

Recensement de la population

Les exploitations principale et complémentaire

Les questionnaires renseignés par les personnes recensées et collectés par les agents recenseurs sont saisis par un dispositif de lecture optique et font l'objet d'un vidéo-codage. L'Insee procède ensuite à une série de traitements statistiques : contrôle de la cohérence des informations collectées, imputation des valeurs manquantes, codification des variables de diffusion, calcul des pondérations en vue d'obtenir un fichier de données individuelles, anonymes et apurées, prêt à être exploité.

Comme lors des recensements généraux traditionnels, l'exploitation statistique des questionnaires collectés est réalisée en deux temps : une exploitation dite « principale », puis une exploitation qualifiée de « complémentaire ». Depuis la rénovation du recensement, le délai qui sépare les deux exploitations est très court et les résultats qui en sont issus sont diffusés simultanément.

1. L'exploitation principale

L'**exploitation « principale »** porte sur l'ensemble des questionnaires collectés. Elle est donc exhaustive pour les communes de moins de 10 000 habitants et porte sur environ 40 % des logements dans les communes de 10 000 habitants ou plus. Elle permet de produire un « fichier détail » contenant l'ensemble des logements et des individus recensés. Ces observations sont assorties d'un poids spécifique de l'exploitation principale¹.

L'exploitation principale traite toutes les informations pouvant être codifiées aisément après la saisie des questionnaires. Les résultats statistiques issus de cette exploitation couvrent la plupart des critères d'étude permis par les questionnaires du recensement (bulletin individuel et feuille de logement). Ils portent sur de nombreuses variables et peuvent être classés en 11 thèmes :

- Population (âge, sexe, état matrimonial, actifs ou inactifs...) ;
- Activité des résidents (taux d'activité par sexe et âge, chômage, statut professionnel...) ;
- Emploi au lieu de travail (salariés ou non, temps partiel ou complet, conditions d'emploi...) ;
- Déplacements domicile - lieu de travail ;
- Formation (scolarisation et lieu d'études, diplômes...) ;
- Migrations (lieu de résidence cinq ans ou un an plus tôt, lieu de naissance) ;
- Nationalité (Français de naissance ou par acquisition, nationalité selon le sexe et l'âge...) ;
- Immigration (pays de naissance, année d'arrivée en France...) ;
- Ménages (taille du ménage, sexe et âge de la personne de référence...) ;
- Logements (type, époque d'achèvement...) ;
- Résidences principales (statut d'occupation, date d'emménagement, nombre de pièces, surface, sanitaires, mode de chauffage, nombre de voitures...).

2. L'exploitation complémentaire

La **seconde phase de l'exploitation statistique, dite « complémentaire »**, est destinée à produire les variables dont l'élaboration est complexe. Il s'agit de celles qui décrivent :

- la structure familiale des ménages : détermination précise de la personne de référence du ménage, identification, le cas échéant, de familles au sein du ménage et composition de ces familles ;
- les secteurs d'activité dans lesquels les emplois sont exercés ;
- la profession et la catégorie socioprofessionnelle des personnes, notamment de celles qui exercent un emploi.

¹ Pour en savoir plus, se reporter à la fiche thématique sur « **Les pondérations** »
<http://www.insee.fr/fr/ppp/bases-de-donnees/recensement/resultats/doc/pdf/fiche-ponderation.pdf>

L'élaboration de ces informations nécessite des traitements complexes. Ces traitements sont constitués de procédures automatiques, qui sont complétés par des interventions humaines pour les cas les plus compliqués ou les cas particuliers. Ces traitements sont longs et coûteux, c'est pourquoi ils ne portent que sur un échantillon des questionnaires collectés.

