

Constitution de l'échantillon-maître pour les extensions régionales : une procédure de tirage d'échantillons équilibrés emboîtés

Marc CHRISTINE¹ et Emmanuel GROS²

La communication de Sébastien FAIVRE aux JMS 2009³ présente une description complète du nouveau système d'échantillonnage des enquêtes ménages⁴ fondé sur les fichiers de logements issus du nouveau recensement « rotatif » mis en place en 2004. C'est dans ce cadre que le projet « Nouveaux systèmes d'échantillonnages », baptisé Octopusse⁵, a été lancé fin 2003. Il repose sur le principe que les listes d'unités échantillonnables sont renouvelées chaque année, via l'apport d'informations actualisées. En contrepartie, cette base n'est plus exhaustive sur l'ensemble du territoire.

Comme les précédents échantillons construits à partir des recensements traditionnels, les nouveaux échantillons seront constitués suivant un schéma directeur commun : celui d'un sondage à deux degrés. Dans un premier temps des unités primaires sont sélectionnées aléatoirement. Elles sont ici baptisées ZAE⁶ et leur construction a nécessité la mise en place d'un processus complexe reposant sur l'affectation (aléatoire) initiale des communes en groupes de rotation pour les besoins du nouveau RP. Dans un second temps, les logements à enquêter sont tirés au sein de ces unités primaires, et ce, pour chaque enquête Ménages.

Les ZAE sélectionnées constituent l'équivalent de l'échantillon-maître traditionnel (EM), qui est la base de sondage principale au sein de laquelle sont sélectionnés les logements à enquêter.

Depuis le recensement de 1999, cette base de sondage nationale s'est vue adjoindre une base supplémentaire : l'échantillon-maître pour les extensions régionales (EMEX). Celui-ci doit être mobilisé dans les cas où des demandeurs régionaux souhaitent la réalisation d'une extension régionale d'enquête nationale afin d'obtenir des résultats spécifiques ou plus précis sur la région.

L'EMEX avait été en fait conçu et construit en 2002, après que l'EM eut été tiré. Ceci a occasionné des difficultés théoriques et pratiques de mise en œuvre non négligeables, nécessitant la conception de solutions statistiques originales⁷. Dans le cadre du nouveau projet

¹ DSDS, Insee - marc.christine@insee.fr

² Dse, Insee - emmanuel.gros@insee.fr

³ « Le projet Octopusse de nouvel échantillon-maître de l'Insee »

⁴ À l'exception de ceux relatifs à l'enquête Emploi, qui reposent sur un système aréolaire et seront désormais construits à partir des fichiers de la taxe d'habitation.

⁵ Organisation Coordonnée de Tirages Optimisés Pour une Utilisation Statistique des Echantillons

⁶ Zones d'Action Enquêteurs.

⁷ « Problèmes théoriques et pratiques de la construction de l'EMEX », Marc CHRISTINE et Laurent WILMS (Insee, DSDS)

Octopusse, il a été décidé, afin d'éviter ces difficultés, de construire l'EMEX de manière coordonnée et simultanée avec la construction de l'EM.

Le présent article détaille la problématique de la construction de l'échantillon-maître pour les extensions régionales : il s'agissait de constituer une base de sondage complémentaire de l'échantillon-maître, c'est-à-dire un ensemble de ZAE additionnelles (celles-ci ayant été construites une fois pour toutes pour l'EM), de même taille, de telle sorte que l'ensemble EM+EMEX possède une représentativité régionale suffisante, tout en préservant les caractéristiques de l'échantillon-maître, en particulier en terme de représentativité nationale. Pour ce faire, une procédure spécifique de tirage en deux phases des ZAE servant à alimenter l'EMEX a été mise au point, conduisant à des échantillons équilibrés emboîtés, représentatifs à chaque niveau de tirage pour les variables d'équilibrage introduites. La constitution de l'échantillon-maître en plusieurs phases plutôt qu'en une seule étape pouvant jouer sur la précision ultérieure des enquêtes, des simulations ont été effectuées afin de quantifier cet impact en terme de précision et de valider la méthode.

Les résultats de ces simulations ont été particulièrement concluants. Forts de ce premiers succès, un raffinement a été introduit sur le concept d'EMEX. En effet, pour la majorité des régions, le choix d'un EMEX de dimension égale à celle de l'EM assure une réserve de logements suffisante pour la réalisation d'enquêtes à extension régionale pendant cinq ans minimum, sans jamais interroger deux fois un même ménage. Cependant, pour certaines régions qui demandent fréquemment d'étendre à leur niveau les résultats d'enquêtes nationales, cette réserve risque de s'avérer insuffisante. En conséquence, il convient de distinguer deux EMEX : un EMEX « restreint », qui est mobilisé seul pour la majeure partie des extensions régionales, et un EMEX « élargi », qui complète la base de logements EM+EMEX restreint, uniquement pour les régions fortement demandeuses d'extensions régionales. La procédure de sélection précédente en deux phases a alors été adaptée sous la forme d'un tirage en trois phases pour permettre la constitution simultanée de l'EM, de l'EMEX restreint et de l'EMEX élargi, garantissant la représentativité des trois ensembles emboîtés obtenus. Des simulations analogues aux précédentes ont confirmé la validité de cette approche en trois phases.