il software SAS in alcuni passaggi.

Per ogni tabella dello schema riportato nella figura 2 occorre applicare il software Argus per ottenere la tabella protetta. L'applicazione del software Argus può essere a sua volta schematizzata in sei punti fondamentali:

1. Scelta della definizione di rischio e dei parametri necessari alla protezione;
2. Predisposizione e lettura dei dati;
3. Definizione delle celle a rischio - soppressioni primarie;
4. Applicazione metodo di protezione - soppressioni secondarie;
5. Salvataggio delle informazioni relative alla tabella protetta;
6. Trasferimento delle informazioni relative al processo di protezione: costruzione e applicazione dell'History file.

### 6.1. Definizione di rischio e parametri adottati per la protezione

La regola per la definizione delle celle a rischio (Istat 2004 utilizzata dall'Istat, è quella della minima frequenza; vengono considerate a rischio, e quindi non pubblicate, tutte quelle celle che presentano un numero di contributori (nel nostro caso imprese o unità funzionali) minore di tre quando le variabili classificatrici della tabella sono variabili presenti in pubblici registri (come nel caso delle variabili attività economica e classi di ampiezza) e la variabile risposta è confidenziale. Per la determinazione delle celle secondarie nel caso SBS viene considerata, come variabile costo, la variabile "Valore Aggiunto" nel caso in cui i contributori sono le imprese mentre viene utilizzata la variabile "Retribuzione" nel caso di unità funzionali.

Per la determinazione delle soppressioni secondarie è stato applicato il metodo modular (Hun-depool, et al. 2009). Tale metodo di soppressione è stato preferito agli altri implementati in Argus in quanto permette di tener in considerazione la natura gerarchica delle variabili. I dati di input per il software Argus (Tau-Argus User's Manual) sono di tipo tabellare. La scelta di considerare tabelle e non microdati è legata a diverse considerazioni. Sicuramente le principali motivazioni sono legate alle procedure di arrotondamento dei pesi campionari. L'utilizzo di microdati in fase di input ad Argus comporterebbe non solo la necessità di arrotondare a posteriori la somma dei pesi in ciascuna cella ma anche una diversa gestione delle celle che, a seguito dell'arrotondamento, raggiungono la soglia delle tre unità. Inoltre sembra più efficiente che la fase di mera gestione dei dati sia effettuata con gli strumenti appositamente sviluppati presso le indagini.

#### *Variabili da leggere nel software Argus*

In relazione alle caratteristiche del software e in base alla regola di definizione del rischio utilizzata, per ogni tabella vengono considerate esclusivamente le variabili classificatrici, la variabile costo (o variabile risposta) e la variabile frequenza.

Quanto appena affermato deriva dalle seguenti considerazioni:

- La regola della minima frequenza, non essendo una regola di "concentrazione", definisce una cella a rischio esclusivamente sulla base delle informazioni fornite dalla variabile frequenza e non dal valore assunto dalla variabile risposta all'interno della cella;
- Nel caso in cui il dato di input per Argus è un dato tabellare (così come nel caso che stiamo analizzando), le uniche variabili costo che possono essere considerate sono: la variabile

frequenza, la variabile unità e la variabile risposta utilizzata per la costruzione della tabella. Quindi la necessità di far coincidere la variabile costo con la variabile risposta. Come variabile costo, nel caso della Tab1.1 è stata considerata la variabile “Valore aggiunto”. Tale scelta è coerente con le precedenti applicazioni del processo di protezione.

### *Valori negativi*

Ulteriore attenzione deve essere posta ai valori della variabile risposta coinvolta nel processo di protezione. La versione di Tau-argus attualmente disponibile non supporta tabelle che presentino valori negativi per tale variabile. Un simile problema viene in qualche modo superato inserendo, e successivamente utilizzando, un'altra variabile che rappresenti il valore assoluto della variabile di interesse. Nel caso delle tabelle che stiamo analizzando tale soluzione non comporta alcun effetto sulla determinazione delle celle a rischio in quanto non viene utilizzata una regola di concentrazione ma esclusivamente la regola della minima frequenza. Un qualche effetto potrebbe esserci nella determinazione delle soppressioni secondarie, in quanto tale variabile viene utilizzata come variabile costo e quindi si potrebbe assegnare un costo troppo elevato a celle con contributi negativi. In letteratura vengono proposti due possibili modi per superare anche questo problema (Hundepool et al. 2009):

- utilizzare un “livello di protezione” (*protection level*) uguale a zero;
- utilizzare una variabile costo differente dalla variabile risposta modificata (nel nostro caso la variabile unità o la variabile frequenza).

### *Minimum frequency range*

Per poter descrivere il parametro del “*minimum frequency range*” bisogna fare alcune considerazioni.

Quando una tabella viene protetta attraverso la soppressione di celle, l'utilizzo delle relazioni lineari che intercorrono tra le stesse e le celle pubblicate (in particolari le celle marginali), permettono di costruire un intervallo (limite superiore e limite inferiore) all'interno del quale cade il valore vero della cella soppressa. Questo intervallo viene indicato come “*feasibility interval*” o “intervallo di fattibilità”.

Definire un’ “ampiezza accettabile” (“*safety range*”) per questo intervallo, può risultare uno strumento ulteriore sul quale agire per proteggere una tabella. Ovvero è possibile richiedere al software Argus di individuare un schema di soppressioni secondarie in modo tale che l'ampiezza dell’ “intervallo di fattibilità” rispetti determinati vincoli.

Nel caso in cui viene considerata come misura del rischio la regola della minima frequenza il parametro che definisce l'ampiezza minima dell'intervallo di fattibilità è il “*minimum frequency range*”. In particolare, se questo valore viene posto uguale a “30%” significa che il pattern delle soppressioni secondarie è determinato in modo tale che, per le celle sopresse, il limite inferiore dell'intervallo di fattibilità è al massimo pari al valore della cella in esame meno il 30% di tale valore ed analogo discorso per il limite superiore (al minimo valore della cella più 30%).

Ad esempio se il valore della cella soppressa è pari a 100 il limite inferiore dell'intervallo di fattibilità sarà minore o uguale di 70 mentre quello superiore sarà maggiore o uguale a 130.

*Va comunque segnalato che nel caso in cui, come regole di definizione del rischio venissero utilizzate le cosiddette regole di “concentrazione” (dipendenti dal valore della cella: regola pq, regola della dominanza), l’ “ampiezza accettabile” dell’intervallo di fattibilità sarebbe automaticamente determinata dalla regola stessa.*

Sempre nella finestra delle regole del rischio è presente un'altra opzione, indicata con “*manul safety range*”. Tale opzione ha la stessa funzione del “*minimum frequency range*” ma viene applicata a tutte quelle celle che manualmente vengono definite come celle a rischio (ovvero la definizione di cella a rischio non dipende da una funzione ma da una scelta del ricercatore).

## 7. Descrizione dell'applicazione del software Argus

In questa sezione sono esaminati nel dettaglio le fasi dell'applicazione vera e propria del software Argus descritte nella sezione precedente. Tali fasi (dalla 1 alla 5) sono esemplificate tramite la descrizione dell'applicazione alla prima tabella “*tabella di partenza*” dello schema della figura 2 (Tab.1.1).

### 7.1 Predisposizione e lettura dati con software Argus

La prima tabella che viene analizzata è la Tab1.1, ovvero la tabella relativa alle imprese a livello nazionale classificate per attività economica prevalente con dettaglio Nace a 4 digit.

I dati che ci sono stati forniti dal servizio di produzione degli stessi, sono relativi a tutte le variabili che devono essere trasferite ad Eurostat indipendentemente dall'Annex in cui le variabili stesse sono coinvolte. La forma è di tipo tabellare e la variabile “Attività economica prevalente” è espressa considerando esclusivamente la sezione e il dettaglio a 4 cifre della classificazione Nace (non è presente nel file l'informazione relativa alle sottosezioni e alle classificazioni 3 e 2-digit).

I dati forniti dal servizio sono stati quindi elaborati in modo tale da permettere di gestire correttamente le informazioni necessarie per il processo di protezione e rendere il formato compatibile con quello di input per il software Argus. In particolare Tau-Argus, gestisce file di dati ASCII o in formato libero con separatore specificato. Per ogni singola cella della tabella, nel file di input devono essere specificati: il valore delle variabili di classificazione; il valore delle variabili risposta; le frequenze ed eventualmente, relativamente alla variabile costo,<sup>6</sup> i tre valori massimi dei contributi alla cella stessa necessari nel caso in cui si adottino misure di rischio che utilizzano il concetto di concentrazione.

L'indicazione al software Argus sulla struttura gerarchia di una variabile classificatrice può essere data in due modi differenti. Il primo utilizza i digit relativi ai codici con cui è costruita la gerarchia della variabile (“*hierarchical levels*”); il secondo si avvale di file costruiti ad hoc (“*file .hrc*”) in cui viene definita l'intera struttura gerarchica.

In entrambi i casi, nel file dei dati devono essere incluse tutte le celle che definiscono la tabella in esame, comprese quindi quelle relative ai totali e ai sottototali definiti dalla struttura gerarchica della variabile stessa.

Nel primo caso (“*hierarchical levels*”) particolare attenzione deve essere posta alla definizione della gerarchia della variabile Nace che, per essere gestita in maniera automatica dal software Argus, deve presentare necessariamente tutti i suoi livelli (sezione, sottosezione, nace 2-digit, nace 3-digit, nace 4-digit). Nel caso in cui non siano presenti, come per le sezioni E, F, G, H, I, K, per automatizzare la gestione della gerarchia, è stata inserita una sottosezione fittizia (che quindi ripete i valori della sezione) denominata Z ottenendo così: sezione F, sottosezione FZ- sezione E, sottosezione EZ ecc.

Nella tavola 3 sono riportati come esempio parte del file dei dati e file dei metadati (file .rda) costruiti per la lettura della tab 1.1 in Tau- argus. In questo caso la gerarchia della variabile Nace viene gestita utilizzando i livelli gerarchici .

