

UN SURVOL DES MÉTHODES D'ESTIMATION EN PRÉSENCE DE VALEURS INFLUENTES DANS LES ENQUÊTES

Cyril Favre-Martinoz¹

¹ *Direction de la Méthodologie et de la Coordination Statistique et Internationale, division Sondage, INSEE, 18 Boulevard Adolphe Pinard, 75014 Paris, cyril.favre-martinoz@insee.fr*

Résumé. Les unités influentes peuvent avoir un effet important sur la qualité des estimations. Nous distinguons les valeurs résultant d'erreurs de mesure des vraies valeurs faisant partie de la population à l'étude, appelées par la suite valeurs influentes. Les premières sont habituellement identifiées à l'étape de vérification et sont corrigées, soit manuellement, soit par imputation. Dans cette présentation, nous nous attacherons uniquement au traitement des valeurs influentes. Le problème des valeurs influentes survient particulièrement dans les enquêtes auprès des entreprises qui collectent des variables économiques dont les distributions sont fortement asymétriques. De plus, des unités exhibant des poids extrêmes et certaines erreurs de la base de sondage sont propices à l'apparition d'unités influentes dans l'échantillon. Les valeurs influentes sont problématiques car elles mènent généralement à des estimateurs instables (c'est-à-dire des estimateurs ayant une grande variance). Même s'il est possible de minimiser leur effet au moyen d'un plan de sondage approprié, il est généralement impossible d'éliminer complètement le problème des unités influentes à cette étape. Il est donc souhaitable de développer des méthodes d'estimation robustes à la présence d'unités influentes. Dans cette présentation, nous tenterons de répondre aux trois questions suivantes : (1) Qu'est-ce qu'une valeur influente dans le contexte des enquêtes ? (2) Comment quantifier l'influence d'une unité sur un estimateur ? (3) Comment réduire l'impact des unités influentes à l'étape de l'estimation? Les réponses aux deux dernières questions s'appuient sur le concept de biais conditionnel d'une unité, qui est une mesure d'influence prenant en compte le plan de sondage. Cette présentation repose sur des travaux réalisés en collaboration avec Jean-François Beaumont et David Haziza.

Bibliographie

- [1] Beaumont, J.-F., Haziza, D. and Ruiz-Gazen, A. (2013). A unified approach to robust estimation in finite population sampling. *Biometrika*, 100, 555–569.
- [2] Favre-Martinoz, C., Haziza, D., & Beaumont, J. F. (2015). A method of determining the winsorization threshold, with an application to domain estimation. *Survey Methodology*, 41(1), 57–77.
- [3] Favre-Martinoz, C., Haziza, D., & Beaumont, J. F. (2016). Robust Inference in Two-phase Sampling Designs with Application to Unit Nonresponse. *Scandinavian Journal of Statistics*.
- [4] Moreno-Rebollo, J.L., Muñoz-Reyez, A.M. and Muñoz-Pichardo, J.M. (1999). Influence diagnostics in survey sampling: conditional bias. *Biometrika*, 86, 923–928.
- [5] Moreno-Rebollo, J.L., Muñoz-Reyez, A.M., Jimenez-Gamero, M.D. and Muñoz-Pichardo, J. (2002). Influence diagnostics in survey sampling: estimating the conditional bias. *Metrika*, 55, 209–214.