Microdata User Guide

Canadian Internet Use Survey

2012

Statistics Canada Statistique Canada

Canada

Table of Contents

1.0 Introduction

The Canadian Internet Use Survey (CIUS) was conducted by the Investment, Science and Technology Division of Statistics Canada on behalf of Industry Canada in 2012. 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

Mark Uhrbach
Investment, Science and Technology Division
150 Tunney's Pasture Driveway
Ottawa, Ontario K1A 0T6
Telephone: 613-951-9214
Fax: 613-951-9920
E-mail: Mark.Uhrbach@statcan.gc.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 on a biennial basis from 1997 to 2003 to measure household Internet use. As growth in the number of households using the Internet leveled off, the survey was redesigned to focus on how individuals, rather than households, are using the Internet. The individual-level CIUS was conducted in 2005, 2007 and 2009.

For 2010, the CIUS was redesigned to meet the measurement needs of the *Broadband Canada: Connecting Rural Canadians Program*, sponsored by Industry Canada. For the first time, the redesigned survey incorporated a hybrid design, consisting of both a Household Component and an Individual Component.

The 2012 CIUS was conducted under the 2010 design.

The Household Component includes a short series of questions on the type of Internet connections and devices used by household members, from home, as well as availability of high speed service, and a standard module on household income. The questions may be answered by any knowledgeable member of the household.

Following the Household Component, an individual aged 16 years and older was randomly selected to complete the Individual Component. Respondents were asked about their use of the Internet, and online activities including electronic commerce. While the Household Component covered Internet access at home, the Individual Component covers use of the Internet from any location.

The Individual Component begins with a module on Current User (CU) of the Internet. As in past years, the CIUS asks individuals about their locations of Internet use, frequency and intensity of use, and reasons for non-use.

The Specific Use (SU) module, as in 2010, asks respondents to report their Internet activities. These activities cover a wide range of topics including the use of email, instant messaging, formal education, and the search for employment. As with the 2010 survey, these activities could have taken place from any location and using any device.

The Electronic Commerce (EC) module collects information on the total number, the total cost, and the types of goods and services ordered over the internet. Additional information is also collected on the location of delivery, the means of payment, and the main reason for not participating in e-commerce. This module remained consistent with the 2010 survey.

The Privacy and Security (PS) module includes questions about online behaviour (e.g., use of security software, frequency of backing up files) and experiences related to security (e.g., experienced a computer virus). For the 2012 survey, this module was slightly modified with the addition of questions regarding concern for security while using internet banking or credit cards online.

As the 2010 and the 2012 surveys have two distinct components – household and individual – with revised and streamlined questions, it is not appropriate to make direct comparisons with results from previous years. Data users who have questions about the survey are invited to contact the Investment, Science and Technology Division (please refer to Chapter 1.0 for contact information).

3.0 Objectives

The 2012 Canadian Internet Use Survey (CIUS) was designed to meet the measurement needs of the Broadband Canada: Connecting Rural Canadians Program, sponsored by Industry Canada. The Household Component was developed to provide information on the types of connections and devices used to access the Internet at home, as well as availability of high speed services. Because of the need to obtain statistically reliable estimates of Internet access in rural areas, households located outside of Statistics Canada Census Metropolitan Areas (CMA) and Census Agglomerations (CA) were oversampled (see also Section 5.5).

- The Individual Component measures rates of individual use of the Internet and a variety of online behaviours. Some of the objectives of the Individual Component are to measure: Canadians' access to and use of the Internet for personal reasons 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 any location (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;

- the intensity of individuals' use of the Internet, as measured by their frequency and hours of use from any location and device;

- behaviours and experiences related to online privacy and security;

- who does not use the Internet and why they do not use it.

The information collected is designed to provide policy-makers, researchers and the general public with information about Internet access rates for both households and individuals, the services individuals access online, reasons for non-use, and the characteristics of Internet users and non-users.

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.

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. The industry coding corresponds to the North American Industry Classification System 2007 (NAICS 2007). Occupation codes are based on the National Occupational Classification for Statistics 2006 (NOC-S 2006), January 1987 to present.

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 than 30 hours per week at their main or only job.

4.2 Canadian Internet Use Survey Concepts and Definitions

Access
In the CIUS 2012 Household Component, access refers to the household having an Internet connection at home, regardless of whether it is used.

Back up files
As a security measure, Internet users may back up their files electronically. This may protect files in case of an attack by a computer virus, a hard drive failure, etc. Only electronic copies of files are included (e.g., files the respondent places on a CD, DVD, memory stick or external hard drive, or stored on websites). Printed hard copies of documents are not included.

Blog
A public website where users post informal journals of their thoughts, comments, and philosophies, often updated frequently and normally reflecting the views of the blog's creator. Blogs provide commentary or news on a particular subject such as food, politics, or local news; some function as personal on-line diaries.

Browser history
Internet browsers may keep a history or record of websites visited by Internet users. This is done to improve the efficiency of the browsing experience (i.e. the next time a respondent visits the same page, information that has not changed since the last visit can be retrieved from storage instead of reloaded). It is also done for reasons of convenience – some users may wish to have their browsing history available in case they would like to quickly navigate to a page they had visited in the past. For security and privacy reasons, respondents may wish to delete their browsing history.

Cable line
Many Internet Service Providers offer services by cable. Coaxial cable that was traditionally used to provide television services can be upgraded to also facilitate Internet access. Over time, Internet by cable has become available in many (though not all) areas where cable television is available. For the purposes of the survey, cable connections are considered to be high-speed connections.

Computer virus
For the purposes of the survey, 'computer virus' is a broad term covering any type of program designed to compromise the security of one's computer system. These include traditional computer viruses but also other computer-related infections such as a Trojan Horse, worm, spyware or adware, etc. Not all viruses result in damage (particularly if they are detected by security software). If the respondent indicates that they have ever experienced a computer virus in question PS_Q09, then PS_Q10 asks if the virus (or viruses) resulted in the loss of information or damage to software.

