Outline

1. Creating Input Data
   - Microdata
   - Metadata
2. From primary to secondary suppression
   - Specifying tables
   - Avoiding Primary suppression
   - Computing secondary suppression
   - Audit
3. Outputs
   - Format SBS
   - Excel Formatting
4. Three special cases
   - Linked tables management
   - Controlled rounding
   - Tabulated data

Outline

1. Creating Input Data
   - Microdata
   - Metadata

2. From primary to secondary suppression
   - Specifying tables
   - Avoiding Primary suppression
   - Computing secondary suppression
   - Audit

3. Outputs
   - Format SBS
   - Excel Formatting

4. Three special cases
   - Linked tables management
   - Controlled rounding
   - Tabulated data

Microdata – from a SAS table

VIEWTABLE: TMP1.donnees_demo

	ape	TREFF	dep	poids_sondage	export
1	P8559	tr1	01	1	26
2	M7010	tr2	01	1	96781.3
3	L6820	tr1	01	1	484.8
4	L6820	tr1	01	1	0.3
5	L6820	tr1	01	1	0.3
6	C1712	tr3	01	1	17808.8
7	C2030	tr3	01	1	15063
8	G4674	tr3	01	1	13808.9
9	L6820	tr1	01	1	54.5
10	L6820	tr1	01	1	0.4
11	L6820	tr1	01	1	0.2
12	H5210	tr1	01	1	2866.9
13	H4939	tr3	01	1	122.9
14	C1013	tr1	01	3.3	466.4
15	C1011	tr3	01	1	2309.2
16	L6820	tr1	01	1	0.1
17	L6820	tr1	01	1	0.2
18	Q8610	tr3	01	1	11075.6
19	C2573	tr3	01	1	2377.6
20	F4110	tr1	01	11.1	37.1
21	C2599	tr4	01	1	65422.8

Input variable types :

- ventilation variables
- weigh
- response variables
- *holding* variable (not shown here)

Microdata – converting a SAS table to ascii file (.asc)

```
/** Étape 1 : conversion en caractère **/
data microDonnees;
            set demo.Donnees_demo;
/* 2 variables numériques avec un chiffre après la virgule que l'on convertit en variable caractère */
            poids       =           put(poids_Sondage, 6.1);
            exportt     =           put(export, 10.1);
run;

/** Étape 2 : création du fichier plat **/
filename asc "D:\dossier de travail\Donnees_demo.asc";
data _NULL_;
            set microdonnees;
            file asc;
            put
                        treff                                   1-4
                        APE                                     6-11
                        dep                                    13-14
                        poids                                  16-21
                        exportt                                23-32 ;
run;
```

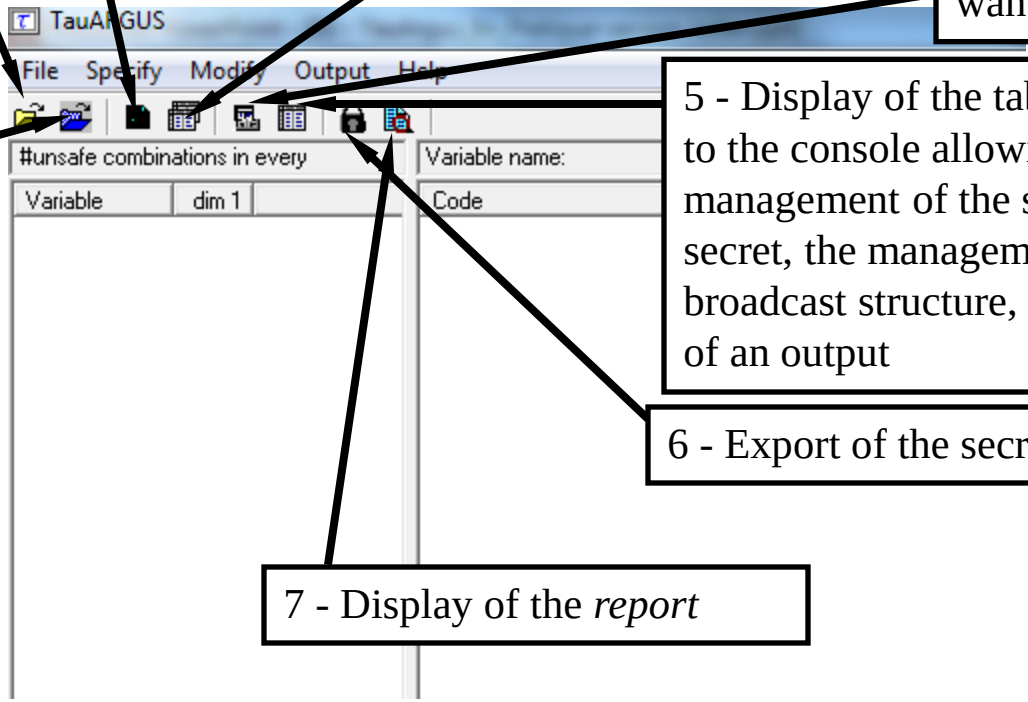
Tau-Argus: Exploring the menu

1 - Opening of microdata (.asc) and possibly associated metadata (.rda)

2 - Modification of the metadata

3 - Specifying tables

4 - Selection of the table on which you want to work



1 - Opening tabulated data (.tab)
And its metadata (.rda)

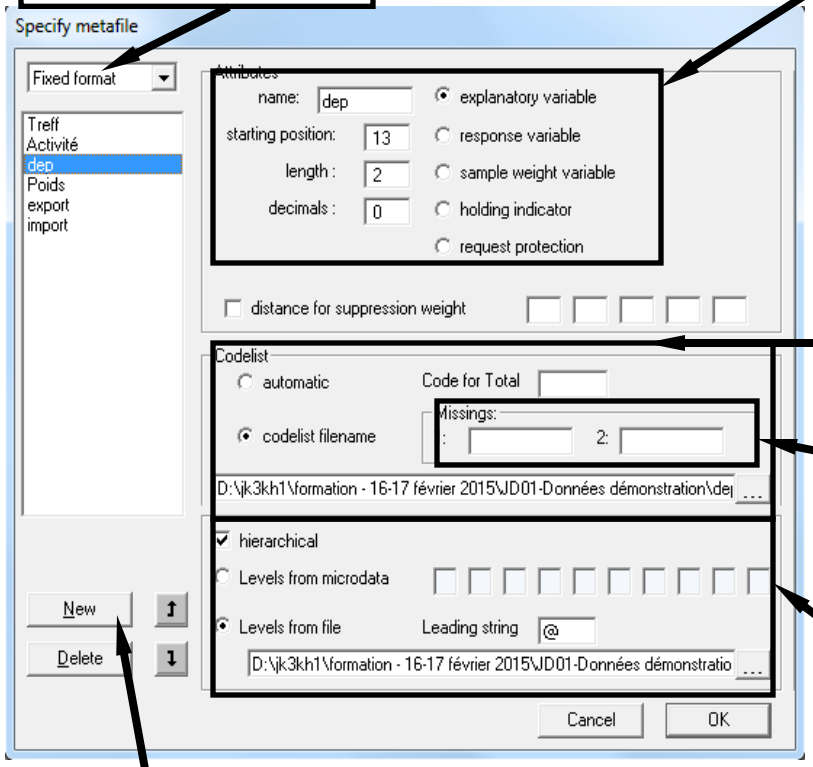
5 - Display of the tables and access to the console allowing the management of the secondary secret, the management of the broadcast structure, the recording of an output

6 - Export of the secret mask

7 - Display of the *report*

Metadata - variables specification

1 - We prefer the fixed format

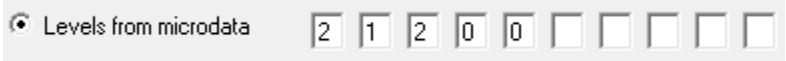


2 - We add the right number of variables

3 - Description of the variable:
- Name
- Position and length in the file (see sas export)
- Number of decimals for numeric variables
- Type

4 – *Codelist*
- management of labels ... little interest
- codes used for missing values

5 – Specifying information about hierarchical variables:
- manually : A88/A272/APE will give for example :



- or via a flat file (.hrc).

Outline

1. Creating Input Data
   - Microdata
   - Metadata

2. From primary to secondary suppression
   - Specifying tables
   - Avoiding Primary suppression
   - Computing secondary suppression
   - Audit

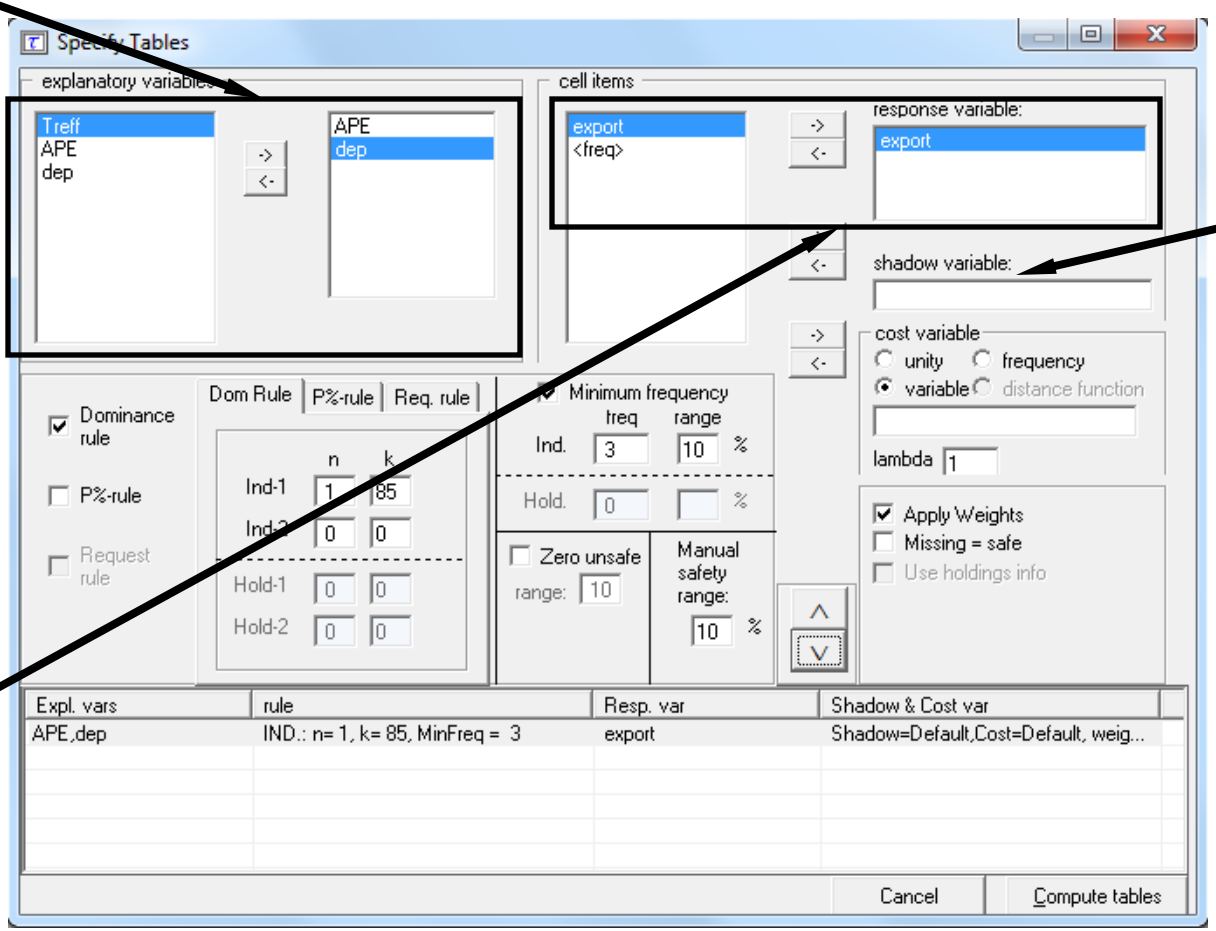
3. Outputs
   - Format SBS
   - Excel Formatting

4. Three special cases
   - Linked tables management
   - Controlled rounding
   - Tabulated data

Specifying tables - selecting variables

Selection of ventilation variables for a given table

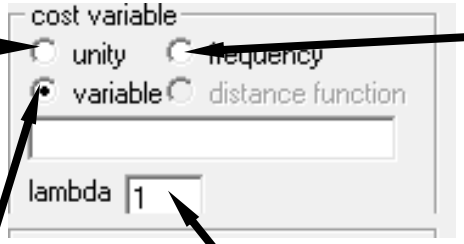
Selecting the Response Variable



Variable on which the primary secret will be calculated

Specifying tables - choosing the cost variable

Unity
Minimize the number of hidden boxes in the secondary secret.



