

Chapter 2 - Data Quality

Chapter 2 provides data quality notes on the data contained in the file. Chapter 2 is composed of two parts:

Information on the characteristics of the sample

The reader will find information on the sample design and the reliability of the estimates, namely:

- A. Target population and geographical limitations of the file
- B. Sample design
- C. Estimation
 - 1. Weighting
 - 2. Types of estimation
- D. Data reliability
 - 1. Acceptability of estimations
 - 2. Sampling variability
 - 3. Non-sampling variability

Information related to other factors affecting data reliability

Non-sampling errors can also have an impact on data quality. The user will find information on these factors, namely for the household maintainer and type of dwelling variables whose data must be interpreted with care.

SAMPLE DESIGN AND ESTIMATION

A. Target Population and Geographical Limitations of the File

The target population for the file includes all private households in occupied private dwellings in Canada outside Indian reserves. Thus, the following population subgroups are not included in the file:

- (a) collective households;
- (b) households outside Canada;
- (c) dwellings occupied by temporary residents and/or foreign residents;
- (d) households located on Indian reserves.

In order to meet confidentiality criteria, the geographical information on the file is limited. Thus, the census metropolitan area (CMA), broken down by rural/urban code, is the lowest level of geographical aggregation of the file. However, a supplementary criterion, minimum population by CMA, is also imposed. Hence, only CMAs with a private off-reserve household population exceeding 50,000 are identified. For this reason, certain CMAs are combined, those are:

Trois-Rivières and Sherbrooke
Sudbury and Thunder Bay

However, two exceptions are made to the minimum population rule: the province of Prince Edward Island and Yukon/Northwest Territories are identified separately.

For a complete list of the identified CMAs, the reader is referred to Chapter 1.

B. Sample Design

The household microdata sample was selected using a two-stage sampling method. The one-fifth (2B data) sample collected during the 1986 Census constitutes the first stage of sampling. In the second stage of sampling, a systematic sample of households was selected within pre-determined strata, with probability proportional to the household weight.

1. First Stage

In the 1986 Census, four out of five households were enumerated using a short questionnaire (2A). This questionnaire contained nine questions of a demographic and ethno-cultural/linguistic nature. One out of five households received a more detailed questionnaire (2B). In addition to the nine 2A questions, the 2B questionnaire contained 23 other questions covering a wide range of topics.

The first stage of sampling for the household microdata file therefore begins with the census one-fifth (2B) sample. The data collected by the census for this population subgroup are weighted. Thus, each household in the sample "represents", in addition to itself, four households which are not part of the sample. The average weighting factor or weight for each household is approximately five, although these weights are modified slightly by a statistical adjustment procedure. For further details concerning the census one-fifth sample, refer to the "Census Handbook", 1988, Catalogue No. 99-104E.

2. Second Stage

As indicated in Section A, the target population for the file was divided into subgroups or geographical strata and a sample of households was drawn from each of these strata. On average, 1.29 households were selected for every 100 households, or approximately 6.45 households for every 100 households in the unweighted 2B population. This sampling ratio varied according to the geographical area concerned in order to ensure, with a few exceptions, a minimum sample size of 2,500 households. Table 1 shows the geographical strata along with their inverse sampling ratios, the size of the target population and the size of the sample obtained.

Table 1: Geographic Stratification of the Sample of Households

Geographic area	Inverse sampling ratio	Size of the target population (weighted 2B)	Sample size
<u>CMA</u>			
1. Halifax	41	103,828	2,532
2. Chicoutimi - Jonquière	20	51,274	2,564
3. Montréal	100	1,115,357	11,154
4. Québec City	87	218,050	2,506
5. Trois-Rivières + Sherbrooke	38	95,992	2,526
6. Hull	28	70,274	2,510
7. Ottawa	92	232,002	2,522
8. Hamilton	80	201,327	2,516
9. Kitchener	44	110,154	2,504
10. London	51	129,402	2,537
11. Oshawa	27	68,010	2,519
12. St. Catharines-Niagara	49	124,576	2,543
13. Sudbury + Thunder Bay	38	95,089	2,503
14. Toronto	100	1,199,754	11,998
15. Windsor	36	91,613	2,545
16. Winnipeg	94	236,319	2,514
17. Regina	27	67,641	2,505
18. Saskatoon	29	73,920	2,549
19. Calgary	99	248,588	2,511
20. Edmonton	100	283,146	2,832
21. Vancouver	100	530,310	5,303
22. Victoria	41	104,654	2,553
<u>Residual areas</u>			
23. Newfoundland	63	159,081	2,525
24. New Brunswick	92	230,807	2,509
25. Prince Edward Island	20	40,594	2,029
26. Nova Scotia (less CMAs)	100	190,813	1,908
27. Quebec (less CMAs)	100	801,583	8,016
28. Ontario (less CMAs)	100	962,317	9,623
29. Prairies (less CMAs)	100	651,719	6,517
30. British Columbia (less CMAs)	100	439,675	4,396
31. Northwest Territories + Yukon	20	21,556	1,078
TOTAL		8,949,425	115,347