Contribute content
Any activity where an Internet user submits content on the Internet for others to see, whether textual (e.g., stories, forum posts, letters, opinion); visual (e.g., images, artwork); or other content (e.g., music or other creative works); excludes e-mail and instant messaging.

Desktop computer
A personal computer (PC) intended for regular use at a single location, as opposed to a mobile laptop or portable computer.

Digital delivery
Some of the products listed in EC_Q02 (for example, software, MP3 music files, videos) may be accessed or downloaded from websites online rather than physically delivered to the home by some other method (such as by mail, courier, etc.). The survey seeks to measure if at least one of the items mentioned in EC_Q02 was digitally delivered over the Internet. If the respondent did not order any of the products listed in EC_Q02, they skip the digital delivery question and continue with the list of other products appearing in EC_Q04.

Discussion group
Also known as an “Internet forum” or “message board”, discussion groups are the modern equivalent of traditional bulletin boards. These are places where users can post and share information on a variety of topics. Some discussion groups and message boards specialize in particular topics.

E-commerce (electronic commerce)
For the purposes of the survey, e-commerce refers to ordering goods or services, for personal or household consumption, on the Internet. To count as ecommerce, the good or service should be ordered online (though not necessarily paid for online). Items that are ordered online, but paid by some other method (e.g., by telephone, cash on delivery (COD), etc.) are still considered to be e-commerce transactions in the CIUS.

More broadly speaking, Internet shopping includes e-commerce as well as window shopping (see definition of “window shopping” below).

Fixed wireless or point-to-point connection (requiring line of site reception)
This form of wireless Internet access, found mostly in rural areas, works by sending an Internet signal from a central distribution tower to a receiver, which may be located on the roof of a dwelling. In order for this service to work, homes receiving the signal must have clear “line-of-sight” reception from the transmission tower. That is, the technology requires a path unobstructed by hills or large trees in the vicinity.

High speed connection
For the purposes of the CIUS, all connection types other than telephone (e.g., cable, satellite, wireless, other) are considered to be high speed connections. In the CIUS 2012 Household Component, respondents with a telephone-only connection are asked whether their household accesses the Internet at home using a high speed connection. Generally, high speed includes services faster than traditional “dial-up” Internet access. The CIUS high speed data are self-reported.

Instant messenger
A common form of online communication where two or more persons exchange text to simulate a conversation. In addition to exchanging text, many instant messenger programs also allow users to share photos, or to display their Webcam.

Some cell phone users subscribe to text messaging services. Text messaging services allow individual to type or receive messages on their phone. However, as this technology relies on cellular networks rather than the Internet, it is excluded from the instant messenger category in the CIUS.

Laptop computer (including Netbooks and Tablet computers) (also known as portable computers)
In recent years, improvements in technology have made lighter, smaller laptops more affordable. Netbooks are very small versions of laptops designed for individuals looking for a lightweight computer that has a keyboard and small monitor, that they can use to surf the Web (particularly while travelling). Tablet computers (e.g., the iPad) often use a touch-screen so that users can navigate without the use of a keyboard (they may use their fingers or a stylus resembling a pen). Tablet computers are designed to lay flat on one's lap.

While becoming smaller in recent years, laptops, netbooks and tablet computers still have the look and feel of a computer. This differentiates them from wireless handheld devices (such as the BlackBerry, iPhone, iPod Touch and Palm Pre) which resemble more closely a cellular telephone or pocket-sized device with Internet access. In fact, many of these devices offer cellular phone service in addition to Internet access (see the definition of “wireless handheld device” below).

Misuse of personal information on the Internet
For the purposes of the survey, misuse of personal information includes any use or undesired sharing of the respondent's personal information over the Internet (e.g. photos, personal data such as their name, phone number, or other details) that the respondent deemed inappropriate.

Mobile Internet service for a BlackBerry, iPhone or other wireless handheld device (e.g. iPod Touch, Palm Pre)
A commercially available access service, typically provided for a fee, that allows users to browse the Internet and/or send and receive e-mail from their wireless handheld devices (see also entry under “wireless handheld device” below). It may also be referred to as a “data plan”, and may be billed monthly or on a per-usage basis.

MP3
A popular file format often used for music and other audio files.

Online games
Online games come in a variety of forms. Some individuals use a video games console with Internet access to play games with others online. Some social networking sites include online games that individuals can play with others. For the purposes of the survey, games played on the computer but not involving use of the Internet are excluded.

Online payment service
Some retailers offer payment using online payment services operated by third-parties. Rather than paying the retailer directly, these online payment services can be used to store the customer's funds or their credit card information, and money is transferred to the retailer from the customer via the customers' account. In doing so, these online payment services are designed to protect the customer's personal financial information (such as their credit card number) from being disclosed to the retailer.

Online telephone calls
These include telephone calls made over the Internet from a PC to a cellular phone, a landline telephone or another PC. Also known as VoIP (Voice over Internet Protocol); VoIP is the generic

name for the technology used to transmit voice conversations over a data network using the Internet Protocol.

Security software
For the purposes of the survey, security software is a general term that refers to any type of software designed to protect the security and privacy of a respondent's computer or other devices while browsing online. For example, the most common forms of security software include anti-virus software, anti-spyware software, spam filters, software firewalls etc. Any type of software that is designed to protect the security of the respondent's personal information online and their computer (and/or other devices used to access the Internet) is included.

Security updates
Most security software provides an option to users to download and install updates automatically at regular intervals, or to update manually. If the updates are done manually, users have the option to review the updates being installed (and to select which particular updates they would like to install) prior to installation.

Updates refer to security updates (for example, they may contain new virus definitions or updates to protect against new or emerging security threats). Updates do not refer to subscription renewals, such as annual or monthly fees required to use the software.

Social Networking Sites
Social Networking Sites are websites which host a variety of content that is shared between users with registered accounts. These sites are designed to allow users to share messages, photos, videos, ideas, activities, events, and interests within their individual networks.