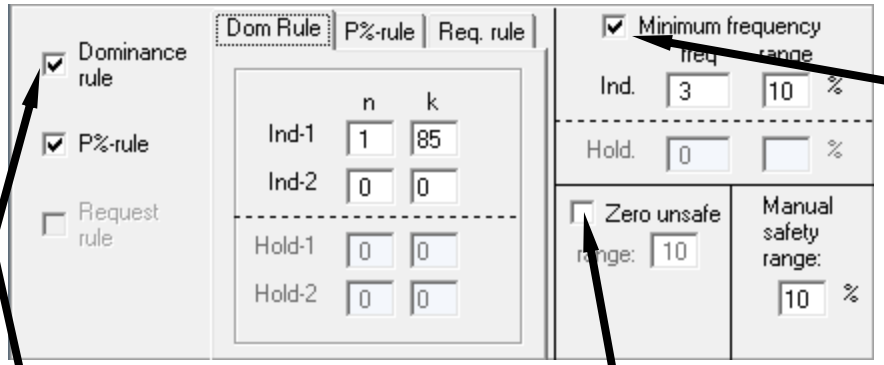
Frequency
Minimization of the number of records concerned by the boxes in secondary secret (number of companies for example).

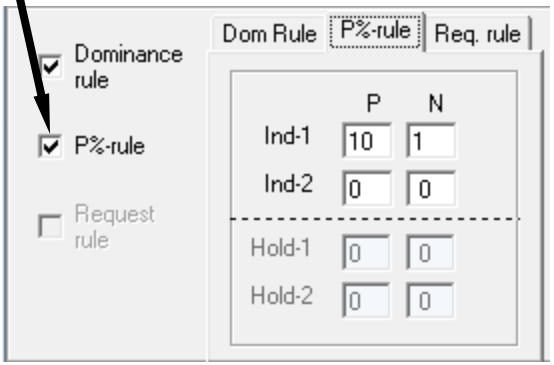
Variable
Minimization of the sum of the variable broken down in the tables (lost information).
Another variable can be chosen to calculate the cost of the loss.

Lambda
Allows to consider the variable x^{λ} , or ln(x) when $\lambda=0$.

Avoiding primary secrecy - dominance, frequency and P%

Rule of dominance or P%
Interval calculated automatically
So as not to be able to retrieve the sensitive individual data (first contributors to the total of the box)

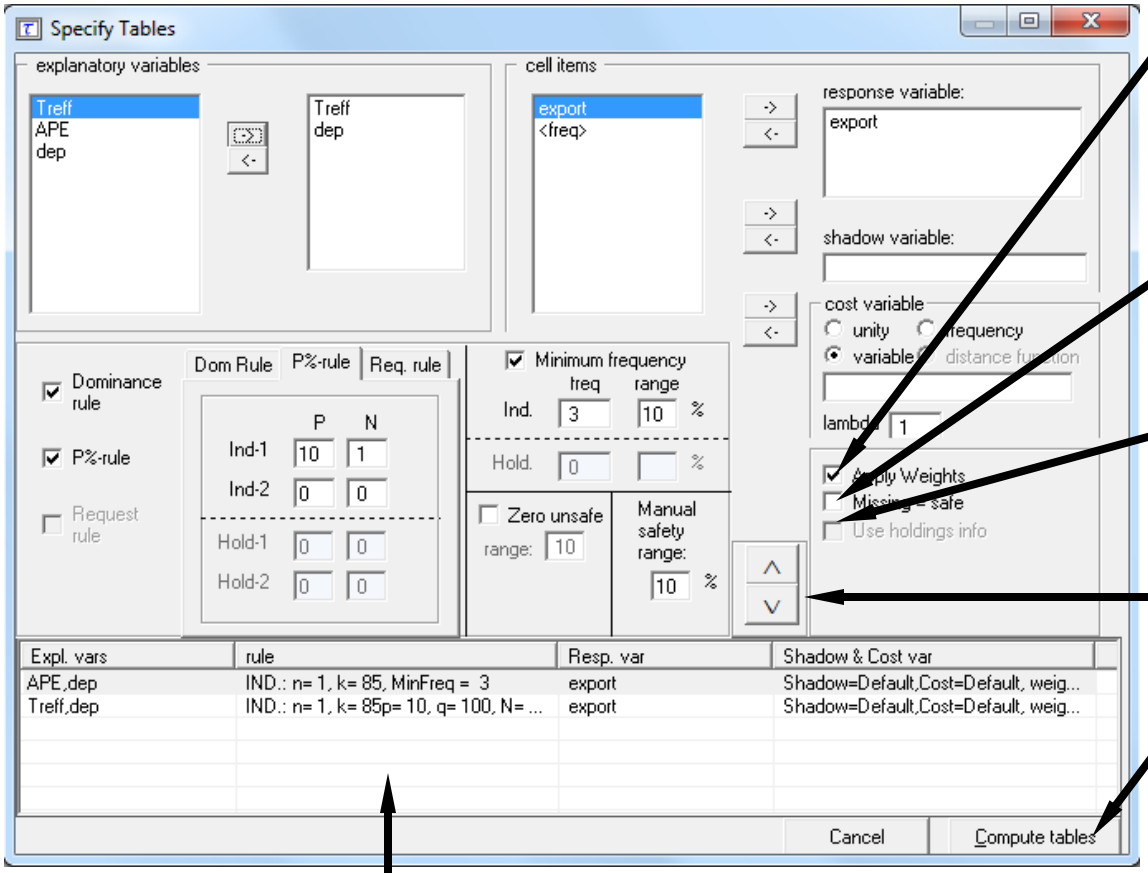




Frequency rule or manually secured box:
- the width of the interval must be defined
- relative to the total of the box
- symmetrical protection interval
- can be defined more precisely by using an *a priori* file

Possibility to hide the boxes equal to 0
Definition of a minimum protection interval
Defined absolutely

Avoiding Primary Secrets - Weight and Holding



Taking weights into account

Missing values considered secure

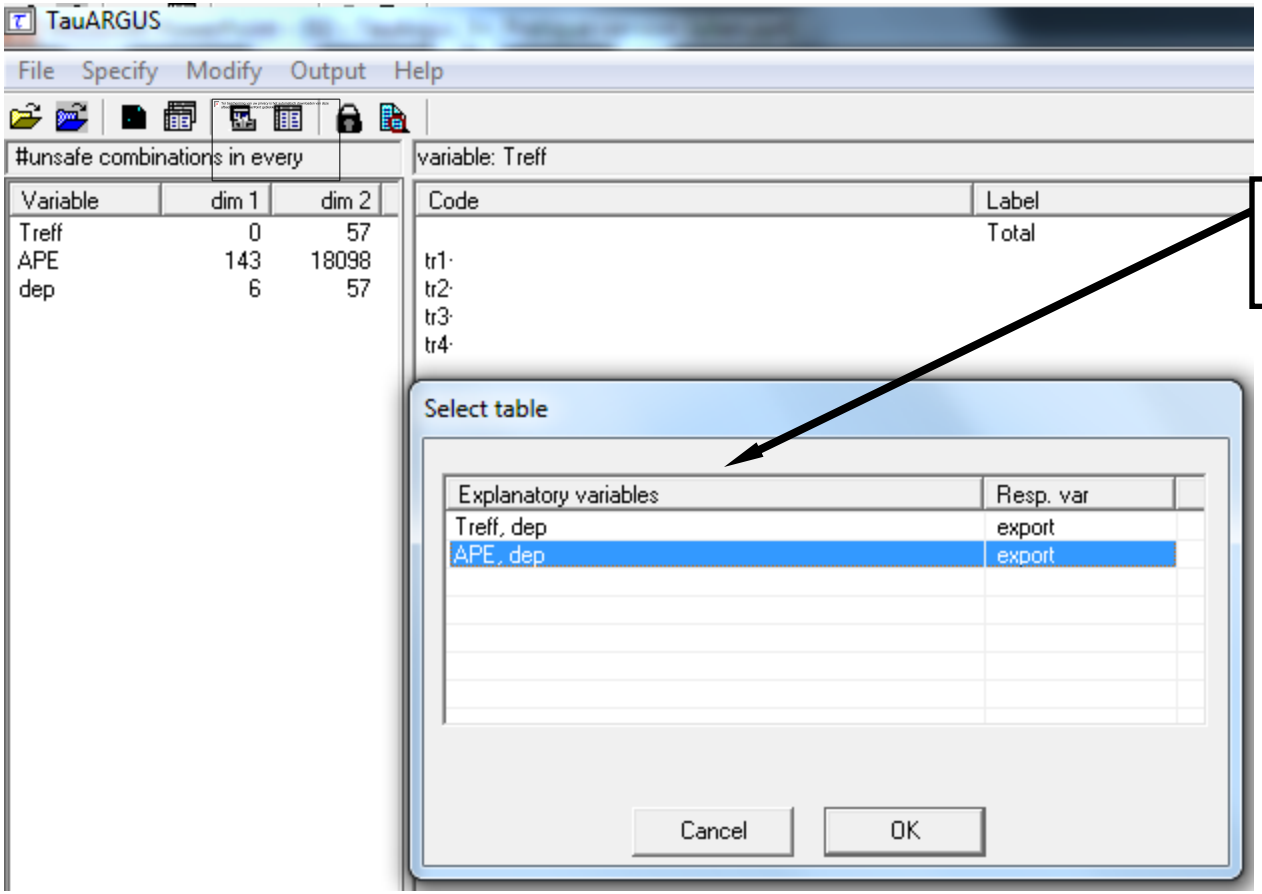
Taking *holding into account*

Tables are dragged

Once all tables are defined, they are generated.

