

La procédure de sous-allocation en cas d'absence d'un enquêteur en début de collecte

Olivier Haag

Chef de la section Echantillonnage



Plan de la présentation

- Présentation des nouvelles contraintes liées au NCEE
- Présentation de la méthode de sous-allocation
- Un exemple de mise en œuvre
- Les premières conclusions



Les nouvelles contraintes liées au NCEE

- Nécessité de faire correspondre les fiches-adresses attribuées à chaque enquêteur à une quotité annuelle préalablement fixée
- Plafond d'emploi à respecter
- 5% de quotité par enquêteur mise en réserve en 2014 pour des remplacements



Les conséquences

- Possibilités de réattribution des fiches-adresses limitées
- Problèmes en cas d'absence prolongée d'un enquêteur sur la période de collecte
- Apparition de zones de collecte dites « orphelines » pour une ou plusieurs enquêtes

A éviter car génère un biais de collecte



Un traitement correctif doit être mis en œuvre

- s'obliger à interroger un minimum de ménages dans chaque ZAE ...
- ... au détriment d'autres ZAE sans problème a priori
- Limiter au maximum la perte de qualité de l'enquête liée à la diminution du nombre d'enquêtes réalisées sur le terrain



Quel traitement mettre en œuvre

Deux cas de figure peuvent se présenter :

- La ZAE devient « orpheline » avant la collecte ou en début de collecte de l'enquête => solution de sous-allocation présentée ici

- La ZAE devient orpheline alors que la collecte de l'enquête est largement avancée: solution à définir dans le calcul de priorisation des FA



Principe de la méthode de sous-allocation

Les informations nécessaires pour pouvoir la mettre en œuvre

- Connaissance de la ZAE orpheline avant la collecte ou en début de collecte
- Connaissance des enquêteurs potentiellement mobilisables
- Connaissance de la quotité restante pour ces enquêteurs
- Temps de déplacement de ces enquêteurs « remplaçants » dans la ZAE orpheline

Informations fournies par les directions régionales

Méthodologie de la sous-allocation

Problème d'optimisation sous contraintes

➤ Critères d'optimisation [1] :

- Minimiser la dispersion des poids de sondage des ZAE de la région pour limiter la dispersion des poids finaux des logements
- Réduire le moins possible la taille de l'échantillon

➤ Les contraintes

- Conserver la quotité de travail des enquêteurs (imposé) [2]
- Contraintes supplémentaires

Un enquêteur fera au moins 50% des FA de sa zone initiale [3]

Un enquêteur remplaçant ne fera pas plus de FA dans la zone orpheline que dans sa zone initiale [4]



Prise en compte de la contrainte de quotité de chaque enquêteur

Le temps de travail des enquêteurs comprend:

- un temps de déplacement,
- un temps d'enquête chez les ménages,

Hypothèses sur ces temps :

- Temps de déplacements : uniquement fonction de la distance domicile enquêteur ZAE de collecte
- Temps d'enquête : reste constant bien que moins d'enquêtes faites in fine



Formalisation du problème (1/2)

Le programme de minimisation région par région

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

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

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

Formalisation du problème (2/2)

Sous les contraintes par enquêteur

[2]

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

Et

[3]

$$n_{ZAE\ rempl}^{final} \geq \frac{n_{ZAE\ rempl}^{initial}}{2}$$

Et

[4]

$$n_{ZAE\ rempl}^{final} \geq n_{ZAE\ orpheline}^{final}$$

Déroulement de l'algorithme

1 – Sélection du scénario

- Choix du nombre d'enquêteurs mobilisés
- Choix des enquêteurs mobilisés

2 – Sélection de la meilleure allocation par scénario

3- Sélection des meilleurs scénarios à proposer aux DR

Hypothèse supplémentaire pour gain de temps de traitement

- Quand plusieurs enquêteurs mobilisés : ils feront le même nombre de FA dans la ZAE orpheline



