

Microdata User Guide
RESIDENTIAL TELEPHONE SERVICE SURVEY
December 2010



Statistics
Canada

Statistique
Canada

Canada

Table of Contents

1.0	Introduction	5
2.0	Background	7
3.0	Objectives	9
4.0	Concepts and Definitions.....	11
5.0	Survey Methodology.....	13
5.1	Modifications to the Labour Force Survey Design for the Residential Telephone Service Survey	13
5.2	Sample Size by Province for the Residential Telephone Service Survey	13
6.0	Data Collection	15
6.1	Interviewing for the Labour Force Survey	15
6.2	Supervision and Quality Control	15
6.3	Non-response to the Labour Force Survey.....	15
6.4	Data Collection Modifications for the Residential Telephone Service Survey	16
6.5	Non-response to the Residential Telephone Service Survey	16
7.0	Data Processing	17
7.1	Data Capture	17
7.2	Editing	17
7.3	Coding of Open-ended Questions	17
7.4	Creation of Derived Variables	17
7.5	Weighting	18
7.6	Suppression of Confidential Information	18
8.0	Data Quality	19
8.1	Response Rates.....	19
8.2	Survey Errors	19
8.2.1	The Frame.....	20
8.2.2	Data Collection.....	20
8.2.3	Data Processing.....	20
8.2.4	Non-response.....	21
8.2.5	Measurement of Sampling Error	21
9.0	Guidelines for Tabulation, Analysis and Release.....	23
9.1	Rounding Guidelines.....	23
9.2	Sample Weighting Guidelines for Tabulation.....	23
9.3	Definitions of Types of Estimates: Categorical and Quantitative	24
9.3.1	Categorical Estimates	24
9.3.2	Quantitative Estimates	24
9.3.3	Tabulation of Categorical Estimates	24
9.3.4	Tabulation of Quantitative Estimates	25
9.4	Guidelines for Statistical Analysis	25
9.5	Coefficient of Variation Release Guidelines	26
9.6	Release Cut-off's for the Residential Telephone Service Survey	28

10.0	Approximate Sampling Variability Tables	29
10.1	How to Use the Coefficient of Variation Tables for Categorical Estimates.....	30
10.1.1	Examples of Using the Coefficient of Variation Tables for Categorical Estimates	31
10.2	How to Use the Coefficient of Variation Tables to Obtain Confidence Limits.....	36
10.2.1	Example of Using the Coefficient of Variation Tables to Obtain Confidence Limits.....	37
10.3	How to Use the Coefficient of Variation Tables to Do a T-test	37
10.3.1	Example of Using the Coefficient of Variation Tables to Do a T-test.....	38
10.4	Coefficients of Variation for Quantitative Estimates.....	38
10.5	Coefficient of Variation Tables	38
10.6	Variance Estimation for Master Files	39
11.0	Weighting	41
11.1	Weighting Procedures for the Labour Force Survey.....	41
11.2	Weighting Procedures for the Residential Telephone Service Survey	42

1.0 Introduction

The latest Residential Telephone Service Survey (RTSS) was conducted by Statistics Canada in December 2010. This manual has been produced to facilitate the manipulation of the microdata file of the survey results.

Any questions about the data set or its use should be directed to:

Statistics Canada

Client Services

Special Surveys Division

Telephone: 613-951-3321 or call toll-free 1 800 461-9050

Fax: 613-951-4527

E-mail: ssd@statcan.ca

2.0 Background

The Residential Telephone Services Survey (RTSS) collects information about telephone service across Canada.

The survey was conducted bi-annually in 2000, 2001, 2002 (May and November) and 2004 (May and December). In 2003, it was conducted once (in May). It was determined that sufficient information could be obtained by conducting the survey once a year so since 2005, the RTSS is conducted in December only. The RTSS was not conducted in 2009.

3.0 Objectives

There are three main objectives. They are:

- 1) to collect information on the telephone service subscription rates across Canada and make them available by province,
- 2) to collect information about the types of households without telephone service, and
- 3) To collect information about households with only wireless telephone service.

To accommodate these goals, and to ensure that the survey is focused on fulfilling these objectives, Statistics Canada developed an analysis plan which outlined data needs. This plan was used to design the questionnaire and to justify the variables required.

4.0 Concepts and Definitions

This chapter outlines concepts and definitions of interest to the users.

Active cell phones

These refer to cellular telephones and mobile phones such as the iPhone and the BlackBerry operated by members of the household. Included are those used for business purposes. Cordless phones are excluded.

Wired telephone service

These are household telephone services which are connected through a physical wire, such as landline, cable, or through Voice over Internet Protocol (VoIP). It excludes wireless services such as cellular telephones and other mobile phone devices.

Landline telephone service

This refers to the traditional telephone service which has been available for decades from local telephone companies.

Cable telephone

Cable television companies offer cable telephone service that delivers and receives telephone calls using the cable network.

VoIP service

“Voice over Internet Protocol” or “VoIP” allows the customer to make and receive calls using the Internet either through a device connected to a telephone or through a computer.

5.0 Survey Methodology

The Residential Telephone Service Survey (RTSS) was administered in December 2010 to a sub-sample of the dwellings in the Labour Force Survey (LFS) sample, and therefore its sample design is closely tied to that of the LFS. A detailed description of the LFS design is available in the Statistics Canada publication entitled *Methodology of the Canadian Labour Force Survey*, Catalogue no. 71-526-XPB. Sections 5.1 and 5.2 describe how the Residential Telephone Service Survey departed from the basic LFS design in December 2010.

5.1 Modifications to the Labour Force Survey Design for the Residential Telephone Service Survey

The Residential Telephone Service Survey used rotation groups 1 and 2 for Ontario and Quebec and rotation groups 1, 2 and 4 for all the other provinces in the December 2010 LFS sample. For the RTSS, the coverage of the LFS was set at the household level. However, unlike the LFS where information is collected for all eligible household members, the RTSS only collected information from one household member who reported about the household.

5.2 Sample Size by Province for the Residential Telephone Service Survey

The following table shows the number of households in the LFS sampled rotations that were eligible for the Residential Telephone Service Survey supplement.

Province	Sample Size
Newfoundland and Labrador	912
Prince Edward Island	614
Nova Scotia	1,400
New Brunswick	1,346
Quebec	3,054
Ontario	4,734
Manitoba	2,156
Saskatchewan	1,824
Alberta	2,426
British Columbia	2,831
Canada	21,297

6.0 Data Collection

Data collection for the Labour Force Survey (LFS) is carried out each month during the week following the LFS reference week. The reference week is normally the week containing the 15th day of the month.

6.1 Interviewing for the Labour Force Survey

Statistics Canada interviewers are employees hired and trained to carry out the LFS and other household surveys. Each month they contact the sampled dwellings to obtain the required labour force information. Each interviewer contacts approximately 75 dwellings per month.

Dwellings new to the sample are usually contacted through a personal visit using the computer-assisted personal interview (CAPI). The interviewer first obtains socio-demographic information for each household member and then obtains labour force information for all members aged 15 and over who are not members of the regular armed forces. Provided there is a telephone in the dwelling and permission has been granted, subsequent interviews are conducted by telephone. This is done out of a centralized computer-assisted telephone interviewing (CATI) unit where cases are assigned randomly to interviewers. As a result, approximately 85% of all households are interviewed by telephone. In these subsequent monthly interviews, the interviewer confirms the socio-demographic information collected in the first month and collects the labour force information for the current month.

In each dwelling, information about all household members is usually obtained from one knowledgeable household member. Such “proxy” reporting, which accounts for approximately 65% of the information collected, is used to avoid the high cost and extended time requirements that would be involved in repeat visits or calls necessary to obtain information directly from each respondent.

If, during the course of the six months that a dwelling normally remains in the sample, an entire household moves out and is replaced by a new household, information is obtained about the new household for the remainder of the six-month period.

At the conclusion of the LFS monthly interviews, interviewers introduce the supplementary survey, if any, to be administered to some or all household members that month.

6.2 Supervision and Quality Control

All LFS interviewers are under the supervision of a staff of senior interviewers who are responsible for ensuring that interviewers are familiar with the concepts and procedures of the LFS and its many supplementary surveys, and also for periodically monitoring their interviewers and reviewing their completed documents. The senior interviewers are, in turn, under the supervision of the LFS program managers, located in each of the Statistics Canada regional offices.

