

Gestion de la confidentialité dans les tableaux de données agrégées

τ -Argus

Alexandre Awad et Maxime Beauté
Département des méthodes statistiques



Plan de présentation

1. Création des données en entrée du logiciel
 - Microdonnées
 - Métadonnées
2. Du secret primaire au secret secondaire
 - Spécification des tables
 - Éviter le secret primaire
 - Calculer le secret secondaire
 - Gestion des tableaux liés
3. Sorties proposées
 - Format SBS
 - Mise en forme Excel
4. Utilisation de données tabulées en entrée du logiciel

Plan de présentation

1. Création des données en entrée du logiciel
 - Microdonnées
 - Métadonnées
2. Du secret primaire au secret secondaire
 - Spécification des tables
 - Éviter le secret primaire
 - Calculer le secret secondaire
 - Gestion des tableaux liés
3. Sorties proposées
 - Format SBS
 - Mise en forme Excel
4. Utilisation de données tabulées en entrée du logiciel

Microdonnées – on part d'une table SAS

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

Types de variable en entrée :

- variables de ventilation
- variable de poids
- variables de réponse
- variable *holding* (non présentée ici)

Microdonnées – conversion d'une table SAS en fichier plat (.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);
```

```
    exporttt   =      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
```

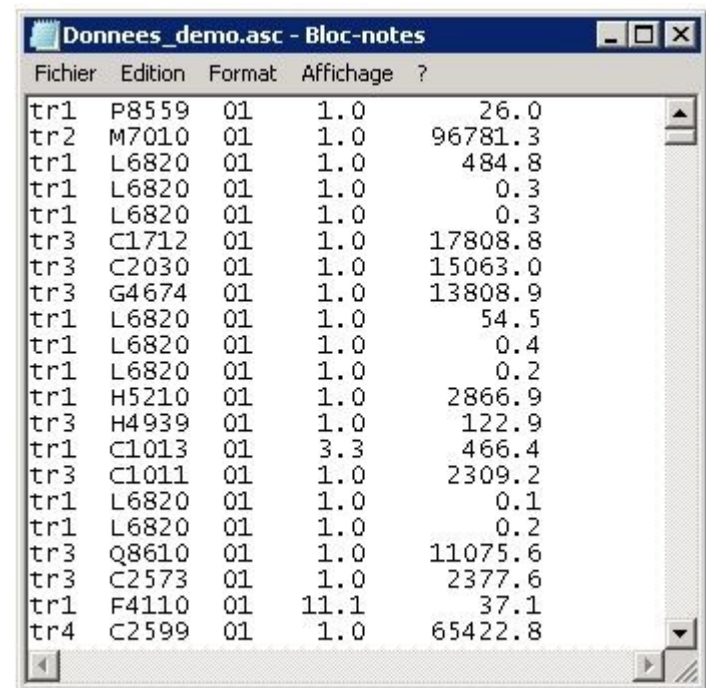
```
        APE            6-11
```

```
        dep            13-14
```

```
        poids          16-21
```

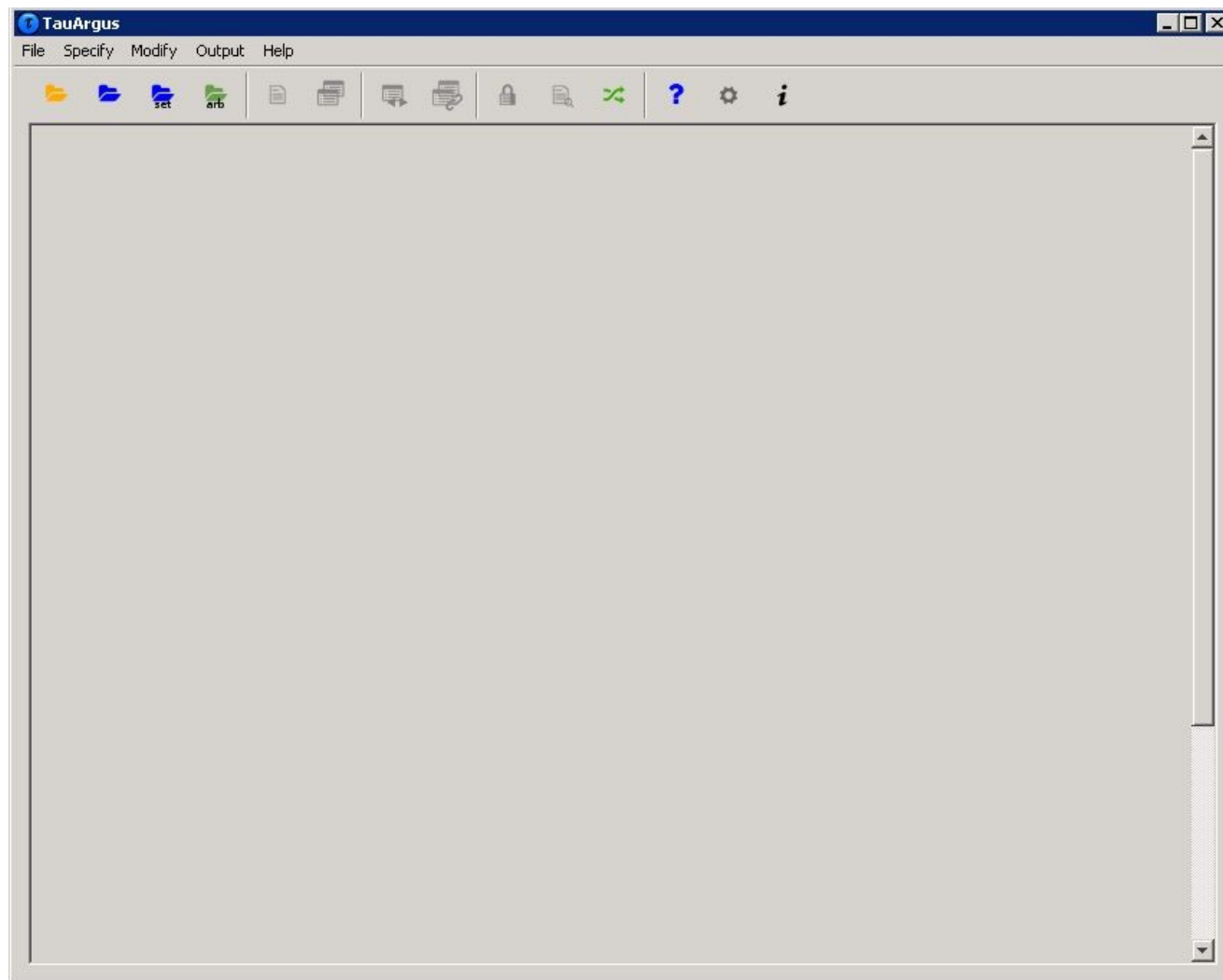
```
        exporttt       23-32 ;
```

