

Les spécificités du plan de sondage ENRJ

Sébastien Faivre & Pierre-Arnaud Pendoli
INSEE - Division Sondages



Mesurer pour comprendre



Plan de la présentation

Introduction

I. L'échantillon pour les jeunes en logements ordinaires

II. L'échantillonnage en communautés

III. Poids de sondage finaux des jeunes adultes

Conclusion

Introduction

Présentation de l'enquête

Champ de l'enquête : les jeunes adultes âgés de 18 à 24 ans inclus au moment de la collecte

Période de la collecte : octobre à décembre 2014

Tirage de l'échantillon : printemps 2014

Base de sondage : la dernière enquête de Recensement disponible au moment du tirage = l'EAR 2013

⇒ unité d'intérêt = l'individu

unité d'échantillonnage = le logement

Deux bases de sondages

Les jeunes de 18 à 24 ans peuvent être atteints de deux manières dans l'EAR 2013 :

- Via les logements ordinaires
- Via les logements en communautés

Conséquence : tirage de deux échantillons distincts

Spécificité d'ENRJ : tirage d'un échantillon de communautés

Impact méthodologique

Les jeunes de 18 à 24 ans sont une population mouvante

Exemple :

- Jeune dans le champ recensé chez ses parents en 2013 (logement ordinaire) ...
- ... ayant déménagé depuis dans une communauté

⇒ Un même individu a plusieurs liens avec la base de sondage s'il réside de façon habituelle dans plusieurs logements. Il peut alors être sélectionné plusieurs fois

⇒ Déterminer le nombre de liens pour mettre en œuvre la méthode du partage des poids

I. L'échantillon pour les jeunes en logements ordinaires

Caractéristique n°1 : plan de sondage stratifié

Deux catégories de jeunes aux ressources a priori très hétérogènes :

- Les jeunes « autonomes »
- Les jeunes « à charge » (personne de référence du ménage de 37 ans ou plus)

⇒ deux strates qui partitionnent comme suit les logements du champ d'après le RP 2010 :

Logements de jeunes	Pourcentage du champ
à charge	70%
autonomes	30%

Objectif n°1 : nombre de répondants par strate (1/2)

Les concepteurs souhaitent 4 000 entretiens réussis représentatifs de la répartition de la population par strate

⇒ 2 800 entretiens réussis pour les logements de jeunes « à charge » (70%)

+

1 200 entretiens réussis pour les logements de jeunes « indépendants » (30%)

Ces cibles doivent être prises en compte pour calculer les tailles de l'échantillon par strate

Objectif n°1 : nombre de répondants par strate (2/2)

Trois éléments doivent être pris en compte pour ne pas manquer l'objectif de nombres de répondants :

- Le taux de hors-champ = taux de logements dans le champ au moment du Recensement qui ne le sont plus lors de la collecte ;
- Le taux de résidences principales au Recensement devenues non principales entre temps ;
- Le taux de non-réponse de la population cible

⇒ Comment prévoir ces différentes quantités ?

Estimation des allocations par strate (1/4)

- Les informations de la base de sondage ont vieilli de près de 2 ans au moment de la collecte au T4 2014 (EAR collectée en janvier et février 2013)
- Estimer le taux de logements où habite(nt) au moins un jeune de 16 à 22 ans l'année N pour lesquels on ne retrouve aucun jeune de 18 à 24 ans l'année N+2
- Cette étude suppose un suivi de la composition des mêmes logements sur plusieurs années successives
- Cela n'est pas possible à l'aide des EAR (Recensement rotatif sur un cycle de cinq années)

Estimation des allocations par strate (2/4)

- Utilisation des sources fiscales (TH et FIP) des années 2009 à 2012

	Taux de logements de même caractéristique (en %)		
	En 2011	En 2010	En 2009
Logements dans le champ en 2012	94.0	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	92.0	88.3