6.3 Non-response to the Labour Force Survey

Interviewers are instructed to make all reasonable attempts to obtain LFS interviews with members of eligible households. For individuals who at first refuse to participate in the LFS, a letter is sent from the Regional Office to the dwelling address stressing the importance of the survey and the household's cooperation. This is followed by a second call (or visit) from the interviewer. For cases in which the timing of the interviewer's call (or visit) is inconvenient, an appointment is arranged to call back at a more convenient time. For cases in which there is no one home, numerous call backs are made. Under no circumstances are sampled dwellings replaced by other dwellings for reasons of non-response.

Each month, after all attempts to obtain interviews have been made, a small number of non-responding households remain. For households non-responding to the LFS and for which LFS information was obtained in the previous month, this information is brought forward and used as the current month's LFS information. No supplementary survey information is collected for these households.

6.4 Data Collection Modifications for the Residential Telephone Service Survey

Information for the Residential Telephone Service Survey (RTSS) was obtained from a knowledgeable household member. Upon completion of the Labour Force Survey interview, the interviewer introduced the RTSS and proceeded with the interview with the respondent's permission.

The Residential Telephone Service Survey was programmed to appear on the list of surveys to be completed after the demographic component for the LFS had been completed. Any RTSS component not completed at the time the LFS was transmitted to one of the Statistics Canada regional offices was left incomplete and transmitted with the LFS.

6.5 Non-response to the Residential Telephone Service Survey

For households responding to the LFS, the next stage of data collection was to administer the Residential Telephone Service Survey. In total, 21,297 households were eligible for the supplementary survey; the RTSS interview was completed for 19,310 of these households for a response rate of 90.7%. More detailed information on response rates is presented in Chapter 8.0 (Data Quality).

7.0 Data Processing

The main output of the Residential Telephone Service Survey (RTSS) is a "clean" microdata file. This chapter presents a brief summary of the processing steps involved in producing this file.

7.1 Data Capture

Responses to survey questions are captured directly by the interviewer at the time of the interview using a computerized questionnaire. The computerized questionnaire reduces processing time and costs associated with data entry, transcription errors, and data transmission. The response data are encrypted to ensure confidentiality and sent via modem to the appropriate Statistics Canada Regional Office. From there they are transmitted over a secure line to Ottawa for further processing. In total 21,297 documents were captured and transmitted for the survey.

Some editing is done directly at the time of the interview. Where the information entered is out of range (too large or small) of expected values, or inconsistent with the previous entries, the interviewer is prompted, through message screens on the computer, to modify the information. However, for some questions interviewers have the option of bypassing the edits, and of skipping questions if the respondent does not know the answer or refuses to answer. Therefore, the response data are subjected to further edit and imputation processes once they arrive in head office.

7.2 Editing

The first stage of survey processing undertaken at head office was the replacement of any "out-of-range" values on the data file with blanks. This process was designed to make further editing easier.

The first type of error treated was errors in questionnaire flow, where questions which did not apply to the respondent (and should therefore not have been answered) were found to contain answers. In this case a computer edit automatically eliminated superfluous data by following the flow of the questionnaire implied by answers to previous, and in some cases, subsequent questions.

The second type of error treated involved a lack of information in questions which should have been answered. For this type of error, a non-response or "not-stated" code was assigned to the item.

7.3 Coding of Open-ended Questions

No data items on the questionnaire were recorded by interviewers in an open-ended format. One partially open-ended question was included in the survey. This related to "Other - specify" responses to question RTS_Q09 (Excluding cell phones, what other type of telephone service do you have in your household? Is it...?). A total of 49 responses were entered in this category of which 22 were manually coded to the correct response.

7.4 Creation of Derived Variables

A number of data items on the microdata file have been derived by combining items on the questionnaire in order to facilitate data analysis. The derived variable CMA1, for example, is actually a combination of census metropolitan area (CMA) and census agglomeration (CA). The CAs have been recoded to 0, while the CMAs remain the same. The derived variable "size of area of residence" (AREASIZE) was also created. This variable provides a population size code

based on the 2006 Census definitions for population centre size classes in the Labour Force Survey (LFS) sample frame.

7.5 Weighting

The principle behind estimation in a probability sample such as the LFS is that each person in the sample "represents", besides himself or herself, several other persons not in the sample. For example, in a simple random 2% sample of the population, each person in the sample represents 50 persons in the population. The same principle also applies to households.

The weighting phase is a step which calculates, for each record, what this number is. This weight appears on the microdata file, and **must** be used to derive meaningful estimates from the survey. For example, if the number of households with one or more telephone numbers for their residence is to be estimated, it is done by selecting the records referring to those households in the sample with that characteristic and summing the weights entered on those records.

Details of the method used to calculate these weights are presented in Chapter 11.0.

7.6 Suppression of Confidential Information

It should be noted that the "Public Use" microdata files described above differ in a number of important respects from the survey "master" files held by Statistics Canada. These differences are the result of actions taken to protect the anonymity of individual survey respondents. Users requiring access to information excluded from the microdata files may purchase custom tabulations. Estimates generated will be released to the user, subject to meeting the guidelines for analysis and release outlined in Chapter 9.0 of this document.

Province - Suppression of Geographic Identifiers

The survey master data file includes explicit geographic identifiers for province and census metropolitan area. It is also possible to obtain, where sample sizes permit, estimates by population centre size (AREASIZE). The survey public use microdata files usually do not contain any geographic identifiers below the provincial level. However, since the RTSS is a household based survey, the variables CMA and population centre size are on the public use microdata file.

8.0 Data Quality

8.1 Response Rates

The following table summarizes the response rates to the Labour Force Survey (LFS) and to the Residential Telephone Service Survey (RTSS) in December 2010.

Province	Household Response Rate for Full LFS* December 2010	Household Response Rate for LFS Rotations 1 and 2 in Ontario and Quebec and 1, 2 and 4 in the other provinces*	Household Response Rate for RTSS**	RTSS Responding Households
	%			
Newfoundland and Labrador	92.8	92.6	94.6	863
Prince Edward Island	91.7	92.1	93.3	573
Nova Scotia	92.5	92.7	92.4	1,294
New Brunswick	92.9	93.4	93.3	1,256
Québec	90.7	91.1	93.2	2,845
Ontario	87.3	88.2	90.1	4,263
Manitoba	90.6	90.7	91.7	1,978
Saskatchewan	90.6	90.9	93.3	1,701
Alberta	86.8	87.5	88.5	2,148
British Columbia	88.1	88.9	84.4	2,389
Canada	89.3	90.0	90.7	19,310

* The LFS response rate is the number of responding households as a percentage of the number of eligible households.

** The RTSS response rate is the number of households responding to the RTSS as a percentage of the number of households responding to the LFS, in the rotations sampled.

8.2 Survey Errors

The estimates derived from this survey are based on a sample of households. Somewhat different estimates might have been obtained if a complete census had been taken using the same questionnaire, interviewers, supervisors, processing methods, etc. as those actually used in the survey. The difference between the estimates obtained from the sample and those resulting from a complete count taken under similar conditions is called the sampling error of the estimate.

Errors which are not related to sampling may occur at almost every phase of a survey operation. Interviewers may misunderstand instructions, respondents may make errors in answering questions, the answers may be incorrectly entered on the questionnaire and errors may be introduced in the processing and tabulation of the data. These are all examples of non-sampling errors.

Over a large number of observations, randomly occurring errors will have little effect on estimates derived from the survey. However, errors occurring systematically will contribute to biases in the survey estimates. Considerable time and effort was made to reduce non-sampling errors in the

survey. Quality assurance measures were implemented at each step of the data collection and processing cycle to monitor the quality of the data. These measures include the use of highly skilled interviewers, extensive training of interviewers with respect to the survey procedures and questionnaire, observation of interviewers to detect problems of questionnaire design or misunderstanding of instructions, procedures to ensure that data capture errors were minimized, and coding and edit quality checks to verify the processing logic.

8.2.1 The Frame

Because the Residential Telephone Service Survey was a supplement to the LFS, the frame used was the LFS frame. Any non-response to the LFS had an impact on the RTSS frame. Because non-response to the LFS is quite low (usually less than 5%) this impact was minimal. The quality of the sampling variables in the frame was very high. The RTSS sample consisted of three rotation groups from the LFS. No records were dropped due to missing rotation group number or any other type of sampling variable.

Note that the LFS frame excludes about 2% of all households in the 10 provinces of Canada. Therefore, the RTSS frame also excludes the same proportion of households in the same geographical area. It is unlikely that this exclusion introduces any significant bias into the survey data.