```
run;
```



tr1	P8559	01	1.0	26.0
tr2	M7010	01	1.0	96781.3
tr1	L6820	01	1.0	484.8
tr1	L6820	01	1.0	0.3
tr1	L6820	01	1.0	0.3
tr3	C1712	01	1.0	17808.8
tr3	C2030	01	1.0	15063.0
tr3	G4674	01	1.0	13808.9
tr1	L6820	01	1.0	54.5
tr1	L6820	01	1.0	0.4
tr1	L6820	01	1.0	0.2
tr1	H5210	01	1.0	2866.9
tr3	H4939	01	1.0	122.9
tr1	C1013	01	3.3	466.4
tr3	C1011	01	1.0	2309.2
tr1	L6820	01	1.0	0.1
tr1	L6820	01	1.0	0.2
tr3	Q8610	01	1.0	11075.6
tr3	C2573	01	1.0	2377.6
tr1	F4110	01	11.1	37.1
tr4	C2599	01	1.0	65422.8

Ouverture de Tau-Argus



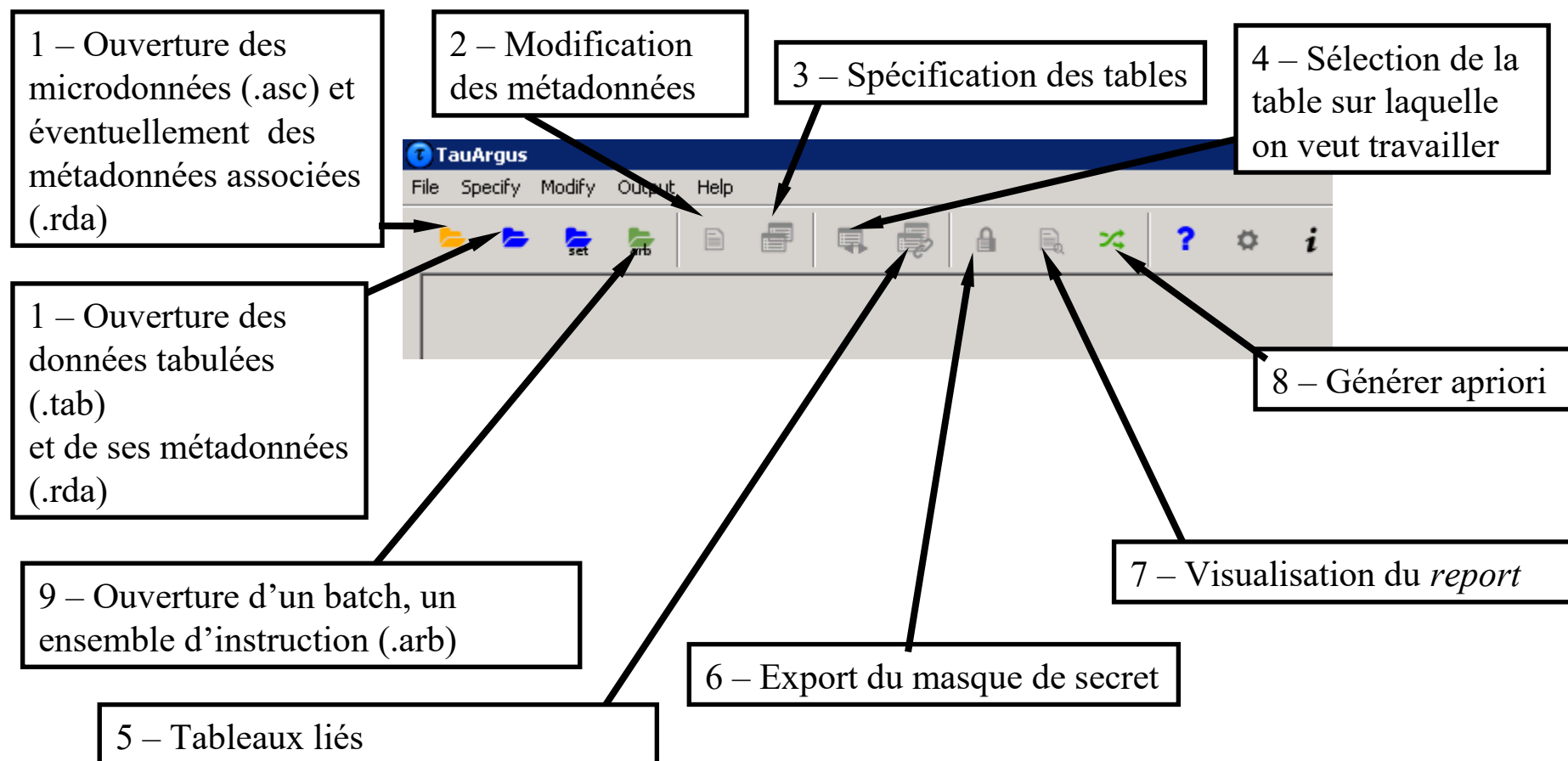
Sur AUSv3, Tau-Argus se trouve dans :

Y:\Logiciels\
TauArgus\
TauArgus4.1.7b4

Sinon, il est
téléchargeable ici :

<http://research.cbs.nl/casc/tau.htm>

Tau-Argus : exploration du menu



Metadonnées – spécification des variables

1 – On préférera le format fixe

2 – On ajoute le bon nombre de variables

3 Description de la variable :

- nom
- position et longueur dans le fichier (voir export sas)
- nombre de décimales pour les variables numériques
- type de variable :
 - *explanatory* = ventilation
 - *response* = réponse
 - *sample weight* = poids
 - *holding indicator* = holding

5 – Codelist
- gestion des labels

6 - Description des valeurs manquantes et des totaux.
Une valeur considérée manquante (ex:N.R) ne sera pas pris en compte dans le secret primaire.

4 – Spécification des informations sur les variables hiérarchiques :

- en manuel : A88/A272/APE donnera par exemple :

☒ Levels from microdata 2 1 2 0 0 ☐ ☐ ☐ ☐

- en passant par un fichier plat (.hrc). Ne pas oublier de spécifier le caractère de différenciation (@ ici)

Exemple de fichier de hiérarchie (.hrc)

Plan de présentation

1. Création des données en entrée du logiciel
 - Microdonnées
 - Métadonnées
2. Du secret primaire au secret secondaire
 - Spécification des tables
 - Éviter le secret primaire
 - Calculer le secret secondaire
 - Gestion des tableaux liés
3. Sorties proposées
 - Format SBS
 - Mise en forme Excel
4. Utilisation de données tabulées en entrée du logiciel

Spécification des tables – sélection des variables

Specify Tables

Explanatory Variables

Tref APE dep

Cell items

export <freq>

Response variable:

export

Shadow variable:

Cost variable

☐ Unity ☐ Frequency

☒ Variable ☐ Distance function

Lambda: 1.0

Parameters

☒ Dominance rule

☐ Disjunctive rule

☐ Request rule

Dom rule %rule Req. rule

	n	k
Ind-1	1	85
Ind-2	0	0
Hold-1	0	0
Hold-2	0	0

☒ Minimum frequency

Ind 3 10 %

Hold 3 10 %

☐ Zero unsafe

Range: 10

☒ Apply weights

☒ Missing=safe

☐ Use holdings info

Manual safety range: 10 %

Expl. vars	Rule	Resp. var	Shadow & cost var
APE, dep	IND.: n=1, k=85, MinFreq = 3	export	Shadow=Default, Cost=Default, w...

Compute tables Cancel

Sélection des variables de ventilation pour un tableau donné

Sélection de la variable de réponse

Variable sur laquelle le secret primaire sera effectivement calculé

Spécification des tables – choix de la variable de coût

Unity

Minimisation du nombre de cases cachées dans le secret secondaire.

Variable

Minimisation de la somme de la variable ventilée dans les tableaux (information perdue).

On peut choisir une autre variable pour calculer le coût de la perte.

cost variable

☒ unity ☐ frequency

☒ variable ☐ distance function

lambda 1

Frequency

Minimisation du nombre d'enregistrements concernés par les cases en secret secondaire (nombre d'entreprises par exemple).

Lambda

Permet de considérer la variable x^λ , ou $\ln(x)$ lorsque $\lambda=0$.

Éviter le secret primaire – dominance, fréquence et $P\%$

Règle de dominance ou du $P\%$

- intervalle calculé automatiquement
- de façon à ne pas pouvoir retrouver la ou les données individuelles sensibles (premiers contributeurs au total de la case)

Dom Rule | P%-rule | Req. rule

☒ Dominance rule

☒ P%-rule

☐ Request rule

	n	k
Ind-1	1	85
Ind-2	0	0
Hold-1	0	0
Hold-2	0	0

Dom Rule | P%-rule | Req. rule

☒ Dominance rule

☒ P%-rule

☐ Request rule

	P	N
Ind-1	10	1
Ind-2	0	0
Hold-1	0	0
Hold-2	0	0

Minimum frequency range

Ind. 3 10 %

Hold. 0 %

Zero unsafe range: 10

Manual safety range: 10 %

Règle de fréquence ou case sécurisée manuellement :

- la largeur de l'intervalle doit être définie
- relativement au total de la case
- intervalle de protection symétrique
- peut être défini plus précisément en utilisant un fichier *a priori*

- si on cache les cases égales à 0
- définition d'un intervalle de protection minimum