Thus, in Halifax, one in every 41 households, or 5 out of 41 households of the unweighted 2B population, were selected. In addition, to make the sample more representative, the population was sorted within each geographical stratum by the following variables:

(a) Tenure

- (i) owner
- (ii) tenant

(b) Type of household

- (i) one-family household
- (ii) multiple-family household
- (iii) non-family household

(c) Household income

- (i) less than \$24,000
- (ii) \$24,000 and over

For households located in CMAs, a supplementary sort is carried out by:

(d1) Area of residence

- (i) urban core
- (ii) urban fringe
- (iii) rural fringe

For households in residual areas, there is a final sort by:

(d2) Area of residence

- (i) population of 100,000 and over
- (ii) population between 30,000 and 99,999
- (iii) population between 10,000 and 29,999
- (iv) population between 2,500 and 9,999
- (v) population less than 2,500 (including rural areas)

The sample was then selected within each stratum as follows: a random number, called the "random start", was selected between 0 and the inverse sampling rate minus 1. The weight of the first household in the stratum as defined by the 2B sample was added to the random start. If the sum obtained was greater than or equal to the inverse sampling ratio, the household was selected; otherwise, the weight of the next household was added, and so on. This procedure is called "systematic sampling with probability proportional to size", with size in this case being the weight of the household. This method makes it possible to obtain a sample in which the distribution of household characteristics is the same as in the total population. In order to avoid any risk of identifying households due to the systematic character of sampling, the order of households is modified randomly within each geographic stratum.

C. Estimation

In this section, the concept of weighting is introduced as it applies to the household microdata file, and then a brief explanation is provided of the types of estimation which may be obtained using the microdata file.

1. Weighting

The household microdata file contains a record for each selected household in the sample. Each record contains a certain number of characteristics or variables described in Chapter 1. Each of those households represents, in addition to itself, several other households which are not part of the sample. Instead of repeating the records in the sample according to the number of households they represent, each record has been assigned a weighting factor or weight. This weight is equivalent to the inverse sampling ratio associated with the household's particular geographic area. These inverse ratios are listed in Table 1 and identified in the microdata file by the WEIGHT variable.

The weight therefore indicates the number of times a particular record must be reproduced to obtain population estimates. For example, if we wish to estimate the number of "semi-detached" dwellings in Canada, we can select the records from the microdata file relating to this category of dwelling, and then add up their weights.

N.B.: Users are advised not to use unweighted data from the microdata file. The file was obtained using a complex sample design, characterized by major differences in sampling ratios among geographical strata (Table 1). Consequently, certain regions are over-represented in the sample relative to their population, to ensure a minimum sample size of 2,500 households in each geographical stratum. This means that the unweighted sample is not representative of the population covered by the file.

2. Types of Estimation

Two types of variables are contained on the file: qualitative variables indicating whether or not a household possesses a given characteristic, and quantitative variables such as income.

(a) Estimation for qualitative variables

Suppose we wish to estimate the number of household maintainers of British ethnic origin residing in Ontario. This estimate is easily obtained by adding up the weights of the records of households possessing this combination of characteristics. This could be expressed as a proportion of all household maintainers in Ontario, the latter being obtained by adding up the weights of all households in the province.

(b) Estimation for quantitative variables

In the case of quantitative variables, estimates may be obtained based directly on the numerical values of these variables, such as averages, medians or totals. For example, the average employment income of female household maintainers in Quebec could be calculated using the following formula:

$$\begin{array}{l} \text{Estimate of the average employment} \\ \text{income of female household} \\ \text{maintainers in Quebec} \end{array} = \frac{X}{Y}$$

where the numerator (X) is a quantitative estimate of the total employment income of female household maintainers in Quebec. This estimate is obtained by multiplying the weights of each record (one per household) of female household maintainers in Quebec by their employment income, and by adding up these products for all the appropriate records. The denominator (Y) is obtained by adding up the weights of the records of female household maintainers in Quebec. It should be noted that the X and Y estimates are obtained independently, then divided one by the other.

(c) More complex analyses

The user may wish to apply certain data analysis techniques, such as linear or logistic regression, or variance analysis, to the household file. However, caution must be exercised in doing so. The microdata file is obtained using a complex sample design and each record is therefore assigned a weight. The user must take these weights into account in any analysis. Thus, the significance of any specific analysis of the data will depend on the extent to which it takes the sampling plan into account.

D. Data Reliability

Since the microdata file is based on a sample of households, we cannot expect perfect agreement between the census data (based on the whole census or on the census one-fifth sample) and estimates based on results obtained using the microdata records. These data will inevitably differ to a certain degree, because of the random character of sample selection; however, the degree of random variation can be calculated.

The methodology used for the production of general tables illustrating measurements of sampling error are presented prior to a discussion of a statistical test designed to evaluate the degree of agreement of the microdata sample with the 2B census data. Finally, there is a brief discussion of non-sampling errors.

1. Acceptability of the Estimations

A frequently used statistical test to verify the degree of agreement between estimates based on a sample and totals for the entire population is the chi-square test. While we are not presenting a mathematical description of this

test here, it concerns the distribution observed in the sample, for each variable in the file, compared to the distribution obtained from the census population (weighted 2B), within each geographic area included in the file. A critical threshold is set, in this instance 5%, and it is to be expected that, on average, 1 test out of 20 will yield significant results due to sampling variability. If this limit is exceeded, the microdata sample is a poor sample of the census population for a given variable.