All variables in the LFS frame are updated monthly.

Some variables on the sampling frame play a critical role with respect to the software application used in the survey. For example, in the RTSS computer-assisted interviewing (CAI) application, each record must have accurate stratum, cluster and rotation group codes. These variables are always of very high quality each month in the LFS.

At times, duplication of records occurs. There were 256 duplicate records in the December 2010 collection.

8.2.2 Data Collection

Interviewer training consisted of reading the RTSS Procedures Manual, practicing with the RTSS training cases on the computer, and discussing any questions with senior interviewers before the start of the survey. A description of the background and objectives of the survey was provided, as well as a glossary of terms and a set of questions and answers. Interviewers collected RTSS information after the LFS information was collected. The collection period ran from the 12th to the 21st of December 2010.

8.2.3 Data Processing

During processing of the data, 399 RTSS records did not match to corresponding records in the LFS. Thus they were coded as out-of-scope and were dropped from further processing. When supplementary survey records do not match to host survey records they must be dropped since a weight cannot be derived for them.

Conversely, 252 records in the LFS were found that should have matched to an RTSS record but did not. These records were coded as in-scope, since they were eligible records from the frame which, for one reason or another, did not have corresponding RTSS records. These records were considered to be non-responding records, and were used in the weighting process to adjust for non-response.

Data processing of the RTSS was straightforward since there were only nine questions on the CAI application. Any record that contained a “Refused” or “Don’t know” response in questions RTS_Q01 and RTS_Q05 was coded as non-response. Note that 3,611 records were treated this way. Since the data was collected using a CAI instrument, data quality before processing was very high. Very few changes were made to the data during editing.

8.2.4 Non-response

A major source of non-sampling errors in surveys is the effect of non-response on the survey results. The extent of non-response varies from partial non-response (failure to answer just one or some questions) to total non-response. Total non-response occurred because the interviewer was either unable to contact the respondent, no member of the household was able to provide the information, or the respondent refused to participate in the survey. Total non-response was handled by adjusting the weight of households that responded to the survey to compensate for those that did not respond.

In most cases, partial non-response to the survey occurred when the respondent did not understand or misinterpreted a question, refused to answer a question, or could not recall the requested information.

Item non-response is usually very low for the RTSS. For questions RTS_Q01 to RTS_Q09 item non-response ranged from a low of 0.088% for RTS_Q01 to a high of 0.725% for RTS_Q08.

8.2.5 Measurement of Sampling Error

Since it is an unavoidable fact that estimates from a sample survey are subject to sampling error, sound statistical practice calls for researchers to provide users with some indication of the magnitude of this sampling error. This section of the documentation outlines the measures of sampling error which Statistics Canada commonly uses and which it urges users producing estimates from this microdata file to use also.

The basis for measuring the potential size of sampling errors is the standard error of the estimates derived from survey results.

However, because of the large variety of estimates that can be produced from a survey, the standard error of an estimate is usually expressed relative to the estimate to which it pertains. This resulting measure, known as the coefficient of variation (CV) of an estimate, is obtained by dividing the standard error of the estimate by the estimate itself and is expressed as a percentage of the estimate.

For example, suppose that, based upon the November 2002 survey results, one estimates that 1.5% of Canadian households did not have telephone service to their residence (RTS_Q01B = 1, None), and this estimate is found to have a standard error of 0.00092. Then the coefficient of variation of the estimate is calculated as:

$$\left(\frac{0.00092}{0.015} \right) \times 100 \% = 6.1 \%$$

There is more information on the calculation of coefficient of variation in Chapter 10.0.

9.0 Guidelines for Tabulation, Analysis and Release

This chapter of the documentation outlines the guidelines to be adhered to by users tabulating, analysing, publishing or otherwise releasing any data derived from the survey microdata files. With the aid of these guidelines, users of microdata should be able to produce the same figures as those produced by Statistics Canada and, at the same time, will be able to develop currently unpublished figures in a manner consistent with these established guidelines.

9.1 Rounding Guidelines

In order that estimates for publication or other release derived from these microdata files correspond to those produced by Statistics Canada, users are urged to adhere to the following guidelines regarding the rounding of such estimates:

- a) Estimates in the main body of a statistical table are to be rounded to the nearest hundred units using the normal rounding technique. In normal rounding, if the first or only digit to be dropped is 0 to 4, the last digit to be retained is not changed. If the first or only digit to be dropped is 5 to 9, the last digit to be retained is raised by one. For example, in normal rounding to the nearest 100, if the last two digits are between 00 and 49, they are changed to 00 and the preceding digit (the hundreds digit) is left unchanged. If the last two digits are between 50 and 99 they are changed to 00 and the preceding digit is incremented by 1.
- b) Marginal sub-totals and totals in statistical tables are to be derived from their corresponding unrounded components and then are to be rounded themselves to the nearest 100 units using normal rounding.
- c) Averages, proportions, rates and percentages are to be computed from unrounded components (i.e. numerators and/or denominators) and then are to be rounded themselves to one decimal using normal rounding. In normal rounding to a single digit, if the final or only digit to be dropped is 0 to 4, the last digit to be retained is not changed. If the first or only digit to be dropped is 5 to 9, the last digit to be retained is increased by 1.
- d) Sums and differences of aggregates (or ratios) are to be derived from their corresponding unrounded components and then are to be rounded themselves to the nearest 100 units (or the nearest one decimal) using normal rounding.
- e) In instances where, due to technical or other limitations, a rounding technique other than normal rounding is used resulting in estimates to be published or otherwise released which differ from corresponding estimates published by Statistics Canada, users are urged to note the reason for such differences in the publication or release document(s).
- f) Under no circumstances are unrounded estimates to be published or otherwise released by users. Unrounded estimates imply greater precision than actually exists.

9.2 Sample Weighting Guidelines for Tabulation

The sample design used for the Residential Telephone Service Survey (RTSS) was not self-weighting. When producing simple estimates, including the production of ordinary statistical tables, users must apply the proper sampling weight.

If proper weights are not used, the estimates derived from the microdata files cannot be considered to be representative of the survey population, and will not correspond to those produced by Statistics Canada.

Users should also note that some software packages may not allow the generation of estimates that exactly match those available from Statistics Canada, because of their treatment of the weight field.

9.3 Definitions of Types of Estimates: Categorical and Quantitative

Before discussing how the RTSS data can be tabulated and analysed, it is useful to describe the two main types of point estimates of population characteristics which can be generated from the microdata file for the RTSS.

9.3.1 Categorical Estimates

Categorical estimates are estimates of the number, or percentage of the surveyed population possessing certain characteristics or falling into some defined category. The number of households which did not have telephone service for their residence during the reference month or the proportion of households which had two or more telephone lines for their residence are examples of such estimates. An estimate of the number of households possessing a certain characteristic may also be referred to as an estimate of an aggregate.

Example of a Categorical Question:

Q: Does anyone in your household currently have an active cell phone? Please include cell phones used for business.
R: Yes / No

9.3.2 Quantitative Estimates

Quantitative estimates are estimates of totals or of means, medians and other measures of central tendency of quantities based upon some or all of the members of the surveyed population. They also specifically involve estimates of the form $\hat{X} / \hat{Y}$ where $\hat{X}$ is an estimate of surveyed population quantity total and $\hat{Y}$ is an estimate of the number of households in the surveyed population contributing to that total quantity.

An example of a quantitative estimate is the average number of cell phones per household reporting at least one active cell phone. The numerator is an estimate of the total number of cell phones recorded in RTS_Q02A, and its denominator is the number of households reporting at least one active cell phone.

Example of a Quantitative Question:

Q: How many active cell phones do the members of your household have?
R: active cell phones

9.3.3 Tabulation of Categorical Estimates

Estimates of the number of people with a certain characteristic can be obtained from the microdata file by summing the final weights of all records possessing the characteristic(s) of interest. Proportions and ratios of the form $\hat{X} / \hat{Y}$ are obtained by:

- a) summing the final weights of records having the characteristic of interest for the numerator $(\hat{X})$,
- b) summing the final weights of records having the characteristic of interest for the denominator $(\hat{Y})$, then
- c) dividing estimate a) by estimate b) $(\hat{X} / \hat{Y})$.

