

SAMPLING VARIABILITY

The Labour Force Survey, on which estimates on this CD are based, collects information from a sample of households. Somewhat different figures might have been obtained if a complete census had been taken using the same questionnaires, interviewers, supervisors, processing methods, etc. as those actually used in the Labour Force Survey.

The sampling error, or standard error, is a measure that quantifies how different the sample estimate might be from the Census value. It is based on the idea of selecting several samples, although in a survey only one sample is drawn and information is collected on units in that sample. Using the same sampling plan, if a large number of samples were to be drawn from the same population, then about 68% of the samples would produce a sample estimate that is within one standard error of the census value and in about 95% of the samples it will be within two standard errors of the census value.

Sampling variability can also be expressed relative to the estimate itself. The standard error as a percentage of the estimate is called the coefficient of variation (CV) or the relative standard error. Probability statements can also be made about CVs; for example, if the CV is 7% then in 68% of the samples the census value will lie within 7% or one CV and in 95% of the samples the census value will lie within 14% or two times the CV of the estimate. For estimates on this CD, the CV is used to give an indication of the accuracy associated with the estimates.

INDICATORS OF SAMPLING VARIABILITY

Small CV's are desirable because they indicate that the sampling variability is small relative to the estimate. The CV depends on the size of the estimates, the sample size the estimate is based on, the distribution of the sample, and the use of auxiliary information in the estimation procedure. The size of the estimates is important because the CV is the sampling error expressed as a percentage of the estimate. The smaller the estimate the larger the CV (all other things being equal). For example, when the unemployment rate is high the CV may be small. If the unemployment rate falls due to improved economic conditions then the corresponding CV will become larger. Typically, of similar estimates, the one with largest sample size will yield the smaller CV. This is because the sampling error is smaller.

Also, estimates referring to characteristics that are more clustered will have a higher CV. For example, persons employed in forestry, fishing, mining, oil and gas in Canada are more clustered geographically than employed women aged 55 to 64 years in Ontario. The latter will have a smaller sampling variability although the estimates are of approximately the same size.

Finally estimates referring to age and sex are usually more reliable than other similar estimates because the LFS sample is calibrated to post-censal population estimates of various age and sex groupings. Continuing the previous example, persons employed part-time in Alberta will have a larger sampling variability than employed men aged 35 to 44 years in British Columbia although the estimates are of similar size.

Variability of monthly estimates

To look up an approximate measure of the CV of an estimate of a monthly total, please consult Table A, which gives the size of the estimate as a function of the geography and the CV. The rows give the geographic area of the estimate while the columns indicate the resulting level of accuracy in terms of the CV, given the size of the estimate. To determine the CV for an estimate of size X in an area A, look across the row for area A, find the estimate that is less than or equal to X. Then the title of the column will give the approximate CV. For example, to determine the sampling error for an estimate of 30.2 thousand unemployed in Newfoundland and Labrador in September 2007, we find the closest but smaller estimate of 27.6 thousand giving a CV of 5%. Therefore, the estimate of 30,200 unemployed in Newfoundland and Labrador has a CV of roughly 5%.

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	1,104.5	329.3	145.4	85.1	48.8	28.5	19.1	13.9	10.7
Newfoundland and Labrador	275.2	72.1	27.6	15.1	8.9	4.9	3.2	2.2	1.7
Prince Edward Island	69.1	20.2	8.5	4.9	2.9	1.7	1.1	0.8	0.6
Nova Scotia	268.5	77.0	32.1	18.4	10.9	6.2	4.1	3.0	2.3
New Brunswick	237.0	66.5	27.0	15.3	9.1	5.1	3.4	2.4	1.9
Quebec	1,103.2	317.1	133.8	76.8	44.6	25.6	17.0	12.3	9.4
Ontario	1,152.9	329.0	138.7	79.4	45.7	26.2	17.3	12.5	9.5
Manitoba	212.4	63.8	28.6	16.8	9.5	5.6	3.8	2.8	2.1
Saskatchewan	181.9	54.2	23.9	14.0	8.0	4.7	3.1	2.3	1.8
Alberta	501.8	151.9	69.6	41.3	22.9	13.6	9.2	6.7	5.1
British Columbia	642.5	192.8	86.9	51.2	28.7	16.9	11.4	8.3	6.4

* Estimates are in thousands.

Table A is supplied as a rough guide to the sampling variability. The sampling variability is modeled so that, given an estimate, approximately 75% of the actual CVs will be less than or equal to the CVs derived from the table. There will, however, be 25% of the actual CVs that will be somewhat higher than the ones given by the table.

The CV values given in Table A are derived from models based on 2004, 2005, 2006 and most of 2007 LFS sample data. It is important to bear in mind that these values are approximations. For more accurate measures of variability, please contact Client Services at 1 866 873-8788 or e-mail us at labour@statcan.ca.

Variability of annual estimates

To look up an approximate measure of the CV of an estimate of an annual average, please consult Table B, which gives the size of the estimate as a function of the geography and the CV. The rows give the geographic level of the estimate while the columns indicate the resulting level of accuracy in terms of the CV, given the size of the estimate. To determine the CV for an estimate of size X in an area A, look across the row for area A, find the estimate that is less than or equal to X. Then the title of the column will give the approximate CV. For example, to determine the sampling error for an annual average estimate of 37.5 thousand unemployed in Newfoundland and Labrador, we find the closest but smaller estimate of 17.1 thousand giving a CV of 2.5%. Therefore, the estimate of 37,500 unemployed in Newfoundland and Labrador has a CV of roughly 2.5%.

