

LES SPECIFICITES DU PLAN DE SONDAGE ENRJ

Sébastien FAIVRE¹, Pierre-Arnaud PENDOLI², Emmanuel GROS³, Olivier HAAG⁴ (*)

(*) INSEE, Département des Méthodes Statistiques

Résumé

L'Enquête Nationale sur les Ressources des Jeunes (ENRJ) se caractérise par un plan de sondage à deux degrés et en deux phases. L'échantillon de logements permettant de cibler la population d'intérêt, les individus âgés de 18 à 24 ans, est issu de l'Enquête Annuelle de Recensement 2013. Il a été collecté par le réseau d'enquêteurs de l'Insee entre octobre et décembre 2014.

Du fait des fortes spécificités de la population étudiée (jeunes de 18 à 24 ans), l'enquête ENRJ a nécessité la mise en place d'une méthodologie adaptée avec :

- des travaux spécifiques à partir de la source fiscale pour déterminer la part des hors champ et calibrer les tailles d'échantillons par strate en logements ordinaires en distinguant les jeunes adultes autonomes des jeunes adultes vivant encore chez leurs parents. L'objectif de ces travaux est de déterminer les allocations par unités primaires qui permettent de minimiser la dispersion des poids de sondage lors du tirage de 2nd degré et qui permettent de planifier la charge de travail des enquêteurs.
- La mise en place d'un échantillonnage en communautés avec un protocole de tirage adapté. L'intérêt des travaux effectués est de couvrir l'ensemble de la population des jeunes âgés de 18 à 24 ans dont les résidences principales ne se limitent pas aux seuls logements ordinaires qui constituent le champ principal des enquêtes auprès des ménages.
- Une pondération adaptée basée sur le partage des poids prenant en compte les multiples possibilités de sélection d'un jeune adulte.

Les résultats observés sur la collecte montrent que les cibles de répondants ont été atteintes, ce qui fournit une première validation ex-post des choix méthodologiques retenus. Il ne s'agit cependant encore à ce stade que d'un premier bilan d'étape. Il reste encore en effet à réaliser l'ensemble des traitements post-collecte avant d'avoir une vision définitive de la qualité de l'enquête.

Abstract

The National Survey on youth resources (ENRJ) targets individuals aged 18 to 24. The sample is a draw of dwellings in the 2013 annual census through a two steps and two degrees sampling design. Thus sampled unities (dwellings) and interest unities (individuals) are not the same. As the targeted population is very specific, one had to estimate non-response rates and out of scope rates in order to calculate the numbers of dwellings to draw which are used to estimate pollsters' workload. The sampling frame is complex because it's the joining of ordinary dwellings (houses, flats) and dwellings constituting a community. Moreover, one young adult can be interviewed in several dwellings if sharing

¹ sebastien.faivre@insee.fr

² pierre-arnaud.pendoli@insee.fr

³ emmanuel.gros@insee.fr

⁴ olivier.haag@insee.fr

its time between several homes. As a consequence, the same person can be drawn several times. These configurations have to be processed by using the Weight Share Method.

Mots-clés

Échantillonnage, partage des poids.

Introduction

L'objectif de l'Enquête Nationale sur les Ressources des Jeunes est d'étudier les ressources et les conditions de vie des personnes âgées de 18 à 24 ans inclus au moment de l'enquête. Il s'agit d'une population très diverse (suivant le statut d'activité : chômeurs, actifs, étudiants) et mouvante (déménagements fréquents, multi-résidence, importance de l'habitat occasionnel, des séjours à l'étranger...), ce qui complexifie le protocole d'enquête. En outre, deux catégories de jeunes sont distinguées eu égard à leurs conditions de vie et de ressources :

- les jeunes dits « à charge » ;
- les jeunes dits « autonomes ».

L'unité d'intérêt pour l'étude est donc l'individu de 18 à 24 ans mais l'unité d'échantillonnage est le logement dans lequel il réside, plus particulièrement sa résidence principale.

La base de sondage retenue pour ENRJ est la dernière EAR disponible dans l'Échantillon-Maître OCTOPUSSE. La collecte de l'enquête a eu lieu à l'automne 2014, ce qui a permis le recours à l'EAR 2013⁵ dont l'ensemble des fichiers détails sont disponibles pour utilisation depuis février 2014.

Les résidences principales recensées chaque année ne sont pas homogènes et ne donnent pas lieu au même protocole de recensement. On distingue ainsi :

- les logements ordinaires constituant une résidence principale ;
- les logements en communautés.

Considérant que les ressources et les conditions de vie des jeunes majeurs de moins de 25 ans sont vraisemblablement corrélées à ces deux catégories de logements (ordinaires / en communautés), un échantillon de logements est sélectionné aléatoirement dans chacune de ces deux bases.

L'article proposera tout d'abord d'étudier les spécificités du plan de sondage d'ENRJ pour les logements ordinaires. Dans une seconde partie, la méthodologie d'échantillonnage utilisée pour les communautés sera explicitée. Il s'agit là en effet d'une particularité d'ENRJ par rapport aux autres enquêtes auprès des ménages de l'Institut dont les échantillons se restreignent très souvent uniquement aux logements ordinaires (ou à des catégories de ceux-ci).

Par ailleurs, au sein des logements sélectionnés, les jeunes adultes du champ sont définis comme ceux résidant de façon « habituelle » dans le logement (au moins un mois par an)⁶, ce qui fait qu'un jeune adulte peut être sélectionné via différents logements de la base de sondage.

Le dernier chapitre de cet article donnera un aperçu de la méthode du « partage des poids » appliquée au calcul des pondérations des individus pour prendre en compte leurs multiples possibilités de sélection dans l'échantillon.

⁵ L'EAR 2013 est collectée en janvier et février 2013. Toutes les informations sur les logements recensés et sur les individus y habitant sont millésimées au 1er janvier 2013.

⁶ A l'exception des jeunes « X-Autonomes » partis de chez leurs parents après janvier 2013 et interrogés lorsque plus aucun jeune du champ ne réside de façon habituelle au domicile parental.

1. L'échantillon d'ENRJ pour les jeunes en logements ordinaires

1.1. Un plan de sondage complexe pour estimer la charge de travail des enquêteurs

Comme la plupart des enquêtes auprès des ménages le plan de sondage d'ENRJ en logements ordinaires est en deux phases et à deux degrés.

- Tirage en deux phases : un échantillon dit « de 1ère phase » est sélectionné dans la base de sondage sans restriction de champ ni sur/sous représentation afin de ne pas déséquilibrer la base complémentaire vis-à-vis de telle ou telle catégorie de logements. Dans l'échantillon de 1ère phase, l'échantillon de 2nde phase est sélectionné en appliquant la restriction de champ et la stratification souhaitées.
- Tirage à deux degrés : depuis avril 2009, la plupart des échantillons des enquêtes auprès des ménages de l'Insee sont sélectionnés via l'application OCTOPUSSE parmi les logements présents dans les unités primaires sélectionnées au 1er degré de tirage appelées Zones d'Action Enquêteur (ZAE)⁷. Les ZAE de l'Échantillon-Maître de l'Insee sont au nombre de 567. En moyenne, un enquêteur est affecté à une ZAE. Les allocations de l'échantillon par ZAE constituent donc l'un des paramètres avec, entre autres, le temps de passation du questionnaire, déterminant le temps total de travail de l'enquêteur pour ENRJ. Le calcul des allocations par ZAE doit être considéré avec attention car il s'agit de définir le vecteur d'allocations qui permette de minimiser la dispersion des poids de sondage des logements.

La collecte d'ENRJ débute en octobre 2014. Cette échéance permet que le tirage de l'échantillon soit réalisé dans l'EAR 2013, dernière EAR collectée disponible depuis février 2014.

Néanmoins, cette enquête faisant partie de la charge de travail des enquêteurs pour l'année 2014, il a été nécessaire depuis la mise en place des Nouvelles Conditions d'Emploi des Enquêteurs de l'Insee (NCEE) que la répartition de l'échantillon par ZAE soit connue dès juin 2013. L'objectif étant que les Directions régionales puissent calculer avant le 1er janvier 2014 la charge de travail par enquêteur induite par les enquêtes collectées en 2014 et vérifier que celle-ci est compatible avec les quotités de chaque enquêteur. La livraison aux Directions régionales des allocations par ZAE pour ENRJ a lieu en juin 2013 alors même que la base de sondage de l'échantillon n'est pas encore disponible à cette échéance.

La première spécificité d'ENRJ est qu'elle constitue la première enquête auprès des ménages dont le tirage de l'échantillon s'effectue avec des allocations par ZAE calculées à l'aide des données d'une EAR antérieure. Nous verrons dans la suite de l'article quelle méthodologie est retenue pour le calcul des allocations par ZAE d'ENRJ.

Nous l'avons évoqué plus haut, entre la date du recensement des individus des logements échantillonnés et la date de l'enquête, presque deux années se sont écoulées.

Or, la population des 18-24 ans est une population *a priori* plus mobile que la moyenne car ils vivent une période transitoire de leur vie passant pour la majorité du statut d'enfant à charge à celui d'adulte indépendant. On peut donc raisonnablement s'attendre à ce que leur taux de déménagement soit supérieur à la moyenne.

