Microdata User Guide

Canadian Internet Use Survey

2005

Statistics Canada Statistique Canada

Canada

Table of Contents

1.0 Introduction

The Canadian Internet Use Survey (CIUS) was conducted by Statistics Canada for Science, Innovation and Electronic Information Division at Statistics Canada, in 2005. This manual has been produced to facilitate the manipulation of the microdata file of the survey results.

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

Statistics Canada

Larry McKeown
Science, Innovation and Electronic Information Division
150 Tunney's Pasture Driveway, Ottawa, Ontario K1A 0T6
Telephone: (613) 951-2582
Fax: (613) 951-0009
E-mail: Larry.McKeown@statcan.ca

2.0 Background

The Canadian Internet Use Survey (CIUS) was conducted for the first time in 2005, replacing the Household Internet Use Survey (HIUS). The HIUS had been conducted annually from 1997 to 2003 to measure household Internet use. As growth in the number of households using the Internet levels off, the survey was redesigned to focus on how individuals, rather than households, are using the Internet.

The shift from a household to individual survey brings the CIUS more in-line with international standards for Internet statistics and allows for more in-depth study of the use of Broadband Internet and how using the Internet may affect individuals and society both socially and economically.

3.0 Objectives

The objective of this survey is to measure the demand for and the use of the Internet by Canadians. This is measured by the intensity of use; the frequency of use and the speed of the Internet connection.

Other objectives of the survey are to measure:

- Canadians' access to and use of the Internet at home, in the workplace and in other locations (e.g., public libraries, schools, cafés);
- the types of services and information people access on the Internet from home (e.g., e-mail, electronic banking, education services, medical and health information);
- the ordering and purchasing of goods and services over the Internet (from any location) for personal or household consumption;
- how *window shopping* on the Internet has influenced our lives; and
- who does not use the Internet and why they do not use it. It is important to understand the characteristics of these individuals as it affects what and when government services will be available on-line.

The information collected will update and expand upon previous studies of Internet use done by Statistics Canada.

4.0 Concepts and Definitions

This chapter outlines concepts and definitions of interest to the users. The concepts and definitions used in the Labour Force Survey (LFS) are described in Section 4.1 while those specific to the Canadian Internet Use Survey (CIUS) are given in Section 4.2. Users are referred to Chapter 12.0 of this document for a copy of the actual survey questionnaire(s) used.

4.1 Labour Force Survey Concepts and Definitions

Labour Force Status
Designates the status of the respondent vis-à-vis the labour market: a member of the non-institutional population 15 years of age and over is either employed, unemployed or not in the labour force.

Employment
Employed persons are those who, during the reference week:

a) did any work[1] at all at a job or business; or
b) had a job but were not at work due to factors such as own illness or disability, personal or family responsibilities, vacation, labour dispute or other reasons (excluding persons on layoff, between casual jobs, and those with a job to start at a future date).

Unemployment
Unemployed persons are those who, during the reference week:

a) were on temporary layoff during the reference week with the expectation of recall and were available for work; or
b) were without work, had actively looked for work in the past four weeks, and were available for work[2]; or
c) had a new job to start within four weeks from the reference week, and were available for work.

Not in the Labour Force
Persons not in the labour force are those who, during the reference week, were unwilling or unable to offer or supply labour services under conditions existing in their labour markets, that is, they were neither employed nor unemployed.

[1] Work includes any work for pay or profit, that is, paid work in the context of an employer-employee relationship, or self-employment. It also includes unpaid family work, which is defined as unpaid work contributing directly to the operation of a farm, business or professional practice owned and operated by a related member of the same household. Such activities may include keeping books, selling products, waiting on tables, and so on. Tasks such as housework or maintenance of the home are not considered unpaid family work.

[2] Persons are regarded as available for work if they:
i) reported that they could have worked in the reference week if a suitable job had been offered; or if the reason they could not take a job was of a temporary nature such as: because of own illness or disability, personal or family responsibilities, because they already have a job to start in the near future, or because of vacation (prior to 1997, those on vacation were not considered available).
ii) were full-time students seeking part-time work who also met condition i) above. Full-time students currently attending school and looking for full-time work are not considered to be available for work during the reference week.

Industry and Occupation
The Labour Force Survey provides information about the occupation and industry attachment of employed and unemployed persons, and of persons not in the labour force who have held a job in the past 12 months. Since 1997, these statistics have been based on the North American Industry Classification System (NAICS) and the Standard Occupational Classification (SOC-91). Prior to 1997, the 1980 Standard Industrial Classification and the 1980 Standard Occupational Classification were used.

Reference Week
The entire calendar week (from Sunday to Saturday) covered by the Labour Force Survey each month. It is usually the week containing the 15th day of the month. The interviews are conducted during the following week, called the Survey Week, and the labour force status determined is that of the reference week.

Full-time Employment
Full-time employment consists of persons who usually work 30 hours or more per week at their main or only job.

Part-Time Employment
Part-time employment consists of persons who usually work less then 30 hours per week at their main or only job.

4.2 Canadian Internet Use Survey Concepts and Definitions

Users
Internet users are persons who accessed the Internet at least once in the last 12 months for personal reasons (non business uses).

Ever users
Internet ever users are persons who have used the Internet in the past but do not currently use it. An “ever user” did not use the Internet in the last 12 months but has used it at some point in the past.

Non-Users
Non-users are persons who have never used the Internet. In some sections, non-users include persons who do not use the Internet from home.

Uses
Uses of the Internet include: communicating with government, finding medical or health information, educational use, specific communication uses (social networks) and electronic commerce.

Window Shop
Persons are window shopping when they compare the characteristics and prices of products and/or services without necessarily buying them.

