
LA DISJONCTION DES ÉCHANTILLONS DES ENQUÊTES AUPRÈS DES MÉNAGES : DE LA THÉORIE À LA PRATIQUE

Nicolas PALIOD

(*) Insee, Direction de la méthodologie et de la coordination statistique et internationale

nicolas.palioid@insee.fr

Mots-clés : Disjonction d'échantillons, marquage, échantillonnage, charge de collecte, tirages à deux degrés

Domaine concerné : Théorie des sondages amont – Échantillonnage et Bases de sondage

Résumé

La gestion de la charge de collecte auprès des ménages fait partie des bonnes pratiques de la statistique européenne. Le choix réalisé par l'Insee est de retirer de la base de sondage les échantillons tirés pour des enquêtes passées. Ce choix est valable sur le plan méthodologique tant que les unités de la base de sondage ont toutes la même probabilité d'être sélectionnées à chaque nouveau tirage d'échantillon.

La disjonction des échantillons au sein d'un millésime donné de la base de sondage pose peu de difficultés sur le plan théorique. Elle implique le tirage d'un échantillon complémentaire à chaque fois qu'un échantillon tiré pour une enquête nécessite une restriction de champ ou une surreprésentation d'une partie de la population. En outre, le recours à un échantillon d'unités primaires pour les enquêtes en face à face conduit à une gestion de la disjonction des échantillons par unité primaire pour les tirages à 2 degrés. Le taux de logements marqués, c'est-à-dire retirés de la base de sondage suite à des tirages antérieurs, diffère par unité primaire.

Ce mécanisme de disjonction était suffisant jusqu'à une période récente puisque les tirages d'échantillons étaient historiquement réalisés dans les enquêtes annuelles de recensement. Le fait que les enquêtes annuelles de recensement soient disjointes assurait qu'un logement donné ne pouvait être échantillonné qu'une fois tous les 5 ans, à condition de garantir la disjonction des échantillons au sein d'un millésime donné de la base de sondage.

Depuis quelques années, l'Insee s'est orienté vers une utilisation accrue des sources fiscales pour le tirage d'échantillons d'enquêtes auprès des ménages, au détriment des enquêtes annuelles de recensement. Cette évolution s'accompagne de nouveaux enjeux autour de la disjonction des échantillons, en raison du caractère exhaustif des sources fiscales d'une part et de la possibilité nouvelle de réaliser des tirages d'individus dans celles-ci d'autre part.

D'abord, la possibilité de tirer des logements ou des individus pose la question de l'interaction entre les disjonctions d'échantillons de logements et d'échantillons d'individus.

Ensuite, l'exhaustivité de la base de sondage conduit à suivre le marquage des unités d'un millésime de la base de sondage au suivant. Cela nécessite une gestion particulière des logements neufs et des naissances ou arrivées depuis l'étranger d'individus. Ces unités, absentes par nature du millésime précédent de la base de sondage, n'ont aucune chance d'avoir été tirées par le passé et doivent faire l'objet d'un tirage d'échantillon (non enquêté), afin que chaque unité de la base de sondage ait la même probabilité d'avoir été tirée par le passé. Cette propriété est en effet essentielle pour permettre de continuer à disjoindre les échantillons sans introduire de biais d'estimation.

Enfin, la gestion de la disjonction des échantillons d'individus présente une problématique supplémentaire par rapport aux échantillons de logements : les individus sont mobiles et déménagent. Donc, lors du passage à un nouveau millésime, les individus peuvent changer d'unités primaires. Conditionnellement à l'échantillon d'unités primaires sélectionné pour les tirages à deux degrés, les individus d'une même unité primaire n'ont plus la même probabilité d'avoir été tirés pour un échantillon dans le millésime précédent. Ainsi, il n'est pas possible de disjoindre les échantillons d'individus lors du passage à un nouveau millésime de la base de sondage, sans traitement statistique supplémentaire.

Cet article présente l'ensemble des solutions adoptées par l'Insee aux problématiques précédemment évoquées : hypothèses-clés pour que la disjonction d'échantillons ne s'accompagne pas de biais d'estimation, gestion des entrées de champ (logements neufs, naissances, arrivées de l'étranger) lors du passage d'un millésime de la base de sondage à un autre, gestion des déménagements d'individus d'une année à l'autre. Il s'achève sur une présentation de la politique actuelle de disjonction des échantillons à l'Insee et propose quelques éléments chiffrés pour évaluer son efficacité.

Bibliographie

- [1] Crenner E., « Fidéli : un fichier démographique d'origine fiscale au service des utilisateurs », *Séminaire de méthodologie statistique de l'Insee*, novembre 2018.
- [2] Hesse C., « Sampling co-ordination : A review by country », *Documents de travail de l'Insee*, Technical Report E9908, 1999.
- [3] McKenzie R., Gross B., « Synchronised sampling », *Proceedings of ICES II The Second International Conference on Establishment Surveys*, pp 237-244, 2001.
- [4] Merly-Alpa T., Pendoli P.A., Vincent L., « Passer du recensement aux sources fiscales pour le nouvel échantillon-maître de l'Insee : le projet Nautile », *Séminaire de méthodologie statistique de l'Insee*, novembre 2018.
- [5] Ohlsson E., « Coordination of samples using permanent random numbers », *Business Survey Methods*, Wiley, New York, chapter 9, pp. 153-169, 1995.
- [6] Paliod N., « Disjonction d'échantillons d'individus dans les tirages de l'Insee », *11^e Colloque International Francophone sur les Sondages*, 2021.
- [7] Sillard P., Faivre S., Paliod N., Vincent L., « Pour les enquêtes auprès des ménages, l'Insee rénove ses échantillons », *Courrier des statistiques*, 4, pp. 81-100, 2020.
- [8] Vincent L., Chevalier M., Costa L., Delta L., Deroyon T., Favre-Martinoz C., Givois S., Guillo C., Merly-Alpa T., Paliod N., Pendoli P.A., Sauvaget T., « Document de travail Nautile », *Série des documents de travail Méthodologie Statistique de l'Insee*, publication à venir.