9.3.4 Tabulation of Quantitative Estimates

Estimates of quantities can be obtained from the microdata file by multiplying the value of the variable of interest by the final weight for each record, then summing this quantity over all records of interest. For example, to obtain an estimate of the total number of cell phones within households which are only used for business purposes, multiply the value reported in question RTS_Q04 (number of cell phones only used for business purposes) by the final weight for the record, then sum this value over all records with RTS_Q01 = 1 (households reporting at least one active cell phone).

To obtain a weighted average of the form $\hat{X} / \hat{Y}$, the numerator $(\hat{X})$ is calculated as for a quantitative estimate and the denominator $(\hat{Y})$ is calculated as for a categorical estimate. For example, to estimate the average number of cell phones only used for business purposes for households reporting at least one active cell phone:

- a) estimate the total number of cell phones within households only used for business purposes $(\hat{X})$ as described above,
- b) estimate the number of households reporting at least one active cell phone $(\hat{Y})$ by summing the final weights of all records with RTS_Q01 = 1, then
- c) divide estimate a) by estimate b) $(\hat{X} / \hat{Y})$.

9.4 Guidelines for Statistical Analysis

The Residential Telephone Service Survey is based upon a complex sample design, with stratification, multiple stages of selection, and unequal probabilities of selection of respondents. Using data from such complex surveys presents problems to analysts because the survey design and the selection probabilities affect the estimation and variance calculation procedures that should be used. In order for survey estimates and analyses to be free from bias, the survey weights must be used.

While many analysis procedures found in statistical packages allow weights to be used, the meaning or definition of the weight in these procedures differ from that which is appropriate in a sample survey framework, with the result that while in many cases the estimates produced by the packages are correct, the variances that are calculated are poor. Approximate variances for simple estimates such as totals, proportions and ratios (for qualitative variables) can be derived using the accompanying Approximate Sampling Variability Tables.

For other analysis techniques (for example linear regression, logistic regression and analysis of variance), a method exists which can make the variances calculated by the standard packages more meaningful, by incorporating the unequal probabilities of selection. The method rescales the weights so that there is an average weight of 1.

For example, suppose that analysis of all Quebec households is required. The steps to rescale the weights are as follows:

- 1) select all households from the file who reported PROVINCE = 24, Quebec;
- 2) calculate the AVERAGE weight for these records by summing the original household weights from the microdata file for these records and then dividing by the number of households who reported PROVINCE = 24;
- 3) for each of these records, calculate a RESCALED weight equal to the original household weight divided by the AVERAGE weight;
- 4) perform the analysis for these households using the RESCALED weight.

However, because the stratification and clustering of the sample's design are still not taken into account, the variance estimates calculated in this way are likely to be under-estimates.

The calculation of more precise variance estimates requires detailed knowledge of the design of the survey. Such detail cannot be given in this microdata file because of confidentiality. Variances that take the complete sample design into account can be calculated for many statistics by Statistics Canada on a cost-recovery basis.

9.5 Coefficient of Variation Release Guidelines

Before releasing and/or publishing any estimate from the Residential Telephone Service Survey, users should first determine the quality level of the estimate. The quality levels are *acceptable*, *marginal* and *unacceptable*. Data quality is affected by both sampling and non-sampling errors as discussed in Chapter 8.0. However for this purpose, the quality level of an estimate will be determined only on the basis of sampling error as reflected by the coefficient of variation as shown in the table below. Nonetheless users should be sure to read Chapter 8.0 to be more fully aware of the quality characteristics of these data.

First, the number of household records that contribute to the calculation of the estimate should be determined. If this number is less than 30, the weighted estimate should be considered to be of unacceptable quality.

For weighted estimates based on sample sizes of 30 or more, users should determine the coefficient of variation of the estimate and follow the guidelines below. These quality level guidelines should be applied to weighted rounded estimates.

All estimates can be considered releasable. However, those of marginal or unacceptable quality level must be accompanied by a warning to caution subsequent users.

Quality Level Guidelines

Quality Level of Estimate	Guidelines
1) Acceptable	Estimates have a sample size of 30 or more, and low coefficients of variation in the range of 0.0% to 16.5%. No warning is required.
2) Marginal	Estimates have a sample size of 30 or more, and high coefficients of variation in the range of 16.6% to 33.3%. Estimates should be flagged with the letter E (or some similar identifier). They should be accompanied by a warning to caution subsequent users about the high levels of error, associated with the estimates.
3) Unacceptable	Estimates have a sample size of less than 30, or very high coefficients of variation in excess of 33.3%. Statistics Canada recommends not to release estimates of unacceptable quality. However, if the user chooses to do so then estimates should be flagged with the letter F (or some similar identifier) and the following warning should accompany the estimates: "Please be warned that these estimates [flagged with the letter F] do not meet Statistics Canada's quality standards. Conclusions based on these data will be unreliable, and most likely invalid."

9.6 Release Cut-off's for the Residential Telephone Service Survey

The following table provides an indication of the precision of population estimates as it shows the release cut-offs associated with each of the three quality levels presented in the previous section. These cut-offs are derived from the coefficient of variation (CV) tables discussed in Chapter 10.0.

For example, the table shows that the quality of a weighted estimate of 10,000 households possessing a given characteristic in Newfoundland and Labrador is marginal.

Note that these cut-offs apply to estimates of population totals only. To estimate ratios, users should not use the numerator value (nor the denominator) in order to find the corresponding quality level. Rule 4 in Section 10.1 and Example 4 in Section 10.1.1 explains the correct procedure to be used for ratios.

Province and Region	Acceptable CV 0.0% to 16.5%	Marginal CV 16.6% to 33.3%	Unacceptable CV > 33.3%
Newfoundland and Labrador	10,200 & over	2,600 to < 10,200	under 2,600
Prince Edward Island	4,000 & over	1,000 to < 4,000	under 1,000
Nova Scotia	14,200 & over	3,600 to < 14,200	under 3,600
New Brunswick	10,800 & over	2,700 to < 10,800	under 2,700
Quebec	94,100 & over	23,600 to < 94,100	under 23,600
Ontario	85,000 & over	21,100 to < 85,000	under 21,100
Manitoba	15,400 & over	3,900 to < 15,400	under 3,900
Saskatchewan	13,300 & over	3,400 to < 13,300	under 3,400
Alberta	37,700 & over	9,500 to < 37,700	under 9,500
British Columbia	43,300 & over	10,800 to < 43,300	under 10,800
Atlantic Provinces	12,000 & over	3,000 to < 12,000	under 3,000
Prairie Provinces	30,000 & over	7,400 to < 30,000	under 7,400
Canada	57,100 & over	14,100 to < 57,100	under 14,100

10.0 Approximate Sampling Variability Tables

In order to supply coefficients of variation (CV) which would be applicable to a wide variety of categorical estimates produced from this microdata file and which could be readily accessed by the user, a set of Approximate Sampling Variability Tables has been produced. These CV tables allow the user to obtain an approximate coefficient of variation based on the size of the estimate calculated from the survey data.

The coefficients of variation are derived using the variance formula for simple random sampling and incorporating a factor which reflects the multi-stage, clustered nature of the sample design. This factor, known as the design effect, was determined by first calculating design effects for a wide range of characteristics and then choosing from among these a conservative value (usually the 75th percentile) to be used in the CV tables which would then apply to the entire set of characteristics.

The table below shows the conservative value of the design effects as well as sample sizes and population counts by province which were used to produce the Approximate Sampling Variability Tables for the Residential Telephone Service Survey (RTSS).

Province and Region	Design Effect	Sample Size	Population
Newfoundland and Labrador	1.22	863	207,431
Prince Edward Island	1.17	573	58,493
Nova Scotia	1.31	1,294	396,978
New Brunswick	1.21	1,256	317,024
Quebec	2.19	2,845	3,425,042
Ontario	1.96	4,263	5,121,451
Manitoba	1.83	1,978	470,186
Saskatchewan	1.56	1,701	410,903
Alberta	1.60	2,148	1,419,224
British Columbia	1.58	2,389	1,826,430
Atlantic Provinces	1.35	3,986	979,927
Prairie Provinces	2.10	5,827	2,300,312
Canada	2.21	19,310	13,653,162

All coefficients of variation in the Approximate Sampling Variability Tables are approximate and, therefore, unofficial. Estimates of actual variance for specific variables may be obtained from Statistics Canada on a cost-recovery basis. Since the approximate CV is conservative, the use of actual variance estimates may cause the estimate to be switched from one quality level to another. For instance a *marginal* estimate could become *acceptable* based on the exact CV calculation.