5.0 Survey Methodology

The Canadian Internet Use Survey (CIUS) was administered from November 13th to December 6th, 2005 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. The LFS design is briefly described in the Sections 5.1 to 5.4.[3] Sections 5.5 and 5.6 describe how the CIUS departed from the basic LFS design in 2005.

5.1 Population Coverage

The LFS is a monthly household survey of a sample of individuals who are representative of the civilian, non-institutionalized population 15 years of age or older in Canada's 10 provinces. Specifically excluded from the survey's coverage are residents of the Yukon, Northwest Territories and Nunavut, persons living on Indian Reserves, full-time members of the Canadian Armed Forces and inmates of institutions. These groups together represent an exclusion of approximately 2% of the population aged 15 or over.

5.2 Sample Design

The LFS has undergone an extensive redesign, culminating in the introduction of the new design at the end of 1994. The LFS sample is based upon a stratified, multi-stage design employing probability sampling at all stages of the design. The design principles are the same for each province. A diagram summarizing the design stages can be found in the document LFS_AppendixA.pdf.

5.2.1 Primary Stratification

Provinces are divided into economic regions (ER) and employment insurance economic regions (EIER). ERs are geographic areas of more or less homogeneous economic structure formed on the basis of federal-provincial agreements. They are relatively stable over time. EIERs are also geographic areas, and are roughly the same size and number as ERs, but they do not share the same definitions. Labour force estimates are produced for the EIERs for the use of Service Canada.

The intersections of the two types of regions form the first level of stratification for the LFS. These ER/EIER intersections are treated as primary strata and further stratification is carried out within them (see Section 5.2.3). Note that a third set of regions, census metropolitan areas (CMA), is also respected by stratification in the current LFS design, since each CMA is also an EIER.

5.2.2 Types of Areas

The primary strata (ER/EIER intersections) are further disaggregated into three types of areas: rural, urban and remote areas. Urban and rural areas are loosely based on the Census definitions of urban and rural, with some exceptions to allow for the formation of strata in some areas. Urban areas include the largest CMAs down to the smallest villages categorized by the 1991 Census as urban (1,000 people or more), while rural areas are made up of areas not designated as urban or remote.

All urban areas are further subdivided into two types: those using an apartment list frame and an area frame, as well as those using only an area frame.

[3] 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.

Approximately 1% of the LFS population is found in remote areas of provinces which are less accessible to LFS interviewers than other areas. For administrative purposes, this portion of the population is sampled separately through the remote area frame. Some populations, not congregated in places of 25 or more people, are excluded from the sampling frame.

5.2.3 Secondary Stratification

In urban areas with sufficiently large numbers of apartment buildings, the strata are subdivided into apartment frames and area frames. The apartment list frame is a register maintained for the 18 largest cities across Canada. The purpose of this is to ensure better representation of apartment dwellers in the sample as well as to minimize the effect of growth in clusters, due to construction of new apartment buildings. In the major cities, the apartment strata are further stratified into low income strata and regular strata.

Where it is possible and/or necessary, the urban area frame is further stratified into regular strata, high income strata, and low population density strata. Most urban areas fall into the regular urban strata, which, in fact, cover the majority of Canada's population. High income strata are found in major urban areas, while low density urban strata consist of small towns that are geographically scattered.

In rural areas, the population density can vary greatly from relatively high population density areas to low population density areas, resulting in the formation of strata that reflect these variations. The different stratification strategies for rural areas were based not only on concentration of population, but also on cost-efficiency and interviewer constraints.

In each province, remote settlements are sampled proportional to the number of dwellings in the settlement, with no further stratification taking place. Dwellings are selected using systematic sampling in each of the places sampled.

5.2.4 Cluster Delineation and Selection

Households in final strata are not selected directly. Instead, each stratum is divided into clusters, and then a sample of clusters is selected within the stratum. Dwellings are then sampled from selected clusters. Different methods are used to define the clusters, depending on the type of stratum.

Within each urban stratum in the urban area frame, a number of geographically contiguous groups of dwellings, or clusters, are formed based upon 1991 Census counts. These clusters are generally a set of one or more city blocks or block-faces. The selection of a sample of clusters (always six or a multiple of six clusters) from each of these secondary strata represents the first stage of sampling in most urban areas. In some other urban areas, census enumeration areas (EA) are used as clusters. In the low density urban strata, a three stage design is followed. Under this design, two towns within a stratum are sampled, and then 6 or 24 clusters within each town are sampled.

For urban apartment strata, instead of defining clusters, the apartment building is the primary sampling unit. Apartment buildings are sampled from the list frame with probability proportional to the number of units in each building.

Within each of the secondary strata in rural areas, where necessary, further stratification is carried out in order to reflect the differences among a number of socio-economic characteristics within each stratum. Within each rural stratum, six EAs or two or three groups of EAs are sampled as clusters.

5.2.5 Dwelling Selection

In all three types of areas (urban, rural and remote areas) selected clusters are first visited by enumerators in the field and a listing of all private dwellings in the cluster is prepared. From the listing, a sample of dwellings is then selected. The sample yield depends on the type of stratum. For example, in the urban area frame, sample yields are either six or eight dwellings, depending on the size of the city. In the urban apartment frame, each cluster yields five dwellings, while in the rural areas and EA parts of cities, each cluster yields 10 dwellings. In all clusters, dwellings are sampled systematically. This represents the final stage of sampling.

5.2.6 Person Selection

Demographic information is obtained for all persons in a household for whom the selected dwelling is the usual place of residence. LFS information is obtained for all civilian household members 15 years of age or older. Respondent burden is minimized for the elderly (age 70 and over) by carrying forward their responses for the initial interview to the subsequent five months in the survey.