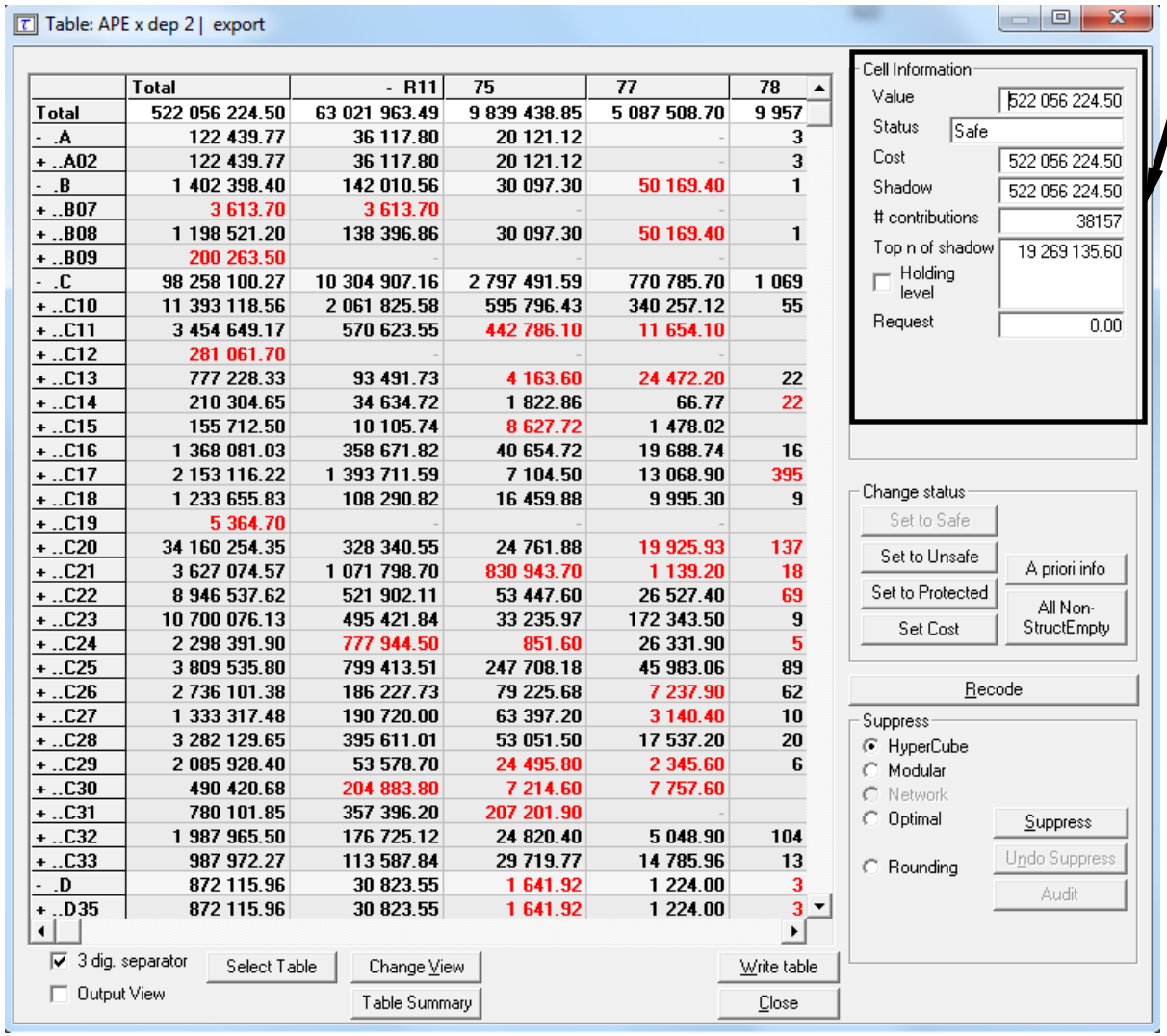
List of tables (maximum 10 at a time)

Calculate the secondary secret - selection of the work table



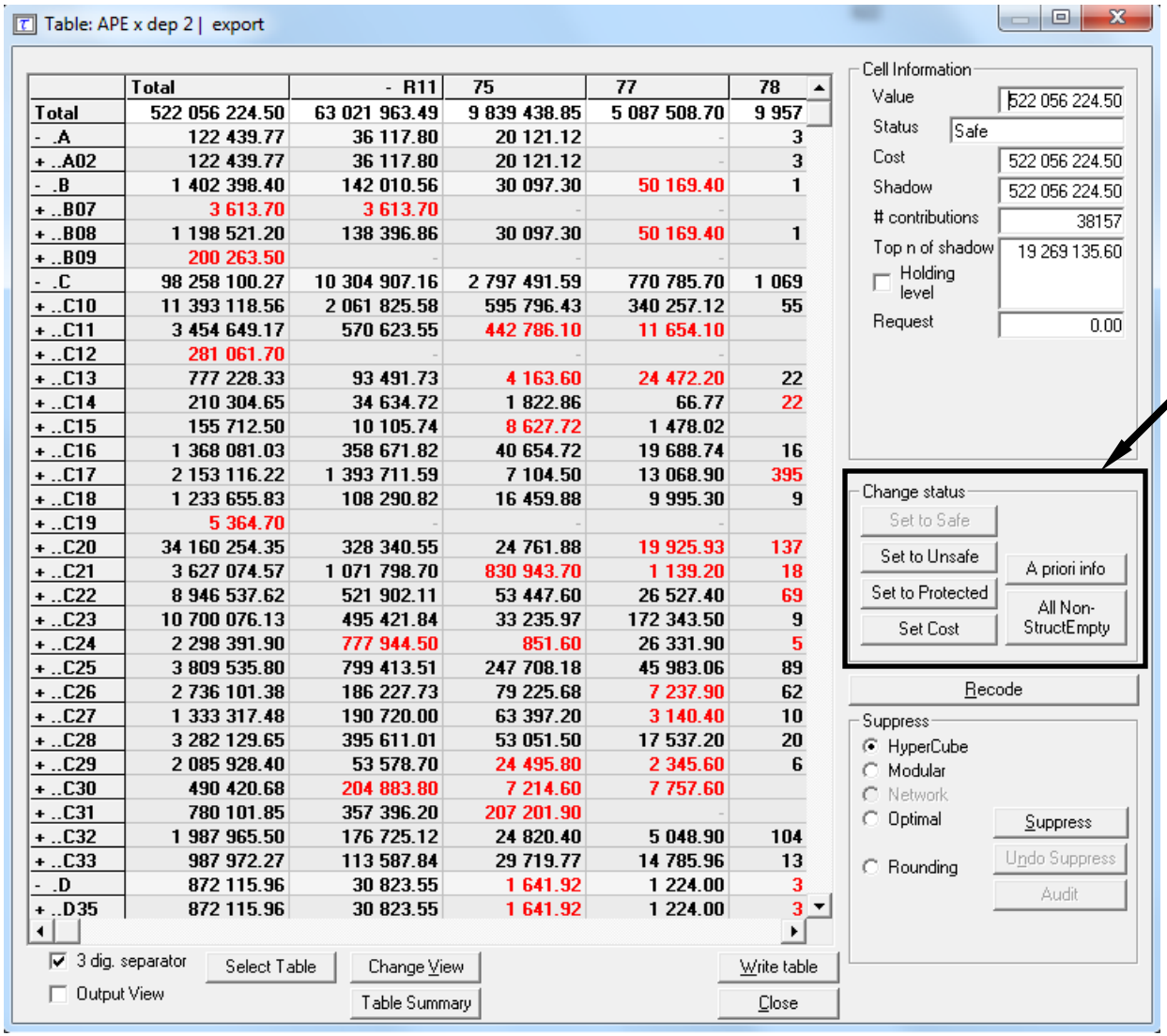
Selection of the table on which you want to work

Calculate the secondary secret - view the table: *cell information*



Information on a given box:
- *status* : *safe/unsafe* (box in black/red)
- *contributions* : number of contributors
- *top n of shadow* : value of the first contributor of the box, to be compared with the value of the box
- *protection interval* : limits of the protection interval defined for the primary secret boxes

Calculate the secondary secret - view the table: *change status*



Changing the status of a box manually:

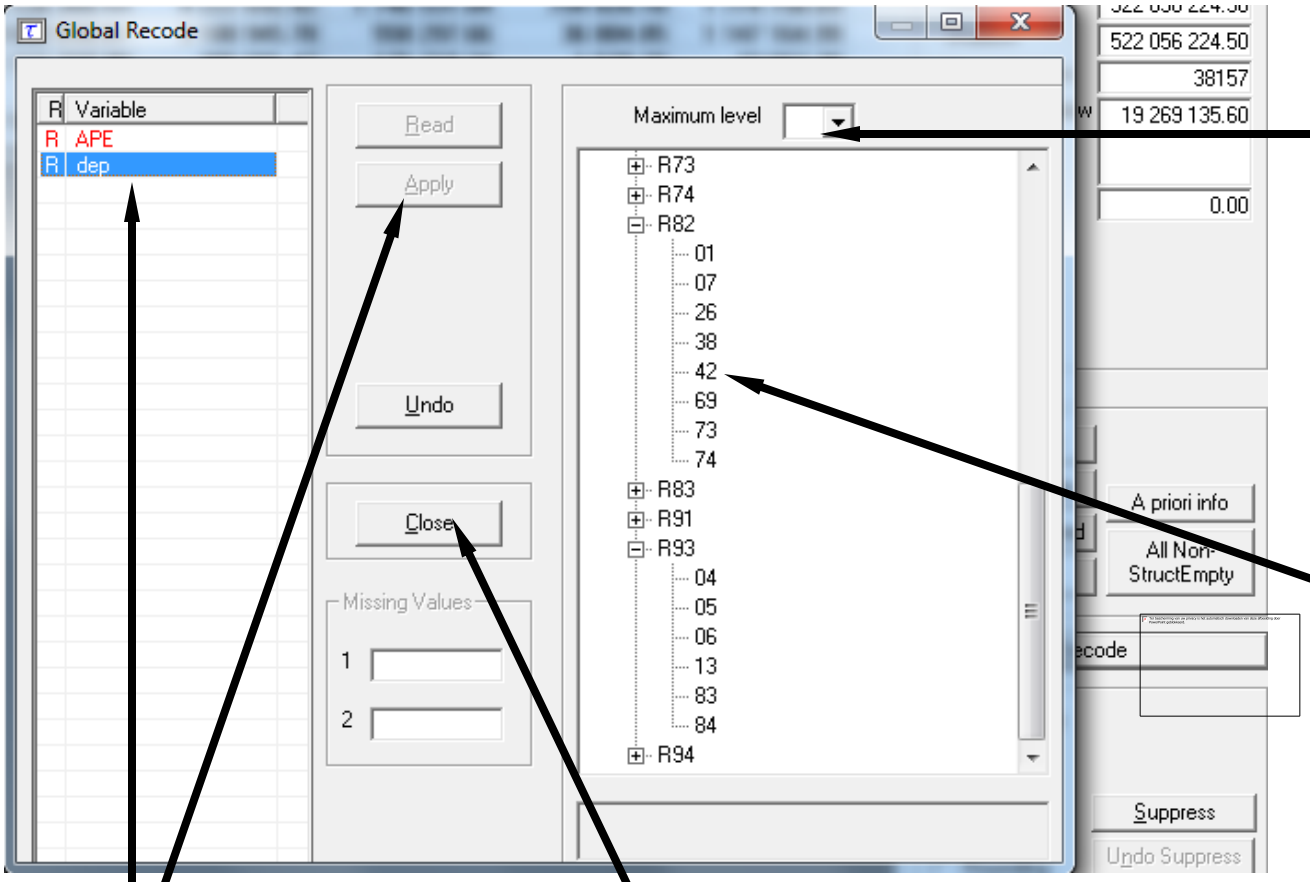
-*set to safe/unsafe* : Force the broadcast / non-broadcast of a box

-- *set to protected* : The box can not be chosen for secondary secrecy (infinite cost)

- *set cost* : Change in cost for secondary secrecy

- *a priori info* : from a flat file (.hst), you change the status of a list of boxes

