



# Labour Force Survey

Statistical Data Documentation System  
Reference Number 3701

## Data Quality Statements



Statistics  
Canada

Statistique  
Canada

Canada

# Data Quality Statement for the Labour Force Survey

## Introduction

This statement provides information to assist users of data from the Labour Force Survey (LFS) in assessing how well the data fits their statistical needs. It describes the quality objectives of the LFS, the factors that affect LFS data quality and assessments of impact of those factors on LFS data. Factors affecting data quality discussed in this statement include both those that impact on the overall quality of LFS estimates and a number of factors that impact on specific LFS data items and classifications.

This statement compliments information contained in the *Guide to the Labour Force Survey*, Catalogue no. 71-543-XIE ([www.statcan.ca/english/concepts/labour/index.htm](http://www.statcan.ca/english/concepts/labour/index.htm)), which provides useful background information on the survey concepts and definitions, the questionnaire, and a brief discussion of the methodology and quality issues. A more complete description of the survey methodology can be found in *Methodology of the Canadian Labour Force Survey*, Catalogue no. 71-526-XPB.

## About the Labour Force Survey

The Labour Force Survey collects monthly information on the labour market activities of Canada's working age population. The data is collected from a sample of 53,500 households each month, involving almost 100,000 respondents age 15 and over. The sample is designed to represent all persons in the population 15 years of age and over, excluding residents of the Territories<sup>1</sup>, persons living on Indian reserves, residents of institutions (prisons, hospitals and nursing homes), and full-time members of the Canadian Armed Forces.

The primary objective of the survey is to provide reliable estimates of employment and unemployment for Canada and the provinces, and to provide descriptive and explanatory data for each group. Data is also available for sub-provincial geographies. Most information collected by the survey is consistent back to 1976, providing a rich source for labour market trend analysis.

The LFS is the only source of monthly estimates of total employment including the self-employed, full and part-time employment, and unemployment. It publishes monthly standard labour market indicators such as the unemployment rate, the employment rate and the participation rate. The LFS is a major source of information on the personal characteristics of the working-age population, including age, sex, marital status, educational attainment, and family characteristics.

Employment estimates include detailed breakdowns by demographic characteristics, industry and occupation, job tenure, and usual and actual hours worked. The survey incorporates questions

---

<sup>1</sup> The Labour Force Survey has been conducted in the Yukon since 1992. Data are available on request. A pilot project to extend collection to Northwest Territories and Nunavut is currently underway.

permitting analyses of many topical issues, such as involuntary part-time employment, multiple job-holding, and absence from work. Since January 1997, it also provides monthly information on the wages and union status of employees, as well as the number of employees at their workplace and the temporary or permanent nature of their job.

Unemployment estimates are produced by demographic group, duration of unemployment, and activity before looking for work. Information on industry and occupation, and reason for leaving last job is also available for persons currently unemployed or not in the labour market with recent labour market involvement.

In addition to providing national and provincial estimates, the LFS also releases estimates of labour force status for sub-provincial areas such as Economic Regions (ERs) and Census Metropolitan Areas (CMAs).

The Labour Force Survey is also specifically mandated under the Employment Insurance Act to produce regional unemployment rates with a specified level of reliability for use in the Employment-Insurance Program to determine the number of hours of work necessary to qualify for employment insurance, and the number of weeks of benefits.

A secondary objective of the survey is to provide the sample base for most household surveys conducted by Statistics Canada (STC).

The results of the survey are used widely by federal and provincial policy makers, program administrators, program evaluators, economic forecasters, and labour market analysts. LFS data are also used indirectly in the System of National Accounts, and other STC program areas.

The importance of the outputs of the survey, and its high visibility require that the program has a well-developed set of quality control measures and procedures, as well as the ability to respond to any ad hoc queries and problems. The survey has undergone significant changes in recent years (redesign of sample and questionnaire, new collection and processing systems, and redevelopment of outputs). In addition, technological advances are continually providing new opportunities for monitoring and analyzing quality. Together, these changes have fostered a climate of flexibility and continual reassessment and re-engineering of the methods and processes used to monitor and promote quality.