5.3 Sample Size

The sample size of eligible persons in the LFS is determined so as to meet the statistical precision requirements for various labour force characteristics at the provincial and sub-provincial level, to meet the requirement of federal, provincial and municipal governments as well as a host of other data users.

The monthly LFS sample consists of approximately 60,000 dwellings. After excluding dwellings found to be vacant, dwellings demolished or converted to non-residential uses, dwellings containing only ineligible persons, dwellings under construction, and seasonal dwellings, about 54,000 dwellings remain which are occupied by one or more eligible persons. From these dwellings, LFS information is obtained for approximately 102,000 civilians aged 15 or over.

5.4 Sample Rotation

The LFS follows a rotating panel sample design, in which households remain in the sample for six consecutive months. The total sample consists of six representative sub-samples or panels, and each month a panel is replaced after completing its six month stay in the survey. Outgoing households are replaced by households in the same or a similar area. This results in a five-sixths month-to-month sample overlap, which makes the design efficient for estimating month-to-month changes. The rotation after six months prevents undue respondent burden for households that are selected for the survey.

Because of the rotation group feature, it is possible to readily conduct supplementary surveys using the LFS design but employing less than the full size sample.

5.5 Modifications to the Labour Force Survey Design for the Canadian Internet Use Survey

The CIUS used five of the six rotation groups in the November 2005 LFS sample. For the CIUS, the coverage of the LFS was modified to include all members of the household aged 18 and older. Unlike the LFS where information is collected for all eligible household members, the CIUS

only collected information from one randomly selected household member and proxy responses were not permitted.

5.6 Sample Size by Province for the Canadian Internet Use Survey

The following table shows the number of household members 18 years and older in the LFS sampled rotations who were eligible for the CIUS supplement. This table includes individuals in households which were non-respondents to the LFS.

Province	Sample Size
Newfoundland and Labrador	1,537
Prince Edward Island	1,127
Nova Scotia	2,282
New Brunswick	2,236
Quebec	7,996
Ontario	12,545
Manitoba	2,926
Saskatchewan	3,092
Alberta	4,257
British Columbia	5,082
Canada	**43,080**

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 Canadian Internet Use Survey

The Canadian Internet Use Survey (CIUS) was administered to one randomly selected individual per household. The random selection was carried out at the time of the interview.

Upon completion of the Labour Force Survey interview, the interviewer asked to speak to the selected person for the CIUS. If the selected person was not available, the interviewer arranged for a convenient time to phone back. Proxy response was not allowed, hence the collection period was extended by more than two weeks to allow the interviewers time to contact the pre-selected individuals.

6.5 Non-response to the Canadian Internet Use Survey

For households responding to the LFS, the next stage of data collection was to administer the CIUS. In total, 43,080 persons were eligible for the supplementary survey; the CIUS interview was completed for 30,466 of these persons for a response rate of 70.7%. More detailed information on response rates is presented in Chapter 8.0 (Data Quality).

7.0 Data Processing

The main output of the Canadian Internet Use Survey (CIUS) 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.

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

A few data items on the questionnaire were recorded by interviewers in an open-ended format. A total of 18 partially or completely open-ended questions were included in the survey. These were items relating to past Internet use, location of use, individual use, government on line, educational use, electronic commerce, non-use from home, ethnic origin and income.

7.4 Imputation

Imputation is the process that supplies valid values for those variables that have been identified for a change either because of invalid information or because of missing information. The new values are supplied in such a way as to preserve the underlying structure of the data and to ensure that the resulting records will pass all required edits. In other words, the objective is not to reproduce the true microdata values, but rather to establish internally consistent data records that yield good aggregate estimates.

We can distinguish between three types of non-response. Complete non-response is when the respondent does not provide the minimum set of answers. These records are dropped and accounted for in the weighting process (see Chapter 11.0). Item non-response is when the respondent does not provide an answer to one question, but goes on to the next question. These are usually handled using the "not stated" code or are imputed. Finally, partial non-response is when the respondent provides the minimum set of answers but does not finish the interview. These records can be handled like either complete non-response or multiple item non-response.

In the case of the CIUS, donor imputation was used to fill in missing data for item and partial non-response. Further information on the imputation process is given in Chapter 8.0 (Data Quality).

7.5 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. An example is the household income quintile variable, HINCQUIN (G_HQUINT in the Public Use Microdata File), which is constructed from income information collected during the interview. An imputation technique was used for records where the income variable was missing (see Section 8.2.4 for more details on the method used to impute income).

7.6 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 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 persons typically using the Internet from home is to be estimated, it is done by selecting the records referring to those individuals 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.

7.7 Suppression of Confidential Information

It should be noted that the "Public Use" Microdata Files (PUMF) may differ from the survey "master" files held by Statistics Canada. These differences usually are the result of actions taken to protect the anonymity of individual survey respondents. The most common actions are the suppression of file variables, grouping values into wider categories, and coding specific values into the "not stated" category. 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.

The survey master file includes the respondent's precise age, while the PUMF contains age groups only.

Where necessary, some of the text codes used in the open-ended questions are aggregated on the PUMF. They are regrouped to suit the major coding schemes.

For certain variables that are susceptible to identifying individuals, the PUMF may have been treated with local suppression, that is, some of the values in the master file may have been coded as "not stated" on the PUMF.

8.0 Data Quality

8.1 Response Rates

The following table summarizes the response rates to the Labour Force Survey (LFS) and to the Canadian Internet Use Survey (CIUS).

