

La gestion de la confidentialité dans les tableaux statistiques.

1- Un cadre légal et réglementaire.

But et enjeux

Le but de la gestion du secret statistique est de :

- protéger les données individuelles.
- empêcher la reconstruction des données individuelles à partir des diffusions.

L'enjeu est de :

- conserver la confiance des répondants et/ou des dépositaires d'informations administratives.
- récolter un maximum de données de la meilleure qualité possible.

La loi du 7 Juin 1951

La loi n°51-711 du 7 Juin 1951 modifiée sur l'obligation, la coordination et le secret en matière de statistiques définit ce qu'est le secret statistique, ses limites et ses conditions.

Elle définit également les sanctions applicables en cas de violation du secret statistique.

- Jusqu'à 1 an d'emprisonnement ;
- Jusqu'à 15 000 € d'amende.

La loi du 7 Juin 1951

Le secret statistique interdit au dépositaire d'informations:

- de communiquer à des tiers des renseignements individuels recueillis par enquête statistique (données d'enquêtes) ;
- de divulguer des informations qui lui ont été transmises par des tiers à des fins exclusives d'établissement de statistiques (données administratives).

Il permet ainsi d'assurer :

- aux personnes physiques que la confidentialité sur leur vie professionnelle et familiale sera garantie ;
- aux entreprises que le secret commercial sera respecté: les informations transmises ne seront pas mises à disposition de leurs concurrents.

Code de Bonnes Pratiques de la Statistique Européenne

Principe 5 : Secret statistique

Le respect de la vie privée ou du secret des affaires des fournisseurs de données (ménages, entreprises, administrations et autres répondants), la confidentialité des informations qu'ils communiquent et l'utilisation de celles-ci à des fins strictement statistiques doivent être absolument garantis.

Indicateurs :

- Le secret statistique est garanti par le droit.
- Le personnel de l'autorité statistique signe un engagement de confidentialité au moment de l'entrée en fonction.
- De lourdes sanctions sont prévues pour toute violation délibérée du secret statistique.
- Des instructions et des lignes directrices sont fournies concernant la protection du secret statistique lors des processus de production et de diffusion. Ces lignes directrices sont formulées par écrit et portées à la connaissance du public.
- Des dispositions matérielles et techniques sont prises en vue de garantir la sécurité et l'intégrité des bases de données statistiques.
- Des protocoles stricts s'appliquent aux utilisateurs extérieurs ayant accès aux micro-données statistiques à des fins de recherche.

Les types de source

Nous nous intéressons dans cette présentation uniquement au cas de tableaux de données agrégées.

Nous allons énoncer dans la suite les règles pratiques pour les trois cas de provenance des informations :

- informations tirées d'enquêtes statistiques (données d'enquêtes) ; les enquêtes statistiques sont celles qui ont reçu le visa prévu par l'article 2 de la loi n°51-711 ;
- informations tirées de sources administratives (données administratives) ; l'article 2 de la loi n°51-711 prévoit que ces données sont couvertes par le secret statistique ;
- informations issues de sources mixtes (données mixtes).

Données d'enquêtes statistiques... *sur les entreprises*

Pour les tableaux de données agrégées sur les entreprises, les deux règles sont les suivantes :

- aucune case ne doit concerner moins de trois unités.

C'est une décision du directeur général de l'Insee datant du 13 Juin 1980.

- aucune case ne doit contenir des données pour lesquelles une entreprise représente plus de 85 % du total.

Cette règle a été élaborée le 7 Juillet 1960 par le Comité de coordination des enquêtes statistiques (Cocoes), ancêtre du Conseil national de l'information statistique (Cnis).

Données d'enquêtes statistiques...

sur les ménages

Pour les tableaux de données agrégées sur les ménages, la seule règle imposée est que :

- l'identification directe ou indirecte des individus est impossible.
- dans la pratique, on considère le secret statistique respecté si la connaissance d'une caractéristique pour un individu ne peut pas entraîner la connaissance d'une autre caractéristique avec laquelle elle est croisée dans un tableau.

