

# L'échantillonnage des enquêtes auprès des ménages dans le recensement annuel de la population

Sébastien Faivre

*Cette fiche s'appuie très largement sur la première partie de l'article d'Olivier Haag « La source fiscale : une base de sondage qui s'impose pour les enquêtes auprès des logements et des individus » présenté au Symposium de Statistique Canada 2013.*

## 1. Description du recensement annuel

Traditionnellement à l'Insee, la base de sondage des enquêtes auprès des ménages est issue du recensement de la population<sup>1</sup>.

Dans le courant de la précédente décennie, l'Insee a décidé de passer du principe de recensement exhaustif de la population française effectué à intervalles de temps quasi-réguliers (7 à 9 ans) à un nouveau mode de recensement rotatif continu. Celui-ci a définitivement été mis en place en janvier 2004.

Désormais, les communes de moins de 10 000 habitants (au dénombrement du RP 1999), ou « petites communes », sont recensées exhaustivement tous les cinq ans par roulement : pour cela, cinq groupes de rotation ont été définis aléatoirement, dans lesquels ont été réparties ces petites communes.

Pour ce qui est des communes comprenant 10 000 habitants ou plus, ou « grandes communes », elles font l'objet d'une enquête de recensement plus complexe, par sondage chaque année au taux moyen de 8 %. Les adresses de ces communes sont réparties aléatoirement entre cinq groupes de rotation disjoints : chaque année, des logements appartenant à un échantillon d'adresses puisé dans le groupe de rotation « actif » sont recensés. Au total, sur un cycle de 5 ans, environ 40 % des logements de la commune sont donc recensés.

## 2. La mise en place d'un nouvel outil de tirage d'échantillon (Octopusse) [1]

Ce changement de méthodologie, s'il induit certes la perte du caractère exhaustif du recensement, offre en contrepartie de nombreux avantages, dont le principal réside dans la fraîcheur des données recueillies : avec cette méthode de collecte, il y a en effet chaque année un recensement exhaustif dans environ 7 000 petites communes et une enquête de recensement par sondage dans environ 900 grandes communes. Du point de vue tant des enquêteurs que des utilisateurs des enquêtes, ce gain de fraîcheur de la base de sondage, obtenu en réduisant au maximum le décalage temporel qui existe entre la date de collecte de l'enquête et la date de recensement, est un plus indéniable.

Par conséquent, **le principe fondamental retenu consiste à sélectionner les échantillons des enquêtes uniquement dans les zones recensées l'année précédente**. Les avantages de ce principe de fraîcheur sont nombreux :

- Il permet de **minimiser le nombre de logements détruits ou en cours de destruction dans les échantillons** ainsi que le nombre de « transformations » d'une résidence principale en résidence secondaire (hors champ), qui sont souvent des points évoqués très négativement par les enquêteurs sur leurs conditions de travail, ainsi que des causes de surcoût pour les enquêtes (car on doit tirer un échantillon plus important pour compenser cela).
- Par ailleurs, la **qualité du ciblage dans les enquêtes de certaines catégories de population** (au moyen de la surreprésentation de ces populations dans les échantillons tirés) **est fortement améliorée** par la fraîcheur de l'information disponible sur les logements de la base de sondage (cela a par exemple permis de mener l'enquête Modes de Garde auprès de parents de jeunes enfants).
- Enfin, il rend possible **une disjonction maximale des échantillons** en garantissant qu'une même feuille de logement ne puisse être sélectionnée qu'une seule fois par Octopusse pour des enquêtes « ménages » au cours d'une période de cinq ans.

---

<sup>1</sup> Dans la suite de la note on parlera de base RP pour désigner cette base de sondage.

Cependant, le changement de contexte lié à la ***mise en place de cette nouvelle méthodologie de recensement*** depuis janvier 2004 entraîne une refonte globale du système actuel d'échantillonnage des enquêtes ménages. En effet, le « nouveau recensement » conduit à une modification radicale des méthodes employées pour la construction des échantillons des enquêtes, parce que la base fournissant les listes d'unités échantillonnables sera renouvelée chaque année en permettant l'apport d'« information fraîche » mais, en contrepartie, ne sera plus exhaustive sur l'ensemble du territoire.

