Journées de méthodologie statistique de l'Insee 2025

COLLECTER ET DIFFUSER DES DONNÉES SENSIBLES : QUESTIONS ET EXEMPLES

Olivier LEFEBVRE (*), Eric RANCOURT (**) , Albane GOURDOL (***), Samuel GIVOIS (****), Kévin MILIN (*)

(*) Insee
(**) Statistique Canada
(***) Ined
(****) SSMSI

olivier.lefebvre@insee.fr

La session – animée par Olivier Lefebvre - vise à partager des éléments de réflexion et des exemples de pratiques au sujet des données sensibles. Le sujet n'est pas étranger aux considérations de méthodologie statistique, car les méthodologies de test, de collecte, de traitement et de valorisation des données doivent être adaptées à ces données particulières.

Pourquoi considère-t-on une donnée comme sensible ? Il peut s'agir de critères juridiques, culturels, ou liés au recueil de l'information (difficile pour les enquêtés d'évoquer le sujet) ; une donnée peut également être sensible car elle touche à des sujets potentiellement polémiques ou clivants ; cela jouera sur la manière de la collecter, mais aussi de la diffuser et de la commenter. Dans quelle mesure la sensibilité d'une donnée est-elle liée au contexte, qu'il s'agisse du contexte national, historique ou culturel, ou encore de l'actualité ?

Le décor étant planté, la session abordera deux angles :

Comment collecter des données sensibles ? Quelles questions se poser, et comment y travailler, en termes de contenu de la collecte, de protocole de recueil et de traitement de l'information, de communication ?

Comment diffuser les données sensibles, et comment les valoriser, les contextualiser, comment gérer les risques de mauvaise interprétation ? Se posent bien sûr les questions de protection de la vie privée, mais aussi celles de prévention des risques de mauvaise interprétation, voire de mésusage des résultats et analyses.