	18-25	26-49	50-59	60 et +
Marié	7	27	0	30
Divorcé	0	11	9	22
Autre	21	12	0	4

Données administratives

Les règles du secret à appliquer sont différentes d'une source à l'autre et propres à chacune d'entre elles.

Dans la diffusion de tableaux de données agrégées issues de :

- données fiscales, aucune case ne doit concerner moins de 11 individus ou ménages.
- données DADS, aucune case ne doit concerner moins de 5 individus ou ménages.
- autres sources, les règles de diffusion sont celles fixées par l'organisme qui a communiqué les informations. En général, ces règles sont inscrites dans la convention qui a permis la transmission des données.

Sources mixtes

On entend par sources mixtes, les sources provenant de combinaisons (appariements) d'enquêtes statistiques et de données administratives.

- Les règles à adopter s'obtiennent par cumul des règles applicables d'une part aux enquêtes statistiques et d'autre part aux fichiers administratifs.

- On peut citer parmi ces sources :

l'enquête revenus fiscaux et sociaux qui apparie les résultats de l'enquête emploi, des données fiscales et des données fournies par les Caf.

2- Les problèmes de diffusion.

Tableaux de fréquence

C'est un tableau dans lequel la valeur d'une cellule correspond au nombre d'unités qui partagent les caractéristiques de la cellule.

- Exemple : nombre d'entreprises dont l'activité a un impact écologique par région.

	Impact sur l'environnement		Total
	Oui	Non	
Nord	9	0	9
Est	14	2	16
Ouest	1	1	2
Sud	7	12	19
Total	31	15	46

- Ce type de tableau est parfois appelé « tableau de comptage ».
- La valeur des cellules n'est pas sensible puisque c'est un comptage.
- Ce sont les variables de ventilation qui sont soit identifiantes, soit sensibles.

Tableaux de fréquence

- Groupe 1 de divulgation : l'ensemble des entreprises émettrices de CO₂ de la région Nord ont un impact environnemental.
⇒ Tous les répondants ne doivent pas avoir la même caractéristique d'une variable sensible.

- Groupe 2 de divulgation : les entreprises de la région Est qui n'ont pas d'impact écologique sont à peu près sûres que les autres entreprises de la même région en ont un.
⇒ Il ne faut pas qu'il y ait trop (en proportion) de répondants dans une modalité sensible d'une variable.

- Groupe 3 de divulgation : l'entreprise de la région Ouest qui n'a pas d'impact écologique sait que l'autre entreprise de la même région en a un.
⇒ Il ne faut pas qu'il y ait trop peu de répondants dans les modalités non sensibles d'une variable.

Tableaux de fréquence

Un critère possible est que :

- Au moins n répondants doivent tomber dans les modalités non sensibles, s'il y a des répondants dans une modalité sensible. Par exemple $n=3$.
 - Dans notre exemple, s'il y avait eu 3 entreprises de l'Est dans la modalité « non » concernant l'impact écologique, ces entreprises n'auraient pas pu savoir quelles entreprises ont un impact sur l'environnement.
 - S'il y avait eu une seule entreprise dans la modalité « oui », cette entreprise qui a effectivement un impact aurait su que les autres n'en ont pas. A priori dans ce sens cette information divulguée n'est pas pénalisante.
- Nous pourrions tout de même imaginer que les entreprises qui n'ont pas d'impact sur l'environnement vont toutes communiquer sur ce fait.
- **En pratique, on impose au moins n répondants dans toutes les modalités.**

Tableaux de volume

C'est un tableau dans lequel la valeur de chaque cellule représente la somme des contributions des répondants qui partagent les caractéristiques de cette cellule.

- Exemple du chiffre d'affaires d'entreprises par région et activité.