### **3. La mise en place d'unités primaires pour le tirage d'échantillon : les ZAE**

Toutefois, les contraintes territoriales françaises liées à l'organisation de la *collecte en face à face*, à savoir, la nécessité de ne pas trop disperser les lieux d'enquêtes afin de limiter les déplacements des enquêteurs, subsistent encore avec ce nouveau mode de recensement. Elles rendent donc toujours nécessaire la mise en œuvre d'un système d'échantillonnage de type « Échantillon-maître » qui demande la mise en place d'unités primaires pour la collecte.

**Une conséquence directe de ce principe de fraîcheur est la nécessité de repenser la construction des unités primaires (ZAE, Zones d'Action Enquêteur) au sein desquelles seront tirés les échantillons de logements.** En effet, ces zones doivent être construites selon les principes suivants :

- comme par le passé, les ZAE doivent être des zones fixes pour pouvoir leur associer un enquêteur stable dans le temps et localisé à proximité ;
- mais, c'est là que réside la principale nouveauté, elles doivent comporter des logements des cinq groupes de rotation pour pouvoir réaliser des enquêtes chaque année sur un échantillon tiré parmi les logements recensés l'année précédente.

Afin de réduire au maximum l'étendue des ZAE, les règles de constitution de celles-ci ont été établies comme suit :

- constitution des ZAE en respectant les frontières régionales,
- séparation ZAE grandes communes (ZAEGC) et ZAE petites communes (ZAEPC),
- une grande commune constitue à elle seule une ZAEGC,
- au moins 300 résidences principales par groupe de rotation dans les ZAEPC,
- avec l'objectif de minimiser leur étendue.

### **4. Le tirage des échantillons [2]**

L'échantillonnage des enquêtes ménage Insee se fait donc en deux phases :

- sélection d'unités primaires (ZAE de l'échantillon maître)
- puis sélection de logements dans ces ZAE en cherchant à minimiser la dispersion des poids des logements ainsi tirés.

#### **4.1. La sélection des unités primaires de l'échantillon maître**

Dans l'ensemble des ZAE mobilisées traditionnellement par l'Insee on distingue deux catégories : les ZAE exhaustives et les ZAE non exhaustives.

- Les ZAE exhaustives sont au nombre de 79 : 34 grandes communes et les arrondissements de Paris, Lyon et Marseille. Elles correspondent aux 37 plus grandes communes de France selon le nombre de résidences principales. Elles concentrent 18 % de l'ensemble des résidences principales. Compte tenu de la taille des échantillons de l'Insee, ces ZAE exhaustives s'interprètent comme celles dont la sélection est jugée indispensable à la qualité finale de l'estimation. Leur probabilité d'inclusion à l'échantillon maître est donc de 1.
- Au final, l'échantillonnage des ZAE pour constituer l'échantillon maître ne doit s'effectuer que sur les ZAE non exhaustives, soit 3 706. Pour des raisons essentiellement liées aux coûts de collecte l'échantillon maître compte 488 ZAE non exhaustives. La sélection de ces ZAE s'est faite par le biais d'un tirage stratifié par région et équilibré sur des totaux régionaux des 5 groupes de rotation.

## **4.2. La sélection des logements au sein des unités primaires**

Dans chaque unité primaire, le nombre de logements à tirer est calculé de façon à obtenir *in fine* une équipondération des logements de l'échantillon. La sélection de ces unités se fait ensuite de la façon décrite ci-dessous.

