

LA METHODOLOGIE DE LA SOUS-ALLOCATION

Pierre-Arnaud Pendoli¹ & Olivier Haag²

¹ INSEE, pierre-arnaud.pendoli@insee.fr

² INSEE, olivier.haag@insee.fr

Résumé. La mise en place d'un nouveau cadre de collecte des enquêtes de l'Insee au 1er janvier 2013, et la nécessité nouvelle de faire correspondre les entretiens attribués à chaque enquêteur à une quotité de travail annuelle préalablement fixée, ont fortement limité les marges de manœuvre en matière de réattribution possible des entretiens d'une zone de collecte devenue « orpheline » en raison de l'absence prolongée d'un enquêteur sur l'ensemble de la durée de la collecte d'une enquête. Cette situation d'abandon complet d'une zone est en effet à éviter impérativement car elle crée un biais de non réponse qu'il n'est pas possible de corriger par la suite au moyen des traitements post-collecte habituels. Ainsi, l'objectif de la sous-allocation est de définir une procédure qui permette de limiter au maximum la perte de qualité de l'enquête liée à la diminution du nombre d'entretiens réalisés sur le terrain, en s'obligeant à interroger un minimum de ménages dans chaque zone de collecte au détriment d'autres zones pour lesquelles aucun problème de collecte n'est a priori diagnostiqué.

1. L'échantillonnage des enquêtes ménages à l'INSEE

Il est usuel en statistique d'enquêtes de recourir au tirage d'échantillons à deux degrés dans le cas d'enquêtes en face-à-face. Les données sont collectées par un enquêteur dont une partie du temps de travail est constituée des déplacements entre son domicile et celui du ménage ou de l'individu à interroger. L'intérêt de l'échantillonnage à deux degrés est de concentrer l'échantillon de ménages ou d'individus dans un nombre limité de zones géographiques peu étendues afin :

- de réduire les coûts de collecte (déplacements, temps de repérage) de l'enquête qui dispose d'un budget contraint ;
- d'embaucher un réseau fixe d'enquêteurs qui reçoivent une formation aux techniques d'enquête et acquièrent une expérience dans le domaine, toutes deux mises à profit sur le long terme.

Cette méthode d'échantillonnage des enquêtes nécessite :

- de définir et construire ces zones géographiques (appelées unités primaires) ;
- de sélectionner les zones dans lesquelles concentrer le tirage des ménages de manière à obtenir un échantillon représentatif d'unités primaires (1^{er} degré du tirage) ;
- et de mettre en œuvre le tirage des ménages ou des individus, i.e. les unités secondaires, dans ces zones sélectionnées (2nd degré du tirage).

a. Les zones d'action enquêteur (ZAE), unités primaires des enquêtes ménages de l'INSEE

La construction des ZAE intègre la méthodologie du Nouveau Recensement. En particulier, on distingue les petites et les grandes communes.

- Une grande commune (GC) est une commune de 10 000 habitants ou plus. Elle est en partie (8% des logements) interrogée chaque année de Recensement, pour un cycle de cinq ans du RP.

Une grande commune forme à elle seule une ZAE.

- Une petite commune (PC) est une commune de moins de 10 000 habitants. Une petite commune est affectée à une année de Recensement donnée, pour un cycle de cinq ans du RP.

Une ZAE PC est un regroupement de petites communes qui satisfait les critères suivants :

- Comporter au moins une commune de chaque année de Recensement ;
- Comporter au moins 300 résidences principales de chaque année de Recensement ;
- Appartenir à une seule région administrative ;
- Avoir une étendue ne dépassant 20 km de rayon.

b. Tirage de 1^{er} degré : constitution de l'Échantillon-Maître

Parmi les 3 832 ZAE qui partitionnent le territoire métropolitain, 567 ZAE sont sélectionnées aléatoirement¹.

c. Tirage de 2nd degré : les logements

Selon le thème de l'enquête, l'unité statistique peut être le ménage ou l'individu. Dans les deux cas, l'unité statistique est atteinte via son logement.

Les grands principes sont les suivants² :

- Tirage des logements dans les ZAE de l'Échantillon-Maître. Ces logements tirés sont communément appelés « fiches adresses » pour les enquêteurs ;
- Calcul d'allocations théoriques par ZAE dont l'objectif est de minimiser la dispersion des poids de sondage des logements ;
- Définition éventuelle d'une borne minimum et d'une borne maximum d'enquêtes à réaliser par ZAE afin de prendre en compte la charge des enquêteurs.