Calculate the secondary secret - view the table: *recode*

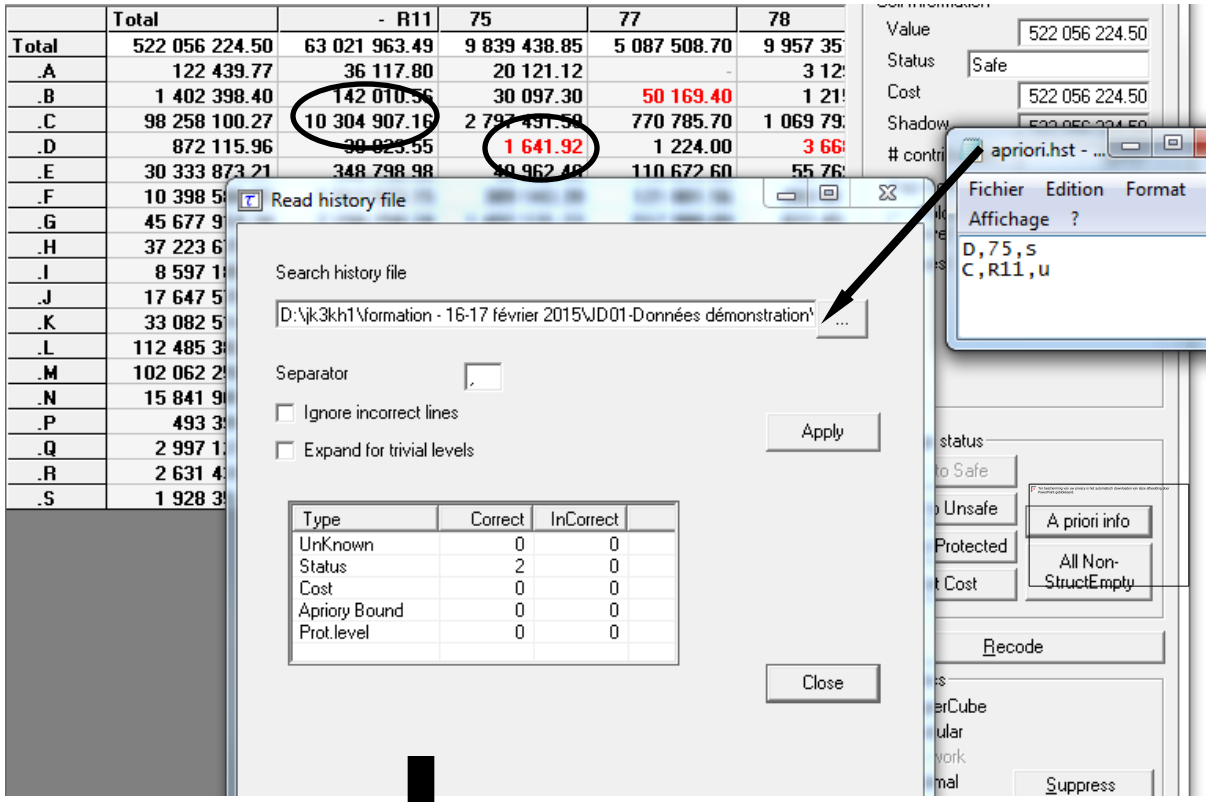


The *maximum level* allows to recode all the modalities at once, making them rise of x notches. Here we go from department to region ...

... But if we wish, for example, to keep the departments of regions 11, 82 and 93, we can modify the structure manually.

First the variable is selected, then recode, then *apply* and then close.

Calculate the secondary secret - view the table: *a priori*

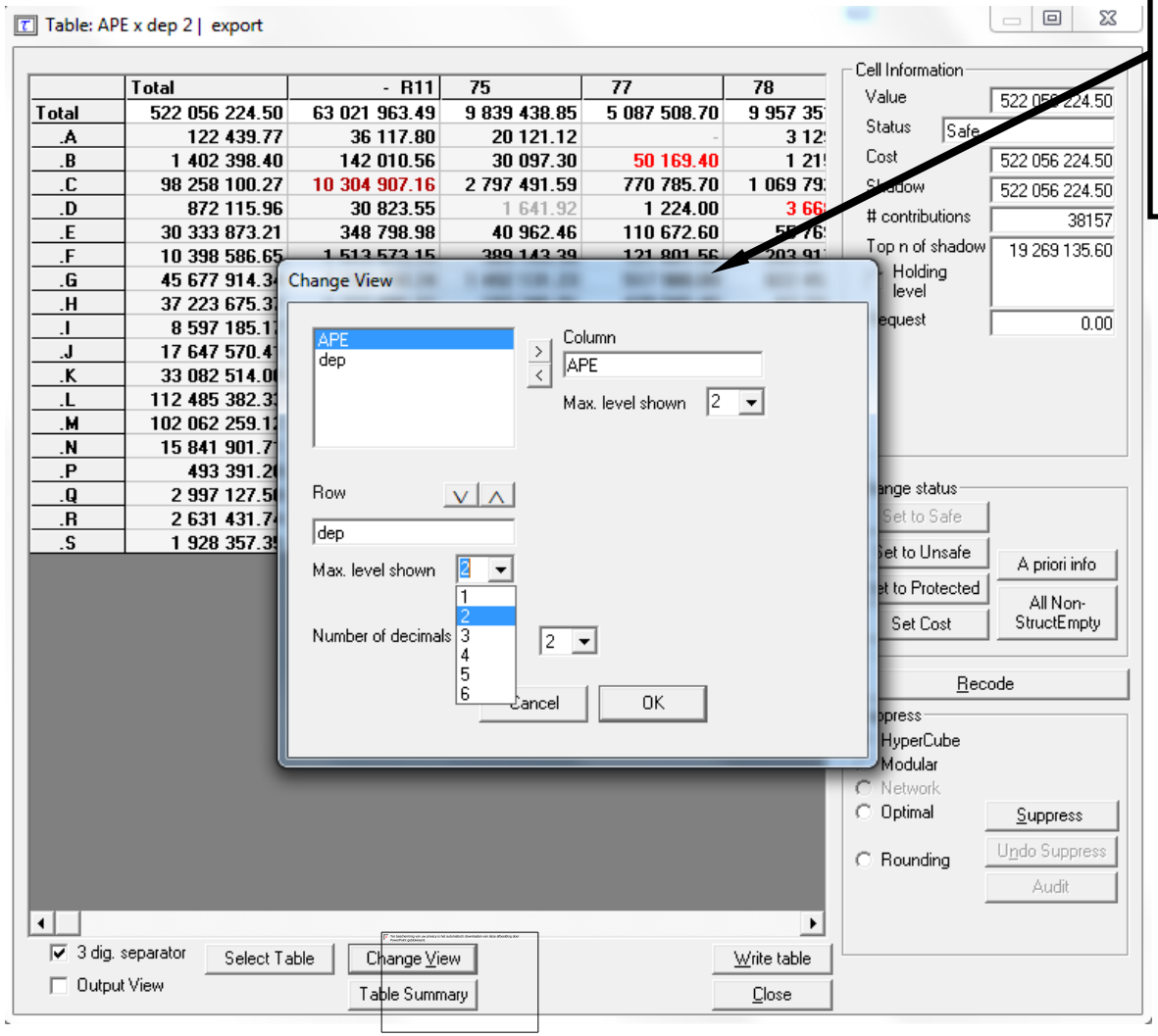


Applying an *a priori* secret structure.
It is not necessary to specify all the boxes in the table.

Be careful not to specify empty boxes.

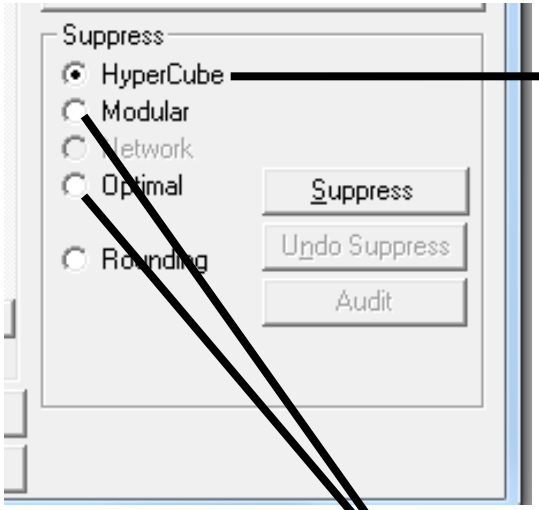
	Total	- R11	75	77	78
Total	522 056 224.50	63 021 963.49	9 839 438.85	5 087 508.70	9 957
.A	122 439.77	36 117.80	20 121.12	-	3
.B	1 402 398.40	142 010.56	30 097.30	50 169.40	1
.C	98 258 100.27	10 304 907.16	2 797 491.59	770 785.70	1 069
.D	872 115.96	30 823.55	1 641.92	1 224.00	3
.E	30 333 873.21	348 798.98	40 962.46	110 672.60	55

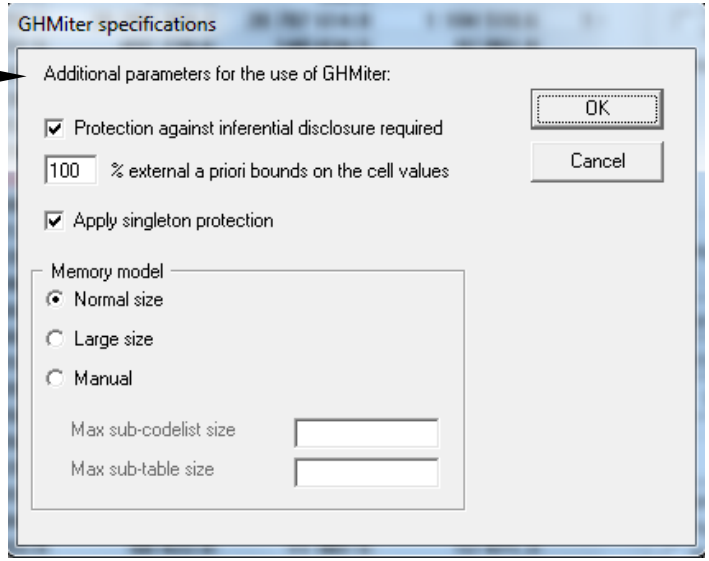
Calculate secondary secret - view table: *change view*

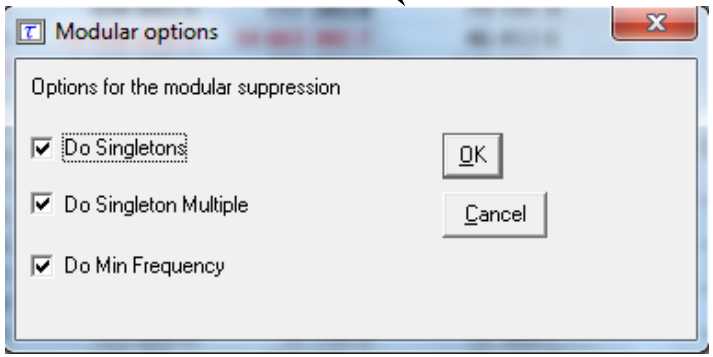


Changing the display of the table on the screen (this does not modify the actual structure of the table).

Calculate Secondary Secret - What Secondary Secret Method?







Hypercube

Sequential treatment of cells under primary secret (search for the least expensive suppression scheme).
Free and fast method.

Optimal

Overall minimization of the cost of hiding boxes
Under the constraint of respecting the protection intervals.
Method sometimes very slow.

Modular

For hierarchical tables, the problem is split: the optimal method is applied on the different sub-tables.
Faster to process than the optimal method.

