Microdata User Guide

Canadian Financial Capability Survey

2014

Statistics Canada Statistique Canada

Canada

How to obtain more information

Specific inquires about this product and related statistics or services should be directed to:
Statistics Canada Client Services
Telephone: 613-951-3321 or call toll-free 1-800-461-9050
Fax: 613-951-4527
E-mail:infostats@statcan.gc.ca

Accessing and ordering information

The Canadian Financial Capability Survey (CFCS) produces two types of microdata files: master files, and public use microdata files (PUMF).

Master files

The master files contain all variables and all records from the survey collected during a collection period. These files are accessible at Statistics Canada for internal use and in Statistics Canada's Research Data Centres (RDC), and are also subject to custom tabulation requests.

Research Data Centre

The RDC Program enables researchers to use the survey data in the master files in a secure environment in several universities across Canada. Researchers must submit research proposals that, once approved, give them access to the RDC. For more information, please consult the following web page: http://www.statcan.gc.ca/rdc-cdr/index-eng.htm

Custom tabulations

Another way to access the master files is to offer all users the option of having staff in Client Services of Special Surveys Division prepare custom tabulations. This service is offered on a cost-recovery basis. It allows users who do not possess knowledge of tabulation software products to get custom results. The results are screened for confidentiality and reliability concerns before release. For more information, please contact Client Services.

Public use microdata files

The public use microdata files are developed from the master files using a technique that balances the need to ensure respondent confidentiality with the need to produce the most useful data possible. The PUMF must meet stringent security and confidentiality standards required by the *Statistics Act* before they are released for public access. To ensure that these standards have been achieved, each PUMF goes through a formal review and approval process by an executive committee of Statistics Canada. Variables most likely to lead to identification of an individual are deleted from the data file or are collapsed to broader categories.

A Microdata Licence Agreement is signed before releasing the file(s). This requires the name of the person who will be responsible for the data file and their contact information.

To obtain a copy of the PUMF contact Client Services.

The Data Liberation Initiative

The Data Liberation Initiative (DLI) Program enables students and researchers to use the public use microdata files in several universities across Canada. For more information, please consult the following web page: http://www.statcan.gc.ca/dli-ild/dli-idd-eng.htm

Table of Contents

1.0 Survey description

The Canadian Financial Capability Survey (CFCS) was conducted by Statistics Canada from May 14, to June 21, 2014 with the cooperation and support of the Financial Consumer Agency of Canada (FCAC). This manual has been produced to facilitate the manipulation of the microdata file of the survey results.

The Canadian Financial Capability Survey was a voluntary survey.

The 2014 version of the Canadian Financial Capability Survey (CFCS) is the second time this survey has been administered. The first CFCS was collected in 2009. The need for this survey was brought about by changing economic conditions and the variety and complexity of financial products available. It can be very difficult for the average investor to understand the risks, fees and potential returns of investing and borrowing. Policy makers foresee a need for more knowledgeable financial consumers who will be able to effectively look after their own interests in increasingly complex financial markets.

2.0 Concepts and definitions

2.1 Canadian Financial Capability Survey concepts and definitions

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

Asset – Anything having a monetary value that is owned by a person (or business). Real estate, stocks, bonds, and money itself are all considered to be assets.

Bankcard – A card issued by a bank that entitles the holder to make electronic payments with a point of sale terminal and to carry out banking transactions via an automatic teller.

Bonds – A certificate of indebtedness, issued by a government or corporation. Interest rates are fixed for the term of the bond but the bond may be sold at more or less than its face value.

Corporate Bonds - A bond issued by a corporation. It carries no claim to ownership and pays no dividends.

Credit card - Method of paying for goods and services, whereby the purchaser defers payment and repays the principal and interest in instalment over time. These include bank-type credit cards (i.e., VISA, Mastercard) as well as retail store and gas station cards.

Credit union - A credit union is a cooperative association, whose membership consists of individuals with some common relationship, usually the same employer, labour union, church, fraternal association or even neighbourhood. Credit unions offer auto, personal and other loans, savings accounts, chequing accounts, credit cards and certificate of deposits to their members.

Debit Card – see Bankcard above

Debt – An amount owed by one party to another for money, goods or services.

Dividend – Monetary amount paid to shareholders of a company from profits made by that company.

Equity – The residual interest in assets after deducting related liabilities. For example the “equity” in a home equals the value of the home minus the amount owed on the mortgage.

Family – a group of two or more persons who live in the same dwelling and are related to each other by blood, marriage, adoption, or common-law.

Guaranteed Investment Certificate (GIC) – A savings vehicle having terms generally ranging from one to five years, during which time the interest rate is guaranteed and the money is usually locked-in until maturity.

Home Buyers' Plan (HBP) – This is a government sponsored plan that allows people to withdraw up to $20,000 tax free from their Registered Retirement Savings Plan (RRSP) to apply towards the purchase of a home.

Household – A household consists of any person or group of persons (related or not) occupying a dwelling who has no usual place of residence elsewhere.

Interest – Payment made at a specified rate for the use of borrowed money.