2. Les nouvelles conditions d'emploi des enquêteurs (NCEE)

L'entrée en vigueur des NCEE en 2013 impose à l'Institut de faire correspondre le temps de collecte annuel d'un enquêteur à une quotité annuelle préalablement fixée. Ce temps de collecte comprend :

- Le temps total de déplacement. Celui-ci est fonction :
 - a. du nombre de logements à interroger ;
 - b. du temps nécessaire pour atteindre la zone d'enquête (y compris le temps de repérage de l'adresse) ;

¹ Pour plus d'informations sur le tirage des ZAE, se référer à l'article de Marc CHRISTINE et Sébastien FAIVRE des JMS 2009 : « OCTOPUSSE : un système d'Échantillon-Maître pour le tirage des échantillons dans la dernière Enquête Annuelle de Recensement » (article disponible sur le site des JMS).

² Pour plus d'informations, se référer à l'article cité ci-dessus.

- c. du nombre moyen de contacts nécessaires avant d'obtenir un rendez-vous.
- Le temps d'enquête total chez les ménages. Celui-ci dépend :
 - a. du nombre de ménages à interroger ;
 - b. du taux de réponse ;
 - c. du temps de passation d'un questionnaire.

Chaque année une planification est donc réalisée afin d'attribuer l'ensemble des échantillons des enquêtes de l'année aux enquêteurs en respectant leur quotité de travail.

Toutefois, par la suite, en cas d'absence prolongée d'un enquêteur sur une période de collecte, il est possible que pour une enquête donnée et une ZAE donnée, très peu voire aucun entretien ne puisse être réalisé. On parle alors de ZAE « orpheline ».

Des marges de manœuvre existent pour prendre en charge ces entretiens dans la ZAE « orpheline ». Il s'agit principalement :

- d'un recours à une réserve, destinée aux remplacements, de la quotité d'autres enquêteurs du réseau³ ;
- d'un recrutement d'un enquêteur occasionnel (sous réserve de disponibilités du plafond d'emploi).

Il peut cependant arriver que, une fois explorées toutes ces possibilités, il ne soit pas possible de dégager de moyens de collecte supplémentaires pour prendre en charge en l'état tout ou partie des entretiens dans la ZAE « orpheline ».

Comme l'abandon complet des enquêtes dans une unité primaire est à proscrire car il créerait par la suite un biais de non réponse qu'il n'est pas possible de corriger au moyen des traitements post-collecte habituels, il est nécessaire de collecter un minimum d'entretiens dans chaque ZAE de l'Echantillon-Maître.

Dans ce cas, la méthodologie mise en œuvre est :

- de réaliser une partie des entretiens prévus initialement sur la ZAE « orpheline ». Ils doivent être attribués à d'autres enquêteurs du réseau, que l'on appellera « enquêteurs remplaçants » ;
- de décharger ces enquêteurs remplaçants d'une partie des entretiens qu'ils devaient réaliser initialement sur leur ZAE d'origine afin de respecter leur quotité de travail⁴.

Se pose alors la question de la méthode à retenir quant à la sélection des logements finalement enquêtés, à la fois dans la ZAE « orpheline » et dans les ZAE d'origine des enquêteurs remplaçants.

C'est pourquoi des travaux ont été mis en œuvre afin de :

- définir une méthodologie qui permette de limiter au maximum la perte de qualité de l'enquête liée à la diminution du nombre d'entretiens réalisés sur le terrain,
- en respectant la contrainte d'interroger un minimum de ménages dans chaque ZAE.

Deux cas de figure peuvent se présenter :

- Un enquêteur est indisponible avant la collecte ou en début de collecte d'une enquête et durant la totalité de la durée de la collecte. La méthode mise en œuvre dans ce cas est celle de la sous-allocation, présentée dans ce document.

³ En 2014, la planification initiale ne prenait en compte que 95% de la quotité des enquêteurs. Les 5% restants étaient destinés à la réserve.

⁴ Bien entendu, comme en général le temps de déplacement vers la ZAE « orpheline » sera supérieur au temps de déplacement de l'enquêteur vers sa ZAE d'origine, on sera contraint de supprimer plus d'entretiens dans sa ZAE d'origine qu'on ne lui en confiera dans la ZAE « orpheline ».

- L'enquêteur arrête brusquement sa collecte en cours d'enquête alors qu'elle est déjà largement avancée, ou ne revient pas en cours de collecte alors qu'il était prévu qu'il le fasse. Une méthodologie différente doit être utilisée afin de prioriser l'interview de certains types de ménages dans le cas où certains profils soient sur ou sous représentés parmi les entretiens réalisés jusque là.