<sup>6</sup> La variabile costo è quella variabile che descrive il costo delle singole celle. Tali costi verranno minimizzati quando vengono individuate le celle secondarie. (Tau-Argus user-manual)

**Tavola 3 - Esempio dati di input in forma di tabella (colonna di sinistra) e file di metadati (colonna di destra)**

<pre>File dati T;2861697050;3740009 C;32058205;3510 CA;23426310;15 CA10;16975;4 CA101;10515;1 CA1010;10515;1 CA102;35;1 CA1020;35;1 CA103;6425;2 CA1030;6425;2 CA11;23409335;11 CA111;23249145;2 CA1110;23249145;2 CA112;160190;9 CA1120;160190;9 CB;8631895;3495 CB13;8770;1 CB132;8770;1 CB1320;8770;1 CB14;8623125;3494 CB141;2924540;1336 CB1411;2294265;1184 CB1412;545025;111 CB1413;85250;41 CB142;4501335;1659 CB1421;4381515;1571 CB1422;119820;88 CB143;55210;5 CB1430;55210;5 CB144;160735;40 CB1440;160735;40 CB145;981305;454 E;98363730;2521 E<del>Z</del>;98363730;2521 <del>EZ40;91528505;1633</del> EZ401;75804065;914 EZ4011;40053440;807 EZ4012;8158345;11 EZ4013;27592280;96 EZ402;14819395;664 EZ4021;148625;79 EZ4022;14670770;585 EZ403;905045;55 EZ4030;905045;55 EZ41;6835225;888 EZ410;6835225;888 EZ4100;6835225;888 . .</pre>	<pre>File .rda &lt;SEPARATOR&gt; ";" nace "999999" &lt;RECODEABLE&gt; &lt;TOTCODE&gt; T &lt;HIERARCHICAL&gt; &lt;HIERLEVELS&gt; 1 1 2 1 1 V_12150_1 &lt;NUMERIC&gt; FreqVar &lt;FREQUENCY&gt;</pre>
---	---

Nella prima colonna del file dei dati (file in formato in formato libero con separatore specificato) è riportata la variabile Nace;<sup>7</sup> nella seconda colonna i valori della variabile risposta modificata (ovvero il valore assoluto della variabile costo), nella terza colonna le frequenze. I valori relativi alla variabile risposta presenti nella tavola 3 non sono i valori effettivamente osservati nell'indagine ma valori fittizi.

<sup>7</sup> Nel caso di due variabili classificatrice il file di input viene organizzato considerando le righe della tabella corrispondente

Nel file dei metadati vengono invece inserite tutte le informazioni necessarie per leggere correttamente il file dei dati. Nel caso particolare:

<SEPARATOR> ", "        indica il tipo di separatore usato nel file dei dati;

**nace "999999"**

**nome della prima variabile e codice usato in caso di valore mancante;**

<RECODEABLE>

istruzione che indica come la prima variabile sia ricodificabile;

<TOTCODE> T

istruzione usata per indicare quale codice viene utilizzato per indicare il totale della prima variabile (T);

<HIERARCHICAL>

istruzione che indica che la prima variabile è gerarchica;

<HIERLEVELS> 1 1 2 1 1

valori che indicano i livelli di gerarchia relativi alla prima variabile.

**V\_12150\_1**

**nome della seconda variabile (in questo caso la variabile V\_12150\_1 rappresenta il valore assoluto del valore aggiunto V\_12150);**

<NUMERIC> >

istruzione che indica che la seconda variabile è numerica.

**FreqVar**

**nome della terza variabile;**

<FREQUENCY>

istruzione che indica che la terza variabile "FreqVar" è la variabile frequenza.<sup>8</sup>

Nella tavola 4 sono riportati come esempio parte del file dei dati (File .csv), file dei metadati (File .rda) e file di gestione della gerarchia (File .hrc) costruiti per la lettura della tab 1.1 in Tau-argus. In questo caso la gerarchia della variabile Nace viene gestita utilizzando il file costruito ad hoc nace4\_tab1.hrc.

<sup>8</sup> Nelle tabelle in cui l'unità di riferimento è l'impresa (Tab1.1, Tab2.1, Tab2.2, Tab2.4) la variabile FreqVar rappresenta la variabile V\_11110=Numero imprese; per le tabelle che fanno riferimento alle unità locali (Tab2.3, Tab3.1) la variabile FreqVar rappresenta la variabile V\_11210= Numero unità locali, infine per la tabella relativa alle unità funzionali (Tab1.2) la variabile FreqVar fa riferimento è la V\_11310= Numero di unità funzionali.

**Tavola 4 - Esempio di file dati di input in forma di tabella (colonna di sinistra); file di metadati e file .hrc (colonna di destra)**

<pre>File dati T;2861697050;3740009 C;32058205;3510 CA;23426310;15 CA10;16975;4 CA101;10515;1 CA1010;10515;1 CA102;35;1 CA1020;35;1 CA103;6425;2 CA1030;6425;2 CA11;23409335;11 CA111;23249145;2 CA1110;23249145;2 CA112;160190;9 CA1120;160190;9 CB;8631895;3495 CB13;8770;1 CB132;8770;1 CB1320;8770;1 CB14;8623125;3494 CB141;2924540;1336 CB1411;2294265;1184 CB1412;545025;111 CB1413;85250;41 CB142;4501335;1659 CB1421;4381515;1571 CB1422;119820;88 CB143;55210;5 CB1430;55210;5 CB144;160735;40 CB1440;160735;40 CB145;981305;454 E;98363730;2521 E40;91528505;1633 E401;75804065;914 E4011;40053440;607 E4012;8158345;11 E4013;27592280;96 E402;14819395;664 E4021;148625;79 E4022;14670770;585 E403;905045;55 E4030;905045;55 E41;6835225;888 E410;6835225;888 E4100;6835225;888 . . . . .</pre>	<pre>File .rda &lt;SEPARATOR&gt; ";" nace "999999"   &lt;RECODEABLE&gt;   &lt;TOTCODE&gt; T   &lt;HIERARCHICAL&gt;   &lt;HIERCODELIST&gt; nace4_tab1.hrc   &lt;HIERLEADSTRING&gt; # V_12150_1   &lt;NUMERIC&gt; FreqVar   &lt;FREQUENCY&gt;</pre> <pre>File .hrc nace4_tab1.hrc C #CA ##CA10 ###CA101 ####CA1010 ####CA102 ####CA1020 ####CA103 ####CA1030 ##CA11 ###CA111 ####CA1110 ####CA112 ####CA1120 #CB ##CB13 ###CB132 ####CB1320 ##CB14 ###CB141 ####CB1411 ####CB1412 ####CB1413 ####CB142 ####CB1421 ####CB1422 ###CB143 ####CB1430 ###CB144 ####CB1440 ###CB145 ####CB1450 D #DA ##DA15 ###DA151 ####DA1511 . . E #E40 ##E401 ###E4011 ####E4012 ####E4013 ##E402 ###E4021 ####E4022 ##E403 .. ..</pre>
---	---

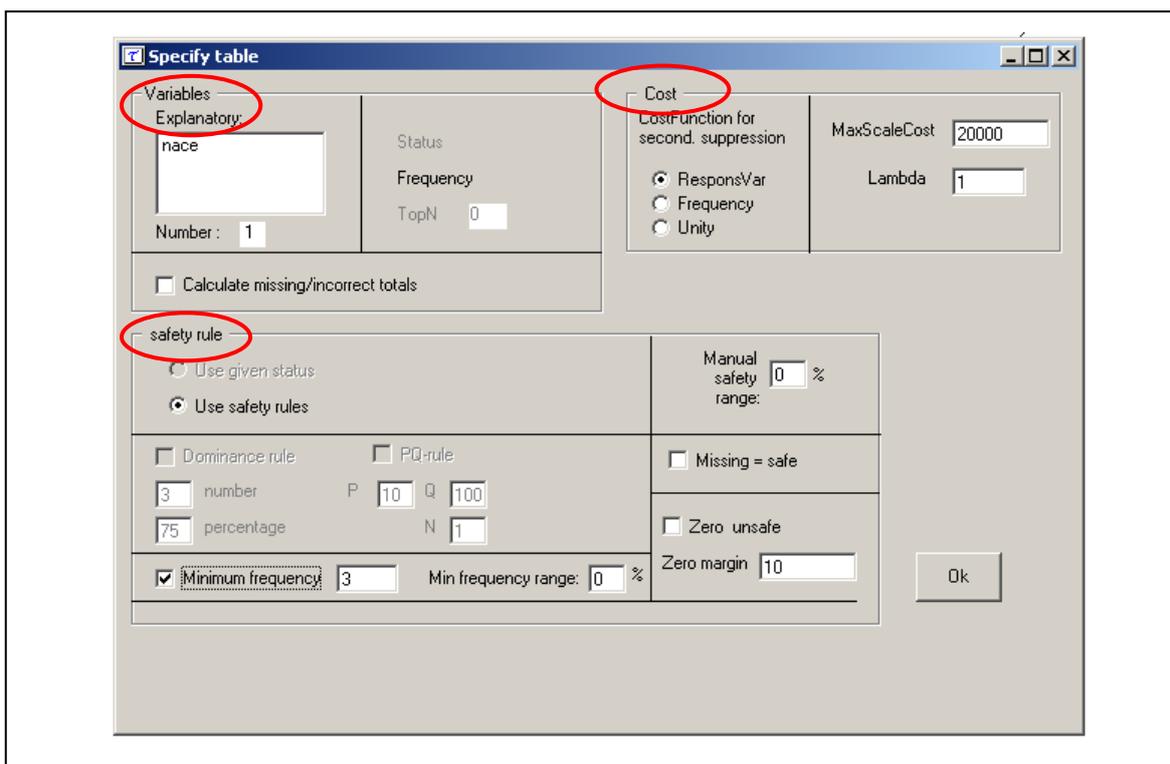
Dal confronto con i dati contenuti nella tavola 3 si può notare che in questo caso non è stata costruita nessun sottosezione fittizia (settosezione EZ) e nel file dei metadati in corrispondenza della variabile gerarchica viene indicato sia il nome del file contenente la struttura gerarchica della variabile (<HIERCODELIST> nace4\_tab1.hrc) che il simbolo (<HIERLEADSTRING> #) attraverso il quale sono gestiti i livelli gerarchici (per ulteriori indicazioni vedi Tau-Argus User's Manual pag. 42-43).

