

(Personal Services Survey) – Concepts, Methodology and Data Quality

The following information should be used to ensure a clear understanding of the basic concepts that define the data provided in this product, of the underlying methodology of the survey, and of key aspects of the data quality. This information will provide you with a better understanding of the strengths and limitations of the data, and of how they can be effectively used and analysed. The information may be of particular importance to you when making comparisons with data from other surveys or sources of information, and in drawing conclusions regarding change over time.

Definitions

A **business entity** is an economic transactor having the responsibility and the authority to allocate resources in the production of goods and services.

A **statistical establishment** is one production entity or the smallest grouping of production entities which produces as homogenous a set of goods and/or services as possible; which does not cross provincial boundaries; and for which records provide data on the value of output together with the cost of principal intermediate inputs used and cost and quantity of labour resources used to produce the output.

A **statistical enterprise** is an organisational unit of a business that directs and controls the allocation of resources relating to its domestic operations, and for which consolidated financial and balance sheet accounts are maintained from which international transactions, an international investment position and a consolidated financial position for the unit can be derived.

A **complex enterprise** is an enterprise that operates in more than one province and/or industry and /or is multi-legal (links to more than one legal entity).

A **simple enterprise** is an enterprise that operates in one province, one industry and is a single legal entity.

In complex businesses, there may be an organisational unit above the establishment but below the enterprise. The **statistical company** is defined as the organisational unit for which income and expenditure accounts and balance sheets are maintained from which operating profit and the rate of return on capital can be derived.

Coverage and Frame

The target population consists of all statistical establishments classified as **personal and laundry services** according to the North American Industrial Classification System (NAICS) during the reference year 1999. 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 Industrial Classification System: Canada 1997” (catalogue number 12-501).

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 statistical 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 firms and incorporated firms below \$50,000 from direct data collection was to achieve major reductions in the response burden. The excluded portion represents a substantial proportion of the whole industry in terms of number (68%) but its contribution to the overall estimate is modest (22%). 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 records as proxy data. Only basic information can be obtained from this source that is; total revenue, expenses, depreciation and wages, salaries and benefits. For the 1999 reference year, the administrative data used to estimate for unincorporated firms was restricted to data that could be directly linked to the frame. This is expected to improve the quality of the unincorporated firm estimate. This change necessitated historical data revisions for some industries. Detailed characteristics such as client base, revenue by type of service and detailed expenses items can only be obtained for firms in the direct data collection portion.

The frame is the list that identifies the firms classified to the industry in question. The frame is maintained by Statistics Canada's Business Register, using taxation account information (i.e. income tax, goods and services tax and payroll deductions records) submitted to Canada Customs and Revenue Agency.

Survey Design

The survey design covered only the portion of the frame subject to direct data collection. Prior to the selection of a random sample, units are grouped in homogeneous groups defined using industrial (NAICS) and geographic (province/territory) attributes. Similar quality requirements are targeted for each group which is then divided into four sub-group called strata: must-take, take-all, large take-some and small take-some.

The take-all stratum includes the largest firms in terms of industrial performance, which are selected in the sample with certainty making such units self-representing. The must-take stratum is also comprised of self-representing units that have a complex structure (multi-establishments, multi-legal, multi-NAICS or multi-province enterprises). Units in the two take-some strata are subjected to a random sample where each sampled firm represents a number of other, similar firms in the industry/province combination according to the inverse of their probability of selection.

Finally, the size of the sample was increased to compensate for such situations as non-response and firms which cannot be contacted because they have moved or gone out of business. The resulting sample size for this survey after removing firms that should not have been included in the frame (out of scope, duplicate records or out of business) was 1497 companies.

Collection

Data is collected through a mail-out/mail-back process, while attempting to provide respondents with the option of telephone or other electronic filing methods as required. Even though the sampling unit was the statistical establishment, the statistical company was chosen as the

collection entity in order to reduce respondent burden and simplify collection procedures. Therefore, 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 collected for each province or territory in which the company operated.

Edit & Imputation

Several checks are performed on the data to verify internal consistency and identify extreme values. For non-response, imputation is performed using a "nearest neighbour" procedure (donor imputation) using available auxiliary information to substitute the data from a company with similar characteristics. In addition to the donor imputation approach, imputation can also be done using historical responses.

Estimation

Prior to estimation, data for companies with production in more than one province or territory were allocated to the provincial level. The survey data collected from the sample were then weighted using the inverse of the probability of selection of each sampled unit to produce estimates representative of the target population. Administrative data were used to estimate the portion that was excluded from survey activity (i.e. unincorporated firms and incorporated firms with revenue less than \$50,000).

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 (e.g. industry and trade associations).

Data Quality

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. These errors can be broken down into two major types: sampling and non-sampling.

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.

The response rate for this survey was 73,9%, after taking into account the fact that some firms were no longer in business, or had changed their primary business activity.

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 computed from the sample itself by using a statistical measure called the coefficient of variation (CV). Over repeated surveys, the relative difference between a sample estimate and the estimate that would have been obtained from an enumeration of all the units would be less than twice the coefficient of variation, 95 times out of 100. Confidence intervals can be constructed around the estimate using the CV's. First, we calculate the

standard error by multiplying the sample estimate by the CV. The sample estimate plus or minus twice the standard error is then referred to as 95 % confidence interval.

For the 1999 **Annual Survey of Personal Services** CVs were calculated for each estimate. Generally, the more commonly reported variables obtained very good CVs (10% or less) while the less commonly reported variables were associated with higher but still acceptable CVs (under 25%). These CVs are available upon request.