Ainsi, comme l'unité d'échantillonnage de l'enquête est le logement, il faut s'attendre à ce que les enquêteurs aient plus de logements hors champ (déchets) que ce que l'on mesure habituellement pour les enquêtes auprès des ménages. Ceci a un double impact :

- sur le temps de travail de l'enquêteur à prévoir pour la collecte ;
- sur la taille de l'échantillon de logements à tirer initialement pour obtenir les 4 000 répondants souhaités par les concepteurs de l'enquête.

⁷ Le tirage des ZAE de l'Échantillon-Maître est détaillé par Marc CHRISTINE et Sébastien FAIVRE (2009).

La seconde spécificité du plan de sondage d'ENRJ est alors d'estimer un taux de déchet pour la collecte de l'échantillon. Ce taux de déchet sera *in fine* pris en compte pour le calcul des allocations par ZAE. Pour réaliser cette estimation, il faut être capable de mesurer sur une période de deux ans quelle proportion de logements appartenant au champ d'ENRJ initialement devient hors champ au cours de la période. Cela implique de pouvoir suivre sur plusieurs années la composition par âge des individus de chaque logement. Un tel suivi n'est pas possible à l'aide des EAR car en France un logement n'est recensé qu'une seule fois sur une période de cinq ans. En revanche, il est possible de suivre le parcours des individus entre leurs résidences successives via les sources fiscales que chaque foyer fiscal est tenu d'informer de sa situation fiscale et de sa composition chaque année.

Enfin, l'autre élément qui impacte le nombre de répondants à l'enquête est le taux de réponse. A partir d'expériences de terrain, il est nécessaire d'estimer puis de prendre en compte le taux de réponse anticipé des jeunes de 18 à 24 ans dans le calcul des allocations par ZAE afin que l'échantillon soit de taille suffisante pour garantir les 4 000 répondants souhaités.

1.2. Taux de hors-champ prévisionnel calculé à partir de la TH

Les bases fiscales utilisées pour cette étude sont la TH de 2009 à 2012, les fichiers d'imposition des personnes (FIP) de 2009 à 2012, et l'impôt sur le revenu de 2008 à 2011.

Pour les besoins de l'étude, nous supposons que l'année de collecte de l'enquête est 2012.

Pour chaque année, les logements ont été caractérisés de la façon suivante⁸.

- Logements du champ de l'enquête 2012 : il s'agit des logements contenant au moins un individu appartenant au champ de l'enquête en 2012 (c'est à dire contenant au moins une personne née entre 1988 et 1994). La définition de ces logements est complexe dans la mesure où il est nécessaire de croiser les données de FIP (qui concernent peu ou prou l'ensemble des jeunes de plus de 18 ans) et ceux de la taxe d'habitation qui contient en plus l'année de naissance des enfants de moins de 18 ans. A été considéré comme à charge d'un logement tout enfant résidant de façon exclusive ou alternée dans l'un des foyers fiscaux du logement.
- Logements du champ de l'enquête l'année concernée : il s'agit des logements contenant au moins un individu appartenant au champ de l'enquête pour l'année concernée (pour l'année 2009 par exemple il s'agit des logements contenant au moins une personne née entre 1985 et 1991). L'ensemble des informations nécessaires pour cette indicatrice est disponible dans FIP.

Ces logements ont ensuite été classés parmi les deux catégories suivantes.

- Logements de jeunes indépendants : il s'agit des logements dont la personne de référence (ou son conjoint) fait partie du champ de l'enquête.
- Logements de jeunes à charge : il s'agit des logements dont la (ou les) personne(s) du champ de l'enquête sont rattachée(s) à la personne de référence de l'un des foyers fiscaux du ménage.

Afin d'estimer le taux de hors champ à prévoir pour ENRJ, il s'agit d'identifier dans les années antérieures les logements contenant des jeunes qui appartiennent au champ de l'enquête l'année de la collecte.

Le tableau suivant donne le taux de logements appartenant au champ de l'enquête en 2012 qui sont repérés comme tels une année antérieure donnée (en 2011, en 2010 ou en 2009).

On distingue également les deux types de logements (logements de jeunes « autonomes » versus logements de jeunes « à charge »).

⁸ Un même logement peut appartenir à aucune, à une ou aux deux catégories.

	Taux de logements de même caractéristique (en %)		
	En 2011	En 2010	En 2009
Logements dans le champ en 2012	81.4	70.3	63.9
Logements dans le champ en 2012 et ayant la même personne de référence	80.8	69.6	63.2
Logements de jeunes « autonomes » en 2012	80.7	64.3	55.1
Logements de jeunes « à charge » en 2012	80.9	70.1	63.4

Lecture : 81.4% des logements de l'année 2011 (toujours résidence principale selon la TH 2012) qui contenaient au moins une personne âgée de 17 à 23 ans en 2011 se retrouvent dans le champ de l'enquête en 2012 (c'est-à-dire qu'ils contiennent au moins une personne âgée de 18 à 24 ans en 2012).

Comme on pouvait s'y attendre le nombre de logements qui appartiennent au champ de l'enquête baisse avec le temps mais on retrouve tout de même plus de 70% des logements du champ après deux ans. Un tel chiffre peut être avancé pour prévoir le taux de hors-champ moyen lors de la collecte d'ENRJ puisque les logements, recensés début 2013, seront enquêtés fin 2014.

En outre, le taux de logements contenant des individus « autonomes » au moment de la collecte diminue plus que la moyenne avec le temps. En effet, plus on remonte dans le passé, plus la borne supérieure de la classe d'âge ciblée baisse (elle est de 22 ans en 2009), et moins les jeunes « autonomes » en 2012 ont de chance d'être répertoriés comme tels.

Le tableau suivant donne des statistiques de dispersion de ce taux calculé par département sur l'année 2010.

	Distribution par département				
	Min	1er quartile	Médiane	3 ^e quartile	Max
Taux de logements dans le champ en 2010 et 2012	67	68.6	69.6	70.4	74.6
Taux de logements dans le champ et ayant la même personne de référence en 2010 et 2012	66.7	68.2	69	69.9	74.4
Taux de logements de jeunes « indépendants » en 2010 et 2012	61.1	66.8	68	71.2	75.4
Taux de logements de jeune « à charge » en 2010 et 2012	66.4	68.2	69.2	70	74.1

Afin de réduire au maximum le nombre de logements hors champ des enquêtes, dont le coût de collecte est très élevé, les concepteurs d'ENRJ modifient à la marge le processus de collecte en permettant d'interroger un jeune de 18-24 ans par téléphone s'il n'habite plus dans le logement de ses parents au moment de la collecte⁹, ces derniers étant toujours interrogés face à face. Cette définition moins stricte du champ conduit donc à considérer dans le champ tout logement qui s'y trouve au moment du tirage et pour lequel il n'y a pas eu déménagement des occupants principaux avant la collecte ou bien dont les nouveaux occupants remplissent les critères d'inclusion dans le champ (au moins une personne du THL¹⁰ âgée de 18 à 24 ans ou à défaut au moins une personne du THL ayant un enfant âgé de 18 à 24 ans n'habitant plus dans le logement).

Le protocole de collecte est donc le suivant :

⁹ Cela implique en contrepartie de modifier le questionnaire afin de disposer de tous les éléments pour pouvoir mettre en œuvre *in fine* le partage des poids.

¹⁰ Le THL est le Tableau des Habitants du Logement. Il joue à peu près le rôle du Tronc Commun des Ménages (TCM) de la plupart des enquêtes auprès des ménages, à la différence que le THL définit les liens de chaque habitant du logement uniquement avec l'individu « Kish », tandis que le TCM définit les liens de tous les habitants du logement entre eux.

- si un ou plusieurs jeunes du champ sont dans le THL (jeune « à charge » chez leurs parents), alors on sélectionne un de ces jeunes lors du tirage « Kish » ;
- s'il n'y a plus de jeune du champ résidant habituellement dans le logement (au sens du THL), on demande aux résidants s'il y a des jeunes dans le champ liés au ménage vivant dans un autre logement et on en sélectionne un par tirage « Kish » pour l'interroger par téléphone (et on interroge toujours les parents en face-à-face).

Ceci doit permettre de privilégier au maximum l'interrogation des jeunes en face-à-face (puisqu'on sélectionne toujours un jeune du THL quand c'est possible), tout en limitant le nombre de hors-champ (qui se limitent au cas où le ménage interrogé ne comprend personne dans le champ et n'a pas d'enfant dans le champ résidant dans un autre logement).

Un tel protocole amène ainsi à ce que certains jeunes « à charge » de leur foyer au moment du tirage de l'échantillon soient en fait enquêtés en tant que jeunes « autonomes ». Ce cas est traité par un partage des poids et, comme on le verra plus loin, reste limité étant donné le protocole retenu.

Avec une telle définition du champ de l'enquête, les résultats des taux de logements dans le champ sont désormais les suivants :

	Taux de logements de même caractéristique (en %)		
	En 2011	En 2010	En 2009
Logements dans le champ en 2012	94	89.5	86.7
Logements de jeunes « autonomes » en 2012	82.2	66.4	57.3
Logements de jeunes « à charge » en 2012	96.4	92	88.3

Globalement, cette modification du protocole de l'enquête permet *a priori* d'obtenir un nombre de logements hors champ moins important.

Comme on pouvait s'y attendre, cette définition moins stricte du champ permet de conserver *in fine* dans le champ de nombreux ménages avec jeunes « à charge » au moment du tirage.