3. Le recours à la technique de sous-allocation

Contexte :

Avant la collecte ou en début de collecte d'une enquête, on a pris connaissance d'une absence prolongée d'un enquêteur sur la totalité de la durée de la collecte, rendant la ZAE « orpheline ».

Critères d'optimisation :

1. **Minimiser la dispersion des poids de sondage** des logements de l'échantillon final (i.e. l'échantillon qui sera finalement collecté) dans la région concernée.
2. **Réduire le moins possible la taille de l'échantillon.**

Contraintes :

Trois contraintes ont été imposées dans le modèle. Elles sont décrites ci-dessous. Alors que la première découle directement du respect des quotités, les deux autres ont été ajoutées pour limiter l'impact de la modification de l'échantillon sur les conditions de travail des enquêteurs concernés.

1. Conserver à l'identique la quotité de travail des enquêteurs.
Hypothèses sur le temps de travail des enquêteurs qui comprend, comme décrit plus haut :
 - *Un temps de déplacement.* On suppose que celui-ci est uniquement fonction de la distance entre le domicile de l'enquêteur et le barycentre de la ZAE de collecte. L'hypothèse implicite est que le nombre moyen de contacts par ménage est le même dans la ZAE d'origine de l'enquêteur remplaçant et dans la ZAE « orpheline » et que les temps de repérage sont identiques bien que l'enquêteur soit moins habitué à la ZAE « orpheline ». Cette hypothèse est donc un peu optimiste en termes de quotité consommée.
 - *Un temps d'enquête.* On suppose qu'il reste constant. Cette hypothèse est en revanche pessimiste dans la mesure où, comme on l'a vu, l'enquêteur se verra confier in fine moins d'entretiens qu'au départ. Toutes choses égales par ailleurs, il réalisera donc moins d'entretiens et son temps d'enquête sera plus faible qu'à l'origine.

Sous ces hypothèses, la contrainte porte uniquement sur la conservation du temps total de déplacement.

2. In fine, un enquêteur remplaçant n'aura pas à sa charge davantage d'entretiens dans la ZAE « orpheline » que dans sa ZAE d'origine.
3. In fine, un enquêteur conservera à sa charge au moins 50 % des entretiens qui lui ont été attribués initialement dans sa ZAE d'origine.

Les informations nécessaires pour mettre en œuvre la méthode :

- Connaître la ZAE « orpheline » avant la collecte ou en début de collecte ;
- Connaître les enquêteurs remplaçants potentiellement mobilisables ;
- Connaître la quotité restante pour chacun de ces enquêteurs ;
- Connaître le temps de déplacement de ces enquêteurs remplaçants entre leur domicile et le barycentre de la ZAE « orpheline ».

Formalisation du problème :

Le programme de minimisation région par région

$$(1) \quad \text{Min}_{n_{ZAE}} \frac{1}{n_{reg}} * \sum_{ZAE \in reg} n_{ZAE} * (Poids_{ZAE}^{log} - Poids_{moy}^{log})^2$$

$$\text{où} \quad n_{reg} = n_{ZAE}^{orpheline} + \sum_{ZAE \in remplaçantes} n_{ZAE} + \sum_{ZAE \in reste reg} n_{ZAE}$$

$$Poids_{ZAE}^{log} = \omega_{ZAE} * \frac{N_{ZAE}}{n_{ZAE}}$$

Sous les contraintes par enquêteur

$$(2) \quad n_{ZAE\ rempl}^{initial} * T_{ZAE\ rempl}^{enq} = n_{ZAE\ rempl}^{final} * T_{ZAE\ rempl}^{enq} + n_{ZAE\ orpheline}^{final} * T_{ZAE\ orpheline}^{enq}$$

$$(3) \quad n_{ZAE\ rempl}^{final} \geq \frac{n_{ZAE\ rempl}^{initial}}{2}$$

$$(4) \quad n_{ZAE\ rempl}^{final} \geq n_{ZAE\ orpheline}^{final}$$

Mise en œuvre pratique :

Les deux ou trois meilleurs scénarios (l'optimum du programme et les scénarios qui s'en approchent le plus) sont proposés aux responsables de la collecte dans la région concernée. Ils choisissent parmi ces scénarios celui qui est le plus adapté aux contraintes de collecte.

