

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 exploité¹.

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.

Sommaire

1	L'exploitation principale.....	2
2	L'exploitation complémentaire.....	2
	2.1 L'échantillon de l'exploitation complémentaire.....	3
	2.2 Le contenu de l'exploitation complémentaire.....	3
	2.3 Les résultats tirés de l'exploitation complémentaire.....	5
3	Choisir entre l'exploitation principale et l'exploitation complémentaire.....	5

¹ Pour en savoir plus, se reporter à « *Le traitement des données du recensement de la population* » sur insee.fr

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 plus². Elle permet de produire un « fichier détail » contenant l'ensemble des logements et des individus recensés. Ces observations sont assorties d'un poids spécifique à l'exploitation principale³. Tous les individus résidant dans un même logement ont le même poids, égal au poids du logement.

L'exploitation principale traite toutes les informations pouvant être codifiées aisément après la saisie des questionnaires (comme les cases à cocher, les réponses précausées et les libellés des lieux géographiques et des nationalités). Les résultats statistiques issus de cette exploitation couvrent la plupart des domaines d'étude permis par les questionnaires du recensement (bulletin individuel et feuille de logement), qui sont classés en 7 thèmes sur insee.fr :

- Évolution et structure de la population (âge, sexe...)
- Couples-familles-ménages (état matrimonial, vie en couple, taille du ménage...)
- Logements-migrations résidentielles (catégorie et type de logement, période d'achèvement, statut d'occupation, nombre de pièces, ancienneté d'emménagement, lieu de résidence cinq ans ou un an plus tôt...)
- Diplômes-formation-mobilités scolaires (scolarisation et lieu d'études, diplôme le plus élevé...)
- Population active-emploi-chômage (taux d'activité par sexe et âge, actifs en emploi, chômage au sens du recensement...)
- Caractéristiques de l'emploi-mobilités professionnelles (statut professionnel, condition d'emploi, temps partiel, lieu de travail, mode de transport pour aller travailler...)
- Étrangers-immigrés (distinction Français / étranger et immigré / non immigré par sexe et âge...).

Depuis l'enquête annuelle de recensement 2018, l'ensemble de l'analyse ménages-familles est réalisé dès l'exploitation principale. À partir du recensement de la population millésimé 2020, les résultats statistiques concernant l'analyse ménages-familles seront disponibles à l'exploitation principale.

2 L'exploitation complémentaire

La seconde phase de l'exploitation statistique, dite « complémentaire », est destinée à produire les variables dont l'élaboration est plus 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, qu'elles occupent un emploi ou en recherchent un.

Elles sont issues de procédures automatiques, complétées par des interventions humaines pour les cas les plus complexes ou les cas particuliers. Ces traitements sont longs et coûteux, c'est pourquoi ils ne portent que sur un échantillon des questionnaires collectés.

² Pour en savoir plus, se reporter au paragraphe 1 de la fiche « *Conseils d'utilisation – synthèse* » (« Une nouvelle méthode depuis 2004 »)

³ Pour en savoir plus, se reporter à la fiche thématique « *Les pondérations* »

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	20 % (*)	100 % des ménages enquêtés (soit environ 40 % du total des ménages)
Individus des communautés	20 % (*)	
Habitations mobiles terrestres et personnes sans abri	20 % (*)	100 %
Habitations mobiles fluviales (bateliers)	100 %	

(*) : jusqu'à l'enquête annuelle de recensement (EAR) de 2013, ce taux de sondage s'élevait à 25 %

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 bateliers.
 - 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 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 à 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 :

- Des variables déjà 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.
- Des 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 aux secteurs d'activité, aux professions et aux catégories socioprofessionnelles des personnes.
- Des variables de l'exploitation principale 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'emploi est exercé.

Ces informations supplémentaires permettent d'améliorer la qualité des variables ainsi recalculées. Dans le fichier issu de l'exploitation complémentaire, seules figurent les variables recodifiées ; les variables correspondantes de l'exploitation principale ne sont pas reportées.

Cela concerne :

- Des variables liées à la localisation géographique de l'emploi et à sa caractérisation (statut professionnel, condition d'emploi) ;
- La détermination de la personne de référence du ménage.
 - À partir des résultats du recensement de la population millésimé 2016, la détermination de la personne de référence du ménage a été modifiée⁴. Elle est à ce stade définie uniquement à l'exploitation complémentaire. La nouvelle définition pourra être utilisée à l'exploitation principale à partir des résultats du recensement millésimé 2020.
 - Jusqu'au recensement millésimé 2015, elle a lieu dans chacune des deux exploitations, selon la même philosophie mais pas avec la même richesse d'informations⁴.

Lors de l'exploitation principale, la personne de référence du ménage est déterminée de façon fruste à partir des seules informations alors disponibles (âge, sexe, état matrimonial, type d'activité des personnes du ménage) ; 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 (identifiée notamment grâce aux apports des liens entre les personnes du ménage décrits dans la liste A de la feuille de logement) ; elle est repérée grâce à la variable « *Lien à la personne de référence du ménage* ».

La personne de référence d'un ménage peut donc éventuellement différer d'une

⁴ Pour plus de détail, se reporter à la fiche thématique « *Ménages et familles* »

exploitation à l'autre, mais les cas restent rares⁵ et concernent davantage les ménages nombreux et complexes (ceux abritant plusieurs familles ou des personnes hors famille). Ce peut être une légère source d'écart dans la répartition des ménages selon les caractéristiques de la personne de référence.

2.3 Les résultats tirés de l'exploitation complémentaire

Les résultats tirés de l'exploitation complémentaire permettent d'enrichir les 7 thèmes déjà évoqués pour l'exploitation principale, grâce aux variables recodifiées et aux nouvelles variables d'analyse, notamment les catégories socioprofessionnelles, les secteurs d'activité, la structure familiale du ménage (couples, familles monoparentales, caractéristiques de la personne de référence de la famille, nombre et âge des enfants...).

En mode consultation sur Insee, ces résultats tirés de l'exploitation complémentaire ne sont toutefois proposés que pour les territoires comptant au moins 2 000 habitants car, en raison du sondage, ils ne sont significatifs que pour des zones géographiques et/ou des sous-populations d'une taille suffisante⁶. 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 et 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 (secteur d'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 statut professionnel et la condition d'emploi, en particulier, sont de bien meilleure qualité à l'issue de l'exploitation complémentaire ; le lieu de travail est aussi dans quelques 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 ou au cinquième (voir le paragraphe 2.1) ; 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* »).

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 pour les territoires dont la population est majoritairement issue de communes de moins de 10 000 habitants⁷;**

⁵ Dans seulement 4 % des ménages de deux personnes ou plus présents dans l'exploitation complémentaire du recensement millésimé 2013

⁶ Pour en savoir plus, se reporter à la fiche thématique sur « *La précision des résultats du recensement* »

⁷ Pour en savoir plus, se reporter à la fiche thématique sur « *La précision des résultats du recensement* »

- **l'exploitation complémentaire fournit des résultats plus complets et améliore la qualité de quelques variables.**