Line of credit account (LOC) – A formal agreement between a borrower and a lender (usually a financial institution) which allows the borrower to borrow as much or as little as they wish up to a pre-specified maximum (or credit limit). For purposes of the survey, the amount to be reported is the amount currently owing on the line of credit.

Line of credit - The maximum amount of credit that a customer of a store, bank, etc. is authorized to use.

Mortgage – Any loan that uses a home or other real estate as collateral.

Mutual Funds – A collection of numerous financial securities that are bought by an investment company and sold as a particular group or fund. Investors purchase units of these funds.

Old Age Security Pension (OAS) –- A monthly benefit available to most Canadians 65 years of age or older who have lived in Canada for at least 10 years.

Overdraft – The amount by which a cheque or other payments exceeds the funds on deposits.

Pawnbroker - A person, whose business is lending money at interest, on personal, movable property, deposited with the lender until redeemed.

Personal line of credit - A formal agreement between a borrower and lender (usually a financial institution), which allows the borrower to borrow through a special account, as much or as little as they require to a pre-specified maximum (or credit limit).

Powers of attorney - A written document given by one person or party to another, authorizing the latter to act for the former.

Pension plan - Retirement benefits that are paid out to employees of companies in the private sector. These benefits are funded by employers and paid to employees when they retire from the company.

Registered Disability Savings Plan (**RDSP) –** A plan that allows funds to be invested tax-free until withdrawal. It is intended to help parents and others to save for the long-term financial security of a child with a disability. *Canadian Financial Capability Survey, 2009 – User Guide Special Surveys Division 13*

Registered Education Savings Plans (**RESP)** – A savings vehicle designed for individuals to accumulate income for post-secondary education. Typically, the plans are entered into by parents seeking to save for their children's post-secondary education. Investing in RESPs can be advantageous since the federal government makes a contribution and income generated is tax sheltered until it is withdrawn for the child's post-secondary education.

Registered Income Fund (**RIF)** – A fund into which RRSP monies may be transferred. Payments from a RIF may be varied, but a minimum amount must be withdrawn annually.

Registered Retirement Savings Plan (RRSP or RSP) – A capital accumulation program designed to encourage savings for retirement. Contributions are tax deductible, within prescribed limits. Investment

income earned in the RRSP is tax-exempt, but benefits are taxable. Amounts in these plans include amounts originally invested plus accrued interest/earnings.

Reverse Mortgage - A type of loan, especially for retired or elderly persons, in which tax-free payments are borrowed from and up to the full equity of a home.

Stocks – Common and preferred shares of corporations (could also be referred to as equities). Related terms: publicly-traded stock, common shares, preferred stock, shares.

Taxable Capital Gain - Incurred when a capital property is considered to have been sold for more than the total of its adjusted cost base plus outlays and expenses incurred to sell the property.

Tax Free Savings Account or Tax Free Savings Plan (TFSA) – Canadian residents age 18 and older can contribute up to $5,000 per year without being taxed on investment income or capital gains.

Term Deposits – A deposit instrument most commonly available from trust companies and chartered banks, requiring a minimum investment at a predetermined rate of interest for a stated term. The interest rate varies according to the amount invested and the term to maturity.

Trust – A trust is an arrangement whereby the right to property is held by one party, the "trustee" (or manager), for the benefit of another, the "beneficiary".

Trust Company – A financial institution that provides financial and trust services to individuals and corporations. A large part of the business of trust companies is acting as trustees for other corporations in handling pension funds, bond issues and the like. They are active financial intermediaries, taking in deposits and making loans of various kinds.

2.2 Content development

In the case of the Canadian Financial Capability Survey, it was proposed from its conception in 2009 that it be collected by telephone interview; an approach that reflected previous successes in other countries with similar subject matter. A first round of cognitive testing, including one-on-one interviews and focus group discussions, across Canada in spring 2007 confirmed that this was indeed the best way to proceed.

With the addition of Finance Canada and the Bank of Canada as active partners, the content was modified to reflect each of the partners data needs. This led to a second round of cognitive testing in only a few selected cities in the spring of 2008. The computer-assisted telephone interviewing (CATI) application was developed and tested during the summer and fall months in 2008.

In an attempt to maintain comparability, the 2014 Canadian Financial Capability Survey has kept the 2009 CFCS's content intact, making only very slight updates where necessary. The 2014 application was then re-developed and tested in the early months of 2014.The 2014 Canadian Financial Capability Survey (CFCS) was designed to collect information about Canadians' knowledge, abilities and behaviour concerning financial decision making. In other words, how Canadians understand their financial situation, the financial services available to them and their plans for the future. The questions asked to determine the respondent's knowledge of financial matters can be classified in three broad groups:

- knowledge of financial issues and instruments,
- financial skills and competence, i.e., the ability to apply knowledge and make financial decisions, and
- financial responsibility, i.e., behaviour in financial matters such as debt re-payment, saving and investing.

In 2009, the CFCS was collected using Random Digit Dialing (RDD). Statistics Canada as a whole is moving away from RDD and towards an address based frame.

3.0 Survey methodology

The Canadian Financial Capability Survey (CFCS) was administered from May 12, 2014 to June 21, 2014. The CFCS sample design is closely tied to that of the Labour Force Survey (LFS) as it was administered to a sub-sample of dwellings that completed their last LFS interview in January and February 2014. 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.

