
ESTIMATION DE VARIANCE DANS LES ENQUÊTES DE L'INSEE : LE PACKAGE R GUSTAVE ET SES APPLICATIONS

Martin CHEVALIER

Insee, Direction de la méthodologie et de la coordination statistique et internationale

martin.chevalier@insee.fr

Mots-clés : calcul de variance, échantillonnage, redressements, calage, méthodologie d'enquête.

Résumé

La mesure de la précision des résultats issus d'enquêtes statistiques constitue un enjeu de plus en plus important dans leur diffusion et leur analyse. Mesure de la qualité du processus de production statistique, les indicateurs de précision figurent dans les rapports communiqués périodiquement à Eurostat ainsi que dans certains règlements européens, en particulier le projet de règlement *Integrated European Social Statistics* (IESS) actuellement en discussion. Ils sont par ailleurs toujours plus mobilisés pour guider le commentaire des résultats d'une enquête, en particulier quand ceux-ci sont ventilés par domaine, secteur et tranche d'effectif dans les enquêtes auprès des entreprises ou région dans les enquêtes auprès des ménages par exemple.

L'imprécision liée à un dispositif d'enquête renvoie à de nombreuses dimensions : celle-ci peut être liée à l'échantillonnage ou à la non-réponse, mais aussi à l'éventuelle imperfection de la base de sondage ou aux erreurs de mesure et d'observation [2]. Cette contribution présente les outils développés au sein du Département des méthodes statistiques de l'Insee pour estimer la variance associée à l'échantillonnage et à la non-réponse dans le cadre méthodologique défini par [5] et [6] et présenté lors des JMS 2015 [4]. Elle porte tout particulièrement sur la mise en œuvre effective de ces méthodes d'estimation sur les enquêtes de l'Insee et leur diffusion auprès des concepteurs d'enquête et des chargés d'études statistiques.

Plusieurs outils ont été proposés pour faciliter la mise en œuvre des estimations de variance dans ce contexte, en particulier la macro SAS Poulpe présentée lors des JMS 1998 ([3] et [7]). Ce type d'outil fait face à une gageure : il doit à la fois présenter un haut niveau de généralité, au sens de sa capacité à prendre en compte des plans de sondage et des redressements complexes, et en même temps rester suffisamment ergonomique pour être utilisé en pratique lors de l'exploitation de l'enquête.

Le package R *gustave*¹ est une nouvelle proposition allant en ce sens. Son ambition est de fournir au méthodologue une « boîte à outils » souple et extensible qui facilite la création de programmes d'estimation de variance autonomes et simples d'utilisation. Ce *package* comporte deux ensembles de fonctions :

- d'une part, des fonctions méthodologiques qui facilitent la mise en œuvre de l'estimation de variance proprement dite (estimateurs de variance, prise en compte du calage sur marges, fonctions de linéarisation [1]) ;
- d'autre part, des fonctions techniques qui créent, à partir des fonctions méthodologiques, des outils d'estimation de variance autonomes et ergonomiques (valeurs par défaut pour les

¹ *Gustave* : a User-oriented Statistical Toolkit for Analytical Variance Estimation, <https://github.com/martinchevalier/gustave>

paramètres techniques, contrôle et validation des données, évaluation non-standard [8] notamment).

Ce faisant, le *package* R *gustave* simplifie la création d'outils d'estimation de variance sans imposer de contrainte *a priori* sur la forme ou la complexité de la fonction de variance utilisée, qui relève de l'expertise du méthodologue. Une fonction de variance « prête-à-estimer » pour les cas les plus simples (tirage à un degré stratifié, prise en compte de la non-réponse et du calage s'il y a lieu) est par ailleurs en cours de développement.

Cette contribution présente le *package* R *gustave* à travers ses applications à deux enquêtes de l'Insee, l'enquête Emploi en continu et le dispositif Statistique sur les revenus et les conditions de vie.

Bibliographie

- [1] Ardilly P., *Les techniques de sondages*, Technip, Paris, 2006, 675 pp.
- [2] Biemer P., « Total survey error: Design, implementation, and evaluation », *Public Opinion Quarterly*, Vol. 74, No. 5, 2010, pp. 817–848
- [3] Caron N., « Le logiciel Poulpe : Aspects méthodologiques », *Actes des journées de méthodologie statistique 1998*, 1998
- [4] Chevalier M., Gros E., Moussallam K., « Les calculs de précision dans Octopusse : Théorie et application à l'enquête Logement 2013 », *Actes des journées de méthodologie statistique 2015*, 2015
- [5] Gros E., Moussallam K. *Les méthodes d'estimation de la précision pour les enquêtes ménages de l'Insee tirées dans Octopusse*, Document de travail de l'Insee M 2015/03, 2015
- [6] Gros E., Moussallam K. *Les méthodes d'estimation de la précision de l'enquête Emploi en continu*, Document de travail de l'Insee M 2016/02, 2016
- [7] Petit J.-N. « Le logiciel Poulpe : Modélisation informatique », *Actes des journées de méthodologie statistique 1998*, 1998
- [8] Wickham H., *Advanced R*, CRC press, 2014, 456 pp. (<http://adv-r.had.co.nz/>)