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

Wi-Fi hotspot (including Internet or cyber café, or similar)
A place that offers Internet access over a wireless network. Hotspots are often found in coffeeshops, airports, and other public establishments. At a Wi-Fi hotspot, people usually bring their own device (such as a laptop) and can browse the Internet using the wireless access provided at the hotspot; at an Internet or cyber café, the device (such as a desktop or laptop computer) is usually provided in addition to Internet access.

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

Wireless connection (including handheld devices, sticks, or fixed wireless)
Wireless connections include mobile Internet services that allow someone to access the Internet over wireless handheld devices (see also entry under “wireless handheld device”). They may also include wireless sticks that can be plugged in to the USB or universal serial bus port of a laptop or are embedded inside a laptop to provide mobile access. Some users will have a wireless connection to their desktop computer in the home as well. Fixed wireless is another type of wireless connection used by some households particularly in rural areas (see “Fixed wireless or point-to-point connection”)

Respondents who indicate they have a wireless connection in HA_Q04 are asked a follow-up question (HA_Q05) asking them to indicate what type of wireless service(s) they use.

In some homes, users are connected to the Internet using some other type of connection (e.g., a “wired” or “wireline” connection such as cable), and use a wireless router within the home to distribute the connection to portable devices (such as laptops). In this case, the home is considered to be connected to the Internet via a cable connection and not by a wireless

connection, since the wireless router is used only to distribute the Internet service within the home (the Internet service is not delivered to the home using a wireless method).

Wireless handheld device
A portable (or “mobile”) electronic device that one can use to access the Internet. Examples include the BlackBerry, the iPhone, iPod Touch, Palm Pilot or Palm Pre. Typically, these devices are small enough to fit in the palm of one's hand (“handheld” device), and often can be used for cellular telephone services in addition to offering Internet services (e.g., Web browsing, e-mail). Only devices that can be used to access the Internet or e-mail are included. If the device is a cellular phone only, and offers no access to the Internet, it is not included.

Wireless router
A device used to distribute an Internet signal within a certain area (e.g., a home). A household may be connected the Internet by a cable line, but use a wireless router to distribute the signal to multiple devices used in the home (such as a desktop computer, laptop computers and other portable devices). As indicated in HA_Q04 and HA_Q05, for the purposes of the survey, a wireless router is not considered an Internet connection and is therefore excluded from responses to these questions.

5.0 Survey Methodology

The Canadian Internet Use Survey (CIUS) was administered in 2012 from October 14th to November 20th 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 CIUS departed from the basic LFS design in 2012.

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

The CIUS used three rotation groups from each of the LFS samples for the months of October and November 2012, for a total sample of six rotation groups. For two rotations each month, the CIUS Household component was first administered. Next, a household member aged 16 and older was randomly selected for the CIUS Individual component; proxy responses to the Individual component were not permitted. To produce sufficient sample for key indicators, rural households were oversampled. The Household component was administered to households in rural clusters in the third LFS rotation group each month. The CIUS target population covers the ten provinces, but excludes the territories.

5.2 Sample Size by Province for the Canadian Internet Use Survey

The following table shows the number of LFS households that were eligible for the CIUS supplement. This table includes households which were non-respondents to the CIUS.

Province	Sample Size for the Household Component	Sample Size for the Individual Component
Newfoundland and Labrador	1,583	1,216
Prince Edward Island	1,067	844
Nova Scotia	2,141	1,762
New Brunswick	2,074	1,652
Quebec	6,983	6,018
Ontario	9,742	8,961
Manitoba	3,375	2,882
Saskatchewan	2,854	2,343
Alberta	4,087	3,392
British Columbia	4,133	3,702
Canada	**38,039**	**32,772**

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 in urban areas are contacted by telephone if the telephone number is available from administrative files, otherwise the dwelling is 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.

Most LFS interviews are conducted using CATI, but some are conducted using CAPI. To reduce collection costs, no CAPI interviews were conducted for the CIUS 2012.

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, 38,039 households were eligible for the CIUS Household component; the CIUS interview was completed for 30,817 of these households for a collection response rate of 81.0%. As well, 32,772 individuals were eligible for the CIUS Individual component; the CIUS interview was completed for 22,615 of these individuals for a collection response rate of 69.0%. 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

Some data items on the questionnaire were recorded by interviewers in an open-ended format. A total of 12 partially or completely open-ended questions were included in the survey. These were items relating to Internet connection types and devices, locations of use, electronic commerce, household income, and non-use.

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 10.0). 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 on the Public Use Microdata Files) 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 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 10.0.

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 data items and grouping values into wider categories. 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.

The survey master file includes geographic identifiers for the 10 provinces as well as for the Employment Insurance Economic Regions, Census Divisions and Census Subdivisions, and other identifiers at the sub-provincial level. The PUMF includes geographic identifiers for the 10 provinces as well as identifiers of urban or rural status in every province except Prince Edward Island. Additionally, the PUMF includes identifiers for each of Montreal, Toronto and Vancouver.

The survey master file includes the respondent's precise age while the PUMF contains age groups only. Detailed industry and occupation, as well as several other detailed variables, are only available on the survey master file.

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.

8.0 Data Quality

8.1 Response Rates

The following table provides the collection response rate for the Household component of the Canadian Internet Use Survey (CIUS). This is the number of CIUS responding households as a percentage of the number of Labour Force Survey (LFS) households eligible for the Household component. The LFS households eligible for the CIUS include LFS respondents from the current month as well as those carried forward from the previous month.

Province	LFS Households Eligible for the CIUS Household Component (CATI only)	CIUS Respondents to the Household Component	CIUS Collection Response Rate for the Household Component (%)
Newfoundland and Labrador	1,583	1,390	87.8
Prince Edward Island	1,067	927	86.9
Nova Scotia	2,141	1,927	90.0
New Brunswick	2,074	1,833	88.4
Quebec	6,983	5,124	73.4
Ontario	9,742	7,586	77.9
Manitoba	3,375	3,033	89.9
Saskatchewan	2,854	2,509	87.9
Alberta	4,087	3,283	80.3
British Columbia	4,133	3,205	77.5
Canada	**38,039**	**30,817**	**81.0**

The following table provides the estimation response rate for the Household component of the CIUS. It is the number of CIUS responding households as a percentage of the number of LFS selected households. The estimation response rate is lower than the collection response rate because it takes into account the LFS non-respondents and the LFS computer-assisted personal interview (CAPI) cases which were excluded from the CIUS collection, in addition to the CIUS non-respondents.