3.1 Target and survey population

The survey population is the population that is actually covered by the survey. It may not be the same as the target population, though, ideally, the two populations should be very similar. It is important to note that conclusions based on the survey results apply only to the survey population.

Various reasons can explain the differences between the two populations. For example, the difficulty and high cost of collecting data in isolated regions may lead to the decision to exclude these units from the survey population. Similarly, members of the target population who are living abroad or are institutionalised may not be part of the survey population if they are too difficult or costly to survey.

The survey population of the 2014 CFCS covers the civilian, non-institutionalised population that is 18 years of age and over. It was conducted in all ten provinces. Excluded from the survey's coverage are: persons living in the territories; persons living on reserves and other Aboriginal settlements; full-time members of the Canadian Armed Forces; and the institutionalized population.

3.2 Sample design

The initial survey frame for the 2014 CFCS was a list of all the dwellings that completed their last LFS interview in January and February 2014. Since the CFCS was conducted as a Computer Assisted Telephone Interview (CATI), all dwellings with missing telephone numbers were first removed from our frame. Then, in order to stay within our collection budget while producing nationally reliable estimates, we selected a systematic sample of dwellings from our frame using the Kish sample allocation method at the provincial level. Upon contact with the dwellings in our sample, one member for each dwelling (subject to the above coverage statement) was randomly chosen to complete the CFCS interview.

3.3 Sample size

The following table shows the number of LFS dwellings that were in the CFCS sample. This table includes dwellings which were non-respondents to the CFCS.

Province	Initial CFCS Sample Size
Newfoundland and Labrador	649
Prince Edward Island	435
Nova Scotia	940

New Brunswick	883
Quebec	1,900
Ontario	2,763
Manitoba	1,202
Saskatchewan	1,029
Alberta	1,382
British Columbia	1,437
Canada	**12,620**

4.0 Data collection

The CFCS is a survey with a cross-sectional design and is based on the current Labour Force Survey (LFS, record number 3701). The LFS has a stratified multi-stage survey design. Each province is divided into large geographic stratum. The first stage of sampling consists of selecting smaller geographic areas, called Primary Sampling Units (PSUs), from within each stratum. The second stage of sampling consists of selecting dwellings from within selected PSU. The LFS covers the civilian, non-institutionalised population 15 years of age and over.

The LFS uses a rotating panel sample design so that selected dwellings remain in the LFS sample for six consecutive months. Each month about 1/6th of the LFS sampled dwellings are in their first month of the survey, 1/6th are in their second month of the survey, and so on. One feature of the LFS sample design is that each of the six rotation groups can be used as a representative sample by itself. The CFCS sample was comprised of dwellings from two rotation groups. More specifically, the CFCS used the two rotation groups that completed their last LFS interviewers in January and February 2014.

Data collection for this reference period: 2014-05-12 to 2014-06-21

Data were collected directly from survey respondents.

Proxy interviews were not permitted.

4.1 Questionnaire Structure

The survey collected a large amount of data for each selected respondent as well as some information about the household. Of particular interest was to identify whether the selected respondent deemed themselves the most knowledgeable person for the household's ongoing expenses and financial management. This self-identification occurred in both of these sections of the survey. The complete questionnaire can be found in Appendix C of this document. Identified below are the sections of the questionnaire.

Introduction: Control form and development of household roster

> The control form guides the interviewer through the opening phase of the interview and provides a table shell to build the household roster. Based on the household roster an eligible household member is randomly selected to complete the survey.

Demography

> This section provides some basic demographic information.

Labour Force

The labour force section identifies current employment status, whether the respondent (and their spouse/partner if appropriate) has worked in the past 12 months and the type of work.

Ongoing Expenses

This section of the survey deals with day-to-day expenses and ongoing bill payments. The information collected in this section pertains to how individuals keep track of their finances and how they manage money.

Financial Management

This section of the questionnaire deals with longer-term financial planning. It involves questions surrounding major purchases, planning for retirement and where appropriate planning for children's post-secondary education.

Assets and Debts

This section of the questionnaire asks questions concerning the assets and debts attributable to the individual or household – as appropriate. This information helps to profile the financial situation of the household.

Income

This information provides contextual information surrounding both the individual and where appropriate the household's income.

Financial Choices

This section of the survey addresses how we approach financial choices and is relevant for issues of planning and responsibility.

Subjective Personal Assessment

In this section of the survey, respondents provide a self-assessment of their comfort with financial matters.

Objective Personal Assessment

The final section of the questionnaire asks respondents to provide answers to a short money quiz.

5.0 Data processing

Processing transforms survey responses obtained during collection into a form that is suitable for tabulation and data analysis. It includes all data handling activities – automated and manual – after collection and prior to estimation.

5.1 Data capture

Computer-assisted interviews:
Responses to survey questions were 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.

5.2 Editing

Editing can occur at several points throughout the survey process and ranges from simple preliminary checks performed by interviewers in the field to more complex automated verifications performed by a computer program after the data have been captured. In general, edit rules are based upon what is logically or validly possible, based upon:

- expert knowledge of the subject matter;
- other related surveys or data;
- the structure of the questionnaire and its questions;
- statistical theory.