Una volta costruiti i file dei dati e dei metadati come sopra descritto, si passa alla lettura della tabella con il software Argus. Nella finestra “file\open table” del software è possibile specificare il percorso per entrambi i file, dati e metadati. Una volta che il software ha “letto” correttamente la tabella in esame si passa alla fase successiva.

## 7.2 Definizione delle celle a rischio (soppressioni primarie)

Per poter procedere alla determinazione delle celle a rischio bisogna specificare la regola che si vuole adottare ed alcuni “parametri” che la caratterizzano. Selezionando il menu “specify” apriamo finestra “specify table” riportate nella figura 3.

Figura 3 - Finestra specify table in Tau-Argus



Questa finestra è suddivisa in tre riquadri relativi rispettivamente alle variabili classificatrici, variabile costo e misura del rischio:

- **variabili classificatrici (o esplicative):** nel caso della Tab1.1 si considera una sola variabile ovvero la variabile Nace (Attività economica prevalente);
- **variabile costo:** in questo riquadro ci sono tre opzioni che permettono di selezionare come variabile costo una tra le seguenti: variabile risposta, variabile frequenza e variabile unità. Nel caso della Tab1.1 come variabile costo viene selezionata la variabile risposta. Tale variabile è il valore assoluto della variabile “valore aggiunto”;<sup>9</sup>
- **la regola di definizione del rischio:** in questo riquadro è possibile scegliere la regola per la definizione delle celle a rischio. Le regole proposte sono: regola della dominanza, la regola pq e la regola della minima frequenza (Tau-Argus User’s Manual). Per ogni misura del rischio è necessario determinare specifici parametri caratterizzanti le regole stesse.

<sup>9</sup> La variabile Valore Aggiunto =V12150 viene utilizzato come variabile costo per la protezione delle tabelle Tab1.1, Tab2.1, Tab2.2 e Tab2.4; per le tabelle Tab2.3, Tab3.1 viene utilizzata la variabile V\_13320=Retribuzioni ; infine per la tabella Tab1.2 viene utilizzata la variabile V\_12110=Turnover

Per le tabelle SBS si è utilizzata la regola della minima frequenza. Tra i parametri da specificare riguardanti tale regola abbiamo:

- il valore soglia e il valore del “*minimum frequency range*”.

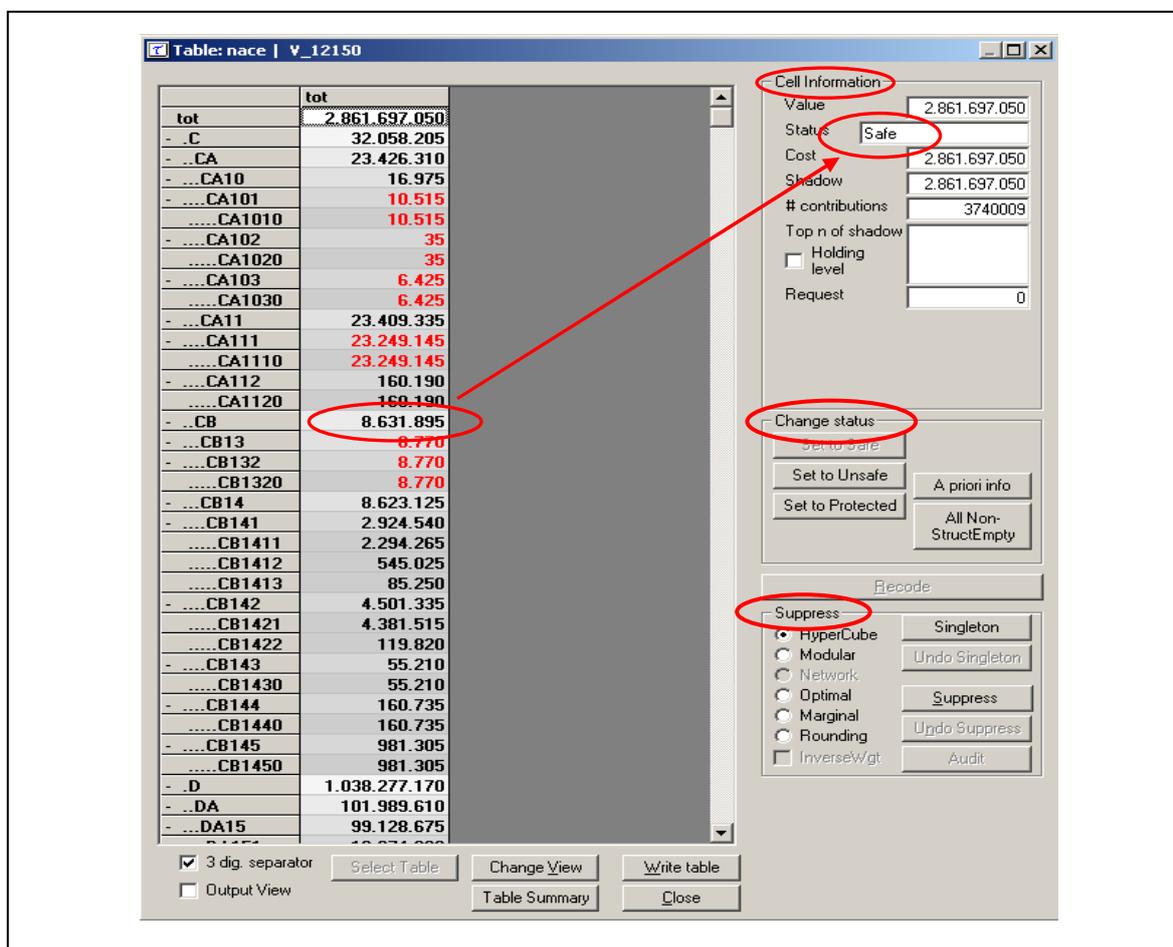
Il **valore soglia** rappresenta il valore minimo che deve essere assunto dalla variabile *frequenza* tale da rendere la cella protetta,<sup>10</sup> nel nostro caso viene fissato pari a tre.

Per poter ottenere risultati uguali a quelli proposti da Eurostat, sia il valore del “*minimum frequency range*” (Si veda sezione 4.) che del “*manul safety range*” (Si rimanda al sezione 4) sono stati posti uguale a 0.

### 7.3 Protezione – Determinazione soppressioni secondarie

Una volta decisa la regola per le soppressioni primarie e le opzioni ad essa associate si può procedere alla fase di protezione ovvero all’individuazione delle celle per le soppressioni secondarie. Selezionando il menu *Modify/View table* si ottiene la finestra riportata nella figura 4.

Figura 4 - Finestra view table in Tau-Argus - Tab1.1



Nel primo riquadro a sinistra viene visualizzata la tabella e il valore della variabile risposta assunto in ogni singola cella. Le celle visualizzate con il colore rosso sono quelle che non soddisfano la regola di definizione del rischio, ovvero le celle con soppressione primaria.

<sup>10</sup> Nel nostro caso la variabile frequenza è la variabile “numero di imprese” quindi il valore soglia rappresenta il numero minimo di contributori che rendono la cella protetta.

Sotto tale riquadro, nella parte bassa della schermata ci sono dei pulsanti relativi alla visualizzazione, al salvataggio e al riepilogo di tutte le informazioni riguardanti il processo di protezione e non della tabella (Tau-Argus User’s Manual). A destra della schermata compaiono tre riquadri.

Nel **primo** vengono sintetizzate una serie di informazioni relative alle singole celle. Per ogni cella selezionata nella tabella che compare a sinistra, vengono visualizzate le seguenti informazioni:<sup>11</sup>

- il valore assunto nella cella dalle variabili costo, esplicativa e “shadow”;<sup>12</sup>
- il numero dei rispondenti;
- lo “status” della cella.

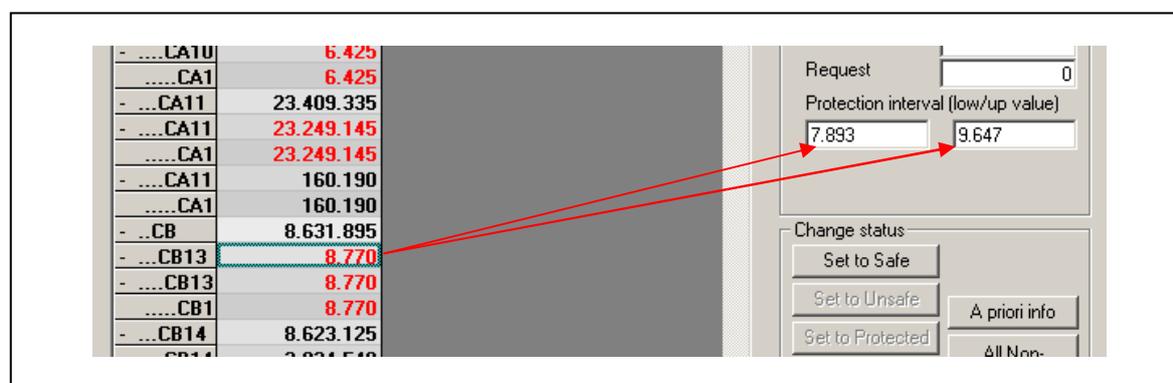
Il software Argus assegna ad ogni cella uno “status” ovvero un **attributo relativo al processo di protezione**. L’insieme dei possibili “status” che una cella può assumere è presentato nella tavola 5:

**Tavola 5 - Descrizione differenti tipi di Status**

STATUS	Descrizione Status
Safe	la cella non è a rischio in quanto non viola la regola del rischio scelta.
Manually safe	la cella viene imposta come cella non a rischio in maniera manuale durante la sessione di lavoro (indipendentemente dalla regola di definizione del rischio).
Unsafe	cella è a rischio in quanto viola la regola del rischio scelta.
Unsafe (frequency)	la cella è considerata a rischio in quanto viola la regola della minima frequenza.
Unsafe (from manual)	la cella è imposta come cella rischio in maniera manuale durante la sessione di lavoro.
Protected	la cella viene considerata protetta e non può essere selezionata come candidata per una soppressione secondaria.
Secondary	la cella è stata selezionata come soppressione secondaria.