La ZAE de Vizille pour l'ENL 2013 (cas théorique) (1/2)

- Seul l'échantillon principal est étudié pour l'exercice (hors ZUS, logements neufs et réserve)
- La ZAE de Vizille a 32 fiches-adresses tirées pour l'ENL échantillon principal
- La région Rhône-Alpes a 2591 fiches-adresses au total pour l'ENL échantillon principal.
- La dispersion des poids dans la région sur l'échantillon complet est de 1954.
- La dispersion des poids dans la région (calculé sous la convention qu'on réalise une FA dans la ZAE orpheline) vaudrait 102 039.



La ZAE de Vizille pour l'ENL 2013 (cas théorique) (2/2)

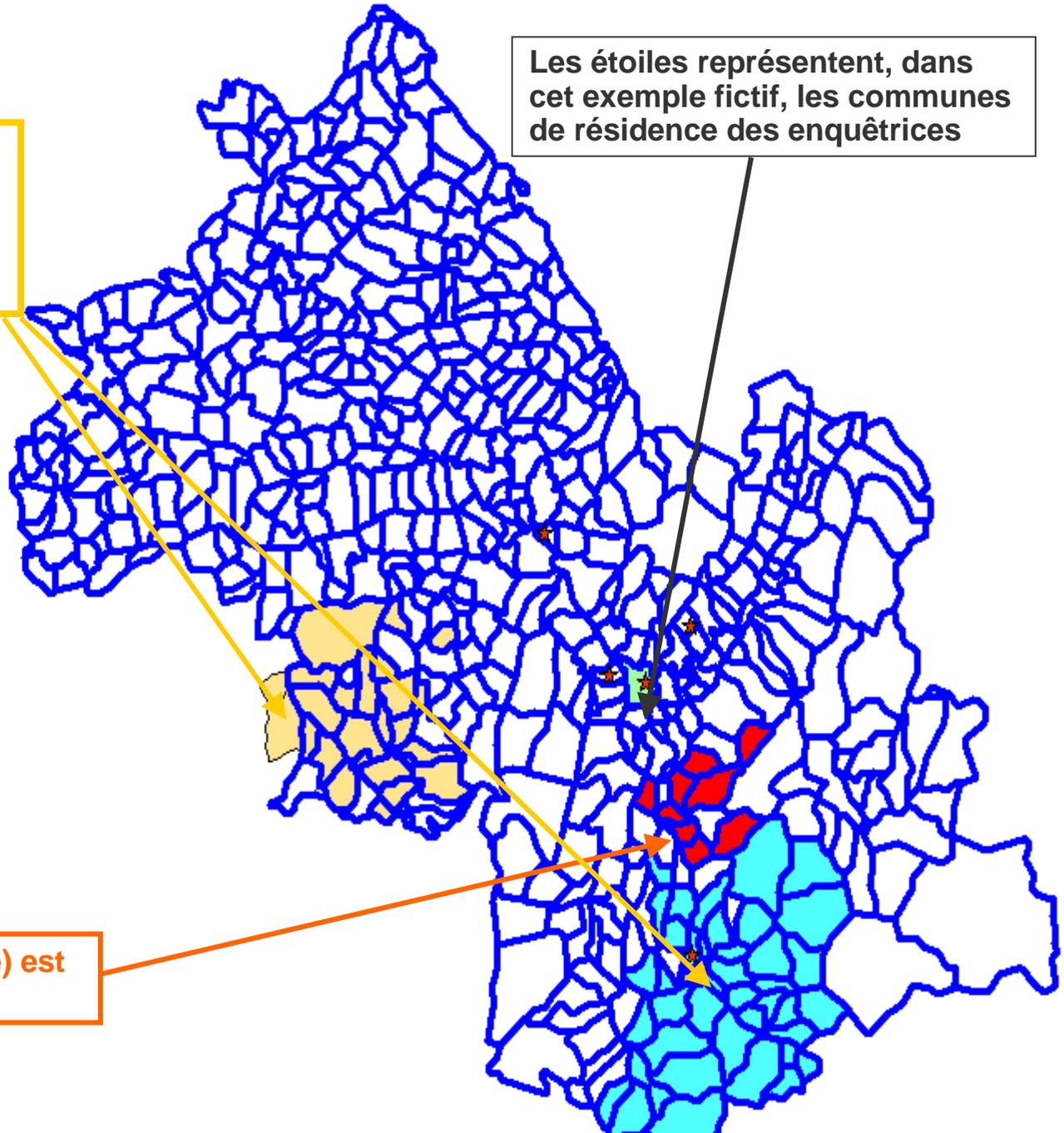
- Les informations fournies par la DR

ZAE	Nombre de FA initiales ENL échantillon principal	Temps accès enquêteur ZAE habituelle (mn)	Temps accès estimé enquêteur ZAE Vizille	Ratio des temps accès
La Mure	52	10	25	2.5
Saint-Marcellin	43	29	55	1.9
Grenoble E1	45	27	47	1.7
Grenoble E2	24	16	36	2.25
Grenoble E3	10	10	25	2.5

La ZAE de **Saint-Marcellin** est en jaune, la ZAE de **Grenoble** est en vert et la ZAE de **La Mure** est en bleu

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

La ZAE orpheline (Vizille) est en rouge.



Résultats obtenus en fonction des ZAE mobilisées

Liste des ZAE mobilisées	Echantillon total région	Dispersion des poids de sondage initiaux
Aucune	2559	102 039
La Mure/Saint-Marcellin	2544	1955
La Mure/Saint-Marcellin/Grenoble	2543	1992
La Mure / Grenoble	2540	2059
La Mure	2544	2233
Saint-Marcellin	2550	2192
Grenoble	2546	2184



Détail du scénario « La Mure/Saint-Marcellin »

- Si on retient ce scénario, l'algorithme donne les allocations suivantes :
- 14 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



La sélection de fiches adresses à réaliser