Province	LFS Selected Households	LFS Response Rate *	LFS Households Eligible for CIUS	CIUS Responding Individuals	CIUS Response Rate **	CIUS Overall Response Rate ***
		%			%	
Newfoundland and Labrador	1,592	90.5	1,537	1,021	66.4	64.1
Prince Edward Island	1,177	91.9	1,127	805	71.4	68.4
Nova Scotia	2,391	91.9	2,282	1,518	66.5	63.5
New Brunswick	2,321	92.0	2,236	1,517	67.8	65.4
Quebec	8,266	91.3	7,996	5,923	74.1	71.7
Ontario	12,968	91.7	12,545	8,653	69.0	66.7
Manitoba	3,020	92.1	2,926	2,094	71.6	69.3
Saskatchewan	3,181	92.6	3,092	2,252	72.8	70.8
Alberta	4,469	90.4	4,257	3,122	73.3	69.9
British Columbia	5,321	90.8	5,082	3,561	70.1	66.9
Canada	**44,706**	**91.5**	**43,080**	**30,466**	**70.7**	**68.1**

Note: The LFS counts are in terms of households while the CIUS counts are in terms of selected individuals within households (only one individual is selected per household). The CIUS overall response rate is based on the LFS household records in sub-sampled panels used by the CIUS. The LFS households eligible for the CIUS include LFS respondents from the current month as well as those carried forward from the previous month.

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

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

*** The CIUS overall response rate is the number of CIUS responding individuals as a percentage of the number of LFS selected households.

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 were taken 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 CIUS was a supplement to the LFS, the frame used was the LFS frame. Any non-response to the LFS had an impact on the CIUS frame. The quality of the sampling variables in the frame was very high. The CIUS sample consisted of five rotation groups from the LFS. The criteria used for the CIUS selection (like rotation group) were not missing for any LFS records.

Note that the LFS frame excludes about 2% of all households in the 10 provinces of Canada. Therefore, the CIUS 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.

Some variables on the sampling frame may play a critical role with respect to the software application used in the survey. For example, in a computer-assisted telephone interview (CATI) application, each record must have an accurate province code. Moreover, it requires accurate coding of the time zone field corresponding to province and each of the telephone number fields. Such analysis of the sampling frame provides important feedback on the quality of the frame used in the survey.

8.2.2 Data Collection

Interviewer training consisted of reading the CIUS Interviewer's Manual, practicing with the CIUS 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 the CIUS information after the LFS information was collected. The collection period ran from November 13th to December 6th, 2005.

8.2.3 Data Processing

During processing of the data, 32 CIUS 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.

Data processing of the CIUS was done in a number of steps including verification, coding, editing, imputation, estimation, confidentiality, etc. At each step a picture of the output files is taken and an easy verification can be made comparing files at the current and previous step. This greatly improved the data processing stage.

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 individuals who responded to the survey to compensate for those who 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. In order to provide complete data concerning the distribution of personal and household income among sampled units and concerning the calculation of totals for internet orders and expenditures, values were imputed when these were missing.

The imputations involved donors that were selected using a score function. For each item non-response or partial non-response records (also called recipient records), we compared certain characteristics to characteristics from all the donors. When the characteristics were the same between a donor and the recipient, a value was added to the score of that donor. The donor with the highest score was deemed the "closest" donor and was chosen to fill in missing pieces of information of the non-respondent. If there was more than one donor with the highest score, a random selection occurred. The pool of donors was made up in such a way that the imputed value assigned to the recipient, in conjunction with other non-imputed items from the recipient, would still pass the edits.

Imputation was done in two steps. First, imputation of personal and household incomes (imputed together whenever necessary, and then always from the same donor) and second, imputation of electronic commerce variables, as both types of imputations did not always make use of the same related auxiliary information. The following table shows the imputation rate for each of the variables.

	Step 1		Step 2		
	Personal Income	**Household Income**	**Electronic Commerce**	**Number of Orders**	**Value of Orders**
Imputed	8,670	10,816	159	216	297
Total	30,466	30,466	30,466	30,466	30,466
Rate (%)	28.5	35.5	0.5	0.7	1.0

The CIUS imputation process worked well and helped to fill incomplete responses with the experience of other respondents with similar or identical characteristics. This will add to the number of units used in any analysis performed by researchers.

Note that the public use microdata file does not contain any of the imputation flags or personal income variables. The impact of this is an additional layer of confidentiality.

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 survey results, one estimates that 28.6% of Canadians had never used the Internet from home, work, school or any other location (EV_Q01 = 2, No), and this estimate is found to have a standard error of 0.017. Then the coefficient of variation of the estimate is calculated as:

$$\left(\frac{0.017}{0.286} \right) X \ 100 \ \% \ = \ 5.9\%$$

There is more information on the calculation of coefficients 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, analyzing, 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 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 ratio) 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 Canadian Internet Use Survey (CIUS) 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 CIUS data can be tabulated and analyzed, 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 CIUS.

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 persons who have never used the Internet or the proportion of people who during the past 12 months have used the Internet at home for e-mail are examples of such estimates. An estimate of the number of persons possessing a certain characteristic may also be referred to as an estimate of an aggregate.

Examples of Categorical Questions:

Q: Have you ever used the Internet (E-mail or World Wide Web) from home, work, school, or any other location for personal non-business use?
R: Yes / No

Q: How often do you use the Internet at home in a typical month?
R: At least once a day / At least once a week (but not every day) / At least once a month (but not every week) / Less than once a month

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 persons in the surveyed population contributing to that total quantity.

An example of a quantitative estimate is the average number of orders for goods or services made by Canadians in 2005 over the Internet. The numerator is an estimate of the total number of orders placed and its denominator is the number of persons making at least one such order.