Remember: If the number of observations on which an estimate is based is less than 30, the weighted estimate is most likely unacceptable and Statistics Canada recommends not to release such an estimate, regardless of the value of the coefficient of variation.

10.1 How to Use the Coefficient of Variation Tables for Categorical Estimates

The following rules should enable the user to determine the approximate coefficients of variation from the Approximate Sampling Variability Tables for estimates of the number, proportion or percentage of the surveyed population possessing a certain characteristic and for ratios and differences between such estimates.

Rule 1: Estimates of Numbers of Households Possessing a Characteristic (Aggregates)

The coefficient of variation depends only on the size of the estimate itself. On the Approximate Sampling Variability Table for the appropriate geographic area, locate the estimated number in the left-most column of the table (headed "Numerator of Percentage") and follow the asterisks (if any) across to the first figure encountered. This figure is the approximate coefficient of variation.

Rule 2: Estimates of Proportions or Percentages of Households Possessing a Characteristic

The coefficient of variation of an estimated proportion or percentage depends on both the size of the proportion or percentage and the size of the total upon which the proportion or percentage is based. Estimated proportions or percentages are relatively more reliable than the corresponding estimates of the numerator of the proportion or percentage, when the proportion or percentage is based upon a sub-group of the population. For example, the proportion of households which did not have telephone service for their residence during the reference period is more reliable than the estimated number of households which did not have telephone service for their residence during the reference period. (Note that in the tables the coefficients of variation decline in value reading from left to right).

When the proportion or percentage is based upon the total population of the geographic area covered by the table, the CV of the proportion or percentage is the same as the CV of the numerator of the proportion or percentage. In this case, Rule 1 can be used.

When the proportion or percentage is based upon a subset of the total population (e.g. those in a particular province or census metropolitan area), reference should be made to the proportion or percentage (across the top of the table) and to the numerator of the proportion or percentage (down the left side of the table). The intersection of the appropriate row and column gives the coefficient of variation.

Rule 3: Estimates of Differences Between Aggregates or Percentages

The standard error of a difference between two estimates is approximately equal to the square root of the sum of squares of each standard error considered separately. That is, the standard error of a difference $(\hat{d} = \hat{X}_1 - \hat{X}_2)$ is:

$$\sigma_{\hat{d}} = \sqrt{(\hat{X}_1 \alpha_1)^2 + (\hat{X}_2 \alpha_2)^2}$$

where $\hat{X}_1$ is estimate 1, $\hat{X}_2$ is estimate 2, and α_1 and α_2 are the coefficients of variation of $\hat{X}_1$ and $\hat{X}_2$ respectively. The coefficient of variation of $\hat{d}$ is given by $\sigma_{\hat{d}} / \hat{d}$. This formula is accurate for the difference between separate and uncorrelated characteristics, but is only approximate otherwise.

Rule 4: Estimates of Ratios

In the case where the numerator is a subset of the denominator, the ratio should be converted to a percentage and Rule 2 applied. This would apply, for example, to the case where the denominator is the number of households which did not have telephone service for their residence during the reference period and the numerator is the number of households which did not have telephone service to their residence during the reference period because they could not afford it.

In the case where the numerator is not a subset of the denominator, as for example, the ratio of the number of households in Quebec with a cell phone as compared to the number of households in Ontario with a cell phone, the standard error of the ratio of the estimates is approximately equal to the square root of the sum of squares of each coefficient of variation considered separately multiplied by $\hat{R}$. That is, the standard error of a ratio ($\hat{R} = \hat{X}_1 / \hat{X}_2$) is:

$$\sigma_{\hat{R}} = \hat{R} \sqrt{\alpha_1^2 + \alpha_2^2}$$

where α_1 and α_2 are the coefficients of variation of $\hat{X}_1$ and $\hat{X}_2$ respectively. The coefficient of variation of $\hat{R}$ is given by $\sigma_{\hat{R}} / \hat{R}$. The formula will tend to overstate the error, if $\hat{X}_1$ and $\hat{X}_2$ are positively correlated and understate the error if $\hat{X}_1$ and $\hat{X}_2$ are negatively correlated.

Rule 5: Estimates of Differences of Ratios

In this case, Rules 3 and 4 are combined. The CVs for the two ratios are first determined using Rule 4, and then the CV of their difference is found using Rule 3.

10.1.1 Examples of Using the Coefficient of Variation Tables for Categorical Estimates

The following examples based on the November 2002 RTSS are included to assist users in applying the foregoing rules. Please note that the data for these examples are different than the results obtained from the current survey and are only to be used as a guide.

Example 1: Estimates of Numbers of Households Possessing a Characteristic (Aggregates)

Suppose that a user estimates that 177,859 households did not have telephone service for their residence during the reference period (Q01B = 1, None). How does the user determine the coefficient of variation of this estimate?

- 1) Refer to the coefficient of variation table for CANADA.
- 2) The estimated aggregate (177,859) does not appear in the left-hand column (the "Numerator of Percentage" column), so it is necessary to use the figure closest to it, namely 200,000.
- 3) The coefficient of variation for an estimated aggregate is found by referring to the first non-asterisk entry on that row, namely, 5.6%.

RESIDENTIAL TELEPHONE SERVICE SURVEY - NOVEMBER 2002													
Approximate Sampling Variability Tables - Canada													
NUMERATOR OF PERCENTAGE (‘000)	ESTIMATED PERCENTAGE												
	0.1%	1.0%	2.0%	5.0%	10.0%	15.0%	20.0%	...	35.0%	40.0%	50.0%	70.0%	90.0%
1	79.3	79.0	78.6	77.4	75.3	73.2	71.0		64.0	61.5	56.1	43.5	25.1
2	56.1	55.8	55.6	54.7	53.2	51.7	50.2		45.3	43.5	39.7	30.7	17.7
3	45.8	45.6	45.4	44.7	43.5	42.3	41.0		36.9	35.5	32.4	25.1	14.5
4	39.7	39.5	39.3	38.7	37.7	36.6	35.5		32.0	30.7	28.1	21.7	12.6
5	35.5	35.3	35.1	34.6	33.7	32.7	31.8		28.6	27.5	25.1	19.4	11.2
.													
.													
50	*****	11.2	11.1	10.9	10.6	10.3	10.0		9.1	8.7	7.9	6.1	3.5
55	*****	10.6	10.6	10.4	10.2	9.9	9.6		8.6	8.3	7.6	5.9	3.4
60	*****	10.2	10.1	10.0	9.7	9.4	9.2		8.3	7.9	7.2	5.6	3.2
65	*****	9.8	9.7	9.6	9.3	9.1	8.8		7.9	7.6	7.0	5.4	3.1
70	*****	9.4	9.4	9.2	9.0	8.7	8.5		7.6	7.3	6.7	5.2	3.0
75	*****	9.1	9.1	8.9	8.7	8.5	8.2		7.4	7.1	6.5	5.0	2.9
80	*****	8.8	8.8	8.6	8.4	8.2	7.9		7.2	6.9	6.3	4.9	2.8
85	*****	8.6	8.5	8.4	8.2	7.9	7.7		6.9	6.7	6.1	4.7	2.7
90	*****	8.3	8.3	8.2	7.9	7.7	7.5		6.7	6.5	5.9	4.6	2.6
95	*****	8.1	8.1	7.9	7.7	7.5	7.3		6.6	6.3	5.8	4.5	2.6
100	*****	7.9	7.9	7.7	7.5	7.3	7.1		6.4	6.1	5.6	4.3	2.5
125	*****	*****	7.0	6.9	6.7	6.5	6.4		5.7	5.5	5.0	3.9	2.2
150	*****	*****	6.4	6.3	6.1	6.0	5.8		5.2	5.0	4.6	3.5	2.0
200	*****	*****	5.6	5.5	5.3	5.2	5.0		4.5	4.3	4.0	3.1	1.8
250	*****	*****	*****	4.9	4.8	4.6	4.5		4.0	3.9	3.5	2.7	1.6
300	*****	*****	*****	4.5	4.3	4.2	4.1		3.7	3.5	3.2	2.5	1.4
350	*****	*****	*****	4.1	4.0	3.9	3.8		3.4	3.3	3.0	2.3	1.3
400	*****	*****	*****	3.9	3.8	3.7	3.5		3.2	3.1	2.8	2.2	1.3
450	*****	*****	*****	3.6	3.5	3.4	3.3		3.0	2.9	2.6	2.0	1.2
500	*****	*****	*****	3.5	3.4	3.3	3.2		2.9	2.7	2.5	1.9	1.1
.													
.													
.													
8000	*****	*****	*****	*****	*****	*****	*****		*****	*****	*****	0.5	0.3
9000	*****	*****	*****	*****	*****	*****	*****		*****	*****	*****	*****	0.3
10000	*****	*****	*****	*****	*****	*****	*****		*****	*****	*****	*****	0.3

NOTE: FOR CORRECT USAGE OF THESE TABLES, PLEASE REFER TO MICRODATA DOCUMENTATION.