To monitor and ensure the continued quality of its data, the LFS has an extensive data quality program. A whole range of quality indicators are produced on a regular basis and carefully analysed. In the presence of unusual values, those responsible for the relevant LFS operations are immediately advised, to guarantee data quality from one month to the next. Some indicators are also monitored in a less regular manner, since their role is to assist in the identification of long-term trends or effects, for example, the consequences of certain operational or sample design changes. This long-term information about the reliability of data is used to improve the general quality of results and help data analysts and users in their interpretation of the data.

## Quality objectives of the Labour Force Survey

The aim of the LFS is to provide highly reliable labour market data at the national and provincial levels, as well as for the Employment Insurance Regions (as required by the Employment Insurance Act). The LFS is also able to produce data, to a lesser degree of reliability, for other sub-provincial regions (e.g. Economic Regions and larger Census Agglomerations), according to standard geographic definitions.

In order to achieve required levels of reliability, maximum limits are set for the sampling errors of estimates that are produced from the survey at various levels of geography. These targets are presented below in the section on “Sampling error”.

## Factors that affect the quality of LFS data

The LFS is a sample survey. All sample surveys are subject to error. Errors in sample surveys come in two broad groups. First, are *sampling errors*. These occur because only a small proportion of the total population of interest is used to produce estimates that represent the total population. Sampling errors can be reliably measured as they are calculated based on the scientific methods used to design and draw survey samples.

Second are *non-sampling errors*. These occur when the survey processes work less effectively than intended. Non-sampling errors can be attributed to numerous sources. Not all persons selected in the LFS respond to the survey. In addition, the survey does not capture all persons it is intended to cover. The persons that do respond do not always fully understand the LFS questions or accurately report the information sought. LFS interviewers occasionally make mistakes in recording the information that is provided. From time to time, errors can occur in processing the data.

All non-sampling errors have the potential to bias survey estimates away from the true value of what is being measured. Some errors may result in a systematic bias to survey data. Other errors may result in one-time changes to a data series (a break in series) or may impact on data series in an ad hoc way.

It is usually easier to identify that a non-sampling error is present in a survey than to quantify its impact on survey estimates. For this reason, a number of indicators of potential bias are produced and monitored on a regular basis. Occasional studies are undertaken to ensure that these non-sampling errors are not having any significant impact on the LFS. The key indicators of non-sampling error and assessments of their impact are included below. Where possible these assessments are based on specific studies. Where this is not possible, assessments are based on the subjective evaluation of survey experts.

Users should also be aware that *survey changes* could influence how data are interpreted. Changes occur from time to time to the survey design and methods, and in the definitions and classifications that are used. Changes are necessary to keep the survey up to date and relevant. In fact, some of these changes are made in order to improve data and therefore changes to data are intended. Other changes are made for operational efficiency. Occasionally, survey changes may unintentionally impact on data. When an impact on data is anticipated steps are taken to measure the impact. Some recent changes to the survey are reported in this statement, along with an assessment of their impact on data.

## Sampling error

The survey divides Canada into its provinces and a number of geographic areas within each province. Within each of these areas a random selection of dwellings is made. Approximately 0.5% of dwellings in Canada are selected for the sample. As the sample is selected at random, there are many possible samples that could have been selected. Estimates from these samples would differ from each other. Sampling error represents the variation in estimates from all possible samples. The sampling error, therefore, indicates how closely an estimate selected from the sample actually approximates the true value for the population. Sampling errors for the LFS levels are generally reported as coefficients of variation, while standard errors are used for estimates of change.