Examples of Quantitative Questions:

Q: During the past 12 months, how many separate orders for goods or services did you place over the Internet?
R: |_|_|_| Number of orders

Q: During the past 12 months, what was the estimated total cost, in Canadian dollars, of the goods and services you ordered over the Internet?
R: |_|_|_|_|_|_| Total cost rounded to nearest dollar value

9.3.3 Tabulation of Categorical Estimates

Estimates of the number of persons 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 orders for goods or services made by Canadians in 2005 over the Internet and paid for directly over the Internet with a credit or debit card, multiply the value reported in question EC_Q03 (number of orders for goods or services) by the final weight for the record, then sum this value over all records with EC_Q07 = 1 (paid directly over the Internet (with a credit or debit card)).

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 orders for goods or services made by Canadians in 2005 over the Internet and paid for directly over the Internet with a credit or debit card,

a) estimate the total number of orders ($\hat{X}$) as described above,
b) estimate the number of persons ($\hat{Y}$) in this category by summing the final weights of all records with EC_Q07 = 1, then
c) divide estimate a) by estimate b) ($\hat{X}/\hat{Y}$).

9.4 Guidelines for Statistical Analysis

The CIUS 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 may 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 male respondents is required. The steps to rescale the weights are as follows:

1) select all respondents from the file who reported CSEX = 1, men;

2) calculate the AVERAGE weight for these records by summing the original person weights from the microdata file for these records and then dividing by the number of respondents who reported CSEX = 1;

3) for each of these respondents, calculate a RESCALED weight equal to the original person weight divided by the AVERAGE weight;

4) perform the analysis for these respondents 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 estimates from the CIUS, 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 respondents who 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 rounded weighted 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-offs for the Canadian Internet Use 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 7,000 people 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 explain 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	24,900 & over	6,400 to < 24,900	under 6,400
Prince Edward Island	7,500 & over	1,900 to < 7,500	under 1,900
Nova Scotia	30,500 & over	7,700 to < 30,500	under 7,700
New Brunswick	27,200 & over	6,900 to < 27,200	under 6,900
Quebec	103,600 & over	25,800 to < 103,600	under 25,800
Ontario	135,400 & over	33,600 to < 135,400	under 33,600
Manitoba	29,900 & over	7,500 to < 29,900	under 7,500
Saskatchewan	21,200 & over	5,300 to < 21,200	under 5,300
Alberta	62,100 & over	15,500 to < 62,100	under 15,500
British Columbia	82,200 & over	20,600 to < 82,200	under 20,600
Atlantic Provinces	28,400 & over	7,100 to < 28,400	under 7,100
Manitoba and Saskatchewan	26,500 & over	6,600 to < 26,500	under 6,600
Prairie Provinces	50,900 & over	12,600 to < 50,900	under 12,600
Canada	**101,000 & over**	**24,900 to < 101,000**	**under 24,900**

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 Canadian Internet Use Survey (CIUS).

Province and Region	Design Effect	Sample Size	Population
Newfoundland and Labrador	1.81	1,021	408,815
Prince Edward Island	1.68	805	105,735
Nova Scotia	1.82	1,518	724,591
New Brunswick	2.03	1,517	583,011
Quebec	2.87	5,923	5,930,348
Ontario	3.36	8,653	9,640,686
Manitoba	2.10	2,094	843,239
Saskatchewan	1.90	2,252	707,607
Alberta	2.22	3,122	2,443,168
British Columbia	2.47	3,561	3,312,097
Atlantic Provinces	2.10	4,861	1,822,152
Manitoba and Saskatchewan	2.06	4,346	1,550,846
Prairie Provinces	2.63	7,468	3,994,014
Canada	**3.41**	**30,466**	**24,699,297**

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 Persons 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 Persons 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 people who have never used the Internet is more reliable than the estimated number of persons who have never used the Internet. (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 sex or age group), 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 $\left(\hat{d} = \hat{X}_1 - \hat{X}_2\right)$ is:

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

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 persons who have access to a computer or other device that could access the Internet at home and the numerator is the number of persons who have access to a computer or other device that could access the Internet at home but currently have no Internet connection at home.

In the case where the numerator is not a subset of the denominator, as for example, the ratio of the number of persons in Quebec who, during the past 12 months, used the Internet at home for electronic banking as compared to the number of persons in Ontario who, during the past 12 months, used the Internet at home for electronic banking, 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 $\left(\hat{R} = \hat{X}_1 / \hat{X}_2\right)$ 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 2005 Canadian Internet Use Survey are included to assist users in applying the foregoing rules.

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

Suppose that a user estimates that 7,064,905 Canadians have never used the Internet (EV_Q01 = 2, No). 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 (7,064,905) 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 7,000,000.

3) The coefficient of variation for an estimated aggregate is found by referring to the first non-asterisk entry on that row, namely, 1.7%.

