

## **Chapter 2 - Data Quality**

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

### **Characteristics of the sample**

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

- Target population
- Sample design
- Estimation
- Data Reliability

### **Other factors affecting data reliability**

Non-sampling errors can also have an impact on data quality. The user will find information on these factors, such as the impact of rounding and adjusting high incomes and losses.

## **SAMPLE DESIGN AND ESTIMATION**

The file consists of two separate samples, a sample of families and a sample of non-family persons. Each of these samples were treated separately through all stages of sampling. Each of the files is discussed in turn below.

### **A. Family Portion**

#### **A.1. Target Population and Geographical Limitations**

The target population for the file includes all families living in private households except for families residing overseas, in collective households, families consisting of temporary and/or foreign residents and families located on partial or total refusal Indian Reserves.

In order to meet confidentiality criteria, the geographical information on the file is limited. Twenty-six geographic areas, that is, the ten provinces, Yukon and the Northwest Territories (considered as one region), and fifteen census metropolitan areas (CMAs) are identified.

#### **A.2. Sample Design**

The sample for the family microdata file was selected using a two-stage sampling method. The one-fifth (2B) sample collected during the 1986 Census constituted the first stage of sampling. The second stage consisted of a systematic sample of families selected within pre-determined strata, with probability proportional to the family weight.

##### **A.2.1 First Stage**

In the 1986 Census of Population, four out of five households were enumerated using a short questionnaire (2A). This questionnaire consisted of 9 questions of demographic and linguistic nature. The remaining households received a more detailed questionnaire (2B) that, in addition to the nine 2A questions, contained 23 other questions covering a wide range of topics.

The first stage of sampling for the family microdata file therefore begins with the census one-fifth (2B) sample. The data collected by the Census for the population of interest are weighted. Thus, each family in the sample "represents", in addition to itself, approximately four other families which are not part of the sample. The average weight for each family is approximately five. These weights were slightly modified by using statistical adjustment procedures, to obtain a more representative sample. Further details concerning the census one in five sample can be obtained from the "Census Handbook," catalogue number 99-104E, or the

User Information Bulletin no. 3 from the 1986 Census: Methodology for Producing Census Sample Estimates (March 1989).

**A.2.2 Second Stage**

In the second stage of sampling the target population was divided into subgroups, or geographical strata and a sample of families was selected from each of these strata.

The population was initially divided into 26 geographic areas, namely:

**Geographic Areas**

1. Halifax
2. Québec
3. Montréal
4. Ottawa-Hull
5. Toronto
6. Hamilton
7. St. Catharines
8. Kitchener
9. London
10. Windsor
11. Winnipeg
12. Calgary
13. Edmonton
14. Vancouver
15. Victoria
16. Newfoundland
17. Prince Edward Island

18. Nova Scotia (exc. Halifax)
19. New Brunswick
20. Quebec (exc. Quebec city, Montreal, Hull)
21. Ontario (exc. Ottawa, Toronto, Hamilton, St. Catharines, Kitchener, London, Windsor)
22. Manitoba (exc. Winnipeg)
23. Saskatchewan
24. Alberta (exc. Calgary, Edmonton)
25. British Columbia (exc. Vancouver, Victoria)
26. Yukon and Northwest Territories

These geographic areas were then further divided into strata formed by classes of Census family structure along with the Census family size. These variables are classified as follows:

Census Family Structure

Husband-wife families

with children present

1. wife under 35 years
2. wife 35-44 years
3. wife 45-54 years
4. wife 55 years and over

without children present

5. wife under 35 years
6. wife 35-44 years
7. wife 45-54 years
8. wife 55 years and over

Lone parent families

with male parent

9. parent of any age

with female parent

10. parent under 35 years
11. parent 35-44 years
12. parent 45-54 years
13. parent 55 years and over

Census Family Size

The size is defined by the number of persons in the census family as follows:

Husband/wife families with or without children present

1. two persons (without children)
2. three persons (with children present)
3. four persons (with children present)
4. five or more persons (with children present)

Lone parent families

1. two persons
2. three persons
3. four or more persons

The combination of 26 geographic regions along with other stratification variables created a total of 677 strata.

