

***Evaluation de la cohérence des codes de réponse à l'Enquête Unifiée auprès des Entreprises
(EUE)***
STATISTIQUE CANADA

RÉSUMÉ

Mon stage s'est déroulé à Statistique Canada qui est l'organisme fédéral de statistique à Ottawa. J'ai travaillé dans la section de l'estimation au sein de la Division des Méthodes d'Enquêtes auprès des Entreprises. Les méthodologistes étudient, créent et améliorent sans cesse les méthodes statistiques qui sont utilisés à Statistique Canada. Ce sont eux qui assurent la qualité des données publiées.

Dans l'Enquête Unifiée auprès des Entreprises (EUE), des codes de réponse sont attribués lors de la collecte à chaque unité et peuvent être modifiés lors des phases de traitement menant à l'estimation. Les codes de réponse indiquent qu'une unité est répondante, non répondante ou hors champ. Les taux de réponse calculés pour les enquêtes se basent sur ces codes de réponse. Des drapeaux d'imputation sont également attribués dès la collecte à toutes les variables de l'enquête pour chaque unité et peuvent également subir des modifications. Le drapeau d'imputation détermine la situation dans laquelle se trouve la valeur de la variable pour les estimations finales.

Les codes de réponse et les drapeaux d'imputation peuvent subir des changements incohérents. Une unité répondante lors de la phase de collecte peut être hors champ lors de la phase d'estimation. C'est un changement qui paraît incohérent. Ce phénomène existe également au niveau des drapeaux d'imputation.

Ma mission était de réaliser une évaluation de la dynamique des codes de réponses au niveau de l'enquête unifiée de la collecte à l'estimation. Le but de l'évaluation était de proposer un nouveau code de réponse dérivé uniquement à partir des drapeaux d'imputation et d'en analyser l'impact sur les taux de réponses.

L'enquête étudiée est celle des services conseils et comporte trois modes de collecte : l'enquête régulière où le questionnaire complet est envoyé aux entreprises, l'enquête sur les caractéristiques où une partie du questionnaire est envoyé mais les données fiscales sont utilisées, et le remplacement fiscal où aucun questionnaire n'est envoyé et les données fiscales sont utilisées.

J'ai travaillé tout au long de mon stage avec le logiciel SAS pour produire des résultats. Mes connaissances en sondage et statistique descriptive m'ont permis d'analyser les résultats et d'en faire ressortir les éléments importants. Enfin, j'ai utilisé une régression logistique pour réaliser des modélisations et une analyse des correspondances multiples pour des analyses graphiques.

Une première évaluation de la cohérence des codes de réponse et des drapeaux d'imputation par mode de collecte au niveau des unités a démontré une grande stabilité dans l'échantillon. Très peu de changements sont incohérents.

Une analyse des statuts des drapeaux d'imputations selon le statut de réponse de l'unité et le mode de collecte a justifié l'idée d'un lien entre le code de réponse et les drapeaux d'imputation. Afin de construire une modélisation de la situation, une régression logistique et une analyse des correspondances multiples ont été utilisées et ont défini les modalités des variables clés qui avaient le plus de poids sur le statut de réponse des unités selon leur mode de collecte. Grâce à ceci, un encadrement des taux de réponse a été construit pour conclure l'étude.