Inoltre, sempre nel primo riquadro, selezionando le celle che non soddisfano la definizione del rischio, è possibile visualizzare i valore degli estremi (inferiore e superiore) del così detto intervallo di fattibilità (Vedi paragrafo 6). Per l’esempio visualizzato nella figura 5 è stato considerato un valore di “*minimum frequency range*” pari al 10%.

**Figura 5 - Finestra view table in Tau-Argus - Tab1.1- “Intervallo di fattibilità”**



Nel **secondo** riquadro, una serie di pulsanti danno la possibilità all’utente di cambiare lo status o a singole celle (set to: safe, protect, unsafe) o ad un insieme strutturato di esse (a priori info). In particolare quest’ultimo caso (**a priori** info) risulta essere molto importante per la nostra procedura nel momento in cui, nel determinare lo schema complessivo delle soppressioni secondarie e non, si vogliono considerare i legami tra più tabelle gerarchiche come varrà spiegato nel seguito.

<sup>11</sup> Le ulteriori informazioni top.shadow, holding level e request non sono rilevanti in questo caso. Per la loro trattazione dettagli si veda Tau-Argus User’s Manual pag. 65.

<sup>12</sup> La variabile Shadow è la variabile che è usata per applicare la regola del rischio. Per maggiori dettagli si rimanda a Tau-Argus User’s Manual pag 55.

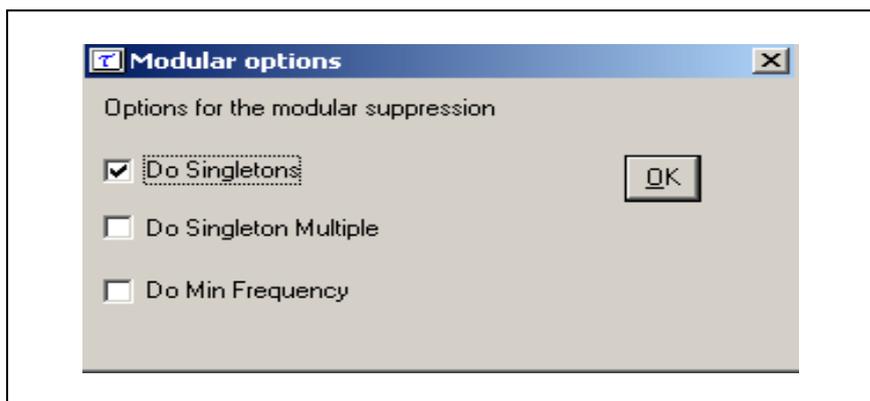
Nel **terzo** riquadro sono invece presenti alcuni pulsanti di opzione che permettono all'utente di selezionare il metodo di soppressione (Hypercube, Modular, Network, Optimal, Marginal e Rounding) che si intende utilizzare per proteggere la tabella inserendo in essa le soppressioni secondarie (si veda Hundepool, et al. 2009 per la descrizione dei singoli metodi). Per ogni metodo sarà poi necessario fissare specifici parametri caratterizzanti i metodi stessi.

Come già indicato, nella procedura di protezione applicata alle tabelle SBS, e quindi alla Tab1.1, per l'individuazione delle soppressioni secondarie è stato utilizzato il metodo "Modular".

Una volta selezionato il metodo si procede all'individuazione delle celle secondarie cliccando il bottone "suppress" in questo stesso riquadro. In relazione al tipo di metodo scelto compare la finestra relativa ai parametri specifici del metodo.

Nel nostro caso ("Modular approach") la finestra è presentata nella figura 6.

**Figura 6 - Finestra delle opzioni relative all'approccio modular**



In questa finestra compaiono tre caselle di controllo selezionando le quali è possibile gestire la protezione dei cosiddetti "singleton".

Una cella viene definita "singleton" nel caso in cui il valore di essa proviene da un **solo** contributore.

In questo caso è possibile richiedere una maggiore protezione in quanto il singolo rispondente può facilmente superare le protezioni adottate. I casi che risultano essere maggiormente a rischio si presentano quando su una stessa riga o una stessa colonna si hanno:

- due celle singleton;
- una cella singleton e una cella a rischio (ovvero, nel nostro caso, con due soli contributori).

Selezionando la prima opzione ("Do Singleton") il metodo modular gestisce le soppressioni secondarie così da evitare il caso 1 ovvero in modo tale che i contributori, relativi alle due celle singleton presenti sulla stessa riga o sulla stessa colonna, non possano risalire ad informazioni riservate. Ciò significa che due celle singleton non possono vicendevolmente proteggersi ma risulta necessaria l'introduzione di una soppressione ulteriore affinché possano considerarsi protette.

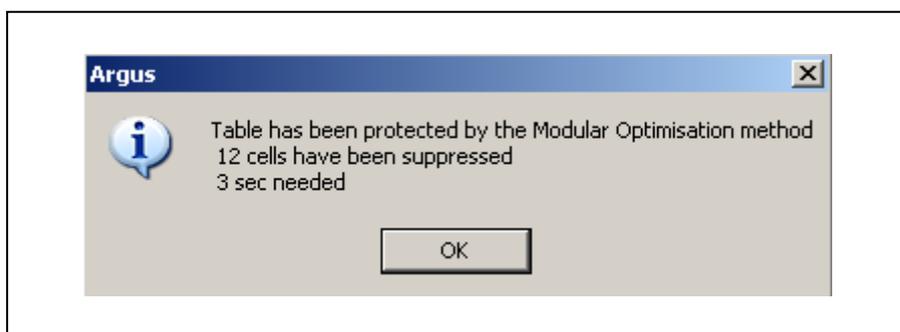
La seconda opzione ("Do Singleton Multiple") prende in considerazione il caso 2) ovvero il caso in cui sulla stessa riga o colonna si presentino una cella singleton ed una a rischio. Anche in questo caso sono necessarie soppressioni ulteriori.

La terza opzione (Do min Frequency) fa riferimento al caso in cui vengano considerate più regole di definizione del rischio contemporaneamente (es. minima frequenza e regola della dominanza). Tale situazione non viene presa in esame nella procedura di protezione delle tabelle Sbs.

Nel caso della procedura applicata alle tabelle Sbs si è scelto di utilizzare esclusivamente l'opzione "Do Singleton" applicando così una maggiore protezione esclusivamente ai casi in cui due singleton si presentano sulla stessa riga o stessa colonna.

Una volta selezionata una o più opzioni relative al metodo di soppressione si procede alla individuazione delle cosiddette celle o soppressioni secondarie.

Alla fine del processo viene visualizzata la seguente finestra in cui è riportata l'informazione relativa al numero di celle selezionate come celle secondarie e al tempo utilizzato dal software per ottenere tale soluzione.



I risultati del processo di protezione, sono riportati nella finestra presentata nella figura 7.

Figura 7 - Finestra specify table in Tau-Argus - Tab1.1 con soppressioni secondarie

	tot
tot	2.861.697.050
C	32.058.205
CA	23.426.310
CA10	16.975
CA101	10.515
CA1010	10.515
CA102	35
CA1020	35
CA103	6.425
CA1030	6.425
CA11	23.409.335
CA111	23.249.145
CA1110	23.249.145
CA112	160.190
CA1120	160.190
CB	8.631.895
CB13	8.770
CB132	8.770
CB1320	8.770
CB14	8.623.125
CB141	2.924.540
CB1411	2.294.265
CB1412	545.025
CB1413	85.250
CB142	4.501.335
CB1421	4.381.515
CB1422	119.820
CB143	55.210
CB1430	55.210
CB144	160.735
CB1440	160.735
CB145	981.305

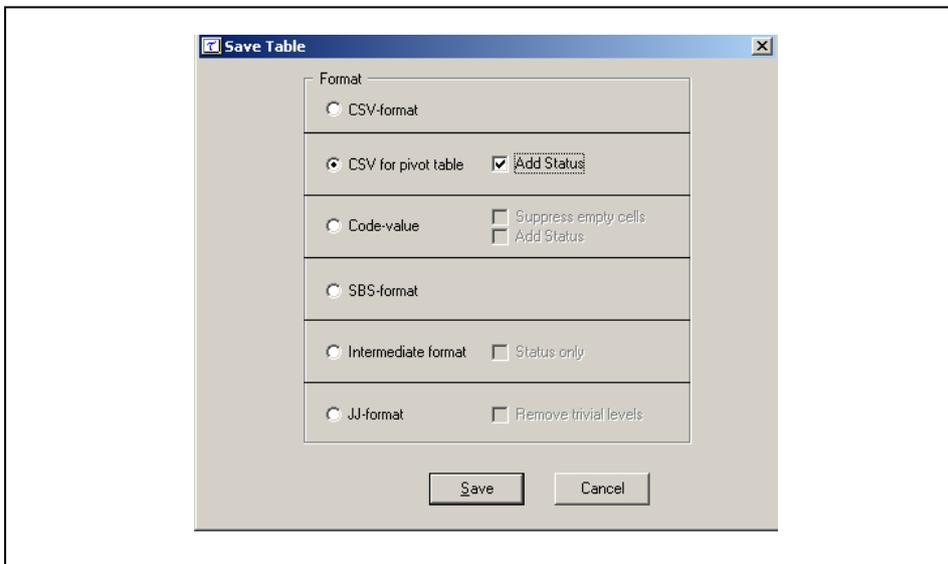
Tale finestra (Figura 7) presenta la stessa struttura di quella commentata nella figura 4. La differenza è la presenza delle soppressioni secondarie che, nel riquadro a sinistra dove è visualizzata la tabella, vengono evidenziate con il colore blu.

Si noti come nel riquadro relativo alle informazioni sulle celle sia cambiato lo status della cella evidenziata nella figura 4 (rispetto alla Figura 7) che da "Safe" è passata a "Secondary".

## 7.4 Salvataggio delle informazioni relative alle soppressioni

Una volta definite le soppressioni primarie e secondarie il risultato può essere salvato selezionando il bottone “Write table”; viene così visualizzata la finestra riportata nella figura 8 che permette di scegliere il formato nel quale si vuole salvare la tabella.

