

L'utilisation des R-indicateurs pour prioriser la collecte des enquêtes ménages : une application à l'enquête Patrimoine 2010

Thomas Merly-Alpa (thomas.merly.alpa@ensae.fr)

Domaine : Méthodologie de collecte

Dans ses enquêtes ménages, réalisées principalement en face-à-face, l'INSEE échantillonne les individus interrogés par un sondage à deux degrés. Chaque unité primaire est une zone d'action enquêteur (ZAE), affectée à un ou plusieurs enquêteurs. L'INSEE peut rencontrer des problèmes logistiques tels que l'absence d'un enquêteur ou la baisse des taux de réponses dans une ZAE en particulier. Il est alors possible de réattribuer d'autres enquêteurs à la ZAE affectée, mais uniquement pour y réaliser un nombre restreint d'enquêtes. Nous avons donc besoin de techniques pour choisir les ménages à interroger en priorité, afin de réduire autant que possible le biais de non-réponse et la perte de précision induite par la baisse du taux de réponse.

Pour ce faire il est possible de s'appuyer sur le concept de R-indicateur, ou indicateur de Représentativité, introduit par Schouten & al. en 2009 [1]. Le R-indicateur permet de mesurer, sans utiliser la variable d'intérêt ou des covariables, le degré de similarité d'un échantillon par rapport à la population de base. Il est basé sur la dispersion des probabilités des ménages échantillonnés à répondre, et, suivant [2], se décline en R-indicateurs partiels permettant d'étudier cette représentativité variable par variable. Ces R-indicateurs sont des outils permettant d'analyser la collecte en isolant des groupes de populations sous-représentées ; de façon analogue à [3], on cible les populations ayant les R-indicateurs les plus grands. Cela permet de lancer une procédure dite de « priorisation » qui consiste à intensifier les efforts de collecte sur les groupes précédemment identifiés.

Cette procédure a été testée sur les données de l'enquête Patrimoine 2010 afin d'une part d'observer l'évolution de la représentativité au cours de la collecte, et d'autre part d'étudier la possibilité de compenser la perte de précision induite par une baisse générale ou localisée des taux de réponse. La procédure étant souple, plusieurs scénarios de priorisation ont été simulés sous des hypothèses de dégradation des taux de collecte, ce qui a permis d'évaluer leurs effets sur la variance de certains indicateurs issus de l'enquête : patrimoine brut moyen, patrimoine net moyen...

Bibliographie

[1] Schouten, B., Cobben, F. and Bethlehem, J. (2009). Indicators for the Representativeness of Survey Response, *Survey Methodol.*, **35**, 101-113.

[2] Schouten, B., Schlomo, N. and Skinner, C. (2011). Indicators for Monitoring and Improving Representativeness of Response, *Journal of Official Statistics*, **27(2)**, 231-253.

[3] Schlomo, N., Schouten, B. and de Heij, V. (2013). Designing Adaptive Survey Design with R-indicators, *New Techniques and Technologies for Statistics Conference 2013*.