Province	LFS Selected Households*	LFS Response Rate (%)	LFS CAPI Cases Excluded from CIUS (%)	CIUS Respondents to the Household Component	CIUS Estimation Response Rate for the Household Component (%)
Newfoundland and Labrador	1,751	89.8	7	1,390	79.4
Prince Edward Island	1,252	89.0	11	927	74.0
Nova Scotia	2,437	91.1	8	1,927	79.1
New Brunswick	2,440	90.0	11	1,833	75.1
Quebec	7,824	91.2	8	5,124	65.5
Ontario	11,502	89.0	12	7,586	66.0
Manitoba	3,843	90.1	9	3,033	78.9
Saskatchewan	3,226	90.2	9	2,509	77.8
Alberta	4,456	89.1	6	3,283	73.7
British Columbia	4,850	87.6	11	3,205	66.1
Canada	**43,581**	**89.7**	**10**	**30,817**	**70.7**

* The LFS selected households count is based on the rotation groups and rural clusters used by the CIUS for the Household component

The following table provides the collection response rate for the Individual component of the Canadian Internet Use Survey (CIUS). This is the number of CIUS responding individuals as a percentage of the number of Labour Force Survey (LFS) households eligible for the Individual component. The LFS households eligible for the CIUS include LFS respondents from the current month as well as those carried forward from the previous month.

Province	LFS Households Eligible for the CIUS Individual Component (CATI only)	CIUS Respondents to the Individual Component	CIUS Collection Response Rate for the Individual Component (%)
Newfoundland and Labrador	1,216	867	71.3
Prince Edward Island	844	605	71.7
Nova Scotia	1,762	1,334	75.7
New Brunswick	1,652	1,183	71.6
Quebec	6,018	4,052	67.3
Ontario	8,961	5,808	64.8
Manitoba	2,882	2,273	78.9
Saskatchewan	2,343	1,820	77.7
Alberta	3,392	2,263	66.7
British Columbia	3,702	2,410	65.1
Canada	**32,772**	**22,615**	**69.0**

The following table provides the estimation response rate for the Individual component of the CIUS. It is the number of CIUS responding individuals as a percentage of the number of LFS selected households. The estimation response rate is lower than the collection response rate because it takes into account the LFS non-respondents and the LFS computer-assisted personal interview (CAPI) cases which were excluded from the CIUS collection, in addition to the CIUS non-respondents.

Province	LFS Selected Households*	LFS Response Rate (%)	LFS CAPI Cases Excluded from CIUS (%)	CIUS Respondents to the Individual Component	CIUS Estimation Response Rate for the Individual Component (%)
Newfoundland and Labrador	1,350	89.3	7	867	64.2
Prince Edward Island	995	88.3	11	605	60.8
Nova Scotia	2,005	90.5	8	1,334	66.5
New Brunswick	1,919	89.9	10	1,183	61.6
Quebec	6,737	90.9	8	4,052	60.1
Ontario	10,557	88.8	12	5,808	55.0
Manitoba	3,276	89.5	8	2,273	69.4
Saskatchewan	2,651	90.0	9	1,820	68.7
Alberta	3,708	88.2	6	2,263	61.0
British Columbia	4,331	87.2	11	2,410	55.6
Canada	**37,529**	**89.2**	**9**	**22,615**	**60.3**

* The LFS selected households count is based on the four sub-sampled rotation groups used by the CIUS for the Individual component

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 six 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 October 14th to November 20th, 2012.

8.2.3 Data Processing

Early in the processing of the data, CIUS records were matched to corresponding records in the LFS. The CIUS records that did not match to corresponding records in the LFS 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), certain characteristics were compared 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 household income, and second, imputation of electronic commerce variables. The following table shows the imputation rate for each of the variables where applicable.

	Step 1	Step 2		
	Household Income	Electronic Commerce	Number of Orders	Value of Orders
Imputed	10,343	133	745	920
Total	30,817	17,610	9,384	9,384
Rate (%)	33.6	0.7	7.9	9.8

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. The impact of this is an additional layer of confidentiality.

8.2.5 Invalid Responses for Connection Type

In addition to gathering information on household connection types, the 2010 CIUS collected supplemental information on the name of the respondents' Internet Service Providers (ISP), the estimated monthly cost of connections, and devices used to access the Internet at home. This supplemental information was used by subject matter analysts to validate responses to questions on connection type. Subject matter analysts also used other sources, such as information gathered from suppliers of Internet services from the Annual Survey of Telecommunications (Integrated Metadatabase (IMDB) record number 2722), as well as supplier websites, to perform data confrontations.

Where respondents provided connection type information that was inconsistent with the services offered by their provider, the original response to connection type was retained but invalid responses were identified using a flag variable (variable name "CONNFLAG"). The flag was developed for respondents who mentioned a unique connection type in HA_Q04 (cable or telephone) that was not part of the available service offerings of the unique ISP name provided in HA_Q07 at the time of collection. Although the original responses could not be imputed, the flag variable was created so that analysts have the option to exclude invalid responses from their analyses. Approximately 4.5% of all households (or 5.8% of those with access at home) are identified by the flag variable. Although the 2010 CIUS was also designed to exclude wireless routers as a valid response option for type of connection (refer to the questionnaire and Section 4.2 above), a small number of households also supplied invalid data by reporting wireless routers as a connection type. Subject matter analysts compared these invalid responses with other collected information on connections and devices used by the household, as well as the ISP name and estimated monthly cost of service. In some cases, there was sufficient information to impute the connection type; for those cases where there was insufficient information to impute, the item was coded as "Not stated". The imputation for invalid responses of wireless router resulted in treatment for the following proportion of respondents in the CIUS 2010 Household Component:

- 0.3% of all households for HA_Q04 (or 0.4% of those reporting a connection type);
- 0.2% of all households for HA_Q05 (or 1.6% of those reporting a wireless connection type); and
- 0.2% of all households for HA_Q06 (or 0.9% of those households reporting on the speed of their connection).