Ainsi, il convient de distinguer au moment du tirage les logements de jeunes « autonomes » des logements de jeunes « à charge » afin de leur attribuer des taux prévisionnels de hors champ différents.

En conclusion de cette section, pour le calcul des allocations, il apparaît nécessaire d'anticiper un taux de hors champ de l'ordre de 33% pour la strate des logements de jeunes « autonomes » et de 8 % pour la strate des logements de jeunes « à charge ».

1.3. Estimation des taux de réponse à partir de l'enquête CVS

Les taux de réponse des 18-24 ans à l'enquête Cadre de Vie et Sécurité pour les exercices 2010, 2011 et 2012 ont été calculés par département.

Contrairement à ENRJ, le champ de CVS couvre l'ensemble des logements ordinaires-résidences principales de Métropole, sans restriction d'âge des habitants. L'individu « Kish » répondant au questionnaire-ménage n'est donc pas nécessairement dans la tranche d'âge de 18 à 24 ans. Pour CVS, l'individu « Kish » correspond à la personne de 14 ans ou plus du logement qui fête le premier son anniversaire après la date de collecte dans l'année d'enquête. Ne disposant des informations exactes sur la personne réellement répondante à l'enquête dans chaque logement, il a donc toutefois été possible de reconstituer le tirage « Kish » au sein du ménage à partir des éléments de la base de sondage de CVS (i.e. les EAR).

Ainsi, à partir des taux de réponse des individus « Kish » âgés de 18 à 24 ans aux éditions 2010 à 2012 de CVS, il est possible d'anticiper un taux de réponse des individus de cette tranche d'âge à ENRJ.

Le taux de réponse moyen observé sur l'enquête CVS est plus faible pour les individus « Kish » âgés de 18 à 24 ans que pour les individus « Kish » des autres classes d'âge.

Le tableau suivant donne des statistiques de dispersion de ce taux de réponse par département et par année. L'écart maximum du taux de réponse entre les trois années considérées a été calculé par département et sa distribution figure également dans le tableau ci-dessous. Pour un département donné dep , cet écart maximum se calcule selon l'expression suivante.

$$ECM_{dep} = \max(tx_rep_{dep}^{2010} ; tx_rep_{dep}^{2011} ; tx_rep_{dep}^{2012}) - \min(tx_rep_{dep}^{2010} ; tx_rep_{dep}^{2011} ; tx_rep_{dep}^{2012})$$

		Distribution par département				
		1 ^{er} décile	1 ^{er} quartile	Médiane	3 ^e quartile	9 ^e décile
Taux de réponse des individus « Kish » de 18 à 24 ans à CVS ...	2010	32.6	45.2	58.3	75	95.8
	2011	30	47.2	57.7	76.2	94.4
	2012	33.3	42.9	58.7	72.7	83.3
Écart maximum (ECM) de ce taux de réponse entre les trois années		7.5	15.7	28.5	46.3	60.4

Si la distribution du taux de réponse au niveau national est stable dans le temps, en revanche, on observe une forte variabilité temporelle du taux de réponse au niveau départemental. En effet, l'écart maximal des taux de réponse entre les trois éditions annuelles de CVS dépasse les 28% dans plus de la moitié des départements.

Ainsi, une différenciation du taux de réponse par département ne semble pas nécessairement pertinente. Pour le calcul des allocations on opte donc pour une estimation au niveau France entière du taux de réponse des 18-24 ans à l'enquête. On estime le taux de réponse attendu au questionnaire d'ENRJ par la médiane des taux de réponse rapportés dans le tableau précédent, soit environ 60%.

1.4. Taille totale et par strate de l'échantillon de logements ordinaires

Comme ceci a été préconisé par le groupe de réflexion sur l'échantillonnage dans le cadre du NCEE, le calcul des allocations d'ENRJ (dont le tirage a lieu au printemps 2014 dans l'EAR 2013) est effectué un an plus tôt, en utilisant les données de l'EAR 2008. Le tirage dans l'EAR 2013 doit respecter exactement les allocations par ZAE calculées dans l'EAR 2008.

Les travaux méthodologiques qui justifient le recours à l'EAR N-5 pour anticiper les allocations théoriques par ZAE qui devraient être calculées dans la dernière campagne de Recensement N, sont disponibles sur le site Internet de l'Insee¹¹.

Le premier élément à définir est la taille de l'échantillon. Les concepteurs de l'enquête souhaitent obtenir *in fine* 4 000 répondants. Le nombre total de ménages à interroger pour obtenir ce nombre de répondants peut être obtenu à l'aide de l'expression suivante :

¹¹ cf. <http://www.insee.fr/fr/insee-statistique-publique/default.asp?page=connaître/colloques/sms/sms-adaptation-methodo-enquetes-menages.htm>

Nombre de répondants = Taille de l'échantillon

$$\begin{aligned} & \times (1 - \text{Taux de résidences principales devenues non principales}) \\ & \times (1 - \text{Taux de hors champ prévu parmi les résidences principales}) \\ & \times \text{Taux de réponse estimé des ménages du champ} \end{aligned}$$

Comme on l'a évoqué plus haut, il convient de stratifier l'échantillon de logements ordinaires d'ENRJ entre les ménages de jeunes « autonomes » et les ménages de jeunes « à charge ». Il s'agit donc de déterminer les tailles d'échantillon par strate sachant que l'on souhaite que l'échantillon de répondants soit représentatif de la structure de la population du champ.

D'après le RP 2010, la strate des ménages de jeunes « autonomes » de 18 à 24 ans représente 30% des logements ordinaires-résidences principales, contre 70% pour la strate des jeunes « à charge ». L'objectif est donc d'obtenir un échantillon de répondants comprenant :

- $4\,000 \times 0.7 = 1\,200$ ménages répondants de jeunes « autonomes » ;
- $4\,000 \times 0.3 = 2\,800$ ménages répondants de jeunes « à charge ».

Le taux de réponse attendu peut être estimé à environ 60% selon les résultats de la section 1.3. présentée ci-avant. Il est indépendant de la stratification choisie (logements de jeunes « autonomes » versus logements de jeunes « à charge »).

L'enquête étant collectée au 3^e trimestre 2014 et la base de sondage utilisée correspondant à des informations millésimées au 1^{er} janvier 2013, on peut estimer que le taux de hors champ sera plus proche du taux observé sur la source fiscale entre trois années (2010-2012) qu'entre deux (2011-2012).

Étant donné que le protocole de l'enquête permet d'étendre le champ aux jeunes de 18 à 24 ans ayant déménagé de chez leurs parents depuis l'EAR 2013, sachant qu'il convient de distinguer les logements de jeunes « autonomes » des logements de jeunes « à charge », d'après **les résultats de la section 1.2. le taux de hors-champ anticipé est de :**

- **34% pour les logements de jeunes « autonomes » ;**
- **8% pour les logements de jeunes « à charge ».**

Enfin, on constate également qu'entre deux campagnes annuelles de relevés de la TH, des logements perdent leur statut de résidence principale. Ces logements sont donc *in fine* à considérer également comme hors champ dans le cadre d'ENRJ. On peut estimer **le taux de résidences principales au 1^{er} janvier 2013 devenues non principales au moment de la collecte d'ENRJ à 6%**¹² (i.e. le taux moyen de résidences principales en 2010 devenues non principales en 2012 d'après la TH).

En agrégeant ces différents résultats, les tailles de l'échantillon d'ENRJ pour les deux strates d'intérêt se calculent de la manière suivante.

- **Pour les ménages de jeunes « autonomes » :**

$$\text{Taille de l'échantillon } n = \frac{1\,200}{(1 - 0.06) \times (1 - 0.34) \times 0.6} = 3\,224$$

- **Pour les ménages de jeunes « à charge » :**

$$\text{Taille de l'échantillon } n = \frac{2\,800}{(1 - 0.06) \times (1 - 0.08) \times 0.6} = 5\,396$$

¹² Ce taux est assez stable par département puisque les quantiles à 10%, 25%, 75% et 90% de la distribution de ce taux par département sont respectivement égaux à 4.8%, 5.4%, 7% et 8%.

L'échantillon de logements ordinaires issu du tirage dans l'EAR 2013 compte donc 8 620 logements.

1.5. Calcul anticipé des allocations pour le tirage

On rappelle que le calcul des allocations par ZAE pour ENRJ, dont l'échantillon est sélectionné dans l'EAR 2013, est effectué à partir des données de l'EAR 2008. On distingue deux cas :

- Les ZAE dites « grandes communes » qui sont constituées d'une seule commune de 10 000 habitants impactée chaque année par l'enquête de Recensement.
- Les ZAE dites « petites communes » composées de plusieurs communes de moins de 10 000 habitants, chacune d'elles étant affectée à un unique groupe de rotation du RP. Chaque ZAE « petites communes » comporte des communes des cinq groupes du RP. Une année donnée, seules les communes du groupe de rotation concerné sont impactées par l'enquête de Recensement. Pour ENRJ, en ZAE « petites communes », le calcul des allocations puis le tirage de l'échantillon de logements ordinaires sont effectués dans les communes du groupe de rotation n°5, mais avec un intervalle de cinq ans (EAR 2008 pour l'un, EAR 2013 pour l'autre).