Calculating the secondary secret – managing singletons problems in Tau-Argus

Singleton

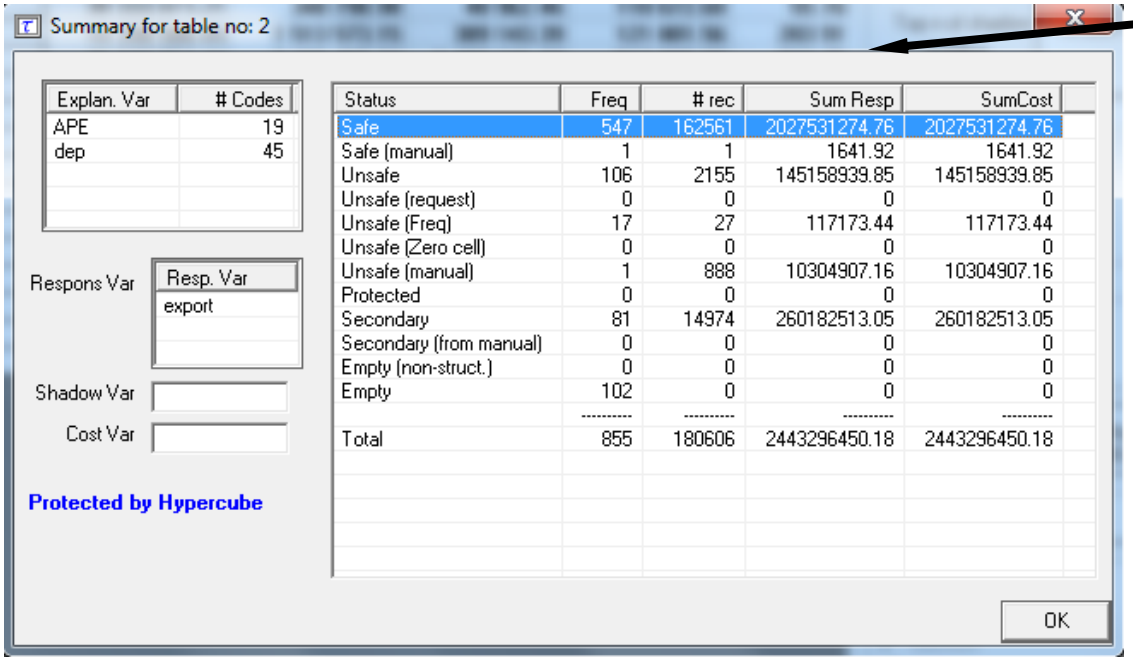
➔ Cell with one contributor, he can know that is the only contributor to the box.

➔ Complementary secondary suppression if :

- two singletons in the same row or column
- un singleton et une cellule cachée en secret primaire
- two cells hidden due to a frequency problem whose sum of the number of contributors is still insufficient (this problem does not arise with the rule of three units)

Method	Singleton management
Hypercube	Yes (automatically)
Modular	Yes (optional)
Network	Not offered
Optimal	Yes (optional)

Calculate the secondary secret - view the table: *summary*



The most interesting information :

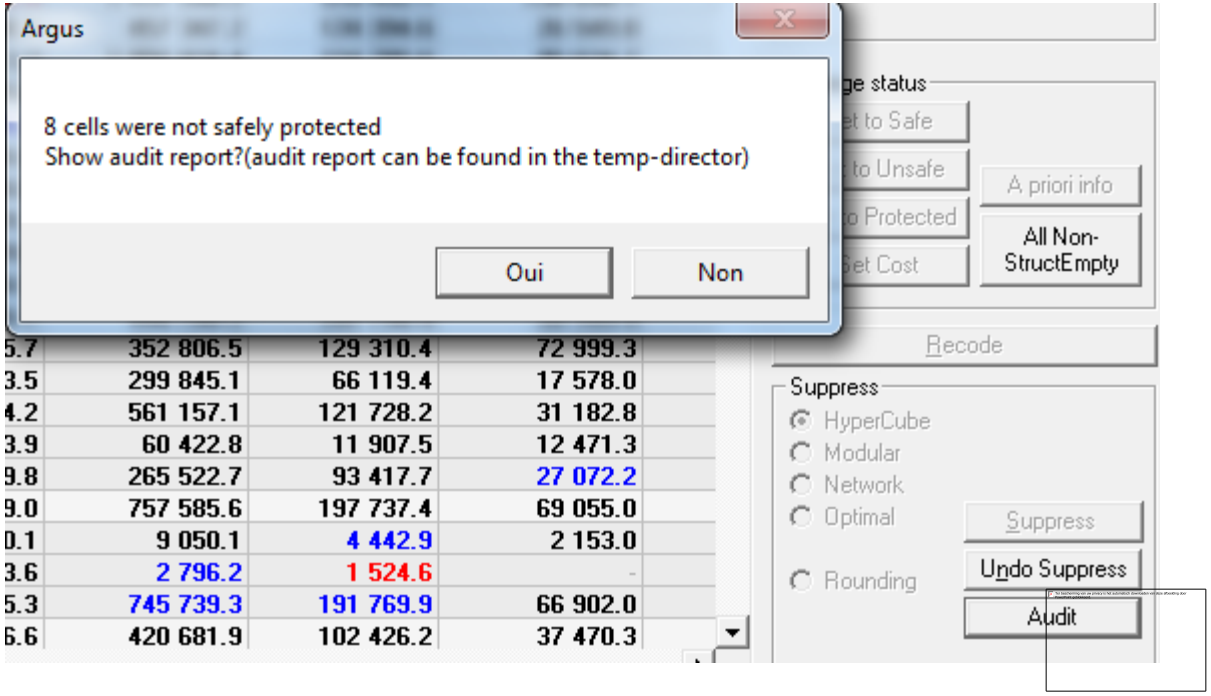
- the primary secret of frequency : *Unsafe(freq)*
- the primary secret of dominance */P%* : *Unsafe*
- manually protected boxes : *Safe(manual)*
- Secondary secrecy : *Secondary*

- The costs of these different categories of secrecy.

Calculate Secondary Secrecy - Different Methods for Different Results

Method	Number of boxes deleted in secondary secret	Cost of boxes deleted	Required calculation time
Hypercube **(free method)**	81 (9.47%)	260 182 513 (10.64%)	1 second
Modular **(paying method)**	64 (7.48%)	26 017 917 (1.06%)	1 minute and 4 seconds
Optimal **(paying method)**	140 (16.37%)	25 461 484 (1.04%)	10 minutes (not "optimal" solution)

Audit


Argus
8 cells were not safely protected
Show audit report?(audit report can be found in the temp-director)
Oui
Non
ge status
et to Safe
to Unsafe
o Protected
Set Cost
A priori info
All Non-StructEmpty
Recode
Suppress
HyperCube
Modular
Network
Optimal
Rounding
Suppress
Undo Suppress
Audit
5.7 352 806.5 129 310.4 72 999.3
3.5 299 845.1 66 119.4 17 578.0
4.2 561 157.1 121 728.2 31 182.8
3.9 60 422.8 11 907.5 12 471.3
9.8 265 522.7 93 417.7 27 072.2
9.0 757 585.6 197 737.4 69 055.0
0.1 9 050.1 4 442.9 2 153.0
3.6 2 796.2 1 524.6 -
5.3 745 739.3 191 769.9 66 902.0
6.6 420 681.9 102 426.2 37 470.3

Audit - Protection Intervals and Audit Intervals

τ-ARGUS Audit-report

Table created date: 09-juin-2015, time 15:25:03

Original file:	*D:\jk3kh1\formation - 16-17 février 2015\JD01-Données démonstration\Donnees_demo.asc*
Meta file:	*D:\jk3kh1\formation - 16-17 février 2015\JD01-Données démonstration\Donnees_demo.rda*
Table file:	

Table generated from microdata

Table structure

Var	Function	# codes
Response var:	*export*	
Explanatory var 1:	*Activité*	*585*
Explanatory var 2:	*dep*	*45*
Shadow variable :	*default (=response var)*	
Cost variable	*default (=response var)*	
Weight variable :	*Poids*	

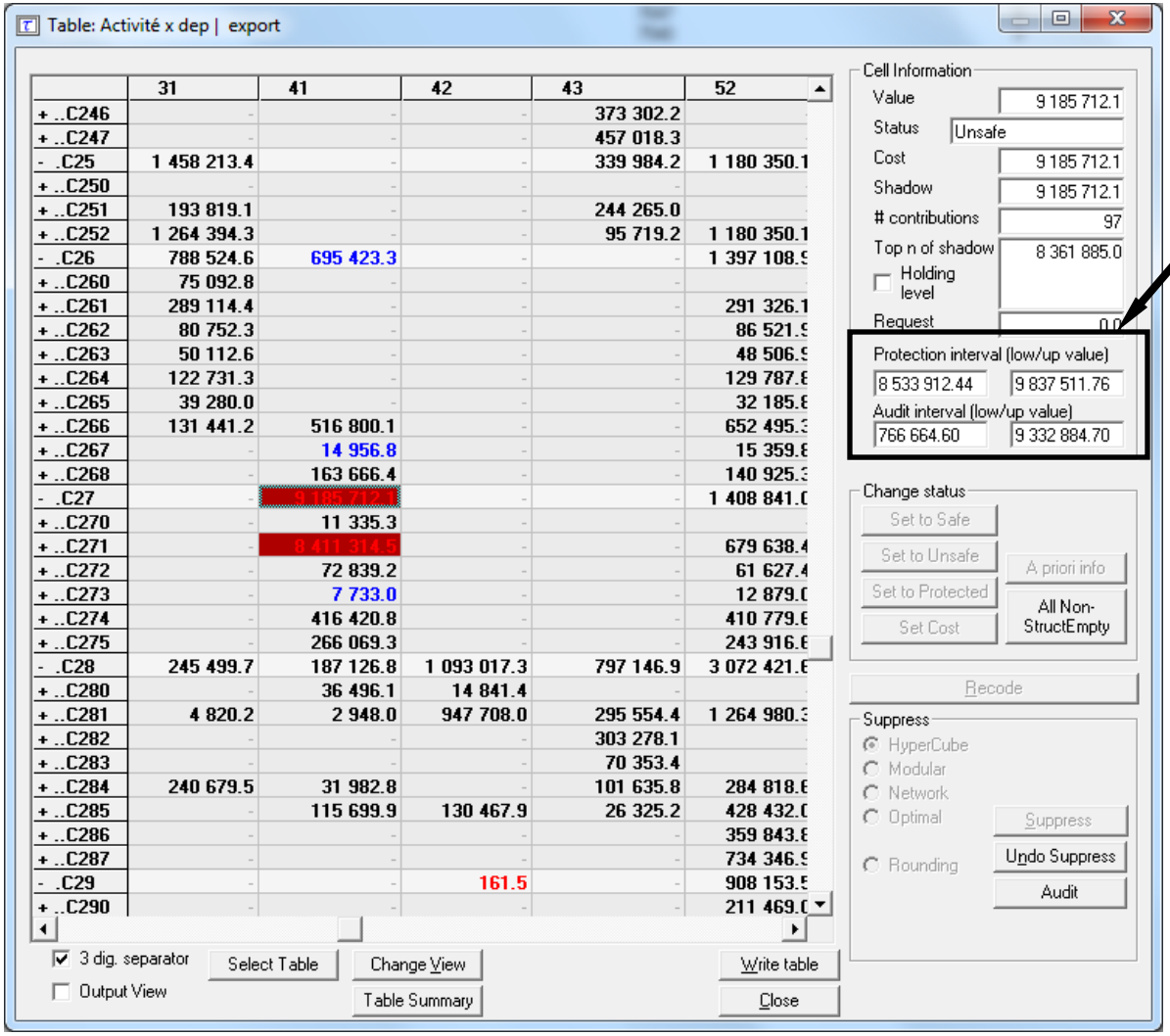
Overview of not-properly protected cells