	Activités			Total
	A	B	C	
Nord	58	47	36	141
Est	71	124	24	219
Ouest	92(5)	157(2)	59(1)	308
Sud	800	934	651	2385
Total	1021	1262	770	3053

- Les données entre parenthèses représentent le nombre de répondants cachés derrière la valeur de la cellule. En effet, même si ce n'est pas un tableau de comptage, il faut tout de même avoir un œil sur le nombre de répondants.

Tableaux de volume

Nous ne devons pas publier d'information individuelle sensible.

- Que faire des cellules avec 1 ou 2 contributeurs ?
 - Si un individu sait que l'entreprise de la région Ouest et d'activité C est la seule à exercer cette activité dans cette région alors cet individu connaît le chiffre d'affaires de l'entreprise.
 - Enfin, parmi les deux entreprises de la région Ouest d'activité B chacune peut déduire le chiffre d'affaires de l'autre.

- Que faire des cellules qui ont un contributeur dominant ?

Tableaux de volume

Nous définissons les règles suivantes pour juger de la sensibilité d'une cellule:

- règle de fréquence minimale
 - Elle permet de contrôler que la cellule est construite à partir d'un nombre minimal de répondants.

- règle de dominance du (n,k)
 - Elle permet de vérifier qu'un certain nombre de répondants (souvent 1) ne dominant pas trop le total la cellule.

- règle du p %
 - Elle permet de vérifier que les répondants entre eux n'ont pas le pouvoir de divulguer de l'information sur un autre partageant les mêmes caractéristiques.

Tableaux de volume

Règle de fréquence minimale

Une cellule est non sécurisée si le nombre de contributeurs au total de la cellule est inférieur à un entier n fixé.

- Par exemple, pour les tableaux issus d'enquêtes statistiques d'entreprises, on prend $n=3$.

Pour des données d'enquêtes par échantillon, il faut considérer les poids d'échantillonnage.

- Une cellule peut être construite à partir de 2 répondants mais en représenter davantage.
- L'échantillon et la pondération associée doivent rester confidentiels.

Tableaux de volume

Règle de dominance du (n,k)

Une cellule C est non sécurisée si la somme des n plus grandes contributions x_i au total de la cellule dépasse k % de ce total T_C .

On note $N(C)$ le nombre de répondants dans la cellule C.

$$x_1 \geq x_2 \geq \dots \geq x_n \geq \dots \geq x_{N(C)}$$

La cellule est non sécurisée si :

$$\sum_{i=1}^n x_i > \frac{k}{100} \sum_{i=1}^{N(C)} x_i$$

⇒ On ne souhaite pas que les n plus grands contributeurs dominent trop le total de la cellule.

Tableaux de volume

Règle de dominance du (n,k)

Prenons l'exemple des 5 entreprises d'activité A de la région Ouest.

- Le total de la cellule est construit comme cela :
 $81 + 5 + 3 \times 2 = 92$ M€
- Regardons si la cellule est sécurisée pour les règles $(1,85)$ ou $(2,90)$...

Vérifications :

- $(1,85)$: $81 / 92 \approx 0.88 > 0.85$. Non sécurisée!
- $(2,90)$: $(81+5) / 92 \approx 0.93 > 0.90$. Non sécurisée!

Tableaux de volume

Règle du p %

Une cellule est non sécurisée si des répondants peuvent estimer la contribution d'un autre répondant « appartenant » à la même cellule avec une précision de p% sur la vraie valeur.

Imaginons que le second plus grand contributeur ait envie d'estimer la plus grande contribution.

Il peut alors procéder de la façon suivante en utilisant sa propre contribution et le total de la cellule: $T_C - x_2$

La cellule est non sécurisée si : $(T_C - x_2) - x_1 < \frac{p}{100} x_1$

⇒ On ne souhaite pas que l'erreur d'estimation soit trop petite par rapport à la vraie valeur x_1 .