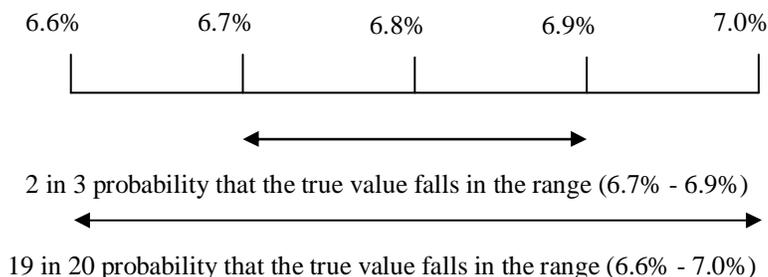
### Standard error

#### *Definition*

A standard error (SE) is the level of sampling error at which there is a 2 in 3 probability that the true estimate will lie between + or - one SE of the survey estimate. Further, there is a 19 in 20 probability that estimates from all possible samples would lie between + or - 2 SEs of the survey estimate. Standard errors apply to all LFS estimates, irrespective of whether they are levels, movements or rates. Example 1 provides an illustration of the application of SEs.

#### Example 1

In January 2000, the unemployment rate for Canada was estimated as 6.8%. The SE for this rate is 0.1 percentage point. The sampling error can therefore be illustrated as:



#### *Implications for data quality*

All LFS estimates are subject to sampling error and the level of sampling error varies from estimate to estimate. Broadly, the larger the number of observations on which an estimate is based, the lower the sampling error. In the same way, the larger the estimate, the lower the relative sampling error. For estimates representing small geographic areas and small population groups, the relative sampling errors for the estimates may be so high that the estimate can not be used with any degree of confidence.

#### *LFS quality criteria based on sampling error*

The LFS is designed so that certain sampling error criteria are met. Sufficient sample is selected to ensure that the following targets for relative standard errors (the standard error of an estimate divided by the estimate, also known as the coefficient of variation or CV) for unemployment rates are met:

- about 2% for Canada
- 4 to 7% for the provinces
- 15% or less for three month moving average estimates for Employment Insurance regions, including Census Metropolitan Areas.

Although there is no formal requirement, a target CV of 25% or less for three month moving averages of unemployment has been set for Economic Regions. In general, CVs are much lower for most Economic Regions, but can be higher in regions with relatively sparse population.

Currently, employment data are published for many of the larger Census Agglomerations on a 3-month moving average basis, but unemployment data are published for only a few of these areas, where CVs are 25% or below.

#### *Interpreting estimates for small areas*

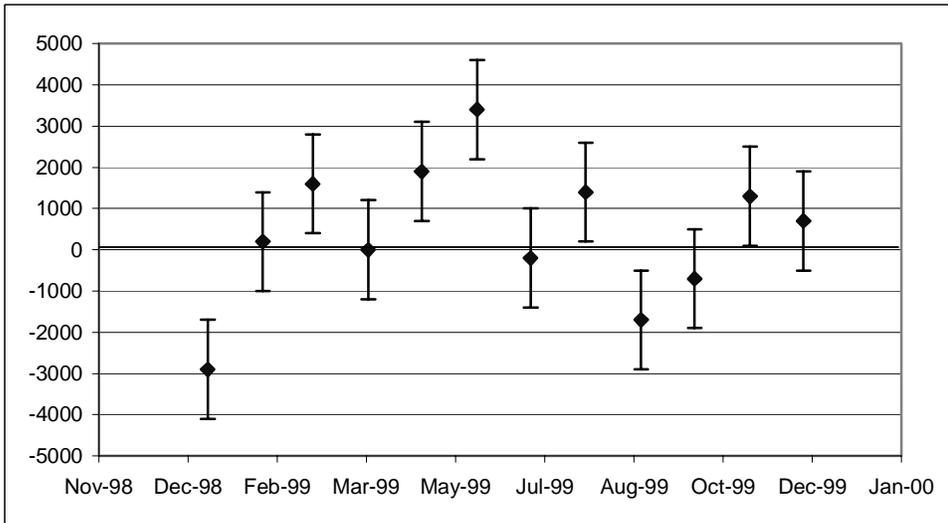
The LFS sample is distributed to meet the CV targets discussed above. To impose CV targets for other (lower) levels of geography, such as cities not designated as Census Metropolitan Areas, would require increases in sample size that would be very expensive and impose a higher level of response burden on Canadian households. Therefore, estimates for areas other than those for which CV targets are set may have CVs approaching or greater than 25%. Because of this level of sampling error, estimates may be very volatile from month to month and users should interpret the movements within the context of trends and other relevant labour market information.

