

Priorités de rappels pour non réponse totale. Revue de la littérature et simulations à partir de l'enquête TIC

Benoît BUISSON¹, Amandine RODRIGUES²

Le pôle ingénierie statistique entreprises (PISE) de l'INSEE a en charge le redressement des enquêtes thématiques entreprises. Il est notamment demandé au PISE d'intervenir durant la collecte pour définir des priorités de rappel, suite à non réponse totale, dans l'objectif d'optimiser les moyens disponibles. Il s'agit bien ici de définir des priorités de rappels personnalisés, par téléphone, toutes les unités enquêtées non répondantes étant systématiquement relancées par voie postale. Dans cette optique, il existe différentes stratégies possibles, par exemple :

- relancer les grandes unités, en tenant compte de leur poids ;
- relancer les unités les plus influentes potentiellement sur la variable cible ;
- relancer des unités dans les strates pour lesquelles on s'attend à une forte réduction de la variance.

Dans un premier temps, il sera fait une revue de la littérature sur le sujet pour établir d'un point de vue théorique les différentes approches possibles en mettant en exergue leurs avantages et leurs limites. Il sera notamment mis en évidence deux grandes classes de méthodes. D'une part celles qui se fondent sur la réduction de l'erreur d'imputation, suite à correction de la non réponse totale. D'autre part celles qui tentent de minimiser l'écart entre la variance cible de l'estimateur, et la variance en présence de non réponse totale (méthode de variance à gagner). Des exemples d'application de ces méthodes, tant en France que dans d'autres pays, seront fournis.

En ce qui concerne l'enquête sur les technologies de l'information et des communications (TIC), la méthode définie par le pôle ISE, actuellement appliquée, est une adaptation d'une méthode mise au point par l'Australian Bureau of Statistics, fondée sur la réduction de l'erreur lié à la correction de la non réponse totale. Cette méthode est notamment employée afin de fiabiliser au maximum les estimateurs relatifs au montant du commerce électronique (achats et ventes électroniques). En considérant le fichier de diffusion de l'enquête nationale TIC 2010, cet article présentera différentes simulations pour apporter un jugement sur la pertinence de la méthode utilisée, sur ses conditions d'utilisation optimales. Il sera nécessaire dans un premier temps de simuler de la non réponse selon différents mécanismes. Nous utiliserons pour cela le logiciel GENESIS (General Simulation System) développé à Statistique Canada sous SAS. Puis différents modèles de correction de la non réponse (aléatoire, par groupe de réponse homogène notamment) seront appliqués. Enfin dans les différents cas mis en évidence, différentes méthodes de priorités de rappels seront testées afin de juger de leur pertinence relative. Pour juger de cette pertinence, il sera nécessaire de calculer des estimateurs, ainsi que les intervalles de confiance associés, dans différents cas : hypothèse de 100% de réponse, en présence de non réponse et avant rappel, en présence de non réponse et après rappel notamment. Cet article tentera d'établir un diagnostic sur les méthodes de priorités de rappel selon les scénarios de référence et les variables analysées.

¹ INSEE, benoit.buisson@insee.fr

² INSEE, amandine.rodriques@insee.fr