

## SURVEY METHODOLOGY

### Coverage and Frame

The frame is the list that allows access to the firms eligible for direct data collection. The frame is maintained by Statistics Canada's Business Register, using taxation account information (i.e., income tax, goods and services tax and payroll deduction records) submitted to Canada Customs and Revenue Agency.

Even though the basic objective of the survey is to produce estimates for the whole industry -- incorporated and unincorporated -- the portion of the population eligible for sampling was defined as all incorporated establishments with revenue above \$50,000.

Some exceptional unincorporated units were also added to direct data collection if their contribution was deemed significant. The same principle applies to unincorporated units belonging to complex enterprises.

The main motivation for the exclusion of unincorporated and incorporated firms below \$50,000 from direct data collection was to achieve major reductions in response burden. Typically, the excluded portion might represent a substantial proportion of the whole industry in terms of number (55% to 60%) but their contribution to the overall estimate is usually modest (less than 10%). Firms below the exclusion thresholds are still part of the universe but their contribution is accounted for in the final estimates through the use of administrative taxation records as proxy data. Only basic information can be obtained from this source, that is, total revenue, total expenses, depreciation, salaries, wages and benefits.

Detailed characteristics such as client base, revenue by type of service and detailed expense items can only be obtained for firms in the direct data collection portion.

The definition and description of the industry is based on the North American Industry Classification System (NAICS), adopted by Statistics Canada and used for the 1997 reference year for the first time. Data users who wish to learn more about NAICS, its underlying principles, and many of the other statistical concepts discussed in this brief summary, are referred to the Introduction section of the Statistics Canada publication "North American Industry Classification System: Canada 1997" (catalogue number 12-501).

## MÉTHODOLOGIE DE L'ENQUÊTE

### Couverture et base de sondage

La base de sondage est la liste qui donne accès aux entreprises qui se prêtent à la collecte directe de données. La base de sondage est tenue à jour par le Registre des entreprises de Statistique Canada à l'aide des données des comptes d'impôt (c.-à-d. les enregistrements de données sur l'impôt sur le revenu, la taxe sur les produits et services et les retenues sur la paie) présentées à l'Agence des douanes et du revenu du Canada.

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 -- on a établi que la partie de la population admissible à l'échantillonnage se composait de tous les établissements constitués ayant un chiffre d'affaires supérieur à 50 000 \$.

On a aussi ajouté par exception à l'étape de la collecte directe des données quelques unités non constituées lorsque leur contribution était jugée suffisamment importante. Le même principe s'applique aux unités non constituées appartenant à des entreprises complexes.

Si l'on a exclu de la collecte directe de données les entreprises constituées de moins de 50 000 \$, c'était surtout pour diminuer substantiellement le fardeau de réponse. De façon typique, le nombre d'entreprises exclues serait susceptible de représenter un pourcentage élevé de toute la branche d'activité (de 55 % à 60 %), cependant leur importance dans l'estimation globale est habituellement modeste (moins de 10 %). Bien que les entreprises en deçà du seuil d'exclusion appartiennent quand même à l'univers, elles figurent cependant dans les estimations finales au moyen de données de substitution provenant des enregistrements administratifs de l'impôt. Seuls des renseignements de base peuvent être obtenus de cette source, à savoir les recettes totales, les dépenses totales, l'amortissement ainsi que les traitements et avantages.

Les caractéristiques détaillées comme la base de clients, les recettes selon le type de service et les postes de dépenses détaillés ne peuvent être obtenues pour les entreprises que par la collecte directe des données.

La définition et la description de cette branche d'activité proviennent du Système de classification des industries de l'Amérique du Nord (SCIAN), adopté par Statistique Canada et utilisé pour la première fois pour l'année de référence 1997. Les utilisateurs de données qui désirent en savoir plus sur le SCIAN, ses principes sous-jacents et nombre d'autres concepts statistiques abordés dans ce sommaire peuvent se reporter à l'introduction de la publication de Statistique Canada intitulée « Système de classification des industries de l'Amérique du Nord : Canada 1997 », n° 12-501 au catalogue).

## Selection of the Sample

The survey design was based on probability sampling and only covered the portion of the frame subject to direct data collection. Each sampled company represented a number of other, similar firms in the industry, based on the probability of being surveyed. The largest firms were included in the sample with certainty due to their significant contribution to industrial performance.

The companies randomly selected for this survey represent only a small sample of the whole survey frame. To make the sample as efficient as possible, the frame was divided into groups of companies based on revenue and province or territory of location. The size of the sample was increased to compensate for non-response (for example, companies which cannot be contacted because they have moved or gone out of business). The resulting sample size for this survey was 1283 companies.

## Collection

This survey collected data through a mail-out/mail-back process, while attempting to provide respondents with the option of telephone or other electronic filing methods as required.

