Traitement de la confidentialité statistique dans les tableaux : expérience de la Direction des Statistiques d'Entreprises

Julien NICOLAS[1]

Le contrôle de la divulgation statistique est un domaine des statistiques de plus en plus présent ces dernières années. D'un côté, les décideurs demandent en effet aux instituts statistiques de publier des données de plus en plus détaillées ; de l'autre, ces mêmes instituts ont l'obligation légale et morale de garantir la confidentialité des informations qui leur ont été confiées par les répondants aux enquêtes. Cette confidentialité est vitale pour garantir une bonne coopération des répondants, auprès de qui il s'agit de collecter un maximum de données, de la meilleure qualité possible.

Les instituts ont notamment l'obligation de contrôler la divulgation statistique dans les informations qu'ils mettent à disposition, en minimisant le risque que des informations sensibles sur des individus ou des entreprises puissent être divulguées à partir des données diffusées.
La communication, ici résumée, traite de la gestion de la confidentialité uniquement dans le cas des tableaux diffusant des statistiques sur les entreprises ; les méthodes statistiques mises en œuvre sont potentiellement les mêmes que dans le cas des tableaux statistiques sur les ménages, mais la jurisprudence française diffère légèrement d'un cas à l'autre, modifiant quelque peu les règles en vigueur.

La première étape du traitement de la confidentialité statistique dans les tableaux est la recherche des cellules ne devant pas être diffusées au regard des règles définies dans la jurisprudence, et d'éventuelles règles supplémentaires que l'on souhaite voir appliquées.
La jurisprudence définit deux règles à appliquer ; une règle de fréquence minimale, qui oblige à vérifier que les cellules diffusées d'un tableau se rapportent à au moins trois entreprises ; et une règle de dominance, qui oblige à vérifier qu'aucune des entreprises contributrices d'une cellule ne doit contribuer à plus de 85% du total de celle-ci.
Au sein de la Direction des Statistiques d'Entreprises, lorsque les cellules ne respectant pas l'un de ces deux critères sont repérées, elles sont supprimées des tableaux diffusés, ce sont les suppressions primaires. Cependant, cette étape est insuffisante pour garantir la confidentialité des données confiées. En effet, la présence dans les tableaux de totaux de ligne ou de colonne, ou encore de totaux intermédiaires dans le cadre de ventilation de données selon la NAF par exemple (variable hiérarchisée), peut permettre la reconstruction par différence des cellules précédemment supprimées. Il faut alors supprimer de nouvelles cellules pour garantir la confidentialité des premières, ce sont les suppressions secondaires.

Pour gérer au mieux le contrôle de la divulgation statistique dans les tableaux, l'Unité Méthodologie Statistique d'Entreprises a fait l'acquisition du logiciel Tau-Argus et d'un optimiseur.

[1] INSEE, Direction des statistiques d'entreprises