To make the sample more representative, sorting was performed within each stratum by the following variable:

Area of Residence

For families in selected CMAs

1. urban core
2. urban fringe
3. rural fringe

For families in other areas

4. population 100,000 or more
5. population 30,000 - 99,999
6. population 10,000 - 29,999
7. population 2,500 - 9,999
8. population under 2,500 including rural.

Finally, within each sorting class the families were put into a random order.

**A.2.3 Sampling**

Within each strata, the sample was selected systematically with a random start and with probability proportional to the first stage weight. Sample sizes within strata were determined such that each family selected represented one-hundred families in the target population. Thus, a one in one-hundred sampling rate was used for sample selection.

**B. Non-family Portion**

**B.1. Target Population and Geographical Limitations**

The target population for the file includes all non-family individuals living in private households except for the non-family individuals living overseas, in collective households, temporary and/or foreign resident and living in partial or total refusal Indian reserves.

For the sake of simplicity, non-family individuals will be referred to as individuals.

Like in family portion of the microdata file, the geographical information on the file is limited to a similar geography, i.e., a total of twenty-six geographic areas.

## B.2. Sample Design

The sample for file was selected using a two-stage sampling method. The one-fifth (2B) sample collected during the 1986 Census constituted the first stage of sampling. The second stage consisted of a systematic sample of individuals selected within pre-determined strata of the 2B sample.

### B.2.1 First Stage

The first stage of sampling for the non-family portion of the file is similar to the first stage of sampling for the family portion of the file. Refer to section A.2.1 for details.

### B.2.2 Second Stage

In the second stage of sampling the target population was divided into subgroups, or geographical strata and a sample of individuals was selected from each of these strata. These geographic areas are the same as identified for the family portion of the file in section A.2.2.

Six groups were then identified in each geographic region using the following classification of age and labour force activity.

#### Age/Labour Force

1. individual is less than 15 years of age
2. individual's age is between 15 and 44 years and in the labour force
3. individual's age is between 15 and 44 years and not in the labour force
4. individual's age is between 45 and 64 years and in the labour force
5. individual's age is between 45 and 64 years and not in the labour force
6. individual is 65 years of age or older

The 26 geographic areas along with other stratification variables created a total of 156 strata.

To make the sample more representative, sorting was performed within each stratum. The first sort was by sex variable.

Within each sex, the records were further sorted by the following classification of the area of residence:

Area of Residence

For individuals in selected CMAs

1. urban core,
2. urban fringe,
3. rural fringe,

For individuals in other areas

4. urban area with 30,000 people or more,
5. urban area with less than 30,000 people, and
6. rural

For individuals in P.E.I., Yukon and N.W.T.

7. no further classes

Finally, within each of these groups the individuals were put into a random order.

**B.2.3** Sampling

Within each strata, the sample was systematically selected with a random start and probability proportional to the first stage weight. Sample sizes within stratum were determined such that each individual selected represented one hundred individuals in the target population. Thus, a one in one-hundred sampling rate was used for sample selection.

Although both family and non-family portions of the microdata file were treated separately through all stages of sampling, the procedures for estimation are identical. The estimation procedures therefore for both portions of the microdata file will be given together. It should, however, be remembered that a Family or a Non-family individual will be referred to as a "unit" unless otherwise specified.



## C. Estimation

There are two types of variables on the microdata file, numeric (quantitative) variables, e.g., income variable and coded (qualitative) variables, e.g., mother tongue variable. Typical estimators for the two types of variables are given below.

### C.1. Coded (Qualitative) Variables

#### C.1.1 Estimates of Total

Estimates of total can be obtained by selecting the "units" possessing the characteristic of interest (e.g., English as the mother tongue), counting them, and multiplying the result by 100.

#### C.1.2 Estimates of Ratios and Percentages

Ratio estimates can be tabulated simply by counting the number of "units" in the numerator, the number in the denominator, and dividing. Percentage estimate can be obtained simply by multiplying the ratio estimate by 100.

### C.2. Numeric (Quantitative) Variables

#### C.2.1 Estimates of Total

Estimates can be obtained by selecting the "units" having the characteristic of interest (e.g., pension income), adding up their values, and multiplying the resulting sum by 100.