There are three main categories of edits: validity, consistency and distribution edits. Validity edits verify the syntax of responses and include such things as checking for non-numeric characters reported in numeric fields and checking for missing values. Validity edits can also check that the coded data lie within an allowed range of values. For example, a range edit might be put on the reported age of a respondent to ensure that it lies between 0 and 125 years.

Consistency edits verify that relationships between questions are respected. Consistency edits can be based on logical, legal, accounting or structural relationships between questions or parts of a question. The relationship between date of birth and marital status is one example where an edit might be: “a person less than 15 years of age cannot have any marital status other than never married.”

Distribution edits are performed by looking at data across questionnaires. These edits attempt to identify records that are outliers with respect to the distribution of the data. Distribution edits are sometimes referred to as statistical edits or outlier detection.

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.

The third type of error treated involved the identification of incoherent entries based on logical relationship between questions.
Electronic text files containing the daily transmissions of completed cases are combined to create the “raw” survey file. At the end of collection, this file should contain one record for each sampled individual. Before further processing, verification is performed to identify and eliminate potential duplicate records and to drop non-response and out-of-scope records.

As a result, editing takes place by modifying the data at the individual variable level. The first step in editing is to determine which items from the survey output need to be kept on the survey master file. Subsequently, invalid characters are deleted and the data items are formatted appropriately. Text fields are stripped off the main files and written to a separate file for coding. The first type of error treated was errors in questionnaire flow, where questions that did not apply to the respondent (and should therefore not have been answered) were found to sometimes 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. For skips based on answered questions, all skipped questions are set to “Valid skip” (6, 96, 996, etc.). For skips based on “Don't know” or “Refusal”, all skipped questions are set to Not stated (9, 99, 999, etc.). The remaining empty items are filled with a numeric value (9, 99, 999, etc. depending on variable length). These codes are reserved for processing purposes and mean that the item was “Not stated”.

5.3 Coding open-ended questions

A few data items on the questionnaire were recorded by interviewers in an open-ended format. This typically occurs when a respondent selects the other please specify response to a question. In these instances the responses are reviewed to determine if they should be allocated to one of the pre-existing categories, be joined with other similar responses to create a new category or whether they remain as just “other”.

5.4 Creation of derived variables

A total of 36 data items on the microdata file have been derived by combining items on the questionnaire in order to facilitate data analysis. Most are continuous variables related to age and the number of years of service a workers has/ had and were grouped in pre-determined intervals as an effort to aid the analytical process.

5.5 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 7.0). 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 CFCS, donor imputation was used to fill in missing data for item and partial non-response for personal and household income amounts. All 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-respondents. 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 of personal and household incomes was performed (together whenever necessary, and then always from the same donor). The following table shows the imputation rate for each of the variables where applicable.

	Household Income	Personal Income
Imputed	3,669	1,875
Total	6,685	6,685
Rate (%)	54.8	28.0

6.0 Data quality

Survey errors come from a variety of different sources. They can be classified into two main categories: non-sampling errors and sampling errors.

6.1 Non-sampling errors

Non-sampling errors can be defined as errors arising during the course of virtually all survey activities, apart from sampling. They are present in both sample surveys and censuses (unlike sampling error, which is only present in sample surveys). Non-sampling errors arise primarily from the following sources: non-response, coverage, measurement and processing.

Non-response

Non-response errors result from a failure to collect complete information on all units in the selected sample.

Non-response produces errors in the survey estimates in two ways. First, is that non-respondents often have different characteristics from respondents, which can result in biased survey estimates if non-response is not corrected properly. Secondly, it reduces the effective size of the sample, since fewer units than expected answered the survey. As a result, the sampling variance increases and the precision of the estimates decrease.

The following table summarizes the response rates to the CFCS.

Province	Total Number of Households in the Sample	Out-of-scope Households	In-scope Households	Total Persons Responding	Overall Response Rate (%)
Newfoundland and Labrador	649	27	622	353	56.8
Prince Edward Island	435	20	415	245	59.0
Nova Scotia	940	38	902	561	62.2
New Brunswick	883	39	844	510	60.4
Quebec	1,900	60	1,840	1,116	60.7
Ontario	2,763	107	2,656	1,357	51.1
Manitoba	1,202	64	1,138	681	59.8
Saskatchewan	1,029	53	976	563	57.7
Alberta	1,382	109	1,273	652	51.2
British Columbia	1,437	80	1,357	647	47.7
Canada	**12,620**	**597**	**12,023**	**6,685**	**55.6**

Coverage errors

Coverage errors consist of omissions, erroneous inclusions, duplications and misclassifications of units in the survey frame. Since they affect every estimate produced by the survey, they are one of the most important types of error; in the case of a census they may be the main source of error. Coverage errors may cause a bias in the estimates and the effect can vary for different sub-groups of the population.

Differences between the target population and the survey population (see Section 3.1) are the result of deliberate restrictions to coverage. Excluded from the CFCS survey coverage are persons living in the territories, persons living on reserves and other Aboriginal settlements, full-time members of the Canadian Armed Forces, and the institutionalized population.