2.1. L'échantillon de l'exploitation complémentaire

L'échantillon destiné à l'exploitation complémentaire est constitué ainsi :

	Communes de moins de 10 000 habitants	Communes de 10 000 habitants ou plus
Ménages	25 % (*)	100 % des ménages enquêtés (soit environ 40 % du total des ménages)
Individus des communautés	25 % (*)	
Habitations mobiles terrestres et personnes sans abri	25 % (*)	100 %
Habitations mobiles fluviales (mariniers)	100 %	

(*) : à partir de l'enquête annuelle de recensement de 2014, le taux de sondage s'élève à 20%

En pratique, le tirage de l'échantillon de l'exploitation complémentaire est réalisé de la façon suivante.

- Sélection des unités statistiques participant au tirage :
 - Les logements ordinaires, y compris les logements de fonction situés dans l'enceinte d'une communauté, ayant au moins un bulletin individuel, c'est-à-dire les **résidences principales** (les résidences secondaires, les logements occasionnels ou vacants ne peuvent donc pas être sélectionnés).
 - Les communautés ayant au moins un bulletin individuel des communautés (BIC) ou un bulletin individuel d'établissement pénitentiaire (BIEP).
 - L'ensemble des habitations mobiles, des personnes sans abri (à chacune on associe un logement fictif) et des mariniers.
 - Les fiches de logement non enquêtées (FLNE) correspondant à des résidences principales.
- Tirage des unités statistiques selon les catégories de population précédemment identifiées :
 - Pour la population hors communauté, l'unité statistique de tirage est le logement ; tous les individus du logement tiré font partie de l'échantillon de l'exploitation complémentaire.
 - Pour la population qui réside en communauté, l'unité statistique de tirage est l'individu. Les communautés dont un BIC au moins est tiré donnent lieu à la création d'un logement fictif.

Cas particuliers :

Les logements et individus enquêtés dans les communes de moins de 10 000 habitants de Corse sont retenus exhaustivement dans l'exploitation complémentaire depuis l'enquête annuelle de 2007. Il en est de même depuis 2006 pour Saint-Pierre-et-Miquelon qui n'est composé que de deux communes de moins de 10 000 habitants.

2.2. Le contenu du fichier de l'exploitation complémentaire

L'exploitation complémentaire consiste à traiter les réponses relatives à l'ensemble des questions posées, notamment celles qui figurent au verso du bulletin individuel et dans les pages centrales de la feuille de logement. Elle produit un fichier qui contient trois types d'observations ou « unités statistiques » : **des logements** et **des individus**, comme dans le fichier issu de l'exploitation principale, ainsi que **des familles**.

Les logements retenus étant des **résidences principales**, le niveau logement accueille aussi les informations relatives au **ménage** : nombre de personnes du ménage, structure familiale du ménage et caractéristiques de la personne de référence du ménage.

Chaque observation est assortie d'un poids spécifique de l'exploitation complémentaire. Les individus et les familles résidant dans un logement ont le même poids, égal au poids du logement.

Le fichier issu de l'exploitation complémentaire contient trois types de variables :

- Les variables présentes dans le fichier de l'exploitation principale et qui, pour les logements et les individus tirés dans l'échantillon de l'exploitation complémentaire, sont recopiées sans modification.
- Les variables « nouvelles », c'est-à-dire spécifiques à l'exploitation complémentaire. Elles se répartissent en deux sous-ensembles :
 - celles relatives à la structure familiale du ménage et aux liens qui unissent les personnes appartenant à un même ménage ;
 - celles relatives à l'activité, à l'emploi et aux professions.
- Les variables de l'exploitation principale qui sont recalculées lors de l'exploitation complémentaire en utilisant les informations supplémentaires disponibles, notamment sur la composition du ménage, la profession ainsi que le secteur d'activité et la localisation de l'établissement où l'activité est exercée. Ces informations supplémentaires permettent d'améliorer la qualité des variables ainsi recalculées.
 - Six variables présentes dans le fichier de l'exploitation principale font l'objet d'une nouvelle codification dans l'exploitation complémentaire : statut professionnel, condition d'emploi, indicateur du lieu de travail, commune du lieu de travail, commune frontalière du lieu de travail et pays du lieu de travail. Dans le fichier issu de l'exploitation complémentaire seules figurent les variables recodifiées ; les variables correspondantes de l'exploitation principale ne sont pas reportées.
 - La détermination de la personne de référence du ménage a lieu dans chacune des deux exploitations. Lors de l'exploitation principale, la personne de référence du ménage est déterminée de façon fruste à partir des seules informations alors disponibles ; elle est repérée par la variable « Indicateur de la personne de référence du ménage ». Lors de l'exploitation complémentaire, la personne de référence du ménage est déterminée de façon plus sophistiquée à partir de la structure familiale du ménage ; elle est repérée grâce à la variable « Lien à la personne de référence du ménage ». Ces deux modes de détermination étant différents, la personne de référence d'un ménage n'est pas toujours la même selon l'exploitation considérée. Ce peut-être une source d'écart dans la répartition des ménages selon les caractéristiques de la personne de référence. L'indicateur calculé lors de l'exploitation principale n'est pas recopié dans le fichier de l'exploitation complémentaire.