#### C.2.2 Estimates of Average

Average estimates involve adding up the values of the characteristic of interest and dividing the resulting sum by the number of "units" possessing the characteristic of interest. For example, the average pension income of female lone parents in Ontario can be tabulated as follows:

$$\begin{array}{l} \text{Estimate of the average pension income} \\ \text{of female lone-parent} \\ \text{in Ontario} \end{array} = \frac{X}{Y}$$

where the numerator X is a numeric estimate of the total pension income of female lone-parents in Ontario and the denominator Y is the total of the number of female lone parents in Ontario who receive pension income.

### C.2.3 Estimates of Ratios and Percentages

Ratios are calculated by obtaining the totals for numerator and denominator and then dividing. Percentage estimate can be obtained by multiplying the ratio estimate by 100. Do NOT calculate the ratio first for each "unit" of interest and then average the resulting ratio.

### C.3. More complex analysis

The microdata file is obtained using a complex sample design where each "unit" represents one-hundred "units". Complex data analysis techniques may be applied to the microdata file. However, it is suggested that sample design must be taken into account to obtain reliable results. Users wishing further information should contact the Chief, Census Data Quality Section, Social Surveys Methods at Statistics Canada.

## D. Data Reliability

Since the microdata file is a sample of "units" enumerated in the Census, one cannot expect the exact agreement between the estimates produced from the microdata file and values that would have been obtained if the information had been collected on a 100% basis. Differences between the microdata file estimates and the "true" values are subject to both sampling and non-sampling errors. However, differences between the microdata file estimates and published estimates should be due only to sampling error. A brief description of each type of error will be discussed in turn below.

### D.1. Sampling Errors

Sampling error is the error attributed to studying a fraction of the population. Each unit selected in the microdata file represents 99 other units who in fact might have different characteristics than the units selected. Such differences are generally presented as sampling variance. Sampling variance can be tabulated, the procedures are outlined below and tabulated in Appendix B.

### D.2. Estimation of Sampling Variance

Sampling variability is frequently measured by the "coefficient of variation (CV)", which is simply the standard error expressed as a percentage of the estimate. In other words it expresses the square root of sampling variance as a percentage of the estimate of interest.

General tables of sampling variability for coded variables are provided in Appendix B. It should however be kept in mind, that due to large number of estimates that can be derived from the microdata file, it is difficult to present accurate CVs for all the possible areas of study. Approximate CVs expressed as

percentages are presented for the geographical areas identified in the microdata file. These tables are easy to use and provide an approximate estimate, thus enabling the user to decide whether a particular estimate may be released for general use or not.

The following guidelines have been established at Statistics Canada relating the amount of sampling variance and whether an estimate should be released.

Category	Coefficient of Variation (%)	Alphabetic Code	Recommendation
Unrestricted	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	
Restricted	16.6 to 25.0	G	The estimates are sufficiently reliable for specific purposes, but must be used with great caution. Anytime they are used, it must be pointed out that their sampling variance is high.
	25.1 to 33.3	H	
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.

### D.3. Rules for Coded (Qualitative) Variables

#### D.3.1 Estimates of Total

The CV depends solely on the value of the estimated total.

1. Locate the appropriate table in Appendix B.
2. Look up the number closest to the estimate of interest in the "Numerator of Percentage" column.
3. The estimated CV is the first number to the right of the "Numerator of Percentage" column different from asterisks (\*).

### **D.3.2 Estimates of Percentages**

The CV of a percentage estimate depends on both the value of the percentage and on the size of the population on which the percentage is based.

1. Locate the appropriate table in Appendix B.
2. Count the number of units possessing the characteristic of interest, multiply the number by 100 and look up the number closest to the estimate in the "Numerator of Percentage" column.
3. Select the column closest to the estimated percentage.
4. The estimated CV is found at the intersection of the row identified in step 2 and the column selected in step 3.

### **D.3.3 Estimates of Ratios**

1. If the numerator is a subset of the denominator, the instructions are identical to those used for percentages (the ratio must be converted to a percent to use the tables).
2. If the numerator is not a subset of the denominator, then the CVs for the numerator and denominator must be obtained using the "estimates of total" instructions above. Each of the CVs is then squared. Then the two squared CVs are added together and the square root taken of the resulting sum. In other words let  $R = X / Y$ , then according to above

$$CV(R) = (CV(X)^2 + CV(Y)^2)^{.5}$$

where  $^2$  is used for square, thus  $CV(X)^2$  means  $CV(X)$  squared and  $^{.5}$  means the square root of the expression within the brackets.