- 4) So the approximate coefficient of variation of the estimate is 5.6%. The finding that 177,859 (to be rounded according to the rounding guidelines in Section 9.1) households did not have telephone service for their residence during the reference period is publishable with no qualifications.

Example 2: Estimates of Proportions or Percentages of Households Possessing a Characteristic

Suppose that the user estimates that $100,791 / 177,859 = 56.7\%$ of households which did not have telephone service for their residence during the reference period (Q01B = 1, None) reported that they could not afford telephone service (Q02 = 2, Can't afford it or Q03 = 1, Can't afford it). How does the user determine the coefficient of variation of this estimate?

- 1) Refer to the coefficient of variation table for CANADA (see above).
- 2) Because the estimate is a percentage which is based on a subset of the total population (i.e., households which did not have telephone service for their residence during the reference period), it is necessary to use both the percentage (56.7%) and the numerator portion of the percentage (100,791) in determining the coefficient of variation.
- 3) The numerator, 100,791 does not appear in the left-hand column (the "Numerator of Percentage" column) so it is necessary to use the figure closest to it, namely 100,000. Similarly, the percentage estimate does not appear as any of the column headings, so it is necessary to use the percentage closest to it, 50.0%.
- 4) The figure at the intersection of the row and column used, namely 5.6% is the coefficient of variation to be used.
- 5) So the approximate coefficient of variation of the estimate is 5.6%. The finding that 56.7% of households which did not have telephone service for their residence during the reference period could not afford telephone service can be published with no qualifications.

Example 3: Estimates of Differences Between Aggregates or Percentages

Suppose that a user estimates that $796,352 / 3,103,651 = 25.7\%$ of households in Quebec (PROV1 = 24, Quebec) reported that their total annual income was less than the low income cut-off (LICO) (Q10 = 1, Less than LICO), while $858,163 / 4,526,715 = 19.0\%$ of households in Ontario (PROV1 = 35, Ontario) reported that their total annual income was less than LICO (Q10 = 1, Less than LICO). How does the user determine the coefficient of variation of the difference between these two estimates?

- 1) Using the QUEBEC and ONTARIO coefficient of variation tables in the same manner as described in Example 1 gives the CV of the estimate for households in Quebec as 2.8%, and the CV of the estimate for households in Ontario as 3.5%.

RESIDENTIAL TELEPHONE SERVICE SURVEY - NOVEMBER 2002
Approximate Sampling Variability Tables - Quebec

NUMERATOR OF PERCENTAGE ('000)	ESTIMATED PERCENTAGE													
	0.1%	1.0%	2.0%	5.0%	10.0%	15.0%	20.0%	25.0%	30.0%	35.0%	40.0%	50.0%	70.0%	90.0%
1	88.9	88.5	88.1	86.7	84.4	82.0	79.6	77.0	74.4	71.7	68.9	62.9	48.7	28.1
2	62.9	62.6	62.3	61.3	59.7	58.0	56.3	54.5	52.6	50.7	48.7	44.5	34.5	19.9
3	51.3	51.1	50.8	50.1	48.7	47.4	45.9	44.5	43.0	41.4	39.8	36.3	28.1	16.2
4	*****	44.3	44.0	43.4	42.2	41.0	39.8	38.5	37.2	35.9	34.5	31.5	24.4	14.1
5	*****	39.6	39.4	38.8	37.7	36.7	35.6	34.5	33.3	32.1	30.8	28.1	21.8	12.6
.														
.														
.														
200	*****	*****	*****	*****	6.0	5.8	5.6	5.4	5.3	5.1	4.9	4.4	3.4	2.0
250	*****	*****	*****	*****	5.3	5.2	5.0	4.9	4.7	4.5	4.4	4.0	3.1	1.8
300	*****	*****	*****	*****	4.9	4.7	4.6	4.4	4.3	4.1	4.0	3.6	2.8	1.6
350	*****	*****	*****	*****	*****	4.4	4.3	4.1	4.0	3.8	3.7	3.4	2.6	1.5
400	*****	*****	*****	*****	*****	4.1	4.0	3.9	3.7	3.6	3.4	3.1	2.4	1.4
450	*****	*****	*****	*****	*****	3.9	3.8	3.6	3.5	3.4	3.2	3.0	2.3	1.3
500	*****	*****	*****	*****	*****	*****	3.6	3.4	3.3	3.2	3.1	2.8	2.2	1.3
750	*****	*****	*****	*****	*****	*****	*****	2.8	2.7	2.6	2.5	2.3	1.8	1.0
1000	*****	*****	*****	*****	*****	*****	*****	*****	*****	2.3	2.2	2.0	1.5	0.9
1500	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	1.6	1.3	0.7
2000	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	1.1	0.6

NOTE: FOR CORRECT USAGE OF THESE TABLES, PLEASE REFER TO MICRODATA DOCUMENTATION.

RESIDENTIAL TELEPHONE SERVICE SURVEY - NOVEMBER 2002
Approximate Sampling Variability Tables - Ontario

NUMERATOR OF PERCENTAGE ('000)	ESTIMATED PERCENTAGE													
	0.1%	1.0%	2.0%	5.0%	10.0%	15.0%	20.0%	25.0%	30.0%	35.0%	40.0%	50.0%	70.0%	90.0%
1	105.6	105.1	104.6	103.0	100.2	97.4	94.5	91.5	88.4	85.2	81.8	74.7	57.9	33.4
2	74.7	74.3	74.0	72.8	70.9	68.9	66.8	64.7	62.5	60.2	57.9	52.8	40.9	23.6
3	61.0	60.7	60.4	59.5	57.9	56.2	54.6	52.8	51.0	49.2	47.2	43.1	33.4	19.3
.														
.														
.														
250	*****	*****	*****	*****	6.3	6.2	6.0	5.8	5.6	5.4	5.2	4.7	3.7	2.1
300	*****	*****	*****	*****	5.8	5.6	5.5	5.3	5.1	4.9	4.7	4.3	3.3	1.9
350	*****	*****	*****	*****	5.4	5.2	5.1	4.9	4.7	4.6	4.4	4.0	3.1	1.8
400	*****	*****	*****	*****	5.0	4.9	4.7	4.6	4.4	4.3	4.1	3.7	2.9	1.7
450	*****	*****	*****	*****	4.7	4.6	4.5	4.3	4.2	4.0	3.9	3.5	2.7	1.6
500	*****	*****	*****	*****	*****	4.4	4.2	4.1	4.0	3.8	3.7	3.3	2.6	1.5
750	*****	*****	*****	*****	*****	*****	3.5	3.3	3.2	3.1	3.0	2.7	2.1	1.2
1000	*****	*****	*****	*****	*****	*****	*****	2.9	2.8	2.7	2.6	2.4	1.8	1.1
1500	*****	*****	*****	*****	*****	*****	*****	*****	*****	2.2	2.1	1.9	1.5	0.9
2000	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	1.7	1.3	0.7
3000	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	1.1	0.6
4000	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	0.5

NOTE: FOR CORRECT USAGE OF THESE TABLES, PLEASE REFER TO MICRODATA DOCUMENTATION.

- 2) Using Rule 3, the standard error of a difference ($\hat{d} = \hat{X}_1 - \hat{X}_2$) is:

$$\sigma_{\hat{d}} = \sqrt{(\hat{X}_1 \alpha_1)^2 + (\hat{X}_2 \alpha_2)^2}$$

where $\hat{X}_1$ is estimate 1 (Quebec), $\hat{X}_2$ is estimate 2 (Ontario), and α_1 and α_2 are the coefficients of variation of $\hat{X}_1$ and $\hat{X}_2$ respectively.