The results agreed with the predictions, since less than 5% of the statistical tests calculated were significant. It may therefore be assumed that the sample obtained is an acceptable representation of the census population.

2. Sampling Variability

The difference between the estimates obtained from the household microdata sample and the census results is defined as the estimate's sampling error. However, even if we do not know the exact value of this discrepancy, it is possible to estimate a statistical measure called the standard error. This measure reflects the variability expected for estimates based on samples of similar size and design and can be estimated from the sample data.

Using the standard error, it is possible to define confidence intervals for the estimates, assuming that they are distributed normally around the true population value. Thus, there is an approximately 68% chance that the difference between an estimate based on a sample and the true population value will be less than one standard error, an approximately 95% chance that this difference will be less than two standard errors and an approximately 99% chance that it will be less than two and a half times the standard error.

It should be noted that these intervals take into account only the sampling variability of the estimate. Thus, errors not due to sampling variability, such as response and processing errors, which can introduce systematic biases in the data, are not accounted for by the confidence interval. Non-sampling errors are discussed in Section D3.

(a) Sampling variability for qualitative variables

A measure frequently used to demonstrate the degree of sampling variability of an estimate is the "coefficient of variation (CV)". This is simply the standard error expressed as a percentage of the estimate.

General tables of sampling variability are provided in Appendix B. It should be noted, however, that because of the large number of estimates which can be derived from the household microdata file, it is impossible to present exact CVs for all the possible areas of study. Approximate CVs expressed as percentages are presented for the national level and for each

CMA and province identified in the microdata file. In addition to being easy to use, these tables, although approximate, enable the user to decide whether a particular estimate may be released for general use or should be discarded.

The following table sets out the rules generally followed at Statistics Canada for deciding, on the basis of its CV, whether or not an estimate should be released.

Category	Coefficient of variation (%)	Alphabetic code	Recommendation
1. Unrestricted	0.0 to 0.5	A	Estimates may be included in a general release without restriction. Use of the alphabetic code is recommended. The letter A indicates that the estimate is very reliable; the letter B indicates that the estimate is reliable, but less so than category A, etc.
	0.6 to 1.0	B	
	1.1 to 2.5	C	
	2.6 to 5.0	D	
	5.1 to 10.0	E	
	10.1 to 16.5	F	
2. Restricted	16.6 to 25.0	G	The estimates are sufficiently reliable for specific purposes, but must be used with great caution. Any time they are used, it must be pointed out that their sampling variability is high.
	25.1 to 33.3	H	
3. Not to be released	33.4 and over	I	The estimates must not be released in any form or under any condition. They should be deleted from statistical tables.

The CVs presented in Tables of Approximated Sampling Variability (in Appendix B) were calculated using the simple random sample formula. However, since the household microdata sample was selected using a more complex sampling plan, each CV was multiplied by a factor called the sample design effect. This factor is defined as the ratio between the actual standard error of the estimate (according to the complex sample design) and the standard error for a simple random sample of the same size. The sample design effect has been calculated for various characteristics for each CMA and province and for Canada as a whole. The design effects used in Tables of Approximated Sampling Variability (in Appendix B) represent typical values selected from these calculated values. Preference was given to high factors rather than low ones, thus introducing a certain conservatism in the tables and generally yielding CVs greater than would have been obtained using more accurate techniques.

A few rules for using these tables are presented below, and should enable users to determine the approximate CVs for estimates of totals, percentages and ratios.

Rule 1. Estimates of Totals

In this case, the CV depends solely on the value of the estimated total. For example, if we estimate the number of "semi-detached" dwellings in Montréal at 532,000; using the table for Montréal, we refer to the row closest to 532,000 (the numerator of the percentage), in this case 500,000, and we move across until we reach the column containing an estimate (any column different from "*****"). For this example, the approximate CV relating to the estimated number of semi-detached dwellings in Montréal is around 1.2%. This estimate can therefore be released without restriction, although use of the alphabetic code "C" is recommended.

Rule 2. Estimation of Percentages

The CV of a percentage estimate depends on the value of the percentage and on the size of the population on which the percentage is based. For example, let us assume that we wish to obtain the CV of the estimate of the number of household maintainers between the ages of 20 and 24 who reside in Winnipeg and were born in Manitoba. The estimate of the number of household maintainers between the ages of 20 and 24 in Winnipeg is 17,578. Of this number, 12,690 were born in Manitoba. The numerator of the percentage in this case is therefore 12,690. The estimated percentage is 12,690/17,578, or approximately 70%. Using the table for Winnipeg, we refer to the row closest to 12,690, in this case 12,000, then to the column closest to 72%, in this case 70%. Looking at the intersection of the corresponding row and column on the table, we find that the approximate CV is 5.7%. The estimate can therefore be released without restriction, although use of the alphabetic code "E" is recommended.