It should be noted that this method is only approximate. The CVs will be overestimated if there is a positive correlation between the numerator and denominator and will be underestimated otherwise.

### **D.4. Numeric (Quantitative) Variables**

The formula given below for calculating approximate sampling variability are themselves approximations. Please refer to any text on sampling theory for precise formulae.

**D.4.1 Estimates of Total and Average**

1. Calculate the sum of squares by taking each record's value, square it and then add all the observations together.
2. Calculate the squared sum by adding up all the values and then square this sum.
3. Divide the sum of squares (step 1) by the squared sum (step 2).
4. Subtract  $1/n$ , where  $n$  is the number of observations, if  $n$  is small ( $< 50$ ).
5. Take the square root of the result of step 4 and multiply by 100.

**D.4.2 Estimates of Ratio and Percentages**

1. Perform steps 1 to 3 from "Estimates of Total and Average" above for the numerator.
2. Perform steps 1 to 3 from "Estimates of Total and Average" above for the denominator.
3. Take each record's value for the numerator, multiply it by the denominator, and add up the resulting products across all records of interest.
4. Add up all the values for the numerator, add up all the values for the denominator, and multiply the two sums together.
5. Multiply the result from step 3 by two and then divide by the result of step 4.
6. Add the results from steps 1 and 2 together and then subtract the result of step 5.
7. Take the square root of the result of step 6 and multiply by 100.

**D.5. Non-sampling error**

Sampling error is only one of the components of the total error in a survey. Further contribution may come from another source called non-sampling errors. Non-sampling errors are introduced, for instance, during imputation for non-reporting or obvious errors in reporting (response error), when individuals are not enumerated or are counted twice (coverage error), or during coding or data capture (processing error), etc. Furthermore, in order to meet confidentiality criteria, some values must be suppressed. The measures of sampling variability

discussed above take into account only variability relative to census data. Thus they do not reflect any potential inaccuracy in a census hence in the sample by non-sampling errors and suppressions.

For estimates of totals representing relatively small proportions of the population, the major component of the total error would be due to sampling error. As the estimated totals approach closely to the true population size, the sampling error decreases. This may not necessarily be true of the non-sampling errors, in fact, the more closely the estimates approach the true population size, the larger are 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 over-counts 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 1986 Census collected income information from all individuals, 15 years and over, in private households and non-institutional residents of collective households. Income statistics for families and non-family persons 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 less than 250, have been suppressed. The users of this microdata file are strongly advised to exercise caution in the interpretation of statistics based on relatively small totals.

### Income Status

Income status refers to the position of the 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<sup>1</sup>. 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. Table 1 shows the 1985 matrix of low income cut-offs.

The concept of an economic family is broader than that of a census family in that an economic family consists of all persons related by blood, marriage or adoption living together. Unattached individuals are persons either living alone or living in a household where they are not related to another person. Where an economic family consists of more than a census family, each of the units making up the economic family carries the income status of the economic family. Low income statistics calculated from this file relate to census families and non-family persons. It should be noted that these statistics will differ from those normally published for economic families and unattached individuals.

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.

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

<sup>1</sup> The census and the Survey of Consumer Finances, from which low income statistics are published annually, 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.

**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. The following rounding and adjustment procedure was adopted.

The individual incomes of non-family persons and all persons in families 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. Government transfer payments were not subject to these rounding and adjustment procedures.

After the individual records had been rounded and adjusted, the income variables at the family 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 the income of families and non-family persons is summarized in the following tables.

Table 4 provides distributions of persons in census families and non-family persons who had one or more sources of income and/or total income outside the limits.

Table 5 provides, at the sample level, a summary of the changes in the aggregate and average incomes, by source, of census families and non-family persons as a result of the rounding and adjustment procedures.

Table 6 provides distributions of the weighted aggregate income of census families and non-family persons in 1985, by province, from the Census and the Public Use Microdata File.