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	421.9	120.5	53.6	31.1	16.6	9.6	6.4	4.6	3.5
Newfoundland and Labrador	58.2	17.1	8.2	4.9	2.4	1.5	1.0	0.7	0.5
Prince Edward Island	17.4	5.5	2.8	1.7	0.9	0.6	0.4	0.3	0.2
Nova Scotia	70.4	21.7	10.7	6.5	3.4	2.0	1.4	1.0	0.8
New Brunswick	57.7	17.4	8.5	5.1	2.6	1.6	1.1	0.8	0.6
Quebec	318.8	96.5	46.3	27.8	14.5	8.7	5.9	4.3	3.3
Ontario	321.1	96.3	46.3	27.8	14.2	8.6	5.8	4.2	3.2
Manitoba	74.3	21.8	10.4	6.2	3.1	1.9	1.3	0.9	0.7
Saskatchewan	60.4	17.8	8.5	5.1	2.6	1.5	1.0	0.7	0.6
Alberta	182.7	54.0	25.9	15.5	7.8	4.7	3.1	2.3	1.7
British Columbia	218.9	66.1	32.2	19.4	9.9	6.0	4.1	3.0	2.3

* Estimates are in thousands.

Table B is supplied as a rough guide to the sampling variability. The sampling variability is modeled so that, given an estimate, approximately 75% of the actual CVs will be less than or equal to the CVs derived from the table. There will, however, be 25% of the actual CVs that will be somewhat higher than the ones given by the table.

The CV values given in Table B are derived from a model based on 2003, 2004, 2005, 2006 and most of 2007 LFS sample data. It is important to bear in mind that these values are approximations. For more accurate measures of variability, please contact Client Services at 1 866 873-8788 or e-mail us at labour@statcan.ca.

Variability of rates

Estimates that are rates and percentages are subject to sampling variability that is related to the variability of the numerator and the denominator of the ratio. The various rates given in this CD are treated differently because some of the denominators are calibrated figures that have no sampling variability associated with them.

Unemployment rate

The unemployment rate is the ratio of X, the total number of unemployed in a group, to Y, which is the total number of participants in the labour force in the same group. Here the group may be a province or CMA and/or it may be an age-sex group. For example, in September 2007, there were approximately 30,200 unemployed persons in Newfoundland and Labrador and 255,000 participants in the labour force, giving an unemployment rate of 11.8%.

To determine the CV for the unemployment rate, the following formula can be used:

$$CV(X/Y) = \sqrt{[CV(X)]^2 - [CV(Y)]^2}$$

where, CV(X) is the CV for the total number of unemployed in a specific geographic or demographic subgroup and CV(Y) is the CV for the total number of participants in the labour force in the same subgroup. Continuing the example for Newfoundland and Labrador, the CV for the unemployment rate of 11.8% would be $\sqrt{(5.0\%)^2 - (2.5\%)^2} = 4.3\%$, where from Table A the CVs of monthly estimates of 30,200 and 255,000 in Newfoundland and Labrador are 5% and 2.5%, respectively.

Participation rate and employment rate

The participation rate represents the number of persons in the labour force expressed as a percentage of the total population size. The employment rate is the total number of employed divided by the total population size. For both the above rates, the numerator and the denominator represent the same geographic and demographic group.

For Canada, the provinces, CMAs and age-sex groups the LFS population estimates are not subject to sampling variability because they are calibrated to independent sources. Therefore, in the case of the participation rate and the employment rate of these geographic and demographic groups, the CV is equal to that of the contributing numerator.

Subgroups of Canada, the provinces and age-sex groups are called domains; for example, persons employed in agriculture in Manitoba are a domain. To determine the CV of rates in the case of domains, the variability of both the numerator and the denominator have to be taken into account because the denominator is no longer a controlled total and is subject to sampling variability. Therefore, for participation rates and employment rates of domains, the CV can be determined similar to the unemployment rate. The totals in the numerator and denominator for the relevant rate should reflect the same domain or subgroup.

Variability of estimate of change

The difference of estimates from two time periods gives an estimate of change that is also subject to sampling variability. An estimate of year-to-year or month-to-month change is based on two samples which may have some households in common. Hence, the CV of change depends on the CV of the estimates for both periods and the sample overlap, ρ , between the periods. The following formula can be used to approximate the CV of the estimate of change:

$$CV(Y_2 - Y_1) = \frac{\sqrt{1 - \rho} \sqrt{Y_1^2 CV(Y_1)^2 + Y_2^2 CV(Y_2)^2}}{(Y_2 - Y_1)}$$

where, Y_1 and Y_2 are the estimates for the two periods; and, the value of ρ is 0.5 for change between consecutive months and ρ is zero for changes over all other time periods. When comparing the annual averages of two years, the CV of the annual estimates (Table B) should be used. For month-to-month change, seasonally adjusted estimates should be used in conjunction with the CVs of the monthly estimates from Table A. Note that the above formula gives approximate estimates of the sampling variability associated to an estimate of change.

How to interpret the CV value

Household surveys within Statistics Canada generally use the following guidelines and reliability categories in interpreting CV values for data accuracy and in the dissemination of statistical information. These guidelines are provided for reference purposes only.

Category 1 - If the CV is $\leq 16.5\%$ - no release restrictions: data are of sufficient accuracy that no special warnings to users or other restrictions are required.

Category 2 - If the CV is $> 16.5\%$ and $\leq 33.3\%$ - release with caveats: data are potentially useful for some purposes but should be accompanied by a warning to users regarding their accuracy.

Category 3 - If the CV $> 33.3\%$ - not for release: data contain a level of error that makes them so potentially misleading that they should not be released in most circumstances. If users insist on inclusion of Category 3 data in a non-standard product, even after being advised of their accuracy, the data should be accompanied by a disclaimer. The user should acknowledge the warnings given and undertake not to disseminate, present or report the data, directly or indirectly, without this disclaimer.