Rule 3. Estimation of Ratios

When the numerator is a subset of the denominator, as in the preceding example, rule 2 must be used. When the numerator is not a subset of the denominator as for example the ratio of the number of female household maintainers to the number of male household maintainers, the following formula is used to calculate the CV:

$$\text{let the ratio } R = \frac{X}{Y}$$

$$CV(R) = (CV(X)^2 + CV(Y)^2)^{1/2}$$

where CV(X) and CV(Y) are obtained using rule 1 or 2. However, this formula will have the effect of overestimating the CV if there is a positive correlation between X and Y and of underestimating it if there is a negative correlation.

(b) Sampling variability for quantitative variables

As explained in Section C. 2., statistics for the quantitative variables can be derived based on their numerical value. Averages, medians and factors of correlation, for example, belong to this category.

Since the microdata file represents a sample of all the private households in occupied private dwellings off reserves in Canada, each record is assigned a weight. These weights must therefore be taken into consideration in all analyses. Although the majority of statistical software programs permit the use of weights, there may be some variation in definitions. Although estimates produced using weights are usually correct, the estimates of sampling variability may not be. To calculate the correct sampling variabilities, we must know all the details of the sampling plan, details which cannot be disclosed here for reasons of complexity and confidentiality.

However, we propose an alternative method, using random groups, for calculating the sampling variability of quantitative variables. Although this method requires additional work, it yields a good estimate of the desired standard error. The following steps should be followed:

- (i) Distribute the units (households) on the microdata file into k subgroups of approximately equal size ($k = 4$ would be sufficient), within each geographical stratum (Table 1). Since the households have already been placed in random order, a systematic distribution is sufficient, placing the first element and each k^{th} element which follows in the first subgroup and so on.
- (ii) Calculate the value of the desired statistics for each subgroup. This requires multiplying the weight by the value k, since each subgroup must represent the entire population. Let " z_i " equal the value of the statistic for the subgroup i ($i = 1, 2, \dots, k$).
- (iii) The standard error of the statistic is calculated using the following formula:

$$s = \frac{1}{k(k-1)} \sum_{i=1}^k (z_i - \bar{z})^2 \quad \text{where} \quad \bar{z} = \sum_{i=1}^k \frac{z_i}{k}$$

3. Non-sampling Variability

Sampling error is only one component of a survey's total error. Non-sampling errors may also contribute to this total error. Errors of this type are introduced, for example, when the respondent provides incorrect information or does not answer a certain question (response error), when a unit of the target population is overlooked or counted more than once (coverage error), or during data processing, for example, coding or data capture errors. Furthermore, in order to meet confidentiality criteria, some values must be suppressed. The measures of sampling variability discussed in the preceding

sections take into account only variability relative to census data. Thus, they do not reflect any inaccuracies introduced in these data (both census and sample) by non-sampling errors and suppressions.

If a specific estimate concerns only a small proportion of the population, the sampling error will be the primary component of the total error. However, the more closely the estimate approximates the total number in the population, the smaller the sampling error becomes, which is not necessarily the case for non-sampling errors. In fact, the more closely the estimate approximates the total of the corresponding population, the larger the non-sampling errors relative to sampling errors.

OTHER FACTORS AFFECTING DATA RELIABILITY

ADJUSTMENTS TO GEOSTATISTICAL AREAS

Users should be aware that census geostatistical areas are subject to change from one census to the next. Therefore, when using data from two or more censuses, the user must be aware of, and take into consideration, any changes to the geographic limits of the areas being compared. Users wishing to obtain additional information in this regard should refer to Chapter 3.

POPULATION COUNTS BASED ON USUAL RESIDENCE

The population counts shown here for a particular area represent the number of Canadians whose usual place of residence is in that area, regardless of where they happened to be on Census Day. Also included are any Canadians staying in a dwelling in that area on Census Day and having no usual place of residence elsewhere in Canada. In most areas, there is little difference between the number of usual residents and the number of people staying in the area on Census Day. For certain places, however, such as tourist or vacation areas, or those including large work camps, the number of people staying in the area at any particular time could significantly exceed the number of usual residents shown here.

IMMIGRANT POPULATION AND POPULATION BORN OUTSIDE CANADA

All persons born outside Canada are not necessarily immigrants to Canada. Individuals who have reported their place of birth outside Canada, but who are Canadian citizens by birth, are not considered immigrants to Canada. Consequently, they do not have a period of immigration or age at immigration when they take up permanent residence in Canada. These individuals will be included in the non-immigrant population. This approach was used in the 1981 Census. By contrast, in the 1971 Census, all persons born outside Canada were categorized as immigrants and required to respond to the question on period of immigration.

MOBILITY STATUS

The geographic areas reflect boundaries as of January 1, 1986, the geographic reference date for the 1986 Census of Canada.

The counts for total "migrants" (a migrant is anyone who, five years earlier, did not have his/her usual place of residence within the census subdivision (CSD) where he/she was enumerated) are additive across any geographic level - e.g., the migrant count at the Canada level is the sum of the migrants at the provincial level.

