

Les enjeux méthodologiques liés à l'usage de bases de sondage imparfaites

Olivier Sautory¹

¹ Insee, 18 boulevard Adolphe Pinard, 75675 Paris Cedex 14, France – olivier.sautory@insee.fr

Thème : Base de sondage, sous-couverture, sur-couverture, doublons, bases multiples

Il est rare que les bases de sondage utilisées par les statisticiens des instituts nationaux possèdent toutes les « bonnes propriétés » requises, telles qu'elles sont présentées au début des ouvrages de sondages. Les imperfections des bases utilisées sont de différents types : différences avec la population-cible de l'enquête (défaut ou excès de couverture), présence de doubles comptes, mauvaise qualité des identifiants, manque d'actualisation des informations présentes dans la base, etc.

Ces imperfections causent bien des difficultés dans la « vraie vie » des statisticiens responsables d'enquête. Malgré la fréquence de ces problèmes, il n'existe pas toujours de méthodologie fermement établie dans ce domaine, et les pratiques des instituts varient beaucoup, en utilisant souvent des méthodes *ad hoc*.

Pour permettre d'obtenir des estimations fiables à partir des échantillons sélectionnés dans ces bases imparfaites, leurs défauts doivent être corrigés au mieux : amélioration de la qualité des bases, mise en œuvre de techniques adaptées au moment de l'échantillonnage, de la collecte, de la phase d'estimation. Certaines enquêtes utilisent simultanément plusieurs bases de sondage, par exemple pour pallier le défaut de couverture.

Ce papier présente les différentes questions posées par ces bases imparfaites, et les principales réponses pratiques ou méthodologiques qui y sont apportées. Les bases de sondage utilisées à l'Insee illustreront le propos.