Many LFS data series are only published as 3-month moving average estimates in order to reduce volatility. Example 2 illustrates SEs for a monthly movement in employment for a CMA and the improvement that can be gained by using a 3-month moving average.

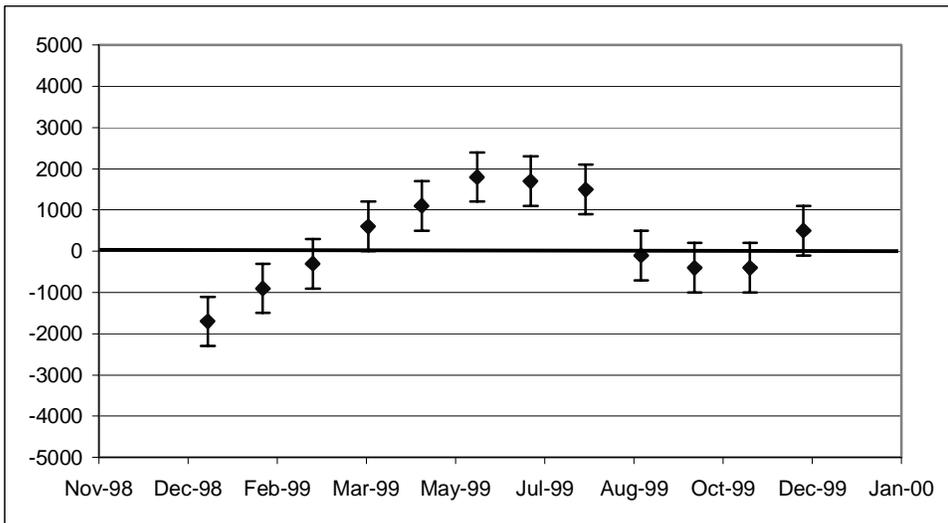
#### Example 2

The two charts below graph monthly employment movements for Trois-Rivières in 1999. In each case the range of + and – 1 SE is displayed for each estimate. Chart 1 shows movements between single month estimates of employment. Chart 2 shows movements between 3-month moving average estimates of employment. As can be seen, the movements in Chart 1 display greater volatility. The SEs for these movements are twice as large as those for 3-month moving average estimates. Although estimates based on the 3-month moving averages will not be as up to date as single month estimates for the latest month, they provide a better indication of the trend in a particular series.

**Chart 1 – Movements in monthly employment, Trois-Rivières**



**Chart 2 – Movements in 3-month moving average employment, Trois Rivières**



Users should also be aware that small area estimates from the LFS are generally not seasonally adjusted. Volatility that can appear in an unadjusted data series may simply be a result of regular seasonal variation. One method of avoiding seasonal variation in unadjusted data is to compare estimates from the same month from year to year. However, while year-over-year comparison eliminates seasonality, estimates may have high standard error. Users should interpret the year-over-year change in light of longer-term trends and other relevant labour market information. For an explanation of seasonal adjustment and procedures used in the LFS see the *Guide to the Labour Force Survey* ([www.statcan.ca/english/concepts/labour/index.htm](http://www.statcan.ca/english/concepts/labour/index.htm)).

Tables of standard errors (often expressed as CVs) for monthly and annual average LFS estimates by province are presented in Tables A and B below and can also be found in Chapter 7 of the *Guide to the Labour Force Survey*. The *Labour Force Historical Review* on CD-ROM (Catalogue no. 71F0004XCB) includes a calculator program to produce CVs for LFS estimates. Standard errors can also be obtained from Labour Statistics Divisions Client Services at 1-866- 873-8788 ([labour@statcan.ca](mailto:labour@statcan.ca)).