That is, the standard error of the difference $\hat{d} = 0.257 - 0.190 = 0.067$ is:

$$\begin{aligned}\sigma_{\hat{d}} &= \sqrt{[(0.257)(0.028)]^2 + [(0.190)(0.035)]^2} \\ &= \sqrt{(0.0000517) + (0.0000442)} \\ &= 0.0098\end{aligned}$$

- 3) The coefficient of variation of $\hat{d}$ is given by $\sigma_{\hat{d}} / \hat{d} = 0.0098 / 0.067 = 0.146$.
- 4) So the approximate coefficient of variation of the difference between the estimates is 14.6%, which is releasable with no qualifications.

Example 4: Estimates of Ratios

Suppose that the user estimates that 796,352 households in Quebec (PROV1 = 24, Quebec) reported that their total annual income was less than the low income cut-off (LICO) (Q10 = 1, Less than LICO), while 858,163 households in Ontario (PROV1 = 35, Ontario) reported that their total annual income was less than LICO (Q10 = 1, Less than LICO). The user is interested in comparing the estimate of Quebec households versus that of Ontario households in the form of a ratio. How does the user determine the coefficient of variation of this estimate?

- 1) First of all, this estimate is a ratio estimate, where the numerator of the estimate ($\hat{X}_1$) is the number of households in Quebec which reported that their total annual income was less than LICO. The denominator of the estimate ($\hat{X}_2$) is the number of households in Ontario which reported that their total annual income was less than LICO.
- 2) Refer to the coefficient of variation tables for QUEBEC and ONTARIO (see above).
- 3) The numerator of this ratio estimate is 796,352. The figure closest to it is 750,000. The coefficient of variation for this estimate is found by referring to the first non-asterisk entry on that row in the QUEBEC CV table, namely, 2.8%.
- 4) The denominator of this ratio estimate is 858,163. The figure closest to it is 750,000. The coefficient of variation for this estimate is found by referring to the first non-asterisk entry on that row in the ONTARIO CV table, namely, 3.5%.
- 5) So the approximate coefficient of variation of the ratio estimate is given by Rule 4, which is:

$$\alpha_{\hat{R}} = \sqrt{\alpha_1^2 + \alpha_2^2}$$

where α_1 and α_2 are the coefficients of variation of $\hat{X}_1$ and $\hat{X}_2$ respectively. That is:

$$\begin{aligned}\alpha_{\hat{R}} &= \sqrt{(0.028)^2 + (0.035)^2} \\ &= \sqrt{0.000784 + 0.001225} \\ &= 0.045\end{aligned}$$

- 6) The obtained ratio of Quebec households versus Ontario households whose total annual income is less than the LICO is 796,352 / 858,163 which is 0.93:1 (to be rounded according to the rounding guidelines in Section 9.1). The coefficient of variation of this estimate is 4.5%, which makes the estimate releasable with no qualifications.

10.2 How to Use the Coefficient of Variation Tables to Obtain Confidence Limits

Although coefficients of variation are widely used, a more intuitively meaningful measure of sampling error is the confidence interval of an estimate. A confidence interval constitutes a statement on the level of confidence that the true value for the population lies within a specified range of values. For example a 95% confidence interval can be described as follows:

If sampling of the population is repeated indefinitely, each sample leading to a new confidence interval for an estimate, then in 95% of the samples the interval will cover the true population value.

Using the standard error of an estimate, confidence intervals for estimates may be obtained under the assumption that under repeated sampling of the population, the various estimates obtained for a population characteristic are normally distributed about the true population value. Under this assumption, the chances are about 68 out of 100 that the difference between a sample estimate and the true population value would be less than one standard error, about 95 out of 100 that the difference would be less than two standard errors, and about 99 out of 100 that the differences would be less than three standard errors. These different degrees of confidence are referred to as the confidence levels.

Confidence intervals for an estimate, $\hat{X}$ are generally expressed as two numbers, one below the estimate and one above the estimate, as $(\hat{X} - k, \hat{X} + k)$ where k is determined depending upon the level of confidence desired and the sampling error of the estimate.

Confidence intervals for an estimate can be calculated directly from the Approximate Sampling Variability Tables by first determining from the appropriate table the coefficient of variation of the estimate $\hat{X}$, and then using the following formula to convert to a confidence interval ($CI_{\hat{X}}$):

$$CI_{\hat{X}} = (\hat{X} - t\hat{X}\alpha_{\hat{X}}, \hat{X} + t\hat{X}\alpha_{\hat{X}})$$

where $\alpha_{\hat{X}}$ is the determined coefficient of variation of $\hat{X}$, and

$t = 1$ if a 68% confidence interval is desired;
 $t = 1.6$ if a 90% confidence interval is desired;
 $t = 2$ if a 95% confidence interval is desired;
 $t = 2.6$ if a 99% confidence interval is desired.

Note: Release guidelines which apply to the estimate also apply to the confidence interval. For example, if the estimate is not releasable, then the confidence interval is not releasable either.

10.2.1 Example of Using the Coefficient of Variation Tables to Obtain Confidence Limits

A 95% confidence interval for the estimated proportion of households which did not have telephone service for their residence during the reference period because they could not afford telephone service (from Example 2, Section 10.1.1) would be calculated as follows:

$$\hat{X} = 56.7\% \text{ (or expressed as a proportion 0.567)}$$

$$t = 2$$

$\alpha_{\hat{X}} = 5.6\%$ (0.056 expressed as a proportion) is the coefficient of variation of this estimate as determined from the tables.

$$CI_{\hat{X}} = \{0.567 - (2)(0.567)(0.056), 0.567 + (2)(0.567)(0.056)\}$$

$$CI_{\hat{X}} = \{0.567 - 0.064, 0.567 + 0.064\}$$

$$CI_{\hat{X}} = \{0.503, 0.631\}$$

With 95% confidence it can be said that between 50.3% and 63.1% of households which did not have telephone service for their residence during the reference period could not afford telephone service.

10.3 How to Use the Coefficient of Variation Tables to Do a T-test

Standard errors may also be used to perform hypothesis testing, a procedure for distinguishing between population parameters using sample estimates. The sample estimates can be numbers, averages, percentages, ratios, etc. Tests may be performed at various levels of significance, where a level of significance is the probability of concluding that the characteristics are different when, in fact, they are identical.

Let $\hat{X}_1$ and $\hat{X}_2$ be sample estimates for two characteristics of interest. Let the standard error on the difference $\hat{X}_1 - \hat{X}_2$ be $\sigma_{\hat{d}}$.

If $t = \frac{\hat{X}_1 - \hat{X}_2}{\sigma_{\hat{d}}}$ is between -2 and 2, then no conclusion about the difference between the

characteristics is justified at the 5% level of significance. If however, this ratio is smaller than -2 or larger than +2, the observed difference is significant at the 0.05 level. That is to say that the difference between the estimates is significant.

10.3.1 Example of Using the Coefficient of Variation Tables to Do a T-test

Let us suppose that the user wishes to test, at 5% level of significance, the hypothesis that there is no difference between the proportion of households in Quebec which reported that their total annual income was less than the low income cut-off (LICO), and the proportion of households in Ontario which reported that their total annual income was less than LICO. From Example 3, Section 10.1.1, the standard error of the difference between these two estimates was found to be 0.0098. Hence,

$$t = \frac{\hat{X}_1 - \hat{X}_2}{\sigma_{\hat{d}}} = \frac{0.257 - 0.190}{0.0098} = \frac{0.067}{0.0098} = 6.84$$

Since $t = 6.84$ is greater than 2, it must be concluded that there is a significant difference between the two estimates at the 0.05 level of significance.

10.4 Coefficients of Variation for Quantitative Estimates

For quantitative estimates, special tables would have to be produced to determine their sampling error. Since all of the variables for the Residential Telephone Service Survey are primarily categorical in nature, this has not been done.

As a general rule, however, the coefficient of variation of a quantitative total will be larger than the coefficient of variation of the corresponding category estimate (i.e., the estimate of the number of persons contributing to the quantitative estimate). If the corresponding category estimate is not releasable, the quantitative estimate will not be either. For example, the coefficient of variation of the total number of weeks absent from work would be greater than the coefficient of variation of the corresponding proportion of paid workers with an absence. Hence, if the coefficient of variation of the proportion is unacceptable (making the proportion not releasable), then the coefficient of variation of the corresponding quantitative estimate will also be unacceptable (making the quantitative estimate not releasable).

Coefficients of variation of such estimates can be derived as required for a specific estimate using a technique known as pseudo replication. This involves dividing the records on the microdata files into subgroups (or replicates) and determining the variation in the estimate from replicate to replicate. Users wishing to derive coefficients of variation for quantitative estimates may contact Statistics Canada for advice on the allocation of records to appropriate replicates and the formulae to be used in these calculations.