At the CSD level, users are advised to exercise caution in the use of data on migrants, particularly for suburban municipalities within large metropolitan areas. Counts for total migrants, including in- and out-migrants, could be distorted due to suspected types of mis-response such as: (a) respondents in metropolitan areas reporting the main city rather than the municipality they actually lived in five years earlier (e.g., reported Toronto instead of Scarborough); (b) respondents failing to indicate a move from a different CSD if they perceived that they were still in the same main city (e.g., moved from Toronto to Scarborough but indicated that they still lived in the same municipality); and (c) respondents reporting moves according to out-of-date boundaries.

The concept of "migrant" is defined at the CSD level. For geographic levels below the CSD, such as enumeration areas (EAs) and census tracts (CTs), please note that the distinction between the migrant and non-migrant population refers to the corresponding CSD of the EA or CT. For example, migrants of a CT are those persons who moved from a different CSD, while non-migrants are those who moved within the same CSD - they moved either between different CTs or within the same CT.

Names and boundaries of particular census subdivisions may undergo trivial or, in some cases, substantial modifications during the five-year intercensal period; therefore, comparisons of data for a specific subprovincial area between any two censuses will not be valid unless these changes, if any, are accounted for.

Details of intercensal boundary changes can be found in the **Standard Geographical Classification** (Cat. No. 12-573).

Boundaries and CSD components of CMAs and CAs will often undergo modifications during the intercensal period; therefore, comparisons of data for specific areas between any two censuses will not be valid unless these changes are accounted for. A publication is available which provides comparisons of 1986 CMAs and CAs, and their 1981 versions. **Census Metropolitan Areas and Census Agglomerations: A 1986 and 1981**

Comparison (Cat. No. 99-105E or F) lists census subdivisions that make up the 1986 version of each CMA and CA, and shows corresponding delineations for 1981.

NUMBER OF WEEKS WORKED

The data for the 40-48 and 49-52 weeks worked categories for 1985 must be interpreted with caution because some respondents tend to exclude their paid leave of absence due to vacation or for other reasons from their work weeks, when in fact such leave of absence should be included. As a result, the 49-52 week category may be understated.

LABOUR FORCE ACTIVITY

The census labour force activity concepts have not changed between 1981 and 1986. However, the processing of the data was modified causing some differences. In the 1986 Census, contrary to previous censuses, a question on school attendance was not asked. This question was used to edit the labour force activity variable, specifically unemployment. Consequently, the processing differences affect the unemployed population and are mostly concentrated among the 15-19-year age group. The table on the following page indicates the magnitude of the effect upon the data, at the Canada level.

COMPARABILITY AND QUALITY OF LANGUAGE DATA

Comparison between 1981 and 1986

Mother tongue and home language. The language questions were the same in the last two censuses, but the instructions to respondents were modified for mother tongue and home language. In 1981, respondents were asked to indicate only one mother tongue and only one home language; nevertheless, 597,980 persons (2.5% of the population) reported more than one mother tongue and 535,735 persons (2.2% of the population) reported more than one home language.

To better reflect the linguistic reality in Canada, these instructions were dropped from the 1986 Census. Under the new guidelines, individuals could report more than one mother tongue if they had learned them at the same time and had spoken one as frequently as the other when they were children. Similarly, respondents could indicate

more than one home language if they were now speaking them equally often at home.

The number of multiple responses given in the 1986 Census was significantly higher than in the 1981 Census. In 1986, 954,940 persons or 3.8% of the population reported a multiple response to the mother tongue question, while 1,159,675 or 4.6% of the population indicated more than one home language.

This increase was the result either of the changes made in the questionnaire, of changes in the way in which the population answers language questions or of an increase in the number of persons who had more than one mother tongue or spoke more than one language at home. A combination of these factors may also explain the increase in multiple responses.

When the 1981 data were processed, only one language was retained for publication, even in cases where the respondent reported more than one. In 1986, responses indicating more than one language were accepted.

In order to facilitate the determination of the trends between the two censuses, the 1986 Census results have been adjusted. In cases where more than one language was reported, the multiple responses were distributed among the component languages in the same proportions as in the 1981 Census. The results have been published in a special document entitled "Adjusted Language Data", April (1988). Also, data from the 1981 Census have been adjusted to show the multiple responses reported at that time. The data are presented in Table 4 of publications 93-102 (mother tongue) and 93-103 (home language). These adjustments to the mother tongue and home language figures make it easier to relate the 1986 data to the 1981 data, but do not make the results of the two censuses entirely comparable. Consequently, considerable care must be exercised in the interpretation of changes between 1981 and 1986.

The 1986 Classification of languages differs from that used in 1981, especially with regard to aboriginal languages. Appendix B of the **1986 Census Dictionary** (Catalogue No. 99-101E) provides a description of the changes.

Official language - Some respondents report speaking English or French or both at home, while on the other hand they indicate in the official language question, that they cannot carry on a conversation in these languages.