**Table A: CVs for estimates\* of monthly totals for Canada and the provinces**

Geographic Area	Coefficient of variation								
	1.0%	2.5%	5.0%	7.5%	10.0%	15.0%	20.0%	25.0%	30.0%
Canada	1,089.6	308.1	132.1	75.6	41.9	24.0	15.8	11.3	8.6
Newfoundland and Labrador	280.1	73.6	29.3	16.2	9.0	4.9	3.2	2.2	1.7
Prince Edward Island	87.2	23.5	9.3	5.2	3.0	1.7	1.1	0.8	0.6
Nova Scotia	245.6	70.7	30.7	17.7	9.9	5.7	3.8	2.7	2.1
New Brunswick	263.2	69.8	27.3	15.1	8.7	4.8	3.1	2.2	1.7
Quebec	1,182.6	324.2	133.0	74.8	42.3	23.8	15.5	11.1	8.4
Ontario	1,024.3	289.6	124.2	71.1	39.4	22.5	14.9	10.7	8.1
Manitoba	223.6	66.2	29.9	17.6	9.6	5.7	3.8	2.7	2.1
Saskatchewan	227.0	63.0	26.7	15.1	8.3	4.7	3.1	2.2	1.7
Alberta	535.6	159.4	72.3	42.6	23.4	13.8	9.2	6.7	5.1
British Columbia	804.3	224.1	95.2	54.1	29.8	16.9	11.1	7.9	6.0

\* Estimates are in thousands.

**Table B: CVs for estimates\* of annual averages for Canada and the provinces**

Geographic Area	Coefficient of variation								
	1.0%	2.5%	5.0%	7.5%	10.0%	15.0%	20.0%	25.0%	30.0%
Canada	407.1	114.8	50.3	28.9	15.5	8.9	5.9	4.2	3.2
Newfoundland and Labrador	77.3	21.1	9.4	5.4	2.7	1.5	1.0	0.7	0.5
Prince Edward Island	19.2	5.8	2.7	1.6	0.8	0.5	0.3	0.2	0.2
Nova Scotia	66.3	20.3	10.0	6.0	3.1	1.9	1.3	0.9	0.7
New Brunswick	60.3	17.3	8.0	4.7	2.4	1.4	0.9	0.7	0.5
Quebec	303.0	89.7	42.3	25.2	13.0	7.7	5.2	3.8	2.9
Ontario	277.7	84.9	41.7	25.3	13.0	7.9	5.4	3.9	3.0
Manitoba	76.5	22.8	11.1	6.6	3.4	2.0	1.4	1.0	0.8
Saskatchewan	71.0	19.1	8.7	5.0	2.4	1.4	0.9	0.6	0.5
Alberta	174.9	52.5	25.5	15.4	7.8	4.7	3.2	2.3	1.8
British Columbia	212.4	63.0	30.3	18.2	9.2	5.5	3.7	2.7	2.0

\* Estimates are in thousands.

## Non-sampling errors

This section presents information about the main non-sampling errors affecting the LFS data in general.

### Non-response

#### *Definition*

In each month of the LFS, there will be a number of people in selected households for whom data are not collected. The non-response rate for the LFS is defined as the number of households from which no data or insufficient data were collected divided by the total number of households in the survey sample.

#### *Implications for data quality*

A high response rate to a survey is important for two reasons. First, the sampling error for the survey is based on the number of responding households in the survey. The higher the response rate, the lower the sampling error for the survey. Second, non-respondents to a survey may have different characteristics on average compared to those who respond to the survey. Depending on the level of non-response and the extent that differences exist, the survey estimates could be biased toward the characteristics of survey respondents.

The LFS receives a high level of co-operation from individuals in Canadian households selected in the survey. Between 1990 and 1999, annual non-response rates for the LFS averaged around 5 to 6%. The average non-response rate for 2000 and 2001 has increased to about 8%.

### *Assessment of impact on quality*

Non-responding households tend to be slightly smaller on average than responding households which could result in a slight bias on some estimates. However, studies have been inconclusive as to whether there is any bias on unemployment estimates. Any bias would affect LFS level estimates and is unlikely to have any impact on month to month movements.