Dans le fichier-détail au niveau Logement provenant de l'EAR 2008, on décompte 552 295 logements ordinaires dans le champ de l'enquête (i.e. contenant au moins un individu de 18 à 24 ans au 01/01/2008). Ces logements, une fois pondérés, représentent 4 212 499 logements en Métropole.

Dans la base de calcul des allocations (logements ordinaires de l'EAR 2008 restreints aux ZAE de l'Échantillon-Maître Insee), on décompte 168 513 logements dans le champ de l'enquête. Ces logements, une fois pondérés, représentent 4 245 991 logements en Métropole.

1.5.1. Stratification

Le tirage d'ENRJ est à deux degrés (tirage de l'échantillon de logements ordinaires dans les ZAE de l'Échantillon-Maître) et stratifié. Dans chaque ZAE, les logements du champ d'ENRJ sont regroupés dans les deux strates suivantes déjà vues plus haut :

- Les logements contenant au moins un individu de 18 à 24 ans et dont la personne de référence aura moins de 37 ans en 2014 (date de naissance à partir du 1^{er} janvier 1978)¹³. Ces logements sont considérés comme contenant des jeunes « autonomes ».
- Les logements contenant au moins un individu de 18 à 24 ans et dont la personne de référence aura 37 ans ou plus en 2014 (année de naissance strictement antérieure à 1978). Ces logements sont considérés comme contenant des jeunes « à charge ».

En population générale, pour un échantillon sans restriction de champ ni sur/sous représentation, ou dans chaque strate pour un échantillon stratifié, le vecteur d'allocations par ZAE est la solution du programme de minimisation de la dispersion des poids de sondage des logements, sous contrainte de taille totale de l'échantillon (et par strate si nécessaire).

Du fait de la proportion de hors champ que l'on anticipe plus importante pour les logements de jeunes « autonomes », ces logements sont surreprésentés dans l'échantillon de logements ordinaires issus du tirage. Leurs poids de sondage sont donc en moyenne moins élevés que ceux des logements de jeunes « à charge ». Toutefois, l'objectif du calcul anticipé des allocations par ZAE à partir de l'EAR 2008 est bien d'obtenir un échantillon de logements équipondérés par strate de tirage.

Afin de respecter la double contrainte

- de taille d'échantillon fixe par strate ;

¹³ Afin de simuler dans l'EAR 2008 la même stratification que celle qui sera utilisée pour le tirage dans l'EAR 2013, les jeunes de 18 à 24 ans sont ceux nés entre le 02/09/1984 et le 01/09/1991 inclus, et l'année de naissance de la personne de référence permettant de distinguer les deux strates est 1973.

- d'allocation fixe par ZAE pour garantir les quotités de travail des enquêteurs calculées par les Directions régionales pour ENRJ,
il est nécessaire de déterminer les allocations du tirage pour chaque croisement ZAE x strate.

1.5.2. Formule de calcul des allocations par ZAE x strate

Au sein de chaque strate, l'allocation par ZAE se calcule de la manière suivante :

$$n_{ZAE}^{strate} = n^{strate} \times \frac{N_{ZAE}^{strate}}{\pi_{ZAE} \times N^{strate}}$$

où

n^{strate} est la taille totale de l'échantillon de la strate : 3 224 pour la strate des logements de jeunes « autonomes », 5 396 pour celle des logements de jeunes « à charge ».

N_{zae}^{strate} est l'estimateur du nombre total de logements de la strate dans la ZAE d'après l'EAR 2008.

$$N_{ZAE}^{strate} = \sum_{\log \in ZAE} (P_{\log} \times 1_{\log \in strate})$$

où P_{\log} est le poids de sondage du logement dans l'enquête annuelle de Recensement 2008.

et $1_{\log \in strate}$ est l'indicatrice d'appartenance à la strate donnée

$$1_{\log \in strate} = \begin{cases} 1 & \text{si } \log \in strate \\ 0 & \text{sinon} \end{cases}$$

π_{zae} est la probabilité d'inclusion initiale de la ZAE dans l'Échantillon-Maître.

N^{strate} est l'estimateur de la taille de la strate dans l'EAR 2008 :

$$N^{strate} = \sum_{ZAE \in EM} \frac{N_{ZAE}^{strate}}{\pi_{ZAE}}^{14}$$

1.6. Le bilan de la collecte

Comme on l'a évoqué précédemment, les jeunes « autonomes » sont surreprésentés dans l'échantillon de logements ordinaires eu égard à la répartition de l'EAR 2008, qui a été utilisée pour calculer de manière anticipée les allocations par ZAE, entre les deux strates d'intérêt.

Concernant la base de tirage de l'échantillon, l'EAR 2013 restreinte aux logements ordinaires du champ, la répartition entre les deux strates d'intérêt est donnée dans le tableau suivant. Il confirme la surreprésentation des logements de jeunes « autonomes » dans l'échantillon.

¹⁴ $N^{strate} = 1\,481\,478$ pour la strate des logements de jeunes « autonomes »

$N^{strate} = 2\,757\,513$ pour la strate des logements de jeunes « à charge ».

Strate de logements de ...	Répartition de ...	
	L'échantillon ENRJ	La base de sondage
... jeunes « autonomes »	37,4% (3 224 FA)	30,46%
... jeunes « à charge »	62,6% (5 396 FA)	69,54%

Dans cette section qui conclut le chapitre relatif à l'échantillon de logements ordinaires, il convient d'apporter les réponses aux questions suivantes :

- La collecte d'ENRJ a-t-elle permis d'atteindre l'objectif de nombre total de répondants ?
- Comment apprécier l'hétérogénéité de la collecte au niveau local (ZAE / Directions régionales) ?
- L'échantillon collecté assure-t-il la représentativité par strate ?

1.6.1. Nombre de FA réussies globalement et par type de ménage

A l'issue de la collecte, 4 622 logements de l'échantillon de logements ordinaires sont considérés comme des fiches-adresses « réussies »¹⁵. Ce résultat permet de dépasser l'objectif minimal de 4 000 jeunes répondants.

La répartition des FA réussies par strate de ménages est donnée dans le tableau suivant.

Strate de logements de ...	Nombre de FA réussies	En pourcentages
... jeunes « autonomes »	1 566	33,88%
... jeunes « à charge »	3 056	66,12%

Les 1 566 FA de jeunes « autonomes » réussies intègrent 310 logements initialement tirés dans la strate des ménages de jeunes « à charge » mais correspondant à des cas de jeunes partis de chez leurs parents depuis leur recensement en janvier 2013. Ces 310 logements correspondent donc en réalité à des cas de jeunes vivant en logement autonome, d'où leur classement final dans la strate correspondante.

Parmi les FA tirées dans la strate des ménages de jeunes « à charge » dont le THL a été renseigné, que l'enquête ait ensuite donné lieu ou non à un questionnaire « jeune adulte » réussi, 357 correspondent en fait à des cas de jeunes « autonomes ». Dans l'échantillon de logements ordinaires issu du tirage, ces 357 FA doivent *in fine* être réaffectées à la strate des ménages de jeunes « autonomes ». Une fois prises en compte ces réaffectations, la répartition de l'échantillon issu du tirage est donnée dans le tableau suivant qui permet de calculer les taux de réponse globaux par strate.

Strate de logements de ...	Échantillon issu du tirage		Échantillon de FA réussies		Taux de collecte ¹⁶ par strate
	Effectifs	%	Effectifs	%	
... jeunes « autonomes »	3 581	41,54 %	1 566	33,88 %	43,73 %
... jeunes « à charge »	5 039	58,46 %	3 056	66,12 %	60,65 %
Total	8 620	100 %	4 622	100 %	53,62 %

Comme cela était attendu, le taux d'enquêtes réussies est plus faible pour les jeunes « autonomes » comparativement aux jeunes « à charge ». Toutefois l'échantillon de FA réussies offre une bonne représentativité des deux catégories de ménages considérés.

¹⁵ Une FA est considérée comme réussie lorsqu'elle a donné lieu à au moins un questionnaire « jeune adulte » (QJA) réussi.

¹⁶ Rapport entre le nombre de jeunes adultes répondants et le nombre de logements tirés

Ainsi on peut dire que l'extension du protocole de collecte aux jeunes de 18-24 ans recensés dans le logement de leurs parents mais qui n'y habitent plus au moment de la collecte a permis d'une part d'atteindre l'objectif de répondants pour la strate des jeunes « autonomes » et d'autre part d'assurer la représentativité de l'échantillon de répondants vis-à-vis des deux strates d'intérêt.

1.6.2. Analyse des taux de réponse par strate

Les ménages de jeunes « autonomes » ont un taux de réponse inférieur de 17 points à celui des ménages de jeunes « à charge ». Cela n'est pas sans conséquence sur les taux de réussite de l'enquête par ZAE.

On choisit la typologie suivante de ZAE :

- Petites ZAE : moins de 10 000 habitants (exclusivement des ZAE regroupant des très petites communes) ;
- ZAE moyennes : de 10 000 à moins de 50 000 habitants ;
- Grandes ZAE : de 50 000 à moins de 100 000 habitants (exclusivement des ZAE « grandes communes ») ;
- Métropole : 100 000 habitants et plus (idem).

Comme le montre le tableau suivant, plus la ZAE est de taille importante et plus les ménages « autonomes » sont représentés dans l'échantillon. Ce dernier point s'explique par le fait que les jeunes « autonomes » de 18 à 24 ans sont souvent des étudiants qui ont quitté le domicile de leurs parents pour leurs études qu'ils suivent le plus fréquemment dans les métropoles régionales ou à Paris.