Companies with production at more than one locale were mailed only one questionnaire, and were instructed to report for all their operations in the surveyed industry. Summary data were also collected for each province or territory in which the company operated.

## Estimation

Prior to estimation, companies with production in more than one province or territory were sub-divided in order to generate accurate estimates at that level. The survey data collected from the sample were then weighted to produce estimates representative of the target population. Tax information was used to estimate the portion that was excluded from survey activity (i.e., unincorporated firms and incorporated firms with revenue less than \$50 000).

Non-response was handled using available auxiliary information to substitute the data from a company with similar characteristics.

## Prélèvement de l'échantillon

Le plan d'enquête reposait sur un échantillonnage probabiliste et s'appliquait uniquement à la partie du plan admissible à la collecte directe de données. Chaque société échantillonnée représentait un nombre d'entreprises similaires de la branche d'activité, en fonction de leur probabilité d'échantillonnage. Les grandes entreprises ont été incluses dans l'échantillon avec certitude à cause de leur apport important aux résultats de celle-ci.

Les sociétés prélevées au hasard aux fins de cette enquête représentent uniquement un petit échantillon de toute la base de sondage. Pour accroître au maximum l'efficacité de l'échantillon, on a réparti la base en groupe de sociétés d'après leur chiffre d'affaires et leur province ou territoire d'emplacement. On a augmenté la taille de l'échantillon afin de compenser pour la non-réponse (p. ex., des sociétés qui ne peuvent être jointes parce qu'elles ont déménagé ou fermé leurs portes). La taille finale de l'échantillon pour cette enquête était de 1283 sociétés.

## Collecte

La collecte des données a eu lieu par envoi et retour par la poste, mais les répondants ont aussi eu la possibilité de répondre par téléphone ou d'utiliser d'autres méthodes de déclaration électronique.

Les sociétés dont la production était répartie à plus d'un endroit ont reçu un seul questionnaire par la poste, avec instruction de regrouper dans leur questionnaire toutes leurs activités dans la branche d'activité en question. On a aussi recueilli des données récapitulatives pour chaque province ou territoire dans lequel la société exerçait ses activités.

## Estimation

Avant l'étape de l'estimation, on a subdivisé les sociétés ayant des activités de production dans plus d'une province ou territoire afin d'obtenir des estimations précises à ce niveau. On a ensuite pondéré les données d'enquête recueillies auprès de l'échantillon afin de produire des estimations représentatives de la population visée. Les données fiscales ont servi à calculer les estimations relatives aux entreprises exclues de l'enquête (c.-à-d. les entreprises non constituées et les entreprises constituées ayant un chiffre d'affaires inférieur à 50 000 \$).

La non-réponse est substituée par des données provenant d'une société affichant des caractéristiques semblables.

The combined survey results were analyzed before publication; in general, this included a detailed review of the individual responses (especially for the largest companies), a review of general economic conditions as well as historic trends and comparisons with tax data information and other administrative data sources (i.e., industry and trade associations).

## Sampling and Non-Sampling Errors

While considerable effort was made to ensure high standards throughout all collection and processing operations, the resulting estimates are inevitably subject to a certain degree of error.

Non-sampling errors are not related to sampling and may occur for many reasons. For example, non-response is an important source of non-sampling error. Population coverage, differences in the interpretation of questions, incorrect information from respondents, mistakes in recording, coding and processing of data are other examples of non-sampling errors.

Sampling errors can occur because estimates are derived from a sample of the population rather than the entire population. These errors depend on factors such as sample size, sampling design and the method of estimation. An important property of probability sampling is that sampling errors can be measured from the sample itself.

On a analysé les résultats combinés de l'enquête avant de les publier; en général, cette opération a consisté 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 fiscales et à d'autres sources de données administratives (c.-à-d. les associations industrielles et de métiers).

## Erreurs d'échantillonnage et erreurs non dues à l'échantillonnage

Même si d'immenses efforts ont été faits pour assurer une qualité élevée pendant toutes les activités de collecte et de traitement, les estimations finales sont inévitablement sujettes à un certain degré d'erreur.

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, des renseignements erronés des répondants et des erreurs de consignation, de codage et de traitement des données représentent d'autres exemples d'erreurs non liées à l'échantillonnage.

Des erreurs d'échantillonnage risquent de se produire 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 mesurées à partir de l'échantillon même.

## Data Quality

This survey is based on probability sampling and so the potential error introduced by sampling can be measured. A standard measure of sampling error is the coefficient of variation (CV). If we were to conduct the survey repeatedly, the relative difference between a sample estimate and the estimate that should have been obtained from an enumeration of all the units would be less than twice the coefficient of variation, 95 times out of 100.