- 89,5% des logements contenant au moins un jeune de 16 à 22 ans en 2010 sont dans le champ ENRJ deux ans plus tard

Estimation des allocations par strate (3/4)

- Un taux de hors-champ anticipé différent par strate :
 - Les jeunes « autonomes » : 34%
 - Les jeunes « à charge » : 8%
- Taux de résidences principales devenues non principales entre 2010 et 2012 (TH) = 6%
- Taux de réponse moyen des individus « Kish » de 18 à 24 ans dans les enquêtes CVS 2010 à 2012 = 60%
 - ⇒ Estimation du taux de non-réponse des jeunes de 18 à 24 ans au questionnaire d'ENRJ

Estimation des allocations par strate (4/4)

Pour chaque strate de logements :

Nombre de répondants = Taille de l'échantillon

× (1 – Taux de résidences principales devenues non principales)

× (1 – Taux de hors champ prévu parmi les résidences principales)

× Taux de réponse estimé des ménages du champ

⇒ Les jeunes « autonomes » :

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

Les jeunes « à charge » :

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

Total = 8 620 logements ordinaires à tirer dans l'EAR 2013

Caractéristique n°2 : plan de sondage à deux degrés

Tirage stratifié de l'échantillon dans l'EAR 2013 uniquement dans les ZAE de l'Echantillon-Maître

⇒ Objectif n°2 de l'échantillonnage d'ENRJ : obtenir un échantillon de logements ordinaires équi-pondérés

Objectif n°3 : fournir les allocations du tirage dans l'EAR 2013 au plus tard à l'été 2013 pour que les DR puissent calculer la charge de travail induite des enquêteurs sur l'année 2014 (contrainte du NCEE)

Problème : EAR 2013 indisponible avant février 2014 !

Calcul anticipé des allocations par ZAE x strate (1/2)

- Recommandation du groupe de réflexion sur l'échantillonnage dans le cadre du NCEE : estimer les allocations d'un tirage dans l'EAR N à l'aide des données de l'EAR N-5
- Pour ENRJ : calcul des allocations à partir des données de l'EAR 2008 avec la contrainte des tailles d'échantillon par strate
- Calcul d'allocations pour chaque croisement ZAE x strate
- Les allocations seront respectées strictement lors du tirage dans l'EAR 2013

Calcul anticipé des allocations par ZAE x strate (2/2)

Le vecteur d'allocations qui minimise la dispersion des poids de sondage est défini comme suit :

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

- n^{strate} = taille totale de l'échantillon de la strate
- N_{zae}^{strate} = nombre total de logements de la strate dans la ZAE d'après l'EAR 2008 (nombre pondéré en utilisant les poids de sondage de l'enquête de Recensement 2008)
- π_{zae} = probabilité d'inclusion initiale de la ZAE dans l'Échantillon-Maître
- $N^{strate} = \sum_{ZAE \in EM} \frac{N_{ZAE}^{strate}}{\pi_{ZAE}}$ = estimateur de la taille de la strate dans l'EAR 2008

Bilan de la collecte (1/2)

- Constat au cours de la collecte : 357 logements de la strate des jeunes à charge ne contiennent plus de jeune de 18 à 24 ans. Depuis le RP, ils se sont installés dans leur propre logement
⇒ interrogation de ces jeunes du champ par téléphone et réaffectation à la strate « jeunes autonomes »

Strate de logements	Échantillon issu du tirage		Échantillon d'enquêtes réussies		Taux de collecte 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 %

Bilan de la collecte (2/2)

Conclusions sur la collecte des logements ordinaires :

- Les nombres minimums de répondants attendus par les concepteurs globalement et par strate sont atteints
- Du fait des taux de hors champ différents anticipés par strate, les logements de jeunes « autonomes » sont surreprésentés dans l'échantillon issu du tirage
- *In fine*, l'échantillon d'enquêtes réussies respecte la structure par strate de la population des logements ordinaires du champ