Tableaux de volume

Règle du p %

Reprenons l'exemple des 5 entreprises d'activité A de la région Ouest.

- Pour rappel, le total de la cellule était construit comme cela:

$$81 + 5 + 3 \times 2 = 92 \text{ M€}$$

Regardons si la cellule est sécurisée pour la règle du p% avec p=10...

Vérification:

- $[(92 - 5) - 81] / 81 \approx 0.074 < 0.10$. Non sécurisée!
- Le second contributeur peut estimer le plus grand trop précisément.

Secret primaire

Les cellules qui ne respectent pas les règles de sécurité forment ce que l'on appelle le secret primaire.

On ne peut pas diffuser ces cellules!

On peut par exemple choisir de les supprimer.

- À cause des marges que l'on publie, il est parfois possible de retrouver ces valeurs. Leur simple suppression n'est alors pas suffisante.
- Mais d'autres sources de problèmes peuvent exister.

Les tableaux liés

Les tableaux liés sont les tableaux qui possèdent la même variable ventilée et partagent une ou plusieurs variables de ventilation.

- Imaginons deux tableaux ventilant le chiffre d'affaires d'entreprises, l'un par activité et région, l'autre par activité et tranche d'effectif.
- Ces deux tableaux sont liés puisqu'ils ventilent tous les deux le chiffre d'affaires et qu'ils ont l'activité comme variable de ventilation en commun.

Ces tableaux peuvent être considérés comme des sous-tableaux d'un tableau de dimension plus grande.

Les tableaux liés

Regardons les trois tableaux de comptage ci-dessous, qui recensent des dirigeants d'entreprise suivant trois variables :

- Région d'implantation de l'entreprise
- Sexe du dirigeant
- Activité polluante ou non?

	A	B	Total
Homme	21	12	33
Femme	16	19	35
Total	37	31	68

(polluante?)	Oui	Non	Total
A	11	26	37
B	20	11	31
Total	31	37	68

(polluante?)	Oui	Non	Total
Homme	23	10	33
Femme	8	27	35
Total	31	37	68

Notons les valeurs des cellules du tableau de dimension 3 par x_{SRP} .

S: H (Homme) ou F (Femme)

R: A (région A) ou B (région B)

P: O (Oui) ou N (Non)

$$x_{HAO} + x_{HAN} = 21$$

$$x_{HBO} + x_{HBN} = 12$$

$$x_{FAO} + x_{FAN} = 16$$

$$x_{FBO} + x_{FBN} = 19$$

Les tableaux liés

Il suffit de résoudre ce système d'équations sous une contrainte naturelle qui est que les x_{SRP} sont des entiers positifs.

On obtient alors des données plus précises :

$$x_{HAO} = 11 \quad x_{HAN} = 10$$

$$x_{HBO} = 12 \quad x_{HBN} = 0$$

$$x_{FAO} = 0 \quad x_{FAN} = 16$$

$$x_{FBO} = 8 \quad x_{FBN} = 11$$

On sait du coup que toutes les entreprises de la région B dirigées par des hommes ont une activité polluante.

Les tableaux hiérarchisés

Les tableaux hiérarchisés sont un cas particulier des tableaux liés.

Une ou plusieurs variables de ventilation sont hiérarchisées, c'est-à-dire qu'il existe des sous-totaux.

- Exemple de la Naf, différents niveaux de finesse d'activités avec les :
 - sous-classes (732)
 - classes (615)
 - groupes (272)
 - divisions (88)
 - sections (21)

- Exemple des zones géographiques :
 - départements (96+5)
 - régions (22+5)

Les tableaux hiérarchisés

On ne peut pas se permettre de diffuser ces deux tableaux indépendamment, car il existe des relations d'additivité.

