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. This difference in the figures is called the sampling error of the estimates.

A measure of the sampling error is the standard error. This measurement 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 uncertainty associated with the estimates.

INDICATORS OF SAMPLING VARIABILITY

Small CVs are desirable because they indicate that the sampling variability is small relative to the estimate. The CV depends on the size of the estimate, the sample size that the estimate is based on, and the distribution of the sample; of two estimates, the one with the larger sample will likely have a smaller sampling error; and, of two estimates of the same size the one referring to a characteristic that is more clustered geographically will have a larger variability associated with it. For example, persons employed in forestry and logging in Canada are more clustered geographically than employed widowed women aged 45 years or over in Canada. The latter will have a smaller sampling variability although the estimates are of approximately the same size. In addition, estimates relating to age and sex are usually more reliable than other estimates of comparable sample size because, in the LFS, the sample is calibrated to independent sources. Continuing the previous example, employed widowed women aged 45 years or over in Canada will have a larger sampling variability than unemployed persons aged 55 and over in Canada 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 37.0 thousand unemployed in Newfoundland in September 2000, we find the closest but smaller estimate of 27.7 thousand giving a CV of 5%. Therefore, the estimate of 37,000 unemployed in Newfoundland has a CV of roughly 5%.

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

	Coefficient of variation								
Geographic area	**1%**	**2.5%**	**5%**	**7.5%**	**10%**	**15%**	**20%**	**25%**	**30%**
Canada	1,070.0	302.1	129.4	74.1	48.8	23.4	15.4	11.1	8.4
Newfoundland	243.8	66.3	27.7	15.6	10.1	4.8	3.1	2.2	1.7
Prince Edward Island	64.2	18.5	8.0	4.6	3.1	1.5	1.0	0.7	0.5
Nova Scotia	245.0	70.2	30.3	17.5	11.6	5.6	3.7	2.7	2.0
New Brunswick	188.4	54.0	23.4	13.5	8.9	4.3	2.9	2.1	1.6
Quebec	1,015.6	287.7	123.5	70.8	46.7	22.5	14.8	10.6	8.1
Ontario	1,001.2	281.1	120.0	68.5	45.0	21.6	14.2	10.1	7.7
Manitoba	254.4	73.4	31.9	18.4	12.2	5.9	3.9	2.8	2.2
Saskatchewan	238.4	65.6	27.6	15.6	10.2	4.8	3.2	2.2	1.7
Alberta	597.0	172.0	74.7	43.1	28.6	13.9	9.2	6.6	5.1
British Columbia	787.1	220.1	93.7	53.4	35.1	16.8	11.0	7.9	6.0

* 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 CVs will be less than or equal to the CVs derived from the table. There will, however, be 25% of the estimates that will be somewhat higher than given by the table.

The measures of variability given in Table A are averages for the 1997, 1998, 1999 and most of 2000 sample design and sample size. It is important to bear in mind that these CVs are approximations and that they reflect only the sampling variability. For more accurate measures of variability, please contact Labour Force Survey Sub-Division, Statistics Canada, Ottawa, K1A 0T6.

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 41.0 thousand unemployed in Newfoundland, we find the closest but smaller estimate of 29.8 thousand giving a CV of 2.5%. Therefore, the estimate of 41,000 unemployed in Newfoundland has a CV of roughly 2.5%.

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

	Coefficient of variation								
Geographic area	**1%**	**2.5%**	**5%**	**7.5%**	**10%**	**15%**	**20%**	**25%**	**30%**
Canada	461.7	125.1	52.1	29.3	19.0	8.9	5.8	4.1	3.1
Newfoundland	121.7	29.8	11.6	6.2	3.9	1.7	1.1	0.8	0.6
Prince Edward Island	28.0	7.7	3.3	1.9	1.2	0.6	0.4	0.3	0.2
Nova Scotia	108.4	30.2	12.8	7.3	4.8	2.3	1.5	1.1	0.8
New Brunswick	89.2	23.7	9.7	5.4	3.5	1.6	1.0	0.7	0.6
Quebec	450.4	122.8	51.4	29.0	18.9	8.9	5.8	4.1	3.1
Ontario	452.6	124.0	52.1	29.4	19.2	9.1	5.9	4.2	3.2
Manitoba	128.0	34.5	14.3	8.0	5.2	2.4	1.6	1.1	0.8
Saskatchewan	127.2	30.8	11.9	6.4	4.0	1.7	1.1	0.8	0.6
Alberta	283.6	76.0	31.4	17.6	11.4	5.3	3.4	2.4	1.8
British Columbia	353.9	94.8	39.2	21.9	14.2	6.6	4.3	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 CVs will be less than or equal to the CVs derived from the table. There will, however, be 25% of the estimates that will be somewhat higher than given by the table.

The measures of variability given in Table B are averages for the 1997, 1998 and most of 1999 sample design and sample size. It is important to bear in mind that these CVs are approximations and that they reflect only the sampling variability. For more accurate measures of variability, please contact Labour Force Survey Sub-Division, Statistics Canada, Ottawa, K1A 0T6.

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 2000, there were approximately 37,000 unemployed persons in Newfoundland and 251,000 participants in the labour force, giving an unemployment rate of 14.6%.

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, x is the total number of unemployed in a specific geographic or demographic subgroup and y is the total number of participants in the labour force in the same subgroup. Continuing the example for Newfoundland, the CV for the unemployment rate of 14.6% would be $\sqrt{5^2 - 1^2} = 4.9\%$, where from Table A the CVs of monthly estimates of 37,000 and 251,000 in Newfoundland are 5% and 1%, 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 population is not subject to sampling variability because the population is 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:

$$CV(y_2 - y_1) = \sqrt{1 - r}\ \frac{\sqrt{y_1^2 CV(y_1)^2 + y_2^2 CV(y_2)^2}}{(y_2 - y_1)}$$

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 estimate (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 rough estimates.