10.5 Coefficient of Variation Tables

Refer to RTSS201012_CVTabE.pdf for the coefficient of variation tables.

10.6 Variance Estimation for Master Files

In order to determine the quality of the estimate and to calculate the coefficient of variation (CV), the standard deviation must be calculated. Confidence intervals also require the standard deviation of the estimate. The Residential Telephone Service Survey (RTSS) uses a multi-stage survey design and calibration, which means that there is no simple formula that can be used to calculate variance estimates. Therefore, an approximate method was needed, the bootstrap method. With the use of the bootstrap weights and the BOOTVAR program, discussed in the next section, CV's and other variance estimates can be derived with accuracy.

Bootstrap method for variance estimation

Independently, in each stratum, a simple random sample of $(n-1)$ of the n units in the sample is selected with replacement. Note that since the selection is with replacement, a unit may be chosen more than once. This step is repeated R times to form R bootstrap samples. An average initial bootstrap weight based on the R samples is calculated for each sample unit in the stratum. The entire process (selecting simple random samples, recalculating weights for each stratum) is repeated B times, where B is large, yielding B different initial bootstrap weights.

These weights are then adjusted according to the same weighting process as the regular weights: non-response adjustment, calibration and so on. The end result is B final mean bootstrap weights for each unit in the sample. The variation among the B possible estimates based on the B bootstrap weights are related to the variance of the estimator based on the regular weights and can be used to estimate it.

Statistical packages for variance estimation

Bootvar

Users should note that bootstrap weights are provided and should be used for variance estimation. BOOTVAR is a macro program that can be used to do the variance calculation using the bootstrap weights. The Bootvar program is available in SAS or SPSS format. It is made up of macros that compute variances for totals, ratios, differences between ratios and for linear and logistic regression.

Bootvar may be downloaded from Statistics Canada's Research Data Centre (RDC) website. Users must accept the Bootvar Click-Wrap Licence before they can read the files. There is a document on the site explaining how to adapt the system to meet users' needs.

SAS: http://www.statcan.gc.ca/rdc-cdr/bootvar_sas-eng.htm

SPSS: http://www.statcan.gc.ca/rdc-cdr/bootvar_spss-eng.htm

Other packages

Other than Bootvar, there are different commercial software packages that can carry out some design-based analysis for variance estimation; Stata 9 or 10, SUDAAN and WesVar.

These methods can be adapted for the RTSS from a paper by Owen Phillips "*Using bootstrap weights with Wes Var and SUDAAN*" (Catalogue no. 12-002-X20040027032) in *The Research Data Centres Information and Technical Bulletin, Chronological index*, Fall 2004, vol.1 no. 2 Statistics Canada, Catalogue no. 12-002-XIE.

11.0 Weighting

Since the Residential Telephone Service Survey (RTSS) used a sub-sample of the Labour Force Survey (LFS) sample, the derivation of weights for the survey records is clearly tied to the weighting procedure used for the LFS. The LFS weighting procedure is briefly described below.

11.1 Weighting Procedures for the Labour Force Survey

In the LFS, the final weight attached to each record is the product of the following factors: the basic weight, the cluster sub-weight, the stabilization weight, the balancing factor for non-response, and the province-age-sex and sub-provincial area ratio adjustment factor. Each is described below.

Basic Weight

In a probability sample, the sample design itself determines weights which must be used to produce unbiased estimates of the population. Each record must be weighted by the inverse of the probability of selecting the person to whom the record refers. In the example of a 2% simple random sample, this probability would be 0.02 for each person and the records must be weighted by $1 / 0.02 = 50$. Due to the complex LFS design, dwellings in different regions will have different basic weights. Because all eligible individuals in a dwelling are interviewed (directly or by proxy), this probability is essentially the same as the probability with which the dwelling is selected.

Cluster Sub-weight

The cluster delineation is such that the number of dwellings in the sample increases very slightly with moderate growth in the housing stock. Substantial growth can be tolerated in an isolated cluster before the additional sample represents a field collection problem. However, if growth takes place in more than one cluster in an interviewer assignment, the cumulative effect of all increases may create a workload problem. In clusters where substantial growth has taken place, sub-sampling is used as a means of keeping interviewer assignments manageable. The cluster sub-weight represents the inverse of this sub-sampling ratio in clusters where sub-sampling has occurred.

Stabilization Weight

Sample stabilization is also used to address problems with sample size growth. Cluster sub-sampling addressed isolated growth in relatively small areas whereas sample stabilization accommodates the slow sample growth over time that is the result of a fixed sampling rate along with a general increase in the size of the population. Sample stabilization is the random dropping of dwellings from the sample in order to maintain the sample size at its desired level. The basic weight is adjusted by the ratio of the sample size, based on the fixed sampling rate, to the desired sample size. This adjustment factor is known as the stabilization weight. The adjustment is done within stabilization areas defined as dwellings belonging to the same employment insurance economic region and the same rotation group.

Non-response

For certain types of non-response (i.e. household temporarily absent, refusal), data from a previous month's interview with the household if any, is brought forward and used as the current month's data for the household.

In other cases, non-response is compensated for by proportionally increasing the weights of responding households. The weight of each responding record is increased by the ratio of the number of households that should have been interviewed, divided by the number that were actually interviewed. This adjustment is done separately for non-response areas, which are defined by employment insurance economic region, type of area, and rotation group. It is based on the assumption that the households that have been interviewed represent the characteristics of those that should have been interviewed within a non-response area.

Labour Force Survey Sub-weight

The product of the previously described weighting factors is called the LFS sub-weight. All members of the same sampled dwelling have the same sub-weight.

Sub-provincial and Province-Age-Sex Adjustments

The sub-weight can be used to derive a valid estimate of any characteristic for which information is collected by the LFS. However, these estimates will be based on a frame that contains some information that may be several years out of date and therefore not representative of the current population. Through the use of more up-to-date auxiliary information about the target population, the sample weights are adjusted to improve both the precision of the estimates and the sample's representation of the current population.

Independent estimates are available monthly for various age and sex groups by province. These are population projections based on the most recent census data, records of births and deaths, and estimates of migration. In the final step, this auxiliary information is used to transform the sub-weight into the final weight. This is done using a calibration method. This method ensures that the final weights it produces sum to the census projections for the auxiliary variables, namely totals for various age-sex groups, economic regions, census metropolitan areas, rotation groups, household and economic family size. Weights are also adjusted so that estimates of the previous month's industry and labour status estimates derived from the present month's sample, sum up to the corresponding estimates from the previous month's sample. This is called composite estimation. The entire adjustment is applied using the generalized regression technique.

This final weight is normally not used in the weighting for a supplement to the LFS. Instead, it is the sub-weight which is used, as explained in the following paragraphs.

11.2 Weighting Procedures for the Residential Telephone Service Survey

The principles behind the calculation of the weights for the Residential Telephone Service Survey are nearly identical to those for the LFS. However, this survey is a household-weighted survey, not a person-weighted survey. Also, further adjustments are made to the LFS sub-weights in order to derive a final weight for the individual records on the RTSS microdata file.

- 1) An adjustment to account for the use of a two-sixths sub-sample for Ontario and Quebec and a three-sixths sub-sample for all the other provinces, instead of the full LFS sample.
- 2) An adjustment to account for the additional non-response to the supplementary survey, i.e., non-response to the RTSS for individuals who did respond to the LFS or for which the previous month's LFS data was brought forward. An analysis was undertaken to determine homogeneous non-response groups. The groups accounted for such things as household size, number of attempts to contact the household, whether the interview was done by proxy and other demographic and geographic variables.
- 3) The final adjustment ensured that estimates produced for a province-age-sex group would agree with the known population totals for that province-age-sex group.

Adjustment 1) is taken into account by multiplying the LFS sub-weight (W_s) for each responding RTSS record by:

$$\frac{6}{2} \times \frac{1}{W_s} \text{ or } \frac{6}{3} \times \frac{1}{W_s}$$

to obtain WEIGHT1.

Adjustment 2) is accounted for by a non-response adjustment factor for each non-response group. WEIGHT1 is multiplied by the non-response adjustment factor to obtain WEIGHT2.

Adjustment 3) is accounted for by a calibration factor for each RTSS respondent. WEIGHT2 is multiplied by the calibration factor to obtain the final RTSS weight. The final RTSS weight WTHM appears on the Residential Telephone Service Survey master file.