For Total Revenue for this survey (chosen because it is the estimate most commonly used in measuring and describing an industry), the CVs at the Canada level for the following industries are:

1. Computer Systems Design and Related Services (NAICS 541510) = 3%
2. Software Publishers (NAICS 511210) = 6%
3. Data Processing Services (NAICS 514210) = 6%

CVs are calculated for all estimates; generally, the less commonly reported variables are associated with higher but still acceptable CVs, usually under 25%. These CVs are not included in this release but are available upon request.

## Qualité des données

Comme cette enquête repose sur un échantillonnage probabiliste, il est possible de mesurer l'erreur potentielle attribuable à l'échantillonnage. Le coefficient de variation (CV) est une mesure type de l'erreur d'échantillonnage. Si l'enquête était menée à plusieurs reprises, l'écart relatif entre l'estimation d'un échantillon et l'estimation qu'on aurait dû obtenir d'un dénombrement de toutes les unités serait inférieur à deux fois le coefficient de variation, 95 fois sur 100.

En ce qui concerne le chiffre d'affaires total pour cette enquête (choisi parce qu'il constitue l'estimation la plus souvent utilisée pour mesurer et décrire une branche d'activité), les CV, au niveau du Canada, pour les industries suivantes sont :

1. Conception de systèmes informatiques et services connexes (SCIAN 541510) = 3%
2. Éditeurs de logiciels (SCIAN 511210) = 6%
3. Services de traitement des données (SCIAN 514210) = 6%

On calcule des CV pour toutes les estimations; en général, on associe les variables le moins souvent déclarées à des CV élevés mais quand même acceptables, habituellement inférieurs à 25%. Ces CV ne figurent pas dans ce document mais on peut les obtenir sur demande.

## **North American Industry Classification (NAICS) Description**

### **1. Computer Systems Design and Related Services (NAICS 541510)**

This industry comprises establishments primarily engaged in providing expertise in the field of information technologies through one or more activities, such as writing, modifying, testing and supporting software to meet the needs of a particular customer, including the creation of Internet home pages; planning and designing computer systems that integrate hardware, software and communication technologies; on-site management and operation of clients' computer and data processing facilities; providing advice in the field of information technologies; and other professional and technical computer-related services.

### **2. Software Publishing (NAICS 511210)**

This industry comprises establishments primarily engaged in computer software publishing. Establishments in this industry carry out operations necessary for producing and distributing computer software, such as designing, providing documentation, assisting in installation and providing support services to software purchasers. These establishments may design and publish, or publish only.

### **3. Data Processing Services (NAICS 514210)**

This industry comprises establishments primarily engaged in providing electronic data processing services. These establishments may provide complete processing and preparation of reports from data supplied by the customer; specialized services, such as automated data entry; or they may make data processing resources available to clients on an hourly or time-sharing basis.

## **For More Information**

For more information, or to provide your comments and feedback, please contact Service Industries Division at (613) 951-3863.

## **Les descriptions de système de classification des industries de l'Amérique du Nord (SCIAN)**

### **1. Conception de systèmes informatiques et services connexes (SCIAN 541510)**

Cette classe comprend les établissements dont l'activité principale consiste à fournir une expertise dans le domaine des technologies de l'information. Ces établissements peuvent notamment : concevoir, modifier, tester et prendre en charge des logiciels pour répondre aux besoins d'un client, y compris créer des pages d'accueil Internet; planifier et concevoir des systèmes informatiques intégrant la technologie du matériel, celle des logiciels et celle des communications; gérer et exploiter sur place les installations d'informatique et de traitement des données de clients; donner des conseils dans le domaine des technologies de l'information; fournir d'autres services professionnels et techniques de nature informatique.

### **2. Éditeurs de logiciels (SCIAN 511210)**

Cette classe comprend les établissements dont l'activité principale est l'édition de logiciels. Les établissements de cette classe exécutent les opérations nécessaires à la production et à la distribution de logiciels telles que la conception, la fourniture de la documentation, l'assistance en matière d'installation et la prestation de services de soutien aux acheteurs de logiciels. Ces établissements peuvent se livrer à la conception et à l'édition, ou se consacrer uniquement à l'édition.

### **3. Services de traitement des données (SCIAN 514210)**

Cette classe comprend les établissements dont l'activité principale consiste à fournir des services de traitement électronique des données. Ces établissements peuvent offrir des services complets de traitement et de préparation de rapports à partir des données fournies par les clients, offrir des services spécialisés, comme la saisie automatisée de données, ou mettre à la disposition des clients des ressources de traitement des données sur une base horaire ou sur une base de temps partagé.

## **Pour de plus amples renseignements**

Pour de plus amples renseignements, ou pour nous faire parvenir vos commentaires et rétroactions, veuillez communiquer avec la division des industries de service au (613) 951-3863.