Measurement errors

Measurement errors (or sometime referred to as response errors) occur when the response provided differs from the real value; such errors may be attributable to the respondent, the interviewer, the questionnaire, the collection method or the respondent's record-keeping system. Such errors may be random or they may result in a systematic bias if they are not random.

It is very costly to accurately measure the level of response error and very few surveys conduct a post-survey evaluation. However, interviewer feedback and observation reports usually provide clues as to which questions may be problematic (poorly worded question, inadequate interviewer training, poor translation, technical jargon, no help text available, etc.).

For the CFCS, interviewer training consisted of reading the Supervisor's Manual, Procedures Manual and Interviewer's Manual, practicing with the 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.

Processing errors

Processing error is the error associated with activities conducted once survey responses have been received. It includes all data handling activities after collection and prior to estimation. Like all other errors, they can be random in nature, and inflate the variance of the survey's estimates, or systematic, and introduce bias.

Data processing of the CFCS 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.

6.2 Sampling errors

Sampling error is defined as the error that results from estimating a population characteristic by measuring a portion of the population rather than the entire population. For probability sample surveys, methods exist to calculate sampling error. These methods derive directly from the sample design and method of estimation used by the survey.

The most commonly used measure to quantify sampling error is sampling variance. Sampling variance measures the extent to which the estimate of a characteristic from different possible samples of the same size and the same design differ from one another. For sample designs that use probability sampling, the magnitude of an estimate's sampling variance can be estimated. The key issue is the magnitude of an estimate's estimated sampling variance relative to the size of the survey estimate: if the variance is relatively large, then the estimate has poor precision and is unreliable.

Factors affecting the magnitude of the sampling variance include:

1. The variability of the characteristic of interest in the population: the more variable the characteristic in the population, the larger the sampling variance.
2. The size of the population: in general, the size of the population only has an impact on the sampling variance for small to moderate sized populations.
3. The response rate: the sampling variance increases as the sample size decreases. Since non-respondents effectively decrease the size of the sample, non-response increases the sampling variance.
4. The sample design and method of estimation: some sample designs are more efficient than others in the sense that, for the same sample size and method of estimation, one design can lead to smaller sampling variance than another.

The standard error of an estimator is the square root of its sampling variance. This measure is easier to interpret since it provides an indication of sampling error using the same scale as the estimate whereas the variance is based on squared differences.

However, even standard error might be difficult to interpret in terms of "How big a standard error is acceptable?" What is large depends on the magnitude of the estimate. For example, a standard error of 100 would be considered large for measuring the average weight of people but would not be considered large for estimating average annual income.

It is more useful in many situations to assess the size of the standard error relative to the estimate of the characteristic being measured. The coefficient of variation (CV) provides such a

measure. It is the ratio of the standard error of the survey estimate to the average value of the estimate itself, across all possible samples. The coefficient of variation is usually computed as the estimate of the standard error of the survey estimate to the estimate itself. This relative measure of sampling error is usually expressed as a percentage (10% instead of 0.1). It is very useful in comparing the precision of sample estimates, where their sizes or scale differ from one another.

7.0 Weighting

7.1 Weighting procedures for the LFS

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.

7.2 Weighting procedures for the Canadian Financial Capability Survey

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

1) An adjustment to account for the use of two out of six LFS rotation groups.

2) An adjustment to account for dwellings without a telephone number on our survey frame.

3) An adjustment to account for dwellings that were sampled out.

4) An adjustment to account for the person level and the household level non-respondents. All units selected for the CFCS were modeled 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 household-level non-respondents where it was not known which member was selected to the CFCS; and a second step for person-level non-respondents where more detailed demographic information could be used in the modeling.

5) An adjustment to account for the fact that we interview one individual per dwelling.

6) A final adjustment is done to match the census projections for independent region-age and region-sex groups as well as income distributions at the national and regional levels as projected by the Survey of Labour and Income Dynamics (SLID).

The resulting weight WTPM is the final weight which appears on the CFCS master file as well as WTPP on the public use microdata file.

8.0 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.

8.1 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 ratios) are to be derived from their corresponding unrounded components and then are to be rounded themselves to the nearest 100 units (or the nearest one decimal) using normal rounding.

e) In instances where, due to technical or other limitations, a rounding technique other than normal rounding is used resulting in estimates to be published or otherwise released which differ from corresponding estimates published by Statistics Canada, users are urged to note the reason for such differences in the publication or release document(s).

f) Under no circumstances are unrounded estimates to be published or otherwise released by users. Unrounded estimates imply greater precision than actually exists.

8.2 Sample Weighting Guidelines for Tabulation

The sample design used for the Canadian Financial Capability Survey was not self-weighting. When producing simple estimates (including the production of ordinary statistical tables), users must apply the proper survey weights.

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.

8.3 Definitions of Types of Estimates: Categorical and Quantitative

Before discussing how the CFCS 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 CFCS.

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 Canadian adults that have a household budget or the proportion that are behind in their payments 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: Does your household have a budget?
R: Yes/No

Q: Thinking about the last 12 months, were you ever behind two or more consecutive months in paying a bill?
R: Yes/ No

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 personal bank accounts held by adult women in Canada. The numerator is an estimate of the total number of personal bank accounts and its denominator is the number of adult women (18 years of age and older) in Canada.