Note that the public use microdata file does not contain the information on the ISP of the respondents, estimated monthly cost of Internet connections, nor the flag variable "CONNFLAG". The impact of this is an additional layer of confidentiality.

8.2.6 Data Quality for Wireless Connection Type

Data users should note that the proportion of households reporting access to the Internet at home using a mobile service for a BlackBerry, iPhone or other wireless handheld device in question HA_Q05 is significantly lower than the proportion of all households who indicated that they access the Internet at home using such a device (see question HA_Q03). The difference in rates is due in part to the fact that some households access the Internet at home over a handheld device using another connection (for example, a

home telephone or cable connection connected to a wireless router); some households will in fact have an Internet service (or "data plan") for their handheld device, but only use this service when accessing the Internet from locations other than home. Data users are reminded that questions on Internet access in the Household Component refer only to connections used by household members from home. A mobile Internet service may not be used by all household members, and is often not the only connection present in the household. Some respondents may not consider such services as a means by which their household is connected to the Internet at home.

In addition, a relatively large proportion of households with a wireless connection indicated that they did not know the type of wireless connection used in their household. Survey testing also revealed that many respondents were unfamiliar with fixed wireless (or "point-to-point") Internet services (HA_Q05). Due to the high incidence of "Don't know" responses, it is recommended that results for the wireless connection type questions be used with caution, and that responses of "Don't know" be included whenever results are analyzed and disseminated.

8.2.7 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 2010 survey results, one estimates that 21.1% of households did not access the Internet at home (HA_Q01 = 2, No), and this estimate is found to have a standard error of 0.00328. Then the coefficient of variation of the estimate is calculated as:

$$\left(\frac{0.00328}{0.211}\right) X\ 100\% = 1.6\%$$

There is more information on the calculation of coefficients of variation in Appendix A and B.

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 did not use the Internet in the last 12 months or the proportion of people who used the Internet 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: Did you use the Internet during the past 12 months for personal use?
R: Yes / No

Q: How often do you use the Internet for personal use 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 during the past 12 months 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 did you place over the Internet?
R: |_|_|_| Number of transactions, not articles purchased

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 during the past 12 months over the Internet and paid for directly over the Internet with a credit card, multiply the value reported in question EC_Q06 (number of orders for goods or services) by the final weight for the record, then sum this value over all records with EC_Q10A = 1 (paid directly over the Internet (with a credit 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 during the past 12 months over the Internet and paid for directly over the Internet with a credit 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_Q10A = 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.

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 SEX = 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 SEX = 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 tables provide an indication of the precision of population estimates generated from the household and individual files as they show 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 Appendix B.

For example, the first table shows that the quality of a weighted estimate of 1,800 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 and Example 4 in Appendix B explain the correct procedure to be used for ratios.

Release cut-offs associated with estimates of population totals from the household file

Province and Region	Acceptable CV 0.0% to 16.5%		Marginal CV 16.6% to 33.3%			Unacceptable CV > 33.3%	
Newfoundland and Labrador	8,800	& over	2,200	to <	8,800	under	2,200
Prince Edward Island	3,100	& over	800	to <	3,100	under	800
Nova Scotia	8,500	& over	2,100	to <	8,500	under	2,100
New Brunswick	8,100	& over	2,000	to <	8,100	under	2,000
Quebec	50,400	& over	12,500	to <	50,400	under	12,500
Ontario	49,300	& over	12,200	to <	49,300	under	12,200
Manitoba	8,000	& over	2,000	to <	8,000	under	2,000
Saskatchewan	7,700	& over	1,900	to <	7,700	under	1,900
Alberta	30,700	& over	7,700	to <	30,700	under	7,700
British Columbia	35,200	& over	8,800	to <	35,200	under	8,800
Atlantic Provinces	8,300	& over	2,100	to <	8,300	under	2,100
Manitoba and Saskatchewan	7,900	& over	1,900	to <	7,900	under	1,900
Prairie Provinces	21,900	& over	5,400	to <	21,900	under	5,400
Canada	**38,500**	**& over**	**9,500**	**to <**	**38,500**	**under**	**9,500**

Release cut-offs associated with estimates of population totals from the individual file

Province and Region	Acceptable CV 0.0% to 16.5%		Marginal CV 16.6% to 33.3%			Unacceptable CV > 33.3%	
Newfoundland and Labrador	25,200	& over	6,500	to <	25,200	under	6,500
Prince Edward Island	11,400	& over	3,000	to <	11,400	under	3,000
Nova Scotia	29,600	& over	7,500	to <	29,600	under	7,500
New Brunswick	27,300	& over	6,900	to <	27,300	under	6,900
Quebec	121,600	& over	30,300	to <	121,600	under	30,300
Ontario	177,600	& over	44,100	to <	177,600	under	44,100
Manitoba	25,300	& over	6,300	to <	25,300	under	6,300
Saskatchewan	22,800	& over	5,700	to <	22,800	under	5,700
Alberta	95,800	& over	24,100	to <	95,800	under	24,100
British Columbia	126,600	& over	31,900	to <	126,600	under	31,900
Atlantic Provinces	27,400	& over	6,800	to <	27,400	under	6,800
Manitoba and Saskatchewan	24,200	& over	6,000	to <	24,200	under	6,000
Prairie Provinces	70,400	& over	17,500	to <	70,400	under	17,500
Canada	**128,800**	**& over**	**31,700**	**to <**	**128,800**	**under**	**31,700**

10.0 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.

10.1 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.

10.2 Weighting Procedures for the Canadian Internet Use Survey Household File

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 household weight for the individual records on the CIUS household microdata file.

