



Séminaire de Méthodologie Statistique

mardi 24 juin 2014

14h-17h, Insee - Malakoff 1 - salle 1245

Le traitement de la confidentialité

La gestion de la confidentialité dans les tableaux statistiques

Julien Nicolas - SOeS

Un panorama de la protection des fichiers de données individuelles

Maxime Bergeat - Insee, Département des méthodes statistiques

Traitement de la confidentialité dans la réponse au règlement européen sur les recensements de la population et du logement

Heidi Koumarianos - Insee, Département de la démographie

Résumés des interventions

La gestion de la confidentialité dans les tableaux statistiques

Julien Nicolas - SOeS

En septembre 2008, la gestion de la confidentialité des données est apparue officiellement au sein de l'organigramme de l'Insee. Depuis, l'Insee a accumulé de l'expérience et de l'expertise sur un logiciel, Tau-Argus, pour traiter de façon plus méthodique la question du secret dans les tableaux statistiques. Cet exposé sera l'occasion de rappeler brièvement les règles de confidentialité des données, puis de présenter, de manière théorique et au travers d'exemples, les méthodes à mettre en œuvre pour respecter ces règles lors de la diffusion de données agrégées sous forme de tableaux : suppressions primaires d'une part, c'est-à-dire suppressions des cellules ne respectant pas les règles de confidentialité des données ; suppressions secondaires d'autre part, c'est-à-dire suppressions d'autres cellules afin d'empêcher la reconstitution du tableau initial.

.../...

Le traitement de la confidentialité

Insee - Département des méthodes statistiques

Séminaire du 24 juin 2014

Un panorama de la protection des fichiers de données individuelles

Maxime Bergeat - Insee, Département des méthodes statistiques

La question d'une diffusion plus importante des fichiers de données individuelles se pose de plus en plus à l'Insee, notamment dans le cadre d'une volonté d'ouverture des données. Parallèlement à cela, il est nécessaire de s'assurer que les fichiers de données largement diffusés protègent la vie privée des répondants, afin de conserver leur confiance. Cet exposé présente un cadre théorique concernant l'anonymisation de fichiers de données individuelles. Après une présentation schématique du processus de protection, plusieurs techniques pour estimer le risque de ré-identification sont introduites. Une revue de la littérature sur les méthodes de protection, perturbatrices ou non, est ensuite effectuée. Des mesures de la perte d'information engendrée par la protection sont enfin présentées. La présentation est ponctuée d'exemples de fichiers diffusés par les instituts nationaux de statistique.

Traitement de la confidentialité dans la réponse au règlement européen sur les recensements de la population et du logement

Heidi Koumarianos - Insee, Département de la démographie

Dans le cadre d'un règlement européen, la France a livré à Eurostat 60 hypercubes (tableaux multidimensionnels croisant toutes les subdivisions des variables retenues et donnant la valeur relative à chacun des croisements) aux niveaux France, région, région de travail, département et commune, ce qui représente plus de 400 millions de cases non vides.

En France, la diffusion des résultats du recensement de la population est encadrée par l'arrêté du 19 juillet 2007. En particulier, cet arrêté précise les règles de diffusion à un niveau géographique fin. Les problèmes liés à la diffusion des données du RP européen ne sont pas tout à fait de même nature : en raison de croisements très fins, il y a de nombreuses cases petites, y compris à des niveaux géographiques agrégés comme les régions.

On a donc choisi de limiter le traitement de la confidentialité aux variables dites sensibles citées explicitement dans l'arrêté de 2007 (à savoir le pays de naissance et la nationalité) concernant les hypercubes dont les données ne proviennent que de l'exploitation principale du recensement : au total 2 hypercubes de niveau région sont concernés.

L'exposé détaillera la technique de confidentialité mise en œuvre sur les données du RP européen ainsi que les difficultés rencontrées.

Le traitement de la confidentialité

Insee - Département des méthodes statistiques

Séminaire du 24 juin 2014