- défini de façon absolue

Éviter le secret primaire – poids et *holding*

Specify Tables

Explanatory Variables

Tref
APE
dep

Cell items

export
<freq>

Response variable:
export

Shadow variable:

Cost variable:
☐ Unity ☐ Frequency ☒ Variable ☐ Distance function

Lambda: 1.0

Parameters

☒ Dominance rule

☐ P%-rule

☐ Request rule

Dom rule P%-rule Req. rule

n k

Ind-1 1 85

Ind-2 0 0

Hold-1 0 0

Hold-2 0 0

☒ Minimum frequency

Freq Range

Ind 3 10 %

Hold 3 10 %

☐ Zero unsafe

Range: 10

☒ Apply weights

☒ Missing=safe

☐ Use holdings info

Manual safety range: 10 %

Expl. vars	Rule	Resp. var	Shadow & cost var
APE,dep	IND.: n=1, k=85, MinFreq = 3	export	Shadow=Default, Cost=Default, w...
Tref,dep	IND.: n=1, k=85, MinFreq = 3	export	Shadow=Default, Cost=Default, w...

Compute tables **Cancel**

Prise en compte du poids

Valeurs manquantes considérées comme sécurisées

On fait glisser les tableaux

Prise en compte du *holding*

Une fois tous les tableaux définis, on les génère.

Liste des tableaux (10 maximum à la fois)

Calculer le secret secondaire – sélection de la table de travail

1-Ouverture de la fenêtre « select table »

2-Sélection du tableau sur lequel on veut travailler

APE x dep	- Total	+ R11	+ R21	+ R22	+ R23	+ R24	+ R25
- Total	522056224,50	63021963,49	7938533,92	17750513,95	6477775,18	13010548,43	567948
+ A	122439,77	36117,80	-	-	5344,18	-	-
+ B	1402398,40	142010,56	732,00	38964,83	2123,70	7976,60	628
+ C	98258100,27	10304907,16	2931963,62	2118539,31	1802885,37	2516662,80	5830
+ D	872115,96	30823,55	102021,44	5972,08	5572,42	2739,76	5
+ E	30333873,21	348798,98	46508,60	102381,18	9722,10	2618462,95	39188
+ F	10398586,65	1512572,15	220470,40	1268558,64	617011,80	121201,02	1120
+ G	45677914,34						
+ H	37223675,37						
+ I	8597185,17						
+ J	17647570,41						
+ K	33082514,00						
+ L	112485382,33						
+ M	102062259,12						
+ N	15841901,71						
+ P	493391,20						
+ Q	2997127,50						
+ R	2631431,74						
+ S	1928357,35						

Select table

Explanatory variables	Resp. var
APE, dep	export
Tref, dep	export

Cell Information

Value: 522056224,50

Status:

Shadow: 5220562

Cost: 522056224,50

#contributions: 38157

Top n of shadow: 19269135,60

Change status

Set to safe

Set to unsafe

Set to protected

Set cost

A priori info

All Non-StructEmpty

Recode

Suppress

☒ Hypercube

☐ Modular

☐ Optimal

☐ Network

☐ CTA

Suppress

Undo suppr...

Audit

Ok Cancel

Calculer le secret secondaire – visualisation de la table : *change status*

The screenshot shows the APE x dep software interface. On the left is a data table with columns: - Total, + R11, + R21, + R22, + R23, + R24, + R25. The rows are labeled + A through + S. A text box is overlaid on the table, explaining manual status modifications. On the right is a 'Cell Information' panel with fields for Value, Status, Shadow, Cost, #contributions, and Top n of shadow. Below this is a 'Change status' dialog box with buttons for 'Set to safe', 'Set to unsafe', 'Set to protected', 'Set cost', 'A priori info', and 'All Non-StructEmpty'. At the bottom right is a 'Suppress' panel with radio buttons for Hypercube, Modular, Optimal, Network, CTA, and Rounding, and buttons for 'Suppress', 'Undo suppr...', and 'Audit'.

Modification manuelle du statut d'une case :

- *set to safe/unsafe* : forcer la diffusion/non diffusion d'une case
- *set to protected* : la case ne pourra pas être choisie pour le secret secondaire (coût infini)
- *set cost* : modification du coût pour le secret secondaire
- *a priori info* : à partir d'un fichier plat (.hst), on change le statut d'une liste de cases

Cell Information

Value: 522056224,50
Status: Safe
Shadow: 522056224,50
Cost: 522056224,50
#contributions: 38157
Top n of shadow: 19269135,60

Change status

Set to safe
Set to unsafe
Set to protected
Set cost
A priori info
All Non-StructEmpty

Recode

Suppress

☒ Hypercube
☐ Modular
☐ Optimal
☐ Network
☐ CTA
☐ Rounding

Suppress
Undo suppr...
Audit

Calculer le secret secondaire – visualisation de la table : *a priori*

L'*a priori* peut se gérer par un fichier (.hst) et permet de forcer manuellement le statut de plusieurs cases.

1 ligne de ce fichier correspond à 1 case du tableau.

Attention à ne pas spécifier des cases vides.

The screenshot shows the 'Select Apriori file' dialog box with the file path 'D:\TauArgus en pratique\Donnees demonstration\apriori.hst'. Below it, the 'apriori.hst - Bloc-notes' editor shows the content: 'D,75,s' and 'C,R11,u'. The 'A priori info' button is highlighted with a red box.