Canadian Internet Use Survey, 2005

Approximate Sampling Variability Tables - Canada
ESTIMATED PERCENTAGE

NUMERATOR OF PERCENTAGE ('000)	0.1%	1.0%	2.0%	5.0%	10.0%	15.0%	…	30.0%	35.0%	40.0%	50.0%	70.0%	90.0%
1	166.1	165.3	164.5	162.0	157.6	153.2		139.0	134.0	128.7	117.5	91.0	52.5
2	117.4	116.9	116.3	114.5	111.5	108.3		98.3	94.7	91.0	83.1	64.4	37.2
3	95.9	95.5	95.0	93.5	91.0	88.4		80.3	77.3	74.3	67.8	52.5	30.3
4	83.0	82.7	82.2	81.0	78.8	76.6		69.5	67.0	64.4	58.7	45.5	26.3
5	74.3	73.9	73.6	72.4	70.5	68.5		62.2	59.9	57.6	52.5	40.7	23.5
.													
.													
.													
100	****	16.5	16.4	16.2	15.8	15.3		13.9	13.4	12.9	11.7	9.1	5.3
125	****	14.8	14.7	14.5	14.1	13.7		12.4	12.0	11.5	10.5	8.1	4.7
150	****	13.5	13.4	13.2	12.9	12.5		11.4	10.9	10.5	9.6	7.4	4.3
200	****	11.7	11.6	11.5	11.1	10.8		9.8	9.5	9.1	8.3	6.4	3.7
250	****	****	10.4	10.2	10.0	9.7		8.8	8.5	8.1	7.4	5.8	3.3
300	****	****	9.5	9.4	9.1	8.8		8.0	7.7	7.4	6.8	5.3	3.0
350	****	****	8.8	8.7	8.4	8.2		7.4	7.2	6.9	6.3	4.9	2.8
400	****	****	8.2	8.1	7.9	7.7		7.0	6.7	6.4	5.9	4.6	2.6
450	****	****	7.8	7.6	7.4	7.2		6.6	6.3	6.1	5.5	4.3	2.5
500	****	****	****	7.2	7.0	6.9		6.2	6.0	5.8	5.3	4.1	2.3
750	****	****	****	5.9	5.8	5.6		5.1	4.9	4.7	4.3	3.3	1.9
1,000	****	****	****	5.1	5.0	4.8		4.4	4.2	4.1	3.7	2.9	1.7
1,500	****	****	****	****	4.1	4.0		3.6	3.5	3.3	3.0	2.3	1.4
2,000	****	****	****	****	3.5	3.4		3.1	3.0	2.9	2.6	2.0	1.2
3,000	****	****	****	****	****	****		2.6	2.5	2.4	2.3	2.1	1.7
4,000	****	****	****	****	****	****		2.2	2.1	2.0	1.9	1.4	0.8
5,000	****	****	****	****	****	****		2.0	1.9	1.8	1.7	1.3	0.7
6,000	****	****	****	****	****	****		1.8	1.7	1.7	1.5	1.2	0.7
7,000	****	****	****	****	****	****		1.7	1.6	1.5	1.4	1.1	0.6
8,000	****	****	****	****	****	****		****	1.5	1.4	1.3	1.0	0.6
9,000	****	****	****	****	****	****		****	****	1.4	1.2	1.0	0.6
10,000	****	****	****	****	****	****		****	****	****	1.2	0.9	0.5
12,500	****	****	****	****	****	****		****	****	****	****	0.8	0.5
15,000	****	****	****	****	****	****		****	****	****	****	0.7	0.4
20,000	****	****	****	****	****	****		****	****	****	****	****	0.4

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

4) So the approximate coefficient of variation of the estimate is 1.7%. The finding that there were 7,064,905 Canadians (to be rounded according to the rounding guidelines in Section 9.1) who have never used the Internet is publishable with no qualifications.

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

Suppose that the user estimates that 2,224,386 / 7,064,905 = 31.5% of people who have never used the Internet (EV_Q01 = 2, No) have a computer or other device that could access the Internet at home (NU_Q03 = 1, Yes). 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., people who have never used the Internet), it is necessary to use both the percentage (31.5%) and the numerator portion of the percentage (2,224,386) in determining the coefficient of variation.

3) The numerator, 2,224,386, 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 2,000,000. Similarly, the percentage estimate does not appear as any of the column headings, so it is necessary to use the percentage closest to it, 30.0%.

4) The figure at the intersection of the row and column used, namely 3.1% is the coefficient of variation to be used.

5) So the approximate coefficient of variation of the estimate is 3.1%. The finding that 31.5% of people who have never used the Internet have a computer or other device that could access the Internet at home can be published with no qualifications.

Example 3: Estimates of Differences Between Aggregates or Percentages

Suppose that a user estimates that 2,722,655 / 5,930,348 = 45.9% of people in Quebec (PROVINCE = 24) used the Internet at home for e-mail during the past 12 months (SU_Q01 = 1, Yes), while 5,891,906 / 9,640,686 = 61.1% of people in Ontario (PROVINCE = 35) used the Internet at home for e-mail during the past 12 months (SU_Q01 = 1, Yes). 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 people in Quebec as 1.7%, and the CV of the estimate for people in Ontario as 1.4%.