1) An adjustment to account for the use of a sub-sample, instead of the full LFS sample
2) An adjustment to account for the non-respondents who did not participate in the CIUS but who did in the LFS. All units selected for the CIUS were modelled using a logistic regression to calculate their propensity to respond. This probability was used to group records into weighting classes. The inverse of the observed response rate in each class was used as the adjustment factor. This adjustment was carried out in two stages: a first step for the computer-assisted personal interview (CAPI) respondents to the LFS who were not interviewed for the CIUS (and in which the propensity to respond via CAPI rather than the propensity to respond was modelled), and a second step for the household-level non-respondents.
3) A final adjustment is done to match the census projections for independent province-age groups, census metropolitan area and household size counts (in a calibration exercise).

The resulting weight WTHM is the final weight which appears on the CIUS household master file as well as WTHP on the household public use microdata file.

10.3 Weighting Procedures for the Canadian Internet Use Survey Person File

Adjustments are also made to the LFS sub-weights in order to derive a final person weight for the individual records on the CIUS person microdata file.

1) An adjustment to account for the use of a two thirds 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 non-respondents who did not participate in the CIUS but who did in the LFS. All units selected for the CIUS were modelled using a logistic regression to calculate their propensity to respond. This probability was used to group records into weighting classes. The inverse of the observed response rate in each class was used as the adjustment factor. This adjustment was carried out in three stages: a first step for the computer-assisted personal interview (CAPI) respondents to the LFS who were not interviewed for the CIUS (and in which the propensity to respond via CAPI rather than the propensity to respond was modelled), a second step for the household-level non-respondents where it was not known which member was selected to the CIUS, and a third step for person-level non-respondents where more detailed demographic information could be used in the modelling.
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 person master file as well as WTPP on the person public use microdata file.

11.0 Questionnaires

11.1 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.

11.2 The Canadian Internet Use Survey Questionnaire

The Canadian Internet Use Survey (CIUS) questionnaire was used to collect the information for the supplementary survey. The files CIUS2012_PR_QuestE.pdf and CIUS2012_HH_QuestE.pdf contains the English questionnaires for the Individual and Household Components which are also located on the Integrated Meta Data Base (IMDB) found here: http://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getInstrumentLink&SurvItem_Id=66020&Query_Id=130941&Query=instance&lang=en&db=imdb&adm=8&dis=2.

12.0 Record Layout with Univariate Frequencies

See CIUS2012_PUMFH_CdBk.pdf and CIUS2012_PUMFP_CdBk.pdf for the record layouts with univariate counts.

13.0 Coefficient of Variation Tables

Refer to CIUS2012_CVTabsE_H.pdf and CIUS2012_CVTabsE_P.pdf for the coefficient of variation tables.

Appendix A – Variance Estimation for Master and Share 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 Canadian Internet Use Survey (CIUS) 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.

The bootstrap parameters for CIUS are R = 1 and B = 1,000.

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 CIUS 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.

Appendix B – Variance Estimation for Public Use Microdata Files

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.

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/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 Persons/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 people who have used the Internet during the past 12 months for personal use is more reliable than the estimated number of persons who have used the Internet in the past 12 months for personal use. (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 used the Internet during the past 12 months for personal use and the numerator is the number of persons who used the Internet during the past 12 months for personal use from 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 for electronic banking as compared to the number of persons in Ontario who, during the past 12 months, used the Internet 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.

Examples of Using the Coefficient of Variation Tables for Categorical Estimates

The following examples are based on the 2010 Canadian Internet Use Survey and are included to assist users in applying the foregoing rules. Please note that the data for these examples are only to be used as a guide.

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

Suppose that a user estimates that 5,393,631 Canadians did not use the Internet during the past 12 months for personal use (CU_Q01 = 2, No). How does the user determine the coefficient of variation of this estimate?

1) Refer to the person coefficient of variation table for Canada.

2) The estimated aggregate (5,393,631) 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 5,000,000.

3) The coefficient of variation for an estimated aggregate is found by referring to the first non-asterisk entry on that row, namely, 2.4%.

Canadian Internet Use Survey, 2010 - Person File

Approximate Sampling Variability Tables - Canada
ESTIMATED PERCENTAGE

NUMERATOR OF PERCENTAGE ('000)	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	187.9	187.1	186.1	183.3	178.4	173.4	168.2		151.6	145.7	133	103	59.5
2	132.9	132.3	131.6	129.6	126.1	122.6	118.9		107.2	103	94	72.8	42
3	108.5	108	107.5	105.8	103	100.1	97.1		87.5	84.1	76.8	59.5	34.3
4	94	93.5	93.1	91.6	89.2	86.7	84.1		75.8	72.8	66.5	51.5	29.7
5	84.1	83.7	83.2	82	79.8	77.5	75.2		67.8	65.1	59.5	46.1	26.6
.													
.													
.													
100	*******	18.7	18.6	18.3	17.8	17.3	16.8		15.2	14.6	13.3	10.3	5.9
125	*******	16.7	16.6	16.4	16	15.5	15		13.6	13	11.9	9.2	5.3
150	*******	15.3	15.2	15	14.6	14.2	13.7		12.4	11.9	10.9	8.4	4.9
200	*******	13.2	13.2	13	12.6	12.3	11.9		10.7	10.3	9.4	7.3	4.2
250	*******	11.8	11.8	11.6	11.3	11	10.6		9.6	9.2	8.4	6.5	3.8
300	*******	********	10.7	10.6	10.3	10	9.7		8.8	8.4	7.7	5.9	3.4
350	*******	********	9.9	9.8	9.5	9.3	9		8.1	7.8	7.1	5.5	3.2
400	*******	********	9.3	9.2	8.9	8.7	8.4		7.6	7.3	6.6	5.1	3
450	*******	********	8.8	8.6	8.4	8.2	7.9		7.1	6.9	6.3	4.9	2.8
500	*******	********	8.3	8.2	8	7.8	7.5		6.8	6.5	5.9	4.6	2.7
750	*******	********	********	6.7	6.5	6.3	6.1		5.5	5.3	4.9	3.8	2.2
1,000	*******	********	********	5.8	5.6	5.5	5.3		4.8	4.6	4.2	3.3	1.9
1,500	*******	********	********	********	4.6	4.5	4.3		3.9	3.8	3.4	2.7	1.5
2,000	*******	********	********	********	4	3.9	3.8		3.4	3.3	3	2.3	1.3
3,000	*******	********	********	********	*******	3.2	3.1		2.8	2.7	2.4	1.9	1.1
4,000	*******	********	********	********	*******	2.7	2.7		2.4	2.3	2.1	1.6	0.9
5,000	*******	********	********	********	*******	********	2.4		2.1	2.1	1.9	1.5	0.8
6,000	*******	********	********	********	*******	*********	********		2	1.9	1.7	1.3	0.8
7,000	*******	********	********	********	*******	*********	********		1.8	1.7	1.6	1.2	0.7
8,000	*******	********	********	********	*******	*********	********		1.7	1.6	1.5	1.2	0.7
9,000	*******	********	********	********	*******	*********	********		1.6	1.5	1.4	1.1	0.6
10,000	*******	********	********	********	*******	*********	********		******	1.5	1.3	1	0.6
12,500	*******	********	********	********	*******	*********	********		******	********	1.2	0.9	0.5
15,000	*******	********	********	********	*******	*********	********		******	********	********	0.8	0.5
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 2.4%. The finding that there were 5,393,631 people (to be rounded according to the rounding guidelines in Section 9.1) that did not use the Internet during the past 12 months for personal use is publishable with no qualifications.