Labour Force Activity, 1981 Census of Canada

Canada	1981 Census (as published in 1981)	1981 Census (using 1986 processing)	% change
Labour force 15 years and over	12,054,150	12,081,280	0.23
Employed	11,167,915	11,167,915	no change
Unemployed	886,235	913,365	3.06
Not in the labour force	6,555,135	6,528,005	-0.41
Labour force 15-19 years	1,073,945	1,098,390	2.28
Employed	906,705	906,705	no change
Unemployed	167,240	191,680	14.61
Not in the labour force	1,229,630	1,205,190	-1.99
Labour force 20 years and over	10,980,205	10,982,890	0.02
Employed	10,261,210	10,261,210	no change
Unemployed	718,995	721,685	0.37
Not in the labour force	5,325,505	5,322,815	-0.05

In such cases, in the 1981 Census, the answer to the official language question was considered erroneous. Consequently, during data processing, this answer was changed to show that the person could speak the official language(s) they had reported to the home language question.

In the 1986 Census, not all of these responses were considered erroneous. If the respondent indicated being able to speak only one official language - either English or French - and this language matched the person's mother tongue, no correction was made during processing. Consequently, these response patterns appear as such in the 1986 tabulations.

For further information on language data, contact the Housing, Family and Social Statistics Division, Statistics Canada, Ottawa, Canada K1A 0T6.

COMPARABILITY OF DATA ON ETHNIC ORIGIN

Comparison between 1981 and 1986. The 1981 and 1986 ethnic origin data are not directly comparable.

The 1981 ethnic origin question: To which ethnic or cultural group did you or your ancestors belong on first coming to this continent?, was modified for the 1986 Census. The phrase "on first coming to this continent" was removed from the 1986 version as it was viewed as being inappropriate for persons of aboriginal origin. The 1986 question was: To which ethnic or cultural group(s) do you or did your ancestors belong?

In 1986, respondents were instructed to mark or specify as many groups as apply. This instruction

along with the addition of two more write-in spaces contributed significantly to an increase in multiple ethnic origin responses.

As well, the mark boxes in the question were ordered on the basis of 1981 incidence reporting of single ethnic origins. This changed the relative position of the mark boxes Chinese and Polish.

In light of the recommendations of a Parliamentary Commission on Visible Minorities in Canadian Society in the report Equality Now and the Abella Commission on Equality in Employment, the mark box Black was added to the 1986 ethnic origin question.

The mark boxes for aboriginal peoples were also changed. In 1986, status and non-status Indian categories which had been part of the 1981 ethnic origin question were replaced by North American Indian. It should be noted that persons of non-aboriginal cultural origin but status Indian under the Indian Act of Canada, for example, persons who obtained Indian status at marriage, could have been included in 1981 data for aboriginal peoples. These persons may not have identified their ethnic origin to be North American Indian in 1986 and thus would not be included in the 1986 count of aboriginal peoples. Also, in 1986, an undetermined number of persons of Métis origin could have indicated their ethnic origin as being the multiple response North American Indian and some other ethnic or cultural origin(s).

Single and Multiple Response

A **Single Response** occurs when the respondent provides only one origin. For example, for Canada, 709,585 gave Italian as their only ethnic origin.

A **Multiple Response** occurs when the respondent provides more than one origin. Some 297,325 Canadians gave a response which included Italian and one or more ethnic or cultural origin(s). For example, 31,495 provided the multiple response combination: Italian and French.

In the ethnic origin legend for this profile, the single origins are shown as unique groups. The multiple origins are shown as one group: multiple origins. In the case of the 31,495 Italian and French multiple response combination, it would be included in the multiple origins count (6,986,345 for Canada).

For further information regarding the data on ethnic origin, please contact the Housing, Family and Social Statistics Division, Statistics Canada, Ottawa, Ontario K1A 0T6, telephone (613) 951-2574.

HOUSEHOLD MAINTAINER

Users of data on household maintainers, such as sex of maintainer or mother tongue of maintainer, should be aware of certain limitations which can potentially have a large impact on the use and analysis of these data.

The household maintainer variable is a derived variable, a combination and manipulation of the responses that users have provided to the question on "person responsible for payments" and the question on "relationship to Person 1". The purpose of the household maintainer variable is to classify families within a household as primary (i.e. families of which the maintainer is a member) or secondary (i.e. families of which the maintainer is not a member). The variable is neither designed nor recommended for use as the equivalent of the previous "Household Head" variable for analytical purposes.

The variable itself was not treated, during processing, as a variable to be used in analysis. For example, if a respondent listed more than one name under the "person responsible for payments" question, only the first name inscribed was captured; the others were discarded. In addition, if a respondent indicated that no person in the household made shelter payments, the household was left without a primary family, but Person 1 was arbitrarily assigned to be the household maintainer. The basis for these processing decisions was the priority of categorizing families as primary or secondary, not providing a reference person for the household.

Users are cautioned, therefore, to refrain from making unjustified inferences based solely on direct comparisons of characteristics of household maintainers. For example, one should be careful when comparing female maintainers with male maintainers because an unknown number of each may have been entered as a second entry in the "person responsible for payments" question, and subsequently discarded. Similarly, a number of cases may have occurred in which a person outside the household has been replaced by "Person 1" in the derivation of the household maintainer, resulting in a person of a different sex ending up as the household maintainer.

Misinterpretation of results can also occur when using other maintainer characteristics, such as mother tongue or ethnic origin, to classify a household because these characteristics can be different for the other members of the household. It is suggested that analyses using these variables also take into account the characteristics of the maintainer's spouse.