Typologie de ZAE	Taux de FA de ménages de jeunes « autonomes »	Taux de collecte moyen
Petites ZAE	25,41 %	63,76 %
ZAE moyennes	34,51 %	53,71 %
Grandes ZAE	45,37 %	47,25 %
Métropole	61,94 %	46,58 %

En moyenne, plus la ZAE est de taille importante et plus les ménages de jeunes « autonomes » sont représentés et, à l'inverse, plus le taux de collecte à l'enquête est faible. Cette corrélation n'implique pas nécessairement une causalité, d'autres différences de caractéristiques sociodémographiques des ménages de jeunes entre les quatre type de ZAE pouvant expliquer ces différentiels de taux de réponse. C'est pourquoi il est nécessaire d'analyser les interactions entre la typologie de ZAE et le type de ménage de jeunes.

Typologie de ZAE	Taux de collecte moyen des ménages de jeunes ...	
	... autonomes	... à charge
Petites ZAE	49,80 %	68,32 %
ZAE moyennes	41,29 %	59,35 %
Grandes ZAE	40,91 %	53,25 %
Métropole	40,05 %	56,08 %

Quelle que soit la taille de la ZAE, on observe toujours un différentiel de taux de collecte allant de 13 à 18 points entre les ménages de jeunes « autonomes » et les ménages de jeunes « à charge ». Pour une strate de ménage donnée (jeunes « autonomes » versus « à charge »), les taux de collecte moyens ne semblent significativement différents qu'entre les ZAE de moins de 10 000 habitants et les autres, au bénéfice des premières.

Ces résultats confirment que les ménages de jeunes « autonomes » de 18 à 24 ans répondent intrinsèquement moins que les ménages de jeunes « à charge » de la même tranche d'âge. Leurs modes de vie spécifiques peuvent expliquer que les enquêteurs rencontrent davantage de difficultés à obtenir des questionnaires réussis dans cette strate de ménages : gestion de l'emploi du temps et des

loisirs plus libre, logements plus souvent inoccupés pendant la journée et en soirée (à cause des sorties, des études, du travail étudiant à horaires décalés, ...).

On notera enfin que pour les deux types de ménages, le taux de collecte moyen à l'enquête est supérieur d'environ neuf points chez les ménages résidant dans des ZAE de moins de 10 000 habitants par rapport aux autres ménages. Ce résultat s'explique sans doute par les modes de vie (occasions de sorties moins nombreuses) et les conditions de travail spécifiques (moins d'horaires de travail décalés) dans les zones rurales qui permettent aux enquêteurs de rencontrer moins de difficultés à obtenir des questionnaires réussis que dans les zones plus urbanisées.

1.6.3. Analyse des taux de réponse par DR

Le dernier point rapporté à la section précédente peut expliquer en partie pourquoi la DR d'Île-de-France est la région dont le taux de réussite moyen à l'enquête est le plus faible (mise à part la Corse pour laquelle seuls 23 logements ont été sélectionnés dans l'échantillon ce qui rend toute statistique peu significative). Le tableau ci-dessous récapitule le taux de collecte affiché par chaque Direction régionale ainsi que les pourcentages de logements tirés qui se trouvent dans des ZAE de moins de 10 000 habitants.

Direction régionale Insee	Taux de réponse moyen	Pourcentage de FA appartenant aux ZAE de moins de 10 000 habitants
94 Corse	39,13%	34,78%
11 Île-de-France	41,99%	3,86%
91 Languedoc-Roussillon	46,41%	46,69%
82 Rhône-Alpes	47,92%	44,25%
93 PACA	48,78%	17,02%
26 Bourgogne	50,23%	46,05%
24 Centre	50,46%	44,98%
74 Limousin	51,04%	62,50%
23 Haute-Normandie	51,21%	40,89%
42 Alsace	53,48%	57,14%
72 Aquitaine	54,64%	52,32%
41 Lorraine	55,62%	59,57%
43 Franche-Comté	55,68%	69,19%
73 Midi-Pyrénées	56,30%	45,31%
25 Basse-Normandie	57,53%	60,75%
22 Picardie	57,59%	47,77%
21 Champagne-Ardenne	58,28%	45,31%
83 Auvergne	60,00%	63,78%
54 Poitou-Charentes	61,39%	70,27%
53 Bretagne	62,56%	54,50%
52 Pays de la Loire	63,23%	50,51%
31 Nord-Pas-de-Calais	65,12%	34,88%

On observe que très peu de logements de la DR d'Île-de-France ont été tirés dans des ZAE de moins de 10 000 habitants ; cette DR est la seule qui en compte si peu en proportion.

Les taux de réponse par DR pour ENRJ sont comparables à ceux observés pour les autres enquêtes auprès des ménages réalisées récemment par l'Insee. Les différentiels entre Directions régionales peuvent s'expliquer localement par des difficultés de recrutement d'enquêteurs, mais la structure urbaine des départements¹⁷ constituant la DR d'Île-de-France explique en partie le faible taux de réponse observé comparativement aux autres DR.

¹⁷ Paris, Hauts-de-Seine, Seine-Saint-Denis, Val-de-Marne et Yvelines.

2. Le tirage des jeunes adultes en communautés

2.1. Le calcul des allocations des individus en communautés

2.1.1. Base de sondage

La base de sondage est double. On dispose en effet de :

- la liste des individus de 18 à 24 ans qui résident dans une communauté du champ (hors communautés religieuses et établissements pénitentiaires) selon le RP 2012. Pour chacun de ces individus on dispose de l'identifiant de la communauté dans laquelle il vit ;
- la liste des communautés et quelques caractéristiques (catégorie, capacité, adresse etc.) dans laquelle réside au moins un individu de 18 à 24 ans selon le RP.

Pour le calcul des allocations, ce sont les données du RP 2010 (EAR de 2008 à 2012) qui ont été mobilisées.

On a en effet considéré que la distribution des âges des individus pour une communauté donnée était stable dans le temps et qu'en cas de déménagement une chambre occupée par un jeune de 18 à 24 ans en 2012 le serait toujours en 2014.

Les tableaux suivants donnent des informations quantitatives sur la base de sondage d'une part et sur la base de tirage (Échantillon-Maître) d'autre part.

- Dans la base de sondage France entière selon le RP 2010

Catégorie	Nombre de communautés			Nombre d'individus	
	total	dans le champ ¹⁸	ayant au moins 5 individus dans le champ	dans les communautés du champ	de 18 à 24 ans
Services de moyen ou long séjour	6 275	5 592	2 228	293 941	46 105
Casernes	3 296	2 912	851	51 381	29 399
Établissements hébergeant des étudiants	5 572	5 130	4 529	486 831	312 909
Services de court séjour	203	198	42	6 030	716
Autre	20	20	9	828	114
Total	15 366	13 852	7 659	839 011	389 243

¹⁸ Communautés qui ont au moins un jeune de 18 à 24 ans parmi ses résidents

- Dans la base de tirage (communautés appartenant aux ZAE de l'Échantillon-Maître)

Catégorie	Nombre de communauté		Nombre d'individus		Nombre d'individus pondérés	
	dans le champ	ayant au moins 5 individus dans le champ	dans les communautés du champ	de 18-24 ans	dans les communautés du champ	de 18-24 ans
Services de moyen ou long séjour	2 142	1 033	125 622	24 474	276 961	46 538
Casernes	572	305	18 677	9 931	46 379	26 364
Établissements hébergeant des étudiants	1 831	1 660	218 651	147 838	497 859	328 707
Services de court séjour	99	27	4 421	456	5 900	681
Autre	9	6	442	65	552	132
Total	4 653	3 031	367 813	182 764	827 651	402 422

2.1.2. Stratification

Au vu du nombre d'individus dans le champ, les communautés ont été classées selon deux catégories¹⁹ :

- les établissements hébergeant des étudiants ;
- les autres types de communautés.

Le tirage des communautés s'est fait par région.

Dans chaque strate obtenue par croisement entre les 2 catégories et la région, les communautés ont été triées par type de communautés plus détaillé et par ZAE.

2.1.3. Calcul des allocations

On vise un échantillon de 200 communautés avec 6 répondants par communauté (soit donc un nombre théorique de 1 200 répondants) de façon à garantir un objectif minimal de 1 000 répondants en communautés.

L'objectif des allocations est d'obtenir l'équipondération des individus échantillonnés.

Le nombre de communautés par strate a donc été défini de la façon suivante :

- Calcul d'un nombre d'individus à interroger par strate qui garantisse l'équipondération des individus. Ce nombre a été obtenu par une allocation proportionnelle au nombre d'individus.

$$n_{ind}^h = 1200 \times \frac{N_{ind}^h}{\sum_h N_{ind}^h} \text{ avec } N_{ind}^h = \sum_{ZAE \in EM} \left[\frac{1}{\pi_{ZAE}^{5c}} \sum_{i \in ZAE} 1 \mathbb{I}_{i \in h} \right] \quad (1)$$

¹⁹ Une stratification plus fine conduisait à des allocations trop faibles dans certaines strates. Ceci aurait conduit à surreprésenter les catégories les moins représentées. Le fait de trier par catégorie plus fine au sein des strates qui mélangent plusieurs catégories de communautés permet d'assurer leur « bonne » représentativité au final.

où

$$11_{i \in h} = \begin{cases} 1 & \text{si l'individu "i" appartient à la strate "h"} \\ 0 & \text{sinon} \end{cases}$$

π_{ZAE}^{5c} correspond au poids calé de la ZAE sur des totaux de variables issus des 5 dernières campagnes de Recensement.