Example of Quantitative Questions:

Q: How many personal chequing or savings accounts do you currently have with a bank, credit union or trust company?
R: |_|_| accounts

Tabulation of Categorical Estimates

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

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

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 personal bank accounts held by adult women in Canada multiply the value reported in OE_Q02A by the final weight for the record, then sum this value over all records with SEX=2 (female).

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 personal bank accounts held by women.

a) estimate the total number of personal bank accounts for Albertans ($\hat{X}$) as described above,
b) estimate the number of women in Canada ($\hat{Y}$) by summing the final weights of all records in the category with SEX=2, then
c) divide estimate a) by estimate b) ($\hat{X} / \hat{Y}$).

8.4 Guidelines for Statistical Analysis

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

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

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

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

1) select all respondents from the file who reported SEX = male;

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 = male;

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

8.5 Coefficient of Variation Release Guidelines

Before releasing and/or publishing any estimates from the CFCS, 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 6.0. 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 6.0 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."

8.6 Release Cut-offs for the Canadian Financial Capability Survey

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

For example, the table shows that the quality of a weighted estimate of 19,000 people possessing a given characteristic in Newfoundland and Labrador is marginal.

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

Province and Region	**Acceptable CV 0.0% to 16.5%**		**Marginal CV 16.6% to 33.3%**			**Unacceptable CV > 33.3%**	
Newfoundland and Labrador	66,600	& over	18,500	to <	66,600	under	18,500
Prince Edward Island	20,000	& over	5,700	to <	20,000	under	5,700
Nova Scotia	69,000	& over	18,200	to <	69,000	under	18,200
New Brunswick	60,900	& over	16,200	to <	60,900	under	16,200
Quebec	463,200	& over	120,200	to <	463,200	under	120,200
Ontario	646,300	& over	166,200	to <	646,300	under	166,200
Manitoba	85,200	& over	22,400	to <	85,200	under	22,400
Saskatchewan	96,000	& over	25,600	to <	96,000	under	25,600
Alberta	397,300	& over	108,600	to <	397,300	under	108,600
British Columbia	394,700	& over	105,300	to <	394,700	under	105,300
Atlantic Provinces	64,200	& over	16,200	to <	64,200	under	16,200
Prairie Provinces	286,200	& over	73,600	to <	286,200	under	73,600
Canada	**481,800**	**& over**	**119,900**	**to <**	**481,800**	**under**	**119,900**

Appendix A – Variance estimation for master and share files

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

Bootstrap method for variance estimation

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

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

Statistical packages for variance estimation

Bootvar

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

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

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

Other packages

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

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

Rule 2: Estimates of Proportions or Percentages of Persons Possessing a Characteristic
The coefficient of variation of an estimated proportion or percentage depends on both the size of the proportion or percentage and the size of the total upon which the proportion or percentage is based. Estimated proportions or percentages are relatively more reliable than the corresponding estimates of the numerator of the proportion or percentage, when the proportion or percentage is based upon a sub-group of the population. For example, the proportion of Canadian women who have a household budget is more reliable than the estimated number of Canadian women who have a household budget. (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 women with a household budget and the numerator is the number of women with a household budget who report that they always stay within their budget.

In the case where the numerator is not a subset of the denominator, as for example, the ratio of the number of men with a household budget as compared to the number of women with a household budget, 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 based on the 2009 CFCS are included to assist users in applying the foregoing rules.

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

Suppose that a user estimates that 13,233,741 adult Canadians report having a household budget. How does the user determine the coefficient of variation of this estimate?

1) Refer to the coefficient of variation table for CANADA.

2) The estimated aggregate (13,233,741) 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 12,500,000.

3) The coefficient of variation for an estimated aggregate is found by referring to the first non-asterisk entry on that row, namely, 1.0%.

4) So the approximate coefficient of variation of the estimate is 1.0%. The finding that

there were 13,233,741 (to be rounded according to the rounding guidelines in Section 9.1) adult Canadians with a household budget is publishable with no qualifications.

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

Suppose that the user estimates that 4,881,629 / 13,233,741 = 36.9% of Canadian adults with a household budget report that they always stay within this budget. How does the user determine the coefficient of variation of this estimate?

1) Refer to the coefficient of variation table for CANADA.

2) Because the estimate is a percentage which is based on a subset of the total population (i.e., adults with a household budget), it is necessary to use both the percentage (36.9%) and the numerator portion of the percentage (4,881,629) in determining the coefficient of variation.

3) The numerator, 4,881,629, 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. Similarly, the percentage estimate does not appear as any of the column headings, so it is necessary to use the percentage closest to it, 35.0%.

4) The figure at the intersection of the row and column used, namely 1.9% is the coefficient of variation to be used.

5) So the approximate coefficient of variation of the estimate is 1.9%. The finding that 36.9% of adults with a household budget report that they always stay within their budget can be published with no qualifications.