Nord1	21
Nord2	X
Nord3	23
Est1	27
Est2	41
Est3	X
Ouest1	32
Ouest2	54
Ouest3	67
Ouest4	38
Sud1	44
Sud2	39
Pays	400

Nord	46
Est	80
Ouest	191
Sud	83
Pays	400



Nord1	21
Nord2	2
Nord3	23
Nord	46
Est1	27
Est2	41
Est3	12
Est	80
Ouest1	32
Ouest2	54
Ouest3	67
Ouest4	38
Ouest	191
Sud1	44
Sud2	39
Sud	83
Pays	400

Singleton

Ce sont les cases où il y a un seul répondant.

Elles nécessitent une analyse supplémentaire : ne peuvent-elles pas retrouver une case en secret primaire grâce à leur propre information ?

	Âge					Total
	<24	24-30	30-50	50-60	>60	
Oui	1	6	7	5	1	20
Non	7	16	18	15	4	60
Total	8	22	25	20	5	80

	Âge					Total
	<24	24-30	30-50	50-60	>60	
Oui	x	6	7	x	x	20
Non	x	16	18	x	x	60
Total	8	22	25	20	5	80



3- Les méthodes de gestion

Éviter le secret primaire

Restructurer les données en recodant les variables de ventilation :

- fusionner des lignes ou colonnes ;
- recoder les modalités d'une variable de ventilation.

Éviter le secret primaire

Fusion de lignes ou colonnes

Le but est de combiner des lignes et/ou des colonnes pour augmenter le nombre de répondants dans les cellules.

	Région			Total
	A	B	C	
Activité 1	58	47	36	141
Activité 2	71	124	24	219
Activité 3	92 (5)	157 (2)	59 (1)	308
Autres	800 (12)	934 (7)	651 (4)	2385
Total	1021	1262	770	3053

	Région			Total
	A	B	C	
Activité 1	58	47	36	141
Activité 2	71	124	24	219
Autres	892 (17)	1091 (9)	710 (5)	2693
Total	1021	1262	770	3053



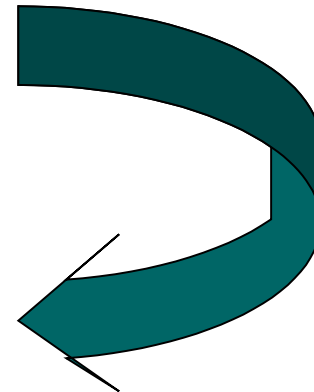
Éviter le secret primaire

Recoder des modalités

Définir de nouvelles modalités pour les variables de ventilation, afin d'accroître le nombre de répondants par cellule.

	Âge				Total
	<25	25-30	30-50	>50	
Oui	2	5	7	6	20
Non	8	15	18	19	60
Total	10	20	25	25	80

	Âge				Total
	<28	28-35	35-55	>55	
Oui	3	6	6	5	20
Non	9	17	19	15	60
Total	12	23	25	20	80



La recherche des bonnes modalités est une exploration à faire manuellement...

Éviter le secret primaire

Mais peut-on vraiment...

Les solutions précédentes conduisent soit à un changement de la structure de diffusion des résultats, soit à une modification des données.

Souvent les structures de diffusion sont rigides (règlements, suivi dans le temps...).

La solution généralement envisagée est donc la non diffusion des agrégats touchés par le secret primaire. On les supprime (cache) de la diffusion. Ce sont les suppressions primaires.

Les suppressions primaires

	Âge				Total
	<25	25-30	30-50	>50	
Oui	2	5	7	6	20
Non	8	15	18	19	60
Total	10	20	25	25	80

	Âge				Total
	<25	25-30	30-50	>50	
Oui	x	5	7	6	20
Non	8	15	18	19	60
Total	10	20	25	25	80



Suppression de la cellule sous secret

Mais cela ne suffit pas car les marges (totaux des lignes et colonnes) sont diffusées (ici et généralement).

On peut donc retrouver la cellule par simple différence en ligne ou colonne.