II. L'échantillonnage en communautés

L'échantillonnage en communauté

Objectif des concepteurs : 1 000 jeunes adultes répondants

Difficultés :

- population très mobile et fortement « multi-résidente »
- recensement parfois ancien

⇒ Pas de base de sondage des individus résidant en communautés

Un échantillonnage à deux degrés partiellement décentralisé

En l'absence de base de sondage « individus », un échantillonnage à deux degrés est mis en place

- Au premier degré, un échantillon de 200 communautés est sélectionné dans le répertoire des communautés
- Au deuxième degré, dans chaque communauté sélectionnée, l'enquêteur détermine sur place les individus - ou les chambres - à interroger, correspondant à des rangs de tirages (principaux et de remplacements) générés aléatoirement.

Premier degré: le tirage des communautés (1/3)

Les communautés du champ : 13 852 communautés avec au moins un résident âgé de 18 à 24 ans au Recensement

En pratique la base de sondage effective a été restreinte aux communautés appartenant à une ZAE de l'Echantillon-Maître

Premier degré: le tirage des communautés (2/3)

Tirage de l'échantillon principal de 202 communautés stratifié par région et par type de communauté (*communauté recevant principalement des étudiants / autres types de communauté*) avec allocation proportionnelle au nombre d'individus du champ donné par le Recensement des Communautés

Tirage complémentaire d'un échantillon de réserve de 4 communautés dans chaque région

Premier degré: le tirage des communautés (3/3)

A l'issue de la pré-enquête, 27 communautés ont été déclarées hors champ

Dans 23 cas, un remplacement par une communauté de la réserve dans la même région (et si possible également dans la même catégorie de communautés) a pu être effectué

A l'issue de cette phase de remplacements, un échantillon définitif de 198 communautés a été constitué

Second degré: le tirage des jeunes adultes au sein des communautés (1/3)

Sélection de 6 jeunes adultes à interroger dans chaque communauté, sur la base de rangs d'interrogation aléatoires

Trois cas de figure possibles pour la sélection:

1. Présence d'un registre des occupants avec leur dates de naissance
2. Présence d'un registre des occupants mais sans leur date de naissance
3. Absence de registre

Second degré: le tirage des jeunes adultes au sein des communautés (2/3)

Lorsqu'un registre est disponible, l'enquêteur sélectionne directement les jeunes adultes à interroger selon l'ordre du registre

Lorsqu'il dispose des dates de naissance, il effectue directement la sélection au sein des personnes du champ

En l'absence de registre, il sélectionne des chambres en se basant sur leur numérotation

Second degré: le tirage des jeunes adultes au sein des communautés (3/3)

L'enquêteur a six rangs d'interrogation principaux et pour chaque rang principal :

- Un rang de remplacement si la communauté a un registre avec âge (sélection directe au sein des jeunes adultes du champ)

En l'absence d'information sur l'âge :

- Deux rangs de remplacement dans les communautés avec une majorité de résidents dans le champ de l'enquête (cités universitaires...)
- Trois rangs de remplacement dans les autres cas

Le résultat de la collecte

1 188 répondants théoriques (6 x 198)

1 039 enquêtes réalisées pour 1 328 rangs de contact générés, soit un taux de collecte de 78,2%

III. Poids de sondage finaux des jeunes adultes

Certains jeunes adultes ont des possibilités multiples de sélection

Dans les logements enquêtés, les jeunes adultes du champ sont ceux dont le logement est une résidence habituelle (au moins un mois de présence par an)

Un jeune en lien avec plusieurs logements a d'autant plus de chances (toutes choses égales par ailleurs) d'être sélectionné pour l'enquête

Les jeunes qui ont plusieurs liens avec la base de sondage n'ont pas les mêmes caractéristiques que ceux qui en ont un seul