STRUCTURAL TYPE

Users of structural type of dwelling data are cautioned about certain limitations of the data. Initial investigation of these data reveals the following limitations which may affect the quality of the data:

- (1) In the 1986 Census, there was a higher rate of non-response to the structural type of dwelling question than in 1981 (2.3% compared with 0.5%). The impact of this higher non-response on overall data quality should be small, except in a limited number of geographic areas where non-responses may have been concentrated. It should also be noted that the information on structural type was reported by the Census Representative in 1986, whereas, in 1981, it was reported by the household respondent.
- (2) Sharp declines between the 1981 and 1986 Censuses were found in every province for mobile homes and other movable dwellings. This is thought to be due to the misclassification of a number of mobile homes as other structural types, primarily single-detached dwellings. For larger geographic areas, this error is not expected to have a significant impact upon other dwelling categories because of the relatively small number of mobiles and movables.

- (3) Apartments in buildings of less than five storeys present some differences with 1981 Census counts, especially in Quebec and particularly in Montréal. Also, high overcounts in 1981 of duplexes, double houses and row houses resulted in sharp declines for these types in 1986 in certain provinces. An initial historical analysis indicated the 1986 counts were quite realistic.

INCOME DATA

The total income concept in the 1986 Census included, for the first time, federal child tax credits. As in 1981, the 1986 Census did not collect income information from institutional residents. Income statistics for families and households are for those in private households only.

Census income statistics are subject to sampling variability. Although such sampling variability may be quite small for large population groups, its effects cannot be ignored in the case of very small subgroups of population in an area or in a particular category. This is because, all other things being equal, the larger the sample size, the smaller is the error. For this reason, published income data for areas below the provincial level, where the non-institutional population was below 250, have been suppressed. However, where statistics are not suppressed but are still based on relatively small totals, the readers are strongly advised to exercise caution in the use and interpretation of these statistics.

Income Status

Income status refers to the position of economic families and unattached individuals in relation to

Statistics Canada's low income cut-offs. These cut-offs are determined separately for families of different sizes and living in areas of different degrees of urbanization. For the 1986 Census, they are based on the revised (1978) cut-offs which were initially estimated from the 1978 National Family Expenditure Survey and then updated to 1985 by the changes in the Consumer Price Index since 1978. The 1985 matrix of low income cut-offs is shown in Table 1.

For the purposes of low income statistics, economic families and unattached individuals in the Yukon and the Northwest Territories and on the Indian reserves are excluded. The low income cut-offs were based on certain expenditure-income patterns which were not available from survey data for the entire population.

The census and the Survey of Consumer Finances differ slightly when applying the "Size of Area" classification to derive incidence of low income. Census takes into account the density of population to designate an enumeration area as urban and the total population of contiguous urban EAs determines the size of area. The survey takes complete CMAs or CAs and classifies these into size of area by total population within the CMA/CA boundaries. The overall impact of this difference is negligible.

For further details on conceptual and coverage aspects, see the 1986 Census publication Family Income, Economic Families, Catalogue No. 93-918. (See also, Income Distributions by Size in Canada, 1985, Catalogue No. 13-207.)

Table 1: Low Income Cut-offs of Family Units, 1985

Size of family unit	Size of area of residence				
	500,000 and over	100,000 - 499,999	30,000 - 99,999	Small urban regions	Rural areas
1985 dollars					
1 person	10,233	9,719	9,117	8,429	7,568
2 persons	13,501	12,815	11,956	11,093	9,891
3 persons	18,061	17,115	15,996	14,880	13,244
4 persons	20,812	19,779	18,490	17,200	15,310
5 persons	24,252	22,963	21,415	19,952	17,803
6 persons	26,488	25,026	23,393	21,758	19,436
7 persons or more	29,155	27,606	25,801	23,994	21,415

Rounding and Adjustment of High Incomes and Losses

In planning this microdata file it was deemed essential to utilize procedures to guard against the possibility of associating a particular income with an identifiable individual, family or household. To accomplish this the income of individuals in households selected for this microdata file were subjected to a rounding and adjustment procedure as described below.

The individual incomes of the members of the households on this file were subjected to two separate operations. Initially, the amounts in wages, self-employment income (farm plus non-farm), investment income, retirement pensions, other money income and total income were rounded to the limits as specified in Table 2. This rounding procedure created certain inconsistencies between the sum of sources of income and total income. These inconsistencies were rectified by applying an adjustment procedure as specified in Table 3. After the individual records had been

rounded and adjusted, the income variables at the family and household level on this microdata file were derived, i.e. Employment Income, Government Transfer Payments, Investment Income, All Other Money Income and Total Income.

The number of records affected by this procedure and its impact on household income are summarized in the following tables.

Table 4 provides a distribution of individuals who had one or more sources of income and/or total income outside the limits imposed by confidentiality consideration.

Table 5 provides a summary of the changes, at the sample level, created by this rounding/adjustment procedure on the aggregate and average household income, by source.

Table 6 provides a distribution of the weighted aggregate household income in 1985, by province, from the main census data base and on the Public Use Microdata File.