Table 7 provides comparable 1985 income size distributions of census families and non-family persons from the Census and the Public Use Microdata File.

**Table 2: Limits Used to Round High and Low Income Records on PUMF (Family), 1986 Census**

(1) In addition to the total income, the following income sources were subject to lower and upper limits for all individuals 15 years and over in the sample on the Family 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>Negative</u>	<u>Positive</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 applicable limit.		
(4) 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 of High Income Records, PUMF (Family), 1986 Census**

---

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

**III. Derivation of Family Income**

Family incomes were derived by summing the incomes of individuals in the family after the adjustments described above.

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

---

**Table 4: Distribution of Persons in Census Families and Non-Family Persons (unweighted sample) with Incomes Outside Positive and Negative Limits(1) in 1985, PUMF (Family), 1986 Census**

Source outside limits	Persons in Census Families(2)		Non-Family Persons	
	Number	Percent	Number	Percent
One source	321	73.5	53	73.6
Wages and salaries	146	33.4	21	29.2
Self-employment income	135	30.9	9	12.5
Investment income	39	8.9	23	31.9
Retirement Income	1	0.2	0	0.0
Two sources	16	3.7	0	0.0
Wages and self-employment	3	0.7	0	0.0
Wages and investment	7	1.6	0	0.0
Self-employment and investment	6	1.4	0	0.0
Three sources	1	0.2	0	0.0
Wages, Investment and Retirement	1	0.2	0	0.0
Total income only	99	22.7	19	26.4
TOTAL	437	100.0	72	100.0

(1) See Table 2 for limits.

(2) There were 437 individuals in 423 families with incomes outside the specified limits.

**Table 5: Number of Census Families and Non-Family Persons, 15 years and over, Their Original and Changed Aggregate and Average Incomes, by Source and Composition of Income in 1985, (unweighted sample), PUMF (Family), 1986 Census**

Number, aggregate income, average income and composition of income	Wages and salaries	Self-employment income	Investment income	Retirement pensions and Other money income	Government transfer payments	Total income
<b>CENSUS FAMILIES</b>						
A. - Number of records	67,326	67,326	67,326	67,326	67,326	67,326
B. Aggregate income (\$'000)						
a. Original	1,908,508	155,990	152,156	83,201	242,996	2,542,851
b. Change (423 Families)	-19,875	-5,325	-6,223	797	0	-32,221
c. Final	1,888,633	150,665	145,933	82,405	242,996	2,510,630
d. Percent change - [(b/a)*100]	-1.0	-3.4	-4.1	-1.0	0.0	-1.3
C. Average per family						
a. Original	28,347	2,317	2,260	1,236	3,609	37,769
b. Change (423 Families)	-46,987	-12,591	-14,711	-1,884	0	-76,173
c. Overall change	-295	-79	-92	-12	0	-479
d. Final	28,052	2,238	2,168	1,224	3,609	37,291
D. Composition of income						
a. Original	75.05	6.13	5.98	3.27	9.56	100.00
b. Final	75.23	6.00	5.81	3.28	9.68	100.00
<b>NON-FAMILY PERSONS</b>						
A. Number of records	34,719	34,719	34,719	34,719	34,719	34,719
B. Aggregate income (\$'000)						
a. Original	342,170	19,526	52,842	27,577	98,947	541,062
b. Change (72 Non-Family Persons)	-1,673	-499	-1,550	-98	0	-3,820
c. Final	340,497	19,027	51,292	27,479	98,947	537,242
d. Percent change - [(b/a)*100]	-0.5	-2.6	-2.9	-0.4	0.0	-0.7
C. Average						
a. Original	9,855	562	1,522	794	2,850	15,584
b. Change (72 Non-Family Persons)	-23,243	-6,921	-21,531	-1,365	0	-53,059
c. Overall change	-48	-14	-45	-3	0	-110
d. Final	9,807	548	1,477	791	2,850	15,474
D. Composition of income						
a. Original	63.24	3.61	9.77	5.10	18.29	100.00
b. Final	63.38	3.54	9.55	5.11	18.42	100.00

**Table 6: Distribution of Aggregate Income of Census Families and Non-Family Persons in 1985, by Province, Census and PUMF (Family) Estimates, 1986 Census**