Canadian Internet Use Survey, 2005

Approximate Sampling Variability Tables - Quebec
ESTIMATED PERCENTAGE

NUMERATOR OF PERCENTAGE ('000)	0.1%	1.0%	2.0%	5.0%	10.0%	15.0%	…	30.0%	35.0%	40.0%	50.0%	70.0%	90.0%
1	169.3	168.6	167.7	165.1	160.7	156.2		141.8	136.6	131.2	119.8	92.8	53.6
2	119.7	119.2	118.6	116.8	113.7	110.5		100.2	96.6	92.8	84.7	65.6	37.9
3	97.8	97.3	96.8	95.3	92.8	90.2		81.8	78.9	75.8	69.2	53.6	30.9
4	84.7	84.3	83.9	82.6	80.4	78.1		70.9	68.3	65.6	59.9	46.4	26.8
5	75.7	75.4	75.0	73.9	71.9	69.9		63.4	61.1	58.7	53.6	41.5	24.0
.													
.													
.													
60	****	****	21.7	21.3	20.8	20.2		18.3	17.6	16.9	15.5	12.0	6.9
65	****	****	20.8	20.5	19.9	19.4		17.6	16.9	16.3	14.9	11.5	6.6
70	****	****	20.0	19.7	19.2	18.7		16.9	16.3	15.7	14.3	11.1	6.4
75	****	****	19.4	19.1	18.6	18.0		16.4	15.8	15.2	13.8	10.7	6.2
80	****	****	18.8	18.5	18.0	17.5		15.8	15.3	14.7	13.4	10.4	6.0
85	****	****	18.2	17.9	17.4	16.9		15.4	14.8	14.2	13.0	10.1	5.8
90	****	****	17.7	17.4	16.9	16.5		14.9	14.4	13.8	12.6	9.8	5.6
95	****	****	17.2	16.9	16.5	16.0		14.5	14.0	13.5	12.3	9.5	5.5
100	****	****	16.8	16.5	16.1	15.6		14.2	13.7	13.1	12.0	9.3	5.4
125	****	****	****	14.8	14.4	14.0		12.7	12.2	11.7	10.7	8.3	4.8
150	****	****	****	13.5	13.1	12.8		11.6	11.2	10.7	9.8	7.6	4.4
200	****	****	****	11.7	11.4	11.0		10.0	9.7	9.3	8.5	6.6	3.8
250	****	****	****	10.4	10.2	9.9		9.0	8.6	8.3	7.6	5.9	3.4
300	****	****	****	****	9.3	9.0		8.2	7.9	7.6	6.9	5.4	3.1
350	****	****	****	****	8.6	8.3		7.6	7.3	7.0	6.4	5.0	2.9
400	****	****	****	****	8.0	7.8		7.1	6.8	6.6	6.0	4.6	2.7
450	****	****	****	****	7.6	7.4		6.7	6.4	6.2	5.6	4.4	2.5
500	****	****	****	****	7.2	7.0		6.3	6.1	5.9	5.4	4.2	2.4
750	****	****	****	****	****	5.7		5.2	5.0	4.8	4.4	3.4	2.0
1,000	****	****	****	****	****	****		4.5	4.3	4.2	3.8	2.9	1.7
1,500	****	****	****	****	****	****		3.7	3.5	3.4	3.1	2.4	1.4
2,000	****	****	****	****	****	****		****	3.1	2.9	2.7	2.1	1.2
3,000	****	****	****	****	****	****		****	****	****	****	1.7	1.0
4,000	****	****	****	****	****	****		****	****	****	****	1.5	0.8
5,000	****	****	****	****	****	****		****	****	****	****	****	0.8

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

Canadian Internet Use Survey, 2005

Approximate Sampling Variability Tables - Ontario
ESTIMATED PERCENTAGE

NUMERATOR OF PERCENTAGE ('000)	0.1%	1.0%	2.0%	5.0%	10.0%	15.0%	…	30.0%	35.0%	40.0%	50.0%	70.0%	90.0%
1	193.3	192.4	191.5	188.5	183.5	178.3		161.8	155.9	149.8	136.8	105.9	61.2
2	136.7	136.1	135.4	133.3	129.7	126.1		114.4	110.3	105.9	96.7	74.9	43.2
3	111.6	111.1	110.5	108.8	105.9	102.9		93.4	90.0	86.5	79.0	61.2	35.3
4	96.6	96.2	95.7	94.2	91.7	89.2		80.9	78.0	74.9	68.4	53.0	30.6
5	86.4	86.1	85.6	84.3	82.1	79.7		72.4	69.7	67.0	61.2	47.4	27.4
.													
.													
.													
60	****	24.8	24.7	24.3	23.7	23.0		20.9	20.1	19.3	17.7	13.7	7.9
65	****	23.9	23.7	23.4	22.8	22.1		20.1	19.3	18.6	17.0	13.1	7.6
70	****	23.0	22.9	22.5	21.9	21.3		19.3	18.6	17.9	16.3	12.7	7.3
75	****	22.2	22.1	21.8	21.2	20.6		18.7	18.0	17.3	15.8	12.2	7.1
80	****	21.5	21.4	21.1	20.5	19.9		18.1	17.4	16.7	15.3	11.8	6.8
85	****	20.9	20.8	20.4	19.9	19.3		17.6	16.9	16.2	14.8	11.5	6.6
90	****	20.3	20.2	19.9	19.3	18.8		17.1	16.4	15.8	14.4	11.2	6.4
95	****	19.7	19.6	19.3	18.8	18.3		16.6	16.0	15.4	14.0	10.9	6.3
100	****	****	19.1	18.8	18.3	17.8		16.2	15.6	15.0	13.7	10.6	6.1
125	****	****	17.1	16.9	16.4	15.9		14.5	13.9	13.4	12.2	9.5	5.5
150	****	****	15.6	15.4	15.0	14.6		13.2	12.7	12.2	11.2	8.6	5.0
200	****	****	****	13.3	13.0	12.6		11.4	11.0	10.6	9.7	7.5	4.3
250	****	****	****	11.9	11.6	11.3		10.2	9.9	9.5	8.6	6.7	3.9
300	****	****	****	10.9	10.6	10.3		9.3	9.0	8.6	7.9	6.1	3.5
350	****	****	****	10.1	9.8	9.5		8.6	8.3	8.0	7.3	5.7	3.3
400	****	****	****	9.4	9.2	8.9		8.1	7.8	7.5	6.8	5.3	3.1
450	****	****	****	8.9	8.6	8.4		7.6	7.4	7.1	6.4	5.0	2.9
500	****	****	****	****	8.2	8.0		7.2	7.0	6.7	6.1	4.7	2.7
750	****	****	****	****	6.7	6.5		5.9	5.7	5.5	5.0	3.9	2.2
1,000	****	****	****	****	****	5.6		5.1	4.9	4.7	4.3	3.3	1.9
1,500	****	****	****	****	****	****		4.2	4.0	3.9	3.5	2.7	1.6
2,000	****	****	****	****	****	****		3.6	3.5	3.3	3.1	2.4	1.4
3,000	****	****	****	****	****	****		****	2.8	2.7	2.5	1.9	1.1
4,000	****	****	****	****	****	****		****	****	****	2.2	1.7	1.0
5,000	****	****	****	****	****	****		****	****	****	****	1.5	0.9
6,000	****	****	****	****	****	****		****	****	****	****	1.4	0.8
7,000	****	****	****	****	****	****		****	****	****	****	****	0.7
8,000	****	****	****	****	****	****		****	****	****	****	****	0.7

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

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

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

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 $\hat{d} = 0.459 - 0.611 = -0.152$ is:

$$\begin{aligned}\sigma_{\hat{d}} &= \sqrt{[(0.459)(0.017)]^2 + [(0.611)(0.014)]^2} \\ &= \sqrt{(0.000061) + (0.000073)} \\ &= 0.012\end{aligned}$$

3) The coefficient of variation of $\hat{d}$ is given by $\sigma_{\hat{d}} / \hat{d} = 0.012 / 0.152 = 0.079$

4) So the approximate coefficient of variation of the difference between the estimates is 7.9%, which is publishable with no qualifications.