2.3. Les résultats tirés de l'exploitation complémentaire

Les résultats socio-économiques qui s'en dégagent peuvent être classés en 12 thèmes :

- les 11 thèmes de l'exploitation principale, modifiés (variables recodifiées) et enrichis de nouvelles variables d'analyse, notamment les catégories socioprofessionnelles, les secteurs d'activité, la structure familiale du ménage ;
- un thème nouveau : **les familles** (couples, familles monoparentales, caractéristiques de la personne de référence de la famille, nombre et âge des enfants...).

En raison du sondage, les résultats tirés de l'exploitation complémentaire ne sont significatifs que pour des **zones géographiques et/ou des sous-populations d'une taille suffisante**².

Cette contrainte, aussi valable pour les résultats issus de l'exploitation principale, est cependant plus forte avec l'exploitation complémentaire dans la mesure où l'ensemble des communes font l'objet d'un sondage et pas seulement les communes de 10 000 habitants ou plus.

3. Choisir entre l'exploitation principale ou l'exploitation complémentaire

Le choix de l'exploitation à privilégier pour les variables recodifiées résulte d'un arbitrage faisant intervenir les critères suivants :

- l'origine des autres variables : si un tableau fait intervenir des variables spécifiques de l'exploitation complémentaire (activité économique, catégorie socioprofessionnelle, structure familiale du ménage...) alors, pour des raisons de cohérence, il convient d'utiliser la variante « complémentaire » des variables recodifiées ;
- la qualité de la variable : le lieu de travail, en particulier, est dans certains cas mieux identifié à l'exploitation complémentaire ;
- la précision des données : pour les communes de moins de 10 000 habitants, elle est meilleure à l'exploitation principale, car l'exploitation complémentaire ne porte que sur un échantillon au quart ; pour les communes de 10 000 habitants ou plus, ce critère n'intervient pratiquement pas dans la mesure où le taux de sondage est le même dans les deux exploitations (sauf pour les individus résidant dans les communautés, qui ne forment qu'une petite minorité de la population) .

Quand les deux exploitations sont mobilisables pour traiter un même thème, le gain de qualité obtenu grâce à l'exploitation complémentaire doit donc être mis en balance avec la perte éventuelle de précision par rapport à l'exploitation principale. C'est notamment le cas pour le lieu de travail (voir la fiche thématique associée : « Les déplacements domicile-lieu de travail » <http://www.insee.fr/fr/ppp/bases-de-donnees/recensement/resultats/doc/pdf/fiche-depl-dom-travail.pdf>).

De façon plus générale, les avantages respectifs de chacune des deux exploitations sont les suivants :

- les résultats issus de l'exploitation principale ont une précision supérieure, surtout pour les communes de moins de 10 000 habitants et les zones qui en englobent³;
- l'exploitation complémentaire fournit des résultats plus complets et améliore la qualité de quelques variables.

² et ³ Pour en savoir plus, se reporter à la fiche thématique sur « **La précision des résultats du recensement** » <http://www.insee.fr/fr/ppp/bases-de-donnees/recensement/resultats/doc/pdf/fiche-precision.pdf>.