Activité	*dep*	*Cell value*	*Lower PI*	*Upper PI*	*Exact*
C27	*41*	*9 185 712.1*	*766 664.6*	*9 332 884.7*	
C271	*41*	*8 411 314.5*	*0.0*	*8 566 220.1*	
C2710	*41*	*8 411 314.5*	*0.0*	*8 566 220.1*	
C27100	*41*	*8 411 314.5*	*0.0*	*8 566 220.1*	
C33	*95*	*1 406 786.8*	*74 360.1*	*1 414 023.3*	
C332	*95*	*1 326 101.5*	*0.0*	*1 339 663.2*	
C3320	*95*	*1 326 101.5*	*0.0*	*1 339 663.2*	
C33200	*95*	*1 326 101.5*	*0.0*	*1 339 663.2*	

8 cells were not safely protected

τ-ARGUS version: 3.5.0 ; build: (27)

Audit - Protection Intervals and Audit Intervals



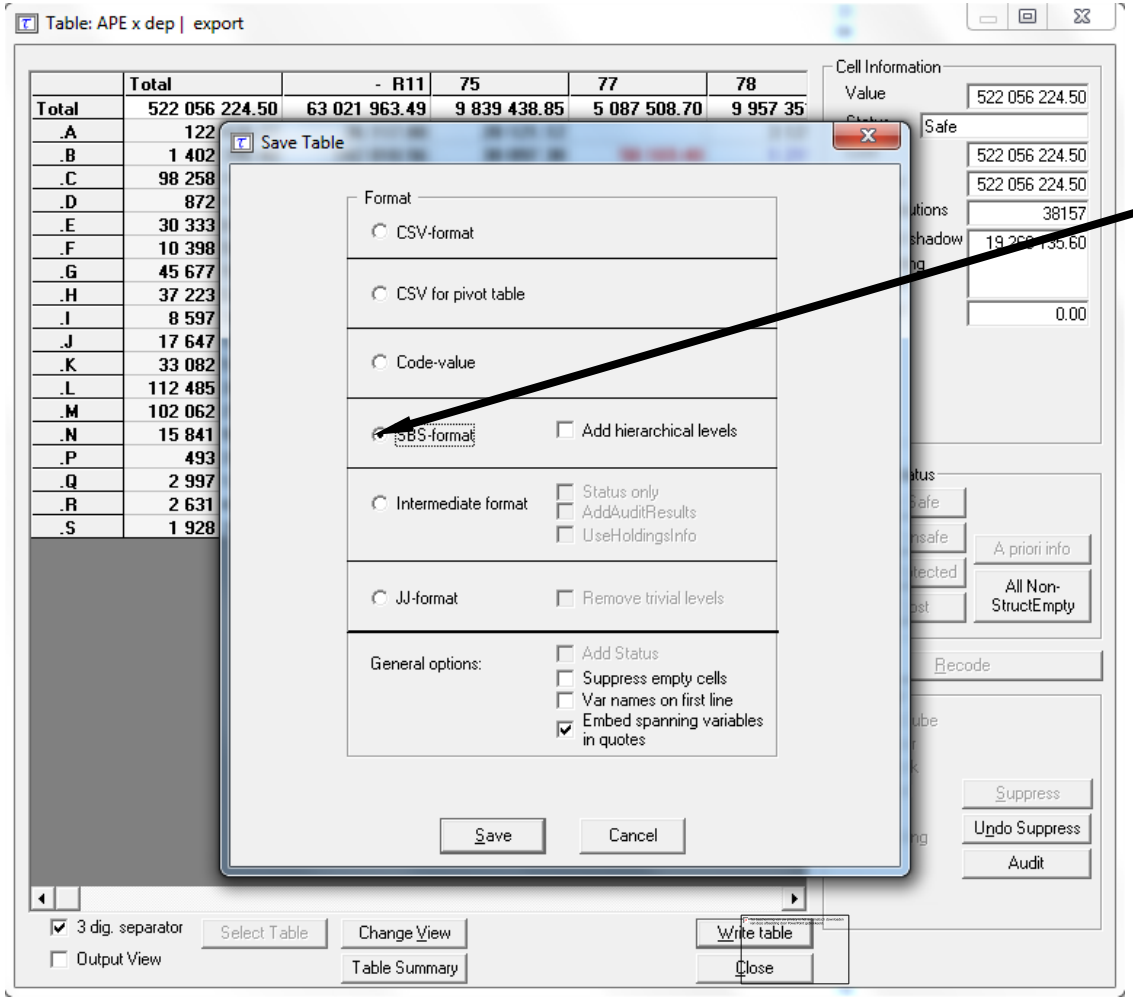
Protection interval
Limits of the protection interval defined for the primary secret boxes

Audit interval
Interval of the possibilities that a user is likely to find thanks to the additivity and positivity relationships of the cells.

Outline

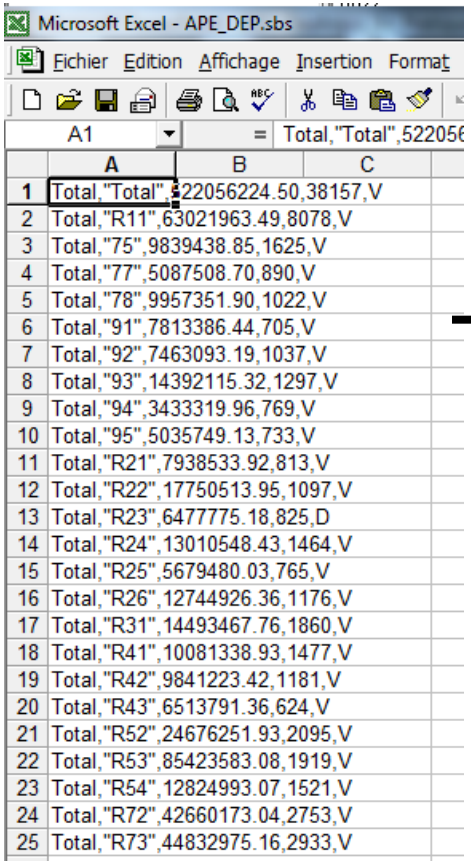
1. Creating Input Data
   - Microdata
   - Metadata
2. From primary to secondary suppression
   - Specifying tables
   - Avoiding Primary suppression
   - Computing secondary suppression
   - Audit
3. Outputs
   - Format SBS
   - Excel Formatting
4. Three special cases
   - Linked tables management
   - Controlled rounding
   - Tabulated data

Exporting your data: different options

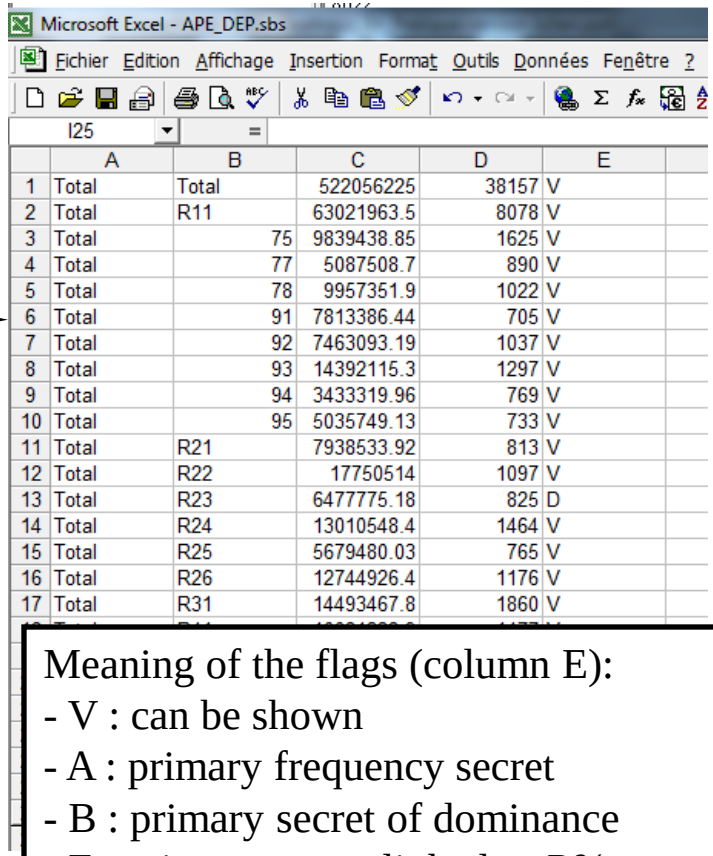


Format requested by Eurostat, which is systematically used in methodology department

Export to the format requested by Eurostat (.sbs): Formatting



➔ selection of column A
➔menu
➔ *data*
➔ *convert*
➔ *next*
➔ tick « , »
➔ *complete*

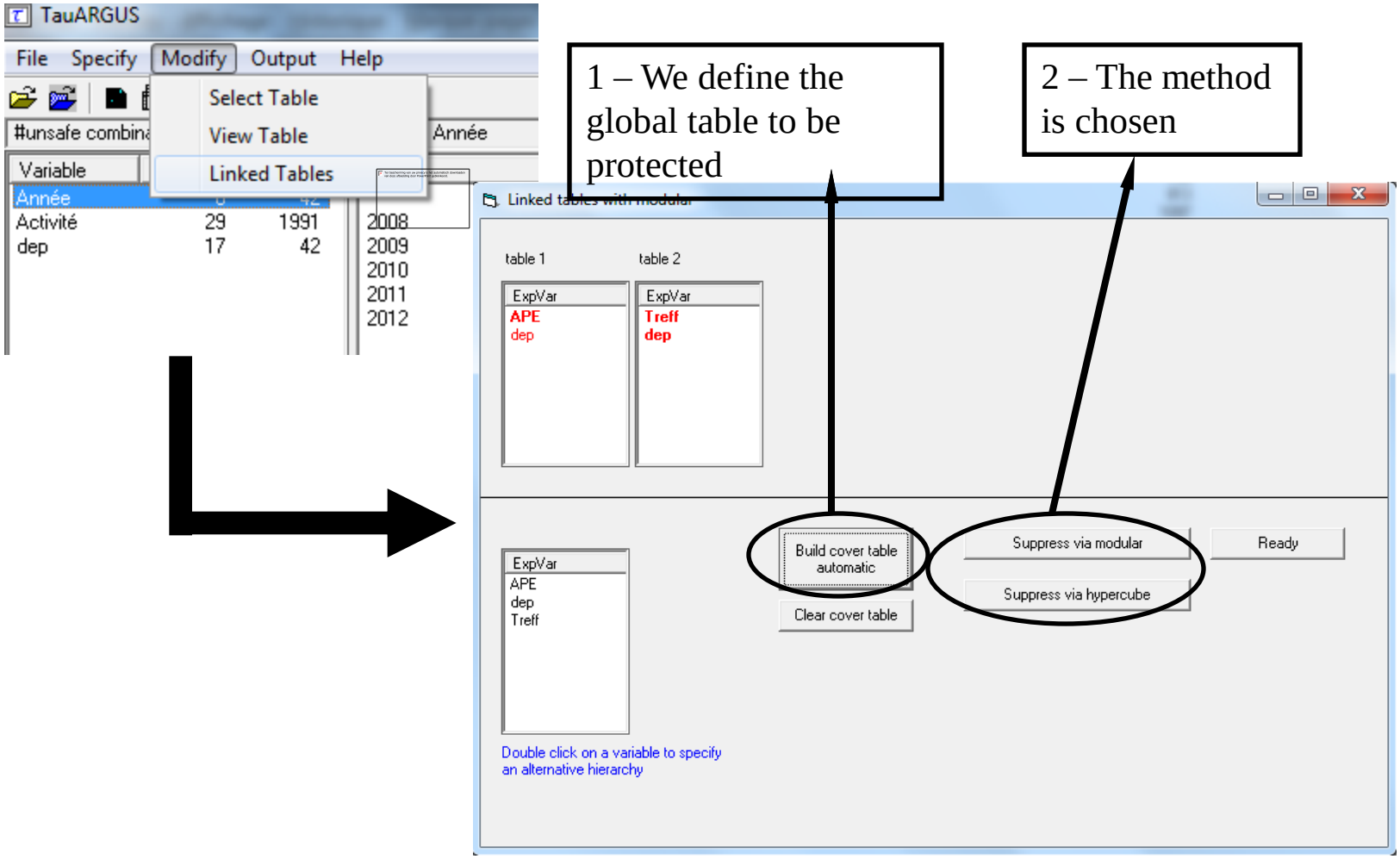


Meaning of the flags (column E):
- V : can be shown
- A : primary frequency secret
- B : primary secret of dominance
- F : primary secret linked to *P%*
- D : secondary secret