**Figura 8 - Finestra relativa alla scelta dei formati di output della tabella protetta**



Il software Argus consente l'utilizzo di sei formati. Per maggiori specifiche su tali formati rimandiamo al manuale del software.

Nel nostro caso viene utilizzato il formato *CSV for pivot table*, che consente di utilizzare le agevolazioni legate alle tabelle pivot di Excel. Viene inoltre selezionata la possibilità di salvare le informazioni sullo “status” delle singole celle. Tale informazione è particolarmente importante in quanto è necessaria per poter gestire in modo coerente il processo di protezione per tabelle collegate.

Nella figura 9 si riporta un esempio del file prodotto utilizzando il formato appena descritto.

**Figura 9 - formato di output derivante dall'opzione csv for pivot con l'opzione add status**

```
"nace","V_12150_1","Status"

"Total",2861697050,1
"C",32058205,1
"CA",23426310,1
"CA10",16975,1
"CA101",10515,5
"CA1010",10515,5
"CA102",35,5
"CA1020",35,5
"CA103",6425,5
"CA1030",6425,5
"CA11",23409335,1
"CA111",23249145,5
"CA1110",23249145,5
"CA112",160190,11
"CA1120",160190,11
"CB",8631895,1
"CB13",8770,5
"CB132",8770,5
"CB1320",8770,5
"CB14",8623125,11
"CB141",2924540,1
```

Come si può notare, non vengono riportate le informazioni sulle frequenze delle celle mentre le informazioni sullo “status” sono indicate con dei valori numerici. La tavola 6 riporta i codici maggiormente utilizzati nel nostro processo e il relativo significato in termini di status.<sup>13</sup>

**Tavola 6 - Insieme dei possibili codici numerici e significato corrispondente**

Codice Argus	Status
1	safe
2	manually safe
5	Unsafe (freq)
9	unsafe (from manual)
11	Secondary
14	Empty

Nel momento in cui si richiede di salvare la tabella protetta, qualsiasi sia il formato scelto, il software Argus crea un file con estensione .saf relativo ai dati (Vedi Figura 9) e un file report, in formato html, in cui vengono sintetizzati tutti i dettagli relativi alla tabella e al processo di protezione applicato.

In particolare vengono registrate le informazioni relative:

- al nome dei file dei dati e dei metadati utilizzati per la costruzione della tabella;
- alla struttura della stessa (variabile risposta, esplicativa e costo);
- alla regola di definizione del rischio utilizzata;
- al metodo di soppressione e le eventuali ricodifiche applicate.

Per ottenere una tabella protetta è necessario così oscurare tutte le celle che corrispondono ai codici 5, 11 e 9 della variabile “status”.

## 7.5 Trasferimento delle informazioni relative al processo di protezione: costruzione e applicazione dell' History file

Un volta prodotta la tabella Tab1.1.saf (tabella protetta), bisogna definire una procedura che permetta di proteggere le ulteriori tabelle (contenute nello schema rappresentato in Figura 2) in modo tale che le soppressioni ad esse applicate siano coerenti con quanto già effettuato. La coerenza delle soppressioni si raggiunge passando le informazioni relative allo status delle celle di una tabella alla tabella successiva come presentato nello schema della figura 1. Le informazioni relative alle celle sono contenute nelle informazioni a priori e il meccanismo di passaggio è la costruzione e applicazione dell'History file come presentato nei paragrafi successivi.

### 7.5.1 Informazioni a Priori

Una delle caratteristiche del software Argus è quella di permette di specificare un insieme di “informazioni a priori” su una collezione di celle definita dall'utente. Un informazione a priori non è altro che un attributo che caratterizza le singole celle di una tabella rispetto al processo di protezione. I codici che è possibile associare alle celle, i relativi significati e le corrispondenti azioni intraprese dal software sono descritti nella tabella seguente.

**Tavola 7 - Insieme dei possibili codici con relativi significati e azioni del software Argus**

Codice Alfanumerico	Significato	Azione del software Argus
U	Unsafe	La cella deve essere protetta <sup>14</sup>
S	Safe	La cella non è a rischio;
P	Protected	La cella non può essere utilizzata come soppressione secondaria

<sup>13</sup> La lista completa di codici Argus con relativo significato è inserita nel file di “report” prodotto dal software dopo aver salvato la tabella protetta.

<sup>14</sup> Quando una cella deve essere protetta significa che non solo deve essere applicata una soppressione primaria alla cella stessa, ma viene analizzata la possibilità che sia necessaria una o più soppressioni secondarie.

Il poter utilizzare delle informazioni a priori per vincolare le azioni del software, in termini di processo di protezione, rappresenta una grande flessibilità del software stesso. Tale flessibilità può essere utilizzata per imporre uno *schema coerente* di soppressioni ad un insieme di tabelle tra loro collegate.

### 7.5.2 History file

Dopo aver protetto la tabella Tab1.1 si passa a considerare le tabelle che risultano ad essa direttamente collegate (Tab2.1 e Tab2.2 come da schema in Figura 2).

Per semplicità di esposizione prendiamo in considerazione solo una delle due tabelle a cui è collegata la Tab1.1. In particolare prendiamo in esame la Tab2.1 che indichiamo con il termine “*tabella corrente*”, mentre indichiamo con il termine “*tabella precedente*” la Tab1.1.

La tabella corrente, per costruzione, presenta alcune celle marginali in comune con quelle della tabella precedente (Vedi Figura 1).

Questo implica che, affinché il processo di protezione sia coerente tra le due tabelle, le soppressioni (primarie e secondarie) applicate alle celle in comune, devono essere riprodotte nella tabella *corrente*. In modo analogo tutte le celle comuni che nella tabella *precedente* non vengono soppresse non devono presentare soppressioni nella tabella *corrente*.

Per ottenere tale risultato viene utilizzata la flessibilità delle *informazioni a priori* che sono inserite in un file, denominato “*History file*” per la tabella *corrente*, contenente le informazioni a priori che il software Argus applicherà come vincoli derivanti dal processo di protezione applicato alla tabella *precedente*.

I vincoli (informazione a priori) che vengono applicati sono differenti a secondo dello *status* associato da Argus alle celle in comune nel processo di protezione della tabella *precedente* (File .saf).

Per ogni codice della variabile status, nella tavola 8 indichiamo: il significato del codice, l’informazione a priori associata da applicare come vincolo alla tabella corrente e l’azione che il software Argus intraprende.

**Tavola 8 - Significato del codice relativo allo Status, l’informazione a priori associata da applicare come vincolo alla tabella corrente e l’azione che il software Argus intraprende**

CODICE VARIABILE STATUS ASSOCIATO ALLA TABELLA PRECEDENTE	Significato	Informazione a priori da applicare al processo di protezione della tabella corrente	Azione applicata da Argus
1	Safe	P	Tale cella non viene selezionata come possibile soppressione secondaria.
2	Manually safe	P	Tale cella non viene selezionata come possibile soppressione secondaria
3	Unsafe	-	Argus per definizione, riconosce tale cella come soppressione primaria; non viene quindi imposto nessun vincolo
4	Unsafe (request)	-	Argus per definizione, riconosce tale cella come soppressione primaria; non viene quindi imposto nessun vincolo
5	Unsafe (freq)	-	Argus per definizione, riconosce tale cella come soppressione primaria; non viene quindi imposto nessun vincolo
6	Unsafe (zero cell)	-	Argus per definizione, riconosce tale cella come soppressione primaria; non viene quindi imposto nessun vincolo
7	Unsafe (singleton)	-	Argus per definizione, riconosce tale cella come soppressione primaria; non viene quindi imposto nessun vincolo
8	Unsafe (singleton manual)	U	Tale cella viene protetta
9	Unsafe (manual)	U	Tale cella viene protetta
10	Protected	P	Tale cella non viene selezionata come possibile soppressione secondaria

**Tabella 8 segue - Significato del codice relativo allo Status, l'informazione a priori associata da applicare come vincolo alla tabella corrente e l'azione che il software Argus intraprende**

CODICE VARIABILE STATUS ASSOCIATO ALLA TABELLA PRECEDENTE	Significato	Informazione a priori da applicare al processo di protezione della tabella corrente	Azione applicata da Argus
11	Secondary suppression	U	Tale cella viene protetta
12	Secondary (from manual)	U	Tale cella viene protetta
13	Empty (non-structural)	-	
14	Empty	-	

Tecnicamente, il file *History* è un file in formato .csv in cui vengono registrate le modalità relative alle variabili classificatrici delle celle in comune tra le tabelle in esame (precedente e corrente) e il codice relativo all'informazione a priori ad esse associate secondo lo schema rappresentato nella tavola 8 segue. Nel nostro caso, a partire dalle informazioni contenute nel file .saf, il file History è stato costruito utilizzando alcuni istruzioni sas.

Con l'ultima versione del software Argus (3.3.1 build 12) è possibile salvare direttamente, in un apposito file, le informazioni necessarie per la costruzione di un file History. Nella finestra "make a priori file", riportata nella figura 10, sono presenti due riquadri. Nel primo viene indicato il nome del file relativo alla tabella protetta rispetto alla quale si vogliono determinare i vincoli, il nome che si vuole assegnare al file history, il numero e il nome delle variabili coinvolte. Nel secondo riquadro, per ogni status di Argus, è possibile associare una informazione a priori corrispondente all'azione che vogliamo applicare durante il processo di protezione.

Con i bottoni "go" e "ready" rendiamo effettiva la creazione del file.

**Figura 10 - Finestra relativa alla determinazione file history**

Status	Omit	Safe	Unsafe	Protect	Weight	Weight Value
Safe	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="text"/>
Safe (manual)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="text"/>
Unsafe	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="text"/>
Unsafe (request)	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="text"/>
Unsafe (Freq)	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="text"/>
Unsafe (Zero cell)	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="text"/>
Unsafe (Singleton)	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="text"/>
Unsafe (Singleton) (manual)	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="text"/>
Unsafe (manual)	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="text"/>
Protected	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="text"/>
Secondary	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="text"/>
Secondary (from man.)	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="text"/>
Empty (non-struct.)	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="text"/>
Empty	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="text"/>

Per una effettiva applicazione di tale file alla tabella Tab2.1 è comunque necessario apportare alcune modifiche. Bisogna infatti selezionare le celle in comune tra la tabella precedente (Tab1.1) e la tabella corrente (Tab2.1).