Example 4: Estimates of Ratios

Suppose that the user estimates that 2,722,655 people in Quebec used the Internet at home for e-mail during the past 12 months (SU_Q01 = 1, Yes), while 5,891,906 people in Ontario used the Internet at home for e-mail during the past 12 months (SU_Q01 = 1, Yes). The user is interested in comparing the estimate of people in Quebec versus those in Ontario 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 persons in Quebec who used the Internet at home for e-mail during the past 12 months. The denominator of the estimate ($\hat{X}_2$) is the number of persons in Ontario who used the Internet at home for e-mail during the past 12 months.

2) Refer to the coefficient of variation tables for QUEBEC and ONTARIO (see above).

3) The numerator of this ratio estimate is 2,722,655. The figure closest to it is 3,000,000. 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, 1.7%.

4) The denominator of this ratio estimate is 5,891,906. The figure closest to it is 6,000,000. 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, 1.4%

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:

$$\begin{aligned}\alpha_{\hat{R}} &= \sqrt{(0.017)^2 + (0.014)^2} \\ &= \sqrt{0.000289 + 0.000196} \\ &= 0.022\end{aligned}$$

6) The obtained ratio of people in Quebec versus people in Ontario who used the Internet at home for e-mail during the past 12 months is 2,722,655 / 5,891,906 which is 0.46:1 (to be rounded according to the rounding guidelines in Section 9.1). The coefficient of variation of this estimate is 2.2%, 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 difference 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 $\left(\hat{X} - k, \hat{X} + k\right)$ 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 ($CI_{\hat{x}}$):

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

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 Canadians who have never used the Internet and have a computer or other device that could access the Internet at home (from Example 2, Section 10.1.1) would be calculated as follows:

$\hat{X}$ = 31.5% (or expressed as a proportion 0.315)

t = 2

$\alpha_{\hat{x}}$ = 3.1% (0.031 expressed as a proportion) is the coefficient of variation of this estimate as determined from the tables.

$CI_{\hat{x}}$ = {0.315 – (2) (0.315) (0.031), 0.315 + (2) (0.315) (0.031)}

$CI_{\hat{x}}$ = {0.315 – 0.020, 0.315 + 0.020}

$CI_{\hat{x}}$ = {0.295, 0.335}

With 95% confidence it can be said that between 29.5% and 33.5% of Canadians who have never used the Internet have a computer or other device that could access the Internet at home.

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 $t = \dfrac{\hat{X}_1 - \hat{X}_2}{\sigma_{\hat{d}}}$ 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 people in Quebec who used the Internet at home for e-mail during the past 12 months and the proportion of people in

Ontario who used the Internet at home for e-mail during the past 12 months. From Example 3, Section 10.1.1, the standard error of the difference between these two estimates was found to be 0.012. Hence,

$$t = \frac{\hat{X}_1 - \hat{X}_2}{\sigma_{\hat{d}}} = \frac{0.459 - 0.611}{0.012} = \frac{-0.152}{0.012} = -12.7$$

Since t = -12.7 is less 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 most of the variables for the CIUS 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 orders for goods or services made by Canadians in 2005 over the Internet would be greater than the coefficient of variation of the corresponding proportion of Canadians who placed an order for goods or services. 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 CIUS2005_CVTabsE.pdf for the coefficient of variation tables.

11.0 Weighting

Since the Canadian Internet Use Survey (CIUS) 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 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 Canadian Internet Use Survey

The principles behind the calculation of the weights for the CIUS are identical to those for the LFS. However, further adjustments are made to the LFS sub-weights in order to derive a final weight for the individual records on the CIUS microdata file.

1) An adjustment to account for the use of a five-sixth sub-sample, instead of the full LFS sample

2) An adjustment to account for the random selection of one respondent from the selected household.

3) An adjustment to account for the additional non-response to the supplementary survey i.e., non-response to the CIUS for individuals who did respond to the LFS or for which previous month's LFS data was brought forward. The procedure is similar to the LFS non-response weight adjustment, but groupings are based on different variables.

4) A final adjustment is done to match the census projections for independent province-sex-age groups and census metropolitan area counts (in a calibration exercise).

The resulting weight WTPM is the final weight which appears on the CIUS microdata file.

12.0 Questionnaires

12.1 The Labour Force Survey Questionnaire

The Labour Force Survey questionnaire (LFS_QuestE.pdf) is used to collect information on the current and most recent labour market activity of all household members 15 years of age or older. It includes questions on hours of work, job tenure, type of work, reason for hours lost or absent, job search undertaken, availability for work, and school attendance.

12.2 The Canadian Internet Use Survey Questionnaire

The Canadian Internet Use Survey (CIUS) questionnaire was used in 2005 to collect the information for the supplementary survey. The file CIUS2005_QuestE.pdf contains the English questionnaire.

13.0 Record Layout with Univariate Frequencies

See CIUS2005_CdBk.pdf for the record layout with univariate counts.