Province	Dollars		Percent		Difference PUMF/Census
	Census(1)	PUMF	Census	PUMF	
	(\$'000,000)		percent		
CENSUS FAMILIES					
Newfoundland	4,104.6	3,986.4	1.6	1.6	-2.9
Prince Edward Island	976.4	956.7	0.4	0.4	-2.0
Nova Scotia	7,591.7	7,625.4	3.0	3.0	0.4
New Brunswick	5,705.7	5,777.6	2.2	2.3	1.3
Quebec	60,560.9	60,467.9	23.8	24.1	-0.2
Ontario	101,944.8	99,386.1	40.0	39.6	-2.5
Manitoba	9,803.7	9,626.7	3.8	3.8	-1.8
Saskatchewan	9,031.9	9,073.9	3.5	3.6	0.5
Alberta	25,104.2	24,512.0	9.9	9.8	-2.4
British Columbia	29,213.7	28,963.9	11.5	11.5	-0.9
Yukon/Northwest Territories	682.6	686.4	0.3	0.3	0.6
CANADA	254,720.2	251,063.0	100.0	100.0	-1.4
NON-FAMILY PERSONS					
Newfoundland	518.4	508.4	1.0	0.9	-1.9
Prince Edward Island	179.6	184.5	0.3	0.3	2.7
Nova Scotia	1,477.7	1,466.5	2.7	2.7	-0.8
New Brunswick	964.3	974.8	1.8	1.8	1.1
Quebec	12,539.4	12,529.6	23.3	23.3	-0.1
Ontario	20,769.9	20,811.1	38.6	38.7	0.2
Manitoba	2,160.1	2,184.0	4.0	4.1	1.1
Saskatchewan	2,012.1	2,097.8	3.7	3.9	4.3
Alberta	5,728.9	5,691.0	10.6	10.6	-0.7
British Columbia	7,282.2	7,116.6	13.5	13.2	-2.3
Yukon/Northwest Territories	164.1	160.1	0.3	0.3	-2.4
CANADA	53,796.5	53,724.2	100.0	100.0	-0.1

(1) 1986 Census of Canada, Family Income: Census Families, Catalogue No. 93-117



**Table 7: Percentage Distribution of Census Families and Non-Family Persons, by 1985 Income Size Groups, Canada, Census and PUMF (Family), 1986 Census**

Income size group	Census(1)	PUMF
	percent	
<b>CENSUS FAMILIES</b>		
Under \$5,000	3.7	3.6
\$5,000-\$9,999	4.7	4.8
\$10,000-\$11,999	2.5	2.5
\$12,000-\$14,999	5.6	5.6
\$15,000-\$19,999	9.1	9.1
\$20,000-\$24,999	8.7	9.0
\$25,000-\$29,999	9.0	9.0
\$30,000-\$34,999	9.5	9.6
\$35,000-\$39,999	9.8	8.7
\$40,000-\$44,999	8.0	8.1
\$45,000-\$49,999	6.5	6.5
\$50,000-\$59,999	9.5	9.6
\$60,000 and over	14.2	14.1
<b>TOTAL</b>	<b>100.0</b>	<b>100.0</b>
Average income	\$37,827	37,291
Median income(2)	\$33,489	33,390
<b>NON-FAMILY PERSONS</b>		
Under \$2,000	9.2	9.1
\$2,000-\$4,999	7.6	7.7
\$5,000-\$6,999	7.1	7.1
\$7,000-\$8,999	16.8	16.7
\$9,000-\$11,999	12.4	12.5
\$12,000-\$14,999	7.8	7.7
\$15,000-\$19,999	11.0	10.8
\$20,000-\$24,999	8.7	8.8
\$25,000-\$29,999	6.4	6.5
\$30,000-\$34,999	4.7	4.7
\$35,000-\$39,999	2.9	2.8
\$40,000 and over	5.3	5.5
<b>TOTAL</b>	<b>100.0</b>	<b>100.0</b>
Average income	\$15,495	15,474
Median income(2)	\$11,243	11,236

(1) 1986 Census of Canada, Family Income: Census Families, Catalogue No. 93-117.

(2) Median Income calculated from the distribution in this table.