

Modèle d'énoncé de méthodologie – Division des industries de service, année de référence 2003

Services personnels – Concepts, méthodologie et qualité des données

Ce document renferme des renseignements permettant de bien comprendre les concepts de base qui sont mesurés, les méthodes qui sous-tendent l'enquête (façon de mesurer les concepts) et les aspects clés de la qualité des données. Ces renseignements vous aideront à mieux comprendre les points forts et les limites des données ainsi que la façon appropriée de les utiliser et de les analyser. Les renseignements pourront vous être particulièrement utiles au moment de comparer les données à celles d'autres enquêtes ou sources d'information et de tirer des conclusions au sujet des changements survenus au fil des ans.

Définitions

Une **entité commerciale** est un agent économique qui a la responsabilité et le pouvoir d'affecter des ressources à la production de biens et de services.

Un **établissement statistique** est une entité de production ou le plus petit groupe d'entités de production qui produit un ensemble de biens et/ou de services le plus homogène possible, dont les activités ne débordent pas les frontières provinciales et qui est en mesure de fournir des données sur la valeur de la production ainsi que sur le coût des principaux facteurs intermédiaires utilisés et le coût et l'importance de la main-d'œuvre affectée à la production.

Une **entreprise statistique** est une unité organisationnelle d'une entité commerciale qui dirige et contrôle l'affectation de ressources intéressant ses activités nationales et pour laquelle on établit des états financiers et des bilans consolidés à partir desquels il est possible de tirer des renseignements sur les transactions internationales, la situation des investissements internationaux et la situation financière générale de l'unité.

Une **entreprise complexe** est une entreprise qui est active dans plus d'une province et/ou dans plus d'un secteur et/ou qui compte plusieurs entités juridiques (liens avec plusieurs entités juridiques).

Une **entreprise simple** est une entreprise qui est active dans une province et dans un secteur et qui compte une seule entité juridique.

Les entreprises complexes peuvent comporter une unité organisationnelle au-dessus du niveau de l'établissement, mais au-dessous du niveau de l'entreprise. La **société statistique** est l'unité organisationnelle pour laquelle on établit des comptes des revenus et des dépenses et des bilans à partir desquels il est possible de calculer les bénéfices d'exploitation et le rendement du capital.

Plan d'échantillonnage

La population cible est constituée de tous les établissements statistiques (parfois appelés entreprises ou unités) considérés comme fournissant des services personnels et de blanchissage dans le Système de classification des industries de l'Amérique du Nord (SCIAN) au cours de l'année de référence 2003. Les utilisateurs de données qui désirent en savoir plus sur le SCIAN, sur ses principes sous-jacents et sur de nombreux autres concepts statistiques abordés dans le présent sommaire peuvent consulter l'introduction de la publication de Statistique Canada intitulée « Système de classification des industries de l'Amérique du Nord : Canada 1997 » (n° 12-501F au catalogue).

Même si l'objectif de l'enquête consiste principalement à produire des estimations pour l'ensemble de la branche d'activité – entreprises constituées et non constituées – les entreprises ne sont pas toutes prises en compte. On prélève plutôt un échantillon et la partie de la population admissible à l'échantillonnage se compose de tous les établissements statistiques ayant des revenus supérieurs à un certain seuil (nota : le seuil varie selon les enquêtes et parfois selon les provinces d'une même enquête). Le segment exclu représente une proportion importante (46%) du nombre d'établissements de la branche, mais il intervient pour environ 5 % seulement de l'ensemble des revenus de la branche. Les établissements exclus figurent dans les estimations finales grâce aux données provenant des fichiers

administratifs. Seuls des renseignements de base peuvent être obtenus de cette source, c'est-à-dire : le total des revenus, les dépenses, l'amortissement et les salaires, traitements et avantages sociaux. Des caractéristiques détaillées comme les catégories de clients, les revenus selon le type de service et les postes de dépenses détaillés ne peuvent être obtenues que pour les établissements participant à la collecte des données.

La **base de sondage** est la liste des établissements dont la partie admissible à l'échantillonnage fait l'objet d'un prélèvement. La base fournit des renseignements fondamentaux sur chaque entreprise, notamment : adresse, classification industrielle et renseignements provenant des sources de données administratives (voir ci-dessus). La base de sondage est tenue dans le Registre des entreprises de Statistique Canada et mise à jour à l'aide des données administratives.

Avant le tirage d'un échantillon aléatoire, les établissements sont classés en groupes homogènes (c.-à-d. des groupes portant les mêmes codes SCIAN, se trouvant dans la même province ou le même territoire et comptant le même type d'entreprise (constituée/non constituée). Des exigences de qualité sont établies puis chaque groupe est réparti en sous-groupes appelés strates : tirage complet, tirage obligatoire et tirage partiel.

La strate à tirage complet englobe les principales sociétés sur le plan du rendement sectoriel (selon les revenus). Chaque entreprise est incluse dans l'échantillon, ce qui signifie que chacune se représente elle-même et reçoit un poids de un. La strate à tirage obligatoire est aussi constituée d'unités autoreprésentatives, mais celles-ci sont prélevées en fonction des caractéristiques d'une structure complexe (entreprises à plusieurs établissements, à plusieurs entités juridiques, ayant plusieurs codes SCIAN ou se trouvant dans plusieurs provinces). Les unités appartenant aux strates à tirage partiel font l'objet d'un tirage aléatoire simple.