Il file history relativo ai vincoli da imporre alla tabella Tab2.1 sulla base del processo di protezione applicato alla tabella Tab1.1 è stato denominato H2\_1.csv.

### 8. Applicazione del software Argus alle tabelle successive alla prima

Una volta protetta la prima tabella dello schema riportato nella Figura 2 si passa alla protezione delle tabelle successive; in particolare in questo caso alla Tab2.1. Tutti i passaggi descritti nei precedenti paragrafi relativi alle fasi di predisposizione e lettura dei dati e definizione delle celle a rischio vengono ripetuti in maniera del tutto analoga a quanto fatto per la tabella Tab1.1. Per completezza nella Tavola 9 sono riportati come esempio parte del file dei dati (File .csv), file dei metadati (File .rda) e file di gestione della gerarchia (File .hrc) costruiti per la lettura della tab 2.1 in Tau- argus.

**Tavola 9 - Esempio di file dati di input in forma di tabella (colonna di sinistra); file di metadati e file .hrc (colonna di destra) tabella tab2.1**

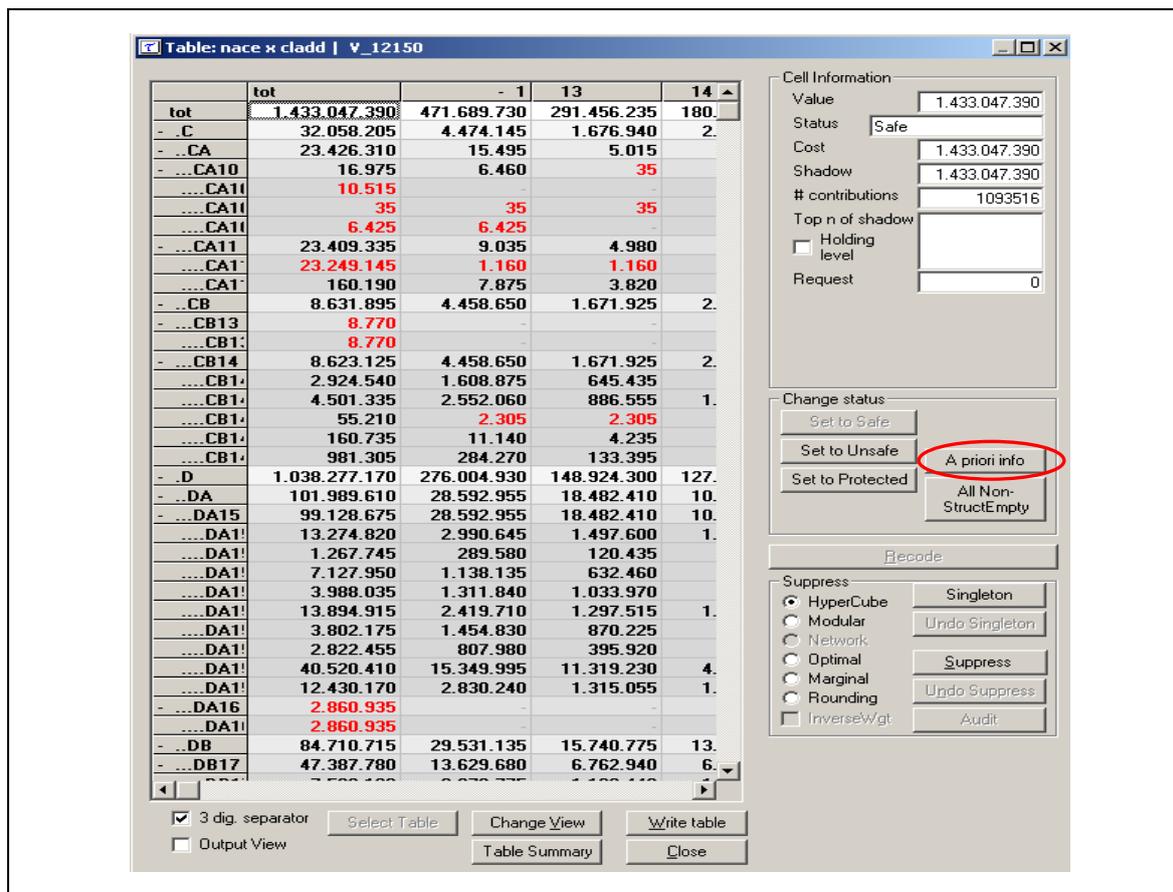
<p>File dati</p> <p>C,1,4474145,3170  C,13,1676940,2563  C,14,2797205,607  C,2,27584060,340  C,21,2406925,273  C,22,1853380,64  C,23,23323755,3  C,T,32058205,3510  CA,1,15495,8  CA,13,5015,5  CA,14,10480,3  CA,2,23410815,7  CA,21,18975,1  CA,22,133340,4  CA,23,23258500,2  CA,T,23426310,15  CA10,1,6460,3  CA10,13,35,1  CA10,14,6425,2  CA10,2,10515,1  CA10,23,10515,1  CA10,T,16975,4  CA101,2,10515,1  CA101,23,10515,1  CA101,T,10515,1  CA102,1,35,1  CA102,13,35,1  CA102,T,35,1  CA103,1,6425,2  CA103,14,6425,2  CA103,T,6425,2  CA11,1,9035,5  CA11,13,4980,4  CA11,14,4055,1  CA11,2,23400300,6  CA11,21,18975,1  CA11,22,133340,4  CA11,23,23247985,1  CA11,T,23409335,11</p>	<p>File .rda</p> <pre>&lt;SEPARATOR&gt; "," nace "99999" &lt;RECODEABLE&gt; &lt;TOTCODE&gt; T &lt;HIERARCHICAL&gt; &lt;HIERCODELIST&gt; nace3_cla1.hrc &lt;HIERLEADSTRING&gt; # cladd "99" &lt;RECODEABLE&gt; &lt;TOTCODE&gt; T &lt;HIERARCHICAL&gt; &lt;HIERCODELIST&gt; cla1.hrc &lt;HIERLEADSTRING&gt; # V_12150 &lt;NUMERIC&gt; FreqVar &lt;FREQUENCY&gt;</pre> <p>File .hrc nace3_cla1.hrc</p> <pre>C #CA ##CA10 ###CA101 ###CA102 ###CA103 ###CA11 ###CA111 ###CA112 #CB ##CB13 ###CB132 ###CB14 ###CB141 ###CB142 ###CB143 ###CB144 ###CB145 .. ..</pre>
--	---

La procedura di protezione per tale tabella presenta delle differenze nella terza fase in quanto, nel determinare le soppressioni secondarie, è necessario tener conto dei vincoli derivanti dalla protezione della tabella Tab1.1 o, nel caso generale, delle soppressioni applicate nelle tabelle precedenti.

#### 8.1 Utilizzo dell'History file

Dopo aver ripetuto i passi descritti nella sezione 6.1 e 6.2 si seleziona il menu *Modify/View table* ottenendo la finestra descritta nella figura 10 (analoga a quella rappresentata nella Figura 2).

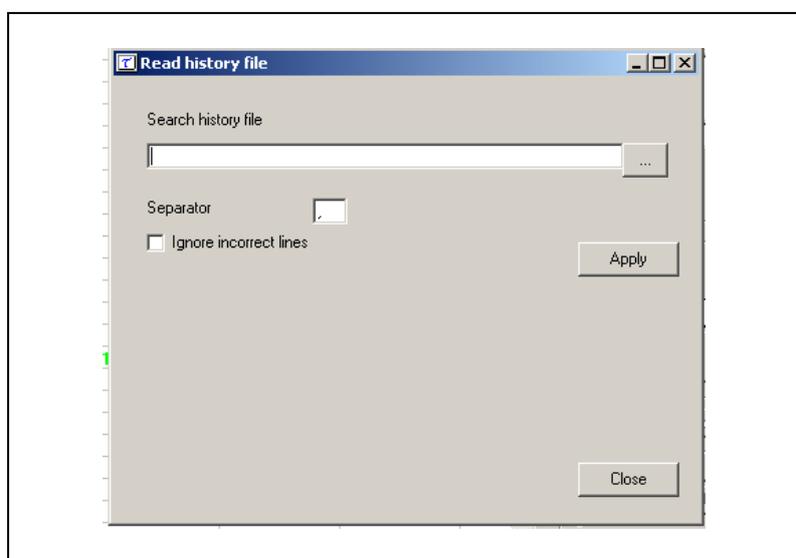
Figura 11 - Finestra view table in Tau-Argus - Tab2.1 con individuazione soppressioni primarie



Come per la tabella Tab1.1 in rosso vengono evidenziate le celle che non soddisfano la regola di definizione del rischio (soppressione primaria).