### Coverage error

#### *Definition*

Coverage errors occur in a sample survey when the sample does not fully represent its target population. For the LFS, the main coverage problem is undercoverage. That is, the survey does not pick up as many people as expected. This can occur for various reasons, such as dwellings not being listed for selection and persons being missed in households. One way of measuring undercoverage is by a ratio known as the slippage rate. The slippage rate for the LFS is determined as the percentage difference between the post-censal population estimates and the population estimates produced from the LFS.

#### *Implications for data quality*

As with non-response, undercoverage may bias survey estimates as those people missed by the survey may have characteristics that differ on average from those people responding to the survey. Again the impact on survey estimates will depend on the degree of difference between the two groups and the proportion of population that are missed. As survey undercoverage is basically constant from month to month, only level estimates are likely to be affected. Monthly movements, for example in employment and the unemployment rate, will therefore be negligible.

The initial sample for the LFS is selected following each decennial Census. Undercoverage tends to increase over the life of the sample. The current sample was introduced in 1995. The average monthly slippage rate in 1995 was 7.5%. In 2000, the average monthly slippage rate had increased to just under 10%.

Slippage rates differ across different age and sex groups and in different parts of Canada. Slippage is highest among 20 to 29 year-olds (17% for 2000), as it is often difficult to find them at home and they can be missed when household members are reported (e.g. away at school). In contrast, for Canadians aged 55 years and over, the slippage rate was 5% in 2000.

### *Assessment of impact on quality*

Despite the best efforts of any survey, coverage problems will occur. To compensate for these problems, sample estimates are adjusted during the estimation process to census-based population control totals. This adjustment has the benefit that the overall quality of the estimates produced by the LFS is greatly improved. However, any difference in the under-covered population and the covered population can still lead to a bias in the estimates.

People in the target population who are missed by the survey sample tend to have an average unemployment rate that tends to be higher than that of those who are covered in the survey sample. For this reason, a great deal of effort is made to reduce the slippage rate.

## Proxy responses

### *Definition*

In order to reduce the overall number of interviews undertaken in the survey, the LFS allows any adult in the household to provide information for other household members. The proxy rate for the LFS is the proportion of all survey responses received that were provided by a proxy respondent.

### *Implications for data quality*

It is possible that proxy respondents may provide some different answers to survey questions than those that would have been provided by the persons for whom they are reporting. The approach is used however, because it is recognized that persons within the same household have a good knowledge of the main characteristics and activities that the survey seeks to measure. Specifically, information used to determine labour force activity i.e. whether a person is working or looking and available to work, is well known. Similarly, demographic characteristics are well known. However, some questions asked in the survey may not be as well known and answers given may be approximate. Items that may not be as well reported by proxy respondents are hours actually worked in the survey reference week, earnings and descriptions of occupation and industry.

Annual proxy interview rates for the LFS over the last 10 years have remained relatively unchanged, averaging 55%.

### *Assessment of impact on quality*

Proxy responses do not appear to substantially impact on key labour force data produced from the survey. However, some LFS data items may be affected such as hours and wages, detailed classifications of industry and occupation.

## **Recent survey changes and their impact**

### Population re-basing

To ensure that the LFS accurately reflects changes that occur in the composition and location of the Canadian population, the LFS responses are re-weighted according to the most current population information following each Census. LFS estimates were re-based to the 1996 Census population counts in January 2000. Detailed information about this change can be found in *Improvements in 2000 to the LFS*, Catalogue no. 71F0031XIE, a free research paper on the Statistics Canada website.

Re-basing of LFS population estimates requires that revisions be made to data for past years. For example, in January 1999 estimates were revised back to January 1976. Over that period, the LFS population was revised downward by 0.35%. However, the revisions were smaller early in the period and larger toward the end of the period (up to 1.35% after June 1996). Note also that population re-basing did not impact consistently on all population groups. Notably, youths were affected differently from adults.

Re-basing of LFS population estimates in the past have slightly modified the long-term trend of employment and unemployment. The changes, however, have not been significant enough to affect how these trends could be interpreted.

