
MÉTHODES DE CARROYAGE DU RECENSEMENT DE LA POPULATION DANS LES COMMUNES MÉTROPOLITAINES DE 10 000 HABITANTS ET PLUS

Martin Chevalier (*), Gabrielle Gallic (**), Clément Guillo (***),
Gaël Guymarc (**), Céline Pilorge (*)

(*) Insee, Département de la Démographie

(**) Insee, Département de l'Action Régionale

(***) Insee, Département des Méthodes Statistiques

martin.chevalier@sante.gouv.fr

gabrielle.gallic@insee.fr

clement.guillo@insee.fr

gael.guymarc@insee.fr

celine.pilorge@insee.fr

Mots-clés : carroyage, estimation sur petits domaines, imputation.

Domaines concernés : Statistique spatiale – Estimation sur petits domaines, carroyage ; Démographie – Recensement.

Résumé

Dans les communes de moins de 10 000 habitants (dénommées également « petites communes »), l'enquête de recensement est réalisée tous les cinq ans, de façon exhaustive. À l'inverse, les communes de 10 000 habitants et plus (« grandes communes ») sont enquêtées tous les ans par sondage, sur un échantillon d'adresses représentant 8 % des logements. En cumulant cinq enquêtes, l'ensemble des habitants des petites communes et environ 40 % des logements des grandes communes ont été recensés. Chaque année, la population légale de l'ensemble des communes est estimée sur cette base, avec pour référence l'année médiane du cycle. Plusieurs indicateurs sont diffusés à des niveaux infra-communaux (à l'échelle des Iris, des quartiers de la politique de la ville) ou produits pour des utilisateurs spécifiques sur des zonages à façon de plus de 1 000 logements, mais aucune donnée ne l'est encore à la maille des carreaux.

Ce projet de carroyage du recensement est impulsé par une demande européenne qui prévoit, dans le cadre du Recensement européen 2021, la fourniture de différentes variables au niveau de carreaux de 1km de côté pour la France métropolitaine. Cette demande de données carroyées s'inscrit dans une tendance de plus long terme : un projet de livraison annuelle à Eurostat à partir de 2025 ou de 2026 est en cours d'élaboration et une demande de données carroyées à une maille infra-communale plus fine que les Iris s'exprime fortement au niveau national.

Un des principaux enjeux pour répondre à ces demandes est de pouvoir réaliser des estimations fiables sur les carreaux des grandes communes, malgré le faible nombre de

logements échantillonnés dans certains carreaux¹. Comment estimer la population dans un carreau où aucun logement (ou presque) n'a été enquêté par le recensement ? Un groupe de travail a cherché à répondre à cette question et à l'automne 2021, trois méthodes d'estimation de la population sur des carreaux de 1 km de côté ont été présélectionnées.

Cet article présente les méthodes d'estimation envisagées, leurs mécanismes ainsi que les résultats qu'elles apportent. Il compare leurs atouts et faiblesses respectifs au regard de plusieurs critères, comme la performance en termes de biais et de variance, la cohérence avec les données du recensement diffusées à d'autres mailles géographiques, la facilité de mise en œuvre et le caractère généralisable ou non de la méthode.

De façon générale, les travaux méthodologiques menés conduisent à appréhender cette problématique d'estimation comme un problème d'imputation de données à l'adresse. Il s'agit alors d'imputer l'ensemble de la base de sondage du recensement à partir des seules adresses recensées et d'informations auxiliaires. Pour réaliser ces imputations, la base de sondage a été préalablement enrichie par des données fiscales. Deux types de méthodes sont envisagées :

- Une méthode d'imputation par modélisation, qui correspond à une méthode d'estimation du type « petits domaines » de niveau individuel : les données des adresses non enquêtées sont estimées à partir d'un modèle de régression mobilisant les données du recensement et les données fiscales ;
- Une méthode d'imputation par hot deck : les données des adresses non enquêtées sont des répliques de celles d'adresses enquêtées, sur la base du plus proche voisin. L'affectation se fait en minimisant, au niveau de chaque Iris, une distance pondérée entre adresses échantillonnées et non échantillonnées, au sens d'une ressemblance en termes de caractéristiques socio-démographiques connues dans les données fiscales, sous certaines contraintes.
 - Modèle avec cible de logements : chaque adresse échantillonnée donne son nombre de logements estimé dans le recensement et chaque adresse de la base de sondage reçoit exactement son nombre de logements au Répertoire des immeubles localisés (RIL).
 - Modèle avec cible de population : la somme de la population fiscale imputée à une adresse donnée doit être égale à la population fiscale de cette adresse.

Les résultats montrent que pour chacune des méthodes, l'estimation de la population au carreau est plus précise que celle fournie par l'estimateur « classique » du recensement². En particulier, les résultats sont fortement améliorés sur les carreaux comptant peu d'observations.

La méthode par modélisation est la plus performante des trois en termes de précision. Elle répond en revanche moins bien au critère de cohérence avec les données du recensement diffusées sur les autres mailles géographiques que les méthodes hot deck, qui intègrent plus de contraintes. Elle présuppose également de disposer au préalable de la liste des indicateurs que l'on souhaite diffuser. La méthode hot deck avec cible de logements, qui assure une cohérence parfaite avec les données du recensement à la maille des Iris et des communes, est à l'inverse la moins performante des trois en termes de précision. Enfin, la méthode hot deck avec cible de population se situe entre les deux autres méthodes, à la fois en termes de précision et de cohérence avec les données diffusées à la commune.

¹En petite commune, le recensement est exhaustif, donc il n'y a pas de difficulté d'estimation.

²Cet estimateur est calculé à partir des poids de sondage des adresses, calés sur le nombre de logement au RIL, au niveau de l'Iris.

Bibliographie

[1] EUROSTAT (2019), “EU legislation on the 2021 population and housing censuses, explanatory notes”. Theme Population and social conditions, Collection Manuals and guidelines. Février 2019.

[2] Commission Implementing Regulation (EU) 2018/1799 of 21 November 2018 on the establishment of a temporary direct statistical action for the dissemination of selected topics of the 2021 population and housing census geocoded to a 1 km² grid.

OJ L 296, 22.11.2018, p. 19–27.

[3] Ardilly P., « Panorama des principales méthodes d'estimation sur les petits domaines », Documents de travail Insee n° M0602, septembre 2006.