Les suppressions secondaires

Nous supprimons de nouvelles cellules (suppressions secondaires) afin de ne pas pouvoir reconstruire la cellule sensible.

	Âge				Total
	<25	25-30	30-50	>50	
Oui	2	5	7	6	20
Non	8	15	18	19	60
Total	10	20	25	25	80

	Âge				Total
	<25	25-30	30-50	>50	
Oui	x	x	7	6	20
Non	x	x	18	19	60
Total	10	20	25	25	80



D'autres cellules sont supprimées.

Maintenant il est impossible de reconstruire la cellule sensible.

Les suppressions secondaires

Toutes les suppressions, primaires et secondaires, sont remplacées par le même symbole. On a alors une structure de suppressions possible.

	Âge				
	<25	25-30	30-50	>50	Total
Oui	x	x	7	6	20
Non	x	x	18	19	60
Total	10	20	25	25	80

Mais attention, cela n'est pas forcément suffisant, car cacher des cases revient à diffuser des intervalles.

L'utilisation des relations d'additivité du tableau, voire d'informations a priori, permettent de calculer ces intervalles.

Les suppressions

Intervalle des possibles

Ces intervalles dont il est question sont appelés intervalles des possibles.

Cela correspond à l'ensemble des valeurs possibles prises par la cellule supprimée.

On le définit comme cela: $[T_C - L_C ; T_C + U_C]$ avec L_C et $U_C \geq 0$.

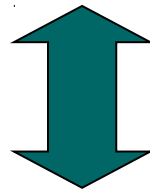
On trouve les bornes de ces intervalles en maximisant et minimisant la valeur cachée de la cellule sous contrainte d'additivité du tableau et de positivité des cellules.

Les suppressions

Intervalle des possibles

Cacher des cellules revient donc à diffuser des intervalles.

	Âge				
	<25	25-30	30-50	>50	Total
Oui	x	x	7	6	20
Non	x	x	18	19	60
Total	10	20	25	25	80



Il y a une totale équivalence entre ces deux tableaux.

	Âge				
	<25	25-30	30-50	>50	Total
Oui	[0 ; 7]	[0 ; 7]	7	6	20
Non	[3 ; 10]	[13 ; 20]	18	19	60
Total	10	20	25	25	80

Les suppressions

Intervalle des possibles

On peut donc ici approcher la case sensible par l'intervalle [0 ; 7].

	Âge				
	<25	25-30	30-50	>50	Total
Oui	[0 ; 7]	[0 ; 7]	7	6	20
Non	[3 ; 10]	[13 ; 20]	18	19	60
Total	10	20	25	25	80

La question est de savoir si cela est satisfaisant en terme de protection.

L'enjeu est donc de définir un intervalle minimal qui satisferait notre exigence en termes de protection.

Il faut alors choisir les suppressions secondaires qui permettent de respecter cet intervalle.

Cet intervalle sera appelé intervalle minimal de protection lié à un niveau de sécurité souhaité.

Niveau de sécurité

L'intervalle des possibles de suppressions primaires doit être suffisamment large.

Exiger une longueur minimale d'intervalles des possibles des suppressions primaires, cela revient à imposer un niveau minimum de sécurité.

Deux possibilités :

- pourcentage fixe de la valeur de la cellule :
 - désavantage: le niveau de sensibilité de la cellule n'est pas pris en compte.
 - mais c'est obligatoirement le cas pour la règle de fréquence minimale.
- lié aux règles de sécurités utilisées :
 - l'intervalle contient au moins une valeur de cellule qui satisfait le secret primaire.
 - le niveau de sensibilité de la cellule détermine le niveau de sécurité.