Canadian Financial Capability Survey, 2009

Approximate Sampling Variability Tables - Canada - All Ages

NUMERATOR OF PERCENTAGE ('000)	ESTIMATED PERCENTAGE													
	0.1%	1.0%	2.0%	5.0%	10.0%	15.0%	20.0%	25.0%	30.0%	35.0%	40.0%	50.0%	70.0%	90.0%
1	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	90.0	52.0
2	#####	#####	#####	#####	#####	#####	#####	#####	97.2	93.7	90.0	82.2	63.7	36.7
3	94.8	94.4	93.9	92.5	90.0	87.5	84.9	82.2	79.4	76.5	73.5	67.1	52.0	30.0
4	82.1	81.8	81.3	80.1	78.0	75.8	73.5	71.2	68.7	66.2	63.7	58.1	45.0	26.0
5	73.5	73.1	72.8	71.6	69.7	67.8	65.7	63.7	61.5	59.3	56.9	52.0	40.3	23.2
6	67.1	66.8	66.4	65.4	63.7	61.9	60.0	58.1	56.1	54.1	52.0	47.4	36.7	21.2
7	62.1	61.8	61.5	60.5	58.9	57.3	55.6	53.8	52.0	50.1	48.1	43.9	34.0	19.6
…	…	…	…	…	…	…	…	…	…	…	…	…	…	…
…	…	…	…	…	…	…	…	…	…	…	…	…	…	…
…	…	…	…	…	…	…	…	…	…	…	…	…	…	…
2,000	*****	******	******	*****	3.5	3.4	3.3	3.2	3.1	3.0	2.8	2.6	2.0	1.2
3,000	*****	******	******	*****	******	2.8	2.7	2.6	2.5	2.4	2.3	2.1	1.6	0.9
4,000	*****	******	******	*****	******	******	2.3	2.3	2.2	2.1	2.0	1.8	1.4	0.8
5,000	*****	******	******	*****	******	******	2.1	2.0	1.9	**1.9**	1.8	1.6	1.3	0.7
6,000	*****	******	******	*****	******	******	*****	1.8	1.8	1.7	1.6	1.5	1.2	0.7
7,000	*****	******	******	*****	******	******	*****	******	1.6	1.6	1.5	1.4	1.1	0.6
8,000	*****	******	******	*****	******	******	*****	******	******	1.5	1.4	1.3	1.0	0.6
9,000	*****	******	******	*****	******	******	*****	******	******	1.4	1.3	1.2	0.9	0.5
10,000	*****	******	******	*****	******	******	*****	******	******	*****	1.3	1.2	0.9	0.5
12,500	*****	******	******	*****	******	******	*****	******	******	*****	******	**1.0**	0.8	0.5
15,000	*****	******	******	*****	******	******	*****	******	******	*****	******	******	0.7	0.4
20,000	*****	******	******	*****	******	******	*****	******	******	*****	******	******	******	0.4

NOTE: for correct usage of these tables please refer to microdata documentation.

Example 3: Estimates of Differences Between Aggregates or Percentages

Suppose that a user estimates that that 2,478,643 / 7,055,176 = 35.1% of women with a household budget report that they always stay within their budget, while 2,402,986 / 6,178,565 = 38.9% of men with a household budget report that they always stay within their budget. How does the user determine the coefficient of variation of the difference between these two estimates?

1) Using the CANADA coefficient of variation table in the same manner as described in Example 2 gives the CV of the estimate for women as 3.0%, and the CV of the estimate for men as 2.8%.

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

3) The coefficient of variation of $\hat{d}$ is given by $\sigma_{\hat{d}} / \hat{d} = 0.015 / (-0.038) = -0.395$

4) So the approximate coefficient of variation of the difference between the estimates is 39.5%. The difference between the estimates is considered unacceptable and Statistics Canada recommends this estimate not be released. However, should the user choose to do so, the estimate should be flagged with the letter F (or some similar identifier) and be accompanied by a warning to caution subsequent users about the high levels of error associated with the estimate.

Example 4: Estimates of Ratios

Suppose that the user estimates that 2,478,643 women with a household budget report that they always stay within their budget, while 2,402,986 men have and stay within their household budget. The user is interested in comparing the estimate of women versus that of men 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 women with a household budget who report that they always stay within their budget.. The denominator of the estimate ($\hat{X}_2$) is the number of men with a household budget who report that they always stay within their budget.

2) Refer to the Person coefficient of variation table for CANADA.

3) The numerator of this ratio estimate is 2,478,643. 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, namely, 3.5%.

4) The denominator of this ratio estimate is 2,402,986. 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, namely, 3.5%

5) So the approximate coefficient of variation of the ratio estimate is given by Rule 4, which is:

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

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

$$\begin{aligned}\alpha_{\hat{R}} &= \sqrt{(0.035)^2 + (0.035)^2} \\ &= \sqrt{0.001225 + 0.001225} \\ &= 0.049\end{aligned}$$

6) The obtained ratio of women versus men with a household budget who report that they always stay within their budget is 2,478,643 / 2,402,986 which is 1.03 (to be rounded according to the rounding guidelines in Section 9.1). The coefficient of variation of this estimate is 4.9%, which makes the estimate releasable with no qualifications.