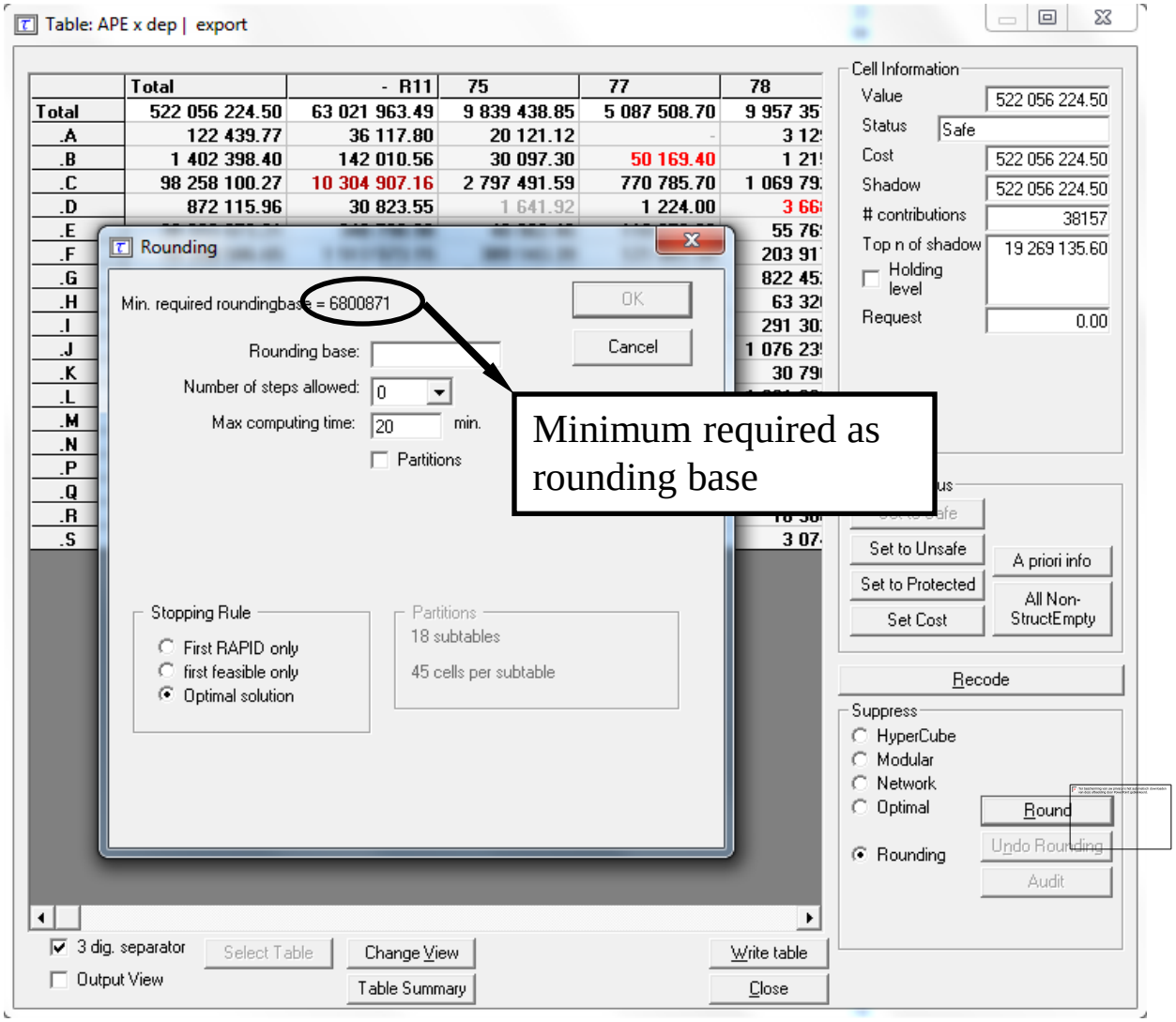
Outline

1. Creating Input Data
   - Microdata
   - Metadata
2. From primary to secondary suppression
   - Specifying tables
   - Avoiding Primary suppression
   - Computing secondary suppression
   - Audit
3. Outputs
   - Format SBS
   - Excel Formatting
4. Three special cases
   - Linked tables management
   - Controlled rounding
   - Tabulated data

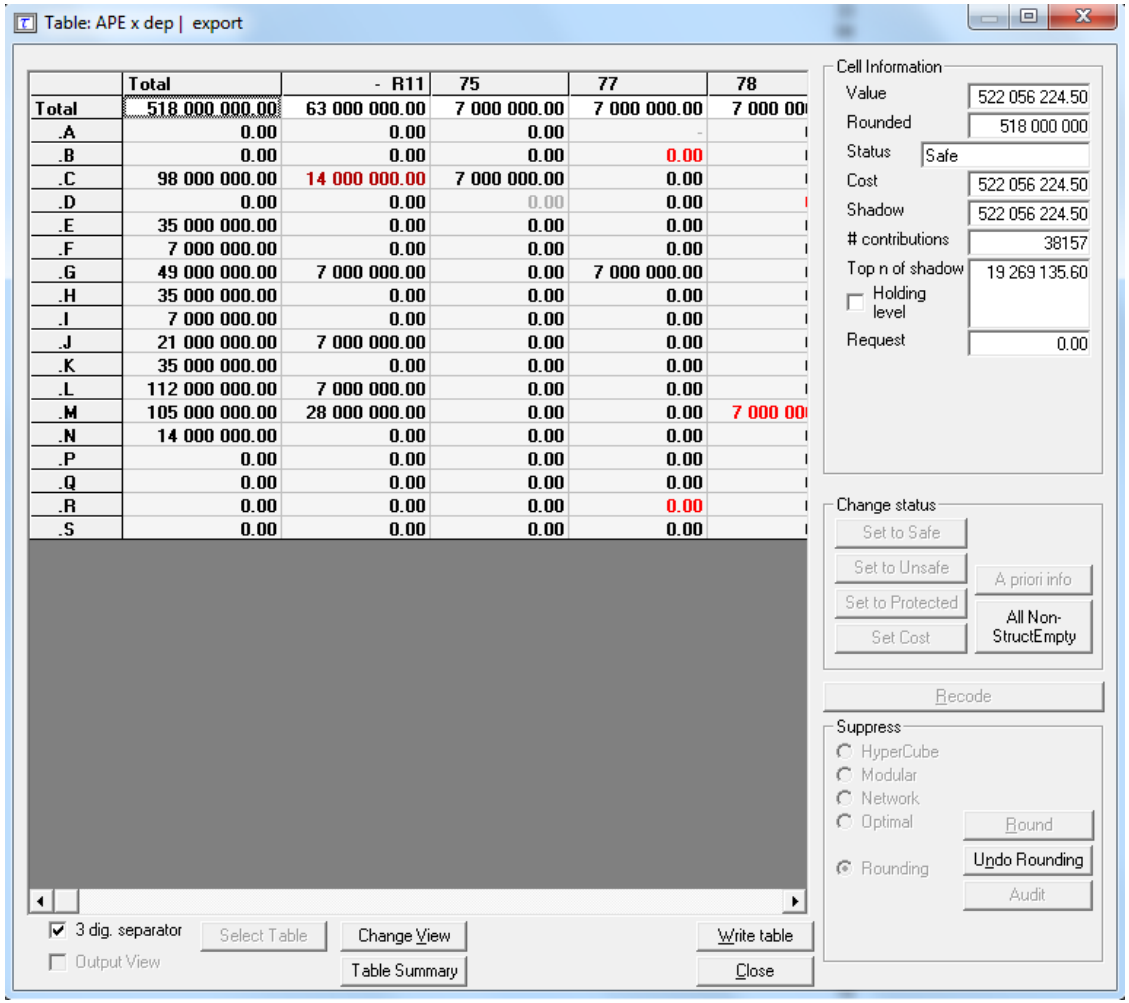
Managing linked tables



Controlled rounding - another way to manage secrecy



Controlled rounding - another way to manage secrecy

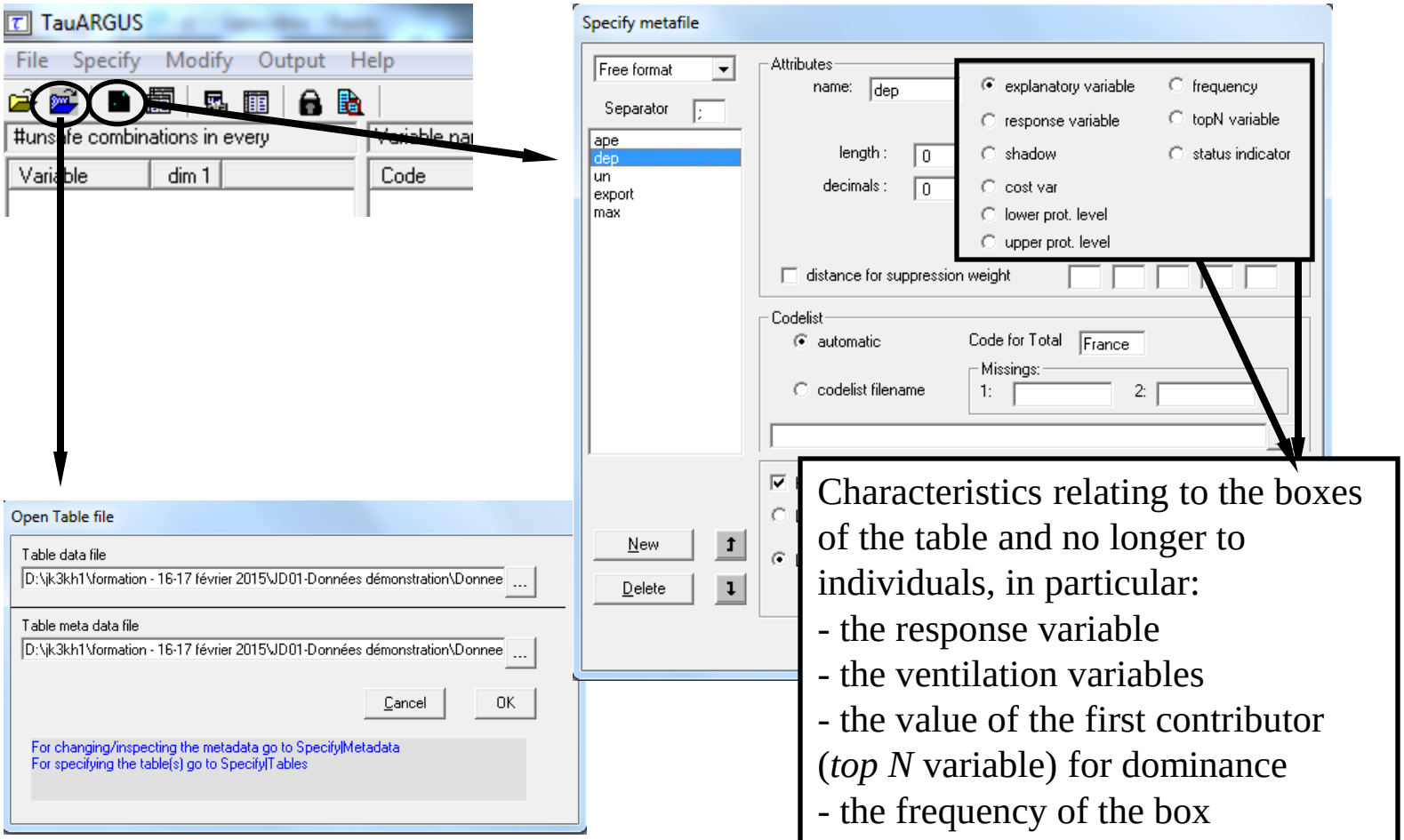


A method sometimes brutal ...