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

Suppose that the user estimates that 488,346 / 5,393,631= 9.1% of the people that did not use the Internet during the past 12 months for personal use (CU_Q01 = 2, No) did not use the Internet due to the cost of service or equipment (CU_Q12A = 1, Yes). How does the user determine the coefficient of variation of this estimate?

1) Refer to the person 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 that did not use the Internet during the past 12 months for personal use), it is necessary to use both the percentage (9.1%) and the numerator portion of the percentage (488,346) in determining the coefficient of variation.

3) The numerator, 488,346, 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 500,000. Similarly, the percentage estimate does not appear as any of the column headings, so it is necessary to use the percentage closest to it, 10.0%.

4) The figure at the intersection of the row and column used, namely 8.0% is the coefficient of variation to be used.

5) So the approximate coefficient of variation of the estimate is 8.0%. The finding that 9.1% of people that did not use the Internet during the past 12 months for personal use because of the cost of service or equipment can be published with no qualifications.

Example 3: Estimates of Differences Between Aggregates or Percentages

Suppose that a user estimates that 2,437,333 / 3,345,020= 72.9% of households in Quebec (PROVINCE = 24) have access to the Internet (HA_Q01 = 1, Yes), 4,057,281 / 4,986,798= 81.4% of households in Ontario (PROVINCE = 35) have access to the Internet (HA_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 for the household file in the same manner as described in Example 1 gives the CV of the estimate for Households in Quebec as 1.3%, and the CV of the estimate for households in Ontario as 0.6%.

Canadian Internet Use Survey, 2010 - Household File

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	109.8	109.3	108.7	107.1	104.2	101.3		91.9	88.6	85.1	77.7	60.2	34.7
2	77.6	77.3	76.9	75.7	73.7	71.6		65	62.6	60.2	54.9	42.5	24.6
3	63.4	63.1	62.8	61.8	60.2	58.5		53.1	51.1	49.1	44.8	34.7	20.1
4	********	54.6	54.4	53.5	52.1	50.6		45.9	44.3	42.5	38.8	30.1	17.4
5	********	48.9	48.6	47.9	46.6	45.3		41.1	39.6	38	34.7	26.9	15.5
.													
.													
.													
60	********	********	14	13.8	13.5	13.1		11.9	11.4	11	10	7.8	4.5
65	********	********	13.5	13.3	12.9	12.6		11.4	11	10.6	9.6	7.5	4.3
70	********	********	********	12.8	12.5	12.1		11	10.6	10.2	9.3	7.2	4.2
75	********	********	********	12.4	12	11.7		10.6	10.2	9.8	9	6.9	4
80	********	********	********	12	11.7	11.3		10.3	9.9	9.5	8.7	6.7	3.9
85	********	********	********	11.6	11.3	11		10	9.6	9.2	8.4	6.5	3.8
90	********	********	********	11.3	11	10.7		9.7	9.3	9	8.2	6.3	3.7
95	********	********	********	11	10.7	10.4		9.4	9.1	8.7	8	6.2	3.6
100	********	********	********	10.7	10.4	10.1		9.2	8.9	8.5	7.8	6	3.5
125	********	********	********	9.6	9.3	9.1		8.2	7.9	7.6	6.9	5.4	3.1
150	********	********	********	8.7	8.5	8.3		7.5	7.2	6.9	6.3	4.9	2.8
200	********	********	********	********	7.4	7.2		6.5	6.3	6	5.5	4.3	2.5
250	********	********	********	********	6.6	6.4		5.8	5.6	5.4	4.9	3.8	2.2
300	********	********	********	********	6	5.8		5.3	5.1	4.9	4.5	3.5	2
350	********	********	********	********	********	5.4		4.9	4.7	4.5	4.2	3.2	1.9
400	********	********	********	********	********	5.1		4.6	4.4	4.3	3.9	3	1.7
450	********	********	********	********	********	4.8		4.3	4.2	4	3.7	2.8	1.6
500	********	********	********	********	********	4.5		4.1	4	3.8	3.5	2.7	1.6
750	********	********	********	********	********	**********		3.4	3.2	3.1	2.8	2.2	1.3
1,000	********	********	********	********	********	**********		2.9	2.8	2.7	2.5	1.9	1.1
1,500	********	********	********	********	********	**********		**********	****	********	2	1.6	0.9
2,000	********	********	********	********	********	**********		**********	****	********	********	1.3	0.8
3,000	********	********	********	********	********	**********		**********	****	********	********	********	0.6