- Calcul du nombre de communautés à tirer par strate. Il est obtenu en prenant l'arrondi du nombre d'individus à tirer par strate divisé par 6. Lorsque l'arrondi donne 0, on tire une communauté dans la strate concernée. On a donc en théorie :

$$n_{com}^h = \frac{n_{ind}^h}{6}$$

In fine, on a donc tiré 202 communautés.

Cependant, il est apparu à l'issue de la pré-enquête que 27 communautés de l'échantillon principal étaient hors champ. Ces communautés ont été remplacées par une communauté sélectionnée au sein de l'échantillon de réserve²⁰, au sein de la même région et autant que possible, au sein de la même catégorie de communautés. Au final, on a constitué un échantillon de 198 communautés du champ avec la ventilation suivante par région de gestion et par catégorie de communauté :

Direction régionale	Établissements hébergeant élèves ou étudiants	Autres types de communautés	Total
11 Île-de-France	7	4	11
21 Champagne-Ardenne	9	1	10
22 Picardie	5	1	6
23 Haute-Normandie	5	1	6
24 Centre	13	2	15
25 Basse-Normandie	6	1	7
26 Bourgogne	4	1	5
31 Nord-Pas-de-Calais	8	2	10
41 Lorraine	7	2	9
42 Alsace	3	1	4
43 Franche-Comté	4	2	6
52 Pays de la Loire	12	3	15
53 Bretagne	10	3	13
54 Poitou-Charentes	5	1	6
72 Aquitaine	10	2	12
73 Midi-Pyrénées	10	1	11
74 Limousin	2	1	3
82 Rhône-Alpes	17	5	22
83 Auvergne	7	1	8
91 Languedoc-Roussillon	5	1	6
93 PACA	8	3	11
94 Corse	1	1	2
Total	158	40	198

²⁰ Comportant 4 communautés par région

2.1.4. Principe de tirage des individus mis en œuvre

Afin de respecter l'objectif final d'équipondération des individus tirés en communauté, le tirage des communautés s'est fait de façon proportionnelle à la taille pondérée de la communauté par strate. Ceci permet en effet d'assurer l'équipondération (tout au moins lorsqu'on interroge bien in fine 6 individus dans les communautés sélectionnées) comme le montre la démonstration ci-dessous.

La probabilité d'inclusion d'un individu peut être décomposée de la façon suivante :

$$\pi_{ind} = (\pi_{ZAE}^{5c} \times \mathbb{1}_{ind \in ZAE}) \times (\pi_{com} \times \mathbb{1}_{ind \in com}) \times \frac{6}{N_{com} \times \mathbb{1}_{ind \in com}} \quad (2)$$

où π_{com} est la probabilité d'inclusion de la communauté ;

et N_{com} est le nombre d'individus dans le champ de la communauté « com ».

Comme le tirage des communautés est stratifié et proportionnel à la taille « pondérée » des communautés on a :

$$\pi_{com} = n_{com}^h \times \frac{N_{com}}{\pi_{ZAE}^{5c} \times \mathbb{1}_{com \in ZAE}} \times \frac{1}{\sum_{ZAE \in EM} \left[\frac{1}{\pi_{ZAE}^{5c}} \sum_{com \in ZAE} N_{com} \cdot \mathbb{1}_{com \in h} \right]} \quad (3)$$

En reprenant les notations de la formule (1) ci-dessus on a :

$$\sum_{ZAE \in EM} \left[\frac{1}{\pi_{ZAE}^{5c}} \sum_{com \in ZAE} N_{com} \cdot \mathbb{1}_{com \in h} \right] = N_{ind}^h. \quad \text{Il s'agit en effet du nombre d'individus du champ dans la strate h.}$$

De plus en remplaçant dans la formule (3) la valeur du $n_{com}^h = \frac{n_{ind}^h}{6}$ d'après la formule (1) on obtient²¹ :

$$\pi_{com} = \frac{1}{6} \times (1200 \times \frac{N_{ind}^h}{\sum_h N_{ind}^h}) \times \frac{N_{com}}{\pi_{ZAE}^{5c} \times \mathbb{1}_{com \in ZAE}} \times \frac{1}{N_{ind}^h} = 200 \times \frac{N_{com}}{\pi_{ZAE}^{5c} \times \mathbb{1}_{com \in ZAE} \times \sum_h N_{ind}^h}$$

Donc en remplaçant ceci dans la formule (2) on obtient :

$$\pi_{ind} = (\pi_{ZAE}^{5c} \times \mathbb{1}_{ind \in ZAE}) \times (200 \times \frac{N_{com}}{\pi_{ZAE}^{5c} \times \mathbb{1}_{com \in ZAE} \times \sum_h N_{ind}^h} \times \mathbb{1}_{ind \in com}) \times \frac{6}{N_{com} \times \mathbb{1}_{ind \in com}}$$

Donc²² finalement :

²¹ A l'exception des communautés, en très faible nombre, ayant moins de 6 individus dans le champ.

²² Les indicatrices d'appartenance aux ZAE se simplifient.

$$\pi_{ind} = \frac{1200}{\sum_h N_{ind}^h} \text{ qui est donc indépendant de l'individu.}$$

2.1.5. La sélection en pratique des jeunes à enquêter en communautés

Pour sélectionner les 6 personnes à enquêter, l'enquêteur dispose de rangs d'interrogation générés par tirage aléatoire à probabilités égales (tirage systématique).

Selon les conditions dans lesquelles se déroule l'enquête, en particulier selon que la communauté met à disposition de l'enquêteur ou pas un registre comportant les dates de naissance des personnes résidentes, ces rangs d'interrogation correspondent à des personnes ou à des logements.

- 1) Si l'enquêteur dispose d'un registre comportant les dates de naissance, il est capable d'identifier les jeunes éligibles à ENRJ au sein de cette liste. Il va donc constituer une sous-liste de jeunes éligibles et directement tirer avec les rangs fournis des individus au sein de cette liste.
- 2) S'il dispose d'un registre sans date de naissance, il va avec les rangs fournis tirer des individus dans ce registre, sans savoir s'ils appartiennent au champ de l'enquête
- 3) S'il ne dispose pas de registre, il va avec les rangs fournis tirer des logements (en utilisant les numéros de chambre ou en comptant les logements).

Dans les deux derniers cas, l'enquêteur doit se renseigner auprès des résidents pour savoir s'ils entrent dans le champ ou non.

La pré-enquête réalisée en amont de la collecte par les directions régionales de l'INSEE permet d'identifier dans quel cas se trouve la communauté.

Les enquêteurs disposent de 6 rangs d'interrogation dits « prioritaires » mais aussi de rangs dits « de remplacement ». Ces derniers doivent servir à remplacer :

- les individus ou logements hors champ (cas où l'enquêteur n'a pas pu disposer d'un registre avec les âges)
- ainsi que les absences et les refus.

Le nombre de rangs de remplacement est d'autant plus grand que l'on anticipe un nombre important de « hors champs » parmi les contacts.

Il dépend donc en premier lieu de la mise à disposition d'un registre comportant les dates de naissance. Dans le cas où il n'y a pas un tel registre (pas de dates de naissance ou absence de registre) le nombre de rangs de remplacement dépend du type de communauté et/ou de la part des jeunes éligibles potentiellement présents en son sein.

Si un registre comportant les dates de naissance est mis à disposition, l'enquêteur dispose de 6 rangs prioritaires et de 6 rangs de remplacement, soit 12 rangs d'interrogation au total, tirés au moment de la collecte, en tenant compte du nombre de jeunes éligibles effectivement présents.

Sinon l'enquêteur dispose :

- de 6 rangs prioritaires et de 12 rangs de remplacement, soit 18 rangs d'interrogation au total, dans les communautés où l'on anticipe que la majorité des personnes accueillies sont éligibles (cités universitaires par exemple).
- de 6 rangs prioritaires et de 18 rangs de remplacement, soit 24 rangs d'interrogation au total, dans les communautés où l'on anticipe peu de jeunes éligibles parmi les personnes accueillies.

2.2. Le résultat de la collecte

Au final, 1 039 enquêtes ont été réalisées en communautés, pour 1 328 rangs de contacts générés par les enquêteurs (soit donc un taux de réussite de 78,2%). L'objectif de 1 000 répondants a donc été atteint.