Canadian Financial Capability Survey, 2009

Approximate Sampling Variability Tables - Canada - All Ages

NUMERATOR OF PERCENTAGE ('000)	ESTIMATED PERCENTAGE													
	0.1%	1.0%	2.0%	5.0%	10.0%	15.0%	20.0%	25.0%	30.0%	35.0%	40.0%	50.0%	70.0%	90.0%
1	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	90.0	52.0
2	#####	#####	#####	#####	#####	#####	#####	#####	97.2	93.7	90.0	82.2	63.7	36.7
3	94.8	94.4	93.9	92.5	90.0	87.5	84.9	82.2	79.4	76.5	73.5	67.1	52.0	30.0
4	82.1	81.8	81.3	80.1	78.0	75.8	73.5	71.2	68.7	66.2	63.7	58.1	45.0	26.0
5	73.5	73.1	72.8	71.6	69.7	67.8	65.7	63.7	61.5	59.3	56.9	52.0	40.3	23.2
6	67.1	66.8	66.4	65.4	63.7	61.9	60.0	58.1	56.1	54.1	52.0	47.4	36.7	21.2
7	62.1	61.8	61.5	60.5	58.9	57.3	55.6	53.8	52.0	50.1	48.1	43.9	34.0	19.6
…	…	…	…	…	…	…	…	…	…	…	…	…	…	…
…	…	…	…	…	…	…	…	…	…	…	…	…	…	…
…	…	…	…	…	…	…	…	…	…	…	…	…	…	…
750	*****	*****	*****	5.8	5.7	5.5	5.4	5.2	5.0	4.8	4.6	4.2	3.3	1.9
1,000	*****	*****	*****	5.1	4.9	4.8	4.6	4.5	4.3	4.2	4.0	3.7	2.8	1.6
1,500	*****	*****	*****	*****	4.0	3.9	3.8	3.7	3.6	3.4	3.3	3.0	2.3	1.3
2,000	*****	*****	*****	*****	**3.5**	3.4	3.3	3.2	3.1	**3.0**	**2.8**	2.6	2.0	1.2
3,000	*****	*****	*****	*****	*****	2.8	2.7	2.6	2.5	2.4	2.3	2.1	1.6	0.9
4,000	*****	*****	*****	*****	*****	*****	2.3	2.3	2.2	2.1	2.0	1.8	1.4	0.8
5,000	*****	*****	*****	*****	*****	*****	2.1	2.0	1.9	1.9	1.8	1.6	1.3	0.7
6,000	*****	*****	*****	*****	*****	*****	*****	1.8	1.8	1.7	1.6	1.5	1.2	0.7
7,000	*****	*****	*****	*****	*****	*****	*****	*****	1.6	1.6	1.5	1.4	1.1	0.6
8,000	*****	*****	*****	*****	*****	*****	*****	*****	*****	1.5	1.4	1.3	1.0	0.6
9,000	*****	*****	*****	*****	*****	*****	*****	*****	*****	1.4	1.3	1.2	0.9	0.5
10,000	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	1.3	1.2	0.9	0.5
12,500	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	1.0	0.8	0.5
15,000	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	0.7	0.4
20,000	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	0.4

NOTE: for correct usage of these tables please refer to microdata documentation.

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

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

> If sampling of the population is repeated indefinitely, each sample leading to a new confidence interval for an estimate, then in 95% of the samples the interval will cover the true population value.
>
> Using the standard error of an estimate, confidence intervals for estimates may be obtained under the assumption that under repeated sampling of the population, the various estimates obtained for a population characteristic are normally distributed about the true population value. Under this assumption, the chances are about 68 out of 100 that the difference between a sample estimate and the true population value would be less than one standard error, about 95 out of 100 that the difference would be less than two standard errors, and about 99 out of 100 that the

differences would be less than three standard errors. These different degrees of confidence are referred to as the confidence levels.

Confidence intervals for an estimate, $\hat{X}$, are generally expressed as two numbers, one below the estimate and one above the estimate, as $\left(\hat{X} - k, \hat{X} + k\right)$ 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 Canadian adults with a household budget who reported that they always stay within their budget (from Example 2, Section 10.1.1) would be calculated as follows:

$\hat{X}$ = 36.9% (or expressed as a proportion 0.369)

t = 2

$\alpha_{\hat{x}}$ = 1.9% (0.019 expressed as a proportion) is the coefficient of variation of this estimate as determined from the tables.

$CI_{\hat{x}}$ = {0.369 – (2) (0.369) (0.019), 0.369 + (2) (0.369) (0.019)}

$CI_{\hat{x}}$ = {0.369 – 0.014, 0.369 + 0.014}

$CI_{\hat{x}}$ = {0.355, 0.383}

With 95% confidence it can be said that between 35.5% and 38.3% of Canadian adults with a household budget report that they always stay within their budget.

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 women with a household budget reporting that they always stay within their budget and the proportion of men with a household budget reporting that they always stay within their budget. From Example 3, Section 10.1.1, the standard error of the difference between these two estimates was found to be 0.015. Hence,

$$t = \frac{\hat{X}_1 - \hat{X}_2}{\sigma_{\hat{d}}} = \frac{0.351 - 0.389}{0.015} = \frac{-0.038}{0.015} = -2.53$$

Since t = -2.53 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.