Syntaxe :

modalité_1,modalité_2,statut

Pour forcer l'état de la case, *statut* peut valoir :

- s : safe (pas de secret primaire)
- u : unsafe (secret primaire)
- p : protected (pas de secret secondaire)
- c, ... : cost (nouveau coût de la case, valeur à préciser)

The main application window displays a table with the following data:

	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	142 810.56	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

Calculer le secret secondaire – visualisation de la table : *recode*

The screenshot shows the 'recode' application window. On the left, a table with columns 'Recoded' and 'Variable' has 'APE' in the first row and 'dep' in the second row. The 'dep' row is selected. To the right of this table are buttons: 'Read', 'Apply', 'Undo', and 'Close'. In the center, a tree view shows a hierarchy starting with 'Total', followed by 'R11' (with sub-items 75, 77, 78, 91, 92, 93, 94, 95), then 'R21', 'R22', 'R23', 'R24' (with sub-items 18, 28, 36, 37, 41, 45), 'R25', 'R26', 'R31' (with sub-items 59, 62), and 'R41'. At the top right, a 'Maximum level:' dropdown menu is set to '2'. On the far right, a list of values is partially visible, including 164, 164, 164, 102, and 102. Below this list are buttons: 'unsafe', 'protected', 'cost', 'Recode', 'Suppre', 'Undo sup', and 'Auditi'.

1-On sélectionne la variable

2-On recode

3-On applique

4-On ferme

Le *maximum level* permet de recoder l'ensemble des modalités d'un coup, en les faisant remonter de x crans. Ici on passe de département à région...

...Mais si l'on désire par exemple conserver les départements des régions 11, 24 et 31, on peut modifier la structure à la main.

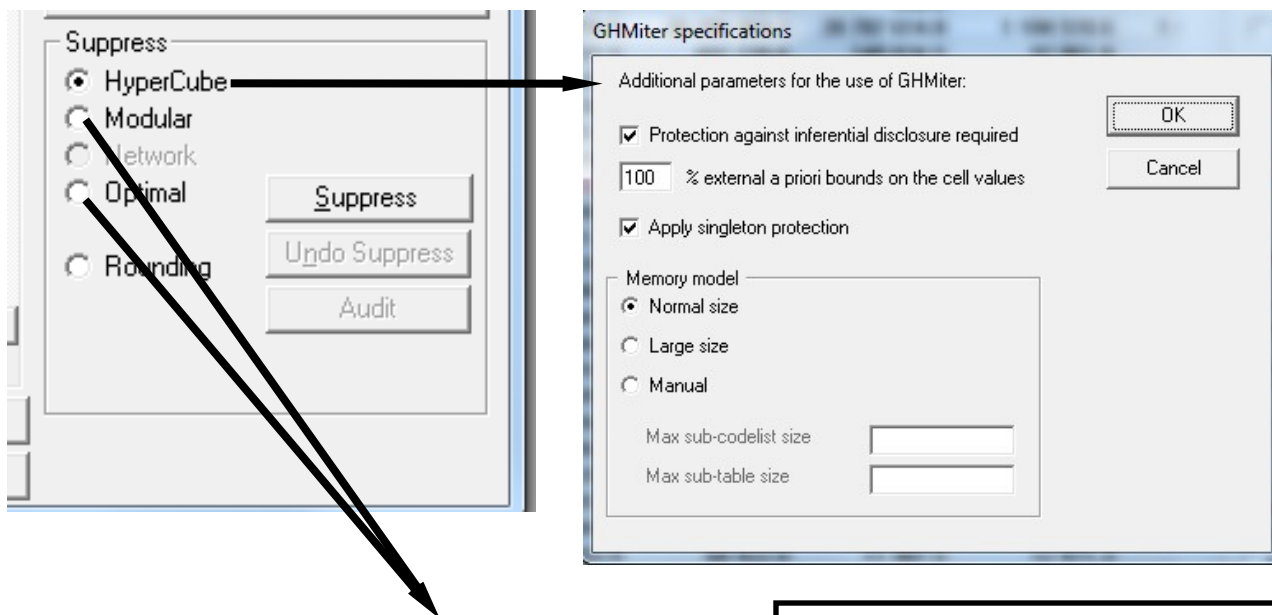
Calculer le secret secondaire – visualisation de la table : *change view*

The screenshot shows the APE software interface. The main window displays a table titled 'dep x APE' with columns for various categories and numerical values. The table includes rows for 'Total' and various 'R' categories (R11 to R94). The values are displayed in a grid format, with some cells highlighted in red. To the right of the table, there are several panels: 'Cell Information' showing details for a selected cell (Value: 522056224,50, Status: Safe, Shadow: 522056224,50, Cost: 522056224,50, #contributions: 38157, Top n of shadow: 19269135,60), 'Change status' with buttons for 'Set to safe', 'Set to unsafe', 'Set to protected', 'Set cost', and 'A priori info', and 'Suppress' with radio buttons for 'Hypercube', 'Modular', 'Optimal', 'Network', 'CTA', and 'Rounding'. At the bottom, there is a 'Select view...' dropdown menu, 'Hor. levels' and 'Vert. levels' dropdowns, 'Number of decimals' set to 2, and an 'Output view' checkbox. A red arrow points from the 'Output view' checkbox to a text box on the right.

	- Total	+ A	+ B	+ C	+ D	+ E	+ F
- Total	522056224,50	122439,77	1402398,40	98258100,27	872115,96	30333873,21	1039858
+ R11	63021963,49	36117,80	142010,56	10304907,16	30823,55	348798,98	151357
+ R21	7938533,92	-	732,00	2931963,62	102021,44	46508,60	23947
+ R22	17750513,95	-	38964,83	2118539,31	5972,08	102381,18	126855
+ R23	6477775,18	5344,18	2123,70	1802885,37	5572,42	9722,10	61791
+ R24	13010548,43	-	7976,60	2516662,80	2739,76	2618462,95	13129
+ R25	5679480,03	-	6284,00	583022,95	55,20	3918827,80	11293
+ R26	12744926,36	470,10	11045,40	1209408,46	4076,82	5272340,10	35754
+ R31	14493467,76	2616,82	107003,76	3659625,54	-	47667,90	33713
+ R41	10081338,93	57,30	5473,60	2345038,46	-	95344,32	38514
+ R42	9841223,42	22644,56	-	3906251,56	-	15203,46	9268
+ R43	6513791,36	-	113205,10	970296,48	-	14430,70	12368
+ R52	24676251,93	-	36187,00	2507051,75	27,80	58252,21	34426
+ R53	85423583,08	1991,60	71931,60	20867923,86	5659,64	351690,44	47022
+ R54	12824993,07	-	73848,15	879885,70	-	7212126,60	5834
+ R72	42660173,04	10409,92	260603,20	3969210,89	48020,45	333905,97	47872
+ R73	44832975,16	25449,38	131771,70	6428226,81	1853,92	941747,42	139949
+ R74	14653092,92	2927,76	-	7771733,84	16925,40	1731958,00	7882
+ R82	67442680,82	6951,85	279684,30	9541480,21	133058,94	246180,08	152348
+ R83	2241576,12	7430,70	-	503842,74	891,04	329049,00	717
+ R91	12668740,55	27,80	75789,20	3767027,76	19039,70	128364,80	24793
+ R93	47071131,78	-	37763,70	9673115,00	495377,80	6503447,40	61017
+ R94	7463,20	-	-	-	-	7463,20	-

