
ESTIMATION DE VARIANCE SIMPLIFIÉE POUR LES ENQUÊTES À PLUSIEURS DEGRÉS

Guillaume CHAUVET

Univ Rennes, ENSAI, CNRS, IRMAR – UMR 6625

guillaume.chauvet@ensai.fr

Mots-clés : calage, enquête ménage, estimation transversale, non-réponse totale.

Résumé

Les enquêtes auprès des ménages sont souvent sélectionnées selon un plan de sondage à plusieurs degrés. Par exemple, le plan de sondage initial de l'enquête Panel Politique de la Ville [1], réalisée entre 2011 et 2014, peut modulo quelques simplifications être vu comme le résultat d'un plan de sondage à deux degrés. Un échantillon de quartiers est tout d'abord sélectionné, en stratifiant selon le degré d'avancement urbain et avec des probabilités de tirage proportionnelles au nombre de résidences principales. Un échantillon de ménages est ensuite sélectionné dans chaque quartier tiré au premier degré, et tous les individus de 3 ans et plus de ces ménages sont théoriquement enquêtés. Les individus sont suivis pendant quatre vagues d'enquête, avec ajout d'échantillons complémentaires lors des vagues suivantes. Ces ajouts sont réalisés afin de compenser de l'attrition et de permettre de produire des estimations transversales à toutes les vagues d'enquête.

Même dans le cas le plus simple d'une estimation lors de la première vague, l'estimation de variance associée est complexe en raison des différents traitements statistiques. Les poids de sondage des ménages sont ajustés de la non-réponse, en général selon la méthode des groupes homogènes de réponse [2], tout comme les poids individuels. Les poids obtenus sont ensuite calés, généralement de façon simultanée [3], sur des totaux auxiliaires au niveau ménage et au niveau individuel. Pour un plan de sondage à d degrés, la variance d'un estimateur se décompose alors en $d+2$ termes. Les d premiers sont dus aux différents degrés d'échantillonnage. Les deux derniers sont dus à la non-réponse, respectivement de niveau ménage et de niveau individuel.

Dans ce travail, nous étudions les performances de deux estimateurs de variance simplifiés à un seul terme permettant de tenir compte de toutes ces sources d'alea. Ces estimateurs ne nécessitent pas de produire d'estimateur de variance à l'intérieur des unités primaires, ce qui les rend particulièrement simples d'utilisation. Dans le cas d'une réponse complète au niveau ménage et au niveau individuel, le premier estimateur estime correctement la variance due au premier degré, mais sous-estime la variance due aux degrés suivants de tirage [4]. Le second estimateur de variance est conservatif : il surestime généralement la variance de premier degré, mais estime correctement la variance due aux degrés suivants de tirage. Ces estimateurs peuvent être adaptés facilement pour tenir compte de la variance de non-réponse, en utilisant une technique de linéarisation [5]. Les méthodes de variance proposées seront illustrées sur un exemple, et testées par des simulations.

Bibliographie

- [1] Sala M., Chauvet G, « Redresser une enquête longitudinale : le panel politique de la ville », *Journées de Méthodologie Statistique 2018*.
- [2] Juillard H., Chauvet G, « Variance estimation under monotone non-response for a panel survey », à paraître dans *Survey Methodology*.
- [3] Le Guennec, J., Sautory, O., « Une nouvelle version de la macro CALMAR de redressement d'échantillon par calage », *Journées de Méthodologie Statistique 2002*.

[4] Chauvet G, Vallée A.-A. « Consistency of variance estimators for two-stage sampling», travail en cours.

[5] Kim J.K., Kim, J.J., « Non-response weighting adjustment using estimated response probability», *Canadian Journal of Statistics*, vol 35, pp. 501-514.