

# ESTIMATION DE VARIANCE SIMPLIFIÉE POUR UNE ENQUÊTE À PLUSIEURS DEGRÉS - APPLICATION AUX ENQUÊTES RÉPÉTÉES DANS LE TEMPS

Guillaume Chauvet <sup>1</sup>

<sup>1</sup> *Ensai (Irmarr), Campus de Ker Lann, 35170 Bruz, email: chauvet@ensai.fr*

**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. On sélectionne tout d'abord un échantillon de quartiers, puis un échantillon de ménages dans chacun, 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 est complexe en raison des différents traitements statistiques. Les poids de sondage des ménages sont ajustés de la non-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.

Nous étudions les performances d'estimateurs de variance 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. 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 [4]. Le second estimateur de variance est conservatif : il surestime la variance de premier degré, mais estime correctement la variance due aux degrés suivants. Ces estimateurs peuvent être adaptés pour tenir compte de la variance de non-réponse, en utilisant une technique de linéarisation [5]. Le second estimateur de variance présente également l'avantage de pouvoir être reproduit par bootstrap.

**Mots-clés.** bootstrap, groupes homogènes de réponse, linéarisation, non-réponse totale.

**Abstract.** Household surveys are often conducted by means of multistage sampling. For example, the sampling design of the "Panel Politique de la Ville" [1] may be viewed as the result of a two-stage sampling design. A sample of districts is selected first, and then a sample of households is drawn inside these districts, and all the individuals aged 3 or more are theoretically surveyed. The individuals are followed during four waves of survey, with the addition of complementary samples at further waves both to compensate for attrition, and to enable producing cross-sectional estimations at each wave.

Even in the simplest case of an estimation at the first wave, variance estimation is intricate due to the various treatments. The sampling weights of the households are adjusted from unit non-response [2], and so are the individual weights. The resulting weights are then calibrated on auxiliary totals at both the household level and the individual level, usually in an integrative way [3]. For a sampling design with  $d$  stages of sampling, the variance may be split into  $d + 2$  terms. The  $d$  first ones are due to the sampling design, while the 2 last ones are due to the non-response of households and individuals.

We study the behaviour of one-term variance estimators, which enable to account for all the sources of randomness. These estimators do not require variance estimators inside the primary sampling units. In case of full response, the first estimator correctly accounts for the variance due to the first stage of sampling, but underestimated the variance due to further stages [4]. The second variance estimator is conservative: the variance due to the first-stage is overestimated, but the variance due to further stages is correctly accounted for. These variance estimators may be adapted to account for unit non-response, by means of a linearization technique [5]. The second variance estimator may also be easily bootstrapped.

**Keywords.** bootstrap, linearization, response homogeneity groups, unit non-response.

## Bibliographie

- [1] Sala M., Chauvet G (2018). Redresser une enquête longitudinale : le panel politique de la ville. Journées de Méthodologie Statistique.
- [2] Juillard H., Chauvet G (2018). Variance estimation under monotone non-response for a panel survey. A paraître dans Survey Methodology.
- [3] Le Guennec, J., Sautory, O. (2002). Une nouvelle version de la macro CALMAR de redressement d'échantillon par calage. Journées de Méthodologie Statistique.
- [4] Chauvet G, Vallée A.-A. (2018). Asymptotic results for multistage sampling designs, travail en cours.
- [5] Kim J.K., Kim, J.J. (2009). Non-response weighting adjustment using estimated response probability. Canadian Journal of Statistics, vol 35, pp. 501-514.