Modification de l'affichage du tableau à l'écran (cela ne modifie pas la structure réelle du tableau).

Calculer le secret secondaire – quelle méthode de secret secondaire ?



Hypercube

Traitement séquentiel des cases sous secret primaire (recherche du schéma de suppression le moins coûteux).

Méthode rapide.

Optimal

Minimisation globale du coût des cases à cacher

Sous contrainte de respecter les intervalles de protection.

Méthode parfois très lente.

Modular

Pour des tableaux hiérarchiques, éclatement du problème : on applique la méthode *optimal* sur les différents sous-tableaux.

Plus rapide à traiter que la méthode *optimal*.

Calculer le secret secondaire – gestion des problèmes de singletons dans Tau-Argus

Singleton

→ Cellule avec un seul contributeur, pouvant savoir qu'il est le seul contributeur de la case.

→ Du secret secondaire supplémentaire :

- deux singletons dans une même ligne ou colonne
- un singleton et une cellule cachée en secret primaire
- deux cellules cachées à cause d'un problème de fréquence dont la somme du nombre de contributeurs est toujours insuffisante (ce problème ne se pose pas avec la règle des trois unités)

Méthode	Gestion des singletons
<i>Hypercube</i>	Oui (automatiquement)
<i>Modular</i>	Oui (en option)
<i>Network</i>	Non proposé
<i>Optimal</i>	Oui (en option)

Calculer le secret secondaire – des méthodes différentes pour des résultats différents

Méthode	Nombre de cases supprimées en secret secondaire	Coût des cases supprimées	Temps de calcul nécessaire
<i>Hypercube</i>	81 (9.47%)	260 182 513 (10.64%)	1 seconde
<i>Modular</i>	64 (7.48%)	26 017 917 (1.06%)	1 minutes et 4 secondes
<i>Optimal</i>	140 (16.37%)	25 461 484 (1.04%)	10 minutes (solution non « optimale »)

Calculer le secret secondaire – visualisation de la table : *table summary*

dep x APE

	- Total	+ A	+ B	+ C	+ D	+ E	+ F	+ G
- Total	522056224,50	122439,77	1402398,40	98258100,27	872115,96	30333873,21	10398586,65	4567791
+ R11	63021963,49	36117,80	142010,56	10304907,16	30823,55	348798,98	1513573,15	715075

Summary for table no: 1 (APE x dep | export)

Expl. var	#Codes	Status	Freq	#rec	Sum resp	Sum cost
APE	830	Safe	8739	462447	5477793538,37	5477793538,37
dep	118	Safe (manual)	0	0	0,00	0,00
		Unsafe	15738	33954	1683417871,70	1683417871,70
		Unsafe (request)	0	0	0,00	0,00
		Unsafe (freq)	2509	4846	42119126,12	42119126,12
		Unsafe (zero cell)	0	0	0,00	0,00
		Unsafe (singleton)	0	0	0,00	0,00
		Unsafe (singleton) (manual)	0	0	0,00	0,00
		Unsafe (manual)	0	0	0,00	0,00
		Protected	0	0	0,00	0,00
		Secondary	5787	71108	627512831,31	627512831,31
		Secondary (from manual)	0	0	0,00	0,00
		Empty (non-struct.)	0	0	0,00	0,00
		Empty	65167	0	0,00	0,00
Total			97940	572355	7830843367,50	7830843367,50

Close

Select view... Hor. levels: .. Number of decimals: 2

Table summary Vert. levels: .. 3 dig. separator

Les informations utiles :

- le secret primaire de fréquence : *Unsafe(freq)*
- le secret primaire de dominance/ *P%* : *Unsafe*
- les cases protégées manuellement : *Safe(manual)*
- le secret secondaire : *Secondary*
- les coûts que représentent ces différentes catégories de secret :
- Le nombre de cellules : *Freq*
- Le nombre de contributeurs : *#rec*
- La somme des valeurs des cellules : *Sum resp*
- La somme du coût choisi : *Sum cost*

Ouvre un tableau récapitulant la perte d'informations engendrée par la pose du secret primaire et secondaire.

Gestion des tableaux liés

On clique sur « linked tables »

On choisit la méthode

	- Total	+ A	+ B	+ C	+ D	+ E	+ F	+ G
- Total	482563838,30	21021,30	1362262,50	95096553,10	807363,80	30296075,00	7617688,70	37055967,...
+ R11	56968939,10	6339,50	140964,00	9868169,70	27998,30	339520,00	832875,50	5176232,...
+ R21	7361807,00	-	732,00	2905469,30	71341,00	44893,80	84758,40	1790618,...
+ R22	14286610,20	-	22415,00	1373623,30	3635,90	94210,60	1225309,80	1088921,...
+ R23	5893818,50	943,00	2123,70	1768667,40	3285,70	9722,10	601440,00	561012,...
+ R24	12295052,20	-	7976,60	2461258,20	1625,80	2618294,60	111098,20	1605650,...
+ R25	5227951,70	-	-	-	-	-	-	-
+ R26	12527216,20	470,10	-	-	-	-	-	-
+ R31	12875839,30	144,70	-	-	-	-	-	-
+ R41	9012973,90	57,30	-	-	-	-	-	-
+ R42	9106430,00	7921,60	-	-	-	-	-	-
+ R43	6379389,60	-	-	-	-	-	-	-
+ R52	23867216,40	-	-	-	-	-	-	-
+ R53	74091217,40	1991,60	-	-	-	-	-	-
+ R54	12451677,20	-	-	-	-	-	-	-
+ R72	41122267,50	557,50	-	-	-	-	-	-
+ R73	41824322,50	1436,30	-	-	-	-	-	-
+ R74	14449147,00	254,00	-	-	-	-	-	-