Direction régionale	Nombre d'enquêtes tentées	Nombre d'enquêtes réussies	Taux de réussite
11 Île-de-France	77	51	66,2%
21 Champagne-Ardenne	67	57	85,1%
22 Picardie	42	36	85,7%
23 Haute-Normandie	42	30	71,4%
24 Centre	104	88	84,6%
25 Basse-Normandie	43	35	81,4%
26 Bourgogne	35	28	80,0%
31 Nord-Pas-de-Calais	69	58	84,1%
41 Lorraine	62	51	82,3%
42 Alsace	27	23	85,2%
43 Franche-Comté	40	32	80,0%
52 Pays de la Loire	105	82	78,1%
53 Bretagne	85	72	84,7%
54 Poitou-Charentes	38	31	81,6%
72 Aquitaine	84	64	76,2%
73 Midi-Pyrénées	76	54	71,1%
74 Limousin	15	6	40,0%
82 Rhône-Alpes	148	110	74,3%
83 Auvergne	53	38	71,7%
91 Languedoc-Roussillon	36	30	83,3%
93 PACA	66	52	78,8%
94 Corse	14	11	78,6%
Total	1 328	1 039	78,2%

3. Le calcul du poids de sondage final des jeunes adultes par la méthode du partage des poids

Du fait de la grande mobilité des jeunes adultes, la notion de résidence principale n'est pas adaptée. Les jeunes adultes du champ sont ainsi ceux dont le logement est une « résidence habituelle » (au moins un mois de résidence par an). De ce fait, un jeune adulte pourra être en lien avec plusieurs logements (ordinaires ou en communauté) de la base de sondage.

D'autre part, afin d'éviter un biais de collecte et de rentabiliser l'effort de collecte sur les logements enquêtés, si un logement du champ (donc avec au moins un jeune adulte au RP 2013) ne contient plus de jeune adulte au moment de la collecte (le ou les jeunes adultes présents étant partis depuis du domicile parental), alors un des jeunes partis depuis janvier 2013 est sélectionné pour être interrogé (jeune nommé « X-Autonomie »). Dans ce cas de figure, les parents sont interrogés en face à face et le jeune adulte par téléphone.

3.1. Pondérations du jeune

Les différents configurations possibles sont les suivantes :

Configuration	TYPE DE JEUNE	Parent 1	Parent 2	SITUATION DES PARENTS
1	Vit avec ses parents	vit dans le logement	vit dans le logement (ou parent hors champ)	Parents ensemble (ou un seul parent dans le champ)
2	Vit avec parent 1	vit dans le logement	vit ailleurs	Parents séparés
3	Autonome	vit ailleurs		Parents ensemble (ou un seul parent dans le champ)
4	Autonome	vit ailleurs	vit ailleurs	Parents séparés
5	X-Autonome	vit dans le logement	vit dans le logement (ou parent hors champ)	Parents ensemble (ou un seul parent dans le champ)
6	X-Autonome	vit dans le logement	vit ailleurs	Parents séparés
7	Jeune temporairement à l'étranger	vit dans le logement		

On note $\Pr(L)$ la probabilité de sélection d'un logement L de la base, n_L le nombre de jeunes de 18 à 24 ans habitant le logement L et présents en France au moment de l'enquête, et m_L le nombre de décohabitants post janvier 2013 présents dans le logement L en janvier 2013.

3.1.1. Approche par le calcul des probabilités d'inclusion

On effectue les hypothèses suivantes :

- ❶ on néglige les cas *a priori* rares où le jeune aurait plus de deux lieux de résidence habituelle simultanés ;
- ❷ on néglige également les cas *a priori* tout aussi rares, pour la tranche d'âge des jeunes du champ de l'enquête, de jeunes résidant avec un de leurs parents en janvier 2013 puis parti ensuite résider avec l'autre parent (cas susceptibles de se produire uniquement dans la configuration 2) ;
- ❸ on néglige aussi les cas *a priori* encore plus rares de jeunes dont les parents sont séparés et qui auraient décohabité, postérieurement au 1^{er} janvier 2013, de chez leurs deux parents ;
- ❹ on néglige de plus les cas *a priori* également très rares de jeunes ayant cessé définitivement d'habiter avec leurs parents, postérieurement au 1^{er} janvier 2013 tandis que leurs parents eux-mêmes déménageaient de leur logement ;
- ❺ enfin, on néglige les cas *a priori* également très rares de jeunes temporairement à l'étranger et résidant en France simultanément chez leurs deux parents séparés.

Des premiers comptages réalisés sur les données d'enquête permettront de valider ou d'infirmer ces hypothèses. À noter que le fait de relâcher une ou plusieurs de ces hypothèses ne changerait pas fondamentalement la méthodologie de calcul des probabilités de sélection présentée ci-dessous. Cela impliquerait juste de distinguer des sous-cas supplémentaires au sein de certaines configurations, et

de compléter les probabilités d'inclusion π_j par l'ajout de termes du type $\frac{\Pr(L_3)}{n_{L_3}}$ ou

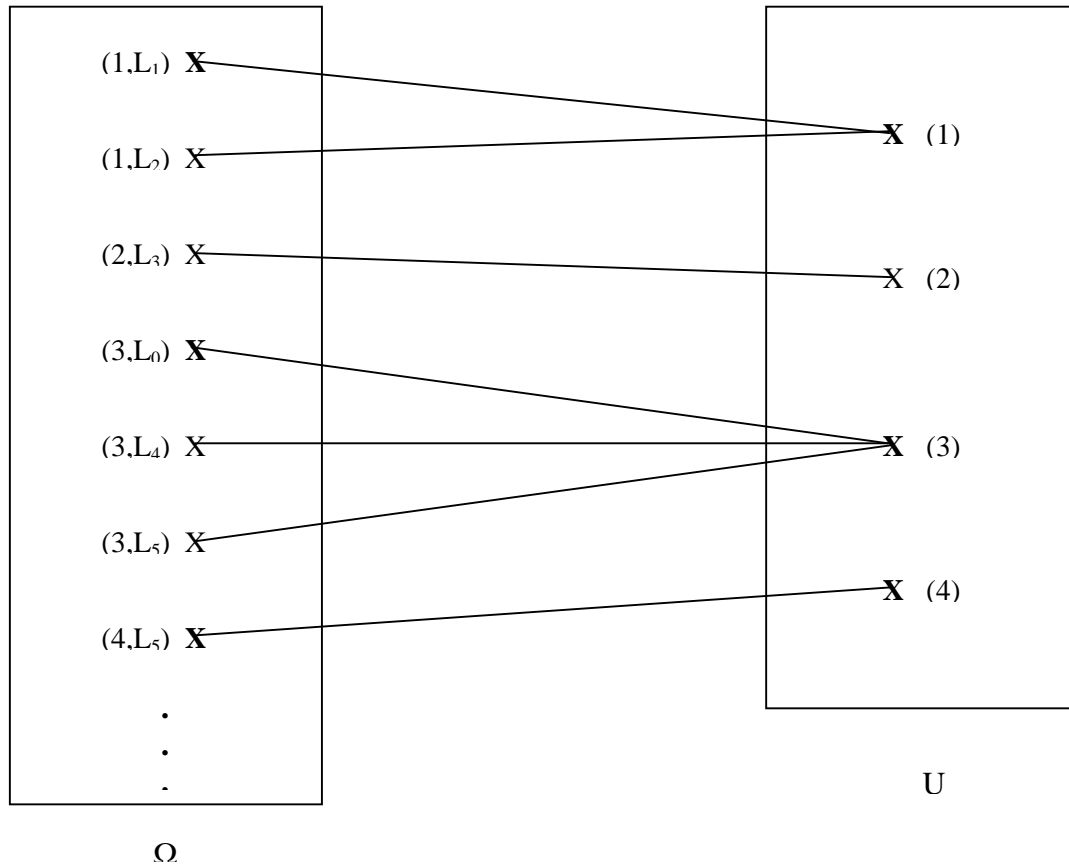
$$\frac{\Pr(L_{0_bis})}{m_{L_{0_bis}}} \prod_{n_{L_{0_bis}}=0}.$$

Remarque : dans la suite, le terme « logement indépendant » désigne un logement hors logement(s) des parents. Il peut donc s'agir soit d'un logement classique, soit d'une communauté.

- Configuration 1 et 2 : les jeunes concernés par ces configurations ne peuvent pas être décohabitants post janvier 2013, par construction dans la configuration 1, et suite à l'hypothèse ② dans la configuration 2. Donc :
 - soit le jeune n'a pas d'autre lieu de résidence habituelle que le logement de ses parents (noté L_1), et sa probabilité de sélection est alors $\pi_i = \frac{\Pr(L_1)}{n_{L_1}}$;
 - soit le jeune a un autre lieu de résidence habituelle (noté L_2), et sa probabilité de sélection est alors $\pi_i = \frac{\Pr(L_1)}{n_{L_1}} + \frac{\Pr(L_2)}{n_{L_2}}$.
- Configuration 3 et 4 :
 - si le jeune n'est pas un « décohabitant post janvier 2013 » :
 - soit le jeune ne réside que dans un seul logement indépendant L_1 , et sa probabilité de sélection est alors $\pi_i = \frac{\Pr(L_1)}{n_{L_1}}$;
 - soit le jeune réside, en sus du logement L_1 , dans un logement indépendant L_2 , et sa probabilité de sélection est alors $\pi_i = \frac{\Pr(L_1)}{n_{L_1}} + \frac{\Pr(L_2)}{n_{L_2}}$;
 - si le jeune est un « décohabitant post janvier 2013 » : en notant L_0 le logement dont le jeune a décohabité après le 1^{er} janvier 2013 – logement des parents dans la configuration 3, logement d'un de ses deux parents dans la configuration 4 – :
 - soit le jeune ne réside que dans le logement indépendant L_1 via lequel il a été sélectionné, et sa probabilité de sélection est alors $\pi_i = \frac{\Pr(L_1)}{n_{L_1}} + \frac{\Pr(L_0)}{m_{L_0}} \mathbb{I}_{n_{L_0}=0}$;
 - soit le jeune réside, en sus du logement L_1 , dans un logement indépendant L_2 , et sa probabilité de sélection est alors $\pi_i = \frac{\Pr(L_1)}{n_{L_1}} + \frac{\Pr(L_2)}{n_{L_2}} + \frac{\Pr(L_0)}{m_{L_0}} \mathbb{I}_{n_{L_0}=0}$.
- Configuration 5 et 6 : en notant L_0 le logement dont le jeune a décohabité après le 1^{er} janvier 2013 – celui de ses parents dans la configuration 5, celui d'un de ses parents dans la configuration 6 – :
 - soit le jeune n'a pas d'autre lieu de résidence habituelle que son logement indépendant L_1 , et sa probabilité de sélection est alors $\pi_i = \frac{\Pr(L_1)}{n_{L_1}} + \frac{\Pr(L_0)}{m_{L_0}} \mathbb{I}_{n_{L_0}=0}$;
 - soit le jeune réside en sus de manière habituelle dans autre logement indépendant L_2 , et sa probabilité de sélection est alors $\pi_i = \frac{\Pr(L_1)}{n_{L_1}} + \frac{\Pr(L_2)}{n_{L_2}} + \frac{\Pr(L_0)}{m_{L_0}} \mathbb{I}_{n_{L_0}=0}$.
- Configuration 7 : en notant L_1 le logement du/des parent(s) sélectionné, la probabilité de sélection du jeune est alors $\pi_i = \Pr(L_1)$.