4. Illustration sur un exemple fictif : la ZAE de Vizille pour la collecte de l'enquête Logement 2013

Informations disponibles dans l'échantillon

- La ZAE de Vizille a une allocation de 32 ménages tirés pour l'enquête Logement 2013.
- La région Rhône-Alpes a une allocation totale de 2 591 ménages pour cette enquête.
- La dispersion des poids de sondage dans l'échantillon pour cette région vaut 1 954.
- La dispersion des poids de sondage finaux dans la région, si l'on ne réalisait qu'un entretien dans la ZAE « orpheline », vaudrait 102 039.

Informations fournies par la Direction régionale de Rhône-Alpes chargée de superviser la collecte dans la ZAE de Vizille

Enquêteur en charge de la ZAE de ...	Allocation initiale dans la ZAE	Temps de déplacement de l'enquêteur vers sa ZAE habituelle (en minutes)	Temps de déplacement estimé de l'enquêteur vers la ZAE de Vizille (en minutes)	Ratio des temps de déplacement
La Mure	52	10	25	2.5
Saint-Marcellin	43	29	55	1.9
Grenoble 1	45	27	47	1.7
Grenoble 2	24	16	36	2.25
Grenoble 3	10	10	25	2.5

Résultats obtenus en fonction des enquêteurs mobilisés

Enquêteur(s) remplaçant(s) mobilisé(s) initialement sur...	Taille de l'échantillon en Rhône-Alpes <i>a posteriori</i>	Dispersion des poids de sondage <i>a posteriori</i>
La Mure / Saint-Marcellin	2 544	1 955
La Mure / Saint-Marcellin / Grenoble	2 543	1 992
La Mure / Grenoble	2 540	2 059
La Mure	2 544	2 233
Saint-Marcellin	2 550	2 192
Grenoble	2 546	2 184
Aucun enquêteur remplaçant mobilisé + Hypothèse : 1 FA réalisée dans la ZAE « orpheline »	2 560	102 039

Détail du scénario « La Mure / Saint-Marcellin » (i.e. l'optimum)

Dans ce scénario, on mobilise les enquêteurs en charge habituellement des ZAE de la Mure et de Saint-Marcellin pour reprendre une partie des entretiens de la ZAE de Vizille « orpheline ». En contrepartie, ces deux enquêteurs remplaçants se voient retirer une partie des entretiens qui leur

avaient été initialement attribués sur leur ZAE d'origine afin de maintenir leur temps de travail constant.

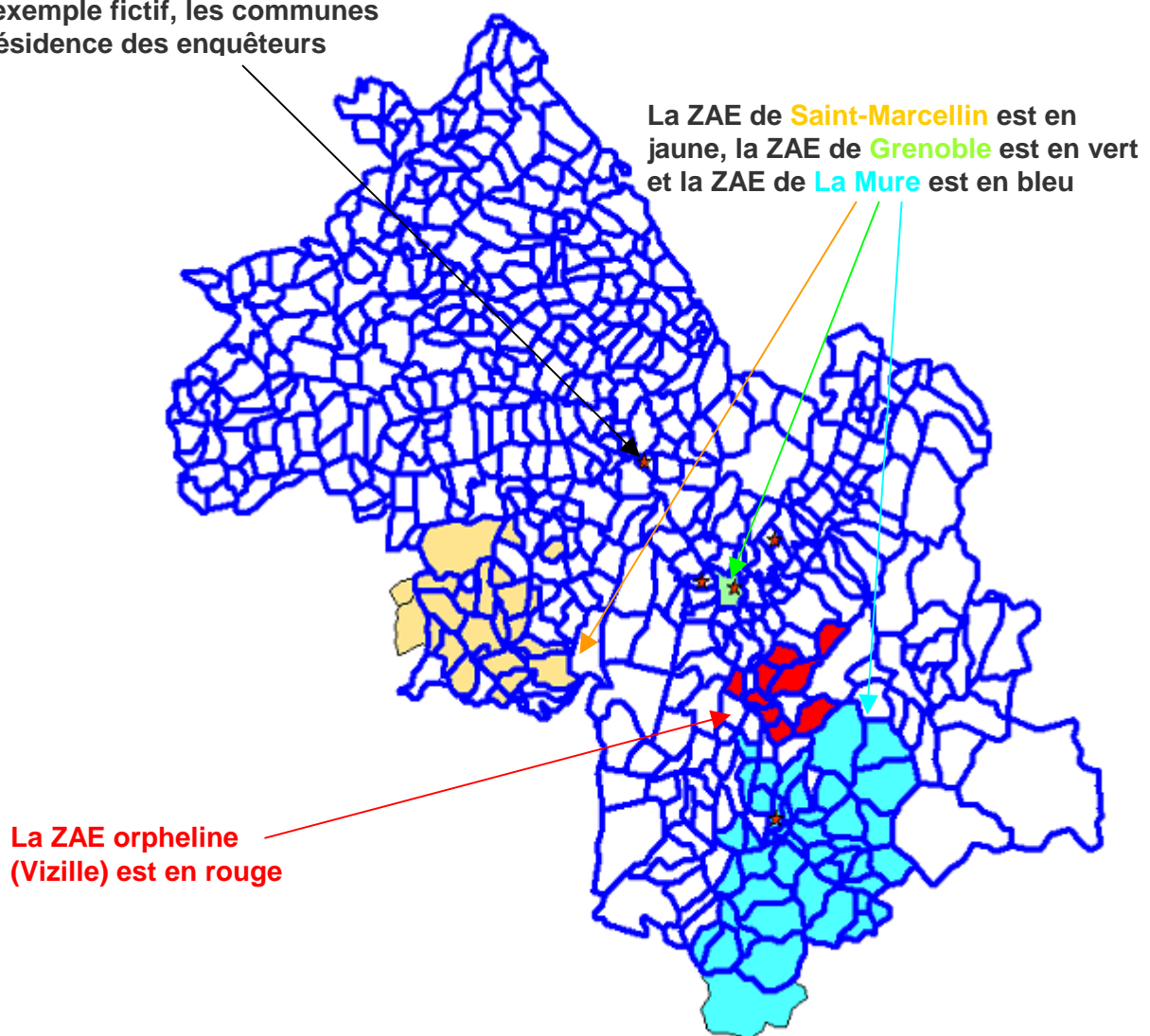
L'algorithme donne les sous-allocations suivantes :