+ R82	63870745,50	478,40	-	-	-	-	-	-
+ R83	2211608,10	399,50	-	-	-	-	-	-
+ R91	11243791,80	27,80	-	-	-	-	-	-
+ R93	45488354,00	-	-	-	-	-	-	-
+ R94	7463,20	-	-	-	-	-	-	-

Plan de présentation

1. Création des données en entrée du logiciel
 - Microdonnées
 - Métadonnées
2. Du secret primaire au secret secondaire
 - Spécification des tables
 - Éviter le secret primaire
 - Calculer le secret secondaire
 - Gestion des tableaux liés
3. Sorties proposées
 - Format SBS
 - Mise en forme Excel
4. Utilisation de données tabulées en entrée du logiciel

Exporter ses données : différentes options

Save table

- ☐ CSV format
- ☐ CSV for pivot table
- ☐ Code-value
- ☒ SBS format ☐ Add hierarchical levels
- ☐ Intermediate format ☐ Status only
☐ Add audit results
☐ Use holding info
- ☐ JJ format ☐ Remove trivial levels

General options

- ☐ Add status
- ☐ Suppress empty cells
- ☐ Variable names on first row
- ☒ Embed spanning variables in quotes

Write table Cancel

Format demandé par Eurostat, que l'on utilise systématiquement au DMS

L'export au format demandé par Eurostat (.sbs) : mise en forme

Clic droit > ouvrir avec > calc

Sélectionner la virgule (,) comme séparateur

Préciser que la colonne Dep est un texte (pour uniformiser l'affichage)

Import de texte - [ape_dep_expor.sbs]

Importer

Jeu de caractères : Europe occidentale (Windows-1252/WinLatin 1)

Langue : Par défaut - Français (France)

À partir de la ligne : 1

Options de séparateur

☐ Largeur fixe ☒ Séparé par

☐ Tabulation ☒ Virgule ☐ Point-virgule ☐ Espace ☐ Autre

☐ Fusionner les séparateurs

Séparateur de texte : "

Autres options

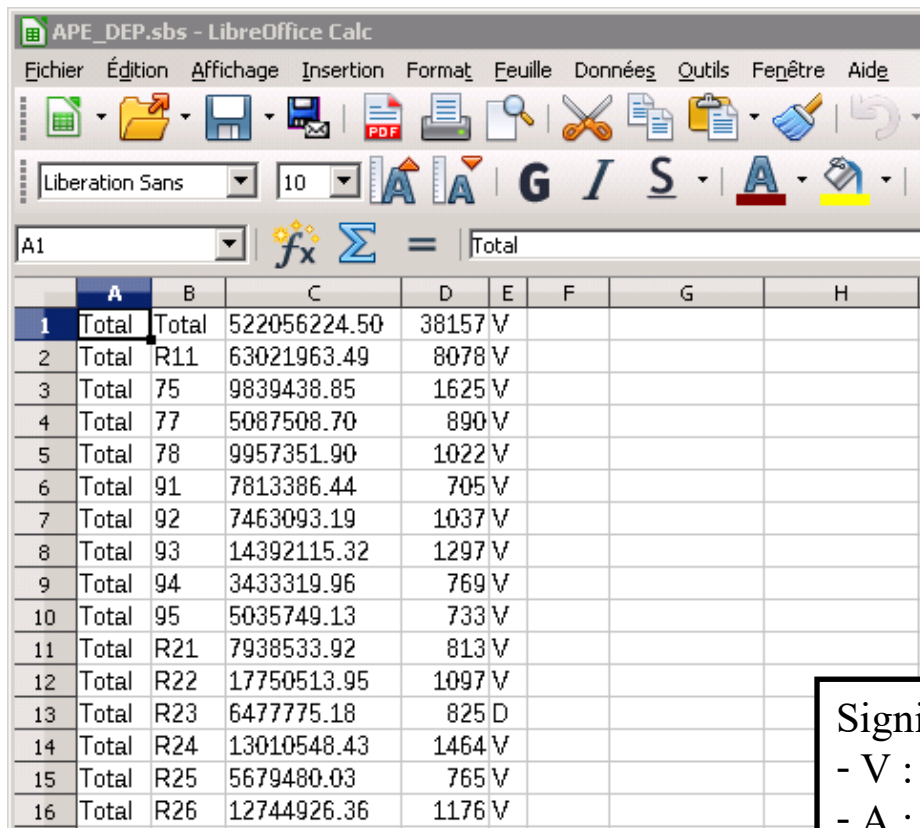
☐ Champ entre guillemets comme texte ☐ Détecter les nombres spéciaux

Champs

Type de colonne : Standard

	Standard	Standard	Standard	Standard	Standard
1	Total	95	5035749.13,733,V	224.50	38157
2	Total	R21	7938533.92,813,V	63.49	8078
3	Total	R22	17750513.65,1097,V	8.85	1625
4	Total	R23	647775.18,825,D	8.70	890
5	Total	R24	13010548.43,1464,V	1.90	1022
6	Total	R25	5679480.03,765,V	6.44	705
7	Total	R26	13744026.26,1176,V	7463093.19	1037
8	Total	R27	14392115.32	1297	V

L'export au format demandé par Eurostat (.sbs) : mise en forme



	A	B	C	D	E	F	G	H
1	Total	Total	522056224.50	38157	V			
2	Total	R11	63021963.49	8078	V			
3	Total	75	9839438.85	1625	V			
4	Total	77	5087508.70	890	V			
5	Total	78	9957351.90	1022	V			
6	Total	91	7813386.44	705	V			
7	Total	92	7463093.19	1037	V			
8	Total	93	14392115.32	1297	V			
9	Total	94	3433319.96	769	V			
10	Total	95	5035749.13	733	V			
11	Total	R21	7938533.92	813	V			
12	Total	R22	17750513.95	1097	V			
13	Total	R23	6477775.18	825	D			
14	Total	R24	13010548.43	1464	V			
15	Total	R25	5679480.03	765	V			
16	Total	R26	12744926.36	1176	V			

Signification des flags (colonne E):

- V : case diffusable
- A : secret primaire de fréquence
- B : secret primaire de dominance
- F : secret primaire lié au $P\%$
- D : secret secondaire

Plan de présentation

1. Création des données en entrée du logiciel
 - Microdonnées
 - Métadonnées
2. Du secret primaire au secret secondaire
 - Spécification des tables
 - Éviter le secret primaire
 - Calculer le secret secondaire
 - Gestion des tableaux liés
3. Sorties proposées
 - Format SBS
 - Mise en forme Excel
4. Utilisation de données tabulées en entrée du logiciel

Données tabulées – constitution d'un fichier plat de données tabulées (.tab)

donnees_demo_tabulees.tab - LibreOffice Calc

Fichier Édition Affichage Insertion Format Feuille Données Outils Fenêtre Aide

Libération Sans 10

A1

	A	B	C	D	E	F
619	B0729	95	1	3613.7	3613.7	
620	B0811	1	1	4498	4498	
621	B0811	27	1	2123.7	2123.7	
622	B0811	30	1	142	142	
623	B0811	33	2	25800.3	17902	
624	B0811	38	1	2898.6	2898.6	
625	B0811	57	1	5473.6	5473.6	
626	B0811	69	3	10514.8	10259.2	
627	B0811	78	1	168.8	168.8	
628	B0811	79	1	39.3	39.3	
629	B0811	83	1	1336.8	1336.8	
630	B0811	95	1	3550.9	3550.9	
631	B0812	1	6	188038.7	107357.3	
632	B0812	6	2	11548.6	8680.8	
633	B0812	7	1	5212.1	5212.1	
634	B0812	10	1	732	732	
635	B0812	12	4	85285	44118	
636	B0812	22	2	13488.5	8004.5	
637	B0812	24	1	7055.4	7055.4	
638	B0812	29	4	48668.4	21394.3	
639	B0812	31	4	32441.4	11418.1	
640	B0812	33	2	19146.9	17710	
641	B0812	34	1	39024.5	39024.5	
642	B0812	35	2	9774.7	8509.8	

donnees_demo_tabulees

Rechercher Tout rechercher

Feuille 1 sur 1 Par défaut Moyenne: ; Somme: ;



donnees_demo_tabulees.tab - Bloc-notes

Fichier Edition Format Affichage ?