Enfin, la taille de l'échantillon est augmentée, surtout pour compenser pour les entreprises qui n'appartiennent plus à ce secteur : c'est-à-dire celles qui ont fermé leurs portes, celles dont l'activité principale a changé, celles qui sont inactives ou celles qui figurent en double dans la base de sondage. Une fois que ces entreprises ont été supprimées, **la taille de l'échantillon de cette enquête s'est établie à 1605 entités de collecte.**

Collecte

Les données sont recueillies par envoi et retour par la poste, mais les répondants ont la possibilité de répondre par téléphone ou d'utiliser d'autres méthodes de déclaration électronique. L'unité d'échantillonnage est l'établissement statistique, cependant les établissements appartenant à la même société et au même secteur sont regroupés pour créer une entité de collecte. Cela réduit le fardeau de réponse et simplifie la collecte. Par conséquent, les sociétés dont la production est répartie dans plus d'un établissement reçoivent un seul questionnaire avec instruction de regrouper toutes leurs opérations au Canada. Les questionnaires comportent une section sur la répartition par province, qu'on utilise par la suite pour déterminer la façon d'attribuer les chiffres de production par province.

Vérification et imputation

Les données recueillies sont assujetties à plusieurs vérifications afin d'en assurer la cohérence interne et de cerner les valeurs extrêmes. Lorsqu'une information manque, l'imputation a lieu à l'aide de la procédure « du plus proche voisin » (imputation par enregistrement donneur), en utilisant des données historiques lorsqu'elles existent, ou, enfin, en recourant à des données administratives à la place des données déclarées.

Estimation

Pour arriver aux chiffres définitifs (qu'on appelle estimations), les données de sociétés opérant dans plus d'une province ou territoire sont désagrégées au niveau de la province. Les données administratives servent à estimer la partie du secteur qui n'a pas fait l'objet de l'enquête (c.-à-d. les petites entreprises dont les revenus sont inférieurs aux seuils d'exclusion). On attribue ensuite aux données échantillonnées des coefficients de pondération afin de produire des estimations représentatives de la population cible.

Avant d'être publiés, les résultats combinés de l'enquête sont analysés pour fins de comparabilité; en général, cette opération consiste à examiner en détail les réponses individuelles (surtout celles des grandes sociétés), à revoir la conjoncture générale et les tendances historiques et enfin à comparer ces résultats aux données administratives (p. ex., impôt sur le revenu, taxe sur les produits et services, enregistrements de retenues sur la paie, sources de données d'associations industrielles et de métiers).

Qualité des données

Malgré les efforts déployés pour que les opérations de collecte et de dépouillement se déroulent en fonction de normes rigoureuses, les estimations qui en résultent comportent inévitablement des erreurs. Celles-ci sont classées en deux grandes catégories : les erreurs d'échantillonnage et les erreurs non dues à l'échantillonnage.

Les erreurs non dues à l'échantillonnage ne sont pas liées à l'échantillonnage et peuvent survenir pour de nombreuses raisons. Par exemple, la non-réponse est une source importante d'erreurs non dues à l'échantillonnage. La couverture de la population, les écarts dans l'interprétation des questions, les mauvaises réponses des répondants et les erreurs de consignation, de codage et de traitement des données représentent d'autres exemples d'erreurs non liées à l'échantillonnage.

Le taux de réponse pour cette enquête a été de 70,5% dans l'année de référence 2003, après avoir tenu compte des facteurs expliqués dans le présent énoncé.

Les erreurs d'échantillonnage se produisent parce que les estimations proviennent d'un échantillon de la population plutôt que de la population entière. Ces erreurs dépendent de divers facteurs, notamment la taille de l'échantillon, le plan d'échantillonnage et la méthode d'estimation. Une des propriétés importantes de l'échantillonnage probabiliste tient au fait que les erreurs d'échantillonnage peuvent être calculées à partir de l'échantillon à l'aide d'une mesure statistique appelée coefficient de variation (CV). Pour des enquêtes menées à plusieurs reprises, l'écart relatif entre l'estimation d'un échantillon et l'estimation qu'on aurait obtenue d'un dénombrement de toutes les unités serait inférieur à deux fois le coefficient de variation, 95 fois sur 100. L'éventail des données acceptables d'un échantillon s'appelle l'intervalle de confiance. Les intervalles de confiance peuvent être calculés autour des estimations en utilisant les CV. Premièrement, nous calculons l'erreur type en multipliant l'estimation d'un échantillon par le CV. L'estimation d'un échantillon plus ou moins deux fois l'erreur type est ce qu'on appelle l'intervalle de confiance de 95 %.

Pour l'Enquête annuelle sur le secteur des services personnels, les CV ont été calculés pour chaque estimation produite. De façon générale, les variables déclarées plus fréquemment ont résulté en de très bons CV (10 % ou moins), tandis que les variables déclarées moins souvent ont été associées à des CV plus élevés mais néanmoins acceptables (moins de 25 %). On peut obtenir ces CV sur demande.