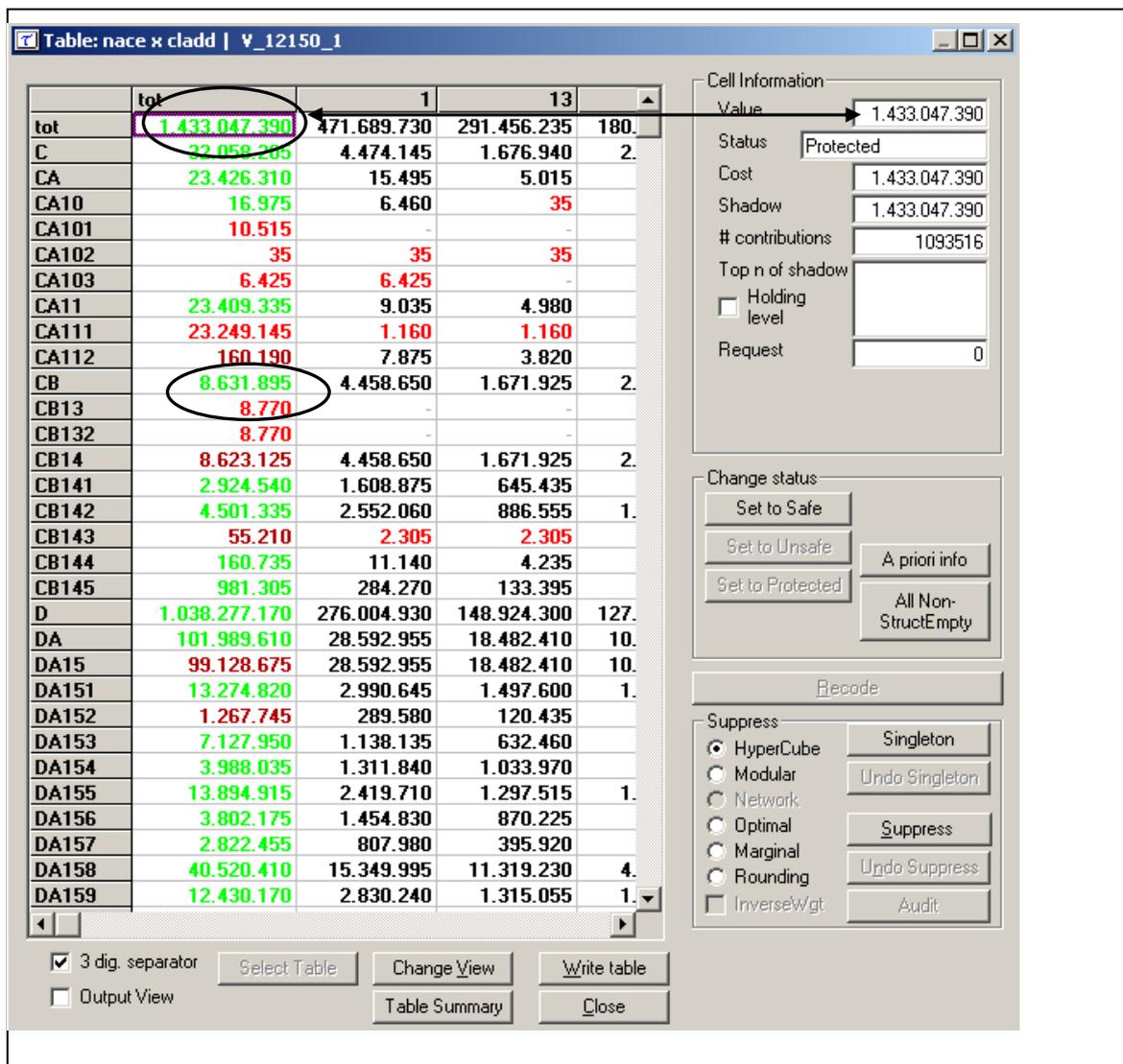
Selezionando il pulsante “A priori info” si apre la finestra seguente (Figura 12). In tale finestra è richiesto di specificare il percorso del file di history che si intende applicare alla tabella in esame (nel nostro caso H2\_1.csv ). Selezionando il pulsante “Apply” i vincoli contenuti nel file history vengono applicati alla tabella.

Figura 12 - Finestra A-priori info in Tau-Argus



Tali vincoli vengono evidenziati nella finestra *view table* così come rappresentato nella figura 13.

**Figura 13 - Finestra *view table* in Tau-Argus - Tab2.1 con individuazione soppressioni primarie - celle "protette" e celle "unsafe"**



Le celle visualizzate con il colore verde sono le celle alle quali è associato lo status "Protected".

Le celle a cui viene associato tale status sono quelle comuni tra la *tabella corrente* (Tab2.1) e quella *precedente* (Tab1.1) e che in quest'ultima, dopo l'applicazione del processo di protezione, non presentano soppressioni. Pertanto, affinché il processo di protezione sia coerente tra le due tabelle, tali celle, nel processo di protezione della *tabella corrente*, non vengono selezionate per la determinazione delle soppressioni secondarie (vengono così rilasciate).

Le celle visualizzate in rosso scuro sono invece quelle a cui è associato lo stato di "Unsafe". In questo caso le celle a cui viene associato tale status sono quelle comuni tra la *tabella corrente* (Tab2.1) e quella *precedente* (Tab1.1) e che in quest'ultima, dopo l'applicazione del processo di protezione, presentano soppressioni. Come conseguenza, anche nella tabella corrente, il valore di tali celle non può essere rilasciato. Il software Argus deve quindi trattarle alla stessa stregua delle celle che non soddisfano la regola del rischio applicata.

Una volta attribuite le informazioni a priori utilizzando il file di history "H2\_1.csv" si procede, in maniera del tutto analoga a quanto precedentemente fatto per la tabella Tab1.1, nella selezione del metodo per l'individuazione delle celle da sottoporre a soppressione secondaria.

Nella costruzione dei file “history” contenenti i vincoli da applicare alle varie tabelle, una particolare attenzione deve essere fatta alle unità di riferimento delle stesse.

Ricordiamo che più unità locali possono far riferimento ad una sola impresa così come una unità funzionale è sempre contenuta in un’impresa.

Nel caso di tabelle che presentano la stessa unità di riferimento (per brevità, nel seguito caso A), qualunque natura essa abbia (impresa, unità locale, unità funzionale), vengono vincolate esclusivamente le celle che durante il processo di protezione della tabella di partenza sono state scelte come “soppressioni secondarie”. Le soppressioni primarie presenti nella tabella di partenza saranno sicuramente presenti anche nella tabella collegata. Questo in quanto, utilizzando per l’individuazione delle celle a rischio (soppressioni primarie) la regola della minima frequenza, se una cella risulta essere a rischio nella tabella di partenza, tale cella presenterà la stessa caratteristica anche nella tabella collegata.

Quando due tabelle non presentano la stessa unità statistica di riferimento, si possono avere due fattispecie:

- caso B: l’unità statistica di riferimento della tabella di partenza “contiene” una o più unità di riferimento della tabella corrente (impresa -unità locale; impresa - unità funzionale);
- caso C: l’unità statistica di riferimento della tabella di partenza è “contenuta” nell’ unità statistica di riferimento della tabella corrente (unità locale-impresa; unità funzionale-impresa).

Nel caso B vengono vincolate sia le celle che nella tabella di partenza sono state scelte come “soppressioni secondarie” sia quelle individuate come soppressioni “primarie”. Infatti, se nella tabella di partenza l’unità di riferimento è l’impresa mentre in quella collegata è l’unità locale, potrebbe verificarsi il caso in cui una cella a rischio (primaria es. frequenza=2) nella tabella di partenza, non venga riconosciuta come tale nella tabella collegata (frequenza associata maggiore di 2). Come conseguenza avremmo che una stessa cella risulterebbe oscurata nella tabella di partenza ma pubblicata nella tabella corrente.

Nel caso C vengono vincolate esclusivamente le celle che durante il processo di protezione della tabella di partenza sono state scelte come “soppressioni secondarie”. Le soppressioni primarie presenti nella tabella di partenza saranno sicuramente presenti anche nella tabella collegata in quanto per caratteristica dell’unità di riferimento di quest’ultima il numero di contributori delle celle risulta inferiore.

## 9. Procedimento generale

Seguendo lo schema di relazioni rappresentato nella figura 1, il procedimento di protezione fin qui descritto viene applicato in maniera analoga a tutte le tabelle presenti nello schema stesso.

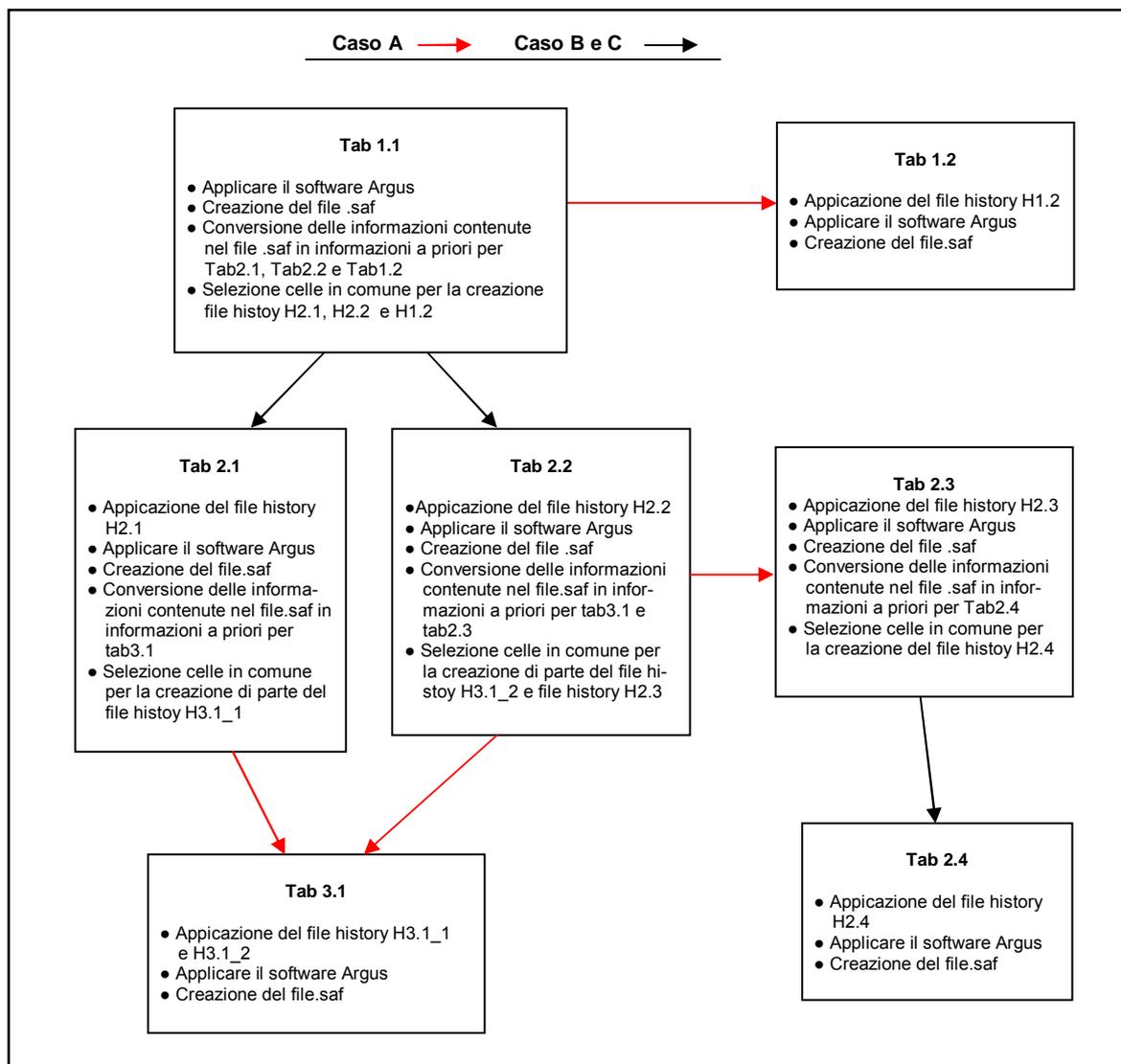
L’intera procedura può essere schematizzata come rappresentato nella figura seguente (Figura 14).

