



# Séminaire de Méthodologie Statistique

**mardi 1<sup>er</sup> avril 2014**

14h-17h, Insee - Malakoff 1 - salle 1245

L'adaptation de la méthodologie des enquêtes ménages à de nouvelles conditions de collecte

## **La nouvelle planification de l'échantillonnage**

**Pierre-Arnaud Pendoli** - *Division Sondages, Insee*

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

**Olivier Haag** - *Division Sondages, Insee*

## **L'utilisation de R-indicateurs pour la priorisation des enquêtes en cours de collecte**

**Thomas Merly-Alpa** - *Division Sondages, Insee*

## **L'évaluation anticipée de biais potentiels en cas de difficultés de collecte : le cas de l'enquête CVS 2013**

**Emmanuel Gros** - *Division Sondages, Insee*

## Résumés des interventions

### **La nouvelle planification de l'échantillonnage**

**Pierre-Arnaud Pendoli** - *Division Sondages, Insee*

L'entrée en vigueur des nouvelles conditions d'emploi des enquêteurs de l'Insee (NCEE) en 2013 a entraîné une adaptation du calendrier et des méthodes de tirage des échantillons des enquêtes ménages. La charge de travail des enquêteurs pour une année donnée doit en effet être fixée au plus tard à l'automne de l'année précédente et elle doit correspondre aux ressources disponibles *a priori*. Cette charge dépend du nombre de Fiches-Adresses attribuées à un enquêteur pour chaque enquête. Pour respecter ce calendrier et permettre aux Directions régionales de s'organiser, la Division Sondages doit donc livrer les allocations par Zone d'Action Enquêteur des échantillons des enquêtes collectées une année N au mois de juin de l'année N-1.

Ainsi, les allocations des échantillons enquêtés au second semestre doivent être connues près d'un an et demi avant le début de la collecte. Afin que cette nouvelle contrainte pénalise le moins possible la qualité de la collecte de ces enquêtes, la Division Sondages a développé une méthodologie permettant de calculer par anticipation les allocations optimales (selon les critères définis par les concepteurs d'enquête) puis de les mettre en œuvre sur une base de sondage plus fraîche disponible seulement en début d'année suivante. Dans ce cas, contrairement à ce qui était fait jusque là, le calcul des allocations et le tirage de l'échantillon ne se font pas à partir de la même base de sondage.

.../...

**La procédure de sous-allocation  
en cas d'absence d'un enquêteur en début de collecte**  
**Olivier Haag - Division Sondages, Insee**

La mise en place des NCEE au 1<sup>er</sup> janvier 2013, et la nécessité nouvelle de faire correspondre les fiches-adresses attribuées à chaque enquêteur à une quotité de travail annuelle préalablement fixée, ont fortement limité les marges de manœuvre en matière de réattribution possible des fiches-adresses d'une ZAE 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 par la suite de corriger 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'enquêtes réalisées sur le terrain, en s'obligeant à interroger un minimum de ménages dans chaque ZAE au détriment d'autres ZAE pour lesquelles aucun problème de collecte n'est *a priori* diagnostiqué.

**L'utilisation de R-indicateurs  
pour la priorisation des enquêtes en cours de collecte**  
**Thomas Merly-Alpa - Division Sondages, Insee**

Le R-indicateur (indicateur de Représentativité) permet de mesurer le degré de similarité d'un échantillon par rapport à la population de base. Il est basé sur la dispersion des probabilités de réponse estimées des ménages échantillonnés, et se décline en R-indicateurs partiels mesurant cette représentativité variable par variable. Ces R-indicateurs sont des outils permettant d'analyser la collecte en isolant des groupes de populations sous-représentées. Il est possible ensuite de lancer une procédure dite de « priorisation » qui consiste à intensifier les efforts de collecte sur les groupes précédemment identifiés.

Cette procédure a été testée sur les données de l'enquête Patrimoine 2010 afin d'une part d'observer l'évolution de la représentativité au cours de la collecte, et d'autre part d'étudier la possibilité de compenser la perte de précision induite par une baisse des taux de réponse. La procédure étant souple, il s'agit de tester plusieurs scénarios de priorisation et d'évaluer leurs effets sur la variance de certains indicateurs issus de l'enquête : patrimoine brut moyen des Français, patrimoine net moyen...

**L'évaluation anticipée de biais potentiels en cas de difficultés de  
collecte : le cas de l'enquête CVS 2013**

**Emmanuel Gros - Division Sondages, Insee**

L'enquête Cadre de Vie et Sécurité (CVS) est une enquête annuelle de l'Insee qui a pour objectif de recenser chaque année les faits de victimation intervenus l'année précédente. En 2013, des problèmes ponctuels, liés à l'entrée en vigueur des nouvelles conditions d'emploi des enquêteurs de l'Insee, sont venus perturber la collecte et ont conduit à une baisse du taux de collecte de l'ordre de 10 points de pourcentage. Cette diminution inhabituelle et non négligeable des taux de collecte pouvait laisser craindre l'apparition de biais de non-réponse que les procédures de redressements habituelles ne seraient pas en mesure de corriger correctement.

Afin d'évaluer l'ampleur de ces biais potentiels, des simulations ont été menées en s'appuyant sur les données issues des millésimes précédents de CVS. Ces simulations ont consisté à reproduire sur les données issues des enquêtes CVS 2012 et 2011 la chute des taux de collecte observée en 2013, puis à effectuer l'ensemble des redressements habituellement mis en œuvre dans les enquêtes CVS, afin de pouvoir mesurer l'impact sur les estimations de la dégradation des conditions de collecte. À titre de comparaison, nous avons également simulé les conditions de collecte de CVS 2011 sur les données de CVS 2012.

La comparaison de ces différents scénarios n'a pas révélé l'existence de biais significatifs concernant les principales variables de victimation habituellement diffusées. Seuls les résultats relatifs aux victimations les plus sensibles, moins fréquentes, présentaient des résultats incitant à une vigilance particulière, et font depuis l'objet d'investigations méthodologiques supplémentaires.