- 15 FA sur 32 sont réalisées dans la ZAE de Vizille
- 35 FA sur 52 sont réalisées dans la ZAE de la Mure
- 30 FA sur 43 sont réalisées dans la ZAE de Saint-Marcellin
- Les enquêteurs grenoblois ne sont pas mobilisés. Il n'y a aucun changement pour les allocations dans les trois ZAE de Grenoble.

Dans chaque ZAE (la ZAE « orpheline » et celles des enquêteurs remplaçants), les ménages qui seront finalement enquêtés sont sélectionnés aléatoirement parmi les ménages tirés initialement dans l'échantillon.

Cartographie schématique de la ZAE de Vizille, des ZAE d'origine des enquêteurs remplaçants et de leurs domiciles

Les étoiles représentent, dans cet exemple fictif, les communes de résidence des enquêteurs



5. Bilan de la mise en œuvre

Cette méthode a été mise en œuvre pour deux enquêtes Ménages de l’Insee.

- Pour l’enquête « Logements 2013 », 6 ZAE étaient orphelines (soit 189 logements). 8 enquêteurs remplaçants ont été mobilisés. Ils ont enquêté 95 logements dans les ZAE « orphelines », tandis que 138 entretiens leur ont été retirés dans leurs ZAE d’origine. Le coût de la méthode a donc été une perte de 43 logements supplémentaires mais en évitant toute ZAE « orpheline ».
- Pour l’enquête « Conditions de Vie et Sécurité 2014 », 6 ZAE étaient orphelines (soit 212 logements). 10 enquêteurs remplaçants ont été mobilisés. Ils ont enquêté 97 logements dans les ZAE « orphelines », tandis que 176 entretiens leur ont été retirés dans leurs ZAE d’origine. Le coût de la méthode a donc été une perte de 79 logements supplémentaires mais en évitant toute ZAE « orpheline ».

6. Conclusion

Cette méthode pragmatique a permis d’éviter d’être confronté à des zones sans collecte et donc d’éviter un biais difficilement corrigeable.

Il est à noter que, plus que la durée d’accès à la zone « orpheline », c’est le ratio entre la durée du trajet pour se rendre dans la ZAE d’origine et la durée du trajet pour se rendre dans la ZAE « orpheline » qui joue un rôle important dans le modèle. Car plus ce ratio est grand et plus la perte d’entretiens est importante.

Cela justifie la mise en place d’une communication adaptée vis-à-vis des enquêteurs. Pour chaque scénario proposé, une validation de la faisabilité sur le terrain par les équipes en charge de la collecte est nécessaire.

Bibliographie

- [1] Marc CHRISTINE et Sébastien FAIVRE (2009), *OCTOPUSSE : un système d’Échantillon-Maître pour le tirage des échantillons dans la dernière Enquête Annuelle de Recensement*.
- [2] Antoine REBECQ (2014), *Heuristique Branch-and-bound pour la sous-allocation et la réallocation*.