```
B0729;95;1;3613.7;3613.7
B0811;01;1;4498;4498
B0811;27;1;2123.7;2123.7
B0811;30;1;142;142
B0811;33;2;25800.3;17902
B0811;38;1;2898.6;2898.6
B0811;57;1;5473.6;5473.6
B0811;69;3;10514.8;10259.2
B0811;78;1;168.8;168.8
B0811;79;1;39.3;39.3
B0811;83;1;1336.8;1336.8
B0811;95;1;3550.9;3550.9
B0812;01;6;188038.7;107357.3
B0812;06;2;11548.6;8680.8
B0812;07;1;5212.1;5212.1
B0812;10;1;732;732
B0812;12;4;85285;44118
B0812;22;2;13488.5;8004.5
B0812;24;1;7055.4;7055.4
B0812;29;4;48668.4;21394.3
B0812;31;4;32441.4;11418.1
B0812;33;2;19146.9;17710
B0812;34;1;39024.5;39024.5
B0812;35;2;9774.7;8509.8
```

Données tabulées – constitution d'un fichier plat de données tabulées (.tab)

/* Étape 1 : création de la tabulation */

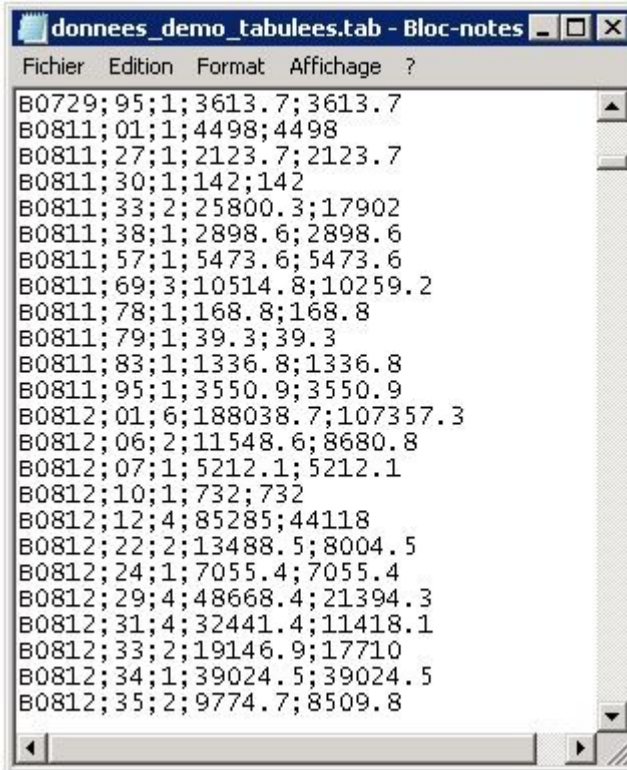
/* Au préalable, on crée un variable de comptage "un" pour que la fréquence sous Tau-Argus prenne en compte le poids */