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

Canadian Internet Use Survey, 2010 - Household File

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	117.8	117.3	116.7	114.9	111.8	108.7		98.6	95	91.3	83.4	64.6	37.3
2	83.3	82.9	82.5	81.2	79.1	76.8		69.7	67.2	64.6	58.9	45.7	26.4
3	68	67.7	67.4	66.3	64.6	62.7		56.9	54.9	52.7	48.1	37.3	21.5
4	58.9	58.6	58.3	57.4	55.9	54.3		49.3	47.5	45.7	41.7	32.3	18.6
5	********	52.5	52.2	51.4	50	48.6		44.1	42.5	40.8	37.3	28.9	16.7
.													
.													
.													
60	********	********	15.1	14.8	14.4	14		12.7	12.3	11.8	10.8	8.3	4.8
65	********	********	14.5	14.3	13.9	13.5		12.2	11.8	11.3	10.3	8	4.6
70	********	********	13.9	13.7	13.4	13		11.8	11.4	10.9	10	7.7	4.5
75	********	********	13.5	13.3	12.9	12.5		11.4	11	10.5	9.6	7.5	4.3
80	********	********	13	12.8	12.5	12.2		11	10.6	10.2	9.3	7.2	4.2
85	********	********	12.7	12.5	12.1	11.8		10.7	10.3	9.9	9	7	4
90	********	********	12.3	12.1	11.8	11.5		10.4	10	9.6	8.8	6.8	3.9
95	********	********	12	11.8	11.5	11.2		10.1	9.8	9.4	8.6	6.6	3.8
100	********	********	********	11.5	11.2	10.9		9.9	9.5	9.1	8.3	6.5	3.7
125	********	********	********	10.3	10	9.7		8.8	8.5	8.2	7.5	5.8	3.3
150	********	********	********	9.4	9.1	8.9		8.1	7.8	7.5	6.8	5.3	3
200	********	********	********	8.1	7.9	7.7		7	6.7	6.5	5.9	4.6	2.6
250	********	********	********	********	7.1	6.9		6.2	6	5.8	5.3	4.1	2.4
300	********	********	********	********	6.5	6.3		5.7	5.5	5.3	4.8	3.7	2.2
350	********	********	********	********	6	5.8		5.3	5.1	4.9	4.5	3.5	2
400	********	********	********	********	5.6	5.4		4.9	4.8	4.6	4.2	3.2	1.9
450	********	********	********	********	5.3	5.1		4.6	4.5	4.3	3.9	3	1.8
500	********	********	********	********	********	4.9		4.4	4.3	4.1	3.7	2.9	1.7
750	********	********	********	********	********	********		3.6	3.5	3.3	3	2.4	1.4
1,000	********	********	********	********	********	**********		3.1	3	2.9	2.6	2	1.2
1,500	********	********	********	********	********	**********		*******	2.5	2.4	2.2	1.7	1
2,000	********	********	********	********	********	**********		**********	****	********	1.9	1.4	0.8
3,000	********	********	********	********	********	**********		**********	****	********	********	1.2	0.7
4,000	********	********	********	********	********	**********		**********	****	********	********	********	0.6

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 $\hat{d} = 0.729 – 0.814 = -0.085$ is:

$$\sigma_{\hat{d}} = \sqrt{[(0.729)(0.013)]^2 + [(0.814)(0.006)]^2}$$
$$= \sqrt{(0.000089) + (0.000024)}$$
$$= 0.0106$$

3) The coefficient of variation of $\hat{d}$ is given by $\sigma_{\hat{d}} / \hat{d} = 0.0106 / 0.085 = 0.125$

4) So the approximate coefficient of variation of the difference between the estimates is 12.5%, which is publishable with no qualifications.

Example 4: Estimates of Ratios

Suppose that the user estimates that 2,437,333 households in Quebec have access to the Internet (HA_Q01 = 1, Yes), while 4,057,281 households in Ontario have access to the Internet (HA_Q01 = 1, Yes). The user is interested in comparing the estimate of households 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 households in Quebec that have access to the Internet. The denominator of the estimate ($\hat{X}_2$) is the number of households in Ontario that have access to the Internet.

2) Refer to the household coefficient of variation tables for Quebec and Ontario (see above).

3) The numerator of this ratio estimate is 2,437,333. The figure closest to it is 2,000,000. 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, 1.3%.

4) The denominator of this ratio estimate is 4,057,281. The figure closest to it is 4,000,000. 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, 0.6%

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:

$$\alpha_{\hat{R}} = \sqrt{(0.013)^2 + (0.006)^2}$$
$$= \sqrt{0.000169 + 0.000036}$$
$$= 0.014$$

6) The obtained ratio of households in Quebec versus households in Ontario that have access to the Internet is 2,437,333 / 4,057,281 which is 0.60:1 (to be rounded according to the rounding guidelines in Section 9.1). The coefficient of variation of this estimate is 1.4%, which makes the estimate releasable with no qualifications.

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.

Example of Using the Coefficient of Variation Tables to Obtain Confidence Limits
A 95% confidence interval for the estimated proportion of Canadians that did not use the Internet during the past 12 months for personal use due to the cost of service or equipment (from Example 2) would be calculated as follows:

$\hat{X}$ = 9.1% (or expressed as a proportion 0.091)

t = 2

$\alpha_{\hat{x}}$ = 8.0% (0.08 expressed as a proportion) is the coefficient of variation of this estimate as determined from the tables.

$CI_{\hat{x}}$ = {0.091 – (2) (0.091) (0.08), 0.091 + (2) (0.091) (0.08)}

$CI_{\hat{x}}$ = {0.091 – 0.015, 0.091 + 0.015}

$CI_{\hat{x}}$ = {0.076, 0.106}

With 95% confidence it can be said that between 7.6% and 10.6% of Canadians did not use the Internet during the past 12 months for personal use due to the cost of service or equipment.

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.

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 that have access to the Internet and the proportion of households in Ontario that have access to the internet. From Example 3, the standard error of the difference between these two estimates was found to be 0.0106. Hence,

$$t = \frac{\hat{X}_1 - \hat{X}_2}{\sigma_{\hat{d}}} = \frac{0.729 - 0.814}{0.0106} = \frac{-0.085}{0.0106} = -8.02$$

Since t = -8.02 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.

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 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.