### Composite estimation

A change in the method of estimation used in the LFS was also introduced in January 2000. The impact of this change was to produce estimates that were slightly smoother over time<sup>2</sup> (i.e. smaller standard errors). Refer to the research paper *Improvements in 2000 to the LFS* (mentioned above) for a discussion on composite estimation.

### Centralized telephone interviewing

Households selected for the LFS remain in the survey for six consecutive months. The first LFS interview is conducted at the home of the respondent. In general, subsequent interviews (months 2 to 6) are conducted by telephone. Until May 2000, the same field interviewers who visit respondent homes to undertake the first interviews conducted telephone interviews. Over the period June to September 2000, Statistics Canada progressively transferred telephone interviews from the field interviewers to telephone interviewing units in Statistics Canada Regional Offices. First interviews continue to be conducted in person by Statistics Canada field interviewers.

This change has two consequences: 1) subsequent interviews are not conducted by the field interviewer who conducted the first interview in a given household and 2) subsequent household interviews can be conducted by different interviewers from one survey month to another. A quality study is currently underway to evaluate whether these changes in interviewing arrangements have an impact on LFS data. An increase in non-response rates was observed during the implementation phase.

### Employment Insurance Region boundary changes

In accordance with the Employment Insurance Act, every five years Human Resources Development Canada (HRDC) conducts a review of the regions defined for the Employment Insurance program. Unemployment rates for these regions are used to administer the Employment Insurance benefits in those regions. The latest changes to Employment Insurance Regions were introduced on July 9, 2000.

For more information on the changes to the Employment Insurance Regions, please go the HRDC website [www.hrdc-drhc.gc.ca/common/news/insur/00-46.shtml](http://www.hrdc-drhc.gc.ca/common/news/insur/00-46.shtml).

---

<sup>2</sup> For more information, see the publication *Survey Methodology*, June 2001, catalogue no. 12-001-XPB (Statistics Canada), vol. 27, no 1, page 65-69.

## **Notes on particular LFS items**

### Employment and employees

There are two major sources of monthly employment data from Statistics Canada – LFS and SEPH (Survey of Employment, Payrolls and Hours). The concept of employment differs between these two sources. The payroll survey measures jobs, rather than persons, and covers all those who are issued a T4 by their employer. While this mostly covers employees, some working owners of incorporated businesses may fall into this category if they pay themselves a salary. Those absent without pay are excluded from the SEPH employment count.

LFS employment refers to persons who are employees or self-employed, and who may or may not have more than one job. Also included in the count are those who are employed, but absent without pay.

### Wages

Wages of employees have been collected in the LFS since January 1997. However, proxy respondents do not always know the wages of their fellow householders and some respondents are reluctant to divulge their wages. As a result, this item has a higher non-response than other LFS data. For the 12 months ending November 2000, the average proportion of employees for whom wages data were not collected was 9%. Based on an analysis of the occupation of employees for whom wages data were not available, it is likely that the level of published wages of employees may be slightly understated. This impact is partly dealt with by imputing wages for employees with wage data missing, based on other employees in the survey with similar characteristics.

The LFS provides the possibility of studying wage distributions by worker and job characteristics. The data are complementary to those available from SEPH, although the concepts are not identical. One major difference is the fact that the LFS asks wage information only during the first interview and carries this information forward for the next five months that the person is in the survey, unless there is a change in job or employment status. This results in a slight lag in the LFS wage estimates, since wage increments that occur without a change in job are not captured. Another difference is that LFS wages exclude overtime, while this is included in the SEPH estimates.

### Hours

All employed people are asked to report their usual regular weekly hours of work and the hours they actually worked in the survey reference week. In both cases, respondents are prone to round the number of hours they report. While it is believed that aggregate total hours and overall average hours are unlikely to be affected, users should be cautious of the grouping of hours when using ranges.

LFS actual hours are conceptually suitable for productivity measures. The concept used in the payroll survey (SEPH) is different and refers to hours paid, whether worked or not.

### **Contact details for further information**

For further information about the quality of LFS data contact Client Services at 1-866-873-8788 (labour@statcan.ca).