La definizione della schema di soppressioni primarie e secondarie, fin qui analizzato, ha preso in considerazione una sola variabile ovvero la variabile risposta che coincide inoltre con la variabile costo.

La definizione delle soppressioni da applicare alle altre variabile contenute nelle tabelle previste nel Regolamento europeo può avvenire in due modi differenti.

Si può applicare l’intera procedura ad ogni singola variabile contenuta nella tabella oppure si può scegliere una variabile come *variabile di riferimento* e applicare, lo schema di soppressioni determinato per tale variabile a tutte le altre contenute nelle tabelle in esame. In quest’ultimo caso inoltre, per ogni variabile contenuta nelle tabelle Sbs, bisogna verificare se le soppressioni secondarie, definite considerando la variabile di riferimento, corrispondono a celle con valore nullo. Se ciò accade, per tale variabile, è necessario applicare l’intera procedura di protezione considerando la stessa come variabile costo.

Figura 14 - “Schema della Procedura di Protezione”



## 10. Conclusioni

La procedura descritta nei precedenti paragrafi è stata applicata alle principali tabelle previste dal precedente regolamento SBS. Le tabelle protette ottenute dall'applicazione della procedura sopra descritta utilizzando il software Tau-Argus sono state confrontate con le tabelle protette prodotte da Eurostat con altra procedura (Feuvrier, Faes-Cannito 2003).

Da questo confronto si è potuto verificare da una parte, una sostanziale equivalenza nei risultati in termini di numerosità di celle soppresse dall'altra invece si può osservare, nella scelta delle celle da sopprimere (soppressioni secondarie), la preferenza di celle che presentano un livello gerarchico inferiore. In particolare, confrontando la tabella tab2.1 e gli annex relativi (2D e 4D; vedi tavola 2) si può notare come, con il software Tau-Argus, si ottenga uno spostamento di soppressioni da celle relative ad un livello gerarchico maggiore a quelle con un livello gerarchico più dettagliato. Viene così maggiormente garantita l'informazione relativa ai totali delle tabelle rispetto a quella contenuta nelle celle interne e, conseguentemente, una maggiore utilità delle tabelle rilasciate. In altri termini, l'utilizzo del software Tau-Argus, causa le routine di ottimizzazione ivi implementate, garantisce una maggiore efficienza in termini di contenuto informativo preservato a parità di protezione conseguita.

L'intero procedimento può essere applicato ad un generico insieme di tabelle collegate. Per tale nuovo insieme, l'intero sistema di relazioni e di legami che intercorrono tra le stesse deve essere studiato ed analizzato così da rendere applicabile la procedura di protezione con il software Argus.

Lo schema concettuale descritto nei precedenti paragrafi non dipende inoltre né dalla particolare funzione del rischio applicata né dalle parametrizzazioni considerate. Questo permette di applicare funzioni del rischio differenti e metodi di ottimizzazioni per la determinazione delle celle secondarie differenti dal modular (De Wolf 2002) quali ad esempio optimal (Fischietti, Salazar 1998) o hypercube (Giessing, Repsilber 2002).

Questo lavoro è stato propedeutico all'effettivo utilizzo del software nella produzione degli aggregati definiti nel regolamento per le Statistiche Strutturali sulle Imprese. Lo schema concettuale sopra descritto è stato concretamente utilizzato per la protezione delle tabelle definite nel nuovo regolamento Eurostat 295/2009 e ciò ha rappresentato per l'Istituto la prima esperienza in questo campo. Pertanto attualmente, così come richiesto dal nuovo regolamento Eurostat 295/2009, l'Istat produce autonomamente tabelle protette.

## Riferimenti bibliografici

- Capobianchi, A., L. Franconi. 2009. Cell suppression in linked tables from structural business statistics using Tau Argus 3.3.0: a conceptual framework. Disponibile su: [http://epp.eurostat.ec.europa.eu/portal/page/portal/research\\_methodology/documents/S18P1\\_CELL\\_SUPPRESSION\\_IN\\_LINKED\\_TABLES\\_CAPOBIANCHI\\_FRAN.pdf](http://epp.eurostat.ec.europa.eu/portal/page/portal/research_methodology/documents/S18P1_CELL_SUPPRESSION_IN_LINKED_TABLES_CAPOBIANCHI_FRAN.pdf) (18-20 febbraio 2009)
- de Wolf, P.P. 2002. HiTaS: A Heuristic Approach to Cell Suppression in Hierarchical Tables. *Inference Control in Statistical Databases*, Domingo-Ferrer,(Eds), Lecture Notes in Computer Science, Vol.2316, Springer.
- de Wolf, P.P. 2007. Cell suppression in a special class of linked tables. Work Session on Statistical confidentiality, Manchester, Luxembourg, December 17-19, 220-226 (2007). Office for Official Publications of the European Communities, ISBN 978-92-79-12055-8.
- de Wolf, P.P., S. Giessing. 2008. How to Make the T-ARGUS Modular Method Applicable to Linked Tables. *Privacy in Statistical Databases 2008*, Domingo-Ferrer, J., Saygin, Y. (EDS), Lecture Notes in Computer Science, Vol.5262, Springer, Heidelberg, 227-238.
- de Wolf, P.P., A. Hundepool. 2010. Three Ways to Deal with a Set of Linked SBS Tables Using t-Argus. *Privacy in Statistical Databases 2010*, Domingo-Ferrer J., Magkos E. (EDS), Lecture Notes in Computer Science, Vol. 6344: Springer Heidelberg, 66-73
- Feuquier, P., F. Faes-Cannito. 2003. Cell suppression in Eurostat on Structural Business Statistics – an example of Statistical Disclosure Control on tabular data. Work Session on Statistical Data Confidentiality, Luxembourg, April 7-9 2003. Disponibile su: <http://www.unece.org/stats/documents/2003/04/confidentiality/wp.21.e.pdf>
- Fischietti, M. J.J.Salazar-Gonzales 1998. Models and Algorithms for Optimizing Cell Suppression in Tabular Data with Linear Constraints. Technical Paper, University of La Laguna, Tenerife.
- Giessing, S. Repsilber D. 2002. *Tools and Strategies to Protect Multiple Tables with the GHQUAR Cell Suppression Engine* Inference Control in statistical Databases. Domingo-Ferrer, (Eds), Lecture Notes in Computer Science, Vol.2316, Springer.
- Giessing, S., Schmidt K. 2010. Techniques for Using T-Argus Modular on sets of Linked Tables – SAS Implementation. Short paper in CD proceedings. Work Session on Privacy in Statistical Databases 2010: Corfù, Greece September 22-24 2010.
- Hundepool, A., J. Domingo-Ferrer , L. Franconi, S.Giessing, R. Lenz, J. Longhurst, E. Schulte Nordholt, G. Seri and P.P De Wolf. (2009) *Handbook on Statistical Disclosure Control*. Disponibile su: [http://neon.vb.cbs.nl/casc/.%5CSDC\\_Handbook.pdf](http://neon.vb.cbs.nl/casc/.%5CSDC_Handbook.pdf)
- Hundepool, A. van de Wetering A., Ramaswamy R. de Wolf P.P., Giessing S., Fischietti M., Salazar J.J., Castro J. Lowthian P. Tau-Argus User's Manual version 3.3. Disponibile su: <http://neon.vb.cbs.nl/casc/Software/TauManualV3.3.pdf>
- Jansson, I., Bernström F., Carlson M.. 2010. The Process of Practicing Statistical Disclosure control in Tabular Data at Statistics Sweden. Work Session on European Conference on Quality in Official Statistics. Helsinki, Finland May 3-6 2010.

## Informazioni per gli autori

La collana è aperta ad autori dell'Istat e del Sistema statistico nazionale, e ad altri studiosi che abbiano partecipato ad attività promosse dal Sistan (convegni, seminari, gruppi di lavoro, ecc.). Da gennaio 2011 essa sostituirà Documenti Istat e Contributi Istat.

Coloro che desiderano pubblicare sulla nuova collana dovranno sottoporre il proprio contributo alla redazione degli Istat Working Papers inviandolo per posta elettronica all'indirizzo [iwp@istat.it](mailto:iwp@istat.it). Il saggio deve essere redatto seguendo gli standard editoriali previsti, corredato di un sommario in italiano e in inglese; deve, altresì, essere accompagnato da una dichiarazione di paternità dell'opera. Per la stesura del testo occorre seguire le indicazioni presenti nel foglio di stile, con le citazioni e i riferimenti bibliografici redatti secondo il protocollo internazionale 'Autore-Data' del *Chicago Manual of Style*.

Per gli autori Istat, la sottomissione dei lavori deve essere accompagnata da una mail del proprio dirigente di Servizio/Struttura, che ne assicura la presa visione. Per gli autori degli altri enti del Sistan la trasmissione avviene attraverso il responsabile dell'ufficio di statistica, che ne prende visione. Per tutti gli altri autori, esterni all'Istat e al Sistan, non è necessaria alcuna presa visione. Tutti i lavori saranno sottoposti al Comitato di redazione, che valuterà la significatività del lavoro per il progresso dell'attività statistica istituzionale. La pubblicazione sarà disponibile su formato digitale e sarà consultabile on line.

Gli articoli pubblicati impegnano esclusivamente gli autori, le opinioni espresse non implicano alcuna responsabilità da parte dell'Istat. Si autorizza la riproduzione a fini non commerciali e con citazione della fonte.