Une année donnée, la base de sondage annuelle Octopusse est constituée à partir des listes de logements suivantes issues de la dernière Enquête Annuelle de Recensement disponible :

- Dans les ZAEGC tirées (ou exhaustives), liste des logements de la commune recensés lors de la dernière Enquête Annuelle de Recensement (environ 8 % des logements de la commune).
- Dans les ZAEPC tirées, liste des logements du recensement exhaustif des communes appartenant à la fraction recensée de la ZAEPC (communes de la ZAEPC appartenant au groupe de rotation impacté par la dernière Enquête Annuelle de Recensement disponible).

Le tirage des logements dans les ZAE se fait par un tirage systématique après tri des logements au sein de chaque ZAE. Ces critères de tri sont au nombre maximum de 5, et dépendent du sujet de l'enquête. Ces critères peuvent correspondre à des informations sur le logement (superficie, nombre de pièces, etc.), ou à des caractéristiques du ménage qui y habite (taille du ménage, catégorie socioprofessionnelle du chef de ménage, nombre d'enfants de moins de 5 ans, etc.). Toutes ces informations sont issues du recensement de la population.

## **5. Les limites actuelles de cette base de sondage**

### **5.1. Une base de sondage fluctuante d'une année sur l'autre**

Le nouveau recensement, et principalement la rotation des zones recensées sur 5 ans, complique quelque peu l'échantillonnage. En effet, comme pour des raisons de fraîcheur de l'information il a été jugé préférable de ne mobiliser que la dernière Enquête Annuelle de Recensement pour constituer la base de sondage, la composition communale de la base RP dans les zones rurales peut fortement varier d'une année sur l'autre en termes de niveau d'urbanisme par exemple.

Des alternatives ont donc dû être trouvées pour les échantillonnages d'enquêtes aréolaires (enquête Emploi par exemple avec le recours à une base alternative issue de la source fiscale [3]), voire d'enquêtes annuelles visant à mesurer des évolutions temporelles et nécessitant donc une stabilité en termes de typologie des zones géographiques couvertes (enquête sur le cadre de vie et sécurité par exemple, tirée dans cinq campagnes de recensement). Pour ces enquêtes, une base de sondage stable dans le temps est en effet préférable.

### **5.2. L'impossibilité de tirer des échantillons « individus »**

Les fichiers du recensement de la population ne contiennent aucune donnée nominative. En outre, le système Octopusse est fondé sur une base de sondage de logements ordinaires<sup>2</sup>. Ce système n'est pas conçu a priori pour les tirages d'échantillons d'individus, tirages pour lesquels il faudrait procéder à certains aménagements.

Il a donc été nécessaire, là encore, de trouver une alternative à partir des fichiers d'imposition des personnes de la DGFIP pour le tirage de l'enquête PIAAC (Program for the International Assessment of Adult Competencies) par exemple.

### **5.3. L'absence de certaines variables de stratification dans le recensement de la population**

Dans le cas d'enquêtes très particulières, il peut manquer dans les données du recensement de la population des variables permettant d'identifier (ou de surreprésenter) la population d'intérêt. Il ne figure, par

---

<sup>2</sup> Il ne couvre pas les personnes vivant dans des communautés (cités universitaires, casernes militaires, etc.)

exemple, aucune variable relative au revenu ou au patrimoine des ménages. Dans ce cas il est donc nécessaire de mobiliser des sources externes pour la constitution de la base de sondage.

## Références

[1] Marc CHRISTINE, Sébastien FAIVRE, « Le projet OCTOPUSSE de nouvel Échantillon-Maître de l'Insee », *actes des journées de méthodologie statistique 23, 24 et 25 mars 2009*.

[2] Fabien GUGGEMOS, « Simulation de tirages de zones d'action enquêteurs pour les enquêtes-ménages de l'Insee », *actes des journées de méthodologie statistique 23, 24 et 25 mars 2009*

[3] Vincent LOONIS, « La construction du nouvel échantillon de l'Enquête Emploi en Continu à partir des fichiers de la Taxe d'Habitation », *actes des journées de méthodologie statistique 23, 24 et 25 mars 2009*.



Département des méthodes statistiques  
Version n°1, diffusée le 10 octobre 2017