- Seules les fiches adresses « non repérées » peuvent être abandonnées au sein des ZAE « remplaçantes »
 - Les FA à abandonner sont sélectionnées de façon aléatoire par la divisions Sondages
- Les FA à réaliser dans la ZAE orpheline sont également sélectionnées de façon aléatoire par la division Sondages



Conclusion sur l'exemple

- Pas forcément les enquêteurs les plus proches retenus pour le scénario optimal
- C'est le ratio temps trajet ZAE initiale / temps de trajet ZAE orpheline qui joue le plus sur la taille de l'échantillon
- Nécessité de prendre en compte les réalités du terrain
- La DR a le dernier mot



Conclusion

- Une méthodologie qui permet de limiter fortement la baisse de la qualité statistique des échantillons
- Travail collaboratif et pragmatique entre la division sondage et la DEM
 - Sélection des enquêteurs potentiels et des données en entrée de l'algorithme par la DEM
 - Plusieurs scénarios « statistiquement » acceptables sont ensuite proposés par la division Sondages
 - Sélection du scénario définitif par la DEM en prenant en compte les contraintes « du terrain »



Les premières expériences

- **Enquête logement :**
 - 5 DR concernées
 - 14 ZAE concernées (dont 6 orphelines)
 - 232 FA abandonnées in fine
 - 95 FA réalisées dans les ZAE orphelines sur les 189 FA de l'échantillon initial

- **Enquête CVS**
 - 5 DR concernées
 - 16 ZAE concernées (dont 6 orphelines)
 - 291 FA abandonnées in fine
 - 97 FA réalisées dans les ZAE orphelines sur les 212 FA de l'échantillon initial

- **Une adaptation de la méthode pour le panel SRCV**
 - Réaliser les quatre 1eres vagues (obligatoires) ...
 - Au détriment des dernières vagues



Merci de votre attention !

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