Nombre de liens en fonction de l'âge

Âge	1 lien avec la base de sondage	2 liens	3 liens ou plus
18 ans	49,9%	44,9%	5,2%
21 ans	71,9%	25,7%	2,4%
24 ans	90,1%	9,6%	0,2%

Nombre de liens en fonction du lieu d'enquête

Lieu d'interrogation	1 lien avec la base de sondage	2 liens	3 liens ou plus
Logement ordinaire	77,7%	21,0%	1,4%
Communauté	26,4%	66,6%	7,1%

L'approche usuelle du partage des poids

Le partage des poids conduit à diviser le poids de sondage du jeune adulte par son nombre de liens avec la base de sondage

Cette technique usuelle permet d'obtenir des estimateurs sans biais sur la population des jeunes adultes

Elle nécessite de connaître le nombre de liens du jeune adulte avec la base de sondage qui peut être déterminé avec les questions suivantes :

- « Dans combien d'autres logements résidez vous de façon habituelle? »
- et, en cas d'arrivée postérieure à janvier 2013, « à votre connaissance, ce (ou ces) logements étaient-ils occupés précédemment par au moins une personne de 18 à 24 ans ».

Le calcul d'un poids « parent » (1/2)

L'enquête ENRJ interroge à la fois les jeunes adultes et leurs parents. On souhaite donc pouvoir diffuser des résultats sur les parents des jeunes adultes

La probabilité d'inclusion d'un adulte « parent » est la somme des probabilités d'inclusion de ses enfants

⇒ Il est donc nécessaire de reconstituer une probabilité d'inclusion pour chacun des enfants (enquêtés ou non enquêtés)

Le calcul d'un poids « parent » (2/2)

Pour calculer les poids « parents », on procède donc de la façon suivante:

- on cherche à reconstituer (au moyen de questions spécifiques dans le questionnaire) la probabilité d'inclusion des jeunes adultes enquêtés
- on impute aux enfants non enquêtés des parents enquêtés la probabilité d'inclusion d'un jeune adulte ayant les mêmes caractéristiques (sexe, âge...) ou bien la valeur moyenne des répondants ayant les mêmes caractéristiques

Le calcul des probabilités d'inclusion des jeunes « enquêtés » (1/2)

Cela suppose une description détaillée des occupants des logements avec lesquels le jeune adulte est en lien, afin de déterminer sa probabilité d'inclusion dans chacun de ses logements (questions ajoutées dans le questionnaire).

Il sera nécessaire de faire des hypothèses simplificatrices sur la probabilité de sélection des autres logements avec lesquels le jeune adulte est en lien.

Le calcul des probabilités d'inclusion des jeunes « enquêtés » (2/2)

Ces probabilités d'inclusion reconstituées fournissent également un second jeu de pondérations sur les jeunes adultes enquêtés

Les estimateurs issus de ce second jeu de pondérations pourront être comparés à ceux issus de la pondération par partage des poids et, autant que possible, à des marges issues de sources externes (Recensement, Etat civil, données de l'Education Nationale) pour validation

Conclusion

Conclusion (1/2)

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 calibrer les tailles d'échantillons par strate en logements ordinaires
- la mise en place d'un échantillonnage en communautés avec un protocole de tirage adapté
- une pondération adaptée basée sur le partage des poids prenant en compte les multiples possibilités de sélection d'un jeune adulte

Conclusion (2/2)

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.

Les spécificités du plan de sondage ENRJ

Merci de votre attention !

Contacts

M. Sébastien Faivre

Tél. : 01 41 17 66 53

Courriel : sebastien.faivre@insee.fr

M. Pierre-Arnaud Pendoli

Tél. : 01 41 17 50 57

Courriel : pierre-arnaud.pendoli@insee.fr

Insee

18 bd Adolphe-Pinard
75675 Paris Cedex 14

www.insee.fr  

Informations statistiques :
www.insee.fr / Contacter l'Insee
09 72 72 4000
(coût d'un appel local)
du lundi au vendredi de 9h00 à 17h00