3.1.2. Approche par partage des poids

Si l'on excepte la configuration 7, pour laquelle le poids du jeune est toujours l'inverse de la probabilité de sélection du logement de ses parents, les configurations 1 à 6 relèvent du cadre classique du partage des poids. Le schéma suivant traduit ces différentes configurations possibles dans le cadre d'un exemple simplifié :



La base de sondage Ω de l'échantillon est composée des couples (jeune, logement), tandis que la population d'intérêt U est celle des jeunes. On a représenté en gras souligné dans Ω un échantillon fictif d'unités sélectionnées lors du tirage de l'échantillon ω , et dans U l'échantillon de jeunes s en résultant.

Le jeune (1) est un jeune multi-résident non décohabitant post janvier 2013, sélectionné dans l'échantillon s via sa sélection dans le logement L_1 lors du tirage de ω . Le jeune (2) est mono-résident et non sélectionné dans s puisque son logement n'a pas été retenu dans ω . Le jeune (3) est un jeune multi-résident décohabitant, appartenant à l'échantillon s via sa sélection dans le logement L_0 de ses parents lors du tirage de ω (jeune X-autonome donc). Enfin, le jeune (4) est un jeune mono-résident appartenant à l'échantillon s via sa sélection dans le logement L_5 lors du tirage de ω .

Lors du tirage de ω , le poids de tirage d'une unité (i, L_j) est :

$$\theta_{i,L_j} = \begin{cases} \frac{n_{L_j}}{\Pr(L_j)} & \text{si le jeune } i \text{ réside dans le logement } L_j \\ \frac{m_{L_j}}{\Pr(L_j)} & \text{si le jeune } i \text{ est décohabitant post janvier 2013 pour } L_j \text{ et si } n_{L_j} = 0 \end{cases}$$

Le poids issu de la méthode du partage des poids pour un jeune i de l'échantillon S est alors :

$$w_i = \frac{\sum_{j \in \omega} \theta_{i, L_j} \mathbb{1}_{i \text{ réside dans } L_j \text{ ou a décohabité de } L_j}}{\underbrace{\sum_{j \in \Omega} \mathbb{1}_{i \text{ réside dans } L_j \text{ ou a décohabité de } L_j}}_{\text{Nombre de logements dans lesquels réside le jeune } i \text{ ou desquels il a décohabité}}}$$

En supposant qu'aucun jeune n'a pu être sélectionné deux fois (ou plus) dans l'échantillon ω via deux (ou plus) logements différents, le poids d'un jeune de l'échantillon ENRJ s'obtient alors – hors cas particulier de la configuration 7 déjà évoqué –, en divisant son poids de tirage – intégrant le tirage Kish – par le nombre de logements auxquels il est lié, soit parce qu'il y réside, soit parce qu'il en a décohabité.

Sous les hypothèses ❶ à ❺, et en notant nb_log_i le nombre de logements auxquels le jeune i est lié, les poids des jeunes issus de la méthode de partage des poids seront donc :

- Pour les configurations 1 à 4, en notant L_1 le logement sélectionné dans lequel réside le jeune i ,

$$w_i = \frac{n_{L_1}}{nb_log_i \cdot Pr(L_1)} \text{ avec } nb_log_i = \begin{cases} 1 & \text{si le jeune est mono-résident non décohabitant} \\ 2 & \text{si le jeune est multi-résident non décohabitant ou mono-résident décohabitant} \\ 3 & \text{si le jeune est multi-résident décohabitant} \end{cases}$$

- Pour les configurations 5 et 6, en notant L_0 le logement sélectionné duquel le jeune a décohabité,

$$w_i = \frac{m_{L_0}}{nb_log_i \cdot Pr(L_0)} \text{ avec } nb_log_i = \begin{cases} 2 & \text{si le jeune est X-autonome mono-résident} \\ 3 & \text{si le jeune est X-autonome multi-résident} \end{cases}$$

- Pour la configuration 7, en notant L_1 le logement du/des parent(s) sélectionné, $w_i = \frac{1}{Pr(L_1)}$.

3.2. Pondérations des parents / ménages de parents/beaux-parents

3.2.1. Pondération d'un parent

Si l'on excepte le cas des parents ayant un enfant de 18 à 24 ans temporairement à l'étranger²³, la probabilité de sélectionner dans l'échantillon de jeunes deux enfants d'un même parent biologique est quasiment nulle, et la probabilité de sélection d'un parent est donc égale à la somme des probabilités de sélection de ses enfants éligibles à l'enquête.

Pour les parents ayant un enfant de 18 à 24 ans temporairement à l'étranger, la probabilité de sélection d'un parent est égale à la probabilité de sélection de son enfant temporairement à l'étranger plus la somme des probabilités de sélection de ses enfants éligibles à l'enquête ne résidant pas dans le même logement que le jeune temporairement à l'étranger.

²³ On néglige ici le cas, a priori extrêmement rare, de parents ayant plusieurs enfants de 18 à 24 ans temporairement à l'étranger.

3.2.2. Pondération d'un ménage de parents/beaux-parents

En suivant le même raisonnement, et en notant J_0 l'ensemble des enfants éligibles à l'enquête communs aux deux parents, et J_1 (resp. J_2) l'ensemble des enfants propres au parent 1 (resp. 2), la probabilité de sélection d'un ménage M n'ayant aucun enfant de 18 à 24 ans temporairement à l'étranger est alors :

$$\pi_M = \sum_{i \in J_0} \pi_i + \sum_{i \in J_1} \pi_i + \sum_{i \in J_2} \pi_i$$

Cette formule est à adapter de façon similaire à ce qui est fait au point 3.2.1. pour le cas des ménages ayant un enfant de 18 à 24 ans temporairement à l'étranger.

3.3. Premiers résultats sur le nombre de liens des jeunes adultes interrogés

Parmi les 5 785 jeunes adultes répondants, 3 870 (66,9%) ont un lien avec la base de sondage, 1 767 (30,5%) ont deux liens avec la base de sondage et 148 (2,6%) ont trois liens ou plus avec la base de sondage. Lorsqu'on s'intéresse à la distribution du nombre de liens sur certaines sous-populations, on obtient les résultats suivants :

Genre / Nombre de liens	1	2	3+
Homme	65,4%	32,0%	2,6%
Femme	68,6%	28,9%	2,5%
Total	66,9%	30,5%	2,6%

Âge / Nombre de liens	1	2	3+
18	49,9%	44,9%	5,2%
19	57,0%	39,7%	3,2%
20	64,4%	33,8%	1,8%
21	71,9%	25,7%	2,4%
22	76,4%	22,5%	1,1%
23	83,9%	15,5%	0,6%
24	90,1%	9,6%	0,3%
Total	66,9%	30,5%	2,6%

Type de logement / Nombre de liens	1	2	3+
Ordinaire	77,7%	21,0%	1,4%
Communauté	26,4%	66,6%	7,1%
Total	66,9%	30,5%	2,6%

On constate notamment un effet important de l'âge sur le nombre de liens, puisque la moitié des jeunes adultes de 18 ans ont au moins deux liens avec la base de sondage contre moins de 10% des jeunes adultes de 24 ans.

D'autre part, le type de logement joue également fortement puisque 74% des jeunes adultes vivant en communauté ont au moins deux liens avec la base de sondage contre seulement 22% des jeunes adultes vivant en logements ordinaires.

A noter que les jeunes X-Autonomes²⁴ ont pour 94% d'entre eux un seul lien avec la base de sondage.

Bibliographie

- [1] CHRISTINE Marc et FAIVRE Sébastien, « Le projet OCTOPUSSE de nouvel Échantillon-Maître de l'Insee », pp. 9 - 12, *JMS*, 2009
- [2] PENDOLI Pierre-Arnaud, « La nouvelle planification de l'échantillonnage », 2014

²⁴ Ayant quitté le domicile parental depuis janvier 2013.