Tabulated data - construction of a flat file of tabulated data (.tab)

```
/* Étape 1 : création de la tabulation */
/* Au préalable, on crée un variable de comptage "un" pour que la fréquence sous Tau-Argus prenne en
compte le poids */

proc means data =  demo.Donnees_demo n sum max noprint ;
            var export ;
            class APE dep ;
            output out= ape_dep n=un sum = export max = max_export ;
run;


data ape_dep ;
            set ape_dep ;
            if ape="" then ape = "Total";
            if dep="" then dep = "Total";
run;


/* Étape 2 : création du fichier plat */

data _null_;
File "Z:\19-20 mai 2016\02-TauArgus_En_Pratique-mai 2016\Données
démonstration\donnees_demo_tabulees.tab" dsd dlm=';' ;
Set  ape_dep (keep=ape dep un export max_export);
Put ( _all_)(+0);
run;
```

Tabulated data - metadata

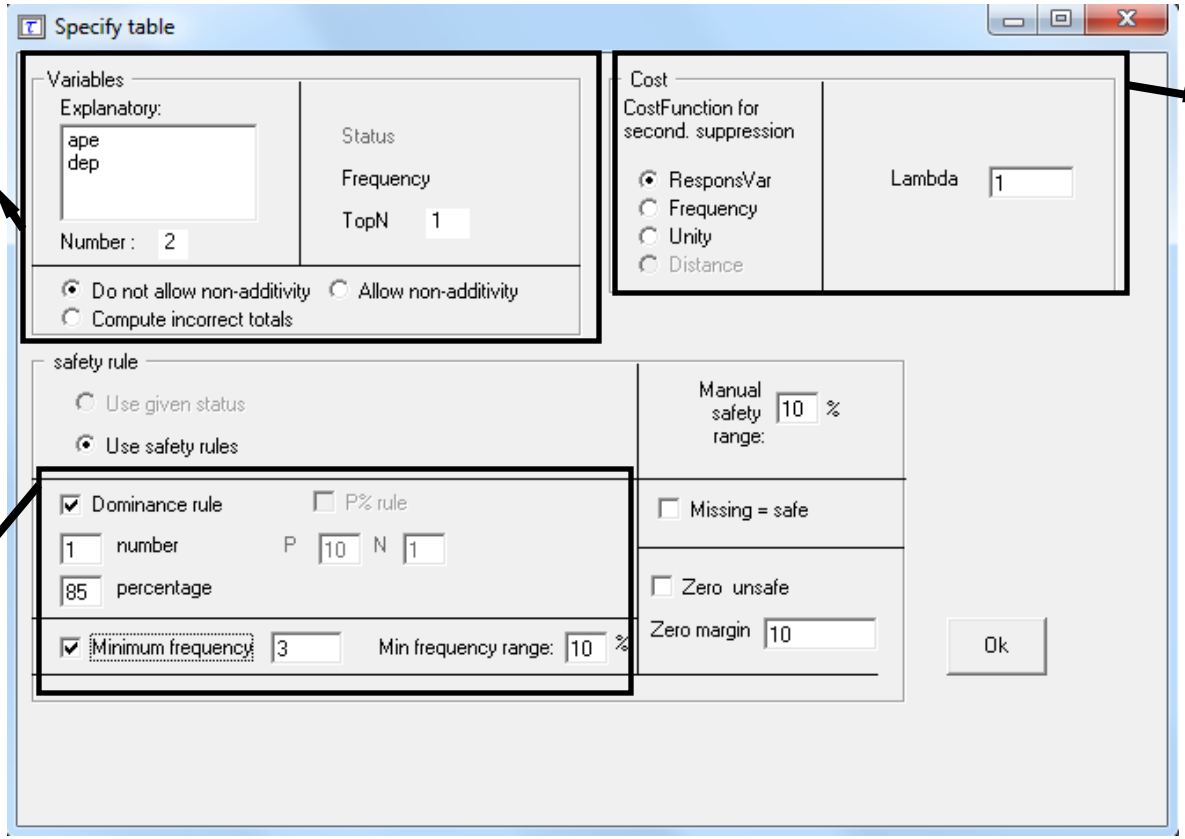


Characteristics relating to the boxes of the table and no longer to individuals, in particular:

- the response variable
- the ventilation variables
- the value of the first contributor (*top N* variable) for dominance
- the frequency of the box (*frequency*)

Tabulated data - table specification

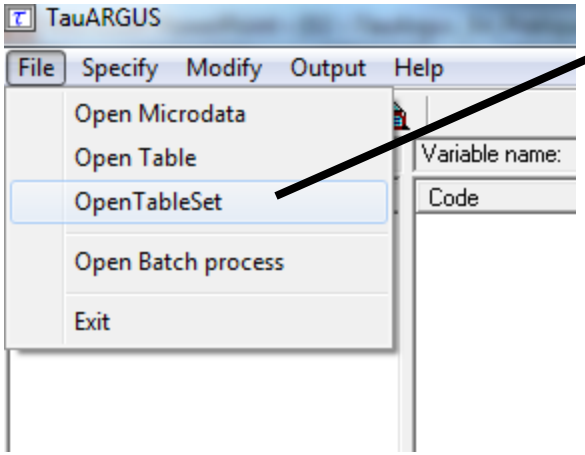
1 - Variables selected for the table

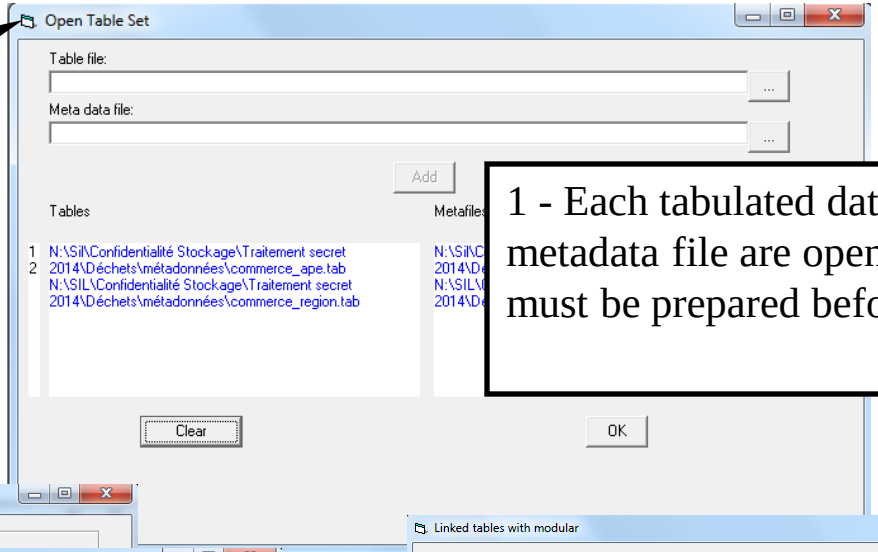


2 - Cost management for secondary secrecy

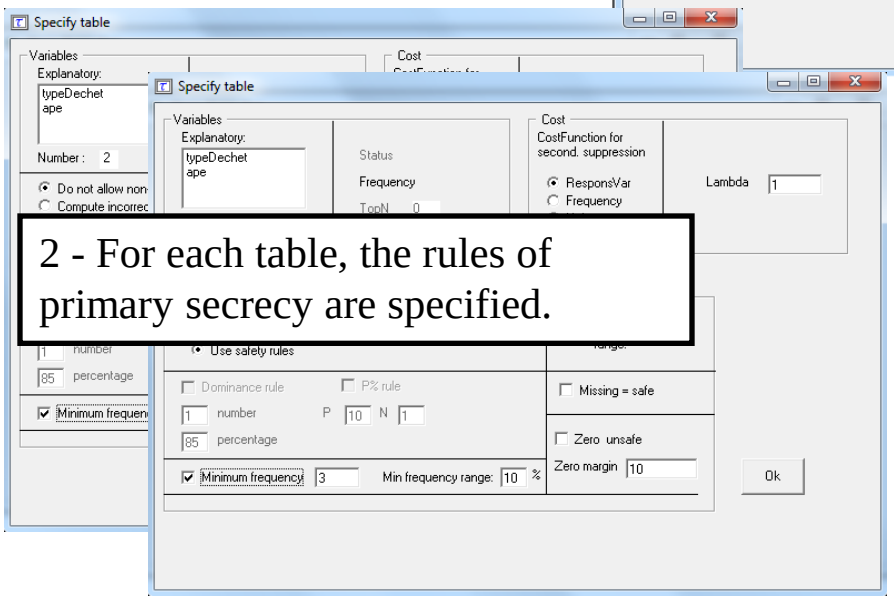
3 - Rules of primary secrecy

Tabulated data - linked tables

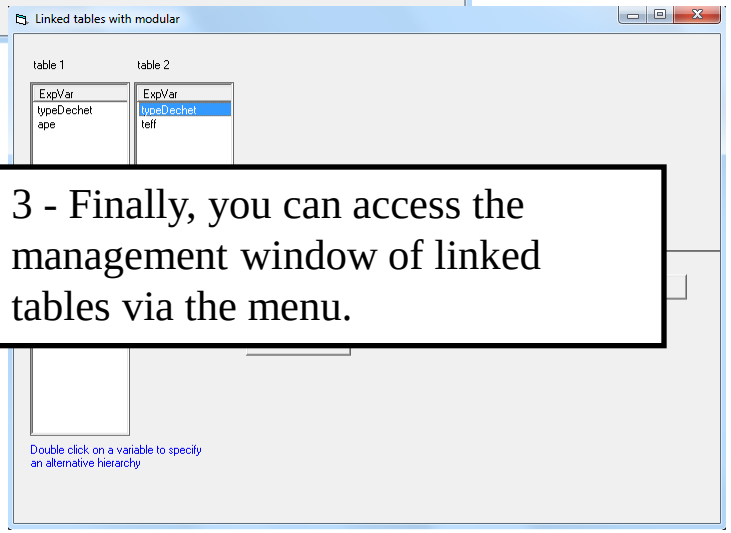




1 - Each tabulated data file and its metadata file are opened, which must be prepared beforehand.



2 - For each table, the rules of primary secrecy are specified.



3 - Finally, you can access the management window of linked tables via the menu.

Confidentiality Management in Aggregate Data Tables

Thanks for your attention ☺

Insee
18 bd Adolphe-Pinard
75675 Paris Cedex 14
www.insee.fr
Informations statistiques :
www.insee.fr / Contacter l'Insee
09 72 72 4000
(coût d'un appel local)
du lundi au vendredi de 9h00 à 17h00

INSEE