Titre : **Découpage optimal d'une variable quantitative pour la stratification :** une expérience sur les données d'entreprises françaises et une mise en œuvre pour l'échantillonnage des Enquêtes Mensuelles de Branches

Auteurs : Laurent Costa (DMCSI) : laurent.costa@insee.fr
Arnaud Fizzala (Ex-DMCSI) : arnaud.fizzala@sante.gouv.fr

Thème : Optimisation d'un plan de sondage

Résumé

Le découpage en tranches d'une variable quantitative pour élaborer des strates fait l'objet de plusieurs articles scientifiques, notamment lorsque cette variable possède une distribution asymétrique, ce qui est le cas des variables d'effectif salarié et de chiffre d'affaires utilisées pour caractériser la taille des entreprises lors de l'élaboration de plans de sondage d'enquêtes auprès des entreprises. Toutefois de telles méthodes de stratification sont aujourd'hui rarement mises en œuvre à l'Insee où les tranches d'effectifs utilisées pour caractériser la taille des entreprises sont en général les mêmes d'une enquête à l'autre.
La première partie de l'article sera consacrée à la présentation des trois méthodes de découpage optimal de strates implémentées dans le package R « *stratification* » à savoir, les méthodes *de* Dalienus, géométrique, Lavalée et Hidiroglou. Puis les résultats d'une première application de ces méthodes sur des données correspondant au champ habituel des enquêtes françaises auprès d'entreprises seront détaillés. L'application consiste à découper la variable « effectif salarié » en sept tranches de façon à minimiser la taille d'échantillon nécessaire pour atteindre une précision fixée a priori de l'estimateur de l'effectif salarié de l'ensemble de la population. L'utilisation de ces méthodes peut permettre de réduire les tailles d'échantillon nécessaires pour atteindre un objectif de précision fixé a priori. Ces réductions sont d'autant plus prononcées que les objectifs de précision sont ambitieux.

La deuxième partie de l'article présentera la mise en œuvre concrète de ces méthodes pour la définition du plan de sondage de 4 produits spécifiques des Enquêtes mensuelles de branche.
Les Enquêtes Mensuelles de Branche (EMB) permettent de suivre l'évolution mensuelle de la production industrielle. Les résultats servent également à calculer l'indice de la production industrielle publié chaque mois. Le suivi des productions s'effectue sur un échantillon d'entreprises et de produits représentatifs.
Actuellement, la méthode de tirage des échantillons des EMB[1] n'est pas aléatoire mais se base sur une méthode de type cut off. Pour un produit donné, la méthode de sélection consiste en effet à :

- trier les entreprises concernées par le produit par valeurs décroissantes de chiffre d'affaires réalisé pour le produit considéré ;
- retenir les entreprises dans l'échantillon jusqu'à couvrir une certaine part (généralement 75%) du chiffre d'affaires total du produit.

Dans le cadre du projet Ocapi (Observation conjoncturelle de l'activité productive industrielle), une étude du plan de sondage a été demandée à la section échantillonnage de la Direction de la méthodologie et de la coordination statistique et internationale de l'Insee avec pour objectif d'étudier l'introduction d'une part aléatoire dans l'échantillonnage des EMB de façon à tenir compte des petites entreprises et pas seulement des plus grandes.
Ainsi, après avoir présenté les indicateurs choisis pour comparer la méthode de tirage actuelle à une méthode de tirage aléatoire[2], l'article présentera les différents plans de sondages mis en œuvre utilisant ces nouvelles méthodes de stratification. Les différentes simulations font varier :

- les critères d'exhaustivité ;
- le nombre de strate ;
- la méthode de définition des strates.

Enfin, le plan de sondage de la solution finalement retenue et les perspectives attendues en termes de précision des résultats seront présentés en détail.

[1] à l'exception de l'EMB concernant la mécanique industrielle

[2] ou comment comparer la réduction du biais de couverture d'une part, avec l'apparition d'une variance d'échantillonnage d'autre part