Table 2: High and Low Income Limits

(1) The following income sources were subject to lower and upper limits for all individuals 15 years and over in the sample on the household microdata file:		
(a)	Wages and salaries	
(b)	Income from self-employment	
(c)	Investment income	
(d)	Retirement pensions	
(e)	Other money income	
(2) The limits were as follows:		
	<u>Lower limit</u>	<u>Upper limit</u>
(a)	Females in all areas and males in the Atlantic region	-\$30,000 \$100,000
(b)	Males in all other areas	-\$50,000 \$140,000
(3) Amounts beyond the limits in (2) above were rounded to the limits.		
(4) In cases where total income was beyond the limit, as a first step, it was rounded to the applicable limit.		
(5) To ensure consistency between the sum of sources and total income, individual records were then subjected to the adjustment procedure described in Table 3.		

Table 3: Adjustments Made to Remove Inconsistencies Introduced by Rounding

After rounding of applicable sources and/or total income as outlined in Table 2, individual sources and total income were subjected to the following adjustment routine in order to ensure consistency between the sum of sources and total income:

I Adjustment of Sources

- (1) If $A > 0$ and $B > 0$ and $A < B$ then $S_f = (S_i) (C/D)$
- (2) If $A < 0$ and $B < 0$ and $A > B$ then $S_{ef} = S_e + A - B$
- (3) No adjustment in all other cases

II Adjustment of Total Income

- (1) $Y = \text{Sum of Sources (after adjustments in I above and including transfer payments)}$
- (2) $Y = 1$ if sum of adjusted sources and transfer payments = 0

A	=	Total income after rounding
B	=	Sum of sources after rounding
C	=	A less transfer payments
D	=	B less transfer payments
S_i	=	Rounded wages, self-employment, investment, retirement and other money income
S_f	=	Final wages, self-employment, investment, retirement and other money income on PUMF
S_e	=	Rounded self-employment income
S_{ef}	=	Final self-employment income on PUMF
Y	=	Final total income on PUMF

Household and family incomes were derived by summing the incomes of individuals in the household/family after the adjustments described above.

Table 4: Distribution (Unweighted Sample) of Individuals with Incomes Outside Positive or Negative Limits (1) in 1985, Household and Housing Public Use Microdata File, 1986 Census

Source outside limits	Sample count	%
One source	436	74.7
Wages and salaries	182	31.2
Self-employment income	181	31.0
Investment income	69	11.8
Retirement pensions	2	.3
Other money income	2	.3
Two sources	7	1.2
Wages and investment	5	.9
Self-employment and investment	2	.3
Total income only	140	24.0
Total (2)	584	100.0

(1) See Table 2 for limits.

(2) There were 584 individuals in 562 households with incomes outside the specified limits.

Table 5: Number of Households, Their Original and Changed Aggregate and Average Incomes in 1985, By Source, Household and Housing Public Use Microdata File, 1986 Census

Number, aggregate income and average income	Employment income	Investment income	Retirement pensions and other money income	Government transfer payments	Total income
1. Number of records	115,347	115,347	115,347	115,347	115,347
2. -Aggregate income (\$'000)					
a. Original, total	3,111,602	251,486	146,169	435,601	3,944,859
b. Changed (562 Households)	-27,662	-9,456	-2,204	0	-39,323
c. Final, total	3,083,941	242,030	143,965	435,601	3,905,537
d. Percentage change - (b/a)	-0.9	-3.8	-1.5	0.0	-1.0
3. Average per household					
a. Original, total	26,976	2,180	1,267	3,776	34,200
b. Changed (562 records)	-49,220	-16,826	-3,922	0	-70,428
c. Overall change	-240	-82	-19	0	-343
d. Final, total	26,736	2,098	1,248	3,776	33,859
e. Percentage change (d/a)	-0.9	-3.8	-1.5	0.0	-1.0

Table 6: Distribution (Amount and Percentage) of Aggregate Income of Household in 1985, By Province, 1986 Census and Estimates From the Households and Housing Public Use Microdata File From the 1986 Census

Province	Dollars		%		Difference
	Census(1)	PUMF	Census(1)	PUMF	PUMF/Census
	(\$'000,000)		%		
Newfoundland	4,610.4	4,567.2	1.5	1.5	-0.9
Prince Edward Island	1,155.5	1,143.9	0.4	0.4	-1.0
Nova Scotia	9,068.3	8,857.1	2.9	2.9	-2.4
New Brunswick	6,657.6	6,672.3	2.2	2.2	0.2
Quebec	73,067.7	72,021.8	23.7	23.8	-1.5
Ontario	122,496.3	120,480.2	39.8	39.8	-1.7
Manitoba	11,952.8	11,560.0	3.9	3.8	-3.4
Saskatchewan	11,029.6	10,893.9	3.6	3.6	-1.2
Alberta	30,766.5	30,214.8	10.0	10.0	-1.8
British Columbia	36,414.8	35,703.4	11.8	11.8	-2.0
Yukon/Northwest Territories	845.3	833.6	0.3	0.3	-1.4
Canada	308,064.8	302,948.7	100.0	100.0	-1.7

(1) 1986 Census data base without random rounding.