Niveau de sécurité

exemple de la règle du p %

L'intervalle des possibles est de la forme : $[T_C - L_C; T_C + U_C]$

La borne supérieure de x_1 déductible par x_2 est : $x_1^U = (T_C + U_C) - x_2$

L'exigence pour satisfaire la règle est que : $x_1^U \geq \left(1 + \frac{p}{100}\right)x_1$

Ainsi, on en déduit la valeur de U_C minimale :

$$U_C \geq \frac{p}{100}x_1 - \sum_{i=3}^{N(C)} x_i = \frac{p}{100}x_1 - (T_C - x_1 - x_2)$$

Niveaux supérieurs de protection

Règle de sensibilité

Niveau supérieur de protection

Règle de dominance du (1,k)
$$U_C \geq \frac{100}{k} \cdot x_1 - T_C$$

Règle de dominance du (n,k)
$$U_C \geq \frac{100}{k} \cdot (x_1 + x_2 + \dots + x_n) - T_C$$

Règle du p %
$$U_C \geq \frac{p}{100} \cdot x_1 - (T_C - x_1 - x_2)$$

Masque de secret

Un masque de secret est une structure possible de suppressions (cellules sous secret primaire + cellules sous secondaires).

Une structure de suppressions est possible lorsque l'intervalle des possibles des suppressions primaires fournit suffisamment de protection, au regard du niveau de sécurité souhaité.

Comment alors choisir entre les masques de secret envisageables lorsqu'il y en a plusieurs?

Minimiser la perte d'information

Pour minimiser cette perte d'information, il faut définir un coût à la suppression d'une cellule.

On peut ainsi choisir de minimiser:

- le nombre de cases supprimées du tableau (coût = unité) ;
- le nombre d'enregistrements (contributeurs) concernés par les cases supprimées ;
- la valeur de la case supprimée (cela permet de plutôt conserver les cellules de plus forte valeur) ;
- la valeur qu'aurait prise la cellule pour une autre variable de réponse (shadow variable).

Méthodes de recherche des masques

Pour avoir le choix entre différents masques de secret et arbitrer en minimisant la perte d'information, il faut avoir un ensemble de solutions (masques) possibles.

Plusieurs méthodes (algorithmes) existent (et ont été implémentées) pour rechercher les suppressions secondaires.

Elles ne s'appliquent pas toutes à tous les cas et n'ont pas les mêmes avantages ou inconvénients.

Méthodes de recherche des masques

Méthode Hypercube

Traitement séquentiel des cases sous secret primaire (recherche du masque le moins coûteux).

Méthode gratuite et rapide

Méthode Network

Modélisation du problème par graphes

Méthode gratuite mais contraignante (uniquement tableau 1H2D)

Méthode Optimal

Minimisation globale du coût

Méthode parfois lente et payante

Méthode Modular

Éclatement en sous tableaux

Plus rapide que la méthode Optimal

Comparaison sur un cas

Méthode	Nombre de cases supprimées	Coût des cases supprimées	Temps de calcul nécessaire
Hypercube (méthode gratuite)	280 (14.1%)	36426200.2 (2.5%)	3 secondes
Modular (méthode payante)	195 (9.8%)	12385301.7 (0.83%)	6 secondes
Optimal (méthode payante)	193 (9.7%)	11177777.2 (0.75%)	16 minutes et 10 secondes

Éléments de bibliographie

M. Bergeat, *Secondary cell suppression in tabular data: a comparison of the methods implemented in τ -Argus*, UNECE/Eurostat Work Session on Statistical Data Confidentiality, 2013.

J. Nicolas, *Traitement de la confidentialité statistique dans les tableaux : expérience de la Direction des Statistiques d'Entreprises*, Journées de Méthodologie Statistique de l'Insee, 2012.

A. Hundepool & al. *Handbook on statistical disclosure control*, version 1.2, 2010.

J. Nicolas, *La gestion du secret dans les tableaux diffusant des statistiques d'entreprises*, La Lettre du SSE, n° 65, septembre 2010.

L. Willenborg and T. de Waal. *Elements of Statistical Disclosure Control*, Lecture Notes in Statistics, vol 155, Springer-Verlag, 2000.