```
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 "V:\formation-confidentialite\02-TauArgus en pratique\Donnees  
donnees_demo_tabulees.tab" dsd dlm=';' ;  
Set ape_dep (keep=ape dep un export max_export);  
Put ( _all_ ) (+0);  
run;
```



B0729;95;1;3613.7;3613.7
B0811;01;1;4498;4498
B0811;27;1;2123.7;2123.7
B0811;30;1;142;142
B0811;33;2;25800.3;17902
B0811;38;1;2898.6;2898.6
B0811;57;1;5473.6;5473.6
B0811;69;3;10514.8;10259.2
B0811;78;1;168.8;168.8
B0811;79;1;39.3;39.3
B0811;83;1;1336.8;1336.8
B0811;95;1;3550.9;3550.9
B0812;01;6;188038.7;107357.3
B0812;06;2;11548.6;8680.8
B0812;07;1;5212.1;5212.1
B0812;10;1;732;732
B0812;12;4;85285;44118
B0812;22;2;13488.5;8004.5
B0812;24;1;7055.4;7055.4
B0812;29;4;48668.4;21394.3
B0812;31;4;32441.4;11418.1
B0812;33;2;19146.9;17710
B0812;34;1;39024.5;39024.5
B0812;35;2;9774.7;8509.8

Données tabulées – les metadonnées

Free format

Separator: ;

ape
dep
un
export
max

Add
Remove
Move Up
Move Down

Attributes

Name: ape

Starting position:

Length: 0

Decimals:

Type

☒ Explanatory
☐ Response
☐ Shadow
☐ Cost
☐ Lower protection level
☐ Upper protection level
☐ Frequency
☐ Top n
☐ Status indicator

☐ Distance for suppression weight

Missings

1:
2:
Code for total:
Total

Codelist

☒ Automatic
☐ Codelist filename

Hierarchy

☒ Non hierarchical
☐ Levels from microdata
☐ Levels from file

Leading string:

SPSS meta

OK Cancel

Des caractéristiques relatives aux cases du tableau et non plus aux individus, notamment :

- le variable de réponse
- les variables de ventilation
- la valeur du premier contributeur (*topN variable*) pour la dominance
- la fréquence de la case (*frequency*)

Les données tabulées – spécification de la table

Specify tables

Variables

Explanatory:

ape
dep

Number: 2 TopN: 1 Status Frequency Cost

☒ Do not allow non-additivity
☐ Compute incorrect totals
☐ Allow non-additivity

Cost function for secondary suppression

☒ Response variable
☐ Cost variable Lambda: 1
☐ Frequency
☐ Unity
☐ Distance

Safety rules

☐ Use given status ☒ Use safety rules

☒ Dominance rule
Number: 3
Percentage: 85 %

☐ P% rule
P: 10
N: 1

☒ Minimum frequency
Frequency: 3
Range: 10 %

☐ Zero unsafe
Zero margin: 10

☐ Missing = safe
Manual safety range: 20 %

OK Cancel

1 – Variables
sélectionnées
pour le
tableau

2 – Gestion du
coût pour le
secret
secondaire

3 – Règles
de secret
primaire

Les données tabulées – tableaux liés

1 – On ouvre chaque fichier de données tabulées et son fichier de métadonnée, que l'on devra préparer en amont.

2 – Pour chaque tableau, on spécifie les règles de secret primaire.

3 – Enfin, on accède par le menu à la fenêtre de gestion des tableaux liés.

	- Total	01	02	03	04	05	06	07
- Total	482563838,3	27508033,4	11268093,1	243972,3	9151260,3	7178790,7	20635873,0	3408720,1
A0220	21021,3	-	-	-	-	-	-	-
B0729	3613,7	-	-	-	-	-	-	-
B0811	56546,8	4498,0	-	-	-	-	-	-
B0812	1089561,7	188038,7	-	-	-	-	11548,6	5212,1
B0892	8337,1	-	-	-	-	-	-	-
B0899	3939,7	-	-	-	-	-	-	-
B0910	200263,5	-	-	-	-	-	-	-
C1011	422795,9	230	-	-	-	-	-	-
C1012	360049,2	-	-	-	-	-	-	-
C1013	1003511,7	110	-	-	-	-	-	-
C1020	299364,8	-	-	-	-	-	-	-
C1031	608696,5	-	-	-	-	-	-	-
C1032	247523,6	-	-	-	-	-	-	-
C1039	597737,8	24	-	-	-	-	-	-
C1041	10725,6	-	-	-	-	-	-	-
C1051	1854933,0	1291	-	-	-	-	-	-
C1052	7058,4	-	-	-	-	-	-	-
C1061	145354,8	225	-	-	-	-	-	-
C1062	146770,2	-	-	-	-	-	-	-
C1071	675620,7	-	-	-	-	-	-	-
C1072	252302,9	66	-	-	-	-	-	-
C1073	134726,9	-	-	-	-	-	-	-
C1081	1045075,7	-	-	-	-	-	-	-
C1082	87151,2	35	-	-	-	-	-	-
C1083	196492,0	-	-	-	-	-	-	-
C1084	696032,8	235	-	-	-	-	-	-
C1085	334289,7	-	-	-	-	-	-	-
C1086	842767,7	3265,4	-	-	-	-	-	-
C1089	145614,6	-	-	-	-	-	45170,4	15173,0
C1091	124195,1	-	-	-	-	-	3769,0	-

Gestion de la confidentialité dans les tableaux de données agrégées

Merci pour votre attention

Contacts

Maxime Beauté

01 87 69 55 43

maxime.beaute@insee.fr

Alexandre Awad

01 87 69 55 14

alexandre.awad